UPPSALA
UNIVERSITET

U.U.D.M. Project Report 2016:25

Teichmüller space and the mapping class group

Erik Orvehed Hiltunen

Examensarbete i matematik, 15 hp
Handledare: Magnus Jacobsson
Examinator: Jörgen Östensson
Juni 2016

Department of Mathematics
Uppsala University

UPPSALA
UNIVERSITET

Teichmüller space and the mapping class group

Erik Orvehed Hiltunen

10th June, 2016

Degree project C in Mathematics, 1MA079
Supervisor: Magnus Jacobsson

Abstract

In this work, we present part of the classical theory of hyperbolic surfaces. We define and study the Teichmüller space, consisting of all hyperbolic metrics up to isometry isotopic to the identity. In particular, we present the Fenchel-Nielsen parametrization of the Teichmüller space of a closed orientable surface of genus $g \geq 2$. Additionally, we study the mapping class group of a closed orientable surface and present explicit generators for the mapping class group (after Dehn, Lickorish).

Contents

1 Background and fundamental questions **3**

2 The hyperbolic plane **4**
2.1 Isometries on the hyperbolic plane . . . 6
2.2 Geodesics of the hyperbolic plane . . . 11
2.3 Classification of the isometries . . . 12

3 Preliminaries from algebraic topology **14**
3.1 Homotopies and the fundamental group . . . 14
3.2 Covering spaces . . . 15
3.3 The action of the fundamental group on a covering space . . . 17
3.4 Isotopies and free-homotopy classes . . . 20
3.5 Closed surfaces . . . 22

4 Hyperbolic surfaces **24**
4.1 Compact hyperbolic surfaces . . . 24
4.2 Classification of complete hyperbolic manifolds . . . 25
4.3 $\mathbb{H}^2$ as the universal covering of a hyperbolic surface . . . 28

5 Teichmüller space **31**
5.1 The parameters of length . . . 31
5.2 The parameters of twist . . . 35
5.3 The full parametrization of τ_g . . . 36

6 Mapping class group **40**
6.1 Generators for Γ_g . . . 41
6.2 Applications to the Riemann moduli space . . . 45

7 Conclusions **45**

8 References **46**

1 Background and fundamental questions

In the nineteenth century, the field of geometry underwent a dramatic change with the introduction of non-Euclidean geometries. In particular, Lobachevsky and Bolyai developed *Hyperbolic geometry*, and in particular the geometry of the *hyperbolic plane* $\mathbb{H}^2$. In $\mathbb{H}^2$, as opposed to the Euclidean plane, Euclid's parallel postulate does not hold. Instead, given a line L and a point p not on L, there are several lines through p not intersecting L. As an example of consequences, triangles in the hyperbolic plane will always have angle sum less that π. In modern language, $\mathbb{H}^2$ can be seen as a two-dimensional Riemannian manifold of constant negative curvature -1.

The topology of the hyperbolic plane coincides with that of $\mathbb{R}^2$. This raises the question whether surfaces of other topologies can be endowed with metrics which locally behaves as the hyperbolic plane. Hilbert showed in 1901 that no complete, smooth surface embedded in the Euclidean space $\mathbb{E}^3$ can have constant negative curvature. However, examples of non-complete or non-regular surfaces with curvature -1 exist, for example the *pseudosphere* shown in figure 1. Observe that the pseudosphere has singularities. In contrast, there are many examples of surfaces embedded in $\mathbb{E}^3$ which locally behave as the Euclidean plane. As an example, the Euclidean plane can be bent into a cylinder without distorting lengths and angles. Note also that the cylinder is topologically inequivalent to the plane.

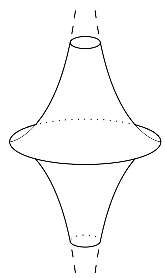


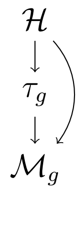
Figure 1: The pseudosphere.

In this work we will consider closed, orientable surfaces, which are characterised by their *genus*. Because of Hilbert's theorem, in order to find hyperbolic metrics we must consider other metrics than those induced by an embedding in $\mathbb{E}^3$. We will see that the only closed orientable surfaces which can have a hyperbolic metric are those of genus $g \geq 2$.

Knowing *which* closed orientable surfaces can be given a hyperbolic metric, the next question is *how many* different hyperbolic metrics can be given a fixed surface M of genus g. More formally, we would like to *parametrize* the set $\mathcal{H}$ of hyperbolic metrics. In other words, we seek a set of parameters which completely describe the different metrics. It is intuitively clear that many of the possible hyperbolic metrics will be in some sense too "similar" to distinguish in a meaningful way. Accordingly, we will not parametrize *all* metrics, but rather the equivalence classes of metrics under a suitable relation. The perhaps most intuitive relation would be that two metrics h_1 and h_2 are related if there is an orientation-preserving isometry $\phi : (M, h_1) \to (M, h_2)$. The set of equivalence classes under this relation is known as the *Riemann moduli space* of M, denoted $\mathcal{M}_g$.

To study $\mathcal{M}_g$, we will begin by studying a different identification of the metrics. We will construct the *Teichmüller space*, τ_g, where we identify metrics under the additional assumption that the isometry is isotopic to the identity. We will parametrize this space, and gain some understanding of its structure.

Using τ_g, we can construct $\mathcal{M}_g$ in a different way. Instead of identifying metrics in $\mathcal{H}$, we will identify metric classes in τ_g. This is done using the action of the *mapping class group* Γ_g on τ_g. The mapping class group is the group of equivalence classes of orientation-preserving diffeomorphisms of M, under the isotopy relation. This way, the study of $\mathcal{M}_g$ reduces to the study of τ_g and Γ_g. The different identifications are illustrated in the following figure.



The problem described above is known as a *moduli problem*. The aim in such a problem is to find a *parameter space* which completely describe the objects of interest. To illustrate the ideas of a moduli problem, we characterize all circles in $\mathbb{R}^2$. We know that a circle is uniquely determined by the center, i.e. a point in $\mathbb{R}^2$, and the radius, i.e. a positive real number. The parameter space in this case is therefore $\mathbb{R}^2 \times \mathbb{R}_+$. Note that the parameters have a *geometrical* interpretation, they describe the location and shape of the circles. Using the parameter space, we can also *compare* the circles, i.e. give a topology on the set of circles. The parameter space $\mathbb{R}^2 \times \mathbb{R}_+$ has a natural topology, which can be induced to the set of circles. All these ideas apply to the Teichmüller space, we will find a parameter space for the equivalence classes of metrics. As in the example, the parameters carry a geometrical interpretation.

Consider now the following moduli problem: to characterize all congruent circles in $\mathbb{R}^2$. In this case the centre point is immaterial. We can thus solve this problem using the solution to the previous problem, namely take the parameter space $\mathbb{R}^2 \times \mathbb{R}_+$ and identify all centre points. We obtain the parameter space $\mathbb{R}_+$ for the set of congruent circles. Observe that we solved this problem by taking objects we already had described, and identifying them in a suitable way. This idea applies to the study of the Riemann moduli space. If we first can solve the moduli problem for τ_g, we can then identify points in τ_g, compare how this identification behaves in the parameter space, and then gain insight in the parameter space of $\mathcal{M}_g$.

Hyperbolic geometry also has numerous applications outside the field of mathematics. For example, in 1910 Varićak found an interpretation of Einstein's law for velocity addition from special relativiy. While ordinary vector addition follow the triangle rule, relativistic velocity addition follow a similar rule, but with hyperbolic triangles. Many biological formations, such as lettuce leafs and coral reefs, can be seen as surfaces with constant negative curvature. Hyperbolic geometry is also found in art, made famous by the artist M. C. Escher.

The aim of this work is to give an introduction to Teichmüller theory at a level accessible to advanced undergraduate students. The necessary prerequisites are basic differential geometry, topology and complex analysis. The material in this work is mostly based on [1]. Section 3 is based on [4], while section 6 is based on [7] and [8].

2 The hyperbolic plane

We now define the hyperbolic plane. For $p, q \in \mathbb{R}^3$, we use the notation $\langle p, q\rangle$ for the bilinear form $\langle p, q\rangle = p_1q_1 + p_2q_2 - p_3q_3$ of signature $(2, 1)$. We remark that the condition $\langle p, p\rangle = -1$ defines a two-sheeted hyperboloid. We will denote the upper sheet, i.e. the sheet with positive third component, by I^2. Furthermore, the tangent plane at a point p is given by all vectors $v \in \mathbb{R}^3$ such that $\langle p, v\rangle = 0$. Indeed, if we parametrize I^2 as $r(x, y) = (x, y, \sqrt{x^2 + y^2 + 1})$, then

$$e_1 = \frac{\partial r}{\partial x} = \left(1, 0, \frac{x}{\sqrt{x^2+y^2+1}}\right)$$
$$e_2 = \frac{\partial r}{\partial y} = \left(0, 1, \frac{y}{\sqrt{x^2+y^2+1}}\right)$$

is a basis for the tangent plane at the point $p = r(x, y)$. It is clear that $\langle p, e_1\rangle = \langle p, e_2\rangle = 0$.

In the basis e_1, e_2, the matrix for the restriction of $\langle\cdot,\cdot\rangle$ to T_pI^2 is given by

$$g = \begin{pmatrix} \langle e_1, e_1\rangle & \langle e_1, e_2\rangle \\ \langle e_1, e_2\rangle & \langle e_2, e_2\rangle \end{pmatrix} = \frac{1}{x^2+y^2+1}\begin{pmatrix} y^2+1 & -xy \\ -xy & x^2+1 \end{pmatrix}$$

Both leading principal minors are positive, so the restriction of $\langle\cdot,\cdot\rangle$ to the tangent plane is positive definite. Hence $\langle\cdot,\cdot\rangle$ defines an inner product on the tangent plane at every point of I^2, so it defines a metric on I^2.

Definition 2.1. The *hyperbolic plane*, denoted $\mathbb{H}^2$, is defined as the set

$$\mathbb{H}^2 = \{x = (x_1, x_2, x_3) \in \mathbb{R}^3 \mid \langle x, x\rangle = -1, x_3 > 0\}$$

together with the metric induced from the restriction of $\langle\cdot,\cdot\rangle$ to the tangent planes.

Remark 2.2. This definition easily extends to more dimensions. To define $\mathbb{H}^n$, take instead the standard bilinear form of signature $(n, 1)$ and proceed as above.

Remark 2.3. We define also the Riemannian manifolds S^n, the *sphere* and $\mathbb{E}^n$, the *Euclidean space*. $\mathbb{E}^n$ is defined as $\mathbb{R}^n$ together with the standard inner product and S^n is defined as the n-sphere together with the metric induced from the inclusion in $\mathbb{E}^{n+1}$.

Remark 2.4. We will next define two global coordinate charts for $\mathbb{H}^2$, called the *disc model* and the *half-plane model* respectively. Because these charts will be globally defined, we could equivalently have defined $\mathbb{H}^2$ as one of these instead. We will call the above definition of $\mathbb{H}^2$ the *hyperboloid model.* When working with intrinsic properties of $\mathbb{H}^2$, we can choose any model to perform computations in. Choosing the best model for our needs will often simplify proofs.

Define $\pi : \mathbb{H}^2 \to \mathbb{R}^2$ by

$$\pi(x_1, x_2, x_3) = \frac{1}{x_3+1}(x_1, x_2),$$

then π describes the stereographic projection with respect to $(0, 0, -1)$ of $\mathbb{H}^2$ onto the open unit disc D^2. Indeed, the line l between a point $(x_1, x_2, x_3) \in \mathbb{H}^2$ and $(0, 0, -1)$ can be parametrized as

$$l : (x, y, z) = (0, 0, -1) + t(x_1, x_2, x_3 + 1), t \in \mathbb{R}.$$

The intersection of l with the plane $z = 0$ is given by $t = \frac{1}{x_3+1}$, and the formula for π follows. This stereographic projection is illustrated in figure 2.

Furthermore, π is a diffeomorphism onto its image. Indeed, π can be extended to $\mathbb{R}^3\backslash\{(0, 0, -1)\}$, where π is obviously smooth, so π is also smooth as a function on $\mathbb{H}^2$. Furthermore, π is invertible, with inverse

$$\pi^{-1}(y_1, y_2) = \frac{2}{1-(y_1^2+y_2^2)}(y_1, y_2, 1 + y_1^2 + y_2^2),$$

which is smooth on D^2.

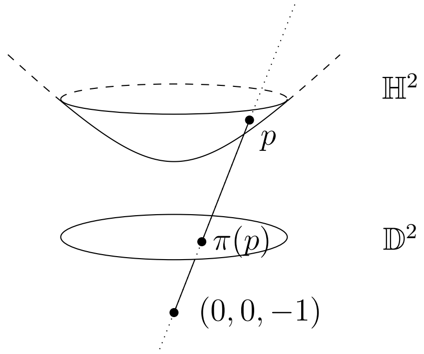


Figure 2: The stereographic projection $\pi : \mathbb{H}^2 \to \mathbb{D}^2$

Definition 2.5. The *disc model* for $\mathbb{H}^2$ is defined as the chart

$$\mathbb{D}^2 = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 < 1\}$$

together with the coordinate map π^{-1}, and the pull-back metric induced from π^{-1}.

Remark 2.6. The pull-back metric is defined using the condition that $\pi : \mathbb{H}^2 \to \mathbb{D}^2$ is an isometry. This way we can consider $\mathbb{D}^2$ not only as a chart for $\mathbb{H}^2$, but as a Riemannian manifold in its own right. Of course this Riemannian manifold will be isometric to $\mathbb{H}^2$

We shall now define a second chart for $\mathbb{H}^2$. Identify $\mathbb{R}^2 \cong \mathbb{C}$ and $\mathbb{D}^2 \subset \mathbb{C}$, and set $\psi : \mathbb{D}^2 \to \{z \in \mathbb{C} \mid \mathrm{Im}(z) > 0\}$,

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

Then ψ is a diffeomorphism between the disc and the upper half plane (as subsets of $\mathbb{R}^2$). Indeed, ψ is the composition of a Möbius transformation with the complex conjugation, and both are diffeomorphisms.

Definition 2.7. The *half-plane model* for $\mathbb{H}^2$ is defined as the chart

$$\mathbb{\Pi}^{+,2} = \{z \in \mathbb{C} \mid \mathrm{Im}(z) > 0\}$$

together with the coordinate map $\pi^{-1} \circ \psi^{-1}$ and corresponding pull-back metric.

Remark 2.8. Often we shall identify $\mathbb{R}^2 \cong \mathbb{C}$ and think of the disc model and the half-plane model as being subsets of $\mathbb{C}$.

2.1 Isometries on the hyperbolic plane

We begin our study of $\mathbb{H}^2$ by describing its isometries. We will do so in all three models defined above. While this may seem like an abstract place to start, knowledge of the isometries will aid our study of more concrete geometrical properties, such as geodesics. We will use the notation $\mathcal{I}(X)$ to denote the group of all isometries of the manifold X onto itself.

The following proposition will be useful.

Proposition 2.9. *Let M, N be Riemannian manifolds of the same dimension, suppose M is connected, and let $\phi_1, \phi_2 : M \to N$ be local isometries onto their images. If $\phi_1(x_0) = \phi_2(x_0)$ and $d_{x_0}\phi_1 = d_{x_0}\phi_2$ for some $x_0 \in M$, then $\phi_1 = \phi_2$.*

Proof. We know that in any Riemannian manifold, locally there is a unique geodesic through any point in any given direction. This can be used to parametrize a neighbourhood U of x_0, the parameters being the direction and the length of the geodesics. U is known as a *normal neighbourhood*. Because both ϕ_1 and ϕ_2 are local isometries, they map the geodesics through x_0 to geodesics through $y := \phi_1(x_0) = \phi_2(x_0)$. Because $d_{x_0}\phi_1 = d_{x_0}\phi_2$, the image of a geodesic in U under ϕ_1 and ϕ_2 will coincide. Hence $\phi_1|_U = \phi_2|_U$. If $\partial U \neq \varnothing$, ϕ_1 and ϕ_2 will coincide on the boundary as well, so we can repeat the argument for a new point $\tilde{x}_0 \in \partial U$. The conclusion is that ϕ_1 and ϕ_2 coincide on all M. $\square$

We begin with the hyperboloid model. Let $v \in \mathbb{R}^3$ be such that $\langle v, v\rangle \neq 0$, and let $p : \mathbb{R}^3 \to \{v\}^\perp$ be the projection to the orthogonal complement $\{v\}^\perp$ (here orthogonality is measured with respect to $\langle \cdot, \cdot \rangle$). Then the mapping $\rho_v : \mathbb{R}^3 \to \mathbb{R}^3, \rho(u) = 2p(u) - u$ is called the *reflection* parallel to v. We note the similarity to the usual notion of reflection, where the bilinear form instead is the standard euclidean inner product.

Observe the following property of the reflection parallel to v: If u is parallel to v then $\rho_v(u) = -u$ and if u is orthogonal to v then $\rho_v(u) = u$.

Let $O(\mathbb{R}^3, \langle \cdot, \cdot \rangle) = \{A \in \mathbb{R}^{3\times 3} \mid \langle Ax, Ay\rangle = \langle x, y\rangle \ \forall x, y \in \mathbb{R}^3\}$ and let $O(I^2)$ be the subgroup of $O(\mathbb{R}^3, \langle \cdot, \cdot \rangle)$ of linear mappings keeping I^2 invariant. $O(I^2)$ is called the *Lorentz group*, and its elements are called *Lorentz transformations*.

The goal will be to prove the following result.

Theorem 2.10.

1. *The group $\mathcal{I}(\mathbb{H}^2)$ consists precisely of the Lorentz transformations restricted to $\mathbb{H}^2$, and the restriction map $R : O(I^2) \xrightarrow{\sim} \mathcal{I}(\mathbb{H}^2)$ is an isometry.*
2. *$\mathcal{I}(\mathbb{H}^2)$ is generated by reflections.*

We begin by establishing part 1, part 2 will be considered later.

Proof of theorem 2.10, part 1. We will show that the map R taking a Lorentz transformation to its restriction to I^2 is an isomorphism $R : O(I^2) \xrightarrow{\sim} \mathcal{I}(\mathbb{H}^2)$. Take $f \in \mathcal{I}(\mathbb{H}^2)$. Recall that the tangent plane at a point x is characterized by the condition $\langle x, u\rangle = 0$, i.e. $T_x\mathbb{H}^2 = \{x\}^\perp$. We can thus define the linear map

$$A : \mathbb{R}^3 = \mathbb{R}x \oplus \{x\}^\perp \to \mathbb{R}^3, \quad ax + v \mapsto af(x) + d_xf(v).$$

Because $\langle x, x\rangle = \langle f(x), f(x)\rangle = -1$, $\langle f(x), d_xf(v)\rangle = 0$ and $\langle d_xf(v), d_xf(v)\rangle = \langle v, v\rangle$ for every $v \in \{x\}^\perp$, it follows that $\langle Au, Av\rangle = \langle u, v\rangle$ for every $u, v \in \mathbb{R}^3$.

Moreover, $f(x) = A(x)$ and $d_xf = d_xA|_{\{x\}^\perp}$. It follows by proposition 2.9 that f and A coincide on $\mathbb{H}^2$. Hence f is the restriction of the Lorentz transformation $A \in O(I^2)$. This shows that the restriction map R is surjective. R is injective because one can find a basis of $\mathbb{R}^3$ in $\mathbb{H}^2$, so the restriction of distinct elements in $O(I^2)$ are distinct in $\mathcal{I}(\mathbb{H}^2)$. It is clear that R is compatible with the group multiplications, so R is an isomorphism. $\square$

We now turn to part 2, the question of generators for $\mathcal{I}(\mathbb{H}^2)$. We begin with a lemma.

Lemma 2.11. *$O(\mathbb{R}^3, \langle \cdot, \cdot \rangle)$ is generated by reflections.*

Proof. We begin by verifying that any reflection ρ parallel to some $v \in \mathbb{R}^3, \langle v, v\rangle \neq 0$ is an element in $O(\mathbb{R}^3, \langle\cdot,\cdot\rangle)$. Pick $x, y \in \mathbb{R}^3$, write $x = x_v + x_{v^\perp}, y = y_v + y_{v^\perp}$. Then $\rho(x) = x_{v^\perp} - x_v, \rho(y) = y_{v^\perp} - y_v$. It follows that $\langle \rho(x), \rho(y)\rangle = \langle x, y\rangle$.

Now, let $A \in O(\mathbb{R}^3, \langle\cdot,\cdot\rangle)$ and let $v = (0, 0, 1)$. Then $\langle v, v\rangle \neq 0$. We have that $-id$ is a composition of reflections, namely the composition of the reflections parallel to the standard basis vectors. We may assume that $\langle Av - v, Av - v\rangle \neq 0$, because if this vanishes then $\langle -Av - v, -Av - v\rangle \neq 0$. Indeed,

$$\langle -Av - v, -Av - v\rangle = 4\langle v, v\rangle - \langle Av - v, Av - v\rangle,$$

so both cannot vanish. Furthermore,

$$v = \frac{1}{2}(Av + v) - \frac{1}{2}(Av - v)$$

and

$$\langle Av + v, Av - v\rangle = 0.$$

Thus, if ρ is the reflection parallel to $Av - v$, then $\rho(v) = Av$. Because ρ is its own inverse it follows that $\rho \circ A(v) = v$ and because $\rho \circ A \in O(\mathbb{R}^3, \langle\cdot,\cdot\rangle)$ it follows that $\rho \circ A|_{\{v\}^\perp} \in O(\{v\}^\perp, \langle\cdot,\cdot\rangle|_{\{v\}^\perp \times \{v\}^\perp})$. But $v = (0, 0, 1)$, so $\{v\}^\perp = \mathbb{R}^2 \times \{0\} \cong \mathbb{R}^2$ and $\langle\cdot,\cdot\rangle|_{\{v\}^\perp \times \{v\}^\perp}$ is simply the standard inner product in $\mathbb{R}^2$. It follows that $O(\{v\}^\perp, \langle\cdot,\cdot\rangle|_{\{v\}^\perp \times \{v\}^\perp}) = O(2)$, where $O(2)$ is the orthogonal group in dimension 2. We know that $O(2)$ is generated by reflections, so $\rho \circ A|_{\{v\}^\perp}$ is a composition of reflections. Extending all reflections to $\mathbb{R}^3$ we find that A is a composition of reflections. □

Proof of theorem 2.10, part 2. We will show that $O(I^2)$ is generated by the reflections it contains, the theorem then follows.

Every reflection ρ is contained in $O(\mathbb{R}^3, \langle\cdot,\cdot\rangle)$, so for $x \in I^2$ we have $\langle \rho(x), \rho(x)\rangle = \langle x, x\rangle = -1$. It follows that ρ keeps the entire two-sheeted hyperboloid $I^2 \cup -I^2$ invariant. Furthermore, if ρ is the reflection parallel to v, then

$$\rho(x) = x - 2\frac{\langle x, v\rangle}{\langle v, v\rangle}v.$$

The third component of the reflection of $x = (0, 0, 1)$ is thus

$$\rho(x)_3 = \frac{v_1^2 + v_2^2 + v_3^2}{\langle v, v\rangle}.$$

It follows that ρ maps $(0, 0, 1)$ to I^2 precisely when $\langle v, v\rangle > 0$, so by continuity, ρ maps I^2 to I^2 precisely when $\langle v, v\rangle > 0$.

Take $A \in O(I^2)$ and write $A = \rho_{x_1} \circ ... \circ \rho_{x_m}$ as a composition of reflections. If $\langle x_i, x_i\rangle < 0$, complete x_i to an orthogonal basis x_i, w_1, w_2 of $\mathbb{R}^3$ with the property that $\langle w_i, w_i\rangle > 0$ for $i = 1, 2$. Then $\rho_{x_i} = -\rho_{w_1} \circ \rho_{w_2}$ (to see this, note that both mappings agree on the basis x_i, w_1, w_2). Substituting into A we find that $A = \pm\rho_{y_1} \circ ... \circ \rho_{y_n}$ where $\langle y_i, y_i\rangle > 0$ for all $i = 1, ..., n$. But $\rho_{y_1} \circ ... \circ \rho_{y_n}$ keeps I^2 invariant, so $-\rho_{y_1} \circ ... \circ \rho_{y_n}$ exchanges the two sheets of $I^2 \cap -I^2$. It follows that $A = \rho_{y_1} \circ ... \circ \rho_{y_n}$, and we have written A as a composition of reflections in $O(I^2)$. □

Remark 2.12. Note that for $x, y \in \mathbb{H}^2$ we can always find an isometry of $\mathbb{H}^2$ mapping $x \mapsto y$. This follows from the above theorem, and the fact that the reflection parallel to $v = x - y$ maps $x \mapsto y$.

We now turn to the disc and the half-plane model. We begin by stating the theorems, and then present the proofs.

Define the set $\mathcal{C}$ as the set consisting of the identity and the conjugation mapping from $\mathbb{C}$ to $\mathbb{C}$.

Theorem 2.13. *The group $\mathcal{I}(\mathbb{D}^2)$ is given by*

$$\mathcal{I}(\mathbb{D}^2) = \left\{ f \circ C : \mathbb{D}^2 \to \mathbb{D}^2 \mid f(z) = e^{i\theta}\frac{z-\alpha}{1-\bar{\alpha}z}, \theta \in \mathbb{R}, \alpha \in D^2, C \in \mathcal{C} \right\},$$

i.e. $\mathcal{I}(\mathbb{D}^2)$ consists of the Möbius transformations keeping D^2 invariant, possibly composed with complex conjugation.

Theorem 2.14. *The group $\mathcal{I}(\mathbb{\Pi}^{+,2})$ is given by*

$$\mathcal{I}(\mathbb{\Pi}^{+,2}) = \left\{ f \circ C : \mathbb{\Pi}^{+,2} \to \mathbb{\Pi}^{+,2} \mid f(z) = \frac{az+b}{cz+d}, a, b, c, d \in \mathbb{R}, ad - bc \neq 0, C \in \mathcal{C} \right\},$$

i.e., $\mathcal{I}(\mathbb{\Pi}^{+,2})$ consists of the Möbius transformations mapping $\mathbb{R}$ to $\mathbb{R}$,composed with complex conjugation if required to keep $\mathbb{\Pi}^{+,2}$ invariant.

We will need theory of complex analysis for the proofs of theorems 2.13 and 2.14. We state the next result without proof, and refer to [2]. We will throughout use the notation D^2 and $\Pi^{+,2}$ for the subsets of $\mathbb{C}$ endowed with the standard euclidean metric, and $\mathbb{D}^2$ and $\mathbb{\Pi}^{+,2}$ for the charts of $\mathbb{H}^2$. We will furthermore denote by $\mathrm{Conf}(M)$ the group of conformal mappings of a Riemannian manifold, and by $\mathrm{Conf}^+(M)$ the orientation-preserving mapping in $\mathrm{Conf}(M)$. For a set $\mathcal{A}$ of mappings of $\mathbb{C}$, we will denote by $c(\mathcal{A}) = \{f(\bar{z}) \mid f \in \mathcal{A}\}$.

Proposition 2.15. *The orientation-preserving conformal mappings of D^2 are given by*

$$\mathrm{Conf}^+(D^2) = \left\{ f : D^2 \to D^2 \mid f(z) = e^{i\theta}\frac{z-\alpha}{1-\bar{\alpha}z}, \theta \in \mathbb{R}, \alpha \in D^2 \right\},$$

and the conformal mappings are given by $\mathrm{Conf}(D^2) = \mathrm{Conf}^+(D^2) \cup c\left(\mathrm{Conf}^+(D^2)\right)$.

Proof of theorem 2.13. We begin by showing that the stereographic projection π is conformal as a map into D^2. Recall that π is given by

$$\pi(x_1, x_2, x_3) = \frac{1}{x_3+1}(x_1, x_2),$$

so parametrizing I^2 as $r(x,y) = (x, y, \sqrt{x^2+y^2+1})$, then

$$\pi(x,y) = \frac{1}{\sqrt{x^2+y^2+1}+1}(x,y)$$

Simple, but rather long, computations show that, in the basis $e_1 = \frac{\partial r}{\partial x}$, $e_2 = \frac{\partial r}{\partial y}$ we have

$$d_{(x,y)}\pi = \frac{1}{\sqrt{x^2+y^2+1}\left(\sqrt{x^2+y^2+1}+1\right)^2} \begin{pmatrix} y^2+1+\sqrt{x^2+y^2+1} & -xy \\ -xy & x^2+1+\sqrt{x^2+y^2+1} \end{pmatrix}.$$

Further computations show that

$$\left(d_{r(x,y)}\pi\right)^{\mathrm{T}} d_{r(x,y)}\pi = \frac{1}{(x^2+y^2+1)\left(\sqrt{x^2+y^2+1}+1\right)^2} \begin{pmatrix} y^2+1 & -xy \\ -xy & x^2+1 \end{pmatrix}$$

Recall from earlier that we computed the metric g in the basis (e_1, e_2). We see that

$$\left(d_{r(x,y)}\pi\right)^{\mathrm{T}} d_{r(x,y)}\pi = \lambda(x,y)g, \qquad \lambda(x,y) = \frac{1}{\left(\sqrt{x^2+y^2+1}+1\right)^2}$$

The conclusion is that for every $u, v \in T_{r(x,y)}$ we have

$$d_{r(x,y)}\pi(u) \bullet d_{r(x,y)}\pi(v) = \lambda(x,y)\langle u, v\rangle,$$

where $\bullet$ denotes the standard scalar product in D^2. Because $\lambda(x,y)$ is positive for every x, y, this shows that π is conformal.

Now, if $\phi \in \mathcal{I}(\mathbb{D}^2)$ then $\pi^{-1} \circ \phi \circ \pi$ is an isometry of $\mathbb{H}^2$. In particular, it is a conformal mapping of $\mathbb{H}^2$. Because π is a conformal mapping from $\mathbb{H}^2$ into D^2, it follows that ϕ is a conformal mapping of D^2. This shows the inclusion $\mathcal{I}(\mathbb{D}^2) \subset \mathrm{Conf}(D^2)$.

We want to show $\mathcal{I}(\mathbb{D}^2) \supset \mathrm{Conf}(D^2)$. Note first that every rotation $z \mapsto e^{i\theta}z$ must be an isometry because of the radial symmetry of $\mathbb{D}^2$. Next, to reach a contradiction we suppose that for some $\alpha_0 \in D^2$, the mapping

$$f_0(z) = \frac{z - \alpha_0}{1 - \bar{\alpha}_0 z}$$

is not an isometry. Then it follows that no mapping

$$f(z) = e^{i\theta}\frac{z - \alpha_0}{1 - \bar{\alpha}_0 z}$$

is an isometry. But these mappings are precisely the mappings in $\mathrm{Conf}(D^2)$ taking $\alpha_0 \mapsto 0$. It follows that there is no isometry taking $\alpha_0 \mapsto 0$. But if we let $x_0 = \pi^{-1}(\alpha_0)$, we know from theorem 2.10 that we can find an isometry ϕ of $\mathbb{H}^2$ taking x_0 to $(0,0,1)$. Then $\pi \circ \phi \circ \pi^{-1}$ is an isometry of $\mathbb{D}^2$ taking α_0 to 0. This is a contradiction, and shows that $\mathrm{Conf}(D^2) \subset \mathcal{I}(\mathbb{D}^2)$. The theorem then follows. □

Proof of theorem 2.14. Note that the mapping ψ used to define $\mathbb{\Pi}^{+,2}$ is the composition of a Möbius transformation with the complex conjugation. Because compositions of Möbius transformations again are Möbius transformations, $\psi \circ \phi$ is a Möbius transformation, possibly composed with the complex conjugation, for every $\phi \in \mathcal{I}(\mathbb{D}^2)$. This shows the inclusion "$\subset$". For the inclusion "$\supset$", note that $\mathcal{I}(\mathbb{D}^2)$ contains all Möbius transformations from $\mathbb{D}^2$ onto $\mathbb{D}^2$, so $\mathcal{I}(\mathbb{\Pi}^{+,2})$ contains all Möbius transformations from $\mathbb{\Pi}^{+,2}$ onto $\mathbb{\Pi}^{+,2}$. □

Remark 2.16. Note that the isometries of $\mathbb{D}^2$ and $\mathbb{\Pi}^{+,2}$ are precisely the conformal mappings of D^2 and $\Pi^{+,2}$. It follows that the notion of angles coincide in the euclidean metric and in the hyperbolic metric. In the future, we will consider angles without specifying which metric we use.

Remark 2.17. The fact that the isometries of $\mathbb{D}^2$ and $\mathbb{\Pi}^{+,2}$ coincide with the conformal mappings of D^2 and $\Pi^{+,2}$ relates hyperbolic geometry to the theory of *Riemann surfaces*. A Riemann surface is a complex manifold of complex dimension 1. In other words, the charts for a Riemann surface are maps to the complex plane. Viewed as a real manifold, a Riemann surface has real dimension 2.

The holomorphic functions with nonvanishing complex derivative are precisely the conformal mappings of a Riemann surface. The Riemann mapping theorem states that any open, simply connected proper subset U of $\mathbb{C}$ can be conformally mapped to the unit disc. By theorem 2.13, this induces a hyperbolic geometry on every such U. The *Uniformization theorem*, which is a generalization of the Riemann mapping theorem, states that every simply connected Riemann surface is conformally equivalent to either the Riemann sphere $\hat{\mathbb{C}}$, the complex plane $\mathbb{C}$ or the unit disc D^2. Riemann surfaces whose universal covering is conformally equivalent to D^2 are known as *hyperbolic*, and are naturally given a hyperbolic geometry.

2.2 Geodesics of the hyperbolic plane

Using the knowledge of the isometries, we can easily describe the geodesics of $\mathbb{H}^2$. We will do so both in $\mathbb{H}^2$ and in its charts $\mathbb{D}^2$ and $\mathbb{\Pi}^{+,2}$.

Proposition 2.18. *In the hyperboloid model, the geodesic at a point $x \in \mathbb{H}^2$ with tangent unit vector $v \in T_x\mathbb{H}^2$ is given by the parametrized curve $\gamma : \mathbb{R} \to \mathbb{H}^2, \gamma(t) = \cosh(t)x + \sinh(t)v$.*

Proof. We begin by verifying that γ is indeed a curve in $\mathbb{H}^2$. Using $\langle x, x\rangle = -1, \langle x, v\rangle = 0$ and $\langle v, v\rangle = 1$ we have that $\langle \gamma(t), \gamma(t)\rangle = -1$. Furthermore, γ is continuous so all points on the curve will lie on the same sheet of the two-sheeted hyperboloid. Because $\gamma(0) = x \in \mathbb{H}^2$, it follows that $\gamma(t) \in \mathbb{H}^2$ for all t. Differentiating we have $\dot{\gamma}(t) = \sinh(t)x + \cosh(t)v$, so $\langle \dot{\gamma}(t), \dot{\gamma}(t)\rangle = 1$, i.e. γ is parametrized by arclength.

Now, we know that $\mathcal{I}(\mathbb{H}^2)$ is generated by reflections. Let $W = \text{span}(x, v) \subset \mathbb{R}^3$, and define $\phi \in O(I^2)$ as $\phi = -\rho_x \circ \rho_v$. Then $\phi|_W = id_W$ and $\phi|_{W^\perp} = -id_{W^\perp}$. If ω is the geodesic starting at x with tangent vector v, it follows that $\phi(\omega)$ passes through x with tangent vector v. But $\phi(\omega)$ must be a geodesic, hence $\phi(\omega) = \omega$ i.e. ω is invariant under ϕ. Since $\phi(x) = x$, ω is even fixed under ϕ. It follows that $\omega \subset W \cap I^2 = \gamma$. □

Proposition 2.19. *In the disc model, the geodesics are precisely the circle segments and the line segments in $\mathbb{R}^2$ which intersect ∂D^2 orthogonally.*

Some geodesics of $\mathbb{D}^2$ are illustrated in figure 3.

Proof. Geometrically, it is clear that any geodesic in $\mathbb{H}^2$ through $(0, 0, 1)$ will map to a line segment in $\mathbb{D}^2$ through the origin under the stereographic projection p. Hence all such lines must be geodesics in $\mathbb{D}^2$. Now, note that by theorem 2.13, the isometries of $\mathbb{D}^2$ are the Möbius transformations preserving the disc, possibly composed with complex conjugation. It is well-known that Möbius transformations carries circles and lines to circles or lines. Furthermore, a Möbius transformation is uniquely determined by its image in three distinct points. Thus, if we take a line segment through the origin, we can find a Möbius transformation which takes the two points on ∂D^2 to any two points on ∂D^2. By proper choice of the third point this Möbius transformation will preserve the disc.

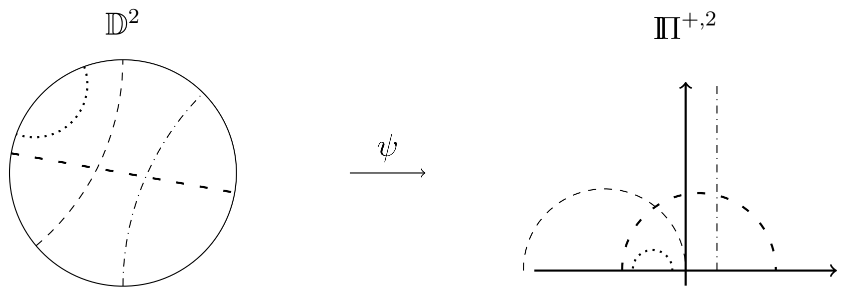


Figure 3: Examples of geodesics in the disc model, and corresponding geodesics in the half-plane model

Furthermore, Möbius transformations are conformal, so the geodesics must intersect ∂D^2 orthogonally.

Finally, note that for any point $x \in D^2$ and any tangent direction v there *is* a circle or line segment through x tangent to v. This shows that the geodesics described in the proposition are all geodesics of $\mathbb{D}^2$. $\square$

Proposition 2.20. *In the upper half-plane model, the geodesics are precisely the vertical line segments and the half circles which intersect* $\mathbb{R}$ *orthogonally.*

Proof. Recall that $\mathbb{\Pi}^{+,2}$ was defined using the isometry $\psi : \mathbb{D}^2 \to \mathbb{\Pi}^{+,2}$, defined as the composition of a Möbius transformation with the complex conjugation. Thus geodesics in $\mathbb{D}^2$ will map to circles or lines in $\mathbb{\Pi}^{+,2}$, intersecting $\mathbb{R}$ orthogonally. $\square$

Remark 2.21. Note that there is a unique geodesic passing through any two points in $\mathbb{H}^2$. To see this, consider the disc model. Then it is clear that any two points defines a circle or line segment intersecting ∂D^2 orthogonally.

2.3 Classification of the isometries

To study the isometries it is convenient to define the *boundary* of $\mathbb{H}^2$, denoted $\partial\mathbb{H}^2$. We will define it in the disc model, and then extend the definition to the hyperboloid and the half-plane models.

Let the boundary $\partial\mathbb{D}^2$ and the closure $\overline{\mathbb{D}^2}$ of the disc model be induced from the inclusion of $\mathbb{D}^2$ as a subset of $\mathbb{R}^2$.

Define now $\partial\mathbb{H}^2 = S^1$ and the space $\overline{\mathbb{H}^2} = \mathbb{H}^2 \cup S^1$, where S^1 is the unit circle. We give $\overline{\mathbb{H}^2}$ a topology using the stereographic projection π. Extend $\pi^{-1} : \mathbb{D}^2 \to \mathbb{H}^2$ to $\overline{\pi}^{-1} : \overline{\mathbb{D}^2} \to \mathbb{H}^2 \cup S^1$ by taking $\overline{\pi}^{-1}|_{S^1} = id_{S^1}$, then let the topology on $\overline{\mathbb{H}^2}$ be defined as the induced topology. We will call points in $\partial\mathbb{H}^2$ *points at infinity.*

To define the boundary of the half-plane model, we extend the map ψ used in the definition of $\mathbb{\Pi}^{+,2}$. Define $\partial\mathbb{\Pi}^{+,2} = \mathbb{R} \cup \{\infty\}$, where ∞ is the formal symbol for a point $\infty \notin \mathbb{C}$. Let $\overline{\psi} : \overline{\mathbb{D}^2} \to \overline{\mathbb{\Pi}^{+,2}}$ be defined as the extension of ψ mapping $-i$ to ∞, and define a topology on $\overline{\mathbb{\Pi}^{+,2}}$ as the topology induced by $\overline{\psi}$. Remark that with this definition, $\partial\mathbb{\Pi}^{+,2}$ is precisely the boundary of $\mathbb{\Pi}^{+,2}$ as a subset of the Riemann sphere $\hat{\mathbb{C}}$.

Intuitively, the boundary of $\mathbb{H}^2$ can be seen as a "circle of infinite radius", and different points on the boundary correspond to "endpoints" of different geodesics in $\mathbb{H}^2$ passing through the origin. This intuitive picture is illustrated in figure 4.

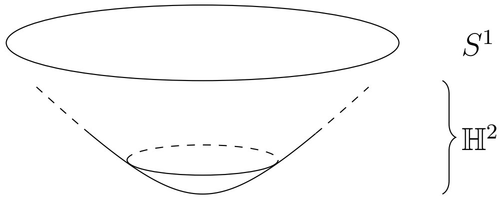


Figure 4: Intuitive picture of the closure of the hyperbolic plane.

Proposition 2.22. *Every isometry on $\mathbb{H}^2$ extends to a homeomorphism of $\overline{\mathbb{H}^2}$, and has a fixed point in $\overline{\mathbb{H}^2}$.*

The proof of the fixed-point property requires the following well-known theorem.

Theorem 2.23 (Brouwer's fixed point theorem for a disc)**.** *If $f : D^2 \to D^2$ is a continuous mapping, then f has a fixed point.*

We omit the proof, and refer to [3].

Proof of proposition 2.22. We show the claim using the disc model. By theorem 2.13, every isometry of $\mathbb{D}^2$ extends to an homeomorphism of $\overline{D^2}$, hence also of $\overline{\mathbb{D}^2}$. It follows from Brouwer's fixed point theorem that every isometry has a fixed point. □

We can now give a classification of the isometries of $\mathbb{H}^2$ using their fixed points in $\overline{\mathbb{H}^2}$.

Proposition 2.24. *Let $f \in \mathcal{I}(\mathbb{H}^2)$. Then either*

1. *f has a fixed point in $\mathbb{H}^2$,*
2. *f has no fixed point in $\mathbb{H}^2$ and precisely one fixed point in $\partial\mathbb{H}^2$,*
3. *f has no fixed point in $\mathbb{H}^2$ and precisely two fixed points in $\partial\mathbb{H}^2$*

Proof. Using proposition 2.22, it remains to show that an isometry with no fixed point in $\mathbb{H}^2$ can have at most two fixed points in $\partial\mathbb{H}^2$. In the half-plane model, take $f(z) \in \mathcal{I}(\Pi^{+,2})$ and assume that $\alpha, \beta \in \partial\Pi^{+,2}$ are fixed points for f. If both $\alpha, \beta \neq \infty$, then defining $h(z) = \frac{z-\alpha}{z-\beta}$ we have that the mapping $g(z) = h \circ f \circ h^{-1}$ has fixed points $0, \infty$. Furthermore, g has no fixed points in $\Pi^{+,2}$, because f does not. By theorem 2.14, h is an isometry, hence also g. It follows that $g(z) = az$ for some $a \in \mathbb{R}\backslash\{1\}$, so g has only $0, \infty$ as fixed points. Hence α, β are the only fixed points of f.

In the case when f has one fixed point $\beta = \infty$, take instead $h = z - \alpha$ and proceed as above. □

We introduce names for the tree cases in proposition 2.24.

Definition 2.25. Let $f \in \mathcal{I}(\mathbb{H}^2)$. In the cases of proposition 2.24, f is called *elliptic* in case 1, *parabolic* in case 2 and *hyperbolic* in case 3.

Remark 2.26. There is an unique geodesic passing through two fixed, distinct points in $\overline{\mathbb{H}^2}$. Indeed, if we consider the disc model, then it is clear that two points in the closed disc will uniquely determine the circle or line segment orthogonal to ∂D^2.

Note also that given a hyperbolic isometry f there is a unique geodesic γ invariant under f. This γ is given by the geodesic through the fixed-points of f in $\partial\mathbb{H}^2$. To see this, note that $f(\gamma)$ will again be a geodesic passing through the same points in the boundary. This uniquely determines the geodesic, so it follows that $f(\gamma) = \gamma$.

We will in the future need the following result concerning the distance between geodesics.

Proposition 2.27. *Two geodesics γ_1, γ_2 with a common point at infinity satisfies*

$$d(\gamma_1, \gamma_2) := \inf \{d(x_1, x_2) \mid x_1 \in \gamma_1, x_2 \in \gamma_2\} = 0$$

In other words, two geodesics intersecting in $\partial\mathbb{H}^2$ come arbitrarily close to each other in $\mathbb{H}^2$.

Proof. In the half-plane model, we let $\alpha \in \partial\mathbb{I}\mathbb{I}^{+,2}$ be the common point at infinity, and $\beta \in \partial\mathbb{I}\mathbb{I}^{+,2}$ the other point at infinity of γ_1. Let $f \in \mathcal{I}(\mathbb{I}\mathbb{I}^{+,2})$ be an isometry keeping γ_1 invariant. As in the proof of proposition 2.24 we consider a conjugate $g = h \circ f \circ h^{-1} = az$ with fixed points $0, \infty$, and choose $a = 2$. Let $\lambda_i = h \circ \gamma_i$ for $i = 1, 2$. Then λ_1 is the imaginary axis, and λ_2 the imaginary axis or a half-circle through the origin. Because g is an isometry, the distance $d(g(\lambda_1), g(\lambda_2)) = d(\lambda_1, \lambda_2)$, and because λ_1 is invariant under g we have $d(g(\lambda_1), g(\lambda_2)) = d(\lambda_1, g(\lambda_2))$. Applying g iteratively we have $d(\lambda_1, \lambda_2) = d(\lambda_1, g^n(\lambda_2))$ for every $n \in \mathbb{N}$. But $g^n(\lambda_2)$ are circles through the origin, with unbounded radius as $n \to \infty$. It is clear that $g(\lambda_2)$ come arbitrarily close to λ_1 for n great enough. Hence $d(\lambda_1, \lambda_2) = d(\gamma_1, \gamma_2) = 0$. □

3 Preliminaries from algebraic topology

Before defining the notion of a hyperbolic surface we will need concepts and results from algebraic topology. In particular, we will need the notion of a *covering space*, and we will define an action of the fundamental group $\pi_1(M)$ of a space M on a covering space of M. We will throughout use the notation I for the unit interval $I = [0, 1] \subset \mathbb{R}$.

3.1 Homotopies and the fundamental group

We begin by recalling a few important definitions and facts which may be already familiar to the reader. Because of this, we will omit proofs of statements in this section.

Definition 3.1. Let X, Y be topological spaces and $f_0, f_1 : Y \to X$ be continuous mappings. A continuous map $H : Y \times I \to X$ is a *homotopy between f_1 and f_2* if $H(\cdot, 0) = f_1$ and $H(\cdot, 1) = f_2$. If $U \subset Y$ and if $H(u, t)$ is constant in t for every $u \in U$, then H is said to be *relative U*.

It is easily verified that "homotopic relative U" forms an equivalence relation on the set of mappings from Y to X.

In the particular case when $Y = I$ in the above definition, both f_0 and f_1 are curves in X. If there is a homotopy from f_0 to f_1 relative $\{0, 1\} \subset I$, we say that f_0 and f_1 are *homotopic relative endpoints*. Denote this equivalence relation by $\sim$. If f_0 is homotopic to the constant path $f_0(0)$ relative endpoints, we say that f_0 is *homotopic to a point*.

Definition 3.2. Let X be a topological space $x_0 \in X$, and let $\mathcal{L} = \{\gamma : I \to X \mid \gamma(0) = \gamma(1) = x_0\}$ be the set of loops in X starting in x_0. Then the *fundamental group of X with base point x_0*, denoted $\pi_1(X, x_0)$ is the set $\mathcal{L}/\sim$, together with the group operation $*$ defined by concatenation of path representatives.

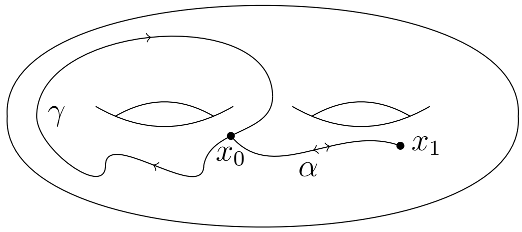


Figure 5: Construction of an isometry between $\pi_1(X, x_0)$ and $\pi_1(X, x_1)$.

Remark 3.3. There are two properties which must be verified for the above definition to make sense. Firstly, the group operation is defined by choosing representatives, so the operation must be independent of this choice. Secondly, the concatenation operation must be a group operation, i.e. it must satisfy the group axioms. Both these properties hold, but the proofs will be omitted here.

Proposition 3.4. *If X is a pathwise connected space and $x_0, x_1 \in X$, then the groups $\pi_1(X, x_0)$ and $\pi_1(X, x_1)$ are isomorphic.*

Remark 3.5. We omit the proof, but describe how such an isomorphism is constructed. Take a curve α from x_0 to x_1, let $[\gamma] \in \pi_1(X, x_0)$ and choose a representative γ of $[\gamma]$. Then the map taking $[\gamma]$ to the class of $\alpha * \gamma * \alpha^{-1}$ in $\pi_1(X, x_1)$, as shown in figure 5, is a well-defined group isomorphism.

Because of proposition 3.4 we often write $\pi_1(X)$ for the fundamental group, without specifying the base point. Note however that the isomorphism is not canonical, in general it does depend on the choice of α.

Definition 3.6. A topological space X is called *simply connected* if X is pathwise connected and $\pi_1(X)$ is trivial, i.e. consists of a single element.

3.2 Covering spaces

Definition 3.7. A map $p : X \to M$ is a *covering map* if X, M are Hausdorff, pathwise connected and locally pathwise connected, and if every $m \in M$ has a neighbourhood $U \subset M$ of m with the property that $p^{-1}(U)$ consists of disjoint sets on which p is a homeomorphism. Such a neighbourhood U is called *elementary*, and X is called a *covering space*. The preimage $p^{-1}(x)$ of a point $x \in X$ is called the *fibre* of x.

Intuitively, small neighbourhoods in a space M are contained, possibly several times, in the covering space X. We will think of the covering space as "lying above" our original space.

Example 3.8. As an example of a covering map, consider $p : \mathbb{R}^2 \to S^1 \times \mathbb{R}$, defined as $(x, y) \mapsto (e^{ix}, y)$. Here S^1 is the unit circle, viewed as a subset of $\mathbb{C}$. This is indeed a covering space. The fibre of $(1, 0) \in S^1 \times \mathbb{R}$ are all points $(2\pi n, 0) \in \mathbb{R}^2$. Moreover, the entire line $(t, 0), t \in \mathbb{R}$ will be mapped to the loop $S^1 \times \{0\}$. This example is illustrated in figure 6.

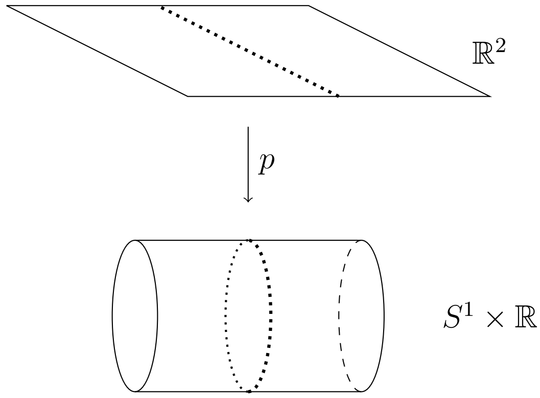


Figure 6: $\mathbb{R}^2$ as a covering space of the cylinder $S^1 \times \mathbb{R}$. A loop in the cylinder is lifted to $\mathbb{R}^2$, where it no longer is a loop.

The above example illustrates many of the coming ideas. We want to *lift* curves or mappings into M, to study them in X instead. For example, if one can find a simply connected covering of M (as in the above example), it is easier to understand properties of curves in M.

Theorem 3.9 (Path lifting theorem)**.** *Let $p : X \to M$ be a covering map, and let $\gamma : I \to M$ be a path in M. Then for every $x_0 \in p^{-1}(\gamma(0))$ there is a unique curve $\tilde{\gamma} : I \to X$ such that $p \circ \tilde{\gamma} = \gamma$ and $\tilde{\gamma}(0) = x_0$.*

In other words, as soon as we fix a lifting of the starting point of a path in M, there is a unique lifting of the entire path.

Proof. Denote by U_m an elementary neighbourhood of $m \in M$. Then $\{\gamma^{-1}(U_m)\}$ is an open cover of $I = [0, 1]$, so there is a finite subcover. Hence there is a $d > 0$ such that any interval in I of diameter $\leq d$ is contained in the finite cover. We define $\tilde{\gamma}$ by induction. By hypothesis, $\tilde{\gamma}(0)$ is defined. If $\tilde{\gamma}$ is defined at a point $x_0 \in X$, then $\gamma([x_0, x_0 + d])$ is contained in an elementary neighbourhood U. On U we can define a unique inverse of p with the property that $p^{-1}(\gamma(x_0)) = \tilde{\gamma}(x_0)$. Hence the definition of $\tilde{\gamma}$ extends uniquely to the entire interval $[x_0, x_0 + d]$. A finite induction then defines $\tilde{\gamma}$ on the whole of I. □

Theorem 3.10 (Homotopy lifting theorem)**.** *Let $p : X \to M$ be a covering map, and let $F : Y \times I \to M$ be a homotopy. If $f : Y \to M$ is a lifting of $F|_{Y\times\{0\}}$ then there is a unique homotopy $G : Y \times I \to X$ such that $p \circ G = F$ and $G|_{Y\times\{0\}} = f$. Moreover, if $F(y, t)$ is relative some $U \subset Y$, also $G(y, t)$ is relative U.*

Similarly as in theorem 3.9, as soon as we fix a lifting of the homotopy at $t = 0$, there is a unique lifting of the entire homotopy.

Proof. We prove the statement by constructing G. For every fixed $y \in Y$, $F|_{\{y\}\times I}$ is a path in M. By theorem 3.9 this can be uniquely lifted to a path in X, starting in $f(y)$. Define $G(y, t)$ as this path evaluated at $t \in I$. It follows that G indeed is a lifting of F which coincide with f on $Y \times \{0\}$. We must show that G is a homotopy, i.e. that G is continuous. Fix $y \in Y$. Then if U_m are the elementary sets in M, $F^{-1}(U_m)$ is an open cover of $\{y\} \times I$, so there is a finite subcover on the form $N_i \times I_i$. Let $N = \cap_i N_i$ and let $d > 0$ be such that any interval of diameter $\leq d$ is contained in some I_i. We proceed by induction. G is continuous on $N \times \{0\}$. If G is continuous on $N \times t_0$ then $F(N, [t_0, t_0 + d])$

is contained in an elementary neighbourhood U, where an inverse of the covering map p is defined. Hence $G|_{N\times[t_0,t_0+d]} = p^{-1} \circ F|_{N\times[t_0,t_0+d]}$ is continuous. A finite induction then shows that G is continuous on the whole of $N \times I$, hence also on $Y \times I$. □

The following corollaries of theorem 3.10 are crucial when defining the action of the fundamental group $\pi_1(M)$ on a covering space X.

Corollary 3.11. *Let γ_1, γ_2 be paths in M which are homotopic relative endpoints. If $\tilde{\gamma}_1$, $\tilde{\gamma}_2$ are liftings with $\tilde{\gamma}_1(0) = \tilde{\gamma}_2(0)$, then $\tilde{\gamma}_1(1) = \tilde{\gamma}_2(1)$*

Proof. Pick a homotopy F from γ_1 to γ_2, and lift F to a homotopy G with $G(t, 0) = \tilde{\gamma}_1(t)$ Then G is a homotopy from $\tilde{\gamma}_1$ to $G(t, 1)$ relative endpoints. Theorem 3.9 then implies that $G(t, 1)$ is the unique lifting of γ_2, i.e. $G(t, 1) = \tilde{\gamma}_2$. □

Corollary 3.12. *Let γ be a loop in M homotopic to a point. Then any lifting $\tilde{\gamma}$ of γ is a loop in X homotopic to a point.*

Proof. By corollary 3.11, any lifting $\tilde{\gamma}$ is homotopic to a lifting of a constant path, which itself is constant. This homotopy fixes endpoints, so $\tilde{\gamma}$ must be a loop. □

We understand that covering spaces are closely linked to the fundamental group. Theorem 3.10 implies that the fundamental group of a covering space cannot be bigger than that of the original space. (More explicitly, the natural group homomorphism $p_\# : \pi_1(X) \to \pi_1(M)$ is injective.) For this reason, simply connected covering spaces are of special interest.

Definition 3.13. A covering space X of M is a *universal cover* if X is simply connected.

Our spaces will be manifolds, and in that case one can show:

Theorem 3.14. *Every manifold M has a universal cover X, and it is unique up to homeomorphism.*

We omit the proof, and refer to [4].

3.3 The action of the fundamental group on a covering space

We can now define the action of $\pi_1(M, m)$ on X, which we will do in steps. First we shall define an action on the fibre $p^{-1}(m)$, using this we shall define another action on the whole space X.

Definition 3.15. For $m \in M$, $x \in p^{-1}(m)$, $\alpha \in \pi_1(M, m)$, we define $x \cdot \alpha$ in the following way. Choose a representative γ of α and pick a lifting $\tilde{\gamma}$ such that $\tilde{\gamma}(0) = x$. Then set $x \cdot \alpha = \tilde{\gamma}(1)$. This is known as the *monodromy action*.

Because γ is a loop, we must have $p(\tilde{\gamma}(0)) = p(\tilde{\gamma}(1))$, i.e. $p(x) = p(x \cdot \alpha)$. Thus the fibre of m is invariant under the action of $\pi_1(M, m)$. A qualitative picture of the monodomy action is given in figure 7.

We verify that definition 3.15 indeed yields a group action:

1. $x \cdot \alpha$ is well-defined, because two different representatives γ_1, γ_2 will lift to curves with the same endpoint according to corollary 3.11.

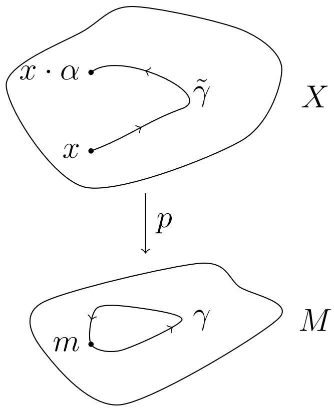


Figure 7: Definition of the monodromy action.

2. $x \cdot 1 = x$ because any loop in M which is homotopic to a point will lift to a loop in X according to corollary 3.12.
3. $(x \cdot \alpha) \cdot \beta = x \cdot (\alpha\beta)$. To see this, choose representatives γ, ρ of α, β, respectively. Lift to curves $\tilde{\gamma}, \tilde{\rho}$ with $\tilde{\gamma}(0) = x$ and $\tilde{\rho}(0) = \tilde{\gamma}(1)$. This can be done because $p(\tilde{\gamma}(1)) = p(x)$. Then $(x \cdot \alpha) \cdot \beta = \tilde{\rho}(1)$ using the definition. But the projection of the concatenation $\tilde{\gamma} * \tilde{\rho}$ is a representative of $\alpha\beta$, so $x \cdot (\alpha\beta) = (\tilde{\gamma} * \tilde{\rho})(1) = \tilde{\rho}(1)$.

Note that if $x \cdot \alpha = x$ for some $x \in p^{-1}$, then $\alpha = 1$. Indeed, in this case some representative γ of α lifts to a loop in X. Because X is a universal covering, this loop is homotopic to a point. If G is such a homotopy, then $p \circ G$ is a homotopy from γ to a point, so $\alpha = 1$.

The intuitive goal is to extend the monodromy action of $\pi_1(M, m)$ from the fibre to the whole space. However, we will see that this is not possible in general. In a sense, the next construction is “as close as you can get” to extending the monodromy action.

Definition 3.16. Let $p : X \to M$ be a covering map. A homeomorphism $D : X \to X$ such that $p \circ D = p$ is called a *deck transformation*. The group of all deck transformations under composition is called the *deck transformation group*, denoted $\Delta(p)$.

Remark 3.17. Note that deck transformations are homeomorphisms keeping every fibre invariant. Also, inverses and compositions of deck transformations are again deck transformations, so $\Delta(p)$ is indeed a group. $\Delta(p)$ is a subgroup of Homeo(X), the group of all homeomorphisms of X onto itself.

Proposition 3.18. *Let $p : X \to M$ be a covering map and $D \in \Delta(p)$. If $D(x) = x$ for some $x \in X$, then $D = id_X$.*

Proof. Note first that if $D(x) = x$ and if U is a neighbourhood of x which projects to an elementary neighbourhood, then $D|_U = id_U$.

Now, let $x_0 \in X$. We want to show $D(x_0) = x_0$. Let α be a path between x and x_0, and let $U_1, ..., U_m$ be an open cover of α such that $p(U_i)$ is elementary for every $i = 1, ..., m$. Then D will be the identity on every U_i, so $D(x_0) = x_0$. □

As a consequence of proposition 3.18 two distinct deck transformations cannot coincide in any point.

Proposition 3.19. *Let $D \in \Delta(p)$, $\alpha \in \pi_1(M, m)$ and $x \in p^{-1}(m)$. Then $(Dx) \cdot \alpha = D(x \cdot \alpha)$.*

Proof. Let γ be a representative of α, and lift γ to $\tilde{\gamma}_1$ starting in x and $\tilde{\gamma}_2$ starting in $D(x)$. On one hand, $\tilde{\gamma}_2$ ends in $(Dx) \cdot \alpha$. On the other hand, we have $\tilde{\gamma}_2 = D(\tilde{\gamma}_1)$, so $\tilde{\gamma}_2$ ends in $D(x \cdot \alpha)$. □

Proposition 3.20. *Let $p : X \to M$ be a universal covering map. Then for every x_1, x_2 in the same fibre there is a $D \in \Delta(p)$ such that $D(x_1) = x_2$.*

Proof. We will construct such a D as follows. Take a path $\tilde{\alpha}_1$ from x_1 to $x \in X$, project to $\alpha = p \circ \tilde{\alpha}_1$. Lift α to $\tilde{\alpha}_2$ starting in x_2, and set $D(x) = \tilde{\alpha}_2(1)$. We must show that D is a deck transformation, i.e. D is independent on choice of $\tilde{\alpha}_1$, D is a homeomorphism and $p \circ D = p$.

$p \circ D = p$. This is clear, because both $\tilde{\alpha}_1$ and $\tilde{\alpha}_2$ are liftings of α, so the endpoints project to the same point.

D is a homeomorphism. Note first that reversing the roles of x_1 and x_2 when constructing D produces D^{-1}. It suffices to show that D is continuous, because D and D^{-1} are created by the same procedure.

Cover the curve $\tilde{\alpha}_2$ by neighbourhoods $U_0, ..., U_m$ such that every U_i projects to an elementary neighbourhood, and such that $D(x) \in U_m$, $x_2 \in U_1$. Let $\{V_i\}_{i=1}^m$ be the covering of $\tilde{\alpha}_1$ which project to the same elementary neighbourhoods. Let $x' \in U_m$, and pick a curve $\tilde{\alpha}_2'$ from x_2 to x' contained in the covering $\{U_1\}_{i=1}^m$. Project to M and lift starting in x_1, then $\tilde{\alpha}_1'$ is contained in the covering $\{V_i\}_{i=1}^m$, so the endpoint x is contained in V_m. This shows that $D^{-1}(U_m) \subset V_m$, and reversing the argument we find that $V_m \subset D^{-1}(U_m)$. The conclusion is that $D^{-1}(U_m)$ is open, so D is continuous.

D is independent of $\tilde{\alpha}_1$. Take another curve $\tilde{\beta}_1$ from x_1 to x. Because X is simply connected, $\tilde{\alpha}_1$ and $\tilde{\beta}_1$ are homotopic relative endpoints, so the projections α and β are homotopic relative endpoints. By corollary 3.11 the liftings $\tilde{\alpha}_2$ and $\tilde{\beta}_2$ have the same endpoints. □

Theorem 3.21. *Let $p : X \to M$ be a universal covering map. Then $\pi_1(M, m)$ and $\Delta(p)$ are isomorphic.*

Proof. Let x_0 be a point in the fibre of m. Define $\Theta : \pi_1(M, m) \to \Delta(p)$ by $\Theta(\alpha) = D_\alpha$, where D_α is the unique deck transformation taking x_0 to $D(x_0) = x_0 \cdot \alpha$. We will show that this is an isomorphism.

Θ is a homomorphism. Indeed, using proposition 3.19, we have $D_{\alpha\beta}(x_0) = x_0 \cdot (\alpha\beta) = (x_0 \cdot \alpha) \cdot (\beta) = D_\alpha(x_0) \cdot (\beta) = D_\alpha \circ D_\beta(x_0)$. Because deck transformations are uniquely determined by its value at a point, $D_{\alpha\beta} = D_\alpha \circ D_\beta$.

Θ is surjective. Any deck transformation D maps x_0 to a point $D(x_0)$. Choose any path from x_0 to $D(x_0)$, project to M to obtain a loop whose homotopy class α satisfies $D = D_\alpha$.

Θ is injective. If $D_\alpha = id$, then $x_0 = x_0 \cdot \alpha$, so any representative γ of α lifts to a loop $\tilde{\gamma}$ in X. Because X is simply connected, $\tilde{\gamma}$ is homotopic to a point. Projecting this homotopy we find that also γ is homotopic to a point. Thus the kernel of Θ is trivial, so Θ is injective. □

Definition 3.22. Let $\Theta : \pi_1(M, m) \to \Delta(p)$, $\alpha \mapsto D_\alpha$ be as above. Then define the action of $\pi_1(M, m)$ on X as the action of $\Delta(p)$ on X.

Remark 3.23. Note that the isomorphism Θ in general depends on the choice of x_0 in the fibre of m. Because of this, we introduce the notion of a *pointed space* (X, x_0), which is a space X along with a choice of x_0. One can show that Θ is independent of this choice precisely when $\pi_1(M, m)$ is abelian. Θ being independent of this choice means precisely that the restriction of D_α coincides with the monodromy action of α on the fibre of m. For the surfaces we will be interested in, the fundamental group is *not* abelian.

We introduce names for some properties of group actions which will be of importance in this work.

Definition 3.24. Let G be a group acting on the space X, and $g \in G$.

1. If for some $x \in X$, $xg = x \Rightarrow g = 1$ then G acts *freely* on X.
2. If every point $x \in X$ has a neighbourhood U such that $Ug \cap U \neq \varnothing \Rightarrow g = 1$ then G acts *properly discontinuously* on X.

We will need the following properties of the action of $\pi_1(M, m)$ on X.

Proposition 3.25. *Let $p : X \to M$ be a universal covering. Then:*

1. *$\pi_1(M, m)$ acts freely on X.*
2. *$\pi_1(M, m)$ acts properly discontinuously on X.*
3. *The quotient map $q : X \to X/\pi_1(M, m)$ is a covering map, and $X/\pi_1(M, m)$ is homeomorphic to M.*

Proof.

1. This is a restatement of proposition 3.18.
2. Every point $x \in X$ has a neighbourhood U which projects to an elementary neighbourhood, and U satisfies the required property.
3. Because $\pi_1(M)$ acts properly discontinuously on X, the images $q(U)$ will be elementary sets in $X/\pi_1(M)$.
   For the second fact, remark that for a point $m \in M$, any two liftings $\tilde{m}_1, \tilde{m}_2$ in X will correspond to the same point in $X/\pi_1(M)$ (i.e. the orbit of an element $x \in X$ is the preimage of $p(x)$). Thus the map $p \circ q^{-1} : X/\pi_1(M) \to M$ is a well-defined bijection, and is continuous with continuous inverse on every elementary set in $X/\pi_1(M)$. Hence $X/\pi_1(M)$ and M are homeomorphic.

□

3.4 Isotopies and free-homotopy classes

We have seen that the fundamental group is independent, up to isomorphism, of the base point. However, this isomorphism is *not canonical*. If we choose different curves between two base points, the corresponding isomorphisms will not coincide. The idea of the next construction is to study loops whose starting points are allowed to vary.

Definition 3.26. Let X be a topological space and $\gamma_1, \gamma_2 : I \to X$ two loops. We call γ_1 and γ_2 *free-homotopic* if there is a homotopy H from γ_1 to γ_2 such that $H(0, s) = H(1, s)$ for every $s \in I$.

Remark 3.27. Note that the homotopy H in this definition is not required to be relative endpoints. This means that the two loops γ_1, γ_2 may have different starting points. However, if we think of H as a "continuous deformation" of γ_1 to γ_2, we still require all intermediate curves to be loops.

"Free homotopy" defines an equivalence relation on the set of loops in X, so we can consider the set of equivalence classes, which we denote by $\tilde{\pi}_1(X)$. In the following result, free-homotopy classes are characterized using the fundamental group.

We will need a notion from general group theory. If G is a group, then two elements $a, b \in G$ are *conjugate* if there is an element $g \in G$ such that $a = gbg^{-1}$. This defines an equivalence relation, and the set of conjugacy classes is denoted by $K(G)$. The conjugacy class of $g \in G$ is denoted by K_g.

Proposition 3.28. *The mapping*

$$\Phi : K(\pi_1(X, x_0)) \to \tilde{\pi}_1(X), \qquad K_{[\gamma]} \mapsto \langle\gamma\rangle$$

is a well-defined bijection.

Proof. Be begin by observing the following fact. If α, is a loop starting in any $x \in X$ and β is a path from x_0 to x, then $\beta * \alpha * \beta^{-1}$ is free-homotopic to α. Indeed, we can create a free-homotopy by contracting along β. This situation is similar to that in figure 5. In particular, this holds if $x = x_0$.

Φ *is well-defined.* Φ is clearly independent of the representative γ of $[\gamma] \in \pi_1(X, x_0)$, because if two loops are homotopic relative endpoints, then they also are free-homotopic. Φ also maps conjugate elements to the same conjugacy class. If $[\lambda][\gamma][\lambda]^{-1}$ is a conjugate of $[\gamma]$, then it follows from above fact that $\lambda * \gamma * \lambda^{-1}$ is free-homotopic to γ. It follows that Φ is well-defined.

Φ *is surjective.* Given any free-homotopy class $\langle\alpha\rangle$, with starting point x, choose a curve β from x_0 to x. From above fact, $\beta * \alpha * \beta^{-1}$ is free-homotopic to α. Furthermore, $\beta * \alpha * \beta^{-1}$ is a loop starting in x_0, so it corresponds to an element in $\pi_1(X, x_0)$.

Φ *is injective.* Suppose $\langle\alpha\rangle = \langle\beta\rangle$ for two loops α, β starting in x_0. This means we can find a free-homotopy $H(t, s)$ from α to β. Let γ be the path traced out by the starting point under this homotopy, i.e. let $\gamma(s) = H(0, s)$. We show that $\gamma * \alpha * \gamma^{-1}$ is homotopic to β relative endpoints.

For a moment, we fix s and introduce some notation. Let $x_s = H(0, s)$, and let γ_s be the path traced out by the starting point up to this fixed s. Let $\lambda_s = H(t, s)$ be the "intermediate loop" starting at x_s and let $\delta_s = \gamma_s * \lambda_s * \gamma_s^{-1}$. Observe that δ_s is a loop starting at x_0.

Now we allow s to vary, and define $\tilde{H}(t, s) = \delta_s(t)$. Then $\tilde{H}$ is a homotopy from $\gamma * \alpha * \gamma^{-1}$ to β relative endpoints. Indeed, at $s = 0$, γ_0 is the constant path and $\lambda_0 = \alpha$, so $\tilde{H}(t, 0) = \alpha(t)$. At $s = 1$, $\gamma_1 = \gamma$ and $\lambda_1 = \beta$, so $\tilde{H}(t, 1) = \gamma * \alpha * \gamma^{-1}$. The claim then follows. □

The following result shows that the bijection Φ can be interpreted using the universal cover X.

Proposition 3.29. *Let $p : (Y, y_0) \to (X, x_0)$ be a universal cover, and let Φ be as above. Let $[\gamma] \in \pi_1(X, x_0)$, and Γ be corresponding deck transformation. Choose $y \in Y$, let $\tilde{\alpha}$ be a curve in Y from y to $\Gamma(y)$ and let $\alpha = p \circ \tilde{\alpha}$. Then $\Phi(K_{[\gamma]} = \langle\alpha\rangle$.*

Proof. We will show that $\langle \alpha \rangle$ is independent of $\tilde{\alpha}$ and y. If this is the case, then choose $y = y_0$ and choose $\tilde{\alpha}$ as the lifting of $[\gamma]$ starting in y_0. Then it follows that $\Phi(K_{[\gamma]}) = \langle \alpha \rangle$.

Φ *is independent of* $\tilde{\alpha}$. Indeed, let $\tilde{\alpha}$, $\tilde{\beta}$ be two curves from y to $\Gamma(y)$. Because Y is simply connected, $\tilde{\alpha}$ and $\tilde{\beta}$ are homotopic relative endpoints, so it follows that $p \circ \tilde{\alpha}$ and $p \circ \tilde{\beta}$ are loops in X homotopic relative endpoints. In particular $p \circ \tilde{\alpha}$ and $p \circ \tilde{\beta}$ are free-homotopic.

Φ *is independent of* y. For $i = 1, 2$ let $y_i \in Y$ and let $\tilde{\alpha}_i$ be a curve from y_i to $\Gamma(y_i)$. Let $\tilde{\beta}_1$ be a path in Y between y_1 and y_2 and $\tilde{\beta}_2 = \Gamma(\tilde{\beta}_2)$ a path between $\Gamma(y_1)$ and $\Gamma(y_2)$. Then $\tilde{\alpha}_1$ is homotopic to $\tilde{\beta}_2^{-1} * \tilde{\alpha}_2 * \tilde{\beta}_1$ relative endpoints. It follows that the loops $p \circ \tilde{\alpha}_1$ and $p \circ \tilde{\beta}_2^{-1} * \tilde{\alpha}_2 * \tilde{\beta}_1$ in X are homotopic relative endpoints. But $p \circ \tilde{\beta}_1 = p \circ \tilde{\beta}_2 =: \beta$, so $p \circ \tilde{\beta}_2^{-1} * \tilde{\alpha}_2 * \tilde{\beta}_1 = \beta^{-1} * \alpha_2 * \beta$. The conclusion is that α_1 is homotopic to $\beta^{-1} * \alpha_2 * \beta$ relative endpoints. Note that β in general will not be a loop, and α_1, α_2 will not have the same starting point. It is clear that $\beta^{-1} * \alpha_2 * \beta$ is free-homotopic to α_2, because we can create a free-homotopy by contracting along β. It follows that α_1 is free-homotopic to α_2. □

Definition 3.30. Let X, Y be topological spaces and $f_0, f_1 : Y \to X$ be continuous mappings. A continuous map $G : Y \times I \to X \times I$ is an *isotopy between* f_1 *and* f_2 if G is a homeomorphism onto its image and if, for every $t \in I$, $H(\cdot, t) = (f_t, t)$ for some $f_t : Y \to X$.

Remark 3.31. Note that two mappings be isotopic is a stronger statement than two mappings be homotopic. Consider for example a non-constant loop in the unit disc $D^2 \subset \mathbb{R}$. Any such loop is homotopic to its starting point d, but it cannot be isotopic to a point. The reason is that no map $G : I \times I \to D^2 \times I$ with $G(s, 1) = d$ can be invertible, hence cannot be an isotopy. Note however that the constant loop is not simple. The next results shows that, in a sense, this is the only example of loops being free-homotopic but not isotopic.

Proposition 3.32. *Let* X *be a pathwise connected space and* γ_1, γ_2 *two simple loops in* X. *Then* γ_1 *and* γ_2 *are isotopic if and only if* γ_1, γ_2 *are free-homotopic.*

Remark 3.33. Note that the only interesting implication is the *if*-part. The *only if*-part is immediate, because any isotopy will be a free-homotopy. The *if*-part was proved by Baer in [5], [6].

3.5 Closed surfaces

We collect some topological properties of compact surfaces, without proofs. Every surface in this section will be closed, connected and oriented.

Theorem 3.34. *Every surface is either diffeomorphic to a the sphere* S^2 *or to a torus* T_g *of some genus* $g \geq 1$.

Remark 3.35. The the surface of genus g, T_g, is the surface with g "holes" called *handles.* The sphere has no handles, so the genus of S^2 is defined as 0.

We have previously studied the fundamental group, so naturally we are interested in the fundamental group of closed surfaces. We begin by recalling a few notions from

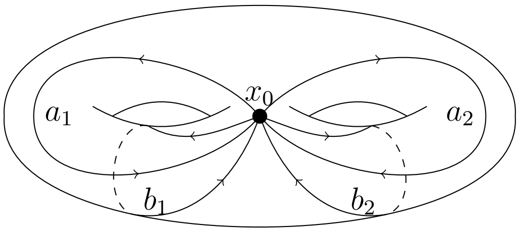


Figure 8: The generators for $\pi_1(T_2, x_0)$.

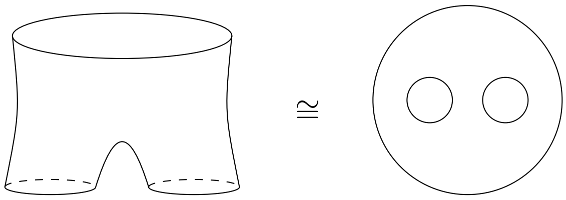


Figure 9: Two pants, identical up to homeomorphism.

group theory. If $w_1, w_2, ..., w_m$ are any symbols, then a *word* is a sequence of symbols $w_{i_1} w_{i_2} ... w_{i_n}$. If R is an equivalence relation on the set of words, then the equivalence classes form a group, denoted $\langle w_1, ..., w_m \mid R \rangle$.

If G is a group, $g, h \in G$, then $[g, h] = ghg^{-1}h^{-1}$ is the *commutator* of g and h.

Theorem 3.36. *Let T_g be a surface of genus g.*

1. *If $g = 0$, then $\pi_1(T_g)$ is trivial.*
2. *If $g > 0$ for every $i = 1, ..., g$ let a_i be an equator and b_i an meridian of the ith handle. Then $\pi_1(T_g) = \langle a_1, b_1, ..., a_g, b_g \mid \prod_{i=1}^{g} [a_i, b_i] = 1 \rangle$.*

As an example of theorem 3.36, the a_i's and b_i's in case $g = 2$ is illustrated in figure 8.

We will be interested in how T_g can be decomposed into parts. The smallest part in our decompositions will be a *pant*, which is a closed disc with two open discs removed, as illustrated in figure 9.

Proposition 3.37. *If M is a surface of genus $g \geq 2$ then there are smooth, pairwise disjoint curves $\alpha_1, ..., \alpha_h$ such that the complement $M \setminus \cup_{i=1}^{h} \alpha_i$ consists of k connected components whose closures are diffeomorphic to pants. The numbers h, k are unique, and given by $h = 3(g-1)$, and $k = 2(g-1)$.*

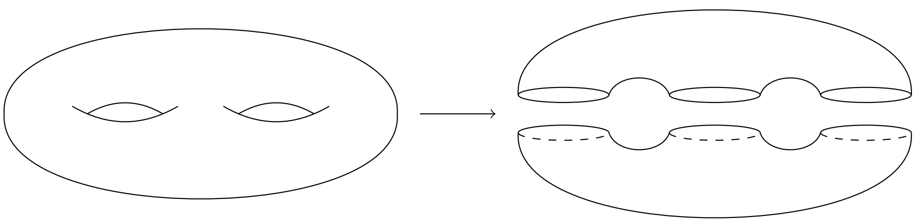


Figure 10: A pant decomposition of T_2.

Remark 3.38. The collection of pants described in proposition 3.37 will be known as a *pant decomposition* of M. Figure 10 shows a pant decomposition of the surface of genus 2.

4 Hyperbolic surfaces

We begin with the definition of a hyperbolic manifold.

Definition 4.1. Let X be a simply connected, oriented n-manifold, and G a group of diffeomorphisms of X. A manifold M has an (X, G)*-structure* if there is an open cover $\{U_i\}$ of M along with differentiable mappings $\phi_i : U_i \to X$ such that

1. $\phi_i : U_i \to \phi_i(U_i)$ is a diffeomorphism
2. Whenever $U_i \cap U_j \neq \varnothing$, the restriction of $\phi_j \circ \phi_i^{-1}$ to connected components of $\phi_i(U_i \cap U_j)$ is the restriction of an element in G.

The collection $\{(U_i, \phi_i)\}$ is called an *atlas* defining the (X, G)-structure and the neighbourhoods U_i are called the *coordinate neighbourhoods.* If $X = \mathbb{H}^n$ and $G = \mathcal{I}(X)$ then a manifold M with an (X, G)-structure is called a *hyperbolic manifold.* If $n = 2$, M is called a *hyperbolic surface.*

Similarily, if $X = S^n$ or $X = \mathbb{E}^n$ and $G = \mathcal{I}(X)$ then a manifold M with an (X, G)-structure is called an *elliptic* or a *flat* manifold, respectively.

4.1 Compact hyperbolic surfaces

We begin our study of hyperbolic surfaces by considering which compact surfaces can be given a hyperbolic structure. The result will be of great importance in the coming sections.

Theorem 4.2. *Let M be a compact, complete surface of genus g. Then*

1. *M can be given an elliptic structure* $\Leftrightarrow g = 0$
2. *M can be given a flat structure* $\Leftrightarrow g = 1$
3. *M can be given a hyperbolic structure* $\Leftrightarrow g \geq 2$

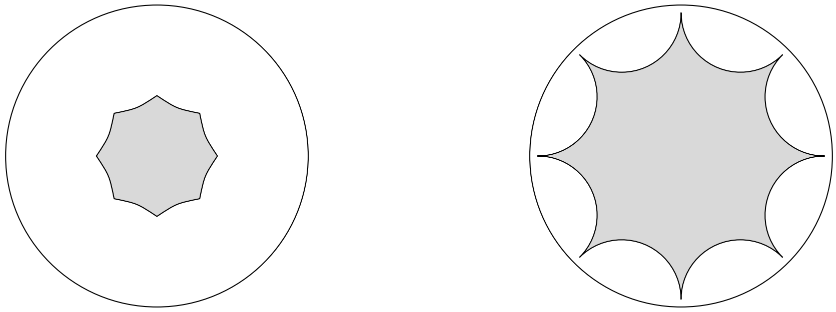
Proof. We begin with the " $\Rightarrow$ "-implications. We will use the Gauss-Bonnet theorem, which for compact surfaces states that

$$\int_M K dS = 2\pi\chi(M),$$

where K is the gaussian curvature and $\chi(M) = 2(1 - g)$ is the Euler characteristic of M. In the three cases of the theorem K constantly equals $1, 0$ or -1, respectively. If $\mathcal{A}(M)$ is the area of M, the Gauss-Bonnet equation states

$$\mathcal{A}(M)K = 4\pi(1 - g),$$

In case 1, the left-hand side is positive, so the only possibility is $g = 0$. In case 2, the left-hand side vanishes, so the only possibility is $g = 1$. In case 3, the left-hand side is negative, so $g \geq 2$.



(a) The angle sum tends to 6π *as* r *tends to* 0.

(b) The angle sum tends to 0 *as* r *tends to* 1.

Figure 11: Illustration of how the angle sum depends on r *in the case of an octagon.*

It remains to show the " $\Leftarrow$ "-implications. *Case 1.* If $g = 0$, then the metric induced by the inclusion of $M = S^2$ in the Euclidian space $\mathbb{E}^3$ gives an elliptic structure.

Case 2. The proof of this fact is highly similar to part 3 of the proof. Take the unit square in $\mathbb{E}^2$ and proceed as below.

Case 3. We recall that a surface of genus g, for $g > 1$, can be obtained by taking a regular $4g$-gon and identifying the sides. Here "identifying" formally means that we choose a suitable equivalence relation, and then consider T_g as the quotient space of the polygon under this relation.

In the disc model, we construct a regular $4g$-gon, i.e. a polygon with geodesic edges with equal lengths and angles. Place the vertices in $re^{\frac{2\pi i}{n}}, n = 0, ..., 4g-1, r > 0$, and let the edges be the unique geodesics between two consecutive vertices. Let $\Theta(r)$ denote the angle sum of the polygon. From the classification of the geodesics and with the aid of figure 11, we see that $\Theta(r) \to 0$ as $r \to 1$ and $\Theta(r) \to \pi(4g-2)$ as $r \to 0$. Θ is a continuous function, so if $g \geq 2$ then for some $0 < r_0 < 1$ we have $\Theta(r_0) = 2\pi$. For this value, we can obtain an atlas for a hyperbolic structure on T_g. We do so explicitly for $g = 2$, the generalization is then clear.

Consider the sets $V_1, .., V_6$ in figure 12. Note that each of these sets can obtained by "cutting" a disc into pieces, and translating the pieces to place. (For this to be possible for V_6 it is crucial that the angle sum is 2π.) If we denote by q the quotient mapping taking the octagon to T_2, then we define the coordinate neighbourhoods in T_2 as $U_i = q(V_i)$. Note that all U_i are connected. Define the diffeomorphisms $\phi_i(u)$ for $u \in U_i$ by choosing a preimage in $q^{-1}(u)$, then translating this point back to the disc. (This is well-defined because the points in the preimage will be mapped to the same point in the disc). In the intersection of two coordinate neighbourhoods U_i, U_j, $\phi_j \circ \phi_i^{-1}$ will be a composition of translations, and therefore an isometry of $\mathbb{H}^2$. We have thus constructed a hyperbolic structure on T_2.

□

4.2 Classification of complete hyperbolic manifolds

We shall describe all complete hyperbolic surfaces M (in fact, all hyperbolic, elliptic or flat manifolds). It will turn out that $\mathbb{H}^2$ is a universal covering of M, and that the covering map is given by the quotient map with respect to some suitable subgroup of $\mathcal{I}(\mathbb{H}^2)$. It

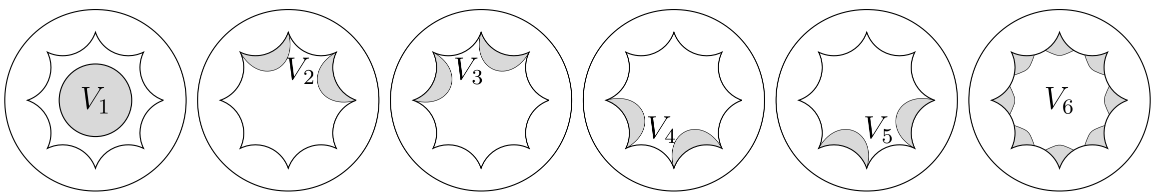


Figure 12: Charts used to define the hyperbolic structure on T_2.

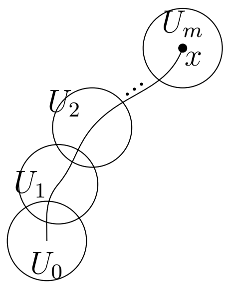
is therefore natural that the cases when M is simply connected or not must be treated separately.

We begin with the case where M is simply connected. We shall need the following proposition.

Proposition 4.3. *Let M be a simply connected space with an (X, G)-structure and let $\phi : U \to X$ be an isometry of the open, connected subset $U \subset M$ onto its image. Then there exists a unique local isometry $D : M \to X$ which extends ϕ.*

The local isometry D in proposition 4.3 is called the *developing function* of M with respect to ϕ.

Proof. We omit the formal proof, and refer to [1] for details. However, we indicate how such a developing function can be constructed. The intuition is to extend ϕ along curves starting in U.



Take a simple curve starting in $x_0 \in U$ and ending in an arbitrary point x. Let U_0 be the intersection of U with a coordinate neighbourhood containing x_0.

Every curve is compact, so it can be covered with a finite number of coordinate neighbourhoods $U_0, U_1, ..., U_m$. Let ϕ_1 be the coordinate function on U_1. We want to extend ϕ to be defined on U_1. We know, from definition 4.1, that $\phi_1 \circ \phi^{-1}$ is the restriction of some element in G, say $g \in G$. Define

$$D|_{U_0} = \phi$$
$$D|_{U_1} = g^{-1} \circ \phi_1$$

This definition matches on the intersection $U_0 \cap U_1$, and both restrictions are isometries onto respective image. Repeat the same process along the entire curve to define $D(x)$.

This way we define the developing function D. It remains to show that D is independent of the choice of covering of the curve, and that D is independent of the curve itself. □

Theorem 4.4. *Let $X = \mathbb{H}^n, \mathbb{R}^n$ or S^n, and let $G \subset \mathcal{I}(X)$. If M is a simply connected, complete (X, G)-manifold, then M is isometric to X.*

Proof. We shall show that any developing function of M is an isometry $D : M \to X$. Because D is a local isometry, it suffices to show that D is a bijection. We will do so by first proving that paths in X can be uniquely lifted to M using D. In other words, given a path γ in $X, \gamma(0) = x \in D(M)$, for a fixed preimage $\tilde{x} \in D^{-1}(x))$ we will show that there exists a unique path $\tilde{\gamma}$ in M such that $\tilde{\gamma}(0) = \tilde{x}$ and $D \circ \tilde{\gamma} = \gamma$.

Because D is a local isometry, such a lifting $\tilde{\gamma}$ must be unique. To show existence, define

$$t_0 = \sup\left\{t \in [0,1] \mid \exists \tilde{\gamma} \in C^1([0,t]), \tilde{\gamma}(0) = \tilde{x}, D \circ \tilde{\gamma} = \gamma\right\},$$

i.e. t_0 is the supremum of the $t \in [0,1]$ such that $\gamma|_{[0,t]}$ can be lifted. Because D is a local isomorphism, some neighbourhood of x is mapped isometrically to a neighbourhood of $\tilde{x}$, so $t_0 > 0$.

We want to show that $t_0 = 1$, and do so by showing that the supremum is a maximum. If this is the case, then because D is an isometry between neighbourhoods of $\tilde{\gamma}(t_0)$ and $\gamma(t_0)$ we must have $t_0 = 1$. The lifting $\tilde{\gamma}(t)$ is well-defined for $t < t_0$. Pick a sequence $\{t_n\}_{n=1}^{\infty}$ converging to t_0, such that $t_n < t_0$ for every $n = 1, 2, \ldots$. We want to show that the sequence $\{\tilde{\gamma}(t_n)\}_{n=1}^{\infty}$ converges, because then $\tilde{\gamma}(t_0)$ is well-defined and the supremum is a maximum. Because M is complete, it suffices to show that $\{\tilde{\gamma}(t_n)\}_{n=1}^{\infty}$ is a Cauchy sequence. To reach a contradiction we suppose not, i.e.

$$\exists \epsilon > 0 \; \forall N > 0 \; \exists m, n > N : d(\tilde{\gamma}(t_m), \tilde{\gamma}(t_n)) > \epsilon.$$

In particular, there is a subsequence $\tilde{\gamma}(t_{n_i})$ such that $d(\tilde{\gamma}(t_{n_i}), \tilde{\gamma}(t_{n_{i+1}})) > \epsilon$ for every i. Then also the length of the arc of $\tilde{\gamma}$ between t_{n_i} and $t_{n_{i+1}}$ is greater than ϵ, so $\tilde{\gamma}$ has infinite length.

But $\tilde{\gamma}$ is covered by neighbourhoods on which D is an isometry, so the length $L(\tilde{\gamma}) = L(\gamma|_{[0,t_0]}) < \infty$. This contradicts the assumption that $\{\tilde{\gamma}(t_n)\}_{n=1}^{\infty}$ is not a Cauchy sequence. Hence $\{\tilde{\gamma}(t_n)\}_{n=1}^{\infty}$ converges. We have thus proven that paths in X can be lifted to M. It follows that homotopies in X can also be lifted to M.

Now, D is surjective, because for any point $x \in X$ there is a curve γ with $\gamma(0) \in D(M), \gamma(1) = x$. This lifts to M, so in particular there is a point which maps to x.

To show injectivity, let $x, x' \in M, D(x) = D(x')$ and take a loop γ starting in $D(x)$. Because X is simply connected, γ is homotopic to a point. By the same logic as in corollary 3.12, any lifting $\tilde{\gamma}$ is a loop, so $x = x'$. □

We now turn to the case when M is not simply connected.

Theorem 4.5. *Let $X = \mathbb{H}^n, \mathbb{R}^n$ or S^n, and let $\mathcal{I}(X)$. If M is a connected, complete (X,G)-manifold, then $\pi_1(M)$ can be identified with a subgroup of $\mathcal{I}(X)$ which acts freely and properly discontinuously on X, and M is isometric to the quotient Riemannian manifold $X/\pi_1(M)$*

Remark 4.6. For a general Riemannian manifold M and a subgroup $G \subset \mathcal{I}(M)$ acting freely and properly discontinuously, the metric on the Riemannian quotient manifold M/G is defined by the condition that the quotient map $q : M \to M/G$ is a local isometry.

Example 4.7. To illustrate this theorem, we return to the construction of a hyperbolic surface of genus 2, given in the proof of theorem 4.2. The universal covering of this surface is the hyperbolic plane, and a lifting of the entire surface will correspond to a tiling of the hyperbolic plane by octagons as shown in figure 13. For any point in T_2, the fibre consists of one point in every octagon. The fundamental group will act as permutations of the octagons. Observe that this action is free and properly discontinuous, and that we indeed regain T_2 as the quotient under this action.

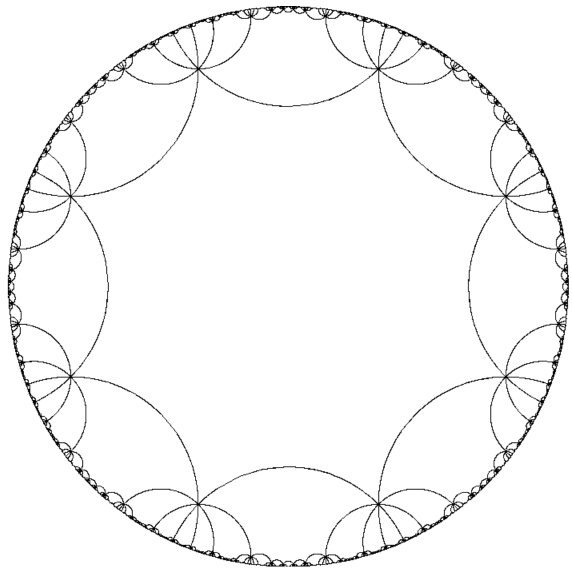


Figure 13: Tiling of the disc model by regular octagons.

Proof of theorem 4.5. Let $\tilde{M}$ be the universal covering of M, and give $\tilde{M}$ an (X,G)-structure by defining the atlas $\{(\tilde{U}_i, \phi_i)\}$ by $\tilde{U}_i = p^{-1}(U_i)$ and $\tilde{\phi}_i = p \circ \phi_i$, where $\{(U_i, \phi_i)\}$ is the atlas defining the (X,G)-structure on M.

We want to apply theorem 4.4 to $\tilde{M}$, so we must show that $\tilde{M}$ is complete. Pick a Cauchy sequence $\{\tilde{x}_i\}_{i=1}^{\infty}$. Then we claim that $\{x_i\}_{i=1}^{\infty}$ is a Cauchy sequence, where $x_i = p(\tilde{x}_i)$. Indeed, the distance $d(\tilde{x}_i, \tilde{x}_j)$ is the length of the shortest curve $\tilde{\gamma}$ between $\tilde{x}_i$ and $\tilde{x}_j$. Let $\gamma = p(\tilde{\gamma})$. Then γ can be covered by neighbourhoods which map isometrically to neighbourhoods of $\tilde{\gamma}$, so γ and $\tilde{\gamma}$ have equal length. Furthermore γ is a curve between x_i and x_j, so $d(\tilde{x}_i, \tilde{x}_j) \geq d(x_i, x_j)$. Consequently, $\{x_i\}_{i=1}^{\infty}$ is a Cauchy sequence, so it converges to some $x \in M$. We will show that $\{\tilde{x}_i\}_{i=1}^{\infty}$ converges to a lifting $\tilde{x}$ of x. Pick an elementary neighbourhood U of x. Then for some N large enough, $x_n \in U$ for all $n > N$. It follows that all $\tilde{x}_n \in p^{-1}(U)$. Moreover, the fibre p^{-1} is discrete, i.e. points in the fibre cannot be arbitrarily close. To see this, note that all points in the fibre have disjoint neighbourhoods isomorphic to U. Because the sequence $\{\tilde{x}_n\}_{n=1}^{\infty}$ is Cauchy, it will come arbitrarily close to the fibre $p^{-1}(x)$. For some $M \geq N$, it follows that $\tilde{x}_n$ lie in the same connected component of $p^{-1}(U)$ for all $n > M$ and because this connected component is homeomorphic to U under p, $\{\tilde{x}_n\}_{n=1}^{\infty}$ will converge to a point in this component. It follows that $\tilde{M}$ is complete.

Now, applying theorem 4.4 we find that $\tilde{M}$ is isometric to X. In the (X,G)-structure defined on $\tilde{M}$, every element in $\pi_1(M)$ will act as an isometry on $\tilde{M}$. Furthermore, by proposition 3.25, $\pi_1(M)$ acts freely and properly discontinuously on $\tilde{M}$. Consequently, $\pi_1(M)$ can be identified with a subgroup of G which acts freely and properly discontinuously on X. Finally, by proposition 3.25 M is homeomorphic to $X/\pi_1(M)$, and this homeomorphism will be an isometry. □

4.3 $\mathbb{H}^2$ as the universal covering of a hyperbolic surface

In this section we will study consequences of theorem 4.5 when applied to hyperbolic surfaces. Theorem 4.5 allows us to view the hyperbolic plane as the universal covering of a hyperbolic surface, so we can apply the topological results in chapter 3 to this setting.

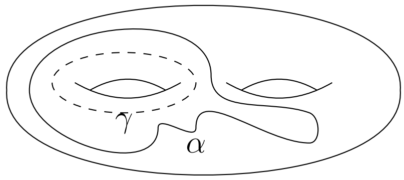


Figure 14: A geodesic loop γ in the same free-homotopy class as α.

Proposition 4.8. *Let M be a connected, complete, compact hyperbolic surface. Then the non-trivial elements of $\pi_1(M)$ acts on $\mathbb{H}^2$ as hyperbolic isometries.*

Recall from proposition 2.24 that an isometry of $\mathbb{H}^2$ is hyperbolic if it has precisely two fixed-points in $\overline{\mathbb{H}^2}$.

Proof. Let f be an isometry of $\mathbb{H}^2$ corresponding to a non-trivial element of $\pi_1(M)$. By theorem 4.5, $\pi_1(M)$ acts freely on $\mathbb{H}^2$, so f can not be elliptic.

To obtain contradiction suppose f is parabolic, and consider the half-plane model (viewed as a subset of $\mathbb{C}$). Similarily as in the proof of proposition 2.24, up to conjugation f has fixed-point ∞, i.e. $f = h \circ g \circ h^{-1}$ for suitable h, where $g(z) = z - b, b \in \mathbb{R}$. Then for the sequence $\{z_n\}_{n=1}^{\infty}$ where $z_n = ni$ we have $d(z_n, g(z_n)) \to 0$ as $n \to \infty$, so by proposition 2.27 also $d(h(z_n), f(h(z_n))) \to 0$ as $n \to \infty$. In general, for any $x \in \mathbb{H}^2$, the distance $d(x, f(x))$ is the length of the geodesic passing through x and $f(x)$. When projecting to M this geodesic will map to a loop which is non-trivial. The conclusion is that the geodesics $\tilde{\gamma}_n$ passing through $h(z_n), f(h(z_n))$ will correspond to non-trivial loops γ_n, such that the length $L(\gamma_n) \to 0$ as $n \to \infty$. But M is compact so the open cover of elementary sets has a finite subcover. Consequently, for large enough n, every γ_n will be completely contained in an elementary set, so γ must be homotopic to a point. This is a contradiction, so the assumption that f be parabolic is incorrect. It follows that f is hyperbolic. □

Proposition 4.9. *Let M be a connected, complete, compact hyperbolic surface. Then any free-homotopy class $\langle\alpha\rangle$ of loops in M contains a unique geodesic loop.*

An example of a loop along with a geodesic loop in the same free-homotopy class is illustrated in figure 14.

Proof. Recall from proposition 3.29 that free-homotopy classes are characterized by conjugacy classes of $\pi_1(X)$. Because $\lambda \in \pi_1(X)$ acts as a hyperbolic isometry on $\mathbb{H}^2$ there is a unique geodesic line $\tilde{\gamma}$ in $\mathbb{H}^2$ invariant under this action. Any conjugate of λ will have the same fixed-points at $\partial\mathbb{H}^2$, so $\tilde{\gamma}$ will also be invariant under the action of any conjugate of λ. Hence any conjugacy class K_λ has a unique invariant geodesic $\tilde{\gamma}$, so the projection γ is the unique geodesic loop in the same class as α. □

Proposition 4.10. *Let M be a connected, complete, compact hyperbolic surface, and γ a non-trivial simple geodesic loop in M. Then two different liftings $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ cannot intersect in $\overline{\mathbb{H}^2}$.*

Proof. It is clear that $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ does not intersect in $\mathbb{H}^2$, because then γ would not be simple.

For contradiction, assume $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ have a common point at infinity. Then we know from proposition 2.27 that $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ come arbitrarily close to each other. It follows that we can find two arbitrarily close points $\tilde{x}_1 \in \tilde{\gamma}_1$ and $\tilde{x}_2 \in \tilde{\gamma}_2$ in the same fibre. But the fibres are discrete, so this is a contradiction. It follows that $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ have no common point at infinity. □

We conclude this section with a proposition which will be of importance when studying the Teichmüller space in the coming section.

Proposition 4.11. *If M is a connected, oriented, compact hyperbolic surface and $\alpha_1, ...\alpha_n$ are non-trivial, pairwise non-intersecting and non-isotopic simple loops in M then there are pairwise non-intersecting and non-isotopic simple geodesic loops $\gamma_1, ..., \gamma_n$ such that γ_i is isotopic to α_i for $i = 1, ..., n$. Furthermore, such $\gamma_1, ...\gamma_n$ are unique.*

Proof. We will show that the γ_i are simple, then by proposition 3.32 it is sufficient if α_i and γ_i are free-homotopic. By proposition 4.9 we define γ_i as the unique geodesic loop free-homotopic to α_i. It remains to show all the required properties for γ_i. Uniqueness is already proven, and the γ_i being pairwise non-isotopic follows by transitivity of the isotopy relation. It remains to show that γ_i are pairwise non-intersecting and simple.

We begin by establishing the following claim. Given $\alpha = \alpha_i$ with a lifting $\tilde{\alpha}$ to $\mathbb{H}^2$, then some lifting $\tilde{\gamma}$ of $\gamma = \gamma_i$ has the same points at infinity as $\tilde{\alpha}$. Conversely, for any lifting $\tilde{\gamma}$ there is a lifting $\tilde{\alpha}$ with the same points at infinity as $\tilde{\gamma}$.

In proposition 4.9 and 3.29, we constructed γ as follows. Given a lifting $\tilde{\alpha}$ there is a unique $\lambda \in \pi_1(X)$ mapping $\tilde{\alpha}(0)$ to $\tilde{\alpha}(1) = \Lambda(\tilde{\alpha}(0))$. Then there is a unique geodesic $\tilde{\gamma}$ invariant under Λ. Now, consider $\tilde{\alpha}$ as defined on $\mathbb{R}$ (this can be done in a natural way; simply define $\tilde{\alpha}|_{[1,2]}$ as the lifting of α starting in $\tilde{\alpha}(1)$, and similarily for any other interval). Iterative application of Λ gives $\tilde{\alpha}(n) = \Lambda^n(\tilde{\alpha}(0))$ and $\tilde{\gamma}(n) = \Lambda^n(\tilde{\gamma}(0))$. Because Λ acts as an isometry,

$$d(\tilde{\alpha}(n), \tilde{\gamma}(n)) = d(\tilde{\alpha}(0), \tilde{\gamma}(0)),$$

so

$$\lim_{n\to\infty} d(\tilde{\alpha}(n), \tilde{\gamma}(n)) < \infty$$

If we consider the disc model, then $\tilde{\gamma}$ will determine a point on the boundary. It follows that the only possibility for $\tilde{\alpha}$ to have bounded distance to $\tilde{\gamma}$ is if $\tilde{\alpha}$ determine the same point on the boundary. We have established our claim.

The γ_i's are pairwise non-intersecting. Indeed, assume γ_i and γ_j intersect at a point $p \in M$. Starting the lift in p, then the lifted curves $\tilde{\gamma}_i$, $\tilde{\gamma}_j$ intersect in $\mathbb{H}^2$. In view of figure 15, we see that $\tilde{\alpha}_j$ must intersect $\tilde{\alpha}_i$. This is a contradiction, and the claim follows.

The γ_i's are simple. Assume γ_i has a self-intersection p. Then the tangent vectors at p are non-parallel, so we can lift γ_i to two different curves $\tilde{\gamma}_1$, $\tilde{\gamma}_2$ intersecting in a point of the fibre of p. Let $\tilde{\alpha}_1$, $\tilde{\alpha}_2$ be liftings of α_i with the same points at infinity. Again by figure 15 we conclude that $\tilde{\alpha}_1$, $\tilde{\alpha}_2$ intersect, which implies that α_i is not simple, which is a contradiction.

□

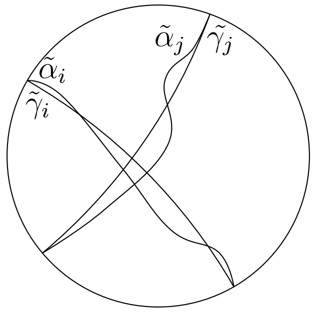


Figure 15: Two pairs of curves in $\mathbb{D}^2$, each pair having the same points in $\partial\mathbb{D}^2$. If two curves, one from each pair, intersects, it follows that the remaining two loops intersects.

5 Teichmüller space

We have now arrived at the main part of this work. The overall question is to describe all possible hyperbolic structures which can be given a compact surface M. We have seen that spheres and 1-tori do not permit even a single hyperbolic structure, so we will only consider the surfaces of genus $g \geq 2$. In this section, all surfaces will be complete, connected, compact and oriented. We will denote the set of all hyperbolic structures on a surface M by $\mathcal{H}$, and we will write (M, h) for the surface M endowed with the structure $h \in \mathcal{H}$.

Definition 5.1. For $h_1, h_2 \in \mathcal{H}$, we define the relation R as follows. $h_1 R h_2$ if there exists an isometry $\phi : (M, h_1) \to (M, h_2)$ which is isotopic to the identity map. Denote the equivalence class of $h \in \mathcal{H}$ by $\langle h \rangle$. We define the *Teichmüller space*, τ_g, as the set $\mathcal{H}/R$.

Remark 5.2. It is clear that R indeed is an equivalence relation, so the definition of τ_g makes sense. We will say "Teichmüller *space*" despite the fact that we have not defined a topology on τ_g yet. For the moment, we consider τ_g simply as a *set*.

We state in words the result we aim to prove. Given a surface M of genus g, we know from proposition 3.37 that M can be decomposed into $2(g-1)$ pants, by "cutting open" along $3(g-1)$ loops. By proposition 4.11 we can choose these loops to be geodesics. The claim is now that a hyperbolic structure, up to equivalence, is completely described by two set of parameters: the length of every such geodesic and the angle with which the pants are "glued together" to form the surface. Accordingly, we will establish a bijection $\tau_g \to \mathbb{R}_+^{3(g-1)} \times \mathbb{R}^{3(g-1)}$, where $\mathbb{R}_+^{3(g-1)}$ are the parameters describing the lengths and $\mathbb{R}^{3(g-1)}$ are the parameters describing the twists. We will see that "twisting" a full turn will not be identical to no "twist" at all, so the twists will indeed range over all $\mathbb{R}$ and not only over the circle S^1. This is known as the Fenchel-Nielsen parametrization of τ_g.

5.1 The parameters of length

We begin by studying the "length"-parameters, as described above. The goal of this section will be to establish a surjection $\tau_g \to \mathbb{R}_+^{3(g-1)}$. In other words, we aim to show that we can find hyperbolic structures giving arbitrary lengths to the geodesic loops in a pant decomposition.

Fix a surface $M = (M, h)$ of genus $g \geq 2$, let $\alpha_1, ..., \alpha_{3(g-1)}$ be the loops in a pant decomposition of M and let $\gamma_1, ..., \gamma_{3(g-1)}$ be the geodesic loops described in proposition

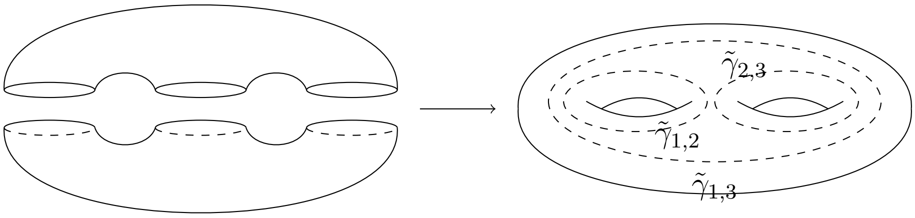


Figure 16: A pant reconstruction of T_2.

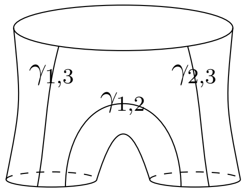


Figure 17: Three geodesic arcs in a pant, splitting the pant into two hexagon with right angles.

4.11. Remark that $\alpha_1, ..., \alpha_{3(g-1)}$ only depend on M, while $\gamma_1, ..., \gamma_{3(g-1)}$ depend on h also. We have seen that a surface can be decomposed into pants. Similarly, pants with geodesic boundary can be joined to form surfaces. Consider two pants P_0, P_1, each with a geodesic boundary component ∂_0 and ∂_1, respectively, of equal length. Then choose points $p_0 \in \partial_0, p_1 \in \partial_1$ and identify the boundaries by identifying p_0, p_1 and then going along the boundaries (in the orientation induced by the pants) and identify points in the boundaries with equal lengths from p_0 and p_1.

We will further decompose our surfaces by decomposing every pant into two hexagons. The next result states that this can be performed using geodesics.

Definition 5.3. A *hyperbolic hexagon* is a hexagon endowed with a hyperbolic structure under which the edges are geodesic segments and the angles are right. A *hyperbolic pant* is a pant endowed with a hyperbolic structure under which the edges are geodesic loops.

Lemma 5.4. *Let $P = (P, k)$ be a hyperbolic pant with boundaries ∂_1, ∂_2 and ∂_3. Then there are unique k-geodesic arcs $\gamma_{i,j}$ from ∂_i to ∂_j. Moreover, these arcs are pairwise disjoint and intersect the boundaries orthogonally.*

Proof. Take two copies of P and "glue" them together as described above, to obtain a surface of genus 2. Then there are three geodesics $\tilde{\gamma}_{i,j}$ each one in the free-homotopy class of an equator, as depicted in figure 16. Then by symmetry $\tilde{\gamma}_{i,j}$ must intersect ∂_i and ∂_j orthogonally. Furthermore, the equators can be chosen to be pairwise disjoint, so by proposition 4.11 also the geodesics in the free-homotopy class are pairwise disjoint. Now, if we restrict $\tilde{\gamma}_{i,j}$ to the pant P, we obtain the geodesic arcs $\gamma_{i,j}$, which shows existence. As for uniqueness, if we are given the geodesic arcs $\gamma_{i,j}$ in P, and glue P to a surface of genus 2 we obtain tree geodesic loops, each one in the free-homotopy class of an equator. These are unique, so it follows that we regain the $\tilde{\gamma}_{i,j}$ as described above. □

Using lemma 5.4 we can cut our pants into hyperbolic hexagons as illustrated in figure 17. We will now parametrize these hexagons under a suitable equivalence relation.

Let E be a hexagon and let $\mathcal{E}$ be the set of all hyperbolic structures on E under which the boundaries are geodesics and the angles are right. Define a relation T on $\mathcal{E}$ by $e_1 T e_2$

if there is an isometry $\phi : (E, e_1) \to (E, e_2)$ isotopic to the identity under an isotopy keeping the boundaries invariant. We introduce names for the boundary segments of E as follows. Pick a vertex, and traverse the boundary segments according to the orientation of E. Then label the first boundary segment a_1, the second b_1, the third a_2, the fourth b_2, the fifth a_3 and the sixth b_3. Then define

$$A : \mathcal{E}/T \to \mathbb{R}^3_+, \quad \langle e \rangle \mapsto \left(L^{(e)}(a_1), L^{(e)}(a_2), L^{(e)}(a_3)\right).$$

Proposition 5.5. *A as defined above is a well-defined bijection.*

Proof. We fix a hexagon E. If $e_1 T e_2$, then there is an isometry $\phi : (E, e_1) \to (E, e_2)$ keeping the edges invariant. It follows that the edges have equal length in both structures, so A is well-defined on $\mathcal{E}/T$. It remains to show that A is a bijection.

A is surjective. We fix three positive numbers $(l_1, l_2, l_3) \in \mathbb{R}^3_+$, and construct a hyperbolic hexagon with the "a"-edges of length l_1, l_2 and l_3.

Fix a point x_1 in $\mathbb{D}^2$, let γ, γ_1 be two geodesics intersecting in x_1 orthogonally. Consider on γ_1 the point x_2 of distance l_1 from x_1, let γ_2 be the geodesic through x_2 intersecting γ_1 orthogonally. This is illustrated in figure 18a.

Choose now $\lambda > 0$, and let $\tilde{x}_1$ be the point along γ of distance λ from x_1. Let $\tilde{\gamma}_1$ be the geodesic intersecting γ orthogonally in $\tilde{x_1}$. Let $\tilde{\gamma}_2$ be the geodesic with a point at infinity in common with $\tilde{\beta}$, intersecting $\tilde{\gamma}_1$ orthogonally in some point $\tilde{x}_2$. This is illustrated in figure 18b.

We have almost constructed a hyperbolic hexagon, it remains to "tie together" γ_2 with $\tilde{\gamma}_2$. If λ tends to 0, $\tilde{\gamma}_1$ will tend to γ_1, so certainly γ_2 and $\tilde{\gamma}_2$ will intersect for small values of λ. At some point $\lambda = \lambda_0$, γ_2 and $\tilde{\gamma}_2$ will have a common point at infinity.

If $0 < \lambda < \lambda_0$, we claim that there can be no geodesic δ intersecting both γ_2 and $\tilde{\gamma}_2$ orthogonally. Indeed, if so is the case, γ_2, $\tilde{\gamma}_2$ and δ form a triangle with geodesic edges and two angles right. But it follows from the Gauss-Bonnet theorem, and the fact that $\mathbb{D}^2$ has negative curvature, that every triangle has angle sum less than π.

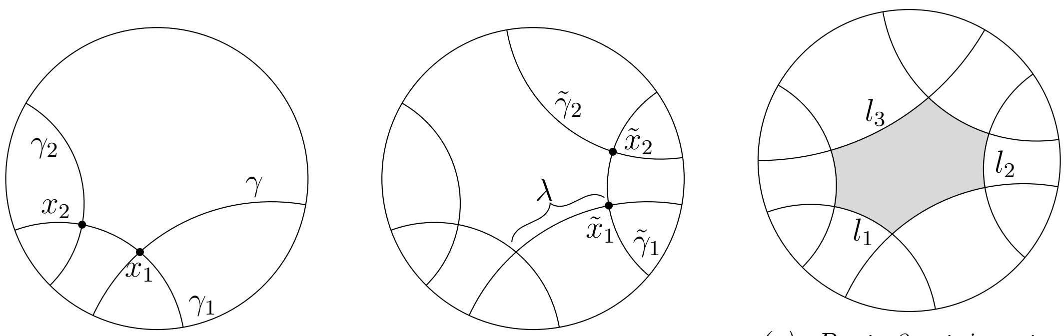
Also in the case $\lambda = \lambda_0$, no such δ can exist. Indeed, consider the half-plane model instead, and suppose γ_2 and $\tilde{\gamma}_2$ have common point ∞ at infinity. It is clear that no geodesic can intersect γ_2 and $\tilde{\gamma}_2$ orthogonally.

In the case $\lambda > \lambda_0$, there is a unique geodesic δ intersecting γ_2 and $\tilde{\gamma}_2$ orthogonally. Denote by $d(\lambda)$ the length of this geodesic. As $\lambda \to \lambda_0$, $d(\lambda) \to 0$ (recall from proposition 2.27 that two geodesics with a common point at infinity come arbitrarily close). As $\lambda \to \infty$, $d(\lambda) \to \infty$. Furthermore, d is a continuous function of λ, so for some $\lambda = \lambda_0$, $d(\lambda_0) = l_3$. This is illustrated in figure 18c.

We have thus constructed a hyperbolic hexagon with all angles right and with tree sides of length l_1, l_2 and l_3. This hexagon can be diffeomorphically mapped to the fixed hexagon E, and we can induce a hyperbolic structure e on E by demanding this diffeomorphism to be an isometry. For this structure we have $A(\langle e \rangle) = (l_1, l_2, l_3)$.

A is injective. Given a hyperbolic hexagon (E, e), we can isometrically embed it in $\mathbb{H}^2$. Indeed, E can be embedded in a hyperbolic surface M, which we know from theorem 4.5 have $\mathbb{H}^2$ as universal covering. Because E is simply connected, we can find a lifting map which lifts E globally to $\mathbb{H}^2$.

We claim that the lengths of the "a"-edges determine uniquely the lengths of the "b"-edges. Indeed, in the previous construction the function $d(\lambda)$ is strictly increasing for



(a) Part 1. *(b) Part 2.* *(c) Part 3, tying together the hexagon.*

Figure 18: Construction of a hyperbolic hexagon with given boundary lengths.

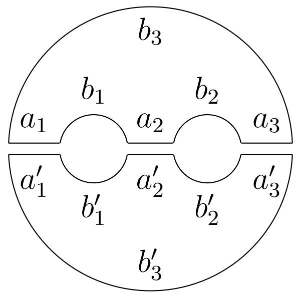


Figure 19: Labelling of the edges of two hexagons obtained from the decomposition of a pant.

$\lambda > 0$. It follows that the choice of λ_0 is unique, so the length of b_1 is unique. Furthermore, the geodesic δ orthogonal to both γ_2 and $\tilde{\gamma}_2$ is unique. Hence also the lengths of b_2, b_3 are unique.

Now, take two hyperbolic structures e_1, e_2 with $A(\langle e_1\rangle) = A(\langle e_2\rangle)$ and embed (E, e_1), (E, e_2) isometrically in the disc $\mathbb{D}^2$. We obtain two hexagons with equal angles and side-lengths. Then there is an isometry $\phi : (E, e_1) \to (E, e_2)$ mapping one hexagon to the other. It follows that $e_1 T e_2$. □

Having parametrized the hexagon, we now return our attention to pants. Let P be a pant with edges ∂_1, ∂_2 and ∂_3, and let $\mathcal{P}$ be the set of all hyperbolic structures on P such that the boundaries of P are geodesics. Define the relation S on $\mathcal{P}$ by $h_1 S h_2$ if there is an isometry $\phi : (P, h_1) \to (P, h_2)$, isotopic to the identity under an isotopy keeping the boundaries invariant. We define the map

$$B : \mathcal{P}/S \to \mathbb{R}^3_+, \quad \langle k\rangle \mapsto \left(L^{(k)}(\partial_1), L^{(k)}(\partial_2), L^{(k)}(\partial_3)\right).$$

Proposition 5.6. *B as defined above is a well-defined bijection.*

Proof. Fix a hyperbolic pant P. If $h_1 S h_2$ then there is an isometry $\phi : (P, h_1) \to (P, h_2)$ keeping the boundaries invariant. In particular, the lengths of the boundaries will be unaltered. It follows that $B(\langle h_1\rangle) = b(\langle h_2\rangle)$. It remains to show that B is a bijection.

By lemma 5.4 we can decompose P into two hyperbolic hexagons. Label the boundaries in these hexagons as in figure 19, then $(L(a_1), L(a_2), L(a_3)) = (L(a_1'), L(a_2'), L(a_3'))$, so it follows from proposition 5.5 that these two hexagons are equivalent under T. In particular $(L(b_1), L(b_2), L(b_3)) = (L(b_1'), L(b_2'), L(b_3') = \frac{1}{2}B(\langle h\rangle)$. But again from proposition 5.5 we

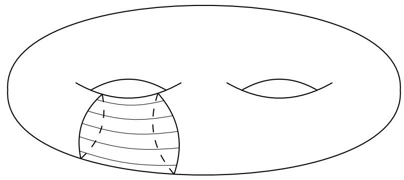


Figure 20: A collar.

know that for any triplet $(l_1, l_2, l_3) \in \mathbb{R}^3_+$ there exists a unique hyperbolic structure $e \in \mathcal{E}$ on a hexagon such that $\left(L^{(e)}(b_1), L^{(e)}(b_2), L^{(e)}(b_3)\right) = \frac{1}{2}(l_1, l_2, l_3)$. It follows that there exist a unique hyperbolic structure h on P such that $B(\langle h \rangle) = (l_1, l_2, l_3)$. $\square$

It now remains to "glue together" pants to form the surface M we are interested in. We define the map

$$L : \tau_g \to \mathbb{R}_+^{3(g-1)}, \quad \langle h \rangle \mapsto \left(L^{(h)}(\gamma_1), ..., L^{(h)}(\gamma_{3(g-1)})\right).$$

Proposition 5.7. *L as defined above is a well-defined surjection.*

Proof. We begin by showing that L is well-defined. If $h_1 R h_2$, then there is an isometry $\phi : (M, h_1) \to (M, h_2)$ isotopic to the identity. Then $\phi(\gamma_i^{(h_1)})$ is a geodesic loop (because ϕ is an isometry) and $\phi(\gamma_i^{(h_1)})$ is isotopic to α_i (because ϕ is isotopic to the identity). Hence $\phi(\gamma_i^{(h_1)}) = \gamma_i^{(h_2)}$, and the lengths of the curves coincide.

To show that L is surjective, let M be a surface and pick any pant decomposition of M using the loops $\alpha_1, ..., \alpha_{3(g-1)}$. Fix $(l_1, ..., l_{3(g-1)}) \in \mathbb{R}_+^{3(g-1)}$, and endow every pant in the decomposition with hyperbolic structures such that for $i = 1, ..., 3(g-1)$ the length of α_i is l_i in both pants which have α_i as boundary. Then identify the boundaries of the pants isometrically. This can be done because any two boundaries which are identified have equal length. We gain a hyperbolic structure h on M such that $L(\langle h \rangle) = (l_1, ..., l_{3(g-1)})$. $\square$

5.2 The parameters of twist

We will now study the "twists" as described earlier. The goal of this section will be to construct a mapping $\Theta_\vartheta : \tau_g \to \tau_g$, depending on the parameter $\vartheta \in \mathbb{R}^{3(g-1)}$, taking a hyperbolic structure and "twists" it with length ϑ_i around the ith geodesic loop.

Θ_ϑ will now be formally defined. Let as before $\gamma_1, ..., \gamma_{3(g-1)}$ be the geodesic loops giving the pant decomposition of the fixed surface (M, h). Fix $1 \leq i \leq 3(g-1)$, and let $x \in \gamma_i$. Then there is a unique geodesic δ_x intersecting γ_i orthogonally in x. If we do this for every $x \in \gamma_i$, and choose $\epsilon > 0$ small enough then

$$r : \gamma_i \times [0, \epsilon] \to M, \quad r(x, t) = \delta_x(t)$$

is a parametrization of the "tube" $C_i = r(\gamma_i \times [0, \epsilon])$ having γ_i as one boundary. C_i will be known as a *collar*. An example of a collar is depicted in figure 20. As always, the geodesics are parametrized by arclength. Intuitively, x describes the "γ_i"-direction and t describes the "δ"-direction of the collar.

On every collar, we can find a diffeomorphism ϕ_i as indicated in figure 21. ϕ_i act as a twist, twisting the entire part closest to γ_i by length ϑ_i and acting as identity on the

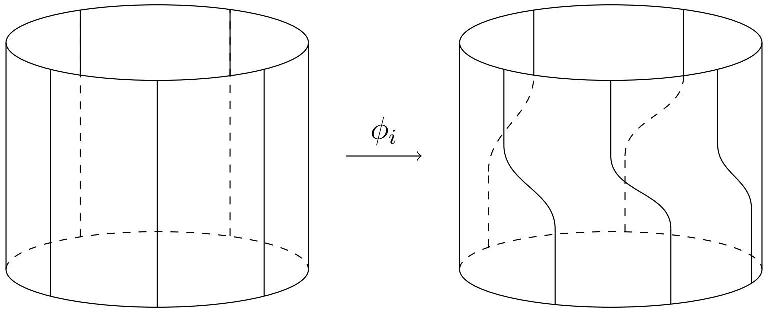


Figure 21: Definition of the diffeomorphism ϕ_i on a collar.

part closest to the other boundary. Using ϕ_i, we can define a new metric $\phi_i^*(h)$ on the collar C_i by demanding ϕ_i to be an isometry. The restrictions of ϕ_i to neighbourhoods of the boundaries of C_i are isometries onto respective image, so we can define a "perturbed" metric h_i' on the entire surface (M, h) by

$$h_i' = \begin{cases} h & \text{outside } C_i \\ \phi_i^*(h) & \text{in } C_i \end{cases}$$

Note that h_i' only differs from h in the collar C_i. If we repeat this construction for all $i = 1, ..., 3(g-1)$, choosing every collar to be pairwise disjoint, we then obtain a new metric on M. Define the equivalence class of this metric to be $\Theta_\vartheta(\langle h\rangle)$. To claim that Θ_ϑ is a well-defined map, we must show that it is independent of all the choices made.

Θ_ϑ is independent of every ϕ_i, because two different diffeomorphisms satisfying the hypotheses are isotopic to the identity. The metrics induced thus belong to the same equivalence class.

It is clear that Θ_ϑ is independent of ϵ when constructing the collars C_i, because different sizes of the collars does not affect the equivalence class of the constructed metric.

Finally, if h_1 and h_2 are two different representatives for $\langle h\rangle$, then there is an isometry $\Phi : (M, h_1) \to (M, h_2)$ isotopic to the identity. If C_i are the collars used when constructing $\Theta_\vartheta(\langle h_1\rangle)$, then $\Phi(C_i)$ and $\Phi \circ \phi_i$ are collars and diffeomorphisms which can be used when constructing $\Theta_\vartheta(\langle h_2\rangle)$. Consequently, Φ is an isometry $\Phi : (M, \Theta_\vartheta(\langle h_1\rangle)) \to (M, \Theta_\vartheta\langle h_2\rangle)$.

Remark 5.8. It is clear that every γ_i is still a geodesic in the new metric $\Theta_\vartheta(\langle h\rangle)$, with the same length. If we recall the map L defined in the previous section, this fact means precisely that Θ_ϑ keep the fibres of L invariant.

Remark 5.9. To simplify notation, we will write $\Theta_\vartheta(h)$, where h is a hyperbolic metric on M. With this we mean $\Theta_\vartheta(\langle h\rangle)$, where $\langle h\rangle$ is corresponding equivalence class. Also, we will sometimes consider $\Theta_\vartheta(h)$ as a metric, then we mean a representative for the class of $\Theta_\vartheta(h)$.

5.3 The full parametrization of τ_g

We are now ready to state the main theorem of this work. Let M, L and Θ_ϑ be as before, and $\sigma : \mathbb{R}_+^{3(g-1)} \to \tau_g$ be an arbitrary mapping such that $L \circ \sigma = id_{\mathbb{R}_+^{3(g-1)}}$. Note that such a mapping σ must be injective.

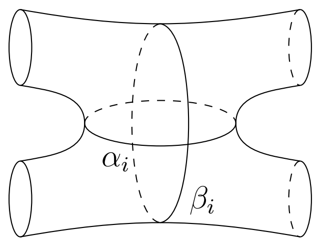


Figure 22: The curve β_i studied in lemma 5.12.

Theorem 5.10. *The mapping*

$$\Psi : \mathbb{R}_+^{3(g-1)} \times \mathbb{R}^{3(g-1)} \to \tau_g, \qquad (l, \vartheta) \mapsto \Theta_\vartheta(\sigma(l))$$

is bijective.

Remark 5.11. Before proving this result, we pause to reflect on the statement, and specifically on the significance of σ. For a fixed metric h on our surface, the theorem claims that given *any* metric h_0 with equal lengths of the geodesic loops in the pant decomposition, we can obtain h by twisting h_0. But the actual value ϑ of the twists may depend on the choice of h_0, i.e. the choice of the function σ.

This can be phrased as follows: Given a metric h, we can find the parameters of h as follows. Begin by measuring the lengths l of the geodesic loops in the pant decomposition. Then *compare* the metric h with a *reference metric*, defined by $\sigma(l)$. To find the twists ϑ, we measure how much h differs from the reference metric.

For the proof we will need the following lemma. For a non-trivial loop β in M, denote by $\tilde{\Lambda}^h(\beta)$ the length, in the metric h, of the h-geodesic loop free-homotopic to β. For a loop α_i in the pant decomposition of M, we denote by β_i a fixed simple, non-trivial loop intersecting α which is not free-homotopic to α_i and is contained in the two pants having α_i as boundary. β_i is illustrated in figure 22.

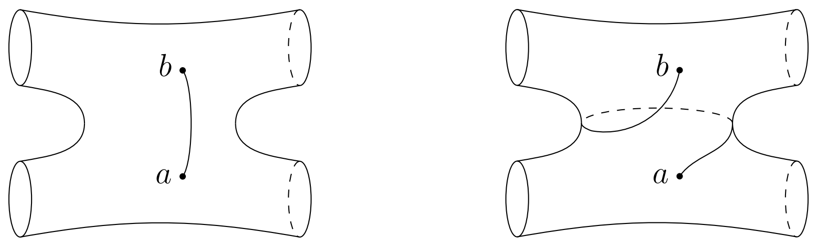
Lemma 5.12. *For every $i = 1, ..., 3(g-1)$ and for every hyperbolic metric h on M, the function*

$$\Lambda_i : \mathbb{R} \to \mathbb{R}, \qquad \Lambda_i(\vartheta_i) = \tilde{\Lambda}^{\Theta_{(0,..,\vartheta_i,...,0)}(h)}(\beta_i)$$

is strictly increasing for large enough arguments.

Remark 5.13. Note that in this lemma, β_i is held fixed and is the same regardless of ϑ_i. However, for different ϑ_i we obtain different geodesic loops free-homotopic to β. The function in the lemma thus varies both the loop we measure the length of, and the metric with which we measure.

Remark 5.14. Before of presenting the proof we discuss the underlying ideas of this lemma. Consider the curve β_i in some metric h. Then let δ_i be the geodesic loop in the same free-homotopy class of β_i, and a, b two points in δ_i. Then one of the arc segments from a to b along δ_i will be the shortest path d from a to b. If we twist the metric h a full turn around α_i, we obtain the metric $\Theta_\vartheta(h)$, where $\vartheta = (0, ..., 0, L_i, 0, ..., 0)$ and L_i is the length of the geodesic γ_i free-homotopic to α_i. The metric $\Theta_\vartheta(h)$ is induced by the requirement that the twist-diffeomorphism ϕ_i is an isometry of the collar. It follows that the shortest path $\tilde{d}$ between a and b in $\Theta_\vartheta(h)$ is the image of d under ϕ_i. Informally we



(a) Shortest path between a and b in some metric h.

(b) Shortest path between a and b in $\Theta_{(0,\dots,L_i,\dots,0)}(h)$.

Figure 23: Illustration of how the shortest path between two points vary under the twist operation.

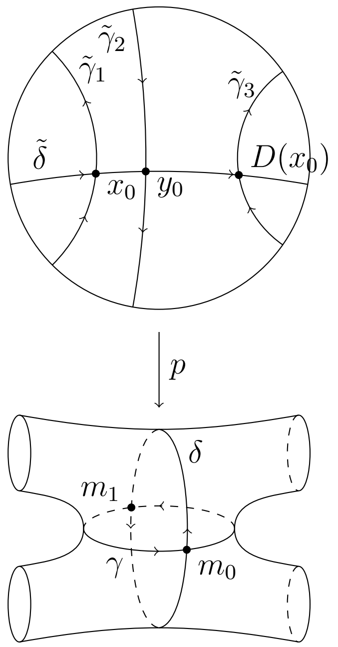


Figure 24: Liftings of geodesic loops to $\mathbb{H}^2$.

can phrase this as "the shortest way across the collar is by going around it once". The situation is illustrated in figure 23.

Now, consider the entire geodesic loop δ_i, and its image $\tilde{\delta}_i$ under ϕ_i. Then δ_i and $\tilde{\delta}_i$ have the same lengths in corresponding metric. *But these two loops are not free-homotopic.* Any loop free-homotopic to the fixed loop β_i will, according to above informal argument, have greater length (in $\Theta_\vartheta(h)$) than δ_i has in h.

Proof. We outline the proof and omit the details. The full proof is given in [1].

Let $\Lambda := \Lambda_i$, $\alpha := \alpha_i$ and $\beta := \beta_i$. The loops are illustrated in figure 22. Let γ be the geodesic loop in the same free-homotopy class as α and let δ be the geodesic loop in the same class as β. Then γ and δ intersect in (at least) two points m_0, m_1. Fix a point x_0 in the fibre of m_0. Let $D \in \mathcal{I}(\mathbb{H}^2)$ be the isometry corresponding to $[\delta] \in \pi_1(M, m_0)$ (in light of remark 3.23, we choose x_0 as base point for $\mathbb{H}^2$). We define three liftings $\tilde{\gamma}_1$, $\tilde{\gamma}_2$ and $\tilde{\gamma}_3$ of γ to $\mathbb{H}^2$ in the following way. Let $\tilde{\gamma}_1$ be the lifting starting in x_0. Let $\tilde{\delta}$ be the lifting of δ starting in x_0, then $\tilde{\delta}$ contains a point y_0 in the fibre of m_1. Let $\tilde{\gamma}_2$ be the lifting of γ starting in y_0. Finally, let $\tilde{\gamma}_3$ be the lifting of γ starting in $D(x_0)$. All these liftings can be naturally extended to maximal geodesics in $\mathbb{H}^2$, as illustrated in figure 24.

Now, let C be a collar with γ as one boundary component. Then we can lift C to

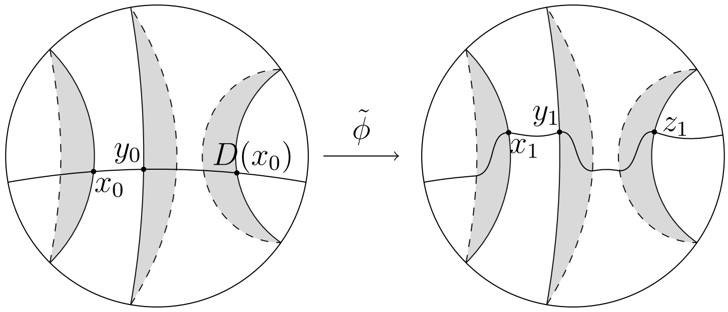


Figure 25: The lifted collars, and the action of $\tilde{\phi}$ on the lifted collars.

three "strips" as indicated in figure 25. Analogously to how we defined the twisted metric $\Theta_\vartheta(h)$ on the surface M, we can define a "twisted" metric on $\mathbb{H}^2$ (in every lifting of the collar, we induce a new metric as the pull-forward of the map $\tilde{\phi}$ which translates points the distance ϑ along the collar). Denote by $\mathbb{H}^2_*$ the hyperbolic plane endowed with this metric, and by d^* corresponding distance function. If x_1, y_1 and z_1 are the images of x_0, y_0 and $D(x_0)$ under the twist map, then $\Lambda(\vartheta) = d^*(x_1, z_1)$. This is illustrated in figure 25.

Now, observe that $z_1 = D(x_1)$, because D maps $\tilde{\gamma}_1$ to $\tilde{\gamma}_3$. Furthermore, we can easily describe the shortest path from x_1 to $D(x_1)$. Inside the collars, it will be the image of $\tilde{\delta}$ under the translation. Outside the collars, the metrics of $\mathbb{H}^2$ and $\mathbb{H}^2_*$ coincide, so the shortest path will run along a geodesic of $\mathbb{H}^2$. This is illustrated in figure 25. It follows that $d^*(x_1, D(x_1)) = d^*(x_1, y_1) + d^*(y_1, D(x_1))$, and because $\mathbb{H}^2$ and $\mathbb{H}^2_*$ coincide outside the collars we have $d^*(x_1, y_1) = d(x_1, y_1)$.

Let x and y be any points on $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$, respectively. Then $d(x_1, y_1) = \inf\{d(x, y)\}$ and $d^*(y_1, D(x_1)) = \inf\{d^*(y, D(x))\}$. Furthermore, if $y + \vartheta$ is the point along $\tilde{\gamma}_2$ of distance ϑ to y (and similar for $D(x)$), then

$$d^*(y, D(x)) = d(y + \vartheta, D(x) + \vartheta).$$

Now, set $f(x, y, \vartheta) = d(x, y) + d(y + \vartheta, D(x) + \vartheta)$. Then we have

$$\Lambda(\vartheta) = \inf_{x \in \tilde{\gamma}_1, y \in \tilde{\gamma}_2} \{f(x, y, \vartheta)\}.$$

Observe that f only depends on points and distances in $\mathbb{H}^2$, so from here on it is sufficient to use properties of the hyperbolic distance. By proposition 4.10 we know that $\tilde{\gamma}_2$ and $\tilde{\gamma}_3$ do not intersect in $\overline{\mathbb{H}^2}$. A property of the hyperbolic distance (which is a key feature of *hyperbolic* geometry, as opposed to Euclidean) is that $d(y+\vartheta, T(x)+\vartheta)$ is strictly convex as a function of ϑ, and has a minimum. In other words, if two points traverse along two geodesics, disjoint in $\overline{\mathbb{H}^2}$, ultimately the distance will be strictly increasing. This implies that Λ is strictly increasing for large enough arguments, so the claim follows. □

Proof of theorem 5.10. We show that Ψ is surjective and injective.

Ψ *is surjective.* Indeed, we know from proposition 5.6 that the metrics on M in the image of Φ produces all metrics of the pants when decomposing M. When reconstructing M from the pants, the only way to obtain different metrics is to twist the boundaries of the pants. By proper choice of ϑ we can obtain all these twists, so Ψ is surjective.

Ψ *is injective.* First, we claim that it is sufficient to show that $\Theta_\vartheta(h)$ is injective as a function of $\vartheta \in \mathbb{R}^{3(g-1)}$. Indeed, if so is the case then $\Psi(l_1, \vartheta_1) = \Psi(l_2, \vartheta_2)$ implies $\Theta_{\vartheta_1}(\sigma(l_1)) = \Theta_{\vartheta_2}(\sigma(l_2))$. Because Θ keep the fibres of L invariant, it follows that $\sigma(l_1) = \sigma(l_2)$, and because σ is injective we have $l_1 = l_2$. If furthermore Θ_ϑ is injective as a function of ϑ we conclude that $\vartheta_1 = \vartheta_2$, which shows the claim.

It remains to show the injectivity of Θ_ϑ. Let $\langle h\rangle \in \tau_g$ be a fixed class of metrics, and assume $\Theta_\vartheta(h) = \Theta_{\vartheta'}(h)$, with $\vartheta \neq \vartheta'$. Then ϑ and ϑ' differ in some coordinate, say the coordinate i. Define $\hat{\vartheta}$, $\hat{\vartheta}'$ as the corresponding vectors with the i:th coordinate replaced with 0, and let $\vartheta_0 = \vartheta - \hat{\vartheta}$, $\vartheta_0' = \vartheta' - \hat{\vartheta}'$. Define

$$h_1 = \Theta_{\hat{\vartheta}}(h)$$
$$h_2 = \Theta_{\hat{\vartheta}'}(h).$$

Note that $\Theta_{\vartheta_0}(h_1) = \Theta_{\vartheta_0}(h_2)$. This follows from the definition of Θ, because Θ only alters the metric in small collars around the geodesic loops. To twist all collars is the same as to first twist all collars but one, and then twist the remaining collar.

Because $\Theta_{\vartheta_0}(h_1) = \Theta_{\vartheta_0}(h_2)$, it follows that for every n,

$$\tilde{\Lambda}^{\Theta_{(0,..,n+\vartheta_i,..0)}(h_1)}(\beta_i) = \tilde{\Lambda}^{\Theta_{(0,..,n+\vartheta_i',..0)}(h_2)}(\beta_i)$$

Furthermore, h_1 and h_2 are "untwisted" in the collar C_i, so there they coincide with h. β_i is disjoint from every collar *but* C_i. It follows that we can replace h_1 and h_2 by h in above formula, so using the notation from lemma 5.12 we have

$$\Lambda_i(n + \vartheta_i) = \Lambda_i(n + \vartheta_i').$$

From lemma 5.12 we know that Λ_i is strictly increasing for large enough arguments. But if $\Lambda_i(n + \vartheta_i) = \Lambda_i(n + \vartheta_i')$, for n large enough we must have $n + \vartheta_i = n + \vartheta_i'$, i.e. $\vartheta_i = \vartheta_i'$. This contradicts the assumptions, so it follows that Θ_ϑ is a injective function of ϑ. The theorem is thus proved. □

6 Mapping class group

Recall from the introduction that we would like to study the equivalence classes of $\mathcal{H}$ under the relation $h_1 \sim h_2$ if there is an orientation-preserving isometry $\phi : (M, h_1) \to (M, h_2)$. We now define and explain formally the notions and ideas presented in the introduction.

Definition 6.1. The *Riemann moduli space* for M, denoted $\mathcal{M}_g$, is defined as the set $\mathcal{H}/\sim$.

Note that if two metrics h_1, h_2 are equivalent under the Teichmüller relation, then $h_1 \sim h_2$. Consequently, instead of identifying metrics in $\mathcal{H}$ we can build the Riemann moduli space by identifying metric classes in τ_g under a suitable relation.

Definition 6.2. Let $\mathrm{Diff}^+(M)$ be the set of orientation-preserving diffeomorphisms of M, and define the relation $\phi_0 \sim_I \phi_1$ if $\phi_0, \phi_1 \in \mathrm{Diff}^+(M)$ are isotopic. Then $\Gamma_g = \mathrm{Diff}^+(M)/\sim_I$ is the *mapping class group* of M.

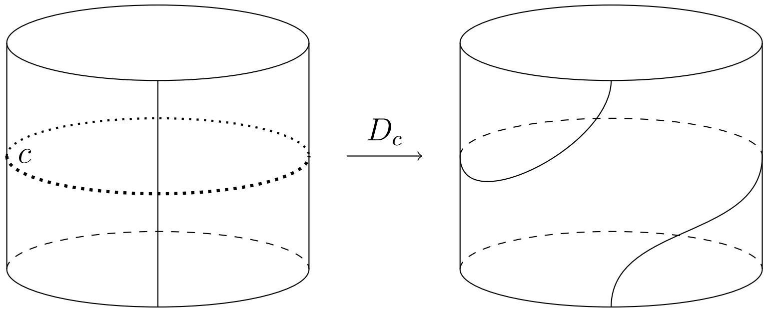


Figure 26: Definition of a Dehn twist.

The mapping class group acts on τ_g as follows. For a class $\langle h \rangle \in \tau_g$, the action of a mapping class $\langle \gamma \rangle \in \Gamma_g$ is defined as the class of metrics $\langle \tilde{h} \rangle$ obtained by the requirement that $\gamma : (M, h) \to (M, \tilde{h})$ is an isometry. This gives a well-defined group action, and the quotient satisfies

$$\mathcal{M}_g = \tau_g \big/ \Gamma_g .$$

6.1 Generators for Γ_g

The goal of this section is to find generators for Γ_g. We begin by defining a particular set of diffeomorphisms, these will turn out to generate Γ_g. Let as before T_g be the complete, connected, compact, oriented surface of genus g.

Definition 6.3. Let c be a simple loop in T_g and let C be a neighbourhood of c homeomorphic to a cylinder. Then a diffeomorphism D_c, acting on C as illustrated in figure 26 and extended as identity on $T_g \backslash C$ is known as a *Dehn twist* about c.

Remark 6.4. The definition of the Dehn twists depends on the orientation of T_g but not on c. A *positive* Dehn twist is defined as follows: If we go along a curve p towards c, then p will be twisted to the right. A *negative* Dehn twist is defined analogously, but twisting to the left. In figure 26, if we choose the normal direction to be outwards from the cylinder then D_c is a positive Dehn twist.

Remark 6.5. We can find the image of a loop p under a Dehn twist about c in the following way. At every intersection of p and c, "cut open" p and c, and glue together the edges of c to the edges of p. Figure 26 shows why this method works. Note that we only consider loops up to isotopy. The precise position of a loop is therefore immaterial, we only care about its isotopy class.

We will prove the following.

Theorem 6.6. *The mapping class group Γ_g is generated by the mapping classes of Dehn twists.*

The proof requires several preparatory lemmas. These will provide an intuition on how Dehn twists behave. We will use the notation $p \sim_C q$ if the loop p is isotopic to q up to Dehn twists, i.e. if p is isotopic to the image of q under a composition of Dehn twists. For the proof of theorem 26 we will need to consider $T_{g,r}$, the surface of genus g with r open discs removed. We let $\mathrm{Diff}^+(T_{g,r})$ be the set of diffeomorphisms of $T_{g,r}$ which restricts to the identity on the boundary components of $T_{g,r}$. If $r > 0$, all isotopies of $T_{g,r}$

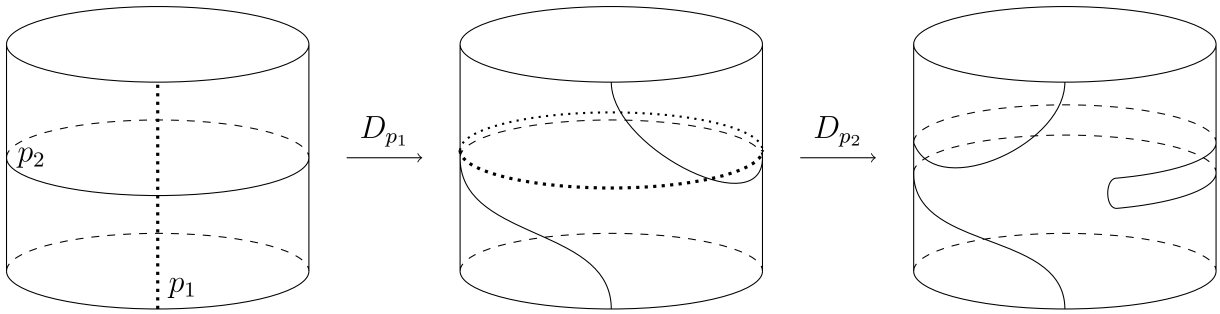


Figure 27: Two loops intersecting once are isotopic up to Dehn twists. Bold dotted lines indicate which loop is twisted about.

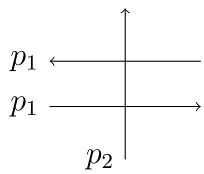
will be relative $\partial T_{g,r}$. We say that a curve p in $T_{g,r}$ is *nonseparating* if $T_{g,r}\backslash p$ is connected. The goal with the lemmas is to show that all non-separating curves are isotopic up to Dehn twists.

Lemma 6.7. *Let p_1, p_2 be loops in $T_{g,r}$ intersecting precisely once. Then $p_1 \sim_C q_2$.*

Proof. Apply a Dehn twist about p_1 to p_2, and then apply a Dehn twist about p_2 to the image of the first Dehn twist. This will be isotopic to p_1, as indicated in figure 27. □

Remark 6.8. It is an easy exercise to show that with the notation of lemma 6.7, the loop $D_{p_1}(D_{p_2})^2 D_{p_1}(p_2)$ is isotopic to p_2, but with the orientation reversed. This loop is the result of first performing the construction in 27, followed by the "reverse" construction. The conclusion is that we can reverse the orientation of any loop by applying Dehn twists.

Definition 6.9. We say that two loops p_1, p_2 have *zero algebraic intersection* either if the loops have no intersection, or if they intersect precisely twice with opposite orientation at the intersection points. The latter situation is illustrated in the following figure.



Lemma 6.10. *Let p, q be simple loops in $T_{g,r}$. Then there is a loop v in T_g such that $v \sim_C p$ and v, q have zero algebraic intersection.*

Proof. The proof is by induction on the number i of intersections. We can assume $i < \infty$, because up to isotopy p does not coincide with q in any open subset of $T_{g,r}$. If p and q have zero algebraic intersection we are done. If p and q intersect precisely once, lemma 6.7 implies $p \sim_C q$. Assume the lemma holds for i intersections, and that p and q have $i + 1$ intersections. We have two cases.

Case 1. p and q have two consecutive intersection points along q, intersecting in the same orientation. We construct the curve c to perfom a Dehn twist about. Start near an intersection point of p and q. Let c run close to p without intersecting p, until the next intersection. Then close c with a straight line segment intersecting p once. This is illustrated in figure 28. Now apply the Dehn twist about c to p. This produces a curve with i intersections with q, so we can apply the induction hypothesis.

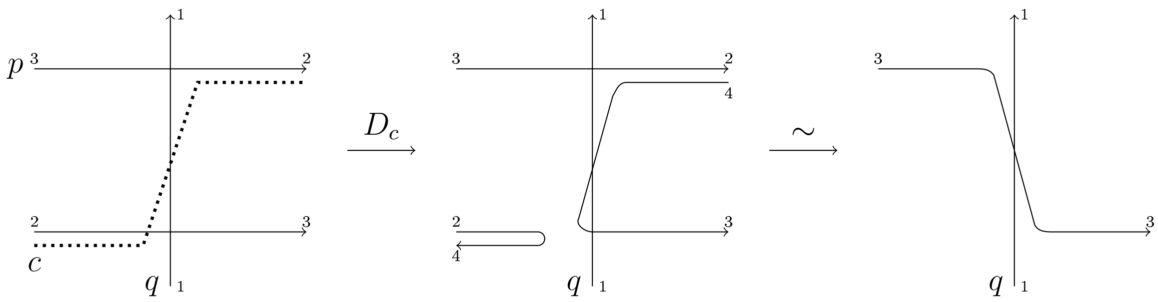


Figure 28: Construction used in the proof of lemma 6.10, case 1. The numbering shows how the segments are connected

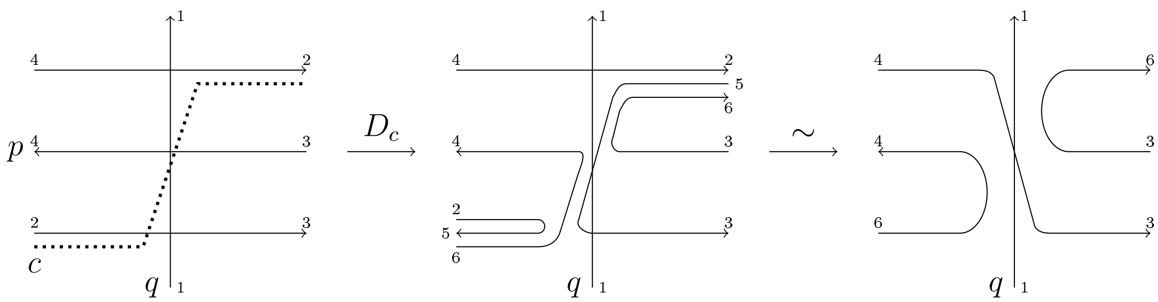


Figure 29: Construction used in the proof of lemma 6.10, case 2. The numbering shows how the segments are connected

Case 2. p and q have three consecutive intersection points along q, intersecting with alternating orientation. In the first and the third of these intersection points, p and q intersect in the same orientation. Using these points, construct c analogously as in case 1. Observe that c will now intersect p twice. Apply a Dehn twist about c. This produces a curve $D_c(p)$ with fewer intersections, so we can apply the induction hypothesis. The procedure is illustrated in figure 29. □

Lemma 6.11. *Let p, q be non-separating curves in $T_{g,r}$. Then $p \sim_C q$.*

Proof. By lemma 6.10 we can find a curve $v \sim_C p$ having zero algebraic intersection with q. There are two cases to consider.

v and q are disjoint. Because p is non-separating, so will v be. Non-separating loops in $T_{g,r}$ are meridians or equators. Because v and q are disjoint, is clear that we can find a curve w which intersects each of v and q precisely once. By lemma 6.7 $v \sim_C w$ and $q \sim_C w$. Consequently $p \sim_C v \sim_C w \sim_C q$, so $p \sim_C q$.

v and q intersect twice. By assumption, v and q have opposite orientation in the two intersection points. Note that v will divide q into two arcs. Define two curves a and b in the following way. Let a start near v and run along v until an intersection between v and q. Let a run along q until the next intersection, then let a run along v. Similarily, let b be the curve which runs along the same part of v, but the opposite part of q. This construction is illustrated in figure 30.

Now, because q is non-separating either a or b is non-separating. Indeed, a, b and q form the boundary of a pant. It is clear that if the "waist" is non-separating then one of the "ankles" also must be non-separating. Say a is non-separating, because a and v

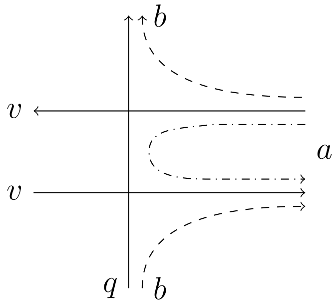


Figure 30: Definition of the loops a and b in the proof of lemma 6.11

does not intersect it follows from above case that $v \sim_C a$. Furthermore, a and q does not intersect, so $a \sim_C q$. The conclusion is that $p \sim_C v \sim_C a \sim_C q$, so $p \sim_C q$.

□

We are now ready for the proof of theorem 6.6.

Proof of theorem 26. The proof relies on induction on the number $k(g,r) = 3g + r$. We begin by establishing the base of the induction, namely the cases $k(g,r) = 2$ and $k(g,r) = 3$ for $g = 0$.

Base of induction. If $k = 2$, then $g = 0$ and $r = 2$, so we have a cylinder. Let $f \in \text{Diff}^+(T_{g,r})$ and let p be a curve between the two boundary components of $T_{g,r}$. Because f fixes the boundary of $T_{g,r}$, $f(p)$ will be a curve between the same points on the boundary, possibly in a different isotopy class relative endpoints. It is clear that f is isotopic to the composition of Dehn twists which carries p to $f(p)$.

If $k = 3$ and $g = 0$, then $r = 3$, so we have a pant. Let p be a curve between two distinct boundary components ∂_1 and ∂_2 of $T_{g,r}$, and let $f \in \text{Diff}^+(T_{g,r})$. We begin by establishing that up to Dehn twists, p and $f(p)$ are isotopic relative endpoints. $f(p)$ will be a curve between the same points on the boundary as p. Let N_1 and N_2 be disjoint cylindrical neighbourhoods of ∂_1 and ∂_2 respectively, such that p intersect each boundary component precisely one. We can assume p and $f(p)$ coincide in $T_{g,r} \backslash (N_1 \cup N_2)$ (if not, we can find an isotopy taking the segment of $f(p)$ in $T_{g,r} \backslash (N_1 \cup N_2)$ to the segment of p). Inside the cylinders N_1 and N_2, p and $f(p)$ are paths between the same points on the boundary. It is clear that p and $f(p)$ are isotopic up to Dehn twists. We have established our claim.

Now, because p and $f(p)$ are isotopic up to Dehn twists, we can find a composition g of Dehn twists and diffeomorphisms isotopic to the identity such that $g(q) = p$. If we let $f_* = g \circ f$, then $f_*(p) = p$. In light of the remark following lemma 6.7 we can even assume f_* to preserve the orientation of p, so (possibly composing with a diffeomorphism isotopic to the identity) $f_*|_p$ is the identity on p. Cut open $T_{g,r}$ in p, then we obtain the cylinder. We may now consider f_* as a element of $\text{Diff}^+(T_{0,2})$, so f_* is isotopic to a composition of Dehn twists in $T_{0,2}$. It follows that f is isotopic to a composition of Dehn twists in $T_{0,3}$.

Induction step. Assume any $f \in \text{Diff}^+(T_{g_0,r_0})$ is isotopic to a composition of Dehn twists for every $2 \leq k(g_0, r_0) < k_0$, with $k_0 > 3$. Let $T_{g,r}$ be a fixed surface such that $k(g,r) = k_0$. We begin with the case $g \geq 1$.

Let $f \in \text{Diff}^+(T_{g,r})$, and let p be a non-separating curve in $T_{g,r}$. Then also $q = f(p)$ is non-separating. We know from lemma 6.11 that $p \sim_C q$. Proceeding as in the base of induction, construct f_* so that f_* is the identity on p. Cut open $T_{g,r}$ in p. Then we reduce the genus by one, and increase the number of boundary components by two. Consequently, we obtain $T_{g-1,r+2}$, satisfying $k(g-1, r+2) = k_0 - 1$. By induction hypothesis f_* is isotopic to a composition of Dehn twists in $T_{g-1,r+2}$, so it follows that f is isotopic to a composition of Dehn twists in $T_{g,r}$.

The case $g = 0$, then $T_{g,r}$ is a sphere with r discs removed. Because $k(g,r) = r = k_0 > 3$, we can find a curve p enclosing precisely two boundary components of $T_{g,r}$. If $f \in \mathrm{Diff}^+(T_{g,r})$, because f preserves the boundaries of $T_{g,r}$ it follows that $f(p)$ encloses the same boundary components as p. Thus $f(p)$ is isotopic to p. As above we define f_* and cut open $T_{g,r}$. The two connected components satisfy $k = 3$ and $k = k_0 - 1$ respectively, so the induction hypothesis applies to both components. The theorem then follows. □

Remark 6.12. Above we only considered the case $g \geq 1$. For $g = 0$, it turns out that the mapping class group is trivial.

6.2 Applications to the Riemann moduli space

We have now studied both the Teichmüller space and the mapping class group. We have seen that τ_g can be parametrized by the lengths and twists of the loops giving a pant decomposition. We further saw that twisting a full turn around a loop gives rise to two different points in τ_g. But the metric obtained from twisting a full turn is precisely the pull-forward metric of a Dehn twist. It follows that two points in τ_g differing by a full twist will be the same point in $\mathcal{M}_g$.

Above we found generators for the mapping class group Γ_g. However, we have not studied the relations between the generators, so our knowledge of Γ_g is quite limited. We have not studied the action of Γ_g on τ_g either, which is central for the study of the Riemann moduli space.

7 Conclusions

In this work we have presented proofs of two main theorems, theorem 5.10 and 6.6. The Teichmüller space τ_g for the surface T_g of genus $g \geq 2$ can be parametrized by $\mathbb{R}_+^{3(g-1)} \times \mathbb{R}^{3(g-1)}$. The parameters in $\mathbb{R}_+^{3(g-1)}$ describe the lengths of the geodesic loops giving rise to a pant decomposition of T_g. The parameters in $\mathbb{R}^{3(g-1)}$ describe how the metric is twisted in the collars around each geodesic in the pant decomposition.

The mapping class group of the surface of any genus is generated by Dehn twists. This helps us understand the action of the mapping class group on the Teichmüller space. It further provides a way of studying the Riemann moduli space as the quotient of the Teichmüller space by the mapping class group.

For further studies, there are several possible directions to take. A direct continuation of this work is to further study properties of mapping class group, and how it acts on the Teichmüller space. Another possible direction is to study the geometry of the Teichmüller space, by defining a metric on the space. One can also study hyperbolic geometry in higher dimensions. The *Mostow rigidity theorem* states that if it exists a hyperbolic structure on a closed manifold of dimension larger than 2, then it is unique. The corresponding Teichmüller space is thus trivial.

8 References

[1] Benedetti, R. & Petronio, C. (1991). *Lectures on Hyperbolic Geometry.* Berlin: Springer-Verlag.

[2] Gamelin, T.W. (2000). *Complex analysis.* New York: Springer.

[3] Munkres, J.R. (2000). *Topology.* 2nd ed. Upper Saddle River: Prentice Hall.

[4] Bredon, G.E. (1993). *Topology and geometry.* New York: Springer.

[5] Baer, R., Kurventypen auf Flächen. *J. reine angew. Math.* 156 (1927), 231-246.

[6] Baer, R., Isotopie von Kurven auf orienterbaren, geschlossenen Flächen und ihr Zuzammenhang mit der topologischen Deformation der Flächen. *J. reine angew. Math.* 159 (1928), 101-111.

[7] Birman, J.S. (1975). *Braids, Links, and Mapping class groups.* Princeton: Princeton University press.

[8] Au, T.K.K, Luo, F. & Yang, T. (2010) *Lectures on the Mapping Class Group of a surface.* [online]. Available from: `http://web.stanford.edu/~yangtian/lecture.pdf` [Accessed 10th June, 2016]