Mapping class groups

Henry Wilton*

March 8, 2021

Contents

*Please send comments and corrections to `h.wilton@maths.cam.ac.uk`.

Lecture 1: Introduction

These notes are closely based on the first few chapters of the book by Farb and Margalit [1].

1.1 Surfaces of finite type

A connected, smooth, oriented surface S is *of finite type* if it can be obtained from a compact surface (possibly with boundary ∂S) by removing a finite number of points. These will be our main objects of study in this course. As well as being of obvious interest to topologists, surface of finite type, in the guise of Riemann surfaces, have been studied since the nineteenth century by complex analysts, algebraic geometers and number theorists.

Fortunately, the classification of surfaces tells us exactly what the possible topological types of the surfaces S of finite type are.

Theorem 1.1 (Classification of surfaces of finite type). *Every connected, orientable surface of finite type is diffeomorphic to some $S_{g,n,b}$, the surface obtained by connect-summing $g \geq 0$ copies of the torus T^2 with the 2-sphere S^2, and removing b open discs and n points.*

The *closed* surfaces $S_g = S_{g,0,0}$ are the ones with no punctures or boundary components. The *compact* surfaces $S_{g,0,b}$ are the ones with no punctures. Another important invariant is the Euler characteristic

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

Example 1.2. If $\chi(S) > 0$ then $g = 0$ and $n + b \leq 1$. It follows that S is one of S^2, $\mathbb{C}$ or the compact 2-disc D^2.

Example 1.3. If $\chi(S) = 0$ then either $g = 1$ and $n + b = 0$ or $g = 0$ and $n + b = 2$. Therefore, S is one of the following: the 2-torus T^2, the twice-punctured sphere (aka the punctured plane $\mathbb{C}_*$), the punctured disc D^2_*, or the annulus $A = S^1 \times [-1, 1]$.

Setting aside these finite lists of of exceptional examples, the remaining surfaces of finite type satisfy $\chi(S) < 0$. Shortly, we will see that that we can study them using hyperbolic geometry.

1.2 Mapping class groups

When studying a surface S, one naturally becomes interested in the group of self-homeomorphisms $\mathrm{Homeo}(S)$. However, $\mathrm{Homeo}(S)$ is huge – an infinite-dimensional topological group – which makes it difficult to study. The idea behind the mapping class group is to rectify this problem by quotienting out by homotopy. But homotopy isn't quite the right notion here; since the group $\mathrm{Homeo}(S)$ consists of homeomorphisms, we say that two self-homeomorphisms ϕ_0, ϕ_1 of S are *isotopic* if they are related by a homotopy ϕ_t that consists of homeomorphisms ϕ_t for every t. Another way to say this is that ϕ_0 and ϕ_1 isotopic if they live in the same path-component of $\mathrm{Homeo}(S)$.

Definition 1.4. Let $\mathrm{Homeo}^+(S, \partial S)$ be the group of orientation-preserving homeomorphisms $S \to S$ that restrict to the identity on ∂S, equipped with the compact-open topology (i.e. the topology of uniform convergence on compact subsets). Let $\mathrm{Homeo}_0(S, \partial S)$ denote the path-component of $\mathrm{Homeo}^+(S, \partial S)$ that contains the identity. That is, $\mathrm{Homeo}_0(S)$ is the set of elements that are isotopic to the identity, where we require isotopies to fix the boundary. Note that $\mathrm{Homeo}_0(S, \partial S)$ is a normal subgroup of $\mathrm{Homeo}^+(S, \partial S)$. The *mapping class group* of S is defined to be the quotient

$$\mathrm{Mod}(S) := \mathrm{Homeo}^+(S, \partial S)/\mathrm{Homeo}_0(S, \partial S)\,.$$

There are several other possible definitions of mapping class groups: one might only consider diffeomorphisms of S up to smooth isotopy, or homeomorphisms of S up to homotopy. In dimension 2, these definitions turn out to give the same result, thanks to the following facts.

Theorem 1.5 (Baer, 1920s)**.** *If two orientation-preserving diffeomorphisms of a surface S of finite type are homotopic relative to ∂S, then they are smoothly isotopic relative to ∂S.*

Theorem 1.6 (Munkres, 1950s)**.** *Every homeomorphism of S (relative to ∂S) is isotopic to a diffeomorphism of S (relative to ∂S).*

Corollary 1.7. *For any smooth S, there are natural isomorphisms*

$$\mathrm{Mod}(S) \cong \mathrm{Diff}^+(S, \partial S)/\mathrm{Diff}_0(S, \partial S) \cong \mathrm{Homeo}^+(S, \partial S)/\simeq$$

(where $\simeq$ denotes homotopy relative to ∂S).

Proof. There is a natural homomorphism

$$\Phi : \mathrm{Diff}^+(S, \partial S)/\mathrm{Diff}_0(S, \partial S) \to \mathrm{Mod}(S)\,,$$

since every diffeomorphism is a homeomorphism and a smooth isotopy is an isotopy. There is also a surjection

$$\Psi : \mathrm{Mod}(S) \to \mathrm{Homeo}^+(S, \partial S)/\simeq$$

since isotopies are homotopies. Theorem 1.6 implies that Φ is a surjection, and Theorem 1.5 implies that $\Psi \circ \Phi$ is injective. The result follows. □

In this course we will pass freely between the continuous and smooth points of view, without much comment. In general, continuous maps are easy to build by gluing maps, while smooth maps have various convenient regularity properties. (For instance, smooth curves have regular neighbourhoods.)

1.3 Context and motivation

Since mapping class groups arise in many different parts of mathematics, many different motivations can be given. We'll give a few here.

From the point of view of topology, mapping classes give a convenient way of constructing bundles. For instance, any self-homeomorphism $\phi : S \to S$ gives rise to a *surface bundle over the circle*:

$$M_\phi := S \times [0,1]/\sim$$

where $\sim$ identifies $(x,1)$ with $(\phi(x),0)$. Note that changing ϕ by an isotopy doesn't change the resulting manifold, so in fact M_ϕ only depends on the mapping class of ϕ. This is a huge source of examples of interesting 3-manifolds. More generally, surface bundles over a connected cell complex B correspond to group homomorphisms $\pi_1 B \to \mathrm{Mod}(S)$.

Next, we'll give the 'high-level' motivation for $\mathrm{Mod}(S)$. Geometers and topologists are interested in studying all possible geometric structures on the genus-g surface S_g, while algebraists and number theorists are interested in studying all possible complex structures on S_g. These turn out to be equivalent, and to be encoded in the points of a space $\mathcal{M}_g$, the *moduli space* of S_g. In particular, we'd like to understand the topology of $\mathcal{M}_g$. Now, moduli space has a contractible 'universal cover' $\mathcal{T}_g$, called *Teichmüller space*.

The moduli space can then be recovered as the quotient of $\mathcal{T}_g$ by the action of the 'fundamental group'.[1] This 'fundamental group' of moduli space is the mapping class group $\mathrm{Mod}(S_g)$. In summary, $\mathrm{Mod}(S_g)$ captures all the topological information about the moduli space $\mathcal{M}_g$.

Our final attempt at motivation is an analogy, which motivates a great deal of research into mapping class groups. When studying the integer lattice $\mathbb{Z}^n$ in $\mathbb{R}^n$, one is naturally led to its group of linear automorphisms $SL_n(\mathbb{Z})$. Our surfaces S correspond to tilings of the (Euclidean or hyperbolic) plane, and placing a point in the centre of each tile describes a kind of lattice in the plane. The mapping class group $\mathrm{Mod}(S)$ plays the role of the group of automorphisms of that lattice. This leads us to think of mapping classes as analogous to integer matrices, and we can try to develop machinery for $\mathrm{Mod}(S)$ that is analogous to the techniques of linear algebra that we use to study $SL_n(\mathbb{Z})$.

Here is a table that locates some of the objects that we will learn about within this analogy.

Surfaces	Tori
S (a surface of finite type)	T^n (the n-torus)
$\pi_1 S$	$\mathbb{Z}^2$
$\mathrm{Mod}(S)$	$SL_n(\mathbb{Z})$
mapping classes	linear maps
closed curves on S	vectors in $\mathbb{Z}^n$
...	...

Lecture 2: Surfaces and hyperbolic geometry

A *closed curve* on a surface S is a map (in the appropriate category) $S^1 \to S$. Closed curves play a role analogous to vectors in vector spaces. Although the surfaces we study are topological objects, in order to compute in them it is often useful to endow them with geometry – that is, a Riemannian metric of constant curvature. For most surfaces, this geometry is hyperbolic.

[1] The scare quotes are because this isn't quite true – it's true in the setting of orbifolds, but not in the setting of manifolds.

2.1 The hyperbolic plane

Recall that the hyperbolic plane can be modelled as the upper half-plane

$$\mathbb{H}^2 = \{x + iy \in \mathbb{C} \mid y > 0\}$$

equipped with the Riemannian metric

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

Geodesics are vertical lines and semicircles that meet the real axis perpendicularly. This is called the *upper half-plane model*. This model makes it easy to see that the group of orientation-preserving isometries $\text{Isom}^+(\mathbb{H}^2)$ is precisely $PSL_2(\mathbb{R})$, the group of Möbius transformations with real coefficients, which acts on the upper half-plane in the natural way.

Another model is the *Poincaré disc model*, which is obtained by conjugating the upper half-plane by the Möbius transformation

$$z \mapsto \frac{z - i}{z + i} .$$

This yields the open unit disc in $\mathbb{C}$, equipped with the Riemannian metric

$$ds^2 = 4\frac{dx^2 + dy^2}{(1 - r^2)^2} .$$

The details of the Riemannian metric itself aren't so important, but note that it is radially symmetric. In this model, the geodesics are still circles and lines that meet the boundary circle at right angles.

The *boundary at infinity* of $\mathbb{H}^2$ is the copy of S^1 that you see as the boundary of the disc model of $\mathbb{H}^2$. We write $\overline{\mathbb{H}}^2 = \mathbb{H}^2 \cup \partial\mathbb{H}^2$, which is homeomorphic to the disc. Note that any isometry $f \in \text{Isom}^+(\mathbb{H}^2)$ of $\mathbb{H}^2$ extend to a Möbius transformation $\bar{f}$ of $\overline{\mathbb{H}}^2$, sending the boundary to itself. This gives us an easy way to classify isometries of $\mathbb{H}^2$, according to the number of fixed points.

Let $f \in \text{Isom}^+(\mathbb{H}^2)$. By the Brouwer fixed point theorem, $\bar{f}$ has at least one fixed point in $\overline{\mathbb{H}}^2$. We can now classify f by the number of fixed points of $\bar{f}$.

If $\bar{f}$ has at least three fixed points then, since it is a Möbius transformation, it is the identity.

If $\bar{f}$ has two fixed points ξ^+, ξ^- then, since it preserves the hyperbolic geodesic between them, ξ^+, ξ^- must lie on $\partial\mathbb{H}^2$. (Otherwise, it fixes the geodesic pointwise and has infinitely many fixed points.) In this case, f is called *hyperbolic* or *loxodromic*. The unique geodesic line in $\mathbb{H}^2$ with endpoints $\xi^\pm$ is denoted by $\mathrm{Axis}(f)$, and f acts on its axis by translation by a fixed distance $\tau(f)$. In the upper half-plane model, f is conjugate to a dilation by $e^{\tau(f)}$. A geometric argument[2] shows that that, for every $x \in \mathbb{H}^2 \smallsetminus \mathrm{Axis}(f)$, $d(x, f(x)) > \tau(f)$.

The case when $\bar{f}$ has a unique fixed point $\xi \in \overline{\mathbb{H}}^2$ splits into two sub-cases. If $\xi \in \mathbb{H}^2$ then f is called *elliptic*, and is conjugate to a rotation in the disc model. If $\xi \in \partial\mathbb{H}^2$ then f is called *parabolic*, and is conjugate to one of the translations $z \mapsto z \pm 1$ in the upper half-plane model. In both cases, there are points $x \in \mathbb{H}^2$ such that $d(x, f(x))$ is arbitrarily small.

2.2 Hyperbolic structures

In this section, we'll assume to start with that S is compact. A *geometric structure* on S is a complete Riemannian metric of constant curvature $\kappa = 1, 0, -1$, in which any boundary components are geodesics. The Gauss–Bonnet theorem asserts that

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

so κ necessarily has the same sign as $\chi(S)$.

In the case when $\chi(S) > 0$, either $S = S^2$ or $S = D^2$. The sphere S^2 admits a well-known geometric structure, while a hemisphere in S^2 gives a geometric structure on the disc.

When $\chi(S) = 0$, the only compact examples are the torus and the annulus. In each case, it's easy to realise S as a quotient of a convex subset of $\mathbb{R}^2$ by isometries. The Euclidean metric on $\mathbb{R}^2$ descends to a metric on S of constant curvature 0.

The (infinitely many) remaining cases all have $\chi(S) < 0$. The following theorem, which also applies to surfaces with punctures, guarantees a hyperbolic metric on S.

Theorem 2.1. *Let S be a connected, oriented surface of finite type with $\chi(S) < 0$. There is a convex subspace $\widetilde{S}$ of the hyperbolic plane $\mathbb{H}^2$ and an*

[2] See Question 4 of Example Sheet 1.

action of the fundamental group $\pi_1 S$ by isometries on $\widetilde{S}$, with finite-area fundamental domain, and a diffeomorphism

$$\pi_1 S \backslash \widetilde{S} \cong S\,.$$

In particular, S carries a metric of curvature -1.

Sketch proof. We will describe the case when S is closed, and leave the reader to work out how to adapt the argument to the other cases. The genus-g surface S can be constructed from a $4g$-gon P by identifying sides in pairs, in such a way that all vertices of P are identified with each other. By a continuity argument, there exists a regular hyperbolic $4g$-gon with interior angles $\pi/2g$; endow P with this metric structure. After gluing, this defines a metric on S so that every point has a neighbourhood isometric to a disc in $\mathbb{H}^2$. The universal cover $\widetilde{S}$ is a complete, simply connected surface of constant curvature -1. A classical theorem of Riemannian geometry (beyond the scope of this course) implies that $\widetilde{S}$ is isometric to $\mathbb{H}^2$. By construction of the metric on $\widetilde{S}$, the action of $\pi_1 S$ is by isometries, and the desired diffeomorphism follows from standard covering-space theory. □

We will call a surface S equipped with a geometric structure modelled on the hyperbolic plane, as in Theorem 2.1, a *hyperbolic surface.*

Lecture 3: Curves and isometries

3.1 Curves on hyperbolic surfaces

Let S be a hyperbolic surface. We shall see that the hyperbolic structure is very useful for analysing elements of $\pi_1 S$ or, equivalently, curves on S. A *closed curve* is a map $\alpha : S^1 \to S$. It is *inessential* if it is homotopic to a point or a puncture, and inessential otherwise.[3]

Standard algebraic topology gives us a bijection between homotopy classes of loops $S^1 \to S$ and conjugacy classes in $\pi_1 S$. By the classification of hyperbolic isometries, these elements of $\pi_1 S$ can be either elliptic, parabolic or hyperbolic. The next lemma tells us when the different types of isometries occur.

[3] Beware! Farb–Margalit call a curve ‘inessential’ if it is homotopic to a point, a puncture or a boundary component.

Lemma 3.1. *Let S be a hyperbolic surface and α a closed curve on S.*

(i) If α is elliptic then α is homotopic to a point.

(ii) If α is parabolic then α is homotopic to a puncture.

(iii) If α is an essential curve then α is hyperbolic.

In particular, $\pi_1 S$ is torsion-free.

Proof. By doubling S along its boundary, we may assume that $\partial S = \varnothing$, and hence that $\widetilde{S} = \mathbb{H}^2$.

Since $\pi_1 S$ acts freely on $\mathbb{H}^2$, if α is elliptic then it acts as the identity on $\mathbb{H}^2$, so α represents the trivial element of $\pi_1 S$ and is homotopic to a point.

For (ii), without loss of generality, α acts on $\mathbb{H}^2$ as the translation $g : x \mapsto x + 1$. Let the loop α be based at x_0, and let $\tilde{x}_0$ be a pre-image of x_0 in $\mathbb{H}^2$. Then α lifts to a path $\tilde{\alpha}$ from $\tilde{x}_0$ to $\tilde{x}_0 + 1$ in $\mathbb{H}^2$. For each $s \in [0, \infty)$, consider the path $\tilde{\alpha}_s : [0,1] \to \mathbb{H}^2$ defined by $t \mapsto \tilde{\alpha}_0(t) + is$. The path $\tilde{\alpha}_s$ descends to a loop α_s in S. As $s \to \infty$, this family defines a homotopy from α to a puncture in S.

By the classification of isometries of $\mathbb{H}^2$, to prove item (iii) it is enough to show that if α is homotopic to a puncture then the corresponding isometry is parabolic. If α is homotopic to a puncture then there is a family of closed curves α_s, all homotopic to α, such that $l(\alpha_s) \to 0$ – if not, then we can construct annuli embedded in S of unbounded area. Fixing a lift $\tilde{x}_0$ of $\alpha(0)$ to the universal cover $\mathbb{H}^2$ and lifting homotopies, we obtain well defined lifts $\tilde{\alpha}_s$ of the α_s. For each s, let $\tilde{x}_s = \tilde{\alpha}_s(0)$, and note that $\alpha.\tilde{x}_s = \tilde{\alpha}_s(1)$. Then

$$\tau(\alpha) \leq d(\tilde{x}_s, \alpha.\tilde{x}_s) = d(\tilde{x}_s, \tilde{\alpha}_s(1)) \leq l(\alpha_s)$$

whence $\tau(\alpha) = 0$. So α is parabolic as required. □

Hyperbolic geometry provides us with canonical representatives for homotopy classes of curves. Note that the uniqueness part of this statement fails when S is the torus.

Lemma 3.2. *Let S be a hyperbolic surface and let α be a closed curve on S which is not homotopic to a point or a puncture. Then there is a unique geodesic curve in the homotopy class of α.*

Proof. Choose a basepoint $x_0 \in S$ for α, and a preimage $\tilde{x}_0$ in $\mathbb{H}^2$. Let $\tilde{\alpha}$ be the unique lift of the composition

$$\mathbb{R} \to S^1 \xrightarrow{\alpha} S$$

(where $\mathbb{R} \to S^1$ is the universal covering map) that sends 0 to $\tilde{x}_0$. Note that the map $\tilde{\alpha}$ is $\mathbb{Z}$-equivariant.

By Lemma 3.1, α acts as a hyperbolic isometry on $\mathbb{H}^2$, preserving an axis $\text{Axis}(\alpha)$. Let $\pi : \mathbb{H}^2 \to \text{Axis}(\alpha)$ be orthogonal projection and, for each $t \in \mathbb{R}$, let $\tilde{\gamma}_t : [0,1] \to \mathbb{H}^2$ be the constant-speed geodesic from $\tilde{\alpha}(t)$ to $\pi \circ \tilde{\alpha}(t)$. By $\mathbb{Z}$-equivariance, $\tilde{\gamma}_t$ descends to a homotopy from α to a path β in the image $\text{Axis}(\alpha)$ in S. Since β is contained in the image of a geodesic line, it is homotopic to a geodesic curve. This proves existence.

For uniqueness, note that if α, β are homotopic geodesics in the same homotopy class, their lifts $\tilde{\alpha}$ and $\tilde{\beta}$ are geodesic lines that stay within a constant distance of one another. It follows that their endpoints on $\partial\mathbb{H}^2$ are equal and hence, by uniqueness of hyperbolic geodesics, $\tilde{\alpha} = \tilde{\beta}$, whence $\alpha = \beta$. □

Lecture 4: Intersections of simple closed curves

4.1 Simple closed curves and intersection numbers

A closed curve on a surface S is called *simple* if the map $\alpha : S^1 \to S$ is injective.

Essential simple closed curves play a role that's analogous to basis vectors in linear algebra. A couple of basic facts about simple closed curves will make it much easier to work with them. As with homeomorphisms of S, rather than working up to homotopy, we want to use the more refined notion of *isotopy*.

Definition 4.1. An *isotopy* between two simple closed curves α_0, α_1 is a homotopy α_t between them so that each α_t is a simple closed curve. An *ambient isotopy* from α_0 to α_1 is an isotopy $\phi_t : S \to S$ so that $\phi_0 = \text{id}_S$ and $\alpha_1 = \phi_1 \circ \alpha_0$.

A priori, isotopies are harder to construct than homotopies. Fortunately, the two notions turn out to be equivalent in this case.

Lemma 4.2. *Two essential simple closed curves α_0, α_1 on a surface S are homotopic (relative to ∂S) if and only if they are ambient isotopic.*

The proof of this lemma is deferred till after we have the bigon criterion (5.1 below).

There are no essential simple closed curves in the sphere or the disc, and only two isotopy classes in the annulus, so the first non-trivial case is the torus. The following result, classifying simple closed curves on the torus, is an easy exercise that provides good motivation. An element h of $\pi_1 S$ is called *primitive* if it is not a proper power h^n for some $n > 1$.

Lemma 4.3. *Let T^2 be the torus. Homotopy classes of essential simple closed curves on T^2 correspond bijectively to the primitive elements of $\pi_1 T^2 \cong \mathbb{Z}^2$.*

Proof. The proof is left as an exercise; see Example Sheet 1, question 8. □

In order to apply the work on geodesic representatives from the last lecture to hyperbolic surfaces, we need the following lemma.

Lemma 4.4. *Let S be an orientable hyperbolic surface, $\alpha : S^1 \to S$ an essential simple closed curve and $\gamma : S^1 \to S$ the unique geodesic representative of its homotopy class. Then γ is also simple.*

Proof. Consider the unwrapped lift $\tilde{\alpha}$ of α to $\mathbb{H}^2$, the image of a copy of $\mathbb{R}$. Covering-space theory implies that α is simple if and only if, for any $g \in \pi_1 S$, $g\tilde{\alpha}$ is disjoint from $\tilde{\alpha}$ unless $g \in \langle \alpha \rangle$. Therefore, if the geodesic representative γ is not simple, there is some $g \notin \langle \gamma \rangle = \langle \alpha \rangle$ such that $g\text{Axis}(\gamma)$ intersects $\text{Axis}(\gamma)$ (recalling that $\text{Axis}(\gamma) = \tilde{\gamma}$). Since $\tilde{\gamma}$ and $\tilde{\alpha}$ have the same endpoints in the boundary at infinity $\partial_\infty \mathbb{H}^2$, this is only possible if g preserves $\text{Axis}(\gamma)$. But now it's easy to see that $g\tilde{\alpha}$ cannot be disjoint from $\tilde{\alpha}$, since g acts by orientation-preserving isometries. □

For hyperbolic surfaces, there's no easy algebraic characterisation of simple closed curves. However, the following necessary condition is useful.

Lemma 4.5. *If α is an essential simple closed curve on S then α represents a primitive element of $\pi_1 S$. Furthermore, if S is hyperbolic then the centraliser $C(\alpha) = \langle \alpha \rangle$.*

Proof. By Lemma 4.3, it suffices to consider the hyperbolic case. Note that the statement about centralisers implies being primitive. By Lemma 3.2, we may assume that α is geodesic. Recall that the action of α on $\mathbb{H}^2$ preserves the geodesic line $\mathrm{Axis}(\alpha)$, and that this line consists of those points that are moved precisely $\tau(\alpha)$.

Suppose that $g \in C(\alpha)$. For any chosen point $x \in \mathrm{Axis}(\alpha)$,

$$d(g(x), \alpha g(x)) = d(x, g^{-1}\alpha g(x)) = d(x, \alpha(x)) = \tau(\alpha)$$

so $g(x)$ is also moved $\tau(\alpha)$, and hence $g(x) \in \mathrm{Axis}(\alpha)$. Therefore, $C(\alpha)$ also preserves the geodesic line $\mathrm{Axis}(\alpha)$.

By the freeness and proper discontinuity of the action, the quotient $C(\alpha)\backslash\mathrm{Axis}(\alpha)$ is a circle. The map $\alpha : S^1 \to S$ now factors as

$$\langle\alpha\rangle\backslash\mathrm{Axis}(\alpha) \to C(\alpha)\backslash\mathrm{Axis}(\alpha) \to S$$

where the first map is a covering map. But α is injective, so $C(\alpha) = \langle\alpha\rangle$, as required. □

It is often useful to put an inner product on a vector space to check if two vectors are linearly independent. Pursuing the analogy with linear algebra, there is a structure on S that resembles a naturally defined inner product – the intersection number. (More precisely, intersection number resembles a symplectic form.)

Definition 4.6. Let α, β be closed curves on a surface S. Their *(geometric) intersection number* is

$$i(\alpha, \beta) = \min_{\alpha' \simeq \alpha\,,\,\beta' \simeq \beta} \#(\alpha' \cap \beta')\,.$$

In particular, if α is homotopic to a simple closed curve then $i(\alpha, \alpha) = 0$.

If $\alpha \cap \beta$ is finite and, at every intersection, each curve locally separates the other curve, then we say that α and β are *transverse*. Two curves can always be made transverse by a small isotopy. We say that two curves α, β are in *minimal position* if $\#(\alpha \cap \beta) = i(\alpha, \beta)$.

4.2 Bigons

In order to compute intersection number, we need to be able to put curves into minimal position. We do this using the *bigons*.

Definition 4.7. Let α, β be transverse simple closed curves on S. A *bigon* is an embedded disc $D^2 \hookrightarrow S$ such the interior is disjoint from $\alpha \cup \beta$ and the boundary ∂D^2 decomposes as a union of closed arcs $a \cup b$, where $a \subseteq \alpha$ and $b \subseteq \beta$.

Evidently, if there exists a bigon then α and β are not in minimal position. The bigon criterion asserts that the absence of bigons is also sufficient to be in minimal position. First, we need a lemma.

Lemma 4.8. *If α, β are transverse, essential simple closed curves on an orientable surface S without bigons, then in the universal cover $\widetilde{S}$, every pair of lifts $\tilde{\alpha}, \tilde{\beta}$ intersect in at most one point.*

Proof. Since α, β are simple, the lifts $\tilde{\alpha}$ and $\tilde{\beta}$ are embeddings in $\widetilde{S}$. Suppose that $\tilde{\alpha}, \tilde{\beta}$ intersect in two points. Then there are subarcs of $\tilde{\alpha}$ and $\tilde{\beta}$ that bound a disc D_0 in $\widetilde{S}$ (since $\widetilde{S}$ is homeomorphic to the sphere or the plane). The preimages of α and β give a cellular decomposition of D_0 into discs. It follows that there is an innermost disc $D \subseteq D_0$, with one boundary arc a contained in the preimage of α, and one b contained in the preimage of β.

It remains to prove that the covering map $\widetilde{S} \to S$ is injective on D. Equivalently, we need to prove that, for any $\phi \in \pi_1 S \smallsetminus 1$, $\phi(D) \cap D = \varnothing$.

If $\phi(D)$ intersects D then either $\phi(D)$ contains D or $\phi(\partial D)$ intersects D. In the former case, ϕ^{-1} sends D to D, so has a fixed point by the Brouwer fixed-point theorem, and hence $\phi = 1$ by freeness of the action of the fundamental group, contradicting the assumption.

Therefore, $\phi(\partial D)$ intersects D, but the interior of D contains no lifts of α or β, so $\phi(\partial D)$ intersects ∂D. The deck transformation ϕ sends lifts of α to lifts of α and lifts of β to lifts of β, so either $\phi(\tilde{\alpha})$ intersects $\tilde{\alpha}$ or $\phi(\tilde{\beta})$ intersects $\tilde{\beta}$; without loss of generality, the former occurs. Because α is simple, if $\phi(\tilde{\alpha})$ intersects $\tilde{\alpha}$ then $\phi \in \langle \alpha \rangle$ (after choosing suitable base points) and preserves $\tilde{\alpha}$.

Now ϕ also preserves the set of intersection points of α with the lifts of β. Therefore, writing x_+ and x_- for the two intersection points on ∂D, we have that either ϕ fixes x_+ or x_-, in which case $\phi = 1$ by the freeness of the action, or, without loss of generality, $\phi(x_-) = x_+$. This last case leads to contradiction, because ϕ is orientation-preserving, but α and β cross with opposite orientations at x_- and x_+. □

Lecture 5: The bigon criterion and topological type

5.1 The bigon criterion

We are now ready to state and prove the bigon criterion.

Proposition 5.1 (Bigon criterion)**.** *Transverse, essential, simple closed curves α, β on a surface S are in minimal position if and only if there are no bigons.*

Proof. One direction is obvious: if there is a bigon, then there is a homotopy that reduces $\#(\alpha \cap \beta)$ by 2.

We prove the reverse direction in the closed hyperbolic case, and leave it to the reader to adapt the argument to the other cases. Suppose that there are no bigons.

Fix a lift $\tilde{\alpha}$ of α to $\mathbb{H}^2$. Consider our fixed lift $\tilde{\alpha}$ of α, and all the lifts $\tilde{\beta}_i$ of β that intersect it. The natural action of $\mathbb{Z}$ on $\tilde{\alpha}$ extends to a map on the $\tilde{\beta}_i$ and, since each lift intersects $\tilde{\alpha}_i$ at most once by Lemma 4.8, the set $\alpha \cap \beta$ is in bijection with the number of $\mathbb{Z}$-orbits of the $\tilde{\beta}_i$. Therefore, to prove the proposition, we need to show that modifying α and β by homotopies doesn't alter whether or not a given pair of lifts $\tilde{\alpha}$ and $\tilde{\beta}$ intersect.

Since S is closed, α acts as a hyperbolic isometry of $\mathbb{H}^2$, and the limits $\xi_\pm = \lim_{t \to \pm\infty} \tilde{\alpha}(t)$ are equal to the endpoints of the axis $\mathrm{Axis}(\alpha)$. Likewise, a lift $\tilde{\beta}$ of β has the same endpoints $\eta_\pm$ as $\mathrm{Axis}(\beta)$.

If $\xi_+ = \eta_+$ and $\xi_- = \eta_-$ then $\tilde{\alpha}$ and $\tilde{\beta}$ are disjoint. Indeed, in this case α, β share a common axis and the isometry α acts on the set of intersections $\tilde{\alpha} \cap \tilde{\beta}$, so if there is one intersection then there are infinitely many, contradicting Lemma 4.8.

Suppose now that $\xi_+ = \xi_-$ but $\eta_+ \neq \eta_-$. Then without loss of generality, we may assume that $\xi_\pm = \infty$ in the upper half-plane model, and a direct computation shows that the commutator $[\alpha, \beta]$ is a non-trivial parabolic element of $\pi_1 S$. This contradicts the claim that S is closed.

In summary, we have seen that if two lifts $\tilde{\alpha}$, $\tilde{\beta}$ intersect, then their endpoints $\xi_\pm$ and $\eta_\pm$ are distinct. Next, note that the parity of $\#(\tilde{\alpha} \cap \tilde{\beta})$ is determined by the pattern of the points $\{\xi_\pm, \eta_\pm\}$ on the circle $\partial\mathbb{H}^2$: if the pair $\{\xi_+, \xi_-\}$ are in different components of $S^1 \smallsetminus \{\eta_+, \eta_-\}$ then the parity is odd; otherwise, the parity is even. Since the lifts intersect at most once, it follows that the arrangement of the endpoints $\{\xi_\pm, \eta_\pm\}$ on $\partial\mathbb{H}^2$ determines whether or not $\tilde{\alpha}$ and $\tilde{\beta}$ intersect.

Changing α by a homotopy doesn't change the endpoints $\xi_\pm$. Indeed, a homotopy $\alpha_\bullet : S^1 \times I \to S$ lifts to a continuous, $\mathbb{Z}$-equivariant map $\tilde{\alpha}_\bullet : \mathbb{R} \times I \to \mathbb{H}^2$. Therefore, if α_1 is homotopic to $\alpha_0 = \alpha$ then the corresponding lift $\tilde{\alpha}_1$ remains within a bounded neighbourhood of $\tilde{\alpha}_0$, and so has the same endpoints.

This completes the proof: changing α and β by homotopies doesn't change the endpoints of their lifts; hence, the same pairs of lifts cross, and so the number of intersection points remains the same. □

Most importantly, this gives us an effective way to see that simple closed curves are in minimal position. It also follows that geodesic representatives are always in minimal position.

Corollary 5.2. *If α, β are distinct simple closed geodesics on a hyperbolic surface S then they are in minimal position.*

Proof. Suppose that $D \to S$ is a bigon for α and β. Since D is simply connected, it lifts to an embedding into the universal cover $D \hookrightarrow \mathbb{H}^2$, bounded by a pair of geodesic lifts $\tilde{\alpha}, \tilde{\beta}$. But there is a unique geodesic on $\mathbb{H}^2$ between the corners of D, so $\tilde{\alpha} = \tilde{\beta}$ and so $\alpha = \beta$. □

5.2 The annulus criterion

It follows from the bigon criterion that homotopic simple closed curves can be made disjoint by an isotopy. We still need to analyse homotopies of disjoint simple closed curves.

Proposition 5.3 (Annulus criterion)**.** *Let α, β be disjoint simple closed curves on a surface S. If α and β are homotopic then α and β bound an embedded annulus in S.*

Proof. Again, we prove this under the assumption that S is a closed, hyperbolic surface and that α and β are essential, leaving the remaining cases to the reader.

Fix a lift $\tilde{\alpha}$ of α to the universal cover $\mathbb{H}^2$. Lifting the homotopy defines a lift $\tilde{\beta}$ of β to $\mathbb{H}^2$, disjoint from $\tilde{\alpha}$ but remaining within a bounded neighbourhood. It follows that $\tilde{\alpha}$ and $\tilde{\beta}$ limit to the same points $\xi_+, \xi_- \in \partial\mathbb{H}^2$. The union of $\tilde{\alpha}$, $\tilde{\beta}$ and $\{\xi_+, \xi_-\}$ forms an embedded circle in $\overline{\mathbb{H}}^2$, which bounds a topological disc $R \subseteq \mathbb{H}^2$. The natural action of $\mathbb{Z} = \langle \alpha \rangle$ preserves $\tilde{\alpha}$ and $\tilde{\beta}$ and hence R. The quotient $A = \mathbb{Z} \backslash R$ is a surface with two boundary components and fundamental group $\mathbb{Z}$, hence is an annulus.

It remains to prove that A embeds in S, or, equivalently, that any covering transformation $g \in \pi_1 S$ such that $g(R) \cap R \neq \emptyset$ is in $\langle \alpha \rangle$. Since α and β are simple and disjoint, $g\tilde{\alpha}$ is disjoint from $\tilde{\beta}$ and either disjoint from or equal to $\tilde{\alpha}$. The same is true, *mutatis mutandis*, for $g\tilde{\beta}$. Therefore, if $g(R)$ intersects R, it follows that g must preserve the set $\{\xi_+, \xi_-\}$, and hence the axis $\mathrm{Axis}(\alpha)$. Unless $g = 1$, it follows that $\mathrm{Axis}(\alpha) = \mathrm{Axis}(g)$, whence $g \in C(\alpha) = \langle \alpha \rangle$ by Lemma 4.5. □

We can now prove that homotopic simple closed curves are ambient isotopic.

Proof of Lemma 4.2. Evidently, if two curves are ambient isotopic then they are homotopic; we prove the converse.

Suppose then that α, β are homotopic. After an isotopy, we may assume that they are also transverse. Since every simple closed curve has intersection-number zero with itself, it follows that $i(\alpha, \beta) = 0$. The bigon criterion implies that every pair of transverse curves can be put into minimal position by an ambient isotopy. We may therefore assume that α, β are disjoint. By Proposition 5.3, it follows that α and β bound an embedded annulus. Pushing across this annulus defines an ambient isotopy taking α to β. □

5.3 Topological type

One of the most useful facts in linear algebra is that any two bases are related by an invertible linear map. To justify the analogy with linear algebra, we would like to understand when two essential simple closed curves are related by a homeomorphism. This turns out to be an easy consequence of the classification of surfaces.

Definition 5.4. Any simple closed curve α on a surface S has a small annular neighbourhood $\overline{N}(\alpha) \cong S^1 \times [-1, 1] \subseteq S$, where α is the core curve $S^1 \times \{0\}$. The *cut surface* associated to α is the surface

$$S_\alpha := S \smallsetminus N(\alpha)$$

obtained by removing the interior $N(\alpha)$ of the annulus. Removing $N(\alpha)$ introduces two new boundary circles $\alpha_+, \alpha_- : S^1 \to \partial S_\alpha$; canonically, we may take α_+ to be the one for which the induced orientation from S agrees with the orientation of α, and α_- to to be the one for which the two orientations disagree. Finally, we may recover S by gluing an annulus $A = S^1 \times I$ along the curves α_- and α_+:

$$S = S_\alpha \cup_{\alpha_+ \sqcup \alpha_-} A\,.$$

The cut surface tells us how to classify simple closed curves.

Definition 5.5. The *topological type* of a simple closed curve α on a surface S is the homeomorphism-type of the cut surface S_α. If S_α is connected then α is called *non-separating.*

Example 5.6. Let S be the closed surface of genus g and α an essential simple closed curve on S. The cut surface S_α has

$$\chi(S_\alpha) = \chi(S) + \chi(S^1) = \chi(S) = 2 - 2g$$

and 2 boundary components. If α is non-separating then, by the classification of surfaces, S_α is homeomorphic to $S_{g-1,2}$ by the classification of surfaces, so all non-separating curves have the same type. If α is separating then, by considering Euler characteristic, we see that

$$S_\alpha \cong S_{k,1} \sqcup S_{g-k,1}$$

for some $k \leq g$. Therefore, there are $\lfloor g/2 \rfloor$ topological types of separating curves.

Lecture 6: Change of coordinates and the Alexander lemma

6.1 Change of coordinates

In fact, topological type is a complete invariant for mapping class group orbits of curves: the classification of surfaces implies that curves have the same type exactly when they are related by a diffeomorphism of the surface. This principle is called *change of coordinates.*

Proposition 6.1 (Change of coordinates)**.** *Two essential simple closed curves α, β on a surface S have the same topological type if and only if there is an orientation-preserving homeomorphism $\phi : S \to S$ that fixes ∂S, with $\phi \circ \alpha = \beta$.*

Proof. One direction is obvious: if ϕ exists then it induces a homeomorphism $S_\alpha \to S_\beta$, so α and β are of the same topological type.

For the other direction, we suppose that a homeomorphism $\phi : S_\alpha \to S_\beta$ exists. The proof consists of making successive modifications to ϕ to make it of the form that we want.

First, note that every orientable surface double covers a non-orientable surface. Hence, S_β admits an orientation-reversing self-homeomorphism, and so we may assume that ϕ is orientation-preserving.

Second, note that $\text{Homeo}^+(S_\beta)$ acts as the full symmetric group on the set of boundary components of each path component of S_β. Hence, we may assume that ϕ preserves the components of ∂S, and sends $\alpha_\pm$ to $\beta_\pm$, with the correct orientations.

Our penultimate task is to extend ϕ across the gluing annuli to recover a homeomorphism $S \to S$. Indeed, note that $\phi \circ \alpha_-$ and $\phi \circ \alpha_+$ are disjoint curves in S, both homotopic to β, and hence to each other. Therefore, by Proposition 5.3, they bound an embedded annulus in S, which because of orientation must be disjoint from the interior of S_β, and so we may extend ϕ across the gluing annulus of S_α to a homeomorphism $S \to S$.

Finally, $\phi \circ \alpha$ and β are homotopic, hence ambient isotopic by Lemma 4.2, so we may compose ϕ with a homeomorphism to ensure that $\phi \circ \alpha = \beta$. $\square$

This makes topological constructions with arbitrary curves much easier: after identifying the topological type, we may apply a homeomorphism and reduce to our favourite representative. The next corollary is a nice example of this.

Corollary 6.2. *If α is a non-separating simple closed curve on S then there is a non-separating simple closed curve β with $i(\alpha, \beta) = 1$.*

The same idea applies to pairs of curves.

Proposition 6.3. *Suppose that α_1, β_1 and α_2, β_2 are pairs of simple closed curves on a surface S with $i(\alpha_1, \beta_1) = i(\alpha_2, \beta_2) = 1$. Then there is a homeomorphism $\phi : S \to S$ with $\phi(\alpha_1) = \alpha_2$ and $\phi(\beta_1) = \beta_2$.*

Proof. The curve β_1 descends to an arc on S_{α_1} with one endpoint on each of the $\alpha_{1\pm}$. Cutting along this arc gives a surface S_{α_1,β_1} with one additional boundary component γ_1 and four marked points, corresponding to the preimages of $\alpha_1 \cap \beta_1$. We can construct S_{α_2,β_2} similarly, including a boundary component γ_2 with four marked points, and the classification of surfaces again gives a diffeomorphism $\phi : S_{\alpha_1,\beta_1} \to S_{\alpha_2,\beta_2}$. As in the proof of Proposition 6.1, we may assume that ϕ sends γ_1 to γ_2. Furthermore, after modifying ϕ by an isotopy, we may assume that ϕ sends the gluing relation on γ_1 to the gluing relation on γ_2. Therefore, ϕ descends to S, and the result follows. □

6.2 The Alexander lemma

Computations of mapping class groups need to start with the simplest example. Consider the closed disc $D^2 = S_{0,0,1}$.

Lemma 6.4 (The Alexander lemma)**.** *The mapping class group of the closed disc,* $\mathrm{Mod}(D^2)$*, is trivial.*

Proof. Let $f : D^2 \to D^2$ be a homeomorphism that restricts to the identity on $\partial D^2 = S^1$. Then

$$f_t(x) = \begin{cases} (1-t)f(x/(1-t)) & 0 \le |x| < 1-t \\ x & 1-t \le |x| \le 1 \end{cases}$$

defines an isotopy between $f = f_0$ and the identity $\mathrm{id}_{D^2} = f_1$. □

Since every f_t leaves the origin fixed, the proof give the same result for the punctured disc $D^2_* = S_{0,1,1}$.

Lemma 6.5. *The mapping class group of the punctured disc,* $\mathrm{Mod}(D^2_*)$*, is trivial.*

6.3 Spheres with few punctures

Next, we will compute the mapping class group of the sphere with 0,1,2 or 3 punctures: $S_{0,n}$ for $n \le 3$. Since there are no essential curves on these surfaces, it is useful instead to work with *arcs*.

Definition 6.6. A *(proper) arc* is a continuous (or smooth) map $\alpha : [0,1] \to S$ so that $\alpha(0)$ and $\alpha(1)$ are either punctures or on ∂S, and $\alpha(0,1)$ is contained in the interior of S. A proper arc is *simple* if it is an embedding on $(0,1)$, and *essential* unless it is homotopic (rel. endpoints) into a puncture. (Note that homotopies are required to hold punctures fixed.)

Many of our previous results about curves also apply to arcs. Homotopies of arcs are always relative to endpoints. Arcs are homotopic if and only if they are isotopic, and we can put transverse arcs into minimal position using the bigon criterion. As with curves, we may cut a surface S along a simple arc α, and S_α denotes the resulting cut surface.

Next, we will characterise the essential arcs on a 3-punctured sphere.

Lemma 6.7. *Let α, β be simple arcs on the 3-punctured sphere with distinct endpoints. If α, β have the same endpoints then they are isotopic.*

Proof. Putting the third puncture at ∞, we may take α, β to be arcs between two points in the plane. After an isotopy, we may make them transverse. If they intersect then an innermost disc argument as in Lemma 4.8 exhibits a bigon. Isotoping one of the curves over this bigon, we can reduce the number of intersections. Therefore, we may assume that α, β are disjoint. Their union is now an embedded circle in the plane, which bounds a disc. Hence, they are isotopic. □

Let $\mathrm{Sym}(n)$ be the symmetric group on n elements. A homeomorphism ϕ of S permutes the n punctures of S, leading to a surjective homomorphism $\mathrm{Mod}(S) \to \mathrm{Sym}(n)$.

Proposition 6.8. *Let $S = S_{0,3}$, the 3-punctured sphere. The natural homomorphism $\mathrm{Mod}(S_{0,3}) \to \mathrm{Sym}(3)$ is an isomorphism.*

Proof. We may identify $S = S_{0,3}$ with $\mathbb{C} \smallsetminus \{0,1\}$. We only need to prove that $\mathrm{Mod}(S_{0,3}) \to \mathrm{Sym}(3)$ is injective. Suppose, therefore, that ϕ is a self-diffeomorphism of $\mathbb{C} \smallsetminus \{0,1\}$ that fixes 0,1 and ∞.

Let α be a simple, smooth arc from 0 to 1. The composition $\phi \circ \alpha$ is also a simple, smooth arc from 0 to 1, so Lemma 6.7 is ambient isotopic to α. Therefore, after an isotopy, we may assume that ϕ fixes α, and so descends to a diffeomorphism $\bar{\phi}$ of the cut surface S_α. But S_α is a punctured disc, and so $\bar{\phi}$ is isotopic to the identity relative to the boundary, by Lemma 6.5. Regluing the boundary, this isotopy descends to an isotopy from ϕ to the identity, as required. □

This result quickly implies the same result for the other low-complexity surfaces.

Corollary 6.9. *If S is either the 2-sphere $S^2 = S_{0,0}$, or the plane $\mathbb{C} = S_{0,1}$, then $\mathrm{Mod}(S)$ is trivial. If $S = \mathbb{C}_* = S_{0,2}$, then $\mathrm{Mod}(S) \cong \mathbb{Z}/2\mathbb{Z}$.*

Proof. In each case, the result asserts that the natural map $\mathrm{Mod}(S) \to \mathrm{Sym}(n)$ is injective, where n is the number of punctures. Suppose, therefore, that $\phi : S \to S$ fixes the punctures. Since $PSL_2(\mathbb{C})$ is connected and acts 3-transitively on the 2-sphere, we may modify ϕ by an isotopy and assume that ϕ fixes three points. Proposition 6.8 now implies that ϕ is isotopic to the identity, as required. □

Lecture 7: Infinite mapping class groups

7.1 The annulus

The *annulus* $A = S^1 \times [0,1] = S_{0,0,2}$ is the first surface we will see for which the mapping class group is infinite. As in Proposition 6.8, an important idea in the proof is to consider the image of a suitable arc.

Proposition 7.1. *For the annulus A, $\mathrm{Mod}(A) \cong \mathbb{Z}$.*

Proof. Identify S^1 with the unit circle in $\mathbb{C}$. The universal cover $\widetilde{A}$ of A is homeomorphic to the infinite strip $[0,1] \times \mathbb{R}$, with covering map $\widetilde{A} \to A$ sending $(x,y) \mapsto (x, \exp(2\pi i y))$.

Consider a homeomorphism $\phi : A \to A$ that restricts to the identity on ∂A. Let $\tilde{\phi}$ be the unique lift of ϕ to $\widetilde{A}$ that fixes the origin, and let $\tilde{\phi}_1$ denote its restriction to $\{1\} \times \mathbb{R}$. Note that $\tilde{\phi}_1$ is a lift of the identity on $S^1 \times \{1\}$, and so is translation by some integer n. The homotopy lifting lemma implies that modifying ϕ by a homotopy changes n continuously, and so n is constant (since $\mathbb{Z}$ is discrete). Therefore, we have a well-defined assignment $\mathrm{Mod}(A) \to \mathbb{Z}$ given by $[\phi] \mapsto n$. It remains to prove that this assignment is an isomorphism of groups.

If ψ is another self-homeomorphism of A relative to ∂A then uniqueness of lifts implies that $\widetilde{\phi \circ \psi} = \tilde{\phi} \circ \tilde{\psi}$, from which it follows immediately that $\mathrm{Mod}(A) \to \mathbb{Z}$ is a group homomorphism.

For each $n \in \mathbb{Z}$, the matrix

$$\tilde{\phi} = \begin{pmatrix} 1 & 0 \\ n & 1 \end{pmatrix}$$

defines a covering diffeomorphism $\widetilde{A} \to \widetilde{A}$ that descends to the identity on each boundary component, and such that $\tilde{\phi}_1$ is translation by n. This shows that $\mathrm{Mod}(A) \to \mathbb{Z}$ is surjective.

Let $\phi : A \to A$ be a self-homeomorphism relative to ∂A, such that $\tilde{\phi}$ fixes $(0,1)$. To prove injectivity, we need to show that ϕ is isotopic to the identity. Let δ be the arc in A defined by $\delta(t) = (t,1)$ and let $\tilde{\delta}$ be its lift starting at the origin. Both $\tilde{\delta}$ and $\tilde{\phi} \circ \tilde{\delta}$ end at $(0,1)$. After a small isotopy, we may assume that δ and $\phi \circ \delta$ are transverse. Since δ and $\phi \circ \delta$ have the same endpoints, Lemma 4.8 implies that δ and $\phi \circ \delta$ form a bigon. If the corners of that bigon are not $(1,0)$ and $(1,1)$ then we may apply an isotopy to ϕ and reduce the number of intersections. Otherwise, δ and $\phi \circ \delta$ together form a bigon, and so we may modify ϕ by an isotopy until it fixes δ.

We now conclude as before. Cutting along δ, ϕ defines a diffeomorphism $\bar{\phi}$ of the cut surface A_δ that fixes the boundary. By the Alexander lemma, $\bar{\phi}$ is isotopic to the identity, and so ϕ is too. □

The generator of $\mathrm{Mod}(A)$ is called a *Dehn twist*. In our analogy with $SL_n(\mathbb{Z})$, they play the role of elementary matrices. Since most surfaces contain many essential annuli, we shall see that they also usually contain many Dehn twists.

7.2 The punctured torus

The torus $T^2 = S^1 \times S^1$ is the first surface with a really interesting mapping class group. We shall see that $\mathrm{Mod}(T^2)$ is a familiar group, but it is also rich enough to give us a sense of what to expect from the mapping class groups of higher-genus surfaces.

Recall that the fundamental group of the torus is $\pi_1 T^2 \cong \mathbb{Z}^2$. We shall identify $\mathrm{Mod}(T^2)$ by making it act by automorphisms on $\pi_1 T^2$. Since π_1 requires a base point, it is more convenient to start with the punctured torus $T^2_* = S_{1,1,0}$, and to think of the puncture as a marked point. Let ϕ be a self-homeomorphism of T^2_*, which we can think of as a homeomorphism of T^2 that fixes the base point. Since it fixes the base point, it induces a based map from the torus to itself, and hence an automorphism of $\pi_1 T^2$, by the functoriality of π_1. So we have defined a group homomorphism

$$\mathrm{Mod}(T^2_*) \to \mathrm{Aut}(\mathbb{Z}^2) = GL_2(\mathbb{Z})\,.$$

Theorem 7.2. *For the once-punctured torus T^2_*, the natural homomorphism to $GL_2(\mathbb{Z})$ induces an isomorphism*

$$\mathrm{Mod}(T^2_*) \cong SL_2(\mathbb{Z})\,.$$

Proof. We have already seen that the map is a homomorphism. We shall show that it is injective, and surjects $SL_2(\mathbb{Z})$.

To show injectivity, we take a self-homeomorphism ϕ of T^2_* acting as the identity on $\pi_1 T^2$. Let α, β be the standard based loops in T^2 that generate the fundamental group. That is, $\alpha : [0,1] \to T^2$ sends $t \mapsto (\exp(2\pi i t), 1)$ and $\beta : [0,1] \to T^2$ sends $t \mapsto (1, \exp(2\pi i t))$. Since ϕ_* is the identity automorphism, $\phi \circ \alpha$ is based homotopic to α and so, the version of Lemma 4.2 for arcs, we may modify ϕ by an isotopy so that $\phi \circ \alpha = \alpha$. Therefore, ϕ induces a self-homeomorphism $\bar{\phi}$ of the cut surface T^2_α, which is an annulus.

On this annulus, the loop β is cut to becomes the standard arc δ that appeared in the proof of Proposition 7.1. Lifting the universal cover of the annulus to the universal cover of the torus gives a diagram

$$\begin{array}{ccc} [0,1]\times\mathbb{R} & \longrightarrow & \mathbb{R}^2 \\ \downarrow & & \downarrow \\ T^2_\alpha & \longrightarrow & T^2 \end{array}$$

which identifies the lifts of $\bar{\phi} \circ \delta$ and $\phi \circ \beta$ at the origin. But $\phi \circ \beta$ is based homotopic to β, so its endpoint is $(1,0)$; hence the same is true of $\bar{\phi} \circ \delta$. Therefore, $\bar{\phi}$ represents the trivial element of the mapping class group of the annulus, by the proof of Proposition 7.1. Thus $\bar{\phi}$ is isotopic to the identity on T^2_α and so, regluing the boundary, ϕ is isotopic to the identity on T^2, as claimed.

To identify the image of the homomorphism $\mathrm{Mod}(T^2) \to GL_2(\mathbb{Z})$, note that any $A \in GL_2(\mathbb{Z})$ preserves the integer lattice $\mathbb{Z}^2$ inside $\mathbb{R}^2$, and hence descends to a continuous self-map ϕ_A of the torus $T^2 = \mathbb{R}^2/\mathbb{Z}^2$. Since A is invertible, this map is a homeomorphism, with inverse $\phi_{A^{-1}}$. To identify the action on the fundamental group, note that the lifts at the origin satisfy

$$\widetilde{\phi_A \circ \alpha} = A.\tilde{\alpha} \text{ and } \widetilde{\phi_A \circ \beta} = A.\tilde{\beta}$$

whence $(\phi_A)_*$ acts as multiplication by A. In particular, every element of $\mathrm{Mod}(T^2)$ is represented by some ϕ_A.

Finally, note that, by considering the action of A on the fundamental square in $\mathbb{R}^2$, we see that ϕ_A acts on $H_2(T^2)$ as multiplication by $\det(A)$. Since ϕ_A is orientation-preserving if and only if this action is trivial, it follows that the image of $\mathrm{Mod}(T^2)$ is $SL_2(\mathbb{R})$. □

Lecture 8: The torus and the Alexander method

8.1 The torus

It turns out that there is a trick that enables us to pass from the punctured torus to the torus. The torus is unique among closed surfaces in that it carries a group structure (namely the quotient group structure on $\mathbb{R}^2/\mathbb{Z}^2$), which we can exploit.

Corollary 8.1. *For the torus T^2,*

$$\mathrm{Mod}(T^2) \cong SL_2(\mathbb{Z})\,.$$

Proof. Think of the puncture on T^2_* as a marked point, which we may take to be the identity 0 of T^2. Forgetting the marked point defines a natural homomorphism

$$\mathrm{Mod}(T^2_*) \to \mathrm{Mod}(T^2)$$

which we will show is an isomorphism. The result then follows from Theorem 7.2.

For any self-homeomorphism ϕ of T^2, choose a continuous path α from 0 to $\phi(0)$. Now,

$$\phi_t(x) = (\alpha(t))^{-1}\phi(x)$$

defines an isotopy from $\phi = \phi_0$ to a homeomorphism ϕ_1 that fixes 0. This shows that the homomorphism $\mathrm{Mod}(T^2_*) \to \mathrm{Mod}(T^2)$ is surjective.

Let ϕ_t be an isotopy between two homeomorphisms ϕ_0 and ϕ_1 of T^2 that each fix 0. Let $\beta(t) = \phi_t(0)$, a loop based at 0. Then

$$\phi'_t(x) = (\beta(t))^{-1}\phi_t(x)$$

is an isotopy from ϕ_0 to ϕ_1 that fixes 0 at every time t. This shows that the homomorphism $\mathrm{Mod}(T^2_*) \to \mathrm{Mod}(T_2)$ is injective. □

Remark 8.2. This is the classical *modular group*, which inspires the notation $\mathrm{Mod}(S)$.

8.2 Pairwise isotopy

A common theme in these computations is to analyse the action of a homeomorphism ϕ via its action on sets of curves or arcs in S. The bigon and annulus criteria together gives a practical method for checking whether or not pairs of curves are isotopic. We would like to extend this to sets of curves or arcs, so it would be convenient if it were true that pairwise isotopic sets were ambient isotopic. Unfortunately, this is not true in general: consider for instance three arcs intersecting pairwise and bounding a triangle. However, this turns out to be essentially the only obstruction.

Another subtlety is that we will need to work with curves and arcs *up to reparametrisaion*. Therefore, let's write $\alpha \approx \beta$ for a pair of curves or arcs that are equal up to an orientation-preserving homeomorphism of their domain.

The next lemma now enables us to determine easily when suitable sets of curves and arcs are isotopic.

Lemma 8.3. *Let S be a surface, with collections $\{\alpha_1, \ldots, \alpha_n\}$ and $\{\beta_1, \ldots, \beta_n\}$ of essential simple closed curves and simple proper arcs on S. Suppose that:*

(i) the $\{\alpha_i\}$ are pairwise in minimal position, and the same for the β_i (we say they have no bigons*);*

(ii) the $\{\alpha_i\}$ are pairwise non-isotopic, and the same for the β_i (we say they have no annuli*);*

(iii) for distinct i, j, k, at least one of $\alpha_i \cap \alpha_j$, $\alpha_j \cap \alpha_k$ and $\alpha_k \cap \alpha_i$ is empty, and the same for the β_i (we say they have no triangles*).*

If α_i is isotopic to β_i for each i then, after an ambient isotopy of S, we have $\alpha_i \approx \beta_i$ for each i.

Proof. The proof is by induction on n. The base case, $n = 1$, follows from Lemma 4.2. By induction on n, we may assume that $\alpha_i \approx \beta_i$ for $i < n$. It remains to show that we can isotope α_n to β_n, while keeping α_i identified with β_i for $i < n$.

If α_n and β_n are not disjoint then, since they are isotopic, they form a bigon D. By hypothesis (i), any $\alpha_i = \beta_i$ that intersects D cannot form a bigon with α_n or β_n; therefore it intersects both α_n and β_n. Furthermore, if $\alpha_i \approx \beta_i$ and $\alpha_j \approx \beta_j$ both intersect D, then hypothesis (iii) implies that they cannot intersect each other. Therefore, the bigon D is crossed by disjoint arcs of the $\alpha_i \approx \beta_i$, and so there is an isotopy of α_n that holds the remaining α_i identified with the β_i, but reduces the number of intersections of α_n and β_n. Therefore, after finitely many of these isotopies, we may assume that α_n and β_n are disjoint.

It follows that α_n and β_n together bound an annulus A in S. By hypothesis (ii), no $\alpha_i \approx \beta_i$ is contained in this annulus. By the same arguments as in the previous paragraph, the arcs of $\alpha_i \approx \beta_i$ that intersect A cross from α_n to β_n. It follows that there is an ambient isotopy taking α_n to β_n while holding the α_i identified with β_i for $i < n$. This completes the proof. □

8.3 The Alexander method

In our computations above, we used the Alexander lemma to certify that certain mapping classes were trivial. This argument generalises to a method that works on all surfaces. Farb and Margalit call it the *Alexander method.* The idea is that if a homeomorphism ϕ moves a 'filling' collection of arcs and curves α_i so that the image curves $\phi \circ \alpha_i$ are isotopic to the α_i, then in fact ϕ is isotopic to the identity.

Definition 8.4. Let S be a surface. A transverse collection of simple closed curves and simple proper arcs $\{\alpha_i\}$ is said to *fill* S if each component of the cut surface $S_{\{\alpha_i\}}$ is a disjoint union of discs and once-punctured discs. For such a collection α_i, the *structure graph* $\Gamma_{\{\alpha_i\}}$ is the graph obtained from $\bigcup_i \alpha_i \cup \partial S$ by placing vertices at the intersection points and punctures.

We are now ready to describe the Alexander method, which gives us a practical method for checking that a homeomorphism is isotopic to the identity.

Proposition 8.5 (The Alexander method)**.** *Let $\{\alpha_i\}$ be a collection of essential simple proper arcs and closed curves on S without bigons, annuli or triangles that fills S.*

(i) If $\phi \in \mathrm{Homeo}^+(S)$ has the property that, for some permutation $\sigma \in \mathrm{Sym}(n)$, $\phi_i \circ \alpha_i$ is isotopic to $\alpha_{\sigma(i)}$ for each i, then ϕ induces an automorphism ϕ_Γ of the structure graph $\Gamma_{\{\alpha_i\}}$.

(ii) If ϕ_Γ is trivial then ϕ is isotopic to the identity.

In particular, under the hypotheses of (i), ϕ has finite order in $\mathrm{Mod}(S)$*.*

Proof. By Lemma 8.3, after composing ϕ with an ambient isotopy, we may assume that ϕ preserves $\bigcup_i \alpha_i$, and hence induces an automorphism ϕ_Γ of $\Gamma_{\{\alpha_i\}}$. This proves (i).

Next, assume that ϕ_Γ is trivial. Because ϕ is orientation-preserving, it follows that ϕ also preserves the complementary regions adjacent to each edge of $\Gamma_{\{\alpha_i\}}$. By the Alexander lemma, ϕ may be modified by an isotopy on each component of the complement of $\Gamma_{\{\alpha_i\}}$ to be the identity on that component. This proves (ii). Finally, $\Gamma_{\{\alpha_i\}}$ is finite, so ϕ_Γ has finite order, and the final assertion follows from (ii). □

This gives a practical method for proving that a mapping class is trivial. Of course, to apply the method we need a suitable collection of curves and proper arcs. Note that, conversely, if $\phi(\alpha_i)$ is not isotopic to α_i for some i, then ϕ represents a non-trivial mapping class.

Exercise 8.6. Every surface S has a filling collection α_i of essential simple closed curves and proper arcs that satisfy the hypotheses of Lemma 8.3. See question 11 on Example Sheet 1.

Lecture 9: Dehn twists

Dehn twists in $\mathrm{Mod}(S)$ play the analogous role to elementary matrices in $SL_n(\mathbb{Z})$.

9.1 Definition and the action on curves

Let A be an annulus $I\times S^1$. Choose an orientation on the S^1 factor; an orientation on A then determines an orientation on the I factor. The proof of Proposition 7.1 tells us that the following homeomorphism generates $\mathrm{Mod}(A)\cong\mathbb{Z}$.

$$\tau(x, e^{2\pi i y}) = (x, e^{2\pi i(x+y)})\,.$$

There is some subtlety about the choices in this definition. The opposite choice of orientation on A would lead to the x coordinate being replaced by $1-x$, leading to the map τ^{-1} instead of τ. However, choosing the opposite orientation on the S^1 factor while keeping the orientation on A fixed leads to the opposite choice of orientation on I also; so y is replaced by $-y$, and x is replaced by $1-x$, and there is no change to τ. In summary, τ depends on the choice of orientation on A, but not on the orientation on the circle S^1. This generator is called a *left* Dehn twist. By embedding annuli in our surface S, we may use this to define many mapping classes of our surfaces.

Definition 9.1. Let α be a simple closed curve in S, and let N be a regular neighbourhood of α. Choose a homeomorphism $\iota : A \to N$, and pull back the orientation on S to an orientation on A. Consider the following homeomorphism of S.

$$\tau_\alpha(x) = \begin{cases} \iota \circ \tau \circ \iota^{-1}(x) & x \in N \\ x & \text{otherwise} \end{cases}$$

The *(left) Dehn twist in* α is the mapping class of τ_α, denoted by T_α.

Lemma 9.2. *The mapping class T_α is independent of the choices made in the definition of τ_α, and also only depends on the isotopy class of α.*

Proof. Fix an orientation on α. We have $\partial N = \alpha_+ \cup \alpha_-$, where α_+ and α_- are isotopic to α and distinguished from each other by the orientation on N. Let α' be an isotopic curve to α and N' a regular neighbourhood of α', with $\partial N' = \alpha'_+ \cup \alpha_-$. Then α_- and α'_- are isotopic, hence ambient isotopic, so we may assume they are equal. Working in the cut surface S_{α_-}, we see that α_+ and α'_+ are ambient isotopic, so we may assume that they too are equal. This gives that N and N' are equal as subsets of S. Now the twists τ_α and $\tau_{\alpha'}$ define the canonical generator of $\mathrm{Mod}(N)$ determined by the orientation. Hence they are isotopic, as required. □

We finish the section by explaining how to draw the result of Dehn twisting a curve. Let β be a simple closed curve on S, and let α intersect β transversely. To draw $T_\alpha^k(\beta)$ for some $k > 0$, we draw $k\#(\alpha \cap \beta)$ parallel copies of α, and then modify the resulting picture by *surgery*. In this case, since T_α is a left Dehn twist, the surgery turns left from β to α. Of course, there is no *a priori* guarantee that the resulting curve cannot be simplified.

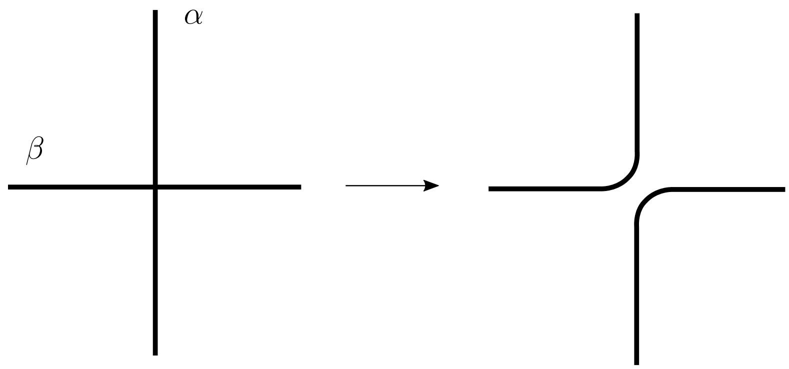

Figure 1: A surgery to produce a left Dehn twist

9.2 Order and intersection number

Proposition 7.1 implies that, in the case of an annulus, a Dehn twist has infinite order. To generalise that fact to more complicated surfaces, we will analyse how intersection numbers change when we Dehn twist.

Lemma 9.3. *Let α be an essential simple closed curve, and let β be any simple closed curve or simple proper arc. Then*

$$i(T_\alpha^k(\beta), \beta) = |k| i(\alpha, \beta)^2$$

for any $k \in \mathbb{Z}$.

Proof. After modifying β by an isotopy, we may assume that α and β are in minimal position, with $i(\alpha, \beta)$ intersection points. Let $\beta' = \tau_\alpha^k(\beta)$, modified by a small isotopy to make it transverse to β. As described above, β' can be constructed as follows: take one parallel copy β_0 of β (moved slightly to the left) and $|k| i(\alpha, \beta)$ parallel copies of α, and perform surgery. Since the surgeries do not introduce any new points of intersection, this shows that $\#(\beta', \beta) = |k| i(\alpha, \beta)^2$. Therefore, it remains to show that β and β' do not form any bigons, and hence are in minimal position.

Suppose therefore that β and β' form a bigon, bounded by arcs $b \subseteq \beta$ and $b' \subseteq \beta'$. Orientation considerations show that the two intersections of b and b' must have opposite orientations. Therefore, b' either both enters and leaves β from the left hand side, or both enters and leaves β from the right hand side.

If b' is on the right hand side of β then it follows that b' is entirely contained inside (some copy of) α, so α and β bound a bigon, contradicting the assumption that they were in minimal position.

If b' is on the left hand side then, if we return to the set-up in the first paragraph and push β_0 slightly to the right of β instead of to the left, we find again that b' is contained entirely inside some copy of α, which again leads to a contradiction. □

It follows easily that Dehn twists have infinite order.

Proposition 9.4. *If α is an essential simple closed curve on S then T_α has infinite order in* $\mathrm{Mod}(S)$*.*

Proof. The result is immediate from Lemma 9.3 together with the claim that there is some simple closed curve or proper arc β on S with $i(\alpha,\beta) > 0$. We construct β according to various cases. By Proposition 7.1, we may assume that S either has $g \geq 1$ or $n + b \geq 3$.

If α is non-separating then the existence of a suitable β follows immediarely from Corollary 6.2.

If α is a boundary component, then it can be taken to lie on an 3-holed sphere in S, and it is then easy to construct a suitable arc β with $i(\alpha,\beta) = 2$.

Finally, if α is separating and not homotopic to a boundary, then it lies on a four-punctured sphere $S_{0,4,0} \subseteq S$, dividing it into two twice-punctured discs. It is again easy to exhibit a simple closed curve β on S with $i(\alpha,\beta) = 2$, as required. □

Lecture 10: Multitwists and pairs of pants

10.1 Basic properties of Dehn twists

Lemma 10.1. *Let α, β be essential simple closed curves on S. The Dehn twists T_α, T_β are equal if and only if α and β are isotopic (up to reversing orientation).*

Proof. If α and β are isotopic then $T_\alpha = T_\beta$ by Lemma 9.2. Suppose therefore that α and β are not isotopic. By Lemma 9.3, it suffices to exhibit a curve or proper arc γ disjoint from β but with $i(\alpha,\gamma) > 0$. If $i(\alpha,\beta) > 0$ then we may take $\gamma = \beta$. Otherwise, after an isotopy, α is disjoint from β, and so is contained in a component Σ of the cut surface S_β. Since α and β are not isotopic, Σ is not a disc, once-punctured disc or annulus. It follows that Σ contains a simple closed curve or proper arc γ with $i(\alpha,\gamma) > 0$, disjoint from β. Regluing along β, we may take γ to be a curve or arc in Σ, and the result follows. □

Remark 10.2. If α is an essential simple closed curve and ϕ is a homeomorphism of S (which we conflate with its mapping class) then $\phi T_\alpha \phi^{-1} = T_{\phi\circ\alpha}$. By Lemma 10.1, T_α and T_β are conjugate if and only if there is a homeomorphism ϕ such that $\alpha^{\pm 1} = \phi \circ \beta$; i.e., if and only if α and β are of the same topological type.

These observations enable us to immediately characterise the centralisers of Dehn twists.

Lemma 10.3. *Let $\phi \in \mathrm{Mod}(S)$ and let α, β be curves on S. Then:*

(i) ϕ commutes with T_α if and only if $\phi \circ \alpha$ is isotopic to $\alpha^{\pm 1}$; and

(ii) T_α commutes with T_β if and only if $i(\alpha,\beta) = 0$.

Proof. Item (i) follows immediately Remark 10.2 and Lemma 10.1.

For (ii), it follows from (i) that T_α commutes with T_β if and only if $T_\beta(\alpha) = \alpha$. This is clearly true if $i(\alpha,\beta) = 0$, since α can be made disjoint from a regular neighbourhood of β by an isotopy. Conversely, if $T_\beta(\alpha) = \alpha$ then

$$0 = i(T_\beta(\alpha), \alpha) = i(\alpha,\beta)^2$$

by Lemma 9.3, and so $i(\alpha,\beta) = 0$ as required. □

10.2 Multitwist subgroups

We often find ourselves not just interested in simple closed curves, but in finite, disjoint collections of them.

Definition 10.4. A *multicurve* on S is a finite set of essential, pairwise non-isotopic, simple closed curves on S. We write $\alpha = \alpha_1 \sqcup \ldots \sqcup \alpha_n$. A mapping class of the form

$$T_{\alpha_1}^{k_1} \ldots T_{\alpha_n}^{k_n}$$

is called a *multitwist*.

Multitwists give natural examples of large abelian subgroups of mapping class groups.

Proposition 10.5. *If α is a multicurve, then the map $\mathbb{Z}^n \to \mathrm{Mod}(S)$ given by*

$$(k_1, \ldots, k_n) \mapsto T_{\alpha_1}^{k_1} \ldots T_{\alpha_n}^{k_n}$$

is an injective homomorphism.

Proof. Because the α_i are disjoint, the map is a homomorphism by Lemma 10.3.

To prove that it is injective, suppose without loss of generality that $k_1 \neq 0$. Consider the cut surface $S_{\alpha_2,\ldots,\alpha_n}$, and let S_0 be the component that contains α_1. Since α is a multicurve, α_1 is an essential simple closed curve in S_0 not parallel to an boundary components labelled $\alpha_2, \ldots, \alpha_n$. It follows that there is a simple closed curve or proper arc β on S_0, with endpoints not on $\alpha_2, \ldots, \alpha_n$, such that $i(\alpha_1, \beta) > 0$. Since β is disjoint from the α_i for $i > 1$ we have

$$T_{\alpha_2}^{k_2} \ldots T_{\alpha_n}^{k_n}(\beta) = \beta\,,$$

while $i(T_{\alpha_1}^{k_1}(\beta), \beta) \neq 0$ by Lemma 9.3. Hence $T_{\alpha_1}^{k_1} \ldots T_{\alpha_n}^{k_n}$ is non-trivial, as required. □

It follows immediately from Proposition 10.5 and Lemma 10.3 that surfaces with boundary have corresponding central subgroups.

Corollary 10.6. *If $S = S_{g,n,b}$, then the multitwist subgroup in the boundary multicurve forms a central subgroup of $\mathrm{Mod}(S)$ isomorphic to $\mathbb{Z}^b$.*

10.3 Pairs of pants

The surface $S_{0,0,3}$ is called the *pair of pants*. It plays an important role, since if we cut a closed surface up maximally along pairwise non-isotopic curves, the resulting pieces will all be pairs of pants. When there are punctures, we may also obtain the punctured annulus $S_{0,1,2}$ and the twice-punctured disc $S_{0,2,1}$. We now know that the mapping class groups of these surfaces contain Dehn twists. In this section, we will see that these account for (almost) the entire mapping class groups.

Theorem 10.7. *Let $S = S_{0,n,b}$ with $n + b = 3$. The kernel of the natural map $\mathrm{Mod}(S) \to \mathrm{Sym}(n)$ is equal to the multitwist subgroup in the boundary. In particular, $\mathrm{Mod}(S_{0,0,3}) \cong \mathbb{Z}^3$, $\mathrm{Mod}(S_{0,1,2}) \cong \mathbb{Z}^2$, and $\mathrm{Mod}(S_{0,2,1}) \cong \mathbb{Z} \times \mathbb{Z}/2$.*

Proof. We may choose a disjoint pair of simple proper arcs $\{\alpha_1, \alpha_2\}$ in S that satisfy the hypotheses of the Alexander method. Embed S into $S_{0,3,0}$ by capping boundary components off with once-punctured discs, and extend the α_i to arcs $\bar{\alpha}_i$ in $S_{0,3,0}$.

Let $\phi \in \mathrm{Mod}(S)$ lie in the kernel of the natural map to $\mathrm{Sym}(n)$, and let $\bar{\phi}$ be the result of extending ϕ to $S_{0,3,0}$ by the identity. Lemma 6.7 shows that each $\bar{\phi} \circ \bar{\alpha}_i$ is isotopic to $\bar{\alpha}_i$. Restricting the isotopy to S, it follows that $\phi \circ \alpha_i$ is isotopic to α_i, *where the isotopy is not required to restrict to the identity on* ∂S.

Let $\widehat{S}$ be obtained by doubling S along ∂S, and likewise double α_i and ϕ to simple arcs or curves $\hat{\alpha}_i$ and a homeomorphism $\hat{\phi}$ in $\widehat{S}$. Then each $\hat{\phi} \circ \hat{\alpha}_i$ is isotopic in $\widehat{S}$ to $\hat{\alpha}_i$, and so, after a small isotopy to make them transverse, if $\hat{\phi} \circ \hat{\alpha}_1$ is not disjoint from $\hat{\alpha}_1$ then there is a bigon D in $\widehat{S}$ bounded by subarcs of $\hat{\alpha}_1$ and $\hat{\phi} \circ \hat{\alpha}_1$.

Consider the intersection of D with the multicurve ∂S in $\widehat{S}$, which we may assume to be transverse. If this intersection is empty, then the bigon D is contained in S, and so we may modify ϕ by an isotopy and reduce the number of intersections between α_1 and $\phi \circ \alpha_1$. If the intersection is non-empty, then a 'corner' of D gives us a *half-bigon* Δ: an innermost disc bounded by subarcs of α_1, $\phi \circ \alpha_1$, and some component δ of ∂S. If there is a half-bigon, the surgery description of Dehn twists shows that, for some Dehn twist $\tau_\delta^{\pm 1}$ we have that

$$\#(\alpha_1 \cap \tau_\delta^{\pm 1} \circ \phi \circ \alpha_1) = \#(\alpha_1 \cap \phi \circ \alpha_1) + 1\,,$$

but $\tau_\delta^{\pm 1} \circ \phi \circ \alpha_1$ and α_1 together bound a bigon, and so we may modify ϕ by a homotopy to reduce the total intersection by 2. By induction, after composing ϕ woth some τ in the boundary multitwist subgroup of S, we may assume that $\psi = \tau \circ \phi$ fixes α_1.

We finish as usual by a cut-and-paste argument. The homeomorphism ψ defines a self-homeomorphism $\bar{\psi}$ of the cut surface S_{α_1}, which has Euler characteristic 0 and at least one boundary component. If S_{α_1} is a punctured disc then $\bar{\psi}$ is isotopic to the identity by the Alexander lemma. Otherwise, S_{α_1} is an annulus, and so $\mathrm{Mod}(S_{\alpha_1}) = \langle T_\epsilon \rangle$ where ϵ is isotopic to the boundary component of S not touched by α_1. Thus, there is $k \in \mathbb{Z}$ such that $\tau_\epsilon^k \circ \bar{\psi}$ fixes α_2 up to isotopy. Regluing, $\tau_\epsilon^k \circ \psi = \tau_\epsilon^k \circ \tau \circ \phi$ fixes both α_1 and α_2 up to isotopy, and so $\phi = \tau^{-1} \circ \tau_\epsilon^{-k}$, as required. □

Lecture 11: Subsurfaces

11.1 The inclusion homomorphism

We have now computed the mapping class groups of all but one of the the oriented finite-type surfaces of Euler characteristic at least -1. (For the one-holed torus $S_{1,0,1}$, see question 11 on problem sheet 2.) The mapping class groups of more complicated surfaces will not usually be isomorphic to well-known groups.

Instead, we will analyse them inductively by relating $\mathrm{Mod}(S)$ to the mapping class groups of the subsurfaces of S. A connected subsurface $\Sigma \subseteq S$ is called *essential* if the induced homomorphism $\pi_1\Sigma \to \pi_1 S$ is injective. By the Seifert–van Kampen theorem, this is equivalent to requiring that no component of the complement $S-\Sigma$ is an open disc, or equivalently, that any simple closed curve in Σ that bounds a disc in S bounds a disc in Σ.

Let Σ be a subsurface of S. There is a natural continuous homomorphism

$$\mathrm{Homeo}^+(\Sigma, \partial\Sigma) \to \mathrm{Homeo}^+(S, \partial S)\,,$$

given by extending each homeomorphism of Σ by the identity on $S - \Sigma$.

Definition 11.1. The induced homomorphism $\iota : \mathrm{Mod}(\Sigma) \to \mathrm{Mod}(S)$ is called the *inclusion homomorphism.*

The first goal of this subsection is to analyse the inclusion homomorphism, and in particular to prove that it is often injective. We need a lemma that helps us to promote isotopies in S to isotopies in Σ.

Lemma 11.2. *Let Σ be a closed,*[4] *essential subsurface in S, and let α, β be essential simple closed curves in Σ that are not isotopic into $\partial\Sigma$. If α is isotopic to β in S then α is isotopic to β in Σ.*

Proof. Make α and β transverse by a small isotopy. As usual, we may reduce $\#(\alpha \cap \beta)$ by removing bigons in S, but since Σ is essential, each bigon is contained in Σ. Therefore, we may assume that α and β are disjoint, and so bound an annulus in S. But since neither α nor β is homotopic into $\partial\Sigma$, this annulus is contained in Σ, which completes the proof. ☐

[4]'Closed' in the sense that the complement is open.

With this lemma in hand, we can completely describe the kernel of the inclusion homomorphism.

Theorem 11.3. *Let S be a surface of finite type, and Σ a connected, essential, closed subsurface. Let $\alpha_1, \ldots, \alpha_m$ be the components of $\partial\Sigma$ that bound punctured discs in $S - \Sigma$. Let $\beta_1^{\pm}, \ldots, \beta_n^{\pm}$ be the components of $\partial\Sigma$ that bound annuli in $S - \Sigma$. The kernel of the inclusion homomorphism*

$$\iota : \mathrm{Mod}(\Sigma) \to \mathrm{Mod}(S)$$

is the central free abelian subgroup

$$\langle T_{\alpha_1}, \ldots, T_{\alpha_1}, T_{\beta_1^+} T_{\beta_1^-}^{-1}, \ldots, T_{\beta_n^+} T_{\beta_n^-}^{-1} \rangle \, .$$

Proof. We will prove that an element ϕ of the kernel of ι is a multitwist in $\partial\Sigma$, from which the result follows by Proposition 10.5.

The proof has various cases, depending on the topological type of Σ. If Σ has genus zero and $n + b \leq 3$, then our descriptions of $\mathrm{Mod}(\Sigma)$ show that any homeomorphism of Σ that fixes the punctures is isotopic to a multitwist in the boundary, and the result follows.

In the remaining cases, Σ either has genus at least one, or has at least four punctures and boundary components. In these cases, it is possible to find a collection of essential, simple closed curves $\{\gamma_i\}$ on Σ without bigons, annuli or triangles, such that no γ_i is homotopic to a boundary component, and such that every component of the cut surface $\Sigma_{\{\gamma_i\}}$ is either a closed disc, a punctured disc or an annulus. Furthermore, the annular components have the property that one side is a component of $\partial\Sigma$, and each γ_i intersects the other side in at most one arc. Finally, the structure graph $\Gamma_{\{\gamma_i\}}$ is connected.

If a homeomorphism ϕ is in the kernel of ι then, for each i, $\phi \circ \gamma_i$ is isotopic to γ_i in S, and hence in Σ by Lemma 11.2. By the Alexander method (Proposition 8.5), ϕ induces an automorphism ϕ_Γ of $\Gamma_{\{\gamma_i\}}$, and we claim that ϕ_Γ is trivial. Indeed, there is nothing to prove unless Σ is a proper subsurface of S, and in this case $\partial\Sigma$ is non-empty; in particular, at least one of the complementary components of $\Gamma_{\{\gamma_i\}}$ is an annulus, and is preserved by ϕ. By construction, ϕ_Γ preserves the cycles γ_i, so since the intersection of a γ_i and an annular component is connected, it follows that ϕ_Γ fixes at least one edge of $\Gamma_{\{\gamma_i\}}$. Since ϕ is orientation-preserving, if ϕ_Γ fixes an edge incident at a vertex then it also fixes all the other edges incident at that vertex. It follows that ϕ_Γ is the identity, because $\Gamma_{\{\gamma_i\}}$ is connected.

As in the Alexander method again, we may isotope ϕ so that it is equal to the identity on the complementary components that are discs or punctured discs. The resulting ϕ is supported on an annular neighbourhood of $\partial\Sigma$, and so the result follows. $\square$

11.2 Cut surfaces and stabilisers

One kind of essential subsurface that has played an important role is the *cut surface* S_α associated to a multicurve α. The cut surface is often not connected, but the definition of mapping class groups still makes sense. In fact, for a cut surface S_α, it is not hard to see that the mapping class group is just the direct product

$$\mathrm{Mod}(S) \cong \prod_{\Sigma\in\pi_0(S)} \mathrm{Mod}(\Sigma)\,,$$

where Σ ranges over the components of S.[5] The next result relates the mapping class group of the cut surface S_α to the *oriented* stabiliser

$$\mathrm{Mod}_\alpha(S) := \{[\phi] \in \mathrm{Mod}(S) \mid \phi \circ \alpha_i \simeq \alpha_i \text{ for all } i\}$$

(Note that $\mathrm{Mod}_\alpha(S)$ may be properly contained in the group of mapping classes that preserve α but are allowed to reverse orientation of the components.)

Proposition 11.4. *Let S be a connected surface of finite type, and α a multicurve on S with n components. There is a central extension*

$$1 \to \mathbb{Z}^n \to \mathrm{Mod}(S_\alpha) \to \mathrm{Mod}_\alpha(S) \to 1$$

where $\mathbb{Z}^n$ is generated by the multitwists $T_{\alpha_i^+}.T_{\alpha_i^-}^{-1}$, as α_i ranges over the components of α.

Proof. By construction of the cut surface, for each component α_i of α, the pair α_i^+ and α_i^- bound an annulus in S, and Theorem 11.4 asserts that the kernel is generated by the differences of these.

[5]For general disconnected surfaces, it may be possible to permute the components. However, in the case of a cut surface S_α, every component contains a boundary component, which has to remain fixed.

It remains to check that the image of $\mathrm{Mod}(S_\alpha)$ under the inclusion homomorphism is equal to $\mathrm{Mod}_\alpha(S)$. By construction, each $\phi \in \mathrm{Mod}(S_\alpha)$ extends to a homeomorphism of S that fixes α, and so its image is contained in $\mathrm{Mod}_\alpha(S)$. Conversely, if a homeomorphism ϕ of S has the property that $\phi \circ \alpha_i \simeq \alpha_i$ for all i then, by Lemma 8.3, we may modify ϕ until it fixes the α_i pointwise; ϕ then extends to a homeomorphism of S_α, and surjectivity follows. □

Lecture 12: Forgetting boundary components and punctures

12.1 Capping

To study all mapping class groups inductively, we need to understand how to remove punctures and boundary components. We will do this in two stages. First, we will *cap off* boundary components by punctured discs. This has the effect of turning a boundary component into a puncture, and is a relatively simple operation, whose result is determined by Theorem 11.3. Second, we remove a puncture and study the result. This operation is more subtle, and is determined by the *Birman exact sequence*.

At both stages of the process, the fact that mapping class groups are allowed to permute punctures introduces problems. For this reason, it is more convenient to work with the *pure* mapping class group – the surface of finite index that fixes the punctures.

Definition 12.1. Let S be a surface of finite type with n punctures. The *pure mapping class group* $\mathrm{PMod}(S)$ is the kernel of the natural map $\mathrm{Mod}(S) \to \mathrm{Sym}(n)$.

So the pure mapping class group is a natural subgroup of index $n!$ in $\mathrm{Mod}(S)$.

A particularly important instance of Theorem 11.3 explains what happens when we *cap* a boundary component with a once-punctured disc.

Corollary 12.2. *Let α be the simple closed curve corresponding to a boundary component of $S_{g,n,b}$. There is a central extension*

$$1 \to \langle T_\alpha \rangle \to \mathrm{PMod}(S_{g,n,b}) \xrightarrow{\iota} \mathrm{PMod}(S_{g,n+1,b-1}) \to 1$$

where ι is the inclusion homomorphism induced by gluing a punctured disc along α.

Proof. The kernel is $\langle T_\alpha \rangle$ by Theorem 11.3. It remains to prove that ι is surjective. Let $\phi \in \mathrm{Homeo}^+(S_{g,n+1,b-1}, \partial S_{g,n+1,b-1})$ fix the punctures. Therefore, $\phi \circ \alpha$ is homotopic to α and so, by Lemma 4.2, we may modify ϕ by an isotopy so that ϕ fixes α. By the Alexander lemma, we may further isotope ϕ to be equal to the identity on the punctured disc bounded by α, and it follows that ϕ is in the image of ι, as required. □

We can use this result to compute mapping class groups of surfaces with boundary from mapping class groups of surfaces with punctures – see, for instance, question 11 on problem sheet 2.

12.2 The Birman exact sequence

In the last section, we understood what happens when a boundary component is filled in by a punctured disc. But what about when the puncture itself is filled in? The effect of this is more subtle, and is explained by the *Birman exact sequence*. We will use the following notation: if S is a surface, then S_* denotes that surface with an additional puncture.

Theorem 12.3 (Birman exact sequence)**.** *Let S be a surface of finite type with $\chi(S) < 0$. Then there is a short exact sequence*

$$1 \to \pi_1(S) \to \mathrm{PMod}(S_*) \to \mathrm{PMod}(S) \to 1$$

The usual proof, which was Birman's original one, uses some sophisticated machinery. One shows that $\mathrm{Diff}_0(S, \partial S)$ is contractible, and that the natural map

$$\mathrm{Diff}(S_*, \partial S_*) \to \mathrm{Diff}(S, \partial S) \to S$$

where the second map is evaluation at $*$, is a fibre bundle. The result then follows from the long exact sequence of homotopy groups for a fibration.

In keeping with the spirit of this course, however, we will give a low-tech proof here, going via the automorphism group of the fundamental group of the surface.

Definition 12.4. Let G be any group, and $\gamma \in G$. A conjugating automorphism

$$\iota_\gamma : g \mapsto \gamma g \gamma^{-1}$$

is called an *inner automorphism* of G. The group of inner automorphisms is denoted by $\mathrm{Inn}(G)$. It is isomorphic to $G/Z(G)$, and is a normal subgroup of $\mathrm{Aut}(G)$. The quotient

$$\mathrm{Out}(G) := \mathrm{Aut}(G)/\mathrm{Inn}(G)$$

is called the *outer automorphism group* of G.

Thinking of the added puncture $*$ as a base point, any homeomorphism ϕ of S that fixes $*$ induces an automorphism ϕ_* of $\pi_1(S,*)$. Since (based) homotopies induced the same automorphism, we obtain a natural map $\mathrm{PMod}(S_*) \to \mathrm{Aut}(\pi_1 S)$. (We used this homomorphism earlier, when computing the mapping class group of the punctured torus.)

Something a little more interesting happens when we forget the puncture. Let ϕ represent an element of $\mathrm{PMod}(S)$. A choice of base point and a path from $*$ to $\phi(*)$ leads to an induced automorphism of $\pi_1 S$, well defined up to conjugation – in other words, ϕ induces an *outer* automorphism of $\pi_1 S$. From this discussion, and using the fact that the fundamental group of a hyperbolic surface group has trivial centre, we see that there is the following commutative diagram with exact rows.

$$\begin{array}{ccccccccc} & & & & \mathrm{PMod}(S_*) & \longrightarrow & \mathrm{PMod}(S) & & \\ & & & & \downarrow & & \downarrow & & \\ 1 & \longrightarrow & \pi_1(S) & \longrightarrow & \mathrm{Aut}(\pi_1 S) & \longrightarrow & \mathrm{Out}(\pi_1 S) & \longrightarrow & 1 \end{array}$$

The idea of our proof of the Birman exact sequence is now to pull back the exact sequence in the bottom row, using the following lemmas.

Lemma 12.5. *For any connected surface S, the natural map* $\mathrm{PMod}(S_*) \to \mathrm{PMod}(S)$ *is surjective.*

Proof. For any $\phi \in \mathrm{Homeo}^+(S, \partial S)$, there is a path from $*$ to $\phi(*)$. Now, we may defined an isotopy $\psi_\bullet$ of S, starting at the identity, so that $\psi_1(*) = \phi(*)$. Now $\psi_\bullet^{-1} \circ \phi$ is an isotopy from ϕ to a homeomorphism that fixes $*$, which proves the lemma. $\square$

Lemma 12.6. *For any connected surface of finite type with empty boundary,* $\mathrm{PMod}(S_*) \to \mathrm{Aut}(S)$ *is injective.*

Proof. There is a filling set of loops $\{\alpha_i\}$ in S based at $*$ (equivalently, proper arcs in S_* with endpoints on $*$) without bigons, annuli or triangles, that generate $\pi_1 S$. If $\phi \in \mathrm{PMod}(S_*)$ and the induced automorphism ϕ_* of $\pi_1 S$ is trivial then, for each i, $\phi \circ \alpha_i$ is homotopic to α_i. Therefore, by the Alexander method, ϕ preserves the structure graph up to isotopy. But the structure graph has a single vertex, and ϕ preserves the orientation of each edge, so the induced graph automorphism ϕ_Γ is trivial. Therefore, ϕ represents the trivial mapping class, as required. □

The final lemma describes the image of $\pi_1 S$ in $\mathrm{Aut}(\pi_1 S)$. Let α be an oriented simple closed curve on S, based at $*$. Let α_+ be the curve α, pushed slightly to the right, and let α_- be the same curve pushed slightly to the left; note that both of these are well-defined simple closed curves on the punctured surface S_*.

Lemma 12.7. *Let α be an oriented simple closed curve on S based at $*$, as above. The map*

$$\pi_1(S) \to \mathrm{Aut}(\pi_1 S)$$

sends α to the automorphism induced by the mapping class

$$T_{\alpha_+} \circ T_{\alpha_-}^{-1} .$$

In particular, as subgroups of $\mathrm{Aut}(\pi_1 S)$, $\pi_1(S) \subseteq \mathrm{Mod}(S_*)$.

Proof. We may extend α to a standard generating set of based curves β for $\pi_1 S$. It now suffices to check that, on each element β of this set, the curve $\tau_{\alpha_+} \circ \tau_{\alpha_-}^{-1} \circ \beta$ is based homotopic to the conjugate $\alpha \cdot \beta \cdot \alpha^{-1}$. There are several cases to consider.

If $\beta = \alpha$ then $\tau_{\alpha_+} \circ \tau_{\alpha_-}^{-1}$ fixes α, which matches up with the fact that $\alpha \cdot \alpha \cdot \alpha^{-1} \simeq \alpha$.

In the remaining cases, β intersects α only at the base point $*$, and we divide into two further cases: either β leaves α on one side and returns on the other, in which case β intersects each of α_+ and α_- exactly once, with opposite orientations, or β leaves α on one side and returns on same side, in which case β intersects either α_+ or α_- twice, once with either orientation. In either case, the result of applying the homeomorphism $\tau_{\alpha_+} \circ \tau_{\alpha_-}^{-1}$ can be drawn explicitly using the surgery description of Dehn twists, and the result is visibly based isotopic to $\alpha \cdot \beta \cdot \alpha^{-1}$. □

Remark 12.8. The subgroup $\pi_1 S \subseteq \mathrm{Mod}(S_*)$ is called the *point-pushing subgroup*, and is an important source of interesting mapping classes on punctured surfaces.

We are now ready to prove Theorem 12.3.

Proof of Theorem 12.3. First, assume that $\partial S = \varnothing$. By Lemma 12.5, $\mathrm{PMod}(S_*) \to \mathrm{PMod}(S)$ is surjective, i.e. that the sequence is exact at $\mathrm{PMod}(S)$.

To prove exactness at $\mathrm{PMod}(S_*)$, we need to prove that a mapping class ϕ is in the kernel of the map to $\mathrm{PMod}(S)$ if and only if it is point-pushing. From Lemma 12.7, the image of each generator α of $\pi_1(S)$ is of the form $T_{\alpha_+} \circ T_{\alpha_-}^{-1}$. Since α_+ and α_- are isotopic in S, the point-pushing subgroup is contained in the kernel of the map to $\mathrm{PMod}(S)$. For the other direction, if ϕ is in the kernel of the map to $\mathrm{PMod}(S)$ then it represents a trivial element of $\mathrm{Out}(\pi_1(S))$ and so the induced automorphism ϕ_* is inner. Therefore ϕ is point-pushing by definition.

The final part of exactness is to prove the injectivity of the map $\pi_1(S) \to \mathrm{PMod}(S_*)$. Since $\pi_1(S)$ has trivial centre (by, for instance, Question 7 of Example Sheet 1), the map $\pi_1(S) \to \mathrm{Aut}(\pi_1(S)$ is injective. But it factors through $\pi_1(S) \to \mathrm{PMod}(S_*)$, so that map must also be injective.

Suppose now that S has $b > 0$ boundary components, and let $\overline{S}$ be the result of capping off the boundary components of ∂S with punctured discs. Likewise, let $\overline{S}_*$ be the result of capping off the boundary components of ∂S_*. We now have the following commutative diagram. The two copies of $\mathbb{Z}^b$ are generated by Dehn twists in the boundary, so the left-hand column is exact. The right-hand column is exact by the case without boundary. The top row is exact because the inclusion $S \to \overline{S}$ induces an isomorphism of fundamental groups, while the middle and bottom rows are exact by inductive application of Corollary 12.2.

$$
\begin{array}{ccccccccc}
 & & & & 1 & & 1 & & \\
 & & & & \downarrow & & \downarrow & & \\
 & & 1 & \longrightarrow & \pi_1(S) & \longrightarrow & \pi_1(\overline{S}) & \longrightarrow & 1 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
1 & \longrightarrow & \mathbb{Z}^b & \longrightarrow & \mathrm{PMod}(S_*) & \longrightarrow & \mathrm{PMod}(\overline{S}_*) & \longrightarrow & 1 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
1 & \longrightarrow & \mathbb{Z}^b & \longrightarrow & \mathrm{PMod}(S) & \longrightarrow & \mathrm{PMod}(\overline{S}) & \longrightarrow & 1 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 1 & & 1 & & 1 & &
\end{array}
$$

A routine diagram chase then shows that the middle column is exact, which completes the proof. □

Lecture 13: The complex of curves

13.1 Generation by Dehn twists in the genus-zero case

As an immediate application of the Birman exact sequence, we can prove a generation result for pure mapping class groups of genus-zero surfaces.

Corollary 13.1 (Dehn, 1938)**.** *Let $S = S_{0,n,b}$ for any n, b. There is a finite collection of simple closed curves on S such that Dehn twists in that collection generate the pure mapping class group* $\mathrm{PMod}(S)$*. In particular,* $\mathrm{Mod}(S)$ *is finitely generated.*

Proof. Suppose first that $b = 0$. The result about $\mathrm{PMod}(S)$ is trivially true in the base cases when $n \le 3$. For larger n, the Birman exact sequence gives us that

$$1 \to \pi_1(S) \to \mathrm{PMod}(S_{0,n,0}) \to \mathrm{PMod}(S_{0,n-1,0}) \to 1\,.$$

Therefore, $\mathrm{PMod}(S_{0,n,0})$ is generated by the images of the generators of $\pi_1(S)$ and any choice of lifts of the generators of $\mathrm{PMod}(S_{0,n-1,0})$. By induction, the latter is generated by a finite collection of Dehn twists, and it is immediate from the definition that any Dehn twist on $S_{0,n-1,0}$ can be lifted to a Dehn twist (in the same curve) on $S_{0,n,0}$. Furthermore, Lemma 12.7 shows that the image of $\pi_1(S)$ is also generated by a finite collection of Dehn twists, and the result follows.

For $b > 0$, the result follows by induction on b, using Corollary 12.2.

Since $\mathrm{Mod}(S)$ has a subgroup of finite index that is finitely generated, $\mathrm{Mod}(S)$ is also finitely generated (for instance by adding coset representatives to the generating set). □

The same proof actually shows that generation by Dehn twists depends only on the genus.

Corollary 13.2. *Let $S = S_{g,n,b}$ for any g, n, b. There is a finite collection of simple closed curves on S such that Dehn twists in that collection generate the pure mapping class group* $\mathrm{PMod}(S)$ *if and only if the same is true for the pure mapping class group of the closed surface* $\mathrm{PMod}(S_g)$.

13.2 Definition and connectivity

Our final major goal for the course is to prove the analogue of Corollary 13.1 for higher-genus surfaces. The strategy of the proof is to relate $\mathrm{Mod}(S)$ to the mapping class groups of its subsurfaces. In order to do this, we need to explain how $\mathrm{Mod}(S)$ is 'built' from the mapping class groups of its subsurfaces. Algebraically, it's not easy to say what it means for a group to be 'built' from its subgroups in this way, but there is a nice topological way to say it: if G acts on a sufficiently nice complex X, the G is 'built' from the stabilisers of the cells of X.

The meaning of 'sufficiently nice' depends on what you want to prove. In this case, wanting to prove results about generating sets of G, we will need to know that X is connected. If we wanted to prove results about presentations for G, we would want to know that X was simply connected.

In any case, our task is then to define a complex on which $\mathrm{Mod}(S)$ acts naturally – the *complex of curves.*

Definition 13.3. Let S be a surface of finite type. The *complex of curves* of S is the following simplicial complex $C(S)$.

(i) The vertices of $C(S)$ are the *unoriented* isotopy classes of essential simple closed curves in S that are not homotopic into ∂S.

(ii) A collection of such isotopy classes $\{[\alpha_0], \dots, [\alpha_n]\}$ spans a simplex whenever the α_i have mutually disjoint representatives; or, equivalently, whenever $i(\alpha_i, \alpha_j) = 0$ for all i, j.

The 1-skeleton of $C(S)$ is called the *curve graph* of S. Note that $\mathrm{Mod}(S)$ acts naturally on $C(S)$ by simplicial automorphisms.

The definition treats punctures and boundary components similarly, and so we will assume for this section that $\partial S = \varnothing$, and write $S = S_{g,n}$.

For the simplest examples, the complex of curves is not very interesting,

Example 13.4. If $g = 0$ and $n \leq 3$ then $C(S) = \varnothing$.

Example 13.5. If $g = 0$ and $n = 4$ or $g = 1$ and $n \leq 1$, every pair of essential simple closed curves on S intersects. Therefore, $C(S)$ is a disjoint union of points.

These examples can be summarised as the cases when $3g+n \leq 4$. Our next result shows us that we get better behaviour for more complicated surfaces.

Theorem 13.6. *If $3g + n \geq 5$ then $C(S)$ is connected.*

Proof. Let α, β be essential, non-boundary-parallel, simple closed curves on S. It is enough to find a finite sequence of essential simple closed curves

$$\alpha = \alpha_0, \alpha_1, \dots, \alpha_n = \beta$$

such that $i(\alpha_i, \alpha_{i+1}) = 0$ for each i. The proof is by induction on $i(\alpha, \beta)$, with the cases $i(\alpha, \beta) = 0$ and $i(\alpha, \beta) = 1$ as the base cases.

If $i(\alpha, \beta) = 0$, we may take $\alpha_0 = \alpha$ and $\alpha_1 = \beta$, and there is nothing to prove.

If $i(\alpha, \beta) = 1$ then a regular neighbourhood N of $\alpha \cup \beta$ is a subsurface with one boundary component and Euler characteristic -1, hence a one-holed torus. Since $3g + n \geq 5$, the complement of N has at least one isotopy classes of essential curve α_1 not parallel to ∂S, which completes the proof.

For the inductive step, assume that $i(\alpha, \beta) > 1$ and that α and β are transverse and in minimal position. Choose orientations on α and β and consider two intersection points $x, y \in \alpha \cap \beta$ that are consecutive on β, joined by a subarc b of β, beginning at x and ending at y. There are two cases to consider, depending on whether the crossings at x and y have the same orientations or different orientations.

Suppose first that they have the same orientations. Let a be either of the subarcs of α from x to y, and let γ be the loop obtained by concatenating a and b. By construction, $i(\alpha, \gamma) = \#(\alpha \cap \gamma) = 1$, so γ is essential. Furthermore, after a small isotopy, we see that γ intersects β in the same points as α does except for x and y, which are replaced with a single intersection point z; so $i(\beta,\gamma) < i(\alpha,\beta)$.

Next, suppose that the two crossings have opposite orientations. Let a_1, a_2 be the two subarcs of β that start at x and end at y, and let γ_i be the simple closed curve obtained by concatenating b with a_i (reversing orientation if necessary). By construction, $i(\alpha,\gamma_i) = 0$ and the intersection points of β with γ_i are (after a small isotopy) a strict subset of the intersection points of α and β, so $i(\beta,\gamma_i) < i(\alpha,\beta)$. Neither γ_i can bound a disc, as otherwise α and β were not in minimal position. If one of the γ_i does not bound a punctured disc, then set γ to be that γ_i.

If both γ_i bound punctured discs, then gluing them together along b, it follows that α bounds a twice-punctured disc. Reversing the roles of α and β, we may also assume that β bounds a twice-punctured disc. It follows that α and β intersect twice, and are contained on an embedded subsurface homeomorphic to $S_{0,3,1}$ – the sphere with three punctures and one boundary component. The boundary component γ is then a simple closed curve disjoint from both α and β, and is essential because $3g + n \geq 5$.

In each case, there is an essential simple closed curve γ on S that does not bound a puncture, and such that $i(\alpha,\gamma) \leq 1$ and $i(\gamma,\beta) < i(\alpha,\beta)$. The result now follows by induction. □

Lecture 14: The complex of curves, continued

14.1 Non-separating curves

This connectivity of $C(S)$ will translate into an algebraic result about the finite generation of $\mathrm{Mod}(S)$. We will have fewer cases to consider in the proof if we restrict our attention to non-separating curves. In light of Corollary 13.2, we only need to consider closed surfaces $S_g \equiv S_{g,0,0}$.

Corollary 14.1. *Let $S = S_g$. If $g \geq 2$ then every pair of isotopy classes of non-separating curves α, β in $C(S)$ is joined by a sequence of non-separating curves α_i such that α_i and α_{i+1} are disjoint for all i.*

Proof. Consider a shortest sequence

$$\alpha = \alpha_0, \alpha_1, \ldots, \alpha_{n-1}, \alpha_n = \beta$$

guaranteed by Theorem 13.6. If α_{n-1} is itself non-separating, then the result follows by induction on n. Otherwise, α_{n-1} is separating. Since n is minimal, α_{n-2} and β must live in the same component of $S_{\alpha_{n-1}}$, or else they are disjoint and α_{n-1} can be removed from the sequence. The other component Σ of $S_{\alpha_{n-1}}$ has positive genus since S is closed, so we may find a non-separating curve α' in it, disjoint from both α_{n-1} and β. Replacing α_{n-1} by α', the result now follows as before by induction. □

14.2 Generation by Dehn twists

How does connectivity of a complex relate to finite generation of a group acting on that complex? The connection is via the following lemma, which is one of the basic ideas in the subject of geometric group theory.

Lemma 14.2. *Let X be a path-connected topological space, and let G be a group acting on X by homeomorphisms. Suppose that Y is an open subset whose G-translates cover X; that is, $GY = X$. Then the set of elements*

$$\{g \in G \mid gY \cap Y \neq \varnothing\}$$

generates G.

Proof. Exercise. □

A sample application is that the fundamental group of a compact manifold is finitely generated. Moreover, using Lemma 14.2 together with our results so far, it is not difficult to show that every mapping class group $\mathrm{Mod}(S)$ is finitely generated. However, we will not use this lemma directly here. Instead, we will use a modified version, adapted to working specifically in our context.

Fix a non-separating simple closed curve α on S, and consider all non-separating curves β on S that are disjoint from, and not isotopic to, α. By the change of coordinates principal, there are at most finitely many $\mathrm{Mod}(S_\alpha)$-orbits of these curves β; fix orbit representatives $\{\beta_1, \dots, \beta_k\}$.

Change of coordinates also tells us that there are homeomorphisms ϕ_j such that $\phi_j \circ \alpha = \beta_j$. (Below, we shall exhibit such ϕ_j explicitly.) The connectivity of $C(S)$ now gives a generating set for $\mathrm{Mod}(S)$.

Lemma 14.3. *If $g \geq 2$ then the set*

$$\mathrm{Stab}_{\mathrm{Mod}(S)}(\alpha) \cup \{\phi_1, \dots, \phi_n\}$$

generates Mod(S).

Proof. Let $g \in \text{Mod}(S)$. By Corollary 14.1, there is an edge-path in $C(S)$, all of whose vertices are non-separating curves, from α to $g(\alpha)$. By change of coordinates, each vertex is in the same orbit as α, and so the vertices of this path are

$$\alpha, g_1\alpha, \ldots, g_{m-1}\alpha, g_m\alpha = g\alpha$$

for some $g_i \in \text{Mod}(S)$. By induction on m, we may assume that g_{m-1} is in the subgroup generated by $\text{Stab}_{\text{Mod}(S)}(\alpha) \cup \{\phi_1, \ldots, \phi_n\}$.

The curve $\gamma = g_{m-1}^{-1} g\alpha$ is a simple closed curve in S_α, non-separating in S, and

$$i(\alpha, \gamma) = i(g_{m-1}\alpha, g_{m-1}\gamma) = i(g_{m-1}\alpha, g\alpha) = 0\,.$$

Therefore γ is disjoint from α (up to isotopy) and so $\gamma = h\beta_j$ for some j and some $h \in \text{Stab}_{\text{Mod}(S)}(\alpha)$. Since $\beta_j = \phi_j(\alpha)$, we have

$$g\alpha = g_{m-1}\gamma = g_{m-1}h\beta_j = g_{m-1}h\phi_j(\alpha)$$

and the result follows. □

To complete the proof that Dehn twists generate, we need to analyse the mapping classes ϕ_j. The next lemma shows that we may take them to be products of Dehn twists.

Lemma 14.4. *For any pair of disjoint, non-separating simple closed curves α, β on S, there is a sequence of Dehn twists taking α to β.*

Proof. First, we claim that there is a simple closed curve γ on S such that

$$i(\alpha, \gamma) = i(\beta, \gamma) = 1\,.$$

Indeed, fix points x on α and y on β, and consider the cut surface $S_{\alpha,\beta}$ obtained by cutting along both curves. Let $x_\pm$ and $y_\pm$ be corresponding points on the boundary components $\alpha_\pm$ and $\beta_\pm$.

Since S_α is connected, α_+ lies in the same component of $S_{\alpha,\beta}$ as β_+ (without loss of generality, after possibly reversing the orientation of β). Let γ_+ be an proper arc from x_+ to y_+. Now α_- is in the same component as β_-, so we may also take γ_- to be a proper arc from y_- to x_-. Furthermore, γ_+ does not disconnect its component of $S_{\alpha,\beta}$, so we may certainly choose γ_- to be disjoint from γ_+. The two arcs $\gamma_\pm$ reglue in S to a simple closed curve in γ in S intersecting each of α and β exactly once, as claimed.

Applying Question 8(a) of Problem Sheet 2 twice, $T_\alpha T_\gamma(\alpha) = \gamma$ and $T_\gamma T_\beta(\gamma) = \beta$, so

$$T_\gamma T_\beta T_\alpha T_\gamma(\alpha) = \beta$$

as required. □

Before giving the final proof, we also need to analyse non-trivial coset representatives of the *oriented* stabiliser $\mathrm{Mod}_\alpha(S)$ in the *unoriented* stabiliser $\mathrm{Stab}_{\mathrm{Mod}(S)}(\alpha)$. The next lemma shows that it can be generated by Dehn twists.

Lemma 14.5. *For any pair of simple closed curves α, β with $i(\alpha, \beta) = 1$,*

$$T_\beta T_\alpha^2 T_\beta(\alpha) = \alpha^{-1}$$

where α^{-1} denotes the same curve with the opposite orientation.

Proof. By change of coordinates, it suffices to perform this computation for the standard pair of curves on the one-holed torus. There, it can be checked by the usual surgery picture for Dehn twists, or by direct computation in $SL_2(\mathbb{Z})$. □

We are now ready to prove our main theorem.

Theorem 14.6. *Let S be any connected, oriented surface of finite type. There is a finite set of simple closed curves on S whose Dehn twists generate* $\mathrm{PMod}(S)$*. Moreover,* $\mathrm{Mod}(S)$ *is finitely generated.*

Proof. By Corollary 13.1, it suffices to prove the theorem for the case of genus $g > 0$. By Corollary 13.2, it suffices to prove the theorem in the case of the closed surface $S = S_g$, in which case $\mathrm{PMod}(S) = \mathrm{Mod}(S)$.

By Corollary 8.1, the case $g = 1$ corresponds precisely to the well known fact from linear algebra that $SL_2(\mathbb{Z})$ is generated by the elementary matrices

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}.$$

For genus $g \geq 2$, we proceed by induction. Fix a non-separating simple closed curve α in S. By Lemma 14.3, $\mathrm{Mod}(S)$ is generated by $\mathrm{Stab}_{\mathrm{Mod}(S)}(\alpha)$ together with the elements $\{\phi_1, \ldots, \phi_k\}$. By Lemma 14.4, there is a finite set of Dehn twists that generates the ϕ_j. The stabiliser $\mathrm{Stab}_{\mathrm{Mod}(S)}(\alpha)$ is a degree-two extension of the oriented stabiliser $\mathrm{Mod}_\alpha(S)$ and, by Lemma 14.5, the non-trivial coset representative can be generated by a finite collection of Dehn twists.

Therefore, it remains to prove that $\mathrm{Mod}_\alpha(S)$ is generated by finitely many Dehn twists. Proposition 11.4 shows that $\mathrm{Mod}_\alpha(S)$ is the image of $\mathrm{Mod}(S_\alpha)$ under the natural inclusion homomorphism (which takes Dehn twists to Dehn twists). Since α is non-separating, the genus of the cut surface S_α is smaller than g, and so the theorem follows by induction. □

References

[1] Benson Farb and Dan Margalit. *A primer on mapping class groups*, volume 49 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ, 2012.