Louis Funar

Surfaces

March 23, 2020

Springer
Berlin Heidelberg New York
Hong Kong London
Milan Paris Tokyo

1

Surfaces and their fundamental groups

1.1 Topological classification of surfaces

Manifolds are Hausdorff topological spaces admitting coverings by open subsets of some Euclidean space. Noncompact manifolds considered here are always assumed to be paracompact, second-countable, regular and in particular metrizable, in order to avoid pathologies. When transition maps between the open charts are constrained, say piecewise linear, Lipschitz or smooth maps, we say that we have a piecewise linear, Lipschitz or smooth manifold.

Manifold classification in dimensions at most two is rather simple and intuitive. One dimensional connected manifolds without boundary are either the circle S^1 or the line $\mathbb{R}$. In the case where we allow boundary we can add the interval $[0,1]$ and the half-line $[0,\infty)$ to them.

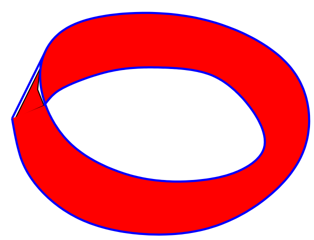
Recall that a *simplicial complex* is a set of simplices which contains along with a simplex all its facets and pairwise intersect along facets if nondisjoint. Their union is the underlying topological space of the complex. A *triangulation* of a surface S is then a simplicial complex endowed with a homeomorphism between the underlying space onto S.

A basic result that will be used freely in this sequel is the existence of triangulations and smooth structures:

Theorem 1. *Every topological surface has a smooth structure and can be triangulated.*

The first proof that surfaces can be triangulated and hence smoothed is attributed to Radó (see [4]), who also gave an explicit classification in the compact case. Under the triangulation assumption, the classification of compact surfaces was essentially known as far as 1860.

Among all curves on the surfaces those which are embedded, namely have no self-intersections, play a proeminent role and are usually called *simple curves*. A surface is orientable if simple closed curves drawn on it are *2-sided*, namely each simple curve admits a small neighborhood homeomorphic to a cylinder. This amounts to the existence of a continuous vector field defined in a neighborhood of the curve which is orthogonal to the curve in every point. Observe that we can always find an orthogonal line field. If we select a vector in this line in a continuous manner as we travel around the curve then we return to the initial point with the same vector or with the opposite vector. In the second alternative the curve is called *1-sided* and a small neighborhood of the curve is homeomorphic to the so-called Moebius band drawn below:

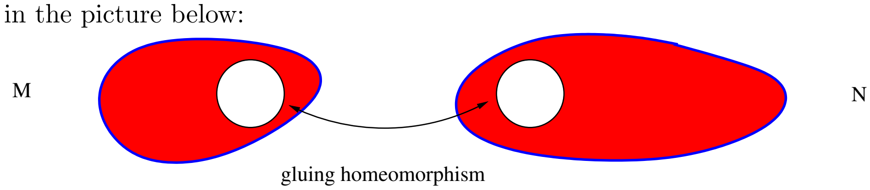


Thus a surface is *nonorientable* if it contains embedded Moebius bands.

The simplest surfaces are the sphere and the torus. In order to construct more orientable surfaces we should first describe a basic operation, the so called *connected sum* of manifolds. Given n-manifolds M and N, embeddings $D^n \to M$, $D^n \to N$ and an orientation reversing homeomorphism φ between the boundary spheres of $M - \mathrm{int}(D^n)$ and $N - \mathrm{int}(D^n)$, we can construct a new manifold

$$M \sharp N = M - \mathrm{int}(D^n) \sqcup N - \mathrm{int}(D^n)/x \sim \varphi(x)$$

by identifying their boundary points by means of the *gluing homeomorphism* φ, as in the picture below:



The connected sum $M \sharp N$ is well-defined up to a homeomorphism and independent on the various choices made. The n-sphere S^n is a unit for the connected sum, as $M \sharp S^n$ is homeomorphic to M.

In particular, we can construct a compact connected orientable surface without boundary by making the connected sum of g copies of the torus $S_g = \sharp_g S^1 \times S^1$. The number g of copies of the torus is called *the genus* of the orientable surface, and we consider $g = 0$ for the 2-sphere. If now S has genus g we can obtain a surface of genus g with b boundary components $S_{g,b} = S_g - \sqcup_1^b \mathrm{int}(D^2)$, as the result of deleting b disjoint open disks from S_g.

The simplest nonorientable surface without boundary is the projective plane, which can be defined as the quotient $S^2/\sim$ by the equivalence relation which identifies antipodal points on the standard 2-sphere. The connected sum of g copies of the projective plane $N_g = \sharp_g RP^2$ is a nonorientable surface without boundary, which is said to have *genus* g. By drilling some circular holes in it we obtain the surface $N_{g,p} = N_g - \sqcup_1^b D^2$ which has genus g and b boundary components.

We can now state the following fundamental result, which essentially says that we constructed all compact surfaces:

Theorem 2 (Radó 1925, Compact surfaces classification). *Compact connected surfaces are homeomorphic if and only if:*

1. *they are both orientable or nonorientable;*
2. *they have the same genus;*
3. *they have the same number of boundary components.*

Proof. We will sketch a proof in the orientable case, following Zeeman ([5] see also [2]) and assuming that compact surfaces can be triangulated. The interested reader might compare with the more recent proof given by Conway (see [1]).

By capping off the boundary components by disks we obtain a closed orientable surface, so it is enough to show that any closed orientable surface S is homeomorphic to some S_g.

Given a triangulation of S, its dual graph comes with a natural embeddeding in S. Choose a maximal tree T in the triangulation and a maximal subgraph $\overline{T}$ disjoint from T in the dual graph.

If f_i denotes the number of i simplices in the triangulation, then T has f_0 vertices and $f_0 - 1$ edges. As every edge of the triangulation belongs either to T or to T^*, we find that T^* has f_2 vertices and $f_1 - f_0 + 1$ edges.

If T^* were disconnected, the boundary of the union of triangles from a connected component will be a cycle in T, contradicting the fact that T is a tree. Therefore, the number of edges of T^* is at least the number of its vertices minus one, namely:

$$f_1 - f_0 + 1 \geq f_2 - 1.$$

It follows that $\chi(S) = f_0 - f_1 + f_2 \leq 2$. We use now induction on $2 - \chi(S)$.

If $\chi(S) = 2$, then T^* has $f_2 - 1$ edges, so that T^* is a tree. A neighborhhod of a tree is homeomorphic to a disk D^2 so that S is the union of two disks along their common boundary, which is S^2, because orientable.

If $2 - \chi(S) > 0$, then T^* cannot be a tree because it has more than $f_2 - 1$ edges and hence contains a nontrivial cycle. Cut S along a minimal cycle, which is a curve drawn on S without self-intersections. This curve cannot disconnect S, as it cannot disconnect the disjoint tree T, so that we obtain a surface with two boundary components. Capping off the boundary components by disks we obtain a closed orientable surface S' such that $S = S' \sharp S^1 \times S^1$. If the minimal cycle has c edges and c vertices then S' has $f_0 + c$ vertices, $f_1 + c$ edges and $f_2 + 2$ faces, so that

$$\chi(S') = \chi(S) + 2.$$

By the induction hypothesis S' is homeomorphic to S_g, for some g, so that S is homeomorphic to S_{g+1}.

An immediate consequence of the classification theorem is:

Corollary 1. *A compact connected and simply connected surface is homeomorphic to the 2-sphere, if it has no boundary and to the 2-disk, otherwise.*

We will consider also a particular case of open surfaces which we call *punctured compact surfaces*. Specifically, $S_{g,b}^p$ and $N_{g,b}^p$ denote the result of deleting p points from the compact surfaces $S_{g,b}$ and $N_{g,b}$, respectively.

Remark 1. Topological manifolds up to the dimension 3 can be triangulated (Moise) and smooth manifolds of any dimension can be triangulated (Whitehead). The existence of topological manifolds which are not triangulable in any dimension higher than 4 is a recent breakthrough by Manolescu (see [3]).

Exercise 1. A more topological view of RP^2 is the result of adjoining a 2-disk D^2 along the boundary circle of a Moebius band B. The Klein bottle K is the genus 2 nonorientable surface $\sharp_2 RP^2$. Show that $M\sharp K = M\sharp S^1 \times S^1$, if M is nonorientable. Further derive from this that $S_g \sharp RP^2$ is homeomorphic to N_{2g+1}.

Exercise 2. Let $c(S)$ be the maximal number of disjoint simple closed curves on S which do not disconnect S. Then show that $c(S)$ is the genus of S, when S is a compact orientable surface. Set further $c'(S)$ for the maximal number of disjoint Moebius bands on S. Prove that $c'(S)$ is the genus of S if S is a compact nonorientable surface and 0, otherwise.

Exercise 3. Prove that an open connected and simply connected surface is homeomorphic to the 2-plane.

References

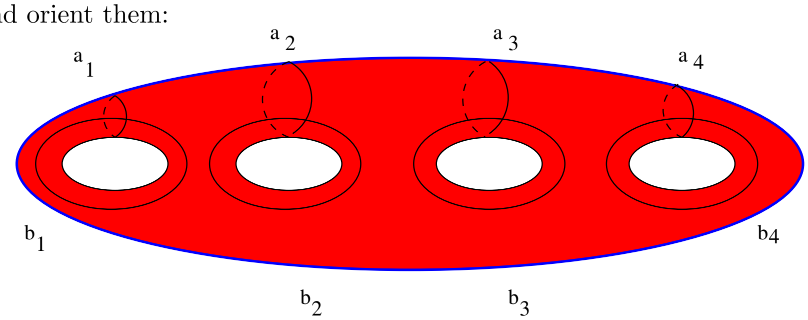
1. G.K.Francis and J.R. Weeks, *Conway's ZIP Proof*, American Mathematical Monthly (5) 106 (1999) 393–399.
2. Jennifer Liu, An extension of Zeeman's An introduction to topology, Master degree thesis, Univ. Santa Barbara, 2012.
3. C. Manolescu, *Lectures on the triangulation conjecture*, Akbulut, Selman (ed.) et al., Proceedings of the 22nd Gökova geometry-topology conference, Gökova, Turkey, May 25–29, 2015. Somerville, MA, International Press, 1–38, 2016.
4. Tibor Radó. *Über den Begriff der Riemannschen Fläche*, Acta Litt. Sci. Szeged 2 (1925), 101–121.
5. E. C. Zeeman, Introduction to topology, unpublished notes (Warwick, 1966); available at http://www.maths.ed.ac.uk/aar/surgery/ecztop.pdf.

1.2 Topological invariants of surfaces

The most interesting part of the homology of the surface S_g lives in degree one:

$$H_1(S_g) \cong \mathbb{Z}^{2g}$$

We obtain a basis of $H_1(S_g)$ by taking one meridian a_j and a longitude b_j, $j = 1, \ldots, g$, for each torus in the connected sum decomposition $S_g = \sharp_g S^1 \times S^1$ and orient them:



Altough all classes in $H_1(S_g)$ can be represented by curves drawn on the surface S_g, curves might have self-intersections, in general.

In degree 2 the homology $H_2(S_g) \cong \mathbb{Z}$ is generated by the fundamental class $[S_g]$, while in higher degrees the homology vanishes.

The Poincaré duality provides us a nondegenerate pairing

$$H^1(S_g) \times H_1(S_g) \to \mathbb{Z}$$

inducing a canonical isomorphism between $H_1(S_g)$ and $H^1(S_g)$.

A fundamental invariant of the surface S_g is the *intersection pairing* ω which is obtained by evaluating the cup produit in cohomology at the fundamental class:

$$H^1(S_g) \times H^1(S_g) \xrightarrow{\cup} H^2(S_g) \xrightarrow{[S_g]} \mathbb{Z}$$

In the dual picture ω is an antisymmetric nondegenerate, i.e. symplectic, pairing

$$\omega : H_1(S_g) \times H_1(S_g) \to \mathbb{Z}.$$

Then $\omega(\alpha, \beta)$ is the algebraic intersection number of the curves α and β representing the given homology classes. One sees then immediately that in the basis $\{a_1, \ldots, a_g, b_1 \ldots, b_g\}$ above ω reads:

$$\omega(a_i, b_j) = \delta_{ij},\ \omega(a_i, a_j) = \omega(b_i, b_j) = 0,\ i, j \in \{1, 2, \ldots, g\}.$$

When this is the case one calls the homology basis a *symplectic basis*.

There is a similar description of the homology of a nonorientable surface:

$$H_1(N_g) \cong \mathbb{Z}^{g-1} \oplus \mathbb{Z}/2\mathbb{Z}$$

where the homology group is generated by the classes of curves a_i which are central curves in the Moebius bands corresponding to the factors RP^2 in the connected sum decomposition $N_g = \sharp_g RP^2$. The intersection form ω arises both from the cup product

$$H^1(N_g; \mathbb{Z}/2\mathbb{Z}) \times H^1(N_g; \mathbb{Z}/2\mathbb{Z}) \xrightarrow{\cup} H^2(N_g; \mathbb{Z}/2\mathbb{Z}) \xrightarrow{[N_g]} \mathbb{Z}/2\mathbb{Z}$$

or in the dual picture as a $\mathbb{Z}/2\mathbb{Z}$-valued form:

$$\omega : H_1(N_g) \times H_1(N_g) \to \mathbb{Z}/2\mathbb{Z}.$$

Now $\omega(\alpha, \beta)$ counts the algebraic intersection of curves mod 2. In particular

$$\omega(a_i, a_j) = \delta_{ij},\ i, j \in \{1, 2, \ldots, g\}.$$

By the functoriality of the cohomology we have the following easy lemma:

Lemma 1. *Any homotopy equivalence $h : S_g \to S_g$ induces an isomorphism $h_* : H_1(S_g) \to H_1(S_g)$ respecting the symplectic form*

$$\omega(h_*(x), h_*(y)) = \omega(x, y)$$

Proof. In fact h induces an isomorphism $h^* : H^1(S_g) \to H^1(S_g)$ respecting the cup product.

In particular homeomorphisms of S_g induce symplectic automorphisms of the homology $H_1(S_g)$. We will show later that all symplectic automorphisms arise this way.

One should note that $H_2(S) = 0$, if S is not closed surface. However the intersection pairing

$$\omega : H_1(S) \times H_1(S) \to \mathbb{Z}$$

is still an antisymmetric bilinear form invariant by the homotopy equivalences, although its kernel might be nontrivial.

A finer invariant of a surface than its homology is the fundamental group. A key fact in surface theory is:

Lemma 2. *If a compact surface has boundary or punctures then its fundamental group is finitely generated free.*

Proof. If the surface has boundary then we can find a spine for it, namely a finite graph embedded in the surface on which the surface deformation retracts. Pick up a triangulation of the surface and start shelling 2-cells starting from the boundary. Recall that shelling is the process of removing a 2-cell with a free face, namely a face which is not a face of another 2-cell of the traingulation. Shellings are (simple) homotopy equivalences. If we get stalked, then all 2-dimensional cells were removed, as otherwise we have a 2-cycle in the homology of the surface and this contradicts the vanishing of its 2-homology. In particular, its fundamental group is the fundamental group of a graph and hence it is free. Eventually observe that a punctured surface is homotopy equivalent to a surface with boundary where each puncture was replaced by a boundary circle surrounding it.

When the surface S is closed there is a well-known description of its fundamental group by generators and relations:

Lemma 3 (Surface group presentation). *According to whether the surface is orientable or not we have:*

$$\pi_1(S_g) = \langle a_1, a_2, \ldots, a_g, b_1, b_2, \ldots, b_g | \prod_{i=1}^{g} [a_i, b_i] = 1 \rangle$$

$$\pi_1(N_g) = \langle a_1, a_2, \ldots, a_g | \prod_{i=1}^{g} a_i^2 = 1 \rangle$$

Proof. We arrange for the curves a_i, b_i above to be disjoint everywhere except at one basepoint. Cutting along these curves we obtain a $4g$-gon whose edges are oriented and labeled by the curves from which they arise. This provides a cell decomposition of S_g with one vertex, $2g$ 1-cells and one 2-cell to be attached on the 1-skeleton according to the desired relation. The proof is similar for the case of N_g.

Exercise 4. We denote by $\chi(M)$ the Euler-Poincaré characteristic of a cellular complex, given by:

$$\chi(M) = f_0 - f_1 + f_2 - \cdots + (-1)^n f_n$$

where f_j is the number of cells of dimension j. Show that $\chi(M \sharp N) = \chi(M) + \chi(N) - 2$, if M and N are compact surfaces.

Exercise 5. A class in $H_1(S_g)$ is called *primitive* if it is not a multiple of some class. Show that primitive classes in homology are precisely those classes which can be represented by simple closed curves on S_g.

Exercise 6. Show that $H_1(S)$ has rank $2 - \chi(S)$.

References

1. Allan Hatcher, Algebraic Topology, Cambridge University Press, 2002, available at https://pi.math.cornell.edu/∼hatcher/AT/ATpage.html

1.3 Coverings of surfaces and Johansson's Theorem

Definition 1. *A* `covering` *of topological spaces $f : M \to N$ is a continuous surjective map such that every point of N has an evenly covered open neighborhood, namely an open set U such that $f^{-1}(U)$ is the disjoint union of open sets in M to which f restricts to a homeomorphism.*

A simply connected covering $\widetilde{M}$ of a space M is a universal covering space M, which has the property that covers any other covering of M. Manifolds and more generally spaces which are connected, locally path-connected and semi-locally simply connected have universal coverings. A covering $f : M \to N$ between path connected spaces induces an injective homomorphism

$$f_* : \pi_1(M, p) \to \pi_1(N, f(p)).$$

Conversely, for any subgroup $\Gamma \subset \pi_1(N)$ we can associate a covering $f : M \to N$ such that $f_*(\pi_1(M)) = \Gamma$. There is a bijection between *equivalence classes of coverings* of a manifold N and the *conjugacy classes* of subgroups of $\pi_1(N)$.

Definition 2. *The number of preimages $f^{-1}(p)$ of a point $p \in N$ under the covering $f : M \to N$ of a path connected space N is called the* `degree` *of the covering.*

The degree does not depend on the choice of the point p and equals the index of the subgroup $f_*(\pi_1(M)) \subseteq \pi_1(N)$. Note that a degree 1 covering is a homeomorphism.

Lemma 4. *If N is a simplicial complex and $M \to N$ is a covering of degree d, then $\chi(M) = d \cdot \chi(N)$.*

Proof. Choose a subdivision of N such that every cell in N is contained in an evenly covered open set. If f_i denotes the number of facets of dimensions i then we have:

$$f_i(M) = d \cdot f_i(N)$$

The desired formula for the Euler-Poincaré characteristic follows.

Corollary 2. *If $S' \to S$ is a nontrivial covering and S' is a compact surface with $\chi(S') \geq 0$, then either $S = S^2$ and $S' = RP^2$ or else S is a Klein bottle and S' is either a torus or a Klein bottle, or S is a torus and S' is torus.*

By language abuse a *surface group* means the fundamental group of a closed surface. The main result of this section is the following:

Theorem 3 (Johansson 1931). *The fundamental group of a noncompact surface is free.*

This result is due to Johansson (see [1]. We will present here the proof due to J.H.C. Whitehead from [3], who proved a more general result valid in any dimension. An alternative proof could be found in the book by Stillwell (see [2]).

We say that a graph is *irreducible* if the complement of every (open) edge has a finite connected component. An infinite connected irreducible graph consists of an infinite ray with finite trees grafted at its vertices.

Lemma 5 (Whitehead's lemma). *A connected infinite locally finite graph contains a set of mutually disjoint infinite connected irreducible graphs whose union contains all the vertices.*

Proof. Let the set of edges of the graph G be ordered. We define inductively the sequence $G_0 = G$ and $G_{n+1} = G_n$, if G_n is irreducible and $G_{n+1} = G_n - e_n$, otherwise, where e_n is the first edge with respect to the chosen order with the property that G_{n+1} has no compact component. Set $G^* = \cap_{n=0} G_n$. Then G^* contains all vertices of G.

Note that G^* has no compact component. Assume the contrary, and let $K \subset G^*$ a compact connected component and denote by A the set of edges of G meeting K which are not in K. The set A is nonempty as G is connected and infinite and no edge in A could be in G^* since K is a component of G^*. Then A is finite since G is locally finite and K is finite. Therefore there exists $n \geq 0$ such that $G_{n+1} = G_n - e_n$, where $e_n \in A$ and G_{n+1} contains no edge from A while G_n is not irreducible. Then the component of G_{n+1} containing K must be finite since its vertices are among those of K, contradicting the choice of the edge e_n.

We claim that G^* is irreducible. Assume the contrary, so that $G_{n+1} \neq G_n$ for every $n \geq 0$ and there exists some edge e such that $G^* - e$ has no compact component. There exists $n \geq 0$ with the property that whenever we have an edge $e_j \in G - G^*$ which precedes e, then $e_j \notin G_n$. It follows that e precedes e_n. Since every component of $G_n - e$ contains a component of $G^* - e$, all these are noncompact. In particular G_{n+1} should be $G_n - e$, thereby contradicting our choice of e_n.

It follows that every connected component of G^* is an infinite irreducible graph.

Proof (Proof of Theorem 3). Assume that the connected surface S is triangulated and consider 1-skeleton G of its the dual cell complex. Let G^* be the irreducible subgraph provided by Whitehead's Lemma 5, when G is infinite and a maximal subtree of G, otherwise. Note that G^* is embedded in S. Let L be the union of ∂S and the edges of G which are not edges of G^*. Then the complement of a regular neighborhood of G^* in S is an open regular neighborhood of L in S. Observe now that a regular neighborhood of a connected component of G^* is homeomorphic to the product $[0,1] \times [0,\infty)$, namely a half-plane. Since a half-plane deformation retracts onto its boundary, it follows that S deformation retracts onto the graph L. In particular, its fundamental group is free.

1.4 Applications of Johansson's Theorem

An immediate consequence is the following:

Corollary 3. *A nontrivial subgroup of a surface group is either a surface group or a free group, depending on whether it is of finite index or not.*

Proof. Consider the unramified covering associated to the given subgroup. If the covering is compact, then its fundamental group is a surface group, necessarily of finite index equal to the degree of the covering. Otherwise Johansson's Theorem 3 shows that the fundamental group of the covering is free, of infinite index.

Furthermore we also have:

Corollary 4. *If S is a closed surface other than RP^2, then $\pi_1(S)$ is torsion-free.*

Proof. If $H \subset \pi_1(S)$ is a cyclic subgroup, then by Corollary 3 H is either free or a surface group. A nontrivial cyclic surface group has cyclic abelianization and hence is $\mathbb{Z}/2\mathbb{Z}$ and the surface is RP^2. Now, if RP^2 finitely covers a surface S, then $\pi_1(S)$ should be a finite group containing $\mathbb{Z}/2\mathbb{Z}$ and the only possibility is $S = RP^2$.

Corollary 5. *If S is a (possibly punctured) surface and $\pi_1(S)$ is virtually abelian, then either $\pi_1(S)$ is abelian and:*

1. *if S is closed, then S is a sphere, a torus or RP^2;*
2. *if S has boundary, then S is a disk, a cylinder or a Moebius band;*
3. *if S is punctured, then S is a once punctured disk or a once punctured RP^2.*

or else S is a Klein bottle.

Proof. Punctured surfaces and surfaces with boundary have free fundamental groups which then have to be cyclic. When S is closed and $\chi(S) \leq -1$, then the rank of $H_1(S)$ is at least 3, so that we can choose two elements $a, b \in \pi_1(S)$, such that the image of the subgroup they generate in $H_1(S)$ has infinite index and rank 2. Then the group $\langle a, b\rangle \subset \pi_1(S)$ has infinite index. By Johansson's Theorem 3 the group $\langle a, b\rangle$ is free and not cyclic hence nonabelian and not virtually abelian, which is a contradiction.

It remains to verify that $\pi_1(N_2)$ is indeed virtually abelian, the other cases being obviously abelian. In the usual presentation of $\pi_1(N_2)$:

$$\langle a, b | bab^{-1} = a^{-1}\rangle$$

the group $\langle a, b^2\rangle$ is a free abelian normal subgroup of $\pi_1(N_2)$ of index 2.

Lemma 6. *The fundamental group $\pi_1(N_2)$ of the Klein bottle is virtually abelian and has center isomorphic to $\mathbb{Z}$.*

Proof. Usual presentations of $\pi_1(N_2)$ are

$$\langle a, b | bab^{-1} = a^{-1}\rangle = \langle b, c | c^2 = b^2\rangle$$

where $c = ab$. It follows that the element b^2 belongs to the center and the quotient

$$\langle b, c | b^2 = c^2 = 1\rangle = \mathbb{Z}/2\mathbb{Z} * \mathbb{Z}/2\mathbb{Z}$$

is virtually $\mathbb{Z}$ as the abelianization map $\mathbb{Z}/2\mathbb{Z} * \mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$ has kernel $\mathbb{Z}$ generated by $[b, c]$. Moreover, the center of this free product is trivial, so that the center of $\pi_1(N_2)$ is indeed the cyclic group generated by b^2. On the other hand the subgroup $\langle a, b^2\rangle \subset \pi_1(N_2)$ is a normal subgroup of index 2 in $\pi_1(N_2)$.

Theorem 4. *The fundamental group of the surface S is centerless if and only if not virtually abelian.*

Proof. If S is noncompact or has nonempty boundary then its fundamental group is free and by our assumptions nonabelian. In particular its center is trivial unless the fundamental group is $\mathbb{Z}$.

Suppose now that S is closed and $\pi_1(S)$ is not virtually abelian. If the center $C \subset \pi_1(S)$ is of finite index, then there exists a closed surface S' with $\pi_1(S') = C$. By the previous Corollary 5 S' is either a sphere, a torus or RP^2 and by Lemma 2 S is also S^2, RP^2 or a torus, contradicting our assumptions.

If $C \subset \pi_1(S)$ is of infinite index, from Johansson's Theorem 3 C must be a free group and since it is abelian it must be cyclic, say generated by c. From Corollary 5 the rank of $H_1(S)$ is at least 3. Take an element $a \in \pi_1(S)$ such that the subgroup generated by the images of a and c in $H_1(S)$ is not cyclic. We claim that the subgroup $G = \langle a, c\rangle \subset \pi_1(S)$ must be free. Otherwise it would be $\pi_1(S')$ for some closed S' covering S. If d denotes the degree of this covering, then

$$\chi(S') = d\chi(S)$$

In particular the rank of $H_1(S')$ would be

$$2 - \chi(S') = 2 - d\chi(S) \geq 2 + d > 2$$

contradicting the fact that $H_1(S')$ has two generators. It follows that G is free and not cyclic hence nonabelian. In particular its center should be trivial, but the center of G should contain C, which is a contradiction.

Now $\pi_1(N_2)$ has the presentation:

$$\langle b, c | c^2 = b^2\rangle$$

Its quotient by the central subgroup $\langle b^2\rangle$ is the free product $\mathbb{Z}/2\mathbb{Z} * \mathbb{Z}/2\mathbb{Z}$, thus center free. Thus the center of $\pi_1(N_2)$ is $\mathbb{Z}$.

Exercise 7 (Whitehead 1961). Prove that a combinatorial n-manifold which is either noncompact or has boundary deformation retracts onto a $(n-1)$-dimensional CW complex.

Exercise 8. Show that a nontrivial subgroup of a finitely generated free group is a free group, which is finitely generated if and only if it is of finite index.

Exercise 9 (Freiheitssatz for surface groups). Prove that a subgroup generated by k elements of $\pi_1(S)$, S a closed orientable surface with $\chi(S) \leq 0$ is free if $k \leq 1 - \chi(S)$.

References

1. I. Johansson, *Topologische Untersuchungen ber unverzweigte Überlagerungsflächen*, Skriftennorske Vidensk.-Akad. Oslo Math.-natur. Kl. 1 (1931), 1–69.

2. J. Stillwell, Classical topology and combinatorial group theory, second edition, Graduate Texts in Mathematics, 72, Springer-Verlag, New York, 1993.
3. J.H.C.Whitehead, *The immersion of an open 3-manifold in Euclidean 3-space*, Proc. London Math. Soc. (3) 11 (1961), 81–90.

2

Surface groups

2.1 Residual finiteness

Definition 3. *A group G is* `residually finite` *if for any element $1 \neq g \in G$, we can find a finite quotient of G for which the image of g is nontrivial.*

Finite groups are obviously residually finite. The first example of an infinite residually finite group is a free group. We indeed have:

Proposition 1. *A free group is residually finite.*

Proof. If g is a nontrivial element of a free group, it can be written as a reduced word in the generators a_i, namely:

$$g = a_{i_n}^{\varepsilon_n} a_{i_{n-1}}^{\varepsilon_{n-1}} \cdots a_{i_1}^{\varepsilon_1}, \text{ where } \varepsilon_{i_s} \in \{-1, 1\}.$$

The word is reduced means here that $\varepsilon_k = \varepsilon_{k+1}$, if $i_k = i_{k+1}$.

Let $\mathfrak{S}_{n+1}$ be the permutation group on the set of $\{1, 2, \ldots, n+1\}$. We will define the permutations $\sigma_{i_k} \in \mathfrak{S}_{n+1}$, for $1 \leq k \leq n$ by asking the following conditions be satisfied:

$$\sigma_{i_k}(k) = k + 1, \text{ if } \varepsilon_k = 1$$

$$\sigma_{i_k}(k+1) = k, \text{ if } \varepsilon_k = -1$$

These requirements determine injective maps σ_{i_k} defined on suitable subsets of $\{1, 2, \ldots, n+1\}$ into $\{1, 2, \ldots, n+1\}$, which can be extended arbitrarily to bijections.

The map which associates to a generator a_{i_k} the element σ_{i_k} extends to a homomorphism of the free group into $\mathfrak{S}_{n+1}$. Then the image of the element g is nontrivial because

$$\sigma_{i_n}^{\varepsilon_n} \sigma_{i_{n-1}}^{\varepsilon_{n-1}} \cdots \sigma_{i_1}^{\varepsilon_1}(1) = n + 1$$

An alternate topological proof was given by Stallings along the following lines. We can assume that the free group is finitely generated, say by m elements

a_i. Then $\langle a_1, a_2, \ldots, a_m \rangle$ is the fundamental group $\pi_1(X)$ of the wedge of m oriented circles X carrying the labels a_i. The universal covering $\widetilde{X}$ is the regular tree of degree $2n$ having its edges labeled by a_i and a_i^{-1}.

Every word

$$a_{i_n}^{\varepsilon_n} a_{i_{n-1}}^{\varepsilon_{n-1}} \cdots a_{i_1}^{\varepsilon_1}, \text{ where } \varepsilon_{i_s} \in \{-1, 1\}.$$

in the generators $a_i^{\pm 1}$ determines an oriented path in $\widetilde{X}$ which follows at the k-th vertex the direction given by the edge labeled $a_{i_k}^{\varepsilon_k}$. The element g determined by this word is nontrivial if and only if the corresponding path is not closed.

Consider an oriented regular graph Y of degree $2m$ which contains the vertices and edges from the path associated to some given nontrivial element g and such that every vertex has m in and m out edges. We can therefore extend the labeling of path edges to Y. Then Y has a simplicial map into X, which sends an edge labeled $a_i^{\pm 1}$ in Y into the oriented edge of X carrying the same label. This is a finite covering of X and the projection $\widetilde{X} \to Y$ sends the lift of the loop representing g into a nonclosed path in Y. It follows that g does not belong to the finite index subgroup corresponding to the image of $\pi_1(Y)$ within $\pi_1(X)$.

Proposition 2. *A finitely generated residually finite group is Hopfian.*

Proof. Consider a surjective homomorphism $\phi : G \to G$ of a finitely generated group G and let $K = \ker \phi$.

As there are only finitely many groups of order m, every finitely generated group G has only finitely many subgroup say H_i, $1 \leq i \leq n$ of a given index m. Now, $\phi^{-1}(H_i) \subset G$ are also subgroups of index m in G. Since $\phi^{-1}(H_i) \neq \phi^{-1}(H_j)$, for distinct $i \neq j$, it follows that the sets $\{\phi^{-1}(H_i)\}_{1 \leq i \leq n}$ and $\{H_i\}_{1 \leq i \leq n}$ coincide. In particular

$$K \subseteq H_i, \text{ for } 1 \leq i \leq n.$$

But m was arbitrary, so that K is contained in every finite index subgroup of G.

But a group G is residually finite if and only if the intersection of all its finite index subgroups is trivial. Thus K must be trivial and hence ϕ is an isomorphism.

Corollary 6. *Free groups are Hopfian.*

2.2 Finitely generated fundamental groups

The main result of this section is the following classification of open surfaces with finitely generated fundamental groups:

Theorem 5. *If S is an open surface and $\pi_1(S)$ is finitely generated then there exists a compact surface F with nonempty boundary ∂F such that S is homeomorphic to* $\mathrm{int}(F)$.

We say that a compact subsurface $F \subset S$ of the possibly noncompact surface S is *incompressible* if no component of $S - F$ is a 2-disk with boundary contained in ∂F.

Lemma 7. *If $F \subset S$ is a compact connected incompressible subsurface of a connected surface S, then the map $\pi_1(F) \to \pi_1(S)$ induced by the inclusion at fundamental group level is injective.*

Proof. When S is noncompact, it is an ascending union of compact subsurfaces S_n so that $S_{n+1} - \mathrm{int}(S_n)$ have no disk components. This reduces the claim to the case when S is compact. Furthermore, we can assume that $S - \mathrm{int}(F)$ is connected, as we can adjoin components one by one.

If $S - \mathrm{int}(F)$ has at least one boundary component outside ∂F, then the map $H_1(F) \to H_1(S)$ induced by the inclusion is injective. In fact a basis of $H_1(F)$ can be completed to a basis of $H_1(S)$. As $\pi_1(F)$ and $\pi_1(S)$ are free groups, we deduce that the image of $\pi_1(F)$ into $\pi_1(S)$ is a free group of rank $H_1(F)$, namely isomorphic to $\pi_1(F)$. As $\pi_1(F)$ is free, it is Hopfian by Corollary 6 and hence the map $\pi_1(F) \to \pi_1(S)$ is injective.

Suppose next that the boundary of $S - \mathrm{int}(F)$ is contained within the boundary ∂F of F. Since $S - \mathrm{int}(F)$ is not a disk, it either contains an annulus or a Moebius band. From above we reduce ourselves to the case when $S - \mathrm{int}(F)$ is either an annulus or a Moebius band.

If $S - \mathrm{int}(F)$ is an annulus, then let $\widehat{F}$ be the surface obtained by gluing two copies of F glued together along the boundary circles of this annulus. Then F embeds in $\widehat{F}$ as one of the two copies, while $\widehat{F}$ double covers S and the inclusion $F \to S$ factors through the covering $\widehat{F}$. Note that $\pi_1(\widehat{F}) \to \pi_1(S)$ is an isomorphism onto a subgroup of index 2, by construction. It remains to show that the map $\pi_1(F) \to \pi_1(\widehat{F})$ is injective using the strategy from above. Observe that $H_1(F)$ is of rank $1 - \chi(F)$ while the rank of $H_1(\widehat{F})$ is either $2 - \chi(\widehat{F})$, when $\widehat{F}$ is closed, or $1 - \chi(\widehat{F})$ otherwise. Note that

$$\chi(S) = \chi(F) + \chi(S - \mathrm{int}(F)) = \chi(F).$$

On the other hand $\widehat{F}$ is a double covering of S, so that

$$\chi(\widehat{F}) = 2\chi(F).$$

By direct inspection the map $H_1(F) \to H_1(\widehat{F})$ induced by the inclusion is injective.

If $\widehat{F}$ is not closed, then the image of $\pi_1(F)$ within $\pi_1(\widehat{F})$ is a free group of rank equal to the rank of $H_1(F)$. The image is isomorphic then to $\pi_1(F)$ and since free groups are Hopfian by Corollary 6, we deduce that $\pi_1(F) \to \pi_1(\widehat{F})$ is injective.

Assume that $\widehat{F}$ is closed. By comparing the ranks we derive that the image of $H_1(F)$ is of infinite index within $H_1(\widehat{F})$, unless F is a Moebius band and $\widehat{F}$ a Klein bottle. In the later case the inclusion map $\pi_1(F) \to \pi_1(N_2)$ is injective, by direct inspection. Eventually, the image of $\pi_1(F)$ within $\pi_1(\widehat{F})$ is a free group of the same rank as $H_1(F)$, because it is of infinite index and we conclude as above.

Eventually let $S-\mathrm{int}(F)$ be a Moebius band. A Moebius band is the quotient of a cylinder by an involution which exchanges the two boundary components. Let $\widehat{F}$ be the union of two copies of F glued together by a cylinder and observe that S is the quotient of $\widehat{F}$ by an involution exchanging the two copies. Then $\widehat{F} \to S$ is a double covering. The injectivity of $\pi_1(F) \to \pi_1(\widehat{F})$ follows along the same lines as in the previous case.

Lemma 8. *Let S be a surface such that $\pi_1(S)$ is finitely generated and $C \subseteq S$ be a compact subset. Then there is a compact, connected and incompressible subsurface $F \subseteq S$ such that the map induced by inclusion $\pi_1(F) \to \pi_1(S)$ is an isomorphism.*

Proof. Choose some based loops representing a finite generating set of classes in $\pi_1(S)$ and let S' be a regular neighborhood of their union. Then S' is a compact connected subsurface and $\pi_1(S') \to \pi_1(S)$ is surjective, by construction. We can enlarge S' in order to contain C. If S' is not incompressible, then some of its boundary components bound embedded disks in S. Define F to be result of adjoining these disks to S'. Then F satisfies all requirements, by Lemma 7.

Lemma 9. *If S is an open connected and simply connected surface then S is homeomorphic to $\mathbb{R}^2$.*

Proof. Lemma 8 tells us that for any compact $C \subset S$ we can find some compact connected subsurface $F \subset S$ containing C in the interior such that $\pi_1(F) = 1$. By the classification of compact surfaces F is homeomorphic to a disk, as S cannot have sphere components. It follows that S is an ascending union of disks, each disk being contained in the interior of the next one. This implies that S is homeomorphic to $\mathbb{R}^2$.

Lemma 10. *If S is an open connected surface and has fundamental group $\mathbb{Z}$, then S is either homeomorphic to an open annulus or else to an open Moebius band.*

Proof. Again by Lemma 8, for any compact $C \subset S$ we can find some compact connected subsurface $F \subset S$ containing C in the interior such that $\pi_1(F) = \mathbb{Z}$. By the classification of compact surfaces F is homeomorphic to either an annulus, or else to a Moebius band. Therefore S is the ascending union of compact surfaces F_i each of which is either an annulus or a Moebius band, such that $F_i \subset \mathrm{int}(F_{i+1})$ and the inclusion map induces an isomorphism $\pi_1(F_i) \to \pi_1(F_{i+1})$.

If infinitely many F_i are annuli, then all F_i must be annuli since they are orientable and so cannot contain Moebius bands. Recall that $\pi_1(F_i) \cong \mathbb{Z}$ is generated by the class of an oriented meridian ℓ_i on the cylinder F_i. It follows that, up to changing the orientation of ℓ_i, the based loops ℓ_i and ℓ_{i+1} are homotopic within F_{i+1}. By the proof of Lemma 20 we can alter ℓ_{i+1} by an isotopy supported on F_{i+1} such that $\ell_{i+1} = \ell_i$. This implies that there exists a collection of homeomorphisms $h_i : F_i \to S^1 \times [-i, i]$ such that $h_{i+1}|_{F_i} = h_i$. Therefore S is homeomorphic to $S^1 \times \mathbb{R}$.

In the remaining case we can assume that all F_i are Moebius bands, so that S is nonorientable. Then the orientable double covering of S is the union of annuli and hence an open cylinder, from above. It follows that S is an open Moebius band.

Lemma 11. *Let S be a surface such that $\pi_1(S)$ is finitely generated and let F be a compact, connected and incompressible subsurface $F \subseteq S$ such that the map induced by inclusion $\pi_1(F) \to \pi_1(S)$ is an isomorphism. Then every connected component of $S - F$ is an annulus.*

Proof. By Lemma 8 S is the ascending union of compact connected incompressible subsurfaces F_i such that $F_i \subset \text{int}(F_{i+1})$ and the inclusion map induces an isomorphism $\pi_1(F_i) \to \pi_1(F_{i+1})$.

If two distinct boundary components a and b of F_i can be connected by an arc u within $F_{i+1} - \text{int}(F_i)$, then we obtain a circle c by joining the endpoints of u by an arc v within F_i. Then c intersects a in a single point, so that the algebraic intersection number between the homology classes of these curves suitably oriented is $\omega(a, c) = \pm 1$. If the class of c in homotopy were in the image of $\pi_1(F_i)$ by the inclusion map, then we would have $\omega(a, c) = 0$, because any curve on F_i can be homotoped off the boundary. This contradicts then the assumptions. It follows that every connected component of $F_{i+1} - \text{int}(F_i)$ only intersects F_i along a single boundary circle of the later.

Let c be a simple loop within some connected component N of $F_{i+1} - \text{int}(F_i)$ which is not nullhomotopic. By hypothesis its homotopy class is in the image of $\pi_1(F_i)$. When crushing all points of $F_{i+1} - \text{int}(N)$ to a single point we obtain a surface $\widehat{N}$. Then the image of c in the $\widehat{N}$ should be nullhomotopic and hence bounds an embedded disk in $\widehat{N}$. Its preimage in N is therefore a cylinder which provides an isotopy between c and the boundary circle $N \cap F_i$. This proves that $\pi_1(N) = \mathbb{Z}$ and further that N is an annulus.

2.3 The Greenberg-Griffiths-Schreier theorem

Theorem 6 (Greenberg-Griffiths-Schreier). *Let S be a connected surface whose fundamental group is not virtually abelian. Let $1 \neq N \triangleleft \pi_1(S)$ be a normal subgroup and G a finitely generated group such that*

$$N \subseteq G \subseteq \pi_1(S).$$

Then G is a a finite index subgroup of $\pi_1(S)$.

The result was first proved by Schreier in [7] in the case of free groups (see also [5]). We follow here [3]. A geometrical proof appeared in [2].

Lemma 12. *If $1 \neq N \triangleleft \pi_1(S)$ is a normal subgroup and $\pi_1(S)$ is not virtually abelian, then N has at least two generators.*

Proof. Assume the contrary, namely that N is cyclic, generated by some element $a \neq 1$. Then for any $x \in \pi_1(S)$ there exists some $k(x) \in \mathbb{Z}$ such that

$$xax^{-1} = a^{k(x)}$$

Since $\pi_1(S)$ is torsion-free, the value of $k(x) \in \mathbb{Z}$ is uniquely determined. Moreover, $k(x) \neq 0$, as $a \neq 1$. It follows that k is a monoid homomorphism $k : \pi_1(S) \to \mathbb{Z}^*$, namely that:

$$k(x)k(y) = k(xy),\ k(1) = 1$$

In particular, we have

$$k(x)k(x^{-1}) = k(1) = 1, \text{ for any } x \in \pi_1(S).$$

As k takes values in $\mathbb{Z}$, $k(x) \in \{-1, 1\}$.

Let $K = \ker k \triangleleft \pi_1(S)$. As the image of k is contained into $\{-1, 1\}$, either $K = \pi_1(S)$, or else K has index 2 in $\pi_1(S)$.

In the first case $k \equiv 1$, so that all elements of $\pi_1(S)$ commute with a, so that $\langle a \rangle$ is central. But $\pi_1(S)$ has trivial center, when it is not virtually abelian.

In the second case a commutes with every element of K and $k(a) = 1$, so that $\langle a \rangle$ is central in K. Let S' be the surface associated to the index two subgroup $K \subset \pi_1(S)$. Then S' is a double covering of S with $\pi_1(S') = K$ and hence having a nontrivial center $\langle a \rangle$. Therefore $\pi_1(S')$ is virtually abelian, and hence S' is either the Klein bottle N_2, or else $\pi_1(S')$ is abelian, isomorphic to either $\mathbb{Z}$, $\mathbb{Z}/2\mathbb{Z}$ or $\mathbb{Z} \oplus \mathbb{Z}$.

If S' were a Klein bottle or a torus then $2\chi(S) = \chi(S') = 0$, hence S should be a Klein bottle or a torus, which contradicts our assumptions. If S' were RP^2, then $\pi_1(S)$ would have 4-elements, which is impossible as S is connected. Eventually, if $\pi_1(S') = \mathbb{Z}$, then S' would be an annulus or a Moebius band. This implies that S is also an annulus or a Moebius band, contradicting our assumptions.

Lemma 13. *Let $X_N \xrightarrow{f} X_G \xrightarrow{g} X$ be coverings associated to the groups $N \subseteq G \subseteq \pi = \pi_1(X)$, where X is locally connected, locally compact and path connected topological space. Assume that $N \triangleleft \pi$ is a normal subgroup of infinite index. Then for every compact sets $A \subseteq X_N$, $B \subseteq X_G$, we have*

$$f(\gamma \cdot A) \cap B = \emptyset$$

for all but finitely many $\gamma \in \pi/N$.

Proof. Consider cosets representatives $\gamma_i \in \pi$ for the classes in π/G, namely such that

$$\pi = \sqcup_{i=1}^{\infty} \gamma_i \cdot G$$

If $\overline{\gamma_i}$ denote their classes in π/N, we can also write

$$\pi/N = \sqcup_{i=1}^{\infty} \overline{\gamma_i} \cdot G/N$$

as $\frac{\pi/N}{G/N}$ is isomorphic to π/G. Every point $x \in X$ has an open neighborhood W which is evenly covered by both $g^{-1}(W)$ and $(g \circ f)^{-1}(W)$. Components of $g^{-1}(W)$ are in bijection with the set π/G and we can write

$$g^{-1}(W) = \sqcup_{i=1}^{\infty} V_i$$

We derive then that

$$(g \circ f)^{-1}(W) = \sqcup_{i=1}^{\infty} f^{-1}(V_i).$$

On the other hand, $g \circ f : X_N \to X$ is a normal covering, so that π/N acts freely and properly discontinuously on X_N as a deck transformation group, so that

$$(g \circ f)^{-1}(W) = \sqcup_{\mu \in \pi/N} \mu \cdot U$$

where U is some connected component of $(g \circ f)^{-1}(W)$. It then follows that

$$f^{-1}(V_i) = \sqcup_{\overline{\mu} \in G/N} \overline{\gamma}_i \overline{\mu} \cdot U$$

Given now any compact $B \subset X_G$, and $\overline{\gamma} \in \pi/N$ there exist only finitely many i such that

$$f(\overline{\gamma}_i \overline{\gamma} \cdot U) \cap B \neq \emptyset$$

because, for distinct values of i the sets $f(\overline{\gamma}_i \overline{\gamma} \cdot U)$ are different connected components V_j.

Furthermore, if $A \subset X_N$ is compact, we can find a finite open covering of $g \circ f(A)$ by evenly covered sets W_t. Then A is covered by finitely many open sets of the form $\overline{\gamma}_r \overline{\mu_s} \cdot U_t$, where $\overline{\mu_j} \in G/N$ and U_s is a connected component of $(g \circ f)^{-1}(W_t)$. From above we derive that for large i we have:

$$f(\overline{\gamma}_i \cdot A) \cap B \subseteq \bigcup_{t,r,s} f(\overline{\gamma_i} \cdot \overline{\gamma}_r \overline{\mu_s} \cdot U_t) \cap B = \emptyset$$

as claimed.

Proof (Proof of Theorem 6). We can assume that S is either a closed surface without boundary or an open surface. Let then $S_N \xrightarrow{f} S_G \xrightarrow{g} S$ be coverings associated to the groups $N \subseteq G \subseteq \pi_1(S)$. Assume that G is of infinite index in $\pi_1(X)$. Then S_G and S_N are open surfaces.

By Lemma 12 the subgroup $N \subseteq G$ has at least two generators; let α and β be based loops in S_G representing two elements which do not generate a cyclic subgroup. By Lemma 8 there exists $F \subset S_G$ a compact connected incompressible subsurface containing α and β such that $\pi_1(F) \to \pi_1(S)$ is an isomorphism. Moreover, since their classes are in N, the loops α and β lift to some based loops α' and β' on S_N. Set A to be their union.

By Lemma 13 for all but finitely many $\gamma \in \pi_1(X)/N$ we have

$$f(\gamma \cdot A) \cap F = \emptyset$$

Note now that the loops $f(\gamma \cdot \alpha')$ and $f(\gamma \cdot \beta')$ belong to the same component of $S_G - F$. However, Lemma 11 shows that every such connected component is homeomorphic to an annulus, in particular its fundamental group is $\mathbb{Z}$. Therefore, the subgroup of $\pi_1(S_G - F)$ generated by the homotopy classes of the based loops $f(\gamma \cdot \alpha')$ and $f(\gamma \cdot \beta')$ is cyclic. It follows that the subgroup $\langle \gamma \cdot \alpha', \gamma \cdot \beta' \rangle \subseteq \pi_1(S_N)$ is cyclic. As γ acts by a homeomorphism of S_N the subgroup $\langle \alpha', \beta' \rangle$ is cyclic and hence the subgroup $\langle \alpha, \beta \rangle \subseteq \pi_1(S_U)$ is cyclic, contradicting our choice.

If $d \geq 1$ then we define the *power subgroup* G^d of the group G as the subgroup generated by the d-th powers of all elements of G. Observe that $G^d \triangleleft G$ is a normal subgroup of G.

Corollary 7. *Let S be a surface whose fundamental group $\pi = \pi_1(S)$ is not virtually abelian. If $G \subseteq \pi_1(S)$ is a finitely generated subgroup then G is of finite index if and only if there exists some d such that G contains the power subgroup π^d.*

Proof. By the Greenberg-Griffiths-Schreier Theorem 6 G is of finite index in π, if it contains some power subgroup, which is normal subgroup. Conversely, if d is the index of G in π, then $x^d \in G$ for every $x \in \pi$, so $G \supseteq G^d$.

Remark 2. The Burnside problem asks whether G^d is of infinite index in G, for a free group G on two generators. S. Adian and P.S. Novikov (see [1]), I.G. Lysenök ([6]) and S. I. Ivanov ([4]) proved this is so for large enough $d \geq 8000$. The case $d = 5$ is still open.

References

1. S. Adian, The Burnside problem and identities in groups, translated from Russian by John Lennox and James Weigold, *Ergebnisse Math. Grenz.* 95, Springer-Verlag 1979.
2. L. Greenberg, *Discrete groups of motions*, Canadian J. Math. 12 (1960), 415–426.
3. H.B.Griffiths, *The fundamental group of a surface, and a theorem of Schreier*, Acta Math. 110 (1963), 1–17.
4. Sergei I. Ivanov, *The free Burnside group of sufficiently large exponents*, Internat. J. Algebra Comput. 4 (1994), no. 1-2.
5. A. Karrass and D. Solitar, *Note on a theorem of Schreier*, Proc. Amer. Math. Soc. 8 (1957), 696.
6. I. G. Lysenök, *Infinite Burnside groups of even period*, Izvest. Math. 60 (1996), 453–654.
7. O. Schreier, *Die Untergruppen der freien Gruppen*, Abh. Math. Sem. Univ. Hamburg 5 (1928).

2.3.1 Locally extended residual finiteness

Definition 4. *A group G is* `locally extended residually finite, abbreviated LERF` *if for any finitely generated subgroup $H \subset G$ and any element $g \in G - H$ there exists a finite index subgroup K of G such that $H \subseteq K$ while $g \notin K$.*

This separability property can be characterized topologically, as follows:

Lemma 14. *Let $\widehat{X} \to X$ be a normal covering with deck group G of a Hausdorff space. Then G is LERF if and only if for any finitely generated subgroup $H \subset G$ and compact $C \subset \widehat{X}/H$ there exists some finite index subgroup $K \subseteq G$ such that C projects by a homeomorphism into the finite covering $X_K = \widehat{X}/K$.*

Proof. To prove the "if" part let $g \in G - H$ and take for C the image of $\{x, g \cdot x\}$ in $\widehat{X}/H$ for some $x \in \widehat{X}$. If X_K is a finite covering as in the statement then $g \notin K$ because the images of x and of $g \cdot x$ are distinct in X_K. Thus G is LERF.

Conversely, select one connected component for the preimage in $\widehat{X}$ of each member a finite open covering of C by evenly covered sets. Intersecting these open sets with the preimage of C provides us with a compact set $\widehat{C} \subset \widehat{X}$ which projects onto C. Since G acts freely and properly discontinuously on $\widehat{X}$, the set

$$A = \{g \in G; g \cdot \widehat{C} \cap \widehat{C} \neq \emptyset\}$$

is finite. Since G is LERF, for each such element $g \notin H$ there exists a finite index subgroup $K_g \triangleleft G$ such that $K_g \supseteq H$ while $g \notin K_g$. Then their intersection $K = \cap_{g \in A-H} K_g$ is still a finite index subgroup of G containing H. If moreover $g \in K \cap A$, then $g \in H$. This implies that the image of $\widehat{C}$ in the quotient X_K is homeomorphic to C, as required.

A subgroup $G \subseteq \pi_1(S)$ of the fundamental group $\pi_1(S)$ of a surface is *geometric* if there exists a compact incompressible subsurface $F \subset S$ such that G is the image of $\pi_1(F)$ by the morphism induced by the inclusion $\pi_1(F) \to \pi_1(X)$.

Lemma 15. *The surface group $\pi_1(S)$ is LERF if and only if for any finitely generated subgroup $G \subset \pi_1(S)$ and element $g \in \pi_1(S) - G$ there exists a finite covering $S' \to S$ such that the image of $\pi_1(S')$ in $\pi_1(S)$ contains G but not g and G is a geometric subgroup of $\pi_1(S')$.*

Proof. For the "only if" part, let S_G be the covering of S corresponding to the subgroup $G \subset \pi_1(S)$ and C be the image of the set $\{x, g \cdot x\}$ in S_G, where x is some point in the universal covering $\widetilde{S}$. By Lemma 7 there exists some compact connected incompressible subsurface $F \subseteq S_G$ containing C and such that the map induced by inclusion induces an isomorphism $\pi_1(F) \to \pi_1(S_G)$. Lemma 14 provides us with an intermediary finite covering $S_G \to S' \to S$ such that F projects homeomorphically onto its image in S'.

The fisrt example of LERF groups comes from the following result proved by Hall ([1]):

Theorem 7 (Hall 1949). *If X is a compact surface with nonempty boundary, $H \subset \pi_1(S)$ a finitely generated subgroup and $g \in \pi_1(X) - H$, then there exists a finite covering $\overline{F} \to S$ such that the image of $\pi_1(\overline{F})$ within $\pi_1(S)$ contains H but not g, while H is a geometric subgroup of $\pi_1(\overline{F})$. In particular, free groups are LERF.*

Proof. Note S is obtained from a polygone D by identifying some of its boundary edges in pairs. Specifically, some edges of D are labeled by signed letters, such that two edges which are paired correspond to labels of the form a and a^{-1}. In order to obtain a finite covering of S we consider several copies of D and glue together pairs of edges whose labels of the form a and a^{-1}.

Let $S_H \to S$ be the covering associated to the subgroup H. By above, there is a tesselation of S_H by copies of D. As $g \notin H$, the lift of a loop having the homotopy class g is a path P in S_H whose endpoints are distinct. By Lemma 7 there exists a compact incompressible surface $F \subset S_H$ containing the path P such that the inclusion induces an isomorphism $\pi_1(F) \to \pi_1(S_H)$. Let $\widehat{F}$ be the union of the finitely many tiles of the tesselation which intersect F nontrivially. By induction on the number of copies, the labels arising on the those edges lying in $\partial\widehat{F}$ come in pairs with opposite signs, i.e. a and a^{-1}. We can therefore glue together these edges according to the label pairings and produce a compact surface with boundary $\overline{F}$ containing $\widehat{F}$. Moreover, $\overline{F}$ is a finite covering of S, by construction. It follows that $\overline{F}$ satisfies all requirements.

Lemma 16. *If the group G is LERF then every subgroup of it is also LERF. Moreover, if $G \subseteq K$ has finite index, then K is LERF.*

Proof. The result for subgroups is immediate. Replacing G with a finite index normal subgroup in K contained in G, we can assume that G is normal in K. Let $H \subseteq K$ be a finitely generated subgroup and $f \in K - H$.

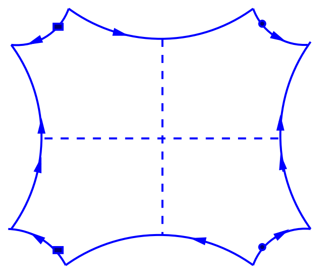
Let GH be the subgroup of K generated by G and H. If $f \notin GH$, then we are done since $GH \subseteq K$ is of finite index. As G is normalized by H we can write $g = gh$, with $g \in G$ and $h \in H$. In particular $g \notin H$. Now $H \cap G \subseteq H$ is a finite index subgroup of H and hence also finitely generated. The LERF property for G implies that there exists some finite index subgroup $L \subset G$ containing $G \cap H$ such that $g \notin L$.

We can assume that L is normalized by H, as H is finitely generated. If $f \in LH$, then we can write $gh = \ell h'$, with $\ell \in L$ and $h' \in H$. This implies that $\ell^{-1}g = h'h^{-1}$, so that $\ell^{-1}g \in H \cap G \subset L$, contradicting the fact that $g \notin L$.

Lemma 17. *There exists a regular pentagon P in the hyperbolic plane $\mathbb{H}^2$ having all angles equal to $\frac{\pi}{2}$.*

Proof. The angles of a regular pentagon go to 0, when the length of its sides goes to infinity and go to $\frac{3\pi}{5}$, when the length of its sides goes to 0 as the pentagon is approximatively Euclidean. The intermediate value theorem yields our claim.

Let Γ be the group of isometries of the plane generated by the reflections in the sides of P. The translates of P by Γ tesselate the plane, because angles of P are $\frac{\pi}{2}$. By the Poincaré Theorem, P is a fundamental domain for the action of Γ. If we reflect the pentagon P along the edges incident to a vertex the four copies so obtained form an octagon Q, as in the figure below.



By the Poincaré Theorem, Q is a fundamental domain for the action of a subgroup $G \subset \Gamma$ of index 4. By identifying edges of Q as in the figure above, we see that G is the fundamental group $\pi_1(N_3)$ of the closed nonorientable surface of genus 3. We denote by L the set of geodesics containing the sides of P and their orbits by the group Γ.

Lemma 18. *The set of points at distance ε from a geodesic γ in the hyperbolic plane $\mathbb{H}^2$ is an arc of circle which passes through the points at infinity of γ and cut the circle at infinity at the same angle.*

Proof. In the uppe half-plane model we use a conformal mapping sending the geodesic γ into the vertical imaginary axis. If θ is a half-line issued from the origin, then the homotethies with center at the origin preserve $\gamma \cup \theta$ and act isometrically and transitively on it. In particular θ is the set of equidistant points to γ. The image of θ by the inverse projective transformation is an arc of circle, as claimed.

The set of points at a fixed distance from a geodesic in $\mathbb{H}^2$ is usually called a *hypercycle*, in the litterature.

The main result of this section is the following theorem due to Scott ([4, 5]):

Theorem 8 (Scott 1978). *Surface groups are LERF.*

Proof (Proof of Theorem 8). The result is trivial when $\chi(S) > 0$. If S is torus, then $\pi_1(S) = \mathbb{Z} \oplus \mathbb{Z}$ and the result is immediate. If S is a Klein bottle, then $\pi_1(S)$ is virtually $\mathbb{Z} \oplus \mathbb{Z}$ and the result follows from Lemma 16. Any closed surface S with $\chi(S) \leq -1$ covers N_3. Thus $\pi_1(S) \subseteq \pi_1(N_3)$ and it suffices to show that $\pi_1(N_3)$ is LERF.

Let now $H \subset G$ be a finitely generated subgroup, $p : \mathbb{H}^2 \to S = \mathbb{H}^2/H$ be the projection and $C \subset S$ be a compact. As in the proof of Lemma 14 there exists a compact $\widehat{C} \subset \mathbb{H}^2$ which projects onto C. We have to prove that there exists a finite index subgroup $K \subseteq G$ which contains H such that

$$\{g \in K; g \cdot \widehat{C} \cap \widehat{C} \neq \emptyset\} \subseteq H.$$

As $\pi_1(S) = H$ is finitely generated, by Lemma 7 there exists a compact connected subsurface $F \subset S$ such that $C \subseteq \mathrm{int}(F)$ and the map $\pi_1(F) \to \pi_1(S)$ induced by the inclusion is an isomorphim. Let $\widehat{F} = p^{-1}(C)$ and define $\overline{F}$ to be the intersection of all closed half-spaces in $\mathbb{H}^2$ which contain $\widehat{F}$ in their interior and are bounded by a line in L. Then $\overline{F}$ is a convex set in $\mathbb{H}^2$ and it is union of translates of the pentagon P. As $\widehat{F}$ is H-invariant, it follows that $\overline{F}$ is also H-invariant and so $p(\overline{F})$ is also a union of pentagons in S.

Assume that $p(\overline{F})$ is compact. Denote by J the group generated by the reflections in the sides of $\overline{F}$. By the Poincaré Theorem $\overline{F}$ is a fundamental domain for the action of the group J on $\mathbb{H}^2$. Set eventually

$$K = \langle J, H\rangle.$$

Then J is a normal subgroup of K because both L and $\overline{F}$ are H-invariant: the image by $h \in H$ of a reflection along a side of $\overline{F}$ is the reflection along another side of $\overline{F}$. Furthermore K/J is isomorphic to H.

If $B \subset \overline{F}$ is a fundamental domain for the action of H on $\overline{F}$ then B is compact. By above B is also a fundamental domain for the action of K on $\mathbb{H}^2$, because

$$\mathbb{H}^2/K = (\mathbb{H}^2/J)/H = \overline{F}/H = p(\overline{F})$$

If $p(\overline{F})$ is compact, then it consists of finitely many pentagons of S and hence K has finite index in the group Γ. If $g \in K$ is such that

$$g \cdot \widehat{C} \cap \widehat{C} \neq \emptyset$$

then

$$g \cdot \mathrm{int}(\overline{F}) \cap \mathrm{int}(\overline{F}) \neq \emptyset$$

and hence $g \in H$. Thus the group $K \cap G \subset G$ satisfies the requirements.

It remains to prove that $p(\overline{F})$ is compact. Every boundary circle of the subsurface $F \subset S$ is nontrivial in homotopy and thus freely homotopic to a unique geodesic in the hyperbolic metric on $S = \mathbb{H}^2/H$. These family of geodesics define a subsurface $N \subset S$ which is homotopic to F. By compacteness, there exists some constant ε so that F is contained in the set N_ε of points in S at distance at most ε from N. Then $\widehat{N} = p^{-1}(N)$ is connected and $\widehat{F} \subseteq \widehat{N}_\varepsilon$. The set $\widehat{N}$ is connected and bounded by geodesics.

Every H-orbit of a line in L intersecting $\mathrm{int}(\overline{F})$ should also intersect $\widehat{F}$ and there are only finitely many such lines because $\widehat{F}/H = F$ is compact. It then suffice to see that projections $p(\ell)$ of such lines ℓ into S are compact in order to find that $p(\overline{F})$ consists of finitely many pentagons.

In order to prove that we show that going far away along ℓ there exists a point and a geodesic orthogonal to ℓ which does not intersect $\widehat{N}_\varepsilon$. This implies that the projection is compact.

Consider a lift γ of a boundary geodesic of $\widehat{N}$ to $\mathbb{H}^2$ which joins the points a and b from the real line in the upper half-plane model. We can assume that the lift of ℓ is a vertical line intersecting the geodesic above. Now a hypercycle is obviously bounding a convex set on one side. Thus there exist geodesics orthogonal to ℓ in the convex side which are arbitrarily closed to the foot of ℓ. This proves the claim.

References

1. M. Hall, *Coset representations in free groups*, Trans. Amer. math. Soc. 67 (1949), 421–432.
2. J. Hempel, *Residual finiteness of surface groups*, Proc. Amer. math. Soc. 32 (1972), 323.
3. J. Hempel, 3-manifolds, Ann. Math. Studies 86, Princeton Univ. Press, 1976.
4. P. Scott, *Subgroups of surface groups are almost geometric*, J. London Math. Soc. (2), 17 (1978), 555–565.
5. P. Scott, *Correction to "Subgroups of surface groups are almost geometric"*, J. London Math. Soc. (2), 32 (1985), 217–220.

3

Homotopy and Isotopy

3.1 Simple curves and homotopy

Definition 5. *Two continuous maps* $f_0, f_1 : X \to Y$ *between topological spaces* X *and* Y *are said to be* `homotopic`, *if there exists a continuous map*

$$f : X \times [0,1] \to Y$$

extending both, namely such that:

$$f(x,0) = f_0(x) \text{ and } f(x,1) = f_1(x), \text{ for } x \in X;$$

Definition 6. `Embeddings` *are continuous maps which are homeomorphisms onto their images. Two embeddings* $f_0, f_1 : X \to Y$ *between topological spaces* X *and* Y *are said to be* `isotopic` *if there exists en embedding*

$$f : X \times [0,1] \to Y \times [0,1]$$

preserving the levels:

$$f(X \times \{t\}) \subseteq Y \times \{t\}$$

and extending both, namely:

$$f(x,0) = f_0(x) \text{ and } f(x,1) = f_1(x), \text{ for } x \in X;$$

Moreover, they are `ambient isotopic` *if there exists a homeomorphism, called an* `ambient isotopy`

$$h : Y \times [0,1] \to Y \times [0,1]$$

preserving the levels:

$$h(Y \times \{t\}) = Y \times \{t\}$$

such that:

$$h(y,0) = y, \text{ for every } y \in Y$$

$$h(f_0(x),1) = f_1(x), \text{for every } x \in X.$$

A second basic result needed in the sequel adds the uniqueness to the existence of smooth structure on surfaces:

Theorem 9. *Every homeomorphism between smooth surfaces is isotopic to a diffeomorphism.*

The first proof of this result is attributed to Epstein 1966, although the case of closed orientable surfaces follows from the work of Baer 1928.

Remark 3. Depending on the context we might consider topological, piecewise linear or smooth maps and structures on the spaces. In dimensions at most 3 Moise proved that all these categories are equivalent, namely topological manifolds have unique smooth structures and their topological classification coincide with their smooth classification. This is not anymore the case in higher dimensions.

We have the following fundamental result due to Baer:

Theorem 10 (Baer 1927, Homotopic curves are isotopic). *Let $f_0, f_1 : S^1 \to S$ be embeddings in the interior int(S) of the compact orientable surface S such that $f_1(S^1)$ does not bound a disk in S. If f_0 is homotopic to f_1 then f_0 and f_1 are ambient isotopic, by an isotopy which is trivial outside a compact of* $\mathrm{int}(S)$.

Remark 4. The analogous result is not true if $f_1(S^1)$ bounds a disk, except when $S = S^2$: take for $f_1(S^1)$ the same curve as $f_0(S^1)$ with opposite orientation. As ambient isotopies preserve the orientation, the two oriented curves and hence the maps f_0 and f_1 are not isotopic. If $f_1(S^1)$ intersects the boundary ∂S then the conclusion is slightly weaker, namely f_0 and f_1 are isotopic.

3.2 Proof of Baer's Theorem

We present here a proof due to D.B.A. Epstein.

Lemma 19 (Disk Lemma). *Let $p : \widehat{S} \to S$ be a unramified covering with noncompact $\widehat{S}$. Suppose that there exists an embedded disk $D^2 \subset \widehat{S}$ such that the restriction of p to its boundary curve ∂D^2 is injective. Then the restriction of p to D^2 is also injective.*

Proof. Consider the common universal covering $\widetilde{S}$ of $\widehat{S}$ and S, which is endowed with the action of the deck transformation group $\pi_1(S)$. The covering map $\widetilde{p} : \widetilde{S} \to S$ can therefore be identified with thenatural quotient map by the action of $\pi_1(S)$ on $\widetilde{S}$.

Let D be a connected component of a lift of D^2 to $\widetilde{S}$, which should still be a 2-disk, as it is simply connected. Assume that there exists a pair of points $x, y \in D$ having the same image $\widetilde{p}(x) = \widetilde{p}(y)$ in S. Then, by the definition of the deck transformation group, there exists $\varphi \in \pi_1(S)$ such that $\varphi(x) = y$. It follows that $\varphi(D) \cup D$ is a compact connected subset of $\widetilde{S}$. On the other hand, since p is injective, when restricted to ∂D^2 we have

$$\varphi(\partial D) \cap \partial D = \emptyset,$$

so that $\varphi(D) \cup D$ is a codimension zero submanifold of $\widetilde{S}$.

The boundary of this submanifold cannot have two components, namely $\varphi(\partial D) \cup \partial D$, because ∂D is contractible to a point in $\varphi(D) \cup D$. Therefore the boundary of $\varphi(D) \cup D$ is either ∂D or else $\varphi(\partial D)$. This implies that either $\varphi(D) \subset D$ or else $D \subset \varphi(D)$. By the symmetry of the situation we can assume the first situation arises. It then follows that

$$\varphi(D) \subset \mathrm{int}(D).$$

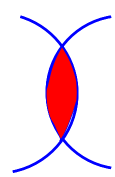
By iterating this argument we obtain:

$$\varphi^k(D) \subset \mathrm{int}(\varphi^{k-1}(D)), \text{ for every } k \geq 1$$

so that infinitely many circles $\varphi^k(\partial D)$ are packed within $\mathrm{int}(D)$. But $\mathrm{int}(D)$ has compact closure and so it can only intersect finitely many components of the preimage of ∂D^2 in $\widetilde{S}$, which is a contradiction.

This shows that $\widetilde{p}$ is injective, when restricted to D, from which the claim follows.

Lemma 20 (Bigon existence). *Consider two homotopic simple closed curves α and β on the surface S, whose intersection is non-empty. Then there exists an embedded disk D^2 whose boundary consists of an arc of α union an arc of β:*



Remark 5. The statement is not true when the two curves are not homotopic, as it can be seen by taking a meridian and a longitude on a torus.

Proof. Let $p : \hat{S} \to S$ be the unramified covering of S associated to the subgroup generated by the homotopy class of α in $\pi_1(S)$. Thus $\pi_1(\hat{S}) = \mathbb{Z}$ is generated by the class of a lift of α and so $\hat{S}$ is a cylinder. Let $\hat{\alpha}$ and $\hat{\beta}$ be simple closed lifts of the curves α and β, respectively. The homotopy between α and β lifts then to a homotopy between $\hat{\alpha}$ and $\hat{\beta}$. Both $\hat{\alpha}$ and $\hat{\beta}$ must separate $\hat{S}$, as otherwise they would be null-homotopic.

Choose an arc b contained in $p^{-1}(\beta) \subset \hat{S}$ which only intersects $\hat{\alpha}$ along its endpoints. It follows that the endpoints of b determine an arc a of $\tilde{\alpha}$ such that the union of a and b bounds an embedded disk D^2 in $\hat{S}$.

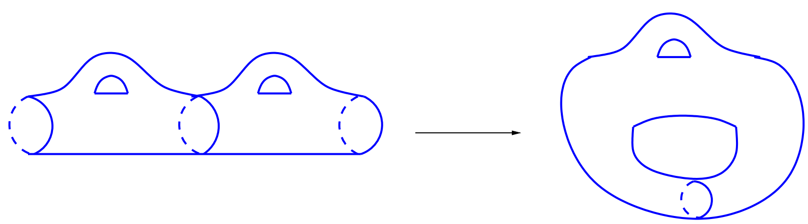
Now, D^2 intersects $p^{-1}(\beta)$ along finitely many arcs having their endpoints in a. Changing $\hat{\beta}$ with another component of $p^{-1}(\beta)$ we can assume that D^2 is an innermost disk, so that $\mathrm{int}(D^2)$ is disjoint from $p^{-1}(\beta)$. Similarly, by replacing the lift $\hat{\alpha}$ with another connected component of $p^{-1}(\alpha)$, we can insure that $\mathrm{int}(D^2)$ is also disjoint from the preimages of $p^{-1}(\alpha)$.

We are now in position for applying Lemma 19.

Lemma 21 (Nullhomotopic curves bound disks). *A nullhomotopic simple closed curve γ in a surface S bounds an embedded disk in S.*

Proof. When $S = S^2$, it is the classical Jordan-Schoenflies Theorem. By Lemma 19 it will suffices to consider S simply connected. By replacing S with $\mathrm{int}(S)$, we can assume that S has no boundary. As γ is nullhomotopic, its class in homology is trivial. We can therefore write γ as the boundary ∂c, of some singular 2-chain c with $\mathbb{Z}/2\mathbb{Z}$ coefficients. Then the union N of all 2-simplices arising in c with nonzero coefficient is a codimension zero submanifold of S whose boundary ∂N is γ.

If N is not a 2-disk, then it is either a Moebius band or else N contains a nonseparating 2-sided simple curve in N. In the first case we can construct double coverings of S and in the second case we can construct infinite cyclic coverings of S. For instance, in the second case, if we N cut along the nonseparating simple curve we obtain a surfaces N' with two boundary components a and b.



Consider infinitely many copies N'_r of N' indexed by $r \in \mathbb{Z}$ and glue them together, by identifying the boundary circle a_r of N'_r with the boundary circle b_{r+1} of N'_{r+1}. We obtain a surface W which covers N by sending each copy N'_r onto N and whose deck transformation group is $\mathbb{Z}$. In both situations we contradict the simple connectivity of S.

A last ingredient is the following relation between orientation and homotopy:

Lemma 22 (Orientation and homotopy). *Suppose that there exists a homotopy in S between the oriented not nullhomotopic curves α and α^{-1}. Then S is either a projective plane or the Klein bottle.*

Proof. Pick up a base point on the curve α and denote by the same letters the homotopy classes of the curves α and α^{-1}. There exists a homotopy between two curves if and only if their homotopy classes are conjugate. There exists then some $\gamma \in \pi_1(S)$ such that

$$\alpha^{-1} = \gamma\alpha\gamma^{-1}.$$

Let $H \subset \pi_1(S)$ be the (nontrivial) subgroup generated by α and γ and $\hat{S}$ be the covering of S associated to H. Then H is either free or a surface group. In the first case $H = \mathbb{Z}$ or H is free abelian, in both situations the relation above cannot hold. In the second case the abelianization of H has at most two generators and hence H is the fundamental group of a projective plane, a Klein bottle or a torus. Now a projective plane can only cover trivially a projective plane so S is a projective plane itself. The relation above cannot be satisfied in the fundemantal group of the torus which is abelian. It remains the case when $H = \pi_1(N_2)$. Since the Euler characteristic χ behaves multiplicatively under unramified coverings, it follows that $\chi(S) = 0$ and S is nonorientable. Therefore S is a Klein bottle.

Proof (Proof of Baer's Theorem 10). Let α and β be the two homotopic simple closed curves $f_0(S^1)$ and $f_1(S^1)$ on S. It suffices to show that α and β are isotopic as unoriented curves, as Lemma 22 forces the isotopy to send orientations accordingly, because S was supposed orientable. If they are not disjoint, then Lemma 20 provides an embedded disk whose boundary consists of an arc of α and an arc of β. By choosing an innermost such disk we can insure that $\text{int}(D^2)$ is disjoint from α and β. There is then an obvious ambient isotopy supported on D^2 which allows the two boundary arcs be separated. Repeatedly use of this procedure makes the curves α and β disjoint.

Suppose first that α is separating. As we supposed S orientable, the simple closed curves are 2-sided. There is a natural map $\pi : S \to S'$, where S' is obtained from S by crushing to a point the closure of the connected component of $S - \alpha$ which does not contain β. As β was homotopic to α, the image $\pi(\beta)$ of β is homotopic in S' to the point $\pi(\alpha)$. Then Lemma 21 says that $\pi(\beta)$ bounds a 2-disk D^2 embedded in S'.

If the disk D^2 does not contain the point $\pi(\alpha)$, then its preimage $\pi^{-1}(D^2)$ would be a disk and hence β would bound a disk in S, which is a contradiction.

Otherwise, D^2 contains $\pi(\alpha)$ and so $\pi^{-1}(D^2)$ is a cylinder embedded in S, whose boundary consists in the disjoint union of the curves α and β. This cylinder provides an ambient isotopy between the two curves.

It remained to settle the case when α is nonseparating. Define S' to be the surface obtained from S by first removing a regular neighborhood $\alpha \times (-1, 1)$ of α and then pinching each one of the two boundary circles $\alpha \times \{\pm 1\}$ of $S - \alpha \times (-1, 1)$ to a point $p_\pm$. The quotient map $S - \alpha \times (-1, 1) \to S'$ extends to a continuous map $\pi : S \to S'$ by setting

$$\pi(x, t) = \gamma(t), \text{ for } (x, t) \in \alpha \times [-1, 1]$$

where $\gamma : [-1, 1] \to S'$ is a curve joining p_+ to p_-. As β was homotopic to α in S, the image $\pi(\beta)$ is homotopic to a point in S'. By Lemma 21 again $\pi(\beta)$ bounds a 2-disk D^2 embedded in S'.

If the disk D^2 does not contain neither p_+ nor p_-, then its preimage $\pi^{-1}(D^2)$ would be a disk and hence β would bound a disk in S, which is a contradiction.

If the disk D^2 contains one of these two points, say p_+, then $\pi^{-1}(D^2)$ would be a cylinder in S with boundary $\alpha \sqcup \beta$, thereby providing an ambient isotopy between these curves.

Eventually, assume that D^2 contains both p_+ and p_-. Then $\pi(\beta)$ will separate S' and hence β will be separating in S. Consider the surface S'' obtained from S by crushing the connected component of the closure of $S - \beta$ which does not contain α to one point and $\pi' : S \to S''$ be the quotient map. Then $\pi'(\alpha)$ is a meridian of the torus S'', while $\pi'(\beta)$ is a point. On the other hand, $\pi'(\alpha)$ should be homotopic to $\pi'(\beta)$. This contradiction settles our claim.

Remark 6. Although stated for orientable surfaces, the proof given above shows that Theorem 10 is valid for all compact surfaces, if we ask $f_1(S^1)$ be 2-sided. Now, using the fact that homeomorphisms of the Moebius band rel boundary are isotopic to the identity, one can show that the statement of Theorem 10 is equally valid when $f_1(S^1)$ is 1-sided without any restrictions.

A map $f : M \to N$ between manifolds *with boundary* will be called *proper* if $f^{-1}(\partial N) = \partial M$. In this context homotopies or isotopies which are the identity on the boundary are also called homotopies or isotopies *rel boundary.* Now Baer's result extends without essential changes to properly embedded arcs, as follows:

Theorem 11 (Homotopic arcs are isotopic). *Let $f_0, f_1 : [0,1] \to S$ be proper embeddings of arcs in a compact orientable surface S. Assume that f_0 is homotopic to f_1 by a homotopy keeping the endpoints fixed. Then f_0 and f_1 are ambient isotopic rel boundary.*

Proof. It suffices to show that whenever we have two properly embedded arcs α and β which are homotopic keeping their endpoints fixed there exists an embedded disk whose boundary consists of an arc of α and an arc of β. For this purpose we consider the universal covering $\widetilde{S} \to S$ and observe that every lift of α or β should separate $\widetilde{S}$. The proof Lemma 20 goes unchanged in this setting.

The case of curves with base points is only slightly more difficult. It is a theorem of Epstein:

Theorem 12 (Epstein 1966). *Let $f_0, f_1 : (S^1, 0) \to (S, p)$ be embeddings of the circle in the interior* $\mathrm{int}(S)$ *of a surface, such that $f_0(0) = f_1(0) = p$ and f_0 is homotopic to f_1 by a homotopy keeping the base point* 0 *fixed. Assume that $f_1(S^1)$ does not bound a disk D^2 or a Moebius band B in S. Then there exist an ambient isotopy f_0 and f_1 which keeps the base point p fixed.*

Exercise 10 (Epstein 1966). Prove that if f_0 and f_1 are circle embeddings in a surface S and $f_0(S^1)$ does not bound an embedded 2-disk if it is 2-sided, then f_0 and f_1 are ambient isotopic.

3.3 Homotopy and homeomorphisms

The main result of this section is the following theorem due to Baer and Epstein:

Theorem 13. *Let h be a homeomorphism of the compact surface $(S, \partial S)$ which is homotopic to the identity. If S is a disk D^2 or a cylinder $S^1 \times [0,1]$, assume that f preserves the orientation. Then h is isotopic to the identity. Moreover, if the homotopy preserved basepoints, then we can allow the isotopy to preserve the basepoints, except in the case of the Moebius band when we should require that the local orientation be preserved at the base point.*

Note that the result is valid equally for picewise linear homeomorphisms and diffeomorphisms, although the proof in the last case should be slightly amended. In fact one key ingredient is the following lemma due to Alexander, whose analog in the smooth case fails in higher dimensions:

Lemma 23 (Alexander's Lemma 1923). *Let $h : (D^n, p) \to (D^n, p)$ be a homeomorphism of the n-ball D^n which restricts to the identity on the boundary sphere ∂D^n. Then there exists an isotopy keeping the point p fixed between h and the identity which is the identity on the boundary ∂D^n.*

Proof. We can assume that D^n is the unit ball and p is its center. Define then the required isotopy $J : D^n \times [0,1] \to D^n$, by:

$$J(x,t) = \begin{cases} th(x/t), & \text{if } 0 \leq \| x \| < t; \\ x, & \text{if } t \leq \| x \| \leq 1. \end{cases}$$

Remark 7. Another version of Alexander's lemma states that every homeomorphism of the sphere extends to a homeomorphism of the ball. Its construction consists in a radial extension through coning and it works as well for piecewise linear maps. The extension of Alexander's Lemma to the smooth category fails in dimensions higher than 6. The smooth isotopy classes of orientation preserving diffeomorphisms of S^{n-1} form an abelian group Γ_n and Milnor proved that $\Gamma_7 = \mathbb{Z}/28\mathbb{Z}$. The nontriviality of Γ_n is at the origin of the existence of exotic spheres, namely smooth manifolds homeomorphic but not diffeomorphic to S^n. In fact Smale has proved that any smooth homotopy sphere of dimension $k \neq 4$ is obtained by means of a connected sum construction, by gluing together two k-disks with opposite orientations by identifying their boundary spheres by means of a diffeomorphism. The homotopy sphere so obtained is diffeomorphic to the standard sphere if and only if the gluing diffeomorphism is smoothly isotopic to the identity.

Proof. When S is closed we consider a family of simple closed curves $\{\alpha_1, \ldots, \alpha_g\}$ such that cutting S along these curves we obtain a disk. The hypothesis tells us that $h(\alpha_i)$ is homotopic to α_i. Baer's Theorem 10 shows that, up to composing h by an ambient isotopy H_1 we can arrange for $h(\alpha_1) = \alpha_1$. There is no loss of generality in assuming that $h(\alpha_i)$ and α_i intersect α_1 in the same point and that the homotopy fixes these intersection points. Cut then S along α_1 and obtain a surface S' with two boundary components. Then there is a homotopy fixing the endpoints between the arcs which are traces of the curves α_i and $h(\alpha_i)$ on the surface S', viewed as a subsurface of S. Theorem 11 shows that, up to composing further the restriction of $h|_{S'}$ by an ambient isotopy of S' rel boundary, we can arrange for the arcs corresponding to $h(\alpha_2)$ and α_2 to coincide. Since ambient isotopies rela boundary lift obviously to S, this amounts to say that $h(\alpha_2) = \alpha_2$, up to composing h with an ambienti isotopy of S. Iterated use of this argument provides an isotopy H between α_i and $h(\alpha_i)$, for all i. It follows that $H \circ h$ induces a homeomorphism of the disk D^2 obtained after cutting along all curves α_i. Now Alexander's Lemma 23 tells us that $H \circ h$ is isotopic to the identity, and the claim follows.

Exercise 11. Prove that a homeomorphism of the Moebius band which is the identity on the boundary is isotopic to the identity by an isotopy which is the identity on the boundary.

Exercise 12. Prove that orientation preserving diffeomorphisms of S^1 are smoothly isotopic to the identity.

References

1. J. W. Alexander, *On the deformation of an n-cell*, Proc. Natl. Acad. Sci. USA 9 (1923), 406–407.
2. Reinhold Baer, *Isotopie von Kurven auf orientierbaren, geschlossenen Flächen und ihr Zusamenhang mit der topologischen Deformation der Flächen*, J. Reine Angew. Math. 159 (1928), 101–116.
3. D.B.A. Epstein, *Curves on 2-manifolds and isotopies*, Acta Math. 115 (1966), 83–107.

4

Maps between surfaces

4.1 The degree

Recall that an *orientation* of a 2-simplex is a cyclic order of its vertices. Thus there are two possible orientations on a 2-simplex. The boundary of an oriented 2-simplex inherits then an orientation of its edges. An *orientation* on a triangulated surface F consists of an orientation on each 2-simplex such that each oriented edge is in the boundary of a unique positive 2-simplex.

Definition 7. *Let $\varphi : F_1 \to F_2$ be a simplicial map between two oriented connected surfaces and $\sigma \subset F_2$ be an open 2-simplex. Let $\varphi^{-1}(\sigma)$ consist of the disjoint union of 2-simplices σ_i and several 0 and 1-simplices. A simplex σ_i is called* `positive` *if its orientation agrees with that of F_1 and* `negative`*, otherwise. Then* `the degree` *of φ is the difference between the number of positive and negative 2-simplices σ_i.*

Lemma 24. *The degree is well-defined and independent on the choice of the 2-simplex σ.*

Proof. Let σ and σ' be two adjacent 2-simplices whose closures contain the edges e, v the vertex of σ opposite to a and v' the vertex of σ' opposite to e. Suppose that $f^{-1}(\sigma) = \sqcup \sigma_i$ and let $e_i \subset \sigma_i$ be an edge of σ_i with $\varphi(e_i) = e$ and v_i the vertex of σ_i opposite to e_i. Denote also by σ_i' the simplex of F_1 which shares the edge e_i with σ_i and v_i' the vertex of σ_i' opposite to e_i.

1. If $\varphi(v_i') = \varphi(v_i) = v$, then σ_i and σ_i' both belong to $f^{-1}(\sigma)$ and they have opposite orientations, so that they do not contribute to the degree of φ.
2. If $\varphi(v_i') = v'$, then $\sigma_i \subseteq f^{-1}(\sigma)$ and $\sigma_i' \subseteq f^{-1}(\sigma')$ have the same contribution to the degree of φ computed using σ and σ', respectively.
3. if $\varphi(v_i') \in a$, then $\varphi(v_i')$ is also the image of vertex of e_i. Consider the edges incident to the other vertex of e_i which inherit a cyclic order. Let e_i' be the edge encountered just before the first edge which does not map under φ into e. Define the simplex σ_i'' to be the 2-simplex sharing the edge e_i' and

the next one. Then the associated simplices σ_i and σ_i'' are in one of the two cases above, because all edges between e_i and e_i' map by φ into e.

In all cases above the value of the degree does not depend on which of σ and σ' is considered, and hence proving the claim.

Remark 8. If $f : M \to N$ is a degree d covering then its degree is $\pm d$ depending on whether f preserves or not the orientations.

Although this definition only works for simplicial maps, it can be extended to arbitrary maps, if we use the following approximation theorem:

Lemma 25 (Simplicial Approximation lemma). *Given a continuous map f between spaces underlying two simplicial complexes, there exist a simplicial map F between sufficiently fine subdivisions of them such that $F(x)$ belongs to the smallest simplex in the target containing $f(x)$, for any x. In particular, F is homotopic to f.*

Proposition 3. *Let $f : S' \to S$ be a map between closed orientable surfaces and $f_* : H_2(S') \to H_2(S)$ be the induced map in homology. Then f_* is the multiplication by the degree of f, once we identify $H_2(S)$ with $\mathbb{Z}$ by using the fundamental class. In particular, the degree is a homotopy invariant.*

Proof. First, the degree does not change if we subdivide the triangulations, so we can assume that f is simplicial. A 2-cycle which realizes the fundamental class $[S']$ is given by the union of all oriented 2-simplexes of the triangulation. Then $f_*([S'])$ is the degree of f times the sum of all oriented 2-simplices of S. Its homological interpretation shows that f is a homotopy invariant.

Remark 9. More generally the map $f_* : H_n(S', \partial S') \to H_n(S, \partial S)$ is the multiplication by the suitably defined degree of a proper map $f : (S' \to S$ between compact oriented manifolds. In the nonorientable case the degree mod 2 is well-defined.

4.2 Ramified coverings

We can extend the notion of orientation from an open simplex to any point of a surface by choosing one of the two cyclic orders on a boundary circle of a small disk neighborhood, called a *local orientation.* The surface is orientable if we can cover it by open disks having compatible local orientations.

Note that a local chart on a surface endowed with local *complex* coordinates has a natural local orientation, by using the cyclic order for which the argument of the coordinate is increasing.

Definition 8. *A map $\varphi : F' \to F$ between oriented surfaces is a* `ramified covering` *if every point $p \in F'$ has an open neighborhood homeomorphic to the* *unit disk U such that the restriction $\varphi : U \to f(U)$ is given in convenient local complex coordinates compatible with the orientations by*

$$\varphi(z) = z^{r_p}$$

for some $r_p \geq 1$. When $r_p > 1$ the point p is called a ramification point of φ and r_p is called the `ramification index` *of p. The image of a ramification point is called a* `branch point`*.*

There are only finitely many ramification points on a compact surface, since they cannot accumulate. If there are no ramification points, then φ is a covering. In every case, the degree of a ramified covering between two oriented surfaces is always positive, since the local orientations do agree.

A basic formula relates the topology of the surfaces and the ramification indices:

Theorem 14 (Riemann-Hurwitz formula). *If φ is a ramified covering of degree d, then*

$$\chi(F') = d\chi(F) - \sum_{p \in F'} (r_p - 1)$$

Proof. If we delete small enough disjoint open disk neighborhoods of the ramification points we obtain a covering map $\varphi : F' - \sqcup_{i=1}^r D_i^2 \to F - \sqcup_{j=1}^b D_j^2$, where r is the total number of ramification points and b the number of branch points. By Lemma 4 we have:

$$\chi(F' - \sqcup_{i=1}^r D_i^2) = d\chi(F - \sqcup_{j=1}^b D_j^2)$$

As $\chi(D^2) = 1$ and χ is additive under connected sums this amounts to:

$$\chi(F') - r = d(\chi(F) - b)$$

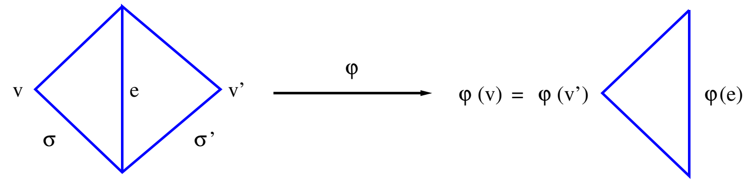
On the other hand, the sum of indices of those ramification points which are mapped on a given branch point is the degree d so that:

$$\sum_{p \in F'} (r_p - 1) = db - r.$$

Let $\varphi : F' \to F$ be a simplicial map of surfaces. We say that φ has a *fold* at the edge e of F' if the triangles σ and σ' sharing the edge e have the same image:

$$\varphi(\sigma) = \varphi(\sigma'),$$

as in the picture below:



Recall that a simplicial map is called *nondegenerate* if it preserves the dimension of the simplices. Recall that the *closed star* of a vertex is the union of all closed cells containing it. Moreover the *link* of a vertex is the union of all cells contained in the star which are disjoint from that vertex. By subdividing each cell we can always assume that vertex stars are homeomorphic to a disk and hence links are homeomorphic to a circle.

Proposition 4. *A simplicial map between oriented triangulated surfaces is a ramified covering after possibly changing the orientation if and only if it is nondegenerate and has no folds.*

Proof. A ramified covering which is also simplicial should be nondegenerate, as it is a local homeomorphism outside the finite set of ramification points. Also the image of a fold is locally modeled on a half-disk and not on a disk, so ramified coverings cannot have folds.

Conversely, a nondegenerate simplicial map is a covering over each open 2-simplex of the target. If there are no folds, then the image of a disk neighborhood of an edge point is the the union of two half-disks having only their diameter in common and hence a disk. Therefore such a map is a local homeomorphism everywhere but possibly at the vertices. Now, triangles containing a vertex v are sent into triangles containing the same vertex. If we travel around the link of the vertex v following the local orientation, then then the image vertices form a cycle in the link of the image vertex. This cycle cannot go back and forth because there are no folds. Thus the map restricts to a finite cover between the links. If the tour direction is contrary to the local orientation, then we change the orientation of one of the two surfaces to remedy. This will work for all vertices, as their local orientations are compatible. The conical extension of a finite cover between circles is given by $z \to z^r$ in local coordinates, if the orientations agree and thus the map is a ramified covering, as claimed.

References

1. A. Hurwitz, *Über Riemann'sche Flächen mit gegebenen Verzweigungspunkten*, Math. Annalen 39 (1891), 1–60.

4.3 Kneser's theorem

The main results of this section is the following inequality due to Kneser:

Theorem 15 (Kneser 1930). *Let $\varphi : F' \to F$ be a degree $d \geq 0$ simplicial map between closed connected oriented surfaces of genera g' and $g \geq 1$. Then the following inequality holds:*

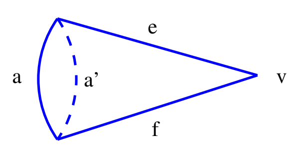
$$\chi(F') \leq d\chi(F).$$

Moreover, if equality holds above and $d \neq 0$, then φ is homotopic to a covering of degree d. Furthermore if $d = 0$, then φ is homotopic to a map sending F' into the 1-skeleton of F.

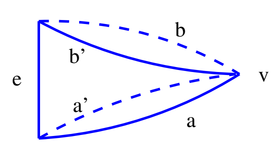
Proof. Before to proceed note that degree of a map $S^2 \to F$ should vanish. Indeed such a map lifts to $S^2 \to \widetilde{F}$ since S^2 is simply connected. The lift is homotopically trivial since $\widetilde{F}$ is homeomorphic to the plane and hence contractible. This implies that the original map $S^2 \to F$ is also nullhomotopic. As the degree is invariant by homotopy, it must be zero.

We can therefore consider only the case when $g' \geq 1$ below. Assume from now on that the triangulation is *clean*, namely edges have different endpoints and any edge appears only once in the boundary of a triangle.

1. If $\varphi : F' \to F$ is a covering, then we already saw in Lemma 4 that
$$\chi(F') = d\chi(F).$$
2. If $\varphi : F' \to F$ is a ramified covering then the Riemann-Hurwitz formula gives us:
$$\chi(F') \leq d\chi(F)$$
and the inequality above is satisfied.
3. Suppose now that φ is nondegenerate. If φ is not a ramified covering, then it has some folds.
We will change the pair (F', φ) so that the degree of the map remains unchanged, $\chi(F')$ increases while the number of 2-simplices in the triangulation diminishes as well as the number of folds.
   a) If $\sigma \cap \sigma' = e$, we cut F' along the boundary of $\sigma \cup \sigma'$ and then identify in the boundary $\partial\sigma - e$ with $\partial\sigma' - e$. Denote by F'' the resulted surface and now observe that φ induces a simplicial nondegenerate map $\varphi' : F'' \to F$ of the same degree as φ.
   b) if $\sigma \cap \sigma'$ consists of two edges e and f having a common vertex v.



   We cut F' along the cone $\sigma \cup \sigma'$ and then identify its two boundary edges denote a and a' in the picture above. Then the surface F'' so obtained has induced a nondegenerate simplicial map $\varphi' : F'' \to F$.
   c) If $\sigma \cap \sigma'$ consists of the union of an edge e and the opposite vertex v in each triangle:



   Cut then F' along a, which is one of the two closed curves formed by two edges having the same endpoints in $\sigma \cap \sigma'$ and further identify the edges corresponding to a and a' in the two boundary components.

If the curve a separates F' and the resulting components are both spheres, then F' would be homeomorphic to a sphere, which is impossible. If only one of the resulting component is a sphere, we define F'' to be the other connected component and then $\chi(F'') = \chi(F')$. If the curve is nonseparating then F' is the resulting surface and we have:

$$\chi(F'') = \chi(F') + 2.$$

In all cases φ induces a simplicial map of the same degree $\varphi' : F'' \to F$.

d) If $\sigma \cap \sigma'$ consists of all three edges of the triangles, then it forms a sphere, which is impossible.

4. Suppose that φ is degenerate, so that one of the following holds:
   a) There exists an edge e of F' such that $\varphi(e)$ is a vertex of F. Let σ and σ' be the two 2-simplices sharing e. We now modify the pair (F', φ) such that the degree and $\chi(F')$ is preserved while the number of faces and the number of collapsed edges e strictly decrease.
      i. If $\sigma \cap \sigma'$ consists of e alone we proceed as in 3a: one collapses e to a vertex by removing $\sigma \cup \sigma'$ and identifying $\partial\sigma - e$ with $\partial\sigma' - e$. Then the surface F'' is endowed with a simplicial map $\varphi' : F'' \to F$.
      ii. If $\sigma \cap \sigma'$ consists of two edges e and f two edges as in 3b, we must have $\varphi(a) = \varphi(a') = \varphi(f)$, because φ is simplicial. Then φ factors through the surface F'' defined at 3b.
      iii. If $\sigma \cap \sigma'$ consists of an edge e and the opposite vertex v, as in 3c, then we define the surface F'' obtained by collapsing the edge e to a vertex and identifying a with b and a' with b' respectively.
      iv. If $\sigma \cap \sigma'$ consists of three edges then the connected component of F' containing them is a sphere and we define F'' as F' deprived of this sphere, as in 3d.
   b) If there exists a 2-simplex σ such that $\varphi(\sigma)$ is a vertex v, then the edges of σ are collapsed by φ to a vertex and we can use 4a above.
   c) If there exists a 2-simplex σ such that $\varphi(\sigma)$ is an edge a, then at least one edge of σ is sent into a vertex and we use again 4a above.

If φ is degenerate, then we change it by the procedure given at point 4. until φ becomes a nondegenerate simplicial map. We further use the transformations from point 3 above to replace (F', φ) by some (F'', φ') such that φ' has no folds. These transformations only could increase $\chi(F')$. Since φ is a nondegenerate simplicial map without folds and hence a ramified covering, the Riemann-Hurwitz inequality gives us:

$$\chi(F') \leq \chi(F'') \leq d\chi(F)$$

This proves the inequality we wanted.

If equality holds above and $g \geq 1$, then one cannot use steps which strictly increase $\chi(F)$. Note now that in steps 3a and 3b and 3c when the homeomorphism type of F' is not affected the maps φ and φ' can be realized by a homotopy on the surface F'. This implies that φ is homotopic to a ramified covering and by the Riemann-Hurwitz formula this should be unramified.

Remark 10. When $g = 0$ and $\chi(F') = d\chi(F)$, then $d = \pm 1$ and $g' = 0$. Then Hopf's theorem shows that a degree ± 1 map of S^2 should be homotopic to a homeomorphism.

References

1. H. Kneser, *Glättung von der Flächenabbildungen*, Math. Annalen 100 (1928), 609–617.
2. H. Kneser, *Die kleinste Bedeckungszahl innerhalb enier Klasse von Flächenabbildungen*, Math. Annalen 103(1930), 347–358.

4.4 Edmonds' theorem

We say that a PL map $f : S \to \Sigma$ between surfaces is a *pinch* map if there exists a subsurface $S' \subset S$ with one boundary component such that Σ is homeomorphic to the quotient S/S' (namely the result of crushing S' to one point) while f is now identified with the quotient map $S \to S/S'$. Equivalently, there exists a 2-disk $D \subset \Sigma$ such that the restriction of f to $f^{-1}(\Sigma - D)$ is one-to-one.

Theorem 16 (Edmonds 1979). *Every PL map between orientable surfaces is homotopic to the composition of a pinch map and a ramified covering.*

Moreover the map $f : S \to \Sigma$ is homotopic to a branched covering if and only if either the map induced in homotopy $f_* : \pi_1(M) \to \pi_1(\Sigma)$ is injective (in which case it is an unramified covering) or else $|deg(f)| > \left|\frac{\pi_1(\Sigma)}{f_*(\pi_1(S))}\right|$, where $deg(f)$ denotes the degree of f. Notice that for any continuous map $f : S \to \Sigma$ of non-zero degree we have $|deg(f)| \geq \left|\frac{\pi_1(\Sigma)}{f_*(\pi_1(S))}\right|$. In particular, f is not homotopic to a ramified covering if and only if

$$|deg(f)| = \left|\frac{\pi_1(\Sigma)}{f_*(\pi_1(S))}\right|, \text{ and } \ker f_* \neq 1$$

References

1. A.L. Edmonds, *Deformation of maps to branched coverings in dimension two*, Ann. Math. 110 (1979), 113–125.
2. H. Kneser, *Glättung von der Flächenabbildungen*, Math. Annalen 100 (1928), 609–617.
3. H. Kneser, *Die kleinste Bedeckungszahl innerhalb enier Klasse von Flächenabbildungen*, Math. Annalen 103(1930), 347–358.

4.5 Gabai-Kazez classification

5

Mapping class groups

5.1 From homeomorphisms to mapping class groups

Definition 9. *The* `mapping class group` $\mathrm{Mod}(M)$ *of a manifold* M *is the quotient of the groups of homeomorphisms* $\mathrm{Homeo}(M)$ *which fix pointwise the boundary* ∂M *by the homotopy equivalence relation rel boundary. When thye manifold is oriented we usually restrict to the quotient of the group* $\mathrm{Homeo}^+(M)$ *of orientation preserving homeomorphisms.*

Remark 11. When S is an (oriented) surface, then Theorem 13 shows that $\mathrm{Mod}(S)$ is the quotient of $\mathrm{Homeo}^+(S)$ by the subgroup $\mathrm{Homeo}_0(S)$ of homeomorphisms isotopic to identity. Thus $\mathrm{Mod}(S)$ is the group of connected components of the topological group $\mathrm{Homeo}^+(S)$.

Remark 12. If we considered instead piecewise linear homeomorphisms or diffeomorphisms, the resulting group would be still isomorphic to $\mathrm{Mod}(S)$. In fact any homeomorphism of a compact surface is isotopic to a diffeomorphism and isotopic diffeomorphisms which are smoothly isotopic.

Some caution is needed when considering surfaces with punctures: if we have a compact surface S and a finite set of points $P \subset S$ then we can either consider points of P as marked points on S or else as punctures of the open surface $S - P$. There are then two possible definitions for the mapping class group of the punctured surface, either as the quotient $\mathrm{Mod}(S, P)$ of the group of homeomorphisms of the pair (S, P) up to homotopy equivalence fixing pointwise P, or else as the mapping class group $\mathrm{Mod}(S - P)$ of $S - P$. Fortunately, the two definitions agree and we will use both instances in the sequel. Notice also that elements of $\mathrm{Mod}(S, P)$ have a well-defined action by permutations on the set P. Those mapping classes which fix the punctures form the *pure* mapping class group $\mathrm{PMod}(S, P)$, which therefore fits into the exact sequence:

$$1 \to \mathrm{PMod}(S, P) \to \mathrm{Mod}(S, P) \to \mathcal{S}_n \to 1$$

where $\mathcal{S}_n$ denotes the symmetric group on $n = |P|$ elements.

We denote $\mathrm{Mod}_{g,b}^k = \mathrm{Mod}(S_{g,b}^k)$ and $\mathrm{PMod}_{g,b}^k = \mathrm{PMod}(S_{g,b}^k)$ and drop the subscript b or the superscript k, when zero. For a few surfaces it is easy to compute directly $\mathrm{Mod}(S)$. Alexander's Lemma 23 tells us that the mapping class groups of the disk D^2 and of the once punctured disk (D^2, P) are trivial:

$$\mathrm{Mod}_{0,1} = \mathrm{Mod}_{0,1}^1 = 1$$

Every orientation preserving homeomorphism of (S^2, P) is homotopic to the identity, since $S^2 - P = \mathbb{R}^2$ is contractible, and hence isotopic to the identity by the Baer-Epstein Theorem 13. Further any homomeomorphism of S^2 can be isotoped to a homeomorphism fixing a point. The mapping class groups of the sphere and the once punctured sphere are also trivial:

$$\mathrm{Mod}_0 = \mathrm{Mod}_0^1 = 1$$

Consider a thrice punctured sphere S_0^3. The induced action of Mod_0^3 on the set of punctures induces a surjective homomorphism $\mathrm{Mod}_0^3 \to \mathcal{S}_3$ onto the group of permutation on three elements. Any two arcs joining two punctures are homotopic rel boundary since the complement of the third puncture is $\mathbb{R}^2$. Thus any mapping class fixing the punctures also fixes the homotopy class of such an arc. Now, cutting along this arc will result in once punctured disk. By Alexander's Lemma 23 we derive that

$$\mathrm{Mod}_0^3 = \mathcal{S}_3$$

Proposition 5. *The mapping class group of the torus* Mod_1 *and of the punctured torus* Mod_1^1 *are both isomorphic to* $SL(2, \mathbb{Z})$.

Proof. Choose a longitude ℓ and a meridian m on the torus intersecting once and providing a basis in homology. Observe that the algebraic intersection of two curves is the area of the parallelogram determined by the two corresponding vectors in $H_1(S; \mathbb{R})$. Thus mapping classes act in homology by matrices from $SL(2, \mathbb{Z})$. Conversely, any integral matrix of determinant one provides a linear map $\mathbb{R}^2 \to \mathbb{R}^2$ which passes to the quotient by $\mathbb{Z}^2$ to an orientation preserving homeomorphism of the torus $\mathbb{R}^2/\mathbb{Z}^2$; the induced action on homology – identified with the lattice $\mathbb{Z}^2$ – is the matrix with which we started.

Eventually, if h is a homeomorphism of the torus $\mathbb{R}^2/\mathbb{Z}^2$ whose action in homology is trivial, then ℓ, m are homotopic to $h(\ell), h(m)$ fixing the base point, respectively since $\pi_1(S_1) \simeq H_1(S_1)$. By Baer-Epstein's Theorem there exists an isotopy fixing the basepoints between them. Then Alexander's Lemma 23 shows that h is isotopic to the identity. The proof works as well for the punctured torus.

The homotopy type of the connected components of homeomorphism groups of surfaces was described in a series of papers by Hamstrom, which culminated with the following description:

Theorem 17 (Hamstrom [2]). *The group* $\mathrm{Homeo}_0(S)$ *is contractible for any compact surface* S *possibly with punctures, except for the case of:*

1. *the closed surfaces* $S = S^2, S^1 \times S^1$*;*
2. *the compact surfaces with boundary* D^2 *and* $S^1 \times [0,1]$*;*
3. *the once punctured* S^2*, the once punctured* D^2 *and the twice punctured* S^2*.*

The description of the homotopy type of the connected components of diffeomorphism groups of surfaces needed additional tools which were developped independently by several authors:

Theorem 18 (Smale ([7]), Earle-Eells ([2]), Earle-Schatz ([3]), Gramain ([4])). *Let* S *be a compact surface without punctures.*

1. *if* $S = S^2$ *or* RP^2*, then* $\mathrm{Diff}_0(S)$ *is homotopy equivalent to* $SO(3)$*;*
2. *if* $S = S^1 \times S^1$*, then* $\mathrm{Diff}_0(S)$ *is homotopy equivalent to* $S^1 \times S^1$*;*
3. *if* S *is a Klein bottle* N_2 *or the Moebius band* B*, then* $\mathrm{Diff}_0(S)$ *is homotopy equivalent to* S^1*;*
4. *if* S *is a cylinder* $S^1 \times [0,1]$ *or the disk* D^2*, then* $\mathrm{Diff}_0(S)$ *is homotopy equivalent to* $O(2)$*;*
5. *Otherwise* $\mathrm{Diff}_0(S)$ *is contractible.*

Note that work of Anderson, Edwards and Kirby show that

Theorem 19. *If* M *is a connected compact manifold, then the group* $\mathrm{Homeo}_0(M)$ *is algebraically simple.*

Problem 1. It is still unknown whether $\mathrm{Diff}(D^4, \partial D^4)$ is connected. Recent results of Watanabe (see [8]) show that $\mathrm{Diff}(D^4, \partial D^4)$ is not contractible. In the topological case this has been proved by Randall and Schweitzer in [6].

References

1. R.D. Anderson, *The algebraic simplicity of certain groups of homeomorphisms*, Amer. J. Math. 80 (1958), 955–963.
2. C.J.Earle and J. Eells, *A fiber bundle description of Teichmüller space*, J. Diff. Geometry 3 (1969), 19–43.
3. C.J.Earle and A.Schatz, *Teichmüller theory for surfaces with boundary*, J. Diff. Geometry 4 (1970), 169–185.
4. A. Gramain, *Le type d'homotopie du groupe des difféomorphismes d'une surface compacte*, Ann. Sci. E.N.S. 4-ème sér. 6(1973), 53–66.
5. M.E. Hamstrom, *Homotopy groups of the space of homeomorphisms on a 2-manifold*, Illinois J. Math. 10 (1966), 563–573.
6. D.Randall and P.A.Schweitzer, *On foliations, concordance spaces and the Smale conjectures*, in "Differential Topology, Foliations and Group Actions", Contemp. Math. 161, 1994, 235–258.
7. S. Smale, *Diffeomorphisms of the 2-sphere*, Proc. Amer. Math. Soc. 10 (1959), 621–626.
8. Tadayuki Watanabe, *Some exotic nontrivial elements of the rational homotopy groups of* $Diff(S^4)$, arxiv 1812.02448.

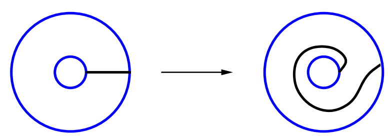
5.2 Dehn twists

In this section all considered surfaces will be orientable and we recall that $\mathrm{Mod}(S)$ is the set of isotopy classes of orientation preserving homeomorphisms of S.

Definition 10. *We identify by an orientation preserving homoeomorphim a regular neighborhood N of the simple closed curve γ on the oriented surface S with the cylinder $S^1 \times [0,1]$. The class $T_\gamma \in \mathrm{Mod}(S)$ of the homeomorphism h_γ of S acting as*

$$h_\gamma(\theta, t) = (\theta + 2\pi t, t)$$

on N and the identity elsewhere is called the `left Dehn twist around` *γ. The action of the homeomorphism h_γ on the arc $\{0\} \times [0,1]$ is described in the picture below:*



Remark 13. The (left) Dehn twist T_γ is welldefined and independent on the choice of identification of N with the cylinder or the orientation of γ. The right Dehn twist in the inverse T_γ^{-1} of the left Dehn twist.

One should note that as $\mathrm{Homeo}(S)$ has a natural action on the curves on the surface, $\mathrm{Mod}(S)$ has an induced action on the set of *isotopy classes* of simple curves and simple properly embedded arcs drawn on S which is fundamental in its study. By language abuse we will denote often by the same letter a homeomorphism and its class in $\mathrm{Mod}(S)$.

Although they follow directly from its definition, the following properties will be frequently used in the sequel:

Lemma 26 (Conjugate Dehn twists are Dehn twists). *If γ is isotopic to γ', then $T_\gamma = T_{\gamma'}$. Further, if $\varphi \in \mathrm{Mod}(S)$, then*

$$T_{\varphi(\gamma)} = \varphi T_\gamma \varphi^{-1}$$

where $\varphi(\gamma)$.

Proof. For the first statement we observe that h_γ is isotopic to $h_{\gamma'}$. Next observe that the image of a regular neighborhood of γ by a homeomorphism in the class φ is a regular neighborhood of $\varphi(\gamma)$ and the conjugate of h_γ acts as required on it.

Lemma 27. *The mapping class group of a cylinder $\mathrm{Mod}(S_{0,2})$ is $\mathbb{Z}$, generated by the Dehn twist along the core.*

Proof. Let α be the arc $\{0\} \times [0,1]$ and γ be the core $S^1 \times \{1/2\}$ of the cylinder $S^1 \times [0,1]$. Choose a homeomorphism whose class is some given $\varphi \in \mathrm{Mod}(S_{0,1})$. Lift it to the homeomorphism of the universal covering $\Phi : \mathbb{R} \times [0,1] \to \mathbb{R} \times [0,1]$ which fixes the origin 0 of $\mathbb{R} \times \{0\}$. Since Φ commutes with the unit translation of the $\mathbb{R}$-factor, the image of the arc $\tilde{\alpha} = \{0\} \times [0,1]$ by Φ is an arc joining $(0,0)$ and $(n,1)$, for some $n \in \mathbb{Z}$. By Lemma 11 there exists only one such arc, up to isotopy. It follows that $\varphi(\alpha)$ is isotopic to $T_\gamma^{-n}(\alpha)$ on the cylinder. Cutting the cylinder along α obtain a square and thus $T_\gamma^n \varphi$ is the class of a homeomorphism acting as the identity on the boundary of the square. By Alexander's Lemma 23 $\varphi = T_\gamma^n$.

5.3 Properties of Dehn twists

We say that two isotopy classes α and β of simple curves are *disjoint* if they admit disjoint representatives. We write then

$$i(\alpha, \beta) = 0$$

More generally, we define the *(geometric) intersection number* $i(\alpha, \beta)$ as the minimal number of intersection points between representatives of α and β respectively. Eventually, two simple closed curves a and b are in *minimal position* if the number $|a \cap b|$ of intersection points is minimal within their isotopy classes, namely the geometric intersection number.

As every simple curve a has a regular neighborhood homeomorphic to a cylinder it follows that:

$$i(a, a) = 0$$

For any two transverse curves a' and b' within the homotopy classes a and b, respectively we have:

$$|a' \cap b'| = \omega(a, b) \pmod 2$$

Therefore, if a and b are transverse and $|a \cap b| = 1$, then a and b are in minimal position, as they cannot be made disjoint. A more general criterion is the following:

Lemma 28. *Two simple closed curves a and b are in minimal position if and only if they have no bigons.*

Proof. Assume that a and a' are isotopic and $|a \cap b| > |a' \cap b|$. Let $H : S^1 \times [0,1] \to S$ be an isotopy between a and a', which we can assume smooth and transverse to b. Then $H^{-1}(b)$ is a submanifold of the cylinder $S^1 \times [0,1]$, namely the union of proper arcs and circles. By our assumption there is an arc b_0 of $H^{-1}(b)$ having both endpoints on a and hence isotopic on the cylinder to an arc a_0 of a. The union $a_0 \cup b_0$ is then null-homotopic in S and hence it is a bigon, by the Disk Lemma 19. The converse is clear.

Lemma 29. *If a and b are simple closed curves then $i(T_a b, a) = i(a, b)$ and $i(T_a b, b) = i(a, b)^2$.*

Proof. Let N be a regular neighborhood of a and $\ell = b \cap N$, which is a proper arc joining the two boundary components. Assume that a and b are in minimal position. Then $T_a b$ consists of $b \setminus (b \cap N)$ union $i(a, b)$ parallel arcs of the form $T_a \ell$. The intersection of a and $T_a \ell$ consists of $i(a, b)$ points. Assume that two intersection points determine a bigon consisting of an arc α of a and an arc β of $T_a b$. The arc β is therefore isotopic to the arc α by keeping its endpoints fixed. Applying then T_a^{-1} we obtain an arc of b joining the two intersection points which is isotopic to an arc of a and this implies that b was not in minimal position with respect to a. This contradiction shows that $T_a b$ and a are in minimal poistion so their geometric intersection number is $i(a, b)$.

Similarly $T_a b \cap b$ consists of $i(a, b)^2$ intersection points. The same argument shows that they are then in minimal position.

Lemma 30. *A Dehn twist T_a is trivial in* $\mathrm{Mod}(S)$ *if and only if a is either nullhomotopic or homotopic to a puncture in S.*

Proof. If a is nonseparating, there is some simple close curve with $|a \cap b| = 1$. By direct inspection $|T_a(b) \cap b| = 1$ and so $i(T_a(b), b) = 1$. It follows $T_a(b) \neq b$, as $i(b, b) = 0$. If a is separating, there is some simple closed curve with $i(a, b) = 2$. By direct inspection $|T_a(b) \cap b| = 4$. From Lemma 28 $T_a(b)$ and b are in minimal position and hence $T_a(b) \neq b$, as above.

Lemma 31. *Show that $T_a = T_b$ if and only if $a = b$.*

Proof. If $i(a, b) \neq 0$, then $i(T_a(a), a) = i(a, a) = 0$, while $i(T_b(a), a) = i(b, a)^2 \neq 0$. This shows that $T_a \neq T_b$.

If $i(a, b) = 0$, we cut the surface along a. The resulting surface is not homeomorphic to an annulus or a disk and thus there exists at least one curve c which has $i(b, c) \neq 0$. It follows that $i(T_a(c), c) = 0$, while $T_b(c), c) = i(b, c)^2 \neq 0$.

Lemma 32. *The mapping class φ commutes with T_a if and only if $\varphi(a) = a$.*

Proof. Assume that $\varphi T_a \varphi^{-1} = T_a$. The left side is equal to $T_{\varphi(a)}$, and the previous Lemma shows that $\varphi(a) = a$.

5.4 Braid relations between Dehn twists

Lemma 33 (Dehn twists commute if supports are disjoint). *We have $T_\alpha T_\beta = T_\beta T_\alpha$ if and only if $i(\alpha, \beta) = 0$ if and only if $T_\alpha \beta = \beta$.*

Proof. One implication is clear, as the corresponding homeomorphisms can be chosen to have disjoint supports thereby commuting. Further, we have $T_\alpha T_\beta T_\alpha^{-1} = T_\beta$ if and only if $T_\alpha \beta = \beta$, by Lemma 32. Now, when this holds, then $i(T_\alpha \beta, \beta) = i(\beta, \beta) = 0$ on one hand and by lemma 29 $i(T_\alpha \beta, \beta) = i(\alpha, \beta)^2$, which implies that $i(\alpha, \beta) = 0$.

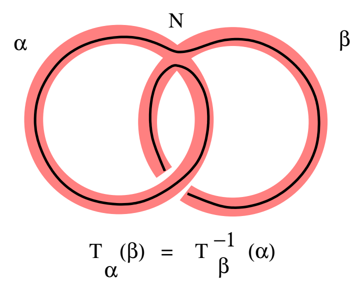
Lemma 34 (Dehn twists satisfy braid relations). *If α and β have intersection number 1, then*
$$T_\beta T_\alpha T_\beta = T_\alpha T_\beta T_\alpha$$

Proof. First, by Lemma 26 above we have the equalities:

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

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

Now, if α and β have a single intersection point, then a regular neighborhood N of $\alpha \cup \beta$ is the union of two annuli intersecting along a square. Observe that the images of both $T_\alpha(\beta)$ and $T_\beta^{-1}(\alpha)$ lie in this regular neighborhood and these isotopy classes of curves on N agree, as can be seen in the picture:



Thus $T_\alpha(\beta) = T_\beta^{-1}(\alpha)$ and hence, by Lemma 26 again the Dehn twists along them should agree, namely:

$$T_\alpha T_\beta T_\alpha^{-1} = T_\beta^{-1} T_\alpha T_\beta$$

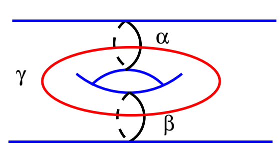
which is equivalent to the desired relation.

Lemma 35 (Chains of curves). *Let α and β be nonseparating simple closed curves on S. Then there exists a sequence of nonseparating simple closed curves $\alpha = \alpha_1, \alpha_1, \alpha_2, \ldots, \alpha_n = \beta$, such that*

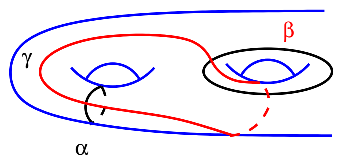
$$i(\alpha_j, \alpha_{j+1}) = 1, \text{ for } j \in \{1, 2, \ldots, n-1\}$$

Proof. We use induction on $i(\alpha, \beta)$. When $i(\alpha, \beta) = 0$ we have two cases:

1. If $\alpha \cup \beta$ separate S, then up to a homeomorphism of the surface S the situation the one pictured below:



2. If $\alpha \cup \beta$ does not separate S then, up to a homeomorphism of S we are as in the picture below:

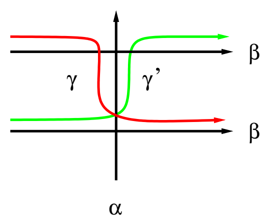


In both cases we can take $n = 3$ and $\alpha = \gamma$.

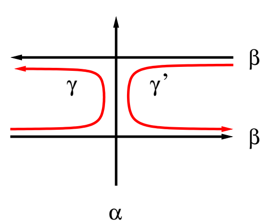
Assume now that our claim holds for all curves α, β satisfying $i(\alpha, \beta) \leq k$ and we want to prove it for two curves with geometric intersection number $k+1$. We can suppose that $k \geq 1$.

We give α and β arbitrary orientations. There are again two cases to consider:

1. When one travels along it α crosses twice β consecutively with the same orientation. We choose then a curve among the two curves γ or γ' in the figure below:



and notice that:
   a) The following equality $\gamma + \gamma' = \beta$ holds in the homology group $H_1(S)$. As a simple curve is nonseparating if and only if its class in homology is nonzero, the it follows that at least one of the two curves γ or γ', say γ, is nonseparating.
   b) $i(\gamma, \beta) = 1$;
   c) $i(\gamma, \alpha) \leq i(\beta, \alpha) - 1$.
2. Otherwise the consecutive intersections of α with β have opposite orientations. We consider the two simple curves γ and γ' drawn in the picture below:



and note again that:

a) Then $\gamma \cup \gamma'$ and β have the same homology class as they cobound a subsurface. As β was nonseparating, either γ or γ', say γ, is nontrivial in homology and hence nonseparating, as well.
b) $i(\gamma, \beta) = 0$;
c) $i(\gamma, \alpha) \leq i(\beta, \alpha) - 2$.

We can therefore apply the induction hypothesis to the pair α, γ and the discussion above for the case $i(\gamma, \beta) = 0$ to conclude.

Proposition 6. *Let α be an oriented nonseparating simple closed curve and $\varphi \in \mathrm{Mod}(S)$. Then there exists a sequence of Dehn twists T_{α_i} such that*

$$\varphi(\alpha) = T_{\alpha_1} T_{\alpha_2} \cdots T_{\alpha_n}(\alpha)$$

Proof. By the proof of Lemma 34 we have

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

for any *nonoriented* simple closed curves α and β with geometric intersection 1. Note that if now α and β are *oriented* curves, then we can only write

$$\alpha^{\omega(\alpha,\beta)} = T_\beta T_\alpha(\beta),\ \beta^{-\omega(\alpha,\beta)} = T_\alpha T_\beta(\alpha)$$

as the orientation of the right hand sides depends on whether the *algebraic intersection* $\omega(\alpha, \beta) \in \{-1, 1\}$ is positive or not. Moreover, if we replace β in the right hand side of the first formula by the right hand side of the second formula we obtain:

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

Therefore, we can change β into β^{-1} by further use of Dehn twists, if necessary.

Now, Lemma 35 provides us with a chain $\alpha = \alpha_1, \ldots, \alpha_n = \varphi(\alpha)$ of simple closed curves with geometric intersection 1. Give arbitrary orientations to α_i, for $1 < i < n$. By above, each α_{i+1} is a product of Dehn twists applied to α_i, as oriented curves. This proves the claim.

5.5 Birman's exact sequence

An essential tool in the theory of mapping class group is the following result (see [2]):

Lemma 36 (Birman's exact sequence). *Let S be a compact surface possibly with punctures and P a point of $\mathrm{int}(S)$ disjoint the punctures. Then there is an exact sequence*

$$\pi_1(\mathrm{Homeo}_0(S)) \to \pi_1(S, P) \to \mathrm{Mod}(S, P) \to \mathrm{Mod}(S) \to 1$$

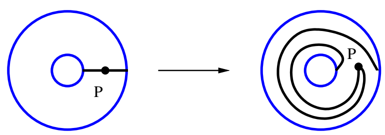
where:

1. *the homomorphism* $J : \pi_1(\mathrm{Homeo}_0(S)) \to \pi_1(S, P)$ *associates to any loop* $\phi_t : [0, 1] \to\in \mathrm{Homeo}_0(S)$ *based at the identity the homotopy class of the loop* $t \in [0, 1] \to \phi_t(P) \in S$.
2. *the homomorphism* $\mathrm{Mod}(S, P) \to \mathrm{Mod}(S)$ *is induced by the tautological homomorphism* $\mathrm{Homeo}^+(S, P) \hookrightarrow Homeo(S)$ *which forgets about the marked point* P*;*
3. *the* push *map* $\iota : \pi_1(S, P) \to \mathrm{Mod}(S, P)$ *is given for any homotopy class of an oriented simple loop* α *on* S *based at* P*, by the formula:*

$$\iota(\alpha) = T_{\alpha^+} T_{\alpha^-}^{-1}$$

where α^+ *and* α^- *are the result of pushing slightly* α *away of* P *by an isotopy of* S *in the left and respective the right of the puncture with respect to the orientation of* α.

Proof. Consider an annulus neighborhood N of α in S. The action of $\iota(\alpha)$ on the once punctured cylinder (N, P) is given in the picture below:



By pushing P back along the loop α we find an isotopy $\iota(\alpha)_t$ between $\iota(\alpha)$ and the identity on S.

Conversely, if the class of the homeomorphism ϕ of (S, P) belongs to the kernel $\ker(\mathrm{Mod}(S, P) \to \mathrm{Mod}(S))$, then $\phi(P) = P$ and ϕ is isotopic to the identity on S through some isotopy ϕ_t, $t \in [0, 1]$. Then the path $J(\phi_t) : [0, 1] \to S$ given by $J(\phi_t)(t) = \phi_t(P)$ is a based loop at P. Its class in $\pi_1(S, P)$ depends on the isotopy ϕ_t, but its image which we denote by $\overline{J}(\phi)$ in $\pi_1(S, P)/J(\pi_1(\mathrm{Homeo}_0(S))$ is welldefined. Moreover, $\overline{J}(\iota(\alpha)) = \alpha$. To prove the exactness of the exact sequence above it suffices to show that

$$\overline{J} : \ker(\mathrm{Mod}(S, P) \to \mathrm{Mod}(S)) \to \pi_1(S, P)/\pi_1(\mathrm{Homeo}_0(S))$$

is injective.

If ϕ is in the kernel of $\overline{J}$ we can change ϕ_t by an element of $\pi_1(\mathrm{Homeo}_0(S))$ and make $\overline{J}_\phi$ be a nullhomotopic simple curve. It then bounds a 2-disk by Lemma 21. Let D_1 be a disk containing it in the interior.

Now, given a pair of 2-disks $D_1 \subset \mathrm{int}(D_2)$ and $P \in \mathrm{int}(D_1)$, there exists a continuous family of homeomorphisms $\psi : D_1^2 \to \mathrm{Homeo}^+(D_2, \partial D_2)$ such that $\psi(q)(q) = P$, for any $q \in D_1$. If $D_2 \subset S$ is embedded, we can extend each $\psi(q)$ by the identity outside D_2 and obtain a continuous map $\psi : D_1^2 \to \mathrm{Homeo}_0(S)$.

Therefore the map $\Phi_{s,t} : [0, 1] \times [0, 1] \to \mathrm{Homeo}_0(S)$,

$$\Phi_{s,t} = \psi_{s\phi_t(P)+(1-s)P} \circ \phi_t$$

provides an isotopy between ϕ and the identity on (S, P).

Remark 14. The proof actually shows that the evaluation map $\mathrm{Homeo}(S) \to S$ sending ϕ to $\phi(P)$ is a locally trivial fiber bundle of $\mathrm{Homeo}(S)$ over S. Then Birman's exact sequence is the long exact sequence in homotopy associated to this fiber bundle.

Lemma 37. *If $\pi_1(S)$ is not virtually abelian (see Corollary 5) then Birman's exact sequence reduces to the short exact sequence:*

$$1 \to \pi_1(S,P) \to \mathrm{Mod}(S,P) \to \mathrm{Mod}(S) \to 1$$

Proof. Choose a lift $\phi_t \in \mathrm{Homeo}(S)$ of the loop α based at P, namely such that $\phi_t(P) = \alpha(t)$, $t \in [0,1]$ and $\phi_0 = \mathrm{id}$. Then $\phi_1 \in \mathrm{Homeo}(S,P)$ acts on the homotopy group $\pi_1(S,P)$ by conjugacy by the class $\alpha \in \pi_1(S,P)$. In particular, if the class $\alpha \in J(\pi_1(\mathrm{Homeo}_0(S)))$, then we can take a lift ϕ_t of the loop α with $\phi_1 = \mathrm{id}$. It follows that $J(\pi_1(\mathrm{Homeo}_0(S))$ is central in $\pi_1(S,P)$.

A more direct proof is as follows. Let ϕ_t be a loop representing a class in $\pi_1(\mathrm{Homeo}_0(S))$ and $\alpha : [0,1] \to S$ be an arbitrary loop in S based at P. Set $\phi = (\phi_t, t) : S \times [0,1] \to S \times [0,1]$ and $\pi : S \times [0,1] \to S$ be the first factor projection. Then the map $\Phi : [0,1] \times [0,1] \to S$ given by $\Phi = \pi \circ \phi \circ (\alpha, \mathrm{id})$ provides a homotopy based at P between the loops $\alpha J(\phi_t)$ and $J(\phi_t)\alpha$.

By Theorem 4 the center of $\pi_1(S)$ is trivial unless S is virtually abelian. The previous arguments are essentially due to Hamstrom ([2]) and Scott ([3]).

Lemma 38. *Let S be a compact surface possibly with punctures and P a point of $\mathrm{int}(S)$ disjoint from the punctures and S° be S deprived of a small open disk D around P with boundary δ. Then there is an exact sequence*

$$1 \to \mathbb{Z} \to \mathrm{Mod}(S^\circ) \to \mathrm{Mod}(S,P) \to 1$$

where the kernel $\mathbb{Z}$ is generated by the Dehn twist T_δ along the the curve δ.

Proof. Let $\varphi^\circ \in \mathrm{Homeo}(S^\circ)$ be a lift of a class with trivial image in $\mathrm{Mod}(S,P)$. Then the extension $\varphi \in \mathrm{Homeo}(S,P)$ by identity on D should be isotopic to the identity on (S,P).

Two simple closed curves or properly embedded arcs in $S^\circ - \delta$ which are homotopic in $S - P$, should also be homotopic and hence isotopic in S°, because $S^\circ - \delta \hookrightarrow S - P$ is a homotopy equivalence. Let N be an open cylindrical neighborhood of δ. Pick up a set $\mathcal{C}$ of curves and arcs so that cutting $S^\circ - N$ along them we obtain a disk. Then $\varphi(C)$ is isotopic to $\mathcal{C}$ in S°. In particular, by Alexander's Lemma 23 we can assume that the restriction of φ to $S^\circ - N$ is the identity. Thus φ is supported on the cylindrical neighborhood N. Lemma 27 shows that the class of φ belongs to the subgroup $\langle T_\delta \rangle$. As T_δ belongs to the kernel, we are done.

References

1. J.S. Birman, Braids, links and mapping class groups, Ann. Math. Studies no.82, Princeton Univ. Press., 1974.

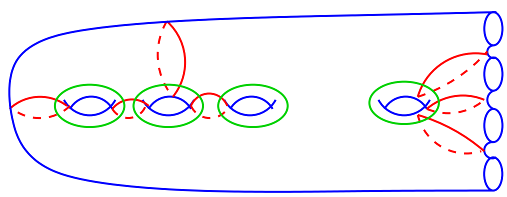
2. M.E. Hamstrom, *Homotopy groups of the space of homeomorphisms on a 2-manifold*, Illinois J. Math. 10 (1966), 563–573.
3. P.Scott, *The space of homeomorphisms of a 2-manifold*, Topology 9 (1970), 97–109.

5.6 Dehn's twists generate the mapping class group

The following result emphasizes the major role of Dehn twists in the theory:

Theorem 20 (Dehn 1938). *The mapping class group* Mod(S) *of a compact oriented surface* S *is generated by the Dehn twists along simple closed curves.*

More efficient set of generators were produced by Lickorish in [3] and later by Gervais. Lickorish's generators are the Dehn twists along the set of curves below:



Thus, there are $2g+1$ Dehn twists generators if $b=0$ (and $2g+b$, if $b>0$). This was shown to be sharp by Humphries (see [2]). If we seek for generators which are not necessarily Dehn twists Wajnryb proved in [4] that two generators suffice to generate Mod_g and $\mathrm{Mod}_{g,1}$ and this is of course the best possible for $g\geq 1$.

Proof. We will use induction on the genus g and the number n of boundary components. If $g=0$ and $n\leq 1$ the mapping class group is trivial, while for $n=2$ the result is given by Lemma 27. If $g=1$ and $n=0$ the mapping class group is isomorhic to $SL(2,\mathbb{Z})$, by Lemma 5. The action of a Dehn twist along a meridian and a longitude in homology is given by $\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$, respectively. These matrices generate $SL(2,\mathbb{Z})$ and hence prove the claim.

Let $\varphi\in\mathrm{Mod}_{g,n}$. Suppose that $n\geq 2$ and $g\geq 1$, as the case $g=1$ and $n\in\{0,1\}$ was already covered in the proof of Lemma 5. The image $\overline{\varphi}\in\mathrm{Mod}_{g,n-1}$ of φ by the map induced by capping off one boundary component can be written as a product of Dehn twists, by the induction hypothesis for $(g,n-1)$. The kernel of $\mathrm{Mod}_{g,n}\to\mathrm{Mod}_{g,n-1}$ is generated by Dehn twists, by Lemmas 37 and 38. Thus φ is a product of Dehn twists.

Suppose now that $g\geq 2$. Then there exists some nonseparating curve α on $S_{g,n}$. By Proposition 6 there exists simple curves α_i such that $\varphi'=\varphi(T_{\alpha_1}\cdots T_{\alpha_p})^{-1}$ keeps α pointwise fixed and hence induces a mapping class

$\overline{\varphi'}$ on the surfaces $S_{g-1,n+2}$ obtained by after cutting S along α. Furthermore, the image of $\overline{\varphi'}$ in $\mathrm{Mod}_{g-1,n}$ by capping off the new boundary components by disks is a product of Dehn twists, by the induction hypothesis for $(g-1,n)$. The kernel $\ker(\mathrm{Mod}_{g,n+2} \to \mathrm{Mod}_{g-1,n})$ is generated by Dehn twists along boundary curves and Dehn twists around simple closed curves. It follows that $\overline{\varphi'}$ is a product of Dehn twists on $S_{g-1,n+2}$ and hence φ, too.

Corollary 8. *The group* $\mathrm{Mod}_{g,n}$ *is generated by finitely many Dehn twists.*

References

1. M. Dehn, *Die Gruppe der Abdilungsklassen*, Acta Math. 69 (1938), 135–206.
2. S. Humphries, *Generators for the mapping class group*, Topology of low dimensional manifolds, Proc. Second Sussex Conf., Chelwood Gate 1977, 44-47, Lecture Notes in Math. no. 722, Springer, 1979.
3. W.B.R. Lickorish, *A finite set of generators for the homeotopy group of a 2-manifold*, Math. Proc. Cambridge Phil. Soc. 60 (1964), 769–778.
4. B. Wajnryb, *Mapping class group of a surface is generated by two elements*, Topology 35 (1996), 377–383.

5.7 Braids

Braid groups were introduced by Artin in 1926 and widely studied as a particular but meaningful case of mapping class groups:

Definition 11. *The classical* `braid group` B_n `on` n `strands` *is the mapping class group* $\mathrm{Mod}_{0,1}^n$ *of a disk with* n *punctures.*

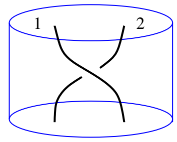
The pure braid subgroup PB_n is $\mathrm{PMod}_{0,1}^n$. Observe that $B_1 = 1$, by Alexander's Lemma, the first nontrivial braid group is B_2. If $\varphi \in \mathrm{Homeo}(S_{0,1}^n)$ is a lift of a braid, then φ is a isotopic to the identity on the disk. let ϕ_t be such an isotopy. The set of properly embedded arcs, called *strands*, in $D^2 \times [0,1]$ which describe the trajectory

$$((\phi_t(P_1), \phi_t(P_2), \ldots, \phi_t(P_n)), t) \subset D^2 \times [0,1]$$

of the marked points P_i under the isotopy is called a *geometric braid*. The defining property of a geometric braid is that it intersects each slice $D^2 \times \{t\}$ in the same number n of (distinct) points. It is clear now that two geometric braid represent the same class in B_n if and only if they are isotopic among geometric braids. This was enhanced by Artin who showed that:

Theorem 21 (Artin 1947). *Braids are in one-to-one correspondence with geometric braids up to isotopy keeping their endpoints fixed within* $D^2 \times [0,1]$.

This three dimensional interpretation provides a convenient way to draw braids, for instance the following *positive standard braid* in B_2 ($D^2 \times [0,1]$ will be often omitted from braid pictures):



Geometric braids can be oriented, letting each strand to go downside in the time direction. Juxtaposition of geometric braids then corresponds to composition of braids. The projection onto a vertical plane restricts to an immersion on a generic geometric braid. Its double points are called *crossings* of the image, which is a *braid diagram*. Once we decide which strands passes above the other at each crossing we uniquely recover the braid from its braid diagram.

Lemma 39. *We have $B_2 = \mathbb{Z}$ generated by a standard braid.*

Proof. The standard braid permutes the two punctures nontrivially and its square is a Dehn twist along the boundary. On the other hand, from Lemma 38 the pure braid group PB_2 is $\mathbb{Z}$, generated by the boundary Dehn twist.

Definition 12. *A standard braid or* `half-twist` *in* $\mathrm{Mod}_{g,b}^n$ *is the image of a standard braid in* B_2 *by an embedding of the twice punctured disk in* $S_{g,b}^n$*.*

In particular, we have a standard braid σ_i associated to the punctures P_i, P_{i+1} of the n-punctured disk. The first presentation of B_n is due to Artin:

Theorem 22. *The group B_n has a presentation with generators σ_i, $1 \le i \le n-1$ and the following set of relations, called braid relations:*

$$\sigma_i\sigma_j = \sigma_j\sigma_i, \text{ if } |i-j| \ge 2, \ 1 \le i,j \le n-1$$

$$\sigma_i\sigma_{i+1}\sigma_i = \sigma_{i+1}\sigma_i\sigma_{i+1}, \ 1 \le i \le n-2$$

Proof. Every braid diagram is a composition of standard braids, so the σ_i generate B_n. Relations are easily proved by drawing the corresponding pictures. To show that these relations suffice, we have to show that the projection onto the vertical plane of any isotopy of geometric braids among geometric braids could only affect the braid diagram by some relation. We skip the details.

A classical reference on braid groups is [2] and a recent comprehensive treatement is [3]. When passing to the full mapping class group $\mathrm{Mod}_{g,b}^n$ of a punctured surface we need to take into account mapping classes which permute punctures, as half-twists. As transpositions generate the symmetric group, we derive from Dehn's Theorem above:

Corollary 9. *If $g \ge 0$, $b \ge 0$ and $n \ge 2$ then $\mathrm{Mod}_{g,b}^n$ is generated by finitely many Dehn twists and half-twists.*

Exercise 13. Show that $\mathrm{Mod}_{0,3} = \mathbb{Z}^3$.

Exercise 14. Show that $\mathrm{Mod}_{1,1}$ is isomorphic to the braid group B_3.

Exercise 15. Show that $\mathrm{PMod}_0^4 = \mathbb{F}_2$ is freely generated by two Dehn twists so that Mod_0^4 is the following extension:

$$1 \to \mathbb{F}_2 \to \mathrm{Mod}_0^4 \to \mathcal{S}_4 \to 1$$

References

1. E. Artin, *Theory of braids*, Ann. Math. 48 (1947), 101–126.
2. J.S. Birman, Braids, links and mapping class groups, Ann. Math. Studies no.82, Princeton Univ. Press., 1974.
3. C. Kassel and V.Turaev, Braid Groups, Springer, 2008.

5.8 Relations between Dehn twists

In order to verify a relation $R = 1$ in $\mathrm{Mod}(S)$ we choose properly embedded arcs and curves λ_i such that $S - \cup_i \lambda_i$ is homeomorphic to a disk. Such a system λ_i will be called a *cut system* on S. We verify that $R(\lambda_i)$ is isotopic to λ_i by an explicit picture and then use Alexander's Lemma 23 to conclude that $R = 1$ in $\mathrm{Mod}(S)$. We say that we use *Alexander's method* to verify the relation and in most cases explicit derivation is immediate and will be omitted.

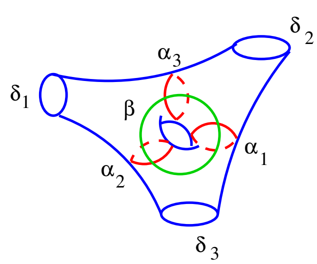
We already encountered in Lemmas 33 and 34 two fundamental relations satisfied by the Dehn twists, which we recall here for reader's convenience:

$$T_\alpha T_\beta = T_\beta T_\alpha, \text{ if } i(\alpha, \beta) = 0$$

$$T_\alpha T_\beta T_\alpha = T_\beta T_\alpha T_\beta, \text{ if } i(\alpha, \beta) = 1$$

Now, using Alexander's method we obtain more relations, as follows:

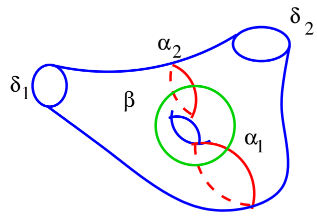
Lemma 40 (Star relation). *Assume that δ_1, δ_2 and δ_3 are curves bounding a 3-holed torsu $S_{1,3} \subset S$. Let α_i be the three pairwise nonequivalent meridians on $S_{1,3}$ and β a longitude, as in the picture below:*



Then the associated Dehn twists satisfy the so-called star *relation:*

$$(T_{\alpha_1} T_{\alpha_2} T_{\alpha_3} T_\beta)^3 = T_{\delta_1} T_{\delta_2} T_{\delta_3}$$

If δ_3 bounds a disk in S, the picture becomes a 2-holed torus:



We then derive the following *chain* relation on $S_{1,2} \subset S$:

$$(T^2_{\alpha_1} T_{\alpha_2} T_\beta)^3 = T_{\delta_1} T_{\delta_2}$$

The braid relations among the curves $\alpha_i, \beta, \delta_j$ read:

$$T_{\alpha_1} T_{\alpha_2} = T_{\alpha_2} T_{\alpha_1},\ T_{\alpha_i} T_{\delta_j} = T_{\delta_j} T_{\alpha_i},\ T_{\alpha_i} T_\beta T_{\alpha_i} = T_\beta T_{\alpha_i} T_\beta$$

By using them above, we can transform the star relation, as follows:

$$\begin{aligned}
(T_{\alpha_2} T_\beta T_{\alpha_1})^4 &= T_{\alpha_2} T_\beta T_{\alpha_1} (T_{\alpha_2} T_\beta) T_{\alpha_1} T_{\alpha_2} T_\beta T_{\alpha_1} T_{\alpha_2} T_\beta T_{\alpha_1} = \\
&= T_{\alpha_2} T_\beta T_{\alpha_1} (T_\beta T_{\alpha_2} T_\beta T_{\alpha_2}^{-1}) T_{\alpha_1} T_{\alpha_2} T_\beta T_{\alpha_1} T_{\alpha_2} T_\beta T_{\alpha_1} = \\
&= T_{\alpha_2} T_\beta T_{\alpha_1} T_\beta T_{\alpha_2} T_\beta (T_{\alpha_2}^{-1} T_{\alpha_1} T_{\alpha_2}) T_\beta T_{\alpha_1} T_{\alpha_2} T_\beta T_{\alpha_1} = \\
&= T_{\alpha_2} (T_\beta T_{\alpha_1} T_\beta) T_{\alpha_2} T_\beta T_{\alpha_1} T_\beta T_{\alpha_1} T_{\alpha_2} T_\beta T_{\alpha_1} = \\
&= T_{\alpha_2} (T_{\alpha_1} T_\beta T_{\alpha_1}) T_{\alpha_2} (T_\beta T_{\alpha_1} T_\beta) T_{\alpha_1} T_{\alpha_2} T_\beta T_{\alpha_1} = \\
&= T_{\alpha_2} T_{\alpha_1} T_\beta T_{\alpha_1} T_{\alpha_2} (T_{\alpha_1} T_\beta T_{\alpha_1}) T_{\alpha_1} T_{\alpha_2} T_\beta T_{\alpha_1} = \\
&= T_{\alpha_2} T_{\alpha_1} T_\beta T^2_{\alpha_1} T_{\alpha_2} T_\beta T_{\alpha_1} T_{\alpha_1} T_{\alpha_2} T_\beta T_{\alpha_1} = \\
&= T_{\alpha_1}^{-1} (T^2_{\alpha_1} T_{\alpha_2} T_\beta)^3 T_{\alpha_1} = T_{\alpha_1}^{-1} (T_{\delta_1} T_{\delta_2}) T_{\alpha_1} = T_{\delta_1} T_{\delta_2}
\end{aligned}$$

We thus obtained a new instance of the star relation:

$$(T_{\alpha_2} T_\beta T_{\alpha_1})^4 = T_{\delta_1} T_{\delta_2}$$

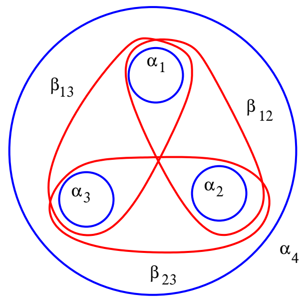
A particularly nice and symmetric presentation of the mapping class group was obtained by Gervais:

Theorem 23 (Gervais 2001). *We add to Lickorish's generators the Dehn twists along curves encircling precisely two boundary circles. Then the braid relations along with the star relations among them form a presentation of the mapping class group of a compact orientable surface with boundary.*

The finite presentability of the mapping class groups was proved by Hatcher and Thurston (see [2]) in 1980 using its action on pants complexes. Later Wajnryb provided explicit presentations in several papers culminating with [5], which were eventually shown to be equivalent to the above symmetric presentation by Gervais in [1].

Another relation which proved to be an important ingredient in many algebraic approaches to the mapping class group is:

Lemma 41 (Lantern relation). *Assume that $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ are simple curves on S bounding together $S_{0,4} \subseteq S$. Consider the simple curves $\beta_{12}, \beta_{13}, \beta_{23}$ on $S_{0,4}$, such that β_{ij}, α_i and α_j bound together a 3-holed sphere, for all distinct $i, j, k \in \{1, 2, 3\}$, as in the figure below:*



Then the associated Dehn twists satisfy the so-called lantern *relation:*

$$T_{\alpha_1} T_{\alpha_2} T_{\alpha_3} T_{\alpha_4} = T_{\beta_{12}} T_{\beta_{13}} T_{\beta_{23}}$$

Corollary 10. *The group $\mathrm{Mod}_{g,n}$ is generated by finitely many Dehn twists along nonseparating curves, if $g \geq 2$ or $g = 1$ and $n \leq 1$.*

Proof. The proof of Dehn's Theorem 20 shows that it suffices to consider Dehn twists along finitely many curves. Note that α^+ and α^- are nonseparating if α is nonseparating. It remains to see that we can get rid of the separating curves arising from Lemma 38.

When $g \geq 2$ and $p \geq 1$ we can embed $S_{0,4}$ into $S_{g,p}$ such that one boundary circle of $S_{0,4}$ is sent into a boundary circle of $S_{g,p}$, while the other boundary circles of $S_{0,4}$ are sent into nonseparating curves on $S_{g,p}$. The lantern relation shows that the Dehn twist along a boundary circle can be expressed as a product of Dehn twists along nonseparating curves.

As PMod_g^n is a quotient of Mod_g^n, by Lemma 38, we immediately derive:

Corollary 11. *The group PMod_g^n is generated by finitely many Dehn twists, which can be taken along nonseparating curves, if $g \geq 2$ or $g = 1$ and $n \leq 1$.*

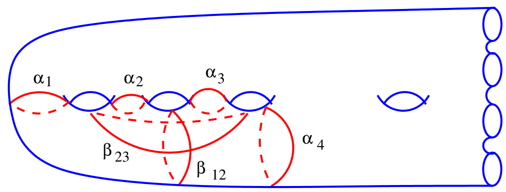
A last consequence of the relations above is the computation of the abelianization:

Corollary 12 (Mumford 1967, Powell 1978). *We have $H_1(\mathrm{Mod}_{g,p}) = 0$, if $g \geq 3$ and $H_1(\mathrm{Mod}_{2,p}) = \mathbb{Z}/10\mathbb{Z}$, $H_1(\mathrm{Mod}_1) = \mathbb{Z}/12\mathbb{Z}$. Furthermore $H_1(\mathrm{Mod}_{1,r}) = \mathbb{Z}^r$, if $r \geq 1$.*

Proof. Suppose that $g \geq 2$. Two nonseparating curves are equivalent by Mod_g, so their Dehn twists are conjugate. Therefore their images in $H_1(\mathrm{Mod}_g)$ coincide. As Mod_g is generated by Dehn twists along nonseparating curves when

$g \geq 2$, we derive that $H_1(\mathrm{Mod}_g)$ is cyclic generated by the class $[T]$ of a nonseparating Dehn twist.

On the other hand, if $g \geq 3$ we can embed $S_{0,4}$ into $S_{g,p}$ such that all involved curves are nonseparating:



It follows now from the Lantern relation that

$$4[T] = 3[T]$$

and hence $H_1(\mathrm{Mod}_{g,p}) = 0$.

When $g = 2$ we embed $S_{1,2}$ such that all five curves from the chain relation are nonseparating. Then the chain relation implies

$$12[T] = 2[T]$$

so that $H_1(\mathrm{Mod}_{2,p})$ is a quotient of $\mathbb{Z}/10\mathbb{Z}$. The explicit presentation of Gervais shows that this is an equality.

When $g = 1$, recall that Mod_1 is isomorphic to $SL(2,\mathbb{Z})$. We can generate $SL(2,\mathbb{Z})$ by the matrices $S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, $T = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$, which verify the relations: $S^4 = 1$, $(ST)^3 = S^2$. The ping-pong lemma shows that $PSL(2,\mathbb{Z}) = SL(2,\mathbb{Z})/\{\pm 1\}$ is a free product $\mathbb{Z}/2\mathbb{Z} \star \mathbb{Z}/3\mathbb{Z}$, the first factor being generated by S and the second by ST. Standard arguments then show that $SL(2,\mathbb{Z})$ has the presentation

$$SL(2,\mathbb{Z}) = \langle S, T; S^4 = 1, (ST)^3 = S^2 \rangle$$

This proves that $H_1(\mathrm{Mod}_1) = H_1(\mathrm{Mod}_1^1) = \mathbb{Z}/12\mathbb{Z}$.

Further, using the exact sequence

$$1 \to \mathbb{Z} \to \mathrm{Mod}_{1,1} \to \mathrm{Mod}_1^1 \to 1$$

we derive that $\mathrm{Mod}_{1,1}$ is isomorphic to the braid group B_3 on 3 strands. Thus $H_1(\mathrm{Mod}_{1,1}) = \mathbb{Z}$.

Let us denote by τ the image of a Dehn twist along a nonseparating curve into $H_1(\mathrm{Mod}_{1,r})$, $r \geq 1$. If α and β are a longitude and a meridian on $S_{1,1}$ and δ the boundary circle, then we derive easily from the chain relation that

$$T_\delta = (T_\alpha T_\beta)^6$$

Therefore τ is a generator of $H_1(\mathrm{Mod}_{1,1}) = \mathbb{Z}$.

Let $\delta_1, \delta_2, \ldots, \delta_r$ denote the boundary components of $S_{1,r}$. The lantern and chain relations show that $\mathrm{Mod}_{1,r}$ is generated by a Dehn twist τ_r along a nonseparating curve and the Dehn twists along $r-1$ boundary components, say $\delta_1, \delta_2, \ldots, \delta_{r-1}$. In particular, $H_1(\mathrm{Mod}_{1,r})$ is a quotient of $\mathbb{Z}^r$.

Assume that $n_0\tau_r + \sum_{i=1}^{r-1} n_i\delta_i = 0 \in H_1(\mathrm{Mod}_{1,r})$, where δ_i is the image of T_{δ_i} in the abelianization.

The surjective homomorphism $\mathrm{Mod}_{1,r} \to \mathrm{Mod}_{1,1}$ induced by the operation of gluing a disk along each one of the boundary curves $\delta_1, \ldots, \delta_{r-1}$ yields a homomorphism $H_1(\mathrm{Mod}_{1,r}) \to H_1(\mathrm{Mod}_{1,1})$. The later sends δ_i into 0 and τ_r onto the generator τ. The relation above implies $n_0 = 0$.

Further glue a disk along each boundary curve δ_i, for $i \neq j$ and also along the last boundary component δ_r, also provides a surjective homomorphism $\mathrm{Mod}_{1,r} \to \mathrm{Mod}_{1,1}$ inducing a homomorphism $H_1(\mathrm{Mod}_{1,r}) \to H_1(\mathrm{Mod}_{1,1})$. The later sends δ_j into 12τ, while δ_i, for $i \neq j$ into 0. From the relation above we derive that $n_j = 0$. Since j was arbitrary it follows that all n_i vanish, so that $H_1(\mathrm{Mod}_{1,r}) = \mathbb{Z}^r$, for $r \geq 1$. This result is due to Gervais, but the previous argument is due to Korkmaz.

Corollary 13. *If $r \geq 2$, then $\mathrm{Mod}_{1,r}$ cannot be generated by Dehn twists along nonseparating curves.*

References

1. S. Gervais, *A finite presentation of the mapping class group of a punctured surface*, Topology 40 (2001), 703–725.
2. A. Hatcher and W. Thurston, *A presentation of the mapping class group of a closed orientable surface*, Topology 19 (1980), 221–237.
3. D. Mumford, *Abelian quotients of the Teichmüller modular group*, J. Analyse Math. 18 (1967), 227–244.
4. J.Powell, *Two theorems on the mapping class group of a surface*, Proc. Amer. Math. Soc. 68 (1978), 347–350.
5. B. Wajnryb, *An elementary approach to the mapping class group of a surface*, Geom. Topology 3 (1999), 405–466.

5.9 The Dehn-Nielsen-Baer theorem

If G is a discrete group we denote by $\mathrm{Aut}(G)$ the group of its automorphisms. The simplest automorphisms are the conjugacies by elements of G, also called inner automorphisms. The later form a normal subgroup $\mathrm{Inn}(G) \triangleleft \mathrm{Aut}(G)$ which is isomorphic to $G/Z(G)$, where $Z(G)$ is the center of G. We denote by $\mathrm{Out}(G)$ the outer automorphisms group, which is the quotient $\mathrm{Aut}(G)/\mathrm{Inn}(G)$.

Proposition 7 (Magnus 1931). *If two elements R and R' have the same normal closure in a free group, then either R' is conjugate to R or else R' is conjugate to R^{-1}.*

Remark 15. Bogopolski extended Magnus' result by proving that it also holds for elements R, R' in a surface group (see [2]).

Consider an automorphism $\varphi : \pi_1(S) \to \pi_1(S)$. We can lift the elements $\varphi(\alpha_i)$, $\varphi(\beta_i)$ to some elements $\tilde{\varphi}(\alpha_i)$, $\tilde{\varphi}(\beta_i)$ in the free group $\mathbb{F}_{2g}$. Then the element $\prod_i [\varphi(\alpha_i), \varphi(\beta_i)]$ should determine the same normal subgroup as the relator $\prod_i [\alpha_i, \beta_i]$. If they are conjugate in the free group, one calls the automorphism *orientation preserving*. We denote by $\mathrm{Aut}^+(\pi_1(S))$ the subgroup of orientation preserving automorphisms of the surface group $\pi_1(S)$.

We say that an outer automorphism of $\pi_1(S)$ preserves the orientation if it lifts to an orientation preserving automorphism of $\pi_1(S)$ and denote by $\mathrm{Out}^+(\pi_1(S))$ the group they form.

Theorem 24 (Dehn-Nielsen-Baer Theorem, [1, 4]). *The homomorphism*

$$\mathrm{Mod}(S) \to \mathrm{Out}^+(\pi_1(S))$$

which associates to the mapping class φ the class of the automorphism $\varphi_* : \pi_1(S, p) \to \pi_1(S, p)$ is an isomorphism.

Proof. If the image of φ is trivial then we can choose the base point $*$ such that there is a homeomorphism lifting φ which induces the identity on $\pi_1(S, *)$. This means that φ is homotopic to the identity and hence it is isotopic to the identity, by the Baer-Epstein theorem. This settles the injectivity part.

Suppose now that the genus of S is at least 2. Fix a base point $*$ on S. Then any automorphism of $\pi_1(S, *)$ can be realised by a continuous map $\varphi : S \to S$ which induces the given automorphism at the level of $\pi_1(S, *)$. This is a general statement, valid for any manifold whose universal covering is contractible. Choose the curves α_i and β_i which only intersect at the base point $*$ and generate the fundemantal group of S. We can see S as the result of gluing a 2-disk along the cycle given by the relator R on the union of the curves α_i and β_i. There is a map φ from the union of curves into S, such that the images of α_i and β_i are curves based at $*$ representing the homotopy classes $\varphi(\alpha_i)$ and $\varphi(\beta_i)$. Since the product $\prod_{i=1}^{g} [\varphi(\alpha_i), \varphi(\beta_i)] = 1$ in $\pi_1(S, *)$ there exists a map from the 2-disk to S extending φ.

By Kneser's theorem we have

$$\chi(S) \leq d\chi(S)$$

so that $d \in \{-1, 0, 1\}$.

If the degree d were zero, then the image of f could be homotoped to lie in the 1-skeleton of S. It follows that $\varphi_* : \pi_1(S, *) \to \pi_1(S, *)$ sends $\pi_1(S)$ into the fundamental group of the 1-skeleton of S, which is a free group. As any subgroup of a free group is free we derive that $\varphi_*(\pi_1(S, *))$ is a free group, which contradicts the fact that φ_* is an automorphism of $\pi_1(S)$, which is not a free group.

Therefore $d = \pm 1$, but if φ preserves the orientation, then $d = 1$. Since we have an equality above, the second part of Kneser's theorem implies that φ is homotopic to a degree one covering, namely a homeomorphism.

The case when the genus $g = 1$ follows from the description of Mod_1 as $SL(2, \mathbb{Z})$.

Remark 16. The Dehn-Nielsen-Baer theorem also holds for nonorientable surfaces.

Corollary 14. *A surface group of genus $g \geq 2$ is co-Hopfian, namely any injective homomorphism into itself should be an isomorphism.*

Proof. Realise the homomorphism by a continuous map $\varphi : S \to S$ fixing the basepoint $*$. The previous arguments show that the degree of φ could only be zero or ± 1. In the first case the image of tjhe surface group would be a free group, which contradicts the injectivity of the group homomorphism. It follows that the degree is ± 1 and hence we conclude as above.

Remark 17. From Kneser's theorem a self map of a closed surface S of genus $g \geq 2$ could only have degree zero or ± 1. This is not true anymore for the case of the torus $S^1 \times S^1$, which admits self-maps of arbitrary degree constructed by using finite coverings of S^1. Further Hopf's theorem states that the degree map from the group $\pi_2(S^2)$ of homotopy classes of based maps $S^2 \to S^2$ into $\mathbb{Z}$ is an isomorphism.

References

1. R. Baer, *Kurventypen auf Flächen*, J. reine angew. Math. 156 (1927), 231–246.
2. O. Bogopoloski, *A surface groups analogue of a theorem of Magnus*, Geometric methods in group theory, Contemporary Math. 372, Amer. math. Soc. 2005, 59–69.
3. W. Magnus, *Über diskontinuerliche Gruppen mit einer definierenden Relation (Der Freiheitsatz)*, J. reine angew. Math. 163 (1930), 141–165.
4. J. Nielsen, *Untersuchungen zur Topologie der geschlossenen zweiseitigen Flächen I, II, III*, Acta Math. 50 (1927), 189–358, 53 (1929); 1–76; 58 (1932), 87–167.

6

The Torelli group

6.1 The symplectic representation

Definition 13. *The* `homology representation` *of the mapping class group* $\mathrm{Mod}(S)$ *is the homomorphism*

$$\rho : \mathrm{Mod}(S) \to \mathrm{Aut}(H_1(S), \omega)$$

which associates to $\varphi \in \mathrm{Mod}(S)$ *the element* $\rho(\varphi) = \varphi_*$ *which is the induced map in homology. The kernel of the homology representation is the* `Torelli group` $\mathcal{I}(S)$.

Note that ω is the classical symplectic form, when $S = S_g$ or $S_{g,1}$ while it has nontrivial kernel as soon as the number of boundary components is at least 2. Once we fix a basis in homology and hence an identification $H_1(S_g) = \mathbb{Z}^{2g}$, we derive an embedding of $\mathrm{Aut}(H_1(S), \omega)$ as a symplectic group of matrices

$$Sp(2g, \mathbb{Z}) = \{A \in GL(2g, \mathbb{Z}) \mid A^{\perp} J A = J\}$$

where $A^{\perp}$ denotes the transpose of A and J denotes the matrix of ω in this basis:

$$J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \oplus \cdots \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

Clebsch and Gordan discovered a generating system for $Sp(2g, \mathbb{Z})$ in 1866. But it is Burkardat ([**?**] who gave the first proof of this fact by showing that these generators are induced by the action of homeomorphisms, namely that:

Theorem 25 (Burkhardt 1890). *The homomorphism* $\mathrm{Mod}_g \to Sp(2g, \mathbb{Z})$ *is surjective.*

Proof. We first claim that the matrices below form generate the symplectic group $Sp(2g, \mathbb{Z})$:

1. $B = B_1 \oplus B_2 \oplus \cdots \oplus B_g$, where $B_i \in SL(2, \mathbb{Z})$;

2. $\mathbf{1} \oplus \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} \oplus \mathbf{1}$
3. $\mathbf{1} \oplus \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \oplus \mathbf{1}$

Consider an integral vector $\mathbf{m} \in \mathbb{Z}^{2g}$. Then there exists some B of type 1. such that

$$B\mathbf{m} = (m'_1, 0, m'_2, 0, \ldots, m'_g, 0)$$

Then a product C of matrices of the form 2. and 3. will reduce the vector at the form:

$$CB\mathbf{m} = (m''_1, 0, 0, 0, \cdots, 0)$$

where $m''_1 = \text{g.c.d.}(m_1, m_2, \ldots, m_{2g})$.

Let now $A \in Sp(2g, \mathbb{Z})$. By the previous reduction there exists some A which is a product of matrices of the form 1., 2. and 3. such that the first line of A_1A has the form $(a_1, 0, 0, \ldots, 0)$. Since symplectic matrices have determinant 1, we have $a_1 = 0$. Adding the first line to the second line a number of times can be obtained by multiplying by some matrix A_2 of type 1. and thus the second line of A_2A_1A starts with 0. Since this matrix is symplectic the first column is $(1, 0, 0, \ldots, 0)$ and the first two diagonal terms are 1. We can therefore reduce the second line of A_2A_1A by the same procedure without changing the first line to the canonical form $(0, 1, 0, 0, \ldots, 0)$ and then being symplectic the second row will have the same form. In particular, we reduced the matrix A to one of the form

$$J = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \oplus A'$$

where A' is also symplectic but of smaller size. Therefore our claim follows by induction on g.

Observe now that matrices of type 1. correspond to homology actions of mapping classes supported on tori. A matrix of type 2. corresponds to exchanging two handles of a surface. In order to realise a matrix of type 3. we consider the following automorphism of the surface group:

$$\alpha_1 \to \alpha_1\alpha_2,\ \beta_1 \to \alpha_2^{-1}\beta_1\alpha_2$$

$$\alpha_2 \to \alpha_2^{-1}\beta_1\alpha_2\beta_1^{-1}\alpha_2,\ \beta_2 \to \beta_2\alpha_2^{-1}\beta_1^{-1}\alpha_2$$

$$\alpha_i \to \alpha_i,\ \beta_i \to \beta_i,\ i \geq 3$$

By the Dehn-Nielsen-Baer theorem this automorphism can be realized as a mapping class and it induces the desired map at homological level.

Corollary 15. *On a closed orientable surface S_g, $g \geq 1$, the class $a \in H_1(S_g)$, $a \neq 0$ can be represented by a a simple closed curve if and only if it is primitive, i.e. not a nontrivial multiple of another class.*

Proof. If we cut S_g along a nonseparating simple curve and then join the boundary curves by a path we obtain another simple curve which intersects the first curve in a single point. Thus the homology class of a nonseparating simple curve is primitive. Conversely, if a is primitive, then we noted in the proof of Theorem 25 that there is some element of $Sp(2g, \mathbb{Z})$ sending a into the class of some generator $\alpha_1 = (1, 0, \ldots, 0)$. Then the claim follows from the Theorem 25, since a mapping class sends simple curves into simple curves.

References

1. H. Burkhardt, *Grundzüge einer allgemeinen Systematik der hyperelliptipschen Funktionen erster Ordnung*, Math. Annalen 35 (1890), 198–296.

6.2 Generators

We will focus mainly on the Torelli groups $\mathcal{I}_g = \mathcal{I}(S_g)$ and $\mathcal{I}_{g,1} = \mathcal{I}(S_{g,1})$ in the sequel.

Definition 14. *A* `BP` *pair is a pair (γ, δ) of simple closed curves on a surface S which bound together a subsurface of S. The* `genus` *of a BP-pair is the minimal genus of a subsurface they bound.*

Remark 18. The homology class of a simple closed curve γ vanishes if and only if γ bounds a subsurface of S, namely it is separating (also called bounding curve). Moreover, two simple closed curves γ and δ determine the same homology class in $H_1(S)$ if and only if they bound together a subsurface of S.

Definition 15. *A* `BP twist` *is the element $T_\gamma T_\delta^{-1}$ of $\mathrm{Mod}(S)$ associated to a BP pair (γ, δ). Further a* `B-twist` *is the Dehn twist T_γ along a bounding simple closed curve γ.*

Lemma 42. *The action of the Dehn twist in homology is given by a transvection:*

$$\rho(T_\gamma)(a) = a + \omega(\gamma, a) \cdot \gamma, \text{ for } a \in H_1(S)$$

Proof. When drawing the image $T_\gamma(\alpha)$ of the simple curve α at each intersection point of α with γ we follow once γ in the right direction if the intresection is positive and in the negative direction, otherwise. Thus its class in homology adds $\omega(\gamma, a) \cdot \gamma$ to the class a of α.

In particular B-twists and BP-twists belong to the group $\mathcal{I}(S)$. The main result of this section is due to D.Johnson:

Theorem 26 (Johnson). *If $g \geq 3$ then $\mathcal{I}_g$ is generated by genus one BP-twists.*

A key ingredient in Johnson's proof is the following result of

Proposition 8 (Birman-Powell). *If $g \geq 3$ then $\mathcal{I}_g$ and $\mathcal{I}_{g,1}$ are generated by BP-twists and B-twists.*

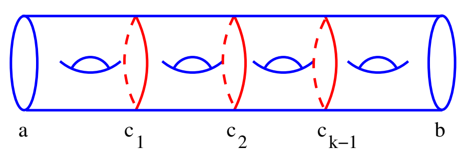
Denote by W_k the subgroup generated by BP-pairs of genus k and by N_k the subgroup generated by B-twists of genus k. They are normal subgroups of $\mathcal{I}_g$ as every homeomorphisms of S_g preserves the genus of curves and pairs of curves.

Then Birman-Powell's result above shows that

$$\mathcal{I}_g = W_1 W_2 \cdots W_g N_1 N_2 \cdots N_g$$

Lemma 43. $W_k \subseteq W_1$.

Proof. If (a, b) is a BP-pair of genus k let $c_1, c_2, \ldots, c_{k-1}$ be separating curves on the subsurface bounded by $a \cup b$ slicing it in subsurfaces of genus one, as in the figure below:



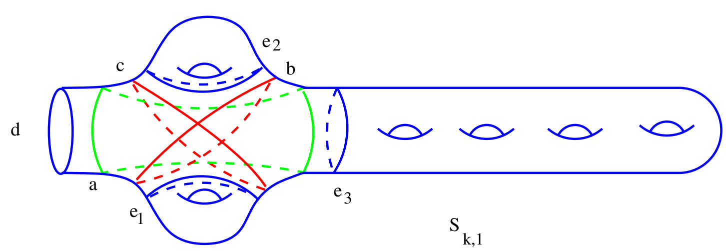
Then we can write:

$$T_a T_b^{-1} = (T_a T_{c_1}^{-1})(T_{c_1} T_{c_2}^{-1}) \cdots (T_{c_{k-1}} T_b^{-1})$$

thereby proving the claim.

Lemma 44. *For $k \geq 2$, $N_k \subseteq N_1 N_2$.*

Proof. Let d bound a subsurface $S_{k,1}$, $k \geq 3$. We then embed the 4-holed sphere $S_{0,4}$ in $S_{k,1}$ as the subsurface bounded by the curves d, e_1, e_2, e_3 in the figure below:



The lantern relation reads:

$$T_aT_bT_c = T_dT_{e_1}T_{e_2}T_{e_3}$$

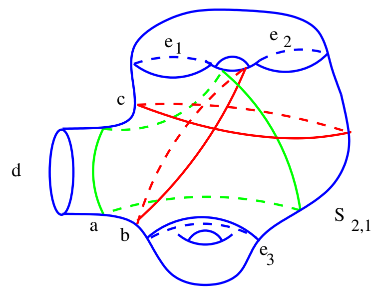
Note now that

$$T_{e_2} \in N_{k-2},\ T_{e_1}, T_{e_3} \in N_1,\ T_a \in N_2, T_b, T_c \in N_{k-1}$$

Therefore $T_d \in N_{k-1}N_1N_2$. By recurrence on k we obtain that $T_d \in N_1N_2$.

Lemma 45. $N_2 \subseteq W_1N_1$.

Proof. We now embed $S_{0,4}$ into $S_{2,1}$ as in the picture below



The lantern relation

$$T_aT_bT_c = T_dT_{e_1}T_{e_2}T_{e_3}$$

gives us

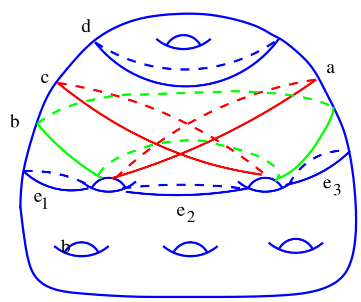
$$T_d = (T_aT_{e_1}^{-1})(T_bT_{e_2}^{-1})T_cT_{e_3}^{-1}$$

where

$$T_aT_{e_1}^{-1} \in W_1,\ T_bT_{e_2}^{-1} \in W_1,\ T_c \in N_1,\ T_{e_3}^{-1} \in N_1.$$

Lemma 46. *If* $g \geq 3$ *then* $N_1 \subseteq W_1$.

Proof. We consider the curve d of genus 1 on S_g, $g \geq 3$ and embed $S_{0,4}$ in S_g as in the picture below:



Then the lantern relation gives us:

$$T_d = (T_{e_1}^{-1}T_a)(T_bT_{e_2}^{-1})(T_cT_{e_3}^{-1})$$

where

$$T_{e_1}^{-1}T_a, T_bT_{e_2}^{-1}, T_cT_{e_3}^{-1} \in N_1.$$

6.3 Johnson's homomorphism

Consider in this section $S = S_{g,1}$, having free fundamental group π on $2g$ generators. Choose a basepoint $*$ on the boundary so that we have a natural embedding:

$$\mathrm{Mod}_{g,1} \to Aut^+(\pi)$$

Denote by H the homology $H_1(S_{g,1})$. If π' denotes the commutator subgroup $[\pi, \pi]$ then have a central extension:

$$0 \to \frac{\pi'}{[\pi, \pi']} \to \frac{\pi}{[\pi, \pi']} \to \frac{\pi}{[\pi, \pi]} = H \to 0$$

The kernel can be described in terms of $\wedge$ products:

Lemma 47. *The map $j : \wedge^2 H \to \frac{\pi'}{[\pi,\pi']}$ given by*

$$j(x \wedge y) = [\widetilde{x}, \widetilde{y}]$$

where $\widetilde{x}, \widetilde{y} \in \frac{\pi}{[\pi,\pi']}$ are arbitrary lifts of x, y respectively, is an isomorphism of abelian groups.

Proof. Note that $\frac{\pi'}{[\pi,\pi']}$ is a free abelian group of rank $\binom{2g}{2}$. By definition j is surjective and

$$j(x, y + z) = j(x, y)j(x, z),\ j(y, x) = j(x, y)^{-1}$$

Remark 19. Every automorphism of π induces an automorphism of $\frac{\pi}{[\pi,\pi']}$ leaving $\frac{\pi'}{[\pi,\pi']}$ invariant and the isomorphism j commutes with the automorphisms. Therefore $\mathcal{I}_{g,1}$ acts trivially on $\frac{\pi'}{[\pi,\pi']}$, the later being a $Sp(2g, \mathbb{Z})$-module isomorphic to $\wedge^2 H$.

Definition 16. *The* `Johnson homomomorphism` *is the map*

$$\delta : \mathcal{I}_{g,1} \to \mathrm{Hom}(H, \wedge^2 H) \subset \wedge^2 H \otimes H^*$$

given by

$$\delta(f)(x) = j^{-1}(f(\tilde{x})\tilde{x}^{-1})$$

where $\tilde{x} \in \frac{\pi}{[\pi,\pi']}$ is a lift of $x \in H$.

Proposition 9. *The map δ is a group homomorphism.*

Proof. As f is an element of the Torelli group, it acts trivially on H so that $f(\tilde{x})\tilde{x}^{-1} \in \frac{\pi'}{[\pi,\pi']}$. If $\tilde{x}\tilde{u}$, with $\tilde{u} \in \frac{\pi'}{[\pi,\pi']}$ is another lift of x, then

$$f(\tilde{x}\tilde{u})(\tilde{x}\tilde{u})^{-1} = f(\tilde{x})f(\tilde{u})\tilde{u}^{-1}\tilde{x}^{-1} = f(\tilde{x})\tilde{x}^{-1}$$

because f acts trivially on $\frac{\pi'}{[\pi,\pi']}$ by the previous Remark 19.

We now verify that

$$\delta f(x+y) = f(\tilde{x}\tilde{y})(\tilde{x}\tilde{y})^{-1} = f(\tilde{x})(f(\tilde{y})\tilde{y}^{-1})\tilde{x}^{-1} = \delta f(x)\delta f(y)$$

because the target group is abelian, so that δf is indeed a homomorphism $\mathrm{Hom}(H, \wedge^2 H)$. Eventually we have

$$\delta(fg)(x) = fg(\tilde{x})\tilde{x}^{-1} = f(g(\tilde{x})\tilde{x}^{-1})f(\tilde{x})x^{-1} = f(\delta g(x))\delta f(x) = \delta g(x)\delta f(x)$$

Now $\delta g(x) \in \frac{\pi'}{[\pi,\pi']}$ and so f acts trivially again by Remark 19, so that

$$f(\delta g(x))\delta f(x) = \delta g(x)\delta f(x)$$

This proves that δ is a homomorphism.

Remark 20. The Johnson homomorphism is $\mathrm{Mod}_{g,1}$-equivariant, namely for $h \in \mathrm{Mod}_{g,1}$ and $f \in \mathcal{I}_{g,1}$ we have:

$$\delta(hfh^{-1}) = h(\delta f)$$

where on the right hand side h acts by means of the homological representation on H and $\wedge^2 H$.

The symplectic form ω on H provides a duality isomorphism

$$\iota : \mathrm{Hom}(\wedge^2 H \otimes H) = \wedge^2 H \otimes \wedge^2 H \otimes H \subset \wedge^3 H$$

which is given by the formula

$$\iota(\nu) = \sum_{i=1}^{g} \nu(a_i) \otimes b_i - \nu(b_i) \otimes a_i$$

where $\{a_i, b_i\}_{i=1,g}$ is a symplectic basis in homology. Its inverse map sends the decomposable element $\theta \otimes x \in \wedge^2 H \otimes H$ into the homomorphism $\nu_{\theta\otimes x} : H \to \wedge^2 H$ given by:

$$\nu_{\theta\otimes x}(y) = \omega(y,x)\theta$$

It is therefore convenient to define another instance of the Johnson homomorphism as the homomorphism $\tau : \mathcal{I}_{g,1} \to \wedge^2 H \otimes H$ obtained by passing to the dual:

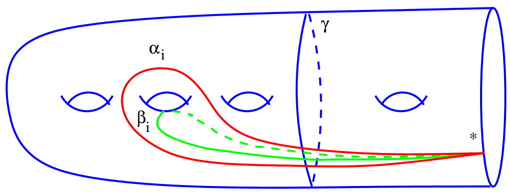
$$\tau(f) = \iota(\delta(f))$$

Definition 17. *The Johnson kernel is the subgroup $\mathcal{K}_{g,1} \subset \mathrm{Mod}_{g,1}$ generated by the B-twists.*

Obviously $\mathcal{K}_{g,1} \subseteq \mathcal{I}_{g,1}$. The following is the main result of this section, due to Johnson:

Theorem 27 (Johnson). *The image of the Johnson homomorphism τ is $\wedge^3 H \subset \wedge^2 H \otimes H$ and its kernel contains the subgroup $\mathcal{K}_{g,1}$.*

Proof. We first show that $\mathcal{K}_{g,1}$ is contained in the kernel of τ. To this purpose let γ be an arbitrary bounding simple curve on $S_{g,1}$. Set $c \in \pi$ be the homotopy class obtained by joining the base point $*$ to γ by an arc disjoint from the curves α_i, β_i which are providing a symplectic basis in homology. Suppose that γ bounds a subsurface $\Sigma_{k,1}$ of genus k.



Then the action on the Dehn twist T_γ on $\pi = \pi_1(S_{g,1}, *)$ can be readily computed:

$$T_\gamma(\alpha_i) = \alpha_i, \; T_\gamma(\beta_i) = \beta_i, \text{ for } i \geq k+1,$$

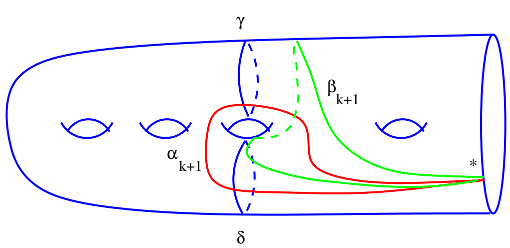
$$T_\gamma(\alpha_i) = c\alpha_i c^{-1}, \; T_\gamma(\beta_i) = c\beta_i c^{-1}, \text{ if } i \leq k.$$

Now c belongs to the commutator subgroup π', because γ is homologically trivial, so for any element x of the free subgroup generated by the α_i, β_i with $i \leq k$ we have

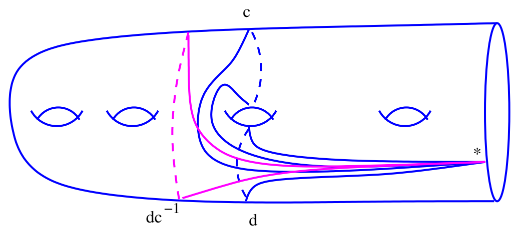
$$T_\gamma(x)x^{-1} = [c, x] \in [\pi, \pi']$$

Therefore $T_\gamma(x)x^{-1} = 0 \in \wedge^2 H$, for any $x \in \pi$, which amounts to say that $\delta(f) = 0$, and hence $\tau(\mathcal{K}_{g,1}) = 0$.

We now proceed with the computation of the image of τ. Consider a BP-twist $f = T_\gamma T_\delta^{-1}$ where the BP-pair (γ, δ) bounds a subsurface $S_{k,2}$ of genus k.



We join γ and δ to the base point $*$ by some arcs and let c and d, respectively denote the homotopy classes of the curves so obtained:



Now, we are able to find the action of f in homotopy:

$$f(x) = \begin{cases} x, & \text{if } x \in \{\alpha_i, \beta_i; i \geq k+1\} \cup \{\beta_{k+1}\}; \\ dxd^{-1}, & \text{if } x \in \{\alpha_i, \beta_i; i \leq k\}; \\ dc^{-1}\alpha_{k+1}, & \text{if } x = \alpha_{k+1} \end{cases}$$

As can be seen on the picture above $dc^{-1} = \prod_{i=1}^{k} [\alpha_i, \beta_i]$.

If we denote by $a_i, b_i \in H$ the homology classes of $\alpha_i \beta_i$ respectively, then the class of dc^{-1} belongs to $\frac{\pi'}{[\pi, \pi']} \cong \wedge^2 H$ and can be identified with the element $\sum_{i=1}^{k} a_i \wedge b_i$. Recall also that the homology class of d is b_{k+1}, so that we can write:

$$\delta f(a_i) = b_{k+1} \wedge a_i,\ \delta f(b_i) = b_{k+1} \wedge b_i,\ \text{for } i \leq k,$$

$$\delta f(a_{k+1}) = \sum_{i=1}^{k} a_i \wedge b_i,$$

while in the remaining cases:

$$\delta f(a_i) = \delta f(b_j) = 0,\ \text{whenever } i \geq k+2,\ \text{and } j \geq k+1.$$

Therefore the image of δf in $\wedge^3 H$ is

$$\tau(f) = \iota(\delta f) = \sum_{i=1}^{g} (\delta f(a_i) \otimes b_i - \delta f(b_i) \otimes a_i) =$$

$$= \left(\sum_{i=1}^{k} (b_{k+1} \wedge a_i) \otimes b_i - b_{k+1} \wedge b_1) \otimes a_i \right) + \sum_{i=1}^{k} (a_i \wedge b_i) \otimes b_{k+1} =$$

$$= \sum_{i=1}^{k} ((b_i \wedge b_{k+1}) \otimes a_i + (b_{k+1} \wedge a_i) \otimes b_i + (a_i \wedge b_i) \otimes b_{k+1})$$

Recall that the following identity from linear algebra holds in $\wedge^2 H \otimes H$ for any triple of vectors $x, y, z \in H$:

$$x \wedge y \wedge z = (x \wedge y) \otimes z + (y \wedge z) \otimes x + (z \wedge x) \otimes y$$

Replacing it in the formula above we obtain the final formula:

$$\tau(f) = \left(\sum_{i=1}^{k} a_i \wedge b_i\right) \wedge b_{k+1}$$

which shows that $\tau(\mathcal{I}_{g,1}) \subset \wedge^3 H$, as claimed.

Now, the image of τ is obviously $Sp(2g,\mathbb{Z})$-invariant. If $g = 2$ we already found above that the $a_1 \wedge b_1 \wedge b_2 \in \tau(\mathcal{I}_{g,1})$. Its image by $\mathbf{1} \oplus \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ is $a_1 \wedge b_1 \wedge a_2$, which also belongs to $\tau(\mathcal{I}_{g,1})$. Using handle permutations we derive that $a_2 \wedge b_2 \wedge b_1, a_2 \wedge b_2 \wedge a_1 \in \tau(\mathcal{I}_{g,1})$, which settles the genus 2 case.

When $g \geq 3$ we consider the follwing symplectic transformations: transformation $a_1 \to a_1 + b_1 - b_3$, $a_3 \to a_3 - b_1 + b_3$ and derive from above that $a_1 \wedge b_1 \wedge b_2 - b_1 \wedge b_2 \wedge b_3 \in \tau(\mathcal{I}_{g,1})$, so that $b_1 \wedge b_2 \wedge b_3 \in \tau(\mathcal{I}_{g,1})$.

Further, handle permutations show that $a_i \wedge b_i \wedge b_j \in \tau(\mathcal{I}_{g,1})$ and $a_i \wedge b_i \wedge a_j \in \tau(\mathcal{I}_{g,1})$, for all distinct indices $i \neq j$. Eventually we use the symplectic action of $\mathbf{1} \oplus \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \oplus 1$ to derive that

$$a_i \wedge a_j \wedge a_k, a_i \wedge a_j \wedge b_k, a_i \wedge b_j \wedge b_k, b_i \wedge b_j \wedge b_k \in \tau(\mathcal{I}_{g,1})$$

for i, j, k pairwise distinct. As these vectors form a basis of $\wedge^3 H$, which is of dimension $\binom{2g}{3}$, this means that the image of τ is indeed $\wedge^3 H$.

The second result improves the previous one:

Theorem 28 (Johnson 1985). *The kernel of the Johnson homomorphism τ coincides with $\mathcal{K}_{g,1}$.*

6.4 Higher Johnson subgroups

Recall that the lower central series of a group π is defined inductively as

$$\pi_1 = \pi,\ \pi_{k+1} = [\pi, \pi_k],\ \text{for } k \geq 2.$$

The mapping class group action $\mathrm{Mod}_{g,1}$ on π leaves π_k invariant, since higher commutators are characteristic subgroups.

Definition 18. *The k-th* `Johnson subgroup` *$\mathrm{Mod}_{g,1}(k)$ is the subgroup of $\mathrm{Mod}_{g,1}$ which acts trivially on the nilpotent quotient $\frac{\pi}{\pi_k}$.*

Thus $\mathrm{Mod}_{g,1}(1) = \mathrm{Mod}_{g,1}$, $\mathrm{Mod}_{g,1}(2) = \mathcal{I}_{g,1}$ and $\mathrm{Mod}_{g,1}(3) = \mathcal{K}_{g,1}$.

The higher Johnson homomorphisms

$$\delta_k : \mathrm{Mod}_{g,1}(k) \to \mathrm{Hom}\left(H, \frac{\pi_{k+1}}{\pi_{k+2}}\right)$$

are defined by means of:

$$\delta_k(f)(x) = f(\tilde{x})\tilde{x}^{-1}$$

for any $x \in H$ and $\tilde{x} \in \pi$ an arbitrary lift.

Right from the definition it follows that

$$\mathrm{Mod}_{g,1}(k+1) = \ker\left(\delta_k : \mathrm{Mod}_{g,1}(k) \to \mathrm{Hom}\left(H, \frac{\pi_{k+1}}{\pi_{k+2}}\right)\right)$$

As $\frac{\pi_{k+1}}{\pi_{k+2}}$ are abelian groups, we have isomorphisms induced by the symplectic duality:

$$\iota : \mathrm{Hom}\left(H, \frac{\pi_{k+1}}{\pi_{k+2}}\right) \cong \frac{\pi_{k+1}}{\pi_{k+2}} \otimes H$$

Note that the graded nilpotent series of a free group π has a well-known description in terms of the *free graded Lie algebra $\mathcal{L}$ on H*. Each element of H has degree one. In small degrees we have the isomorphisms

$$\mathcal{L}_2 \cong \wedge^2 H$$

which sends a commutator $[u, v]$ into $u \wedge v$ and

$$\mathcal{L}_3 \cong \frac{\wedge^2 H \otimes H}{\wedge^3 H}$$

which sends the triple commutator $[[u, v], w]$ into $u \wedge v \otimes w$. Note that the element $u \wedge v \wedge w$ corresponds to the Jacobi identity in the Lie algebra

$$[[u, v], w] + [[w, u], v] + [[v, w], u] = 0 \in \mathcal{L}_3$$

The extension of Johnson's homomorphism τ is then the homomorphism

$$\tau_k : \mathrm{Mod}_{g,1}(k) \to \mathcal{L}_k \otimes H$$

given by

$$\tau_k = \iota \circ \delta_k$$

We previously saw that B-twists are in $\mathrm{Mod}_{g,1}(3)$. The value of τ_3 on B-twists can be computed:

Proposition 10. *Let γ bounds a subsurface $S_{k,1}$ having the symplectic basis a_i, b_i, with $i \leq k$. Then*

$$\tau_3(T_\gamma) = \sum_{i,j=1}^{k} a_j \wedge b_j \otimes a_i \wedge b_i \in \frac{\wedge^2 H \otimes H}{\wedge^3 H} \otimes H$$

Proof. The computations from the previous section show that

$$T_\gamma(\alpha_i)\alpha_i^{-1} = [c, \alpha_i],$$

$$T_\gamma(\beta_i)\beta_i^{-1} = [c, \beta_i]$$

where c is the homotopy class of γ, namely $c = \prod_{i=1}^{k}[\alpha_i, \beta_i]$. The isomorphism $\frac{\pi_3}{\pi_4} \to \mathcal{L}_3$ sends

$$[c, \alpha_i] = [\prod_{j=1}^{k}[\alpha_j, \beta_j], \alpha_i] \to \sum_{j=1}^{k}[[a_j, b_j], a_i]$$

$$[c, \beta_i] = [\prod_{j=1}^{k}[\alpha_j, \beta_j], \beta_i] \to \sum_{j=1}^{k}[[a_j, b_j], b_i]$$

so that using the symplectic duality and the identification $\mathcal{L}_3 \cong \frac{\wedge^2 H \otimes H}{\wedge^3 H}$ we obtain

$$\tau_3(T_\gamma) = \sum_{i,j=1}^{k} a_j \wedge b_j \otimes a_i \otimes b_i - a_j \wedge b_j \otimes b_i \otimes a_i$$

A key property of Johnson's filtration is the fact that it is central:

Proposition 11 (Morita). *Johnson's subgroups form a central filtration, namely:*

$$[\mathrm{Mod}_{g,1}(k), \mathrm{Mod}_{g,1}(n)] \subseteq \mathrm{Mod}_{g,1}(k+n-1).$$

Proof. Let us show first that whenever $f \in \mathrm{Mod}_{g,1}(k)$ if the element $x \in \pi_n$, then $f(x)x^{-1} \in \pi_{k+n-1}$. This is clear for $n = 1$ and we now prove this for $n \geq 2$ by induction on n. Assume that the the claim holds until $n-1$. It suffices to consider elements x of the form $x = [y, z]$, where $y \in \pi_{n-1}$ and $z \in \pi$. By the induction hypothesis we can write

$$f(y)y^{-1} = a \in \pi_{k+n-2}, \text{ and } f(z)z^{-1} = b \in \pi_k.$$

Therefore

$$f(x)x^{-1} = [f(y), f(z)][z, y] = yazba^{-1}y^{-1}b^{-1}yz^{-1}y^{-1} =$$

$$= y[a, zb]z[b, y^{-1}]z^{-1}y^{-1} \in \pi_{k+n-1},$$

because $[a, zb], [b, y^{-1}] \in \pi_{k+n-1}$.

Choose now $f \in \mathrm{Mod}_{g,1}(k)$ and $g \in \mathrm{Mod}_{g,1}(n)$, where $k \geq n$. Given $x \in \pi$ we know that:

$$a = f(x)x^{-1} \in \pi_k \text{ and } b = g(x)x^{-1} \in \pi_n$$

Now

$$f^{-1}(x) = f^{-1}(a^{-1}) \cdot x \text{ and } g^{-1}(x) = g^{-1}(b^{-1}) \cdot x.$$

Therefore

$$[f, g](x) = fgf^{-1}g^{-1}(x) = fgf^{-1}\left(g^{-1}(b^{-1}) \cdot x\right) = fg\left(f^{-1}g^{-1}(b) \cdot f^{-1}(a^{-1}) \cdot x\right) =$$

$$= [f, g](b^{-1}) \cdot fgf^{-1}(a^{-1}) \cdot f(b) \cdot a \cdot x$$

It follows that we can write $[f,g](x)x^{-1}$ as the follwing product:

$$\big([f,g](b^{-1})\cdot b\big)\,\big(b^{-1}\cdot fgf^{-1}(a^{-1})\cdot ab\big)\,\big(b^{-1}a^{-1}\cdot f(b)\cdot b^{-1}ab\big)\cdot[b^{-1},a^{-1}]$$

Now $[f,g]\in \mathrm{Mod}_{g,1}(k)$ and $fgf^{-1}\in \mathrm{Mod}_{g,1}(n)$ since Johnson's groups are normal subgroups of $\mathrm{Mod}_{g,1}$. The claim from above implies that every element of the form

$$[f,g](b^{-1})\cdot b,\ fgf^{-1}(a^{-1}),\ f(b)b^{-1},\ [b^{-1},a^{-1}]$$

belongs to π_{k+n-1}, so in the product above we only encounter conjugates of elements of π_{n+k-1}, namely $[f,g]\in \mathrm{Mod}_{g,1}(k+n-1)$.

An immediate consequence is:

Corollary 16. *The Torelli group $\mathcal{I}_{g,1}$ is residually nilpotent.*

Proof. By Morita's Proposition 11 the k-th term of the lower central series of $\mathcal{I}_{g,1}$ is a subgroup of $\mathrm{Mod}_{g,1}(k+1)$. If f is a mapping class that belongs to all Johnson's subgroups then

$$f(x)x^{-1}\in\bigcap_{k=2}^{\infty}\pi_{k+1}$$

But the free group π is residually nilpotent, so that this intersection is trivial.

Exercise 16. Show that $\tau_1:\mathrm{Mod}_{g,1}\to H\otimes H$ is given by

$$\tau_1(T_\gamma)=c\otimes c$$

where c is the homology class of the simple closed curve γ with an arbitrary orientation.

Exercise 17. If α and β are simple closed curves on $S_{g,1}$ such that $\omega(a,b)=0$, then $[T_\alpha,T_\beta]\in\mathcal{I}_{g,1}$. Show that this commutator might be nontrivial, if α and β are not disjoint.

7

Mapping class group representations

7.1 Symplectic representation

7.2 Homological representations

7.3 Koberda's asymptotic faithfulness

7.4 Hadari's detection of infinite order

7.5 Compact representations: Aramayona-Souto

7.6 Finite index subgroups

7.7 Finite quotients of the mapping class groups, Dunfiedl-Thurston, BooRabee-Leininger

8

Nielsen-Thurston's classification and actions

8.1 Complexes associated to mapping class groups

8.2 The curve complex

8.3 The Masur-Minsky theorem

8.4 Arc complexes after Harer and Hatcher

9

Fuchsian groups

The aim of this chapter is to collect the basic notions and essential results in Fuchsian groups needed later. Most proofs will only be sketched. There are a number of excellent introductions to hyperbolic geometry, including the following standard references:

References

1. A. Beardon, *The geometry of discrete groups*, Springer, 1983.
2. Benedetti, C. Petronio, *Lectures on hyperbolic geometry*, Universitext, Springer 1992.
3. S. Katok, *Fuchsian groups*, Univ. Chicago Press, 1992.
4. W.Thurston, *Three-dimensional geometry and topology*, Vol. 1., Ed. Silvio Levy, Princeton Mathematical Series, 35, 1997.

9.1 The geometry of the hyperbolic plane

9.1.1 The upper half-plane

We denote by $\mathbb{C}$ the field of complex numbers, which can be identified with the set of points of the plane. If $z \in \mathbb{C}$ we also set $\operatorname{Im} z$ for the imaginary part of the complex number z.

The most common model for the hyperbolic plane is the upper half-plane

$$\mathcal{H} = \{z \in \mathbb{C}\ ;\ \operatorname{Im}(z) > 0\},$$

endowed with the Riemannian metric $ds^2 = \frac{dx^2+dy^2}{y^2}$.

If $\mathrm{SL}(2,\mathbb{R})$ denotes the Lie group of 2×2 matrices of determinant 1 then one sets $\mathrm{PSL}(2,\mathbb{R}) = \mathrm{SL}(2,\mathbb{R})/\{\pm 1\}$. There is a natural action of $\mathrm{SL}(2,\mathbb{R})$ on the Riemann sphere $\mathbb{C}\cup\{\infty\}$ by means of so-called Möbius transformations. Given $z \in \mathbb{C}$ and $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{R})$, the action is defined by the following formula:

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

This action factors through an action of $\mathrm{PSL}(2,\mathbb{R})$ on the upper half-plane $\mathcal{H}$. Indeed, if we set $w = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z$, then by direct calculation we obtain:

$$\operatorname{Im} w = \operatorname{Im} z \cdot \frac{1}{(cz+d)^2}$$

and hence $\mathcal{H}$ is invariant by $\mathrm{PSL}(2,\mathbb{R})$. Moreover, $\mathrm{PSL}(2,\mathbb{R})$ acts by homeomorphisms on $\mathcal{H}$. The half-plane is naturally endowed with a geometry whose group of isometries is $\mathrm{PSL}(2,\mathbb{R})$. To be more precise, we have:

Proposition 12. *i) $\mathcal{H}$ is a complete Riemannian manifold of constant curvature -1, when endowed with the Riemannian metric $ds^2 = \frac{dx^2+dy^2}{y^2}$, which will be called the hyperbolic plane.*

ii) The group $\mathrm{Isom}^+(\mathcal{H})$ of orientation preserving isometries of the hyperbolic plane is $\mathrm{PSL}(2,\mathbb{R})$, acting by Möbius transformations.

iii) The geodesics of the hyperbolic plane $\mathcal{H}$ are semi-circles (half-lines) orthogonal to the real axis.

iv) The group $\mathrm{PSL}(2,\mathbb{R})$ maps geodesics into geodesics.

v) The hyperbolic distance $d(z,w)$ between two points $z, w \in \mathcal{H}$ is given by the formulas:

$$d(z,w) = \log \frac{|z - |w|| + |z-w|}{|z-|w|| - |z-w|} = \log[w, z^*, z, w^*]$$

where the geodesic joining z to w intersects the real axis into z^* and w^*. The cross ratio of four points of $\mathbb{C}$ is defined as

$$[z_1, z_2, z_3, z_4] = \frac{(z_1-z_2)(z_3-z_4)}{(z_2-z_3)(z_1-z_4)}.$$

Proof. We see that $\mathrm{PSL}(2,\mathbb{R}) \subset \mathrm{Isom}(\mathcal{H})$ by a direct computation of the pull-back of the Riemannian metric. To show that geodesics of the hyperbolic metric are semi-circles one computes a lower bound for the Riemannian length $l(\gamma)$ of a curve γ which joins $\gamma(0) = ia$ to $\gamma(1) = ib$, for $i = \sqrt{-1}$, where $a, b \in \mathbb{R}_+$:

$$l(\gamma) \geq \int_0^1 \frac{\dot{y}}{y}\, dt = \log \frac{b}{a}$$

Equality is only attained when γ is the vertical segment of half-line joining the two points ia and ib. Further $\mathrm{PSL}(2,\mathbb{R})$ sends semi-circles into half-lines orthogonal to the real axis and in meantime it acts transitively on pairs of points in the half-plane, thus a geodesic joining two arbitrary points is the image of a vertical segment by such a Möbius transformation. Eventually observe that the cross ratio is invariant by $\mathrm{PSL}(2,\mathbb{R})$. The cross ratio formula for the distance follows by using an element of $\mathrm{PSL}(2,\mathbb{R})$ which sends z and w onto two points on the vertical axis, so that $z^* = 0$ and $w^* = \infty$.

Remark 21. The full group of isometries of the upper half-plane is $\mathrm{Isom}(\mathcal{H}) = \mathrm{SL}^*(2,\mathbb{R})/\{\pm 1\}$ where $\mathrm{SL}^*(2,\mathbb{R}) = \{A \in \mathrm{GL}(2,\mathbb{R})\ ;\ \det A \in \{\pm 1\}\}$.

We often use the term metric both for the Riemannian metric ds^2 and for the hyperbolic distance d induces from it.

9.1.2 The unit disk

Another well-known model for the hyperbolic geometry is the unit disk model $D = \{z \in \mathbb{C}\ ;\ |z| < 1\}$, with the metric $dr^2 = \frac{2(dx^2+dy^2)}{1-(x^2+y^2)}$, where we write $z = x + iy$. The map $f : \mathcal{H} \to \mathcal{D}$ given by $f(z) = \frac{z \cdot i+1}{z+i}$ is therefore an isometry. The geodesics are semi-circles orthogonal to the boundary circle (which is also called the principal circle).

Remark 22. 1. The action of $\mathrm{PSL}(2,\mathbb{R})$ on $\mathcal{D}$ can be deduced from above:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} w = \frac{w(a+d+(b-c)i)+(b+c+(a-d)i)}{w(b+c-(a-d)i)+((a+d)-(b-c)i)} = \frac{w\alpha+\bar{\beta}}{w\beta+\bar{\alpha}}$$

where $\alpha, \beta \in \mathbb{C}$, $\alpha\bar{\alpha} - \beta\bar{\beta} = 1$.
2. The $\mathrm{PSL}(2,\mathbb{R})$ action on $\mathcal{D}$ extends continuously to an action on the boundary circle, hence to the closed 2-disk $\overline{\mathcal{D}}$.

The closed disk $\overline{\mathcal{D}}$ is also called the compactified hyperbolic plane and the boundary circle $\partial\overline{\mathcal{D}}$ is called the *circle at infinity*. The corresponding compactification of $\mathcal{H}$ is the union $\overline{\mathcal{H}} = \mathcal{H} \cup \mathbb{R} \cup \{\infty\}$ of the hyperbolic plane with the compactified real axis, also called circle at infinity. The map f extends to a homeomorphism $\overline{\mathcal{H}} \to \overline{\mathcal{D}}$.

9.2 PSL(2, ℝ) and Fuchsian groups

9.2.1 Classification of isometries

Set $\mathrm{Tr} : \mathrm{PSL}(2,\mathbb{R}) \to \mathbb{R}_+$ for the function $\mathrm{Tr}(A) = |\,\mathrm{tr}(\widetilde{A})|$ where $\widetilde{A}$ is an arbitrary lift of A in $\mathrm{SL}(2,\mathbb{R})$. The elements of $\mathrm{PSL}(2,\mathbb{R})$ fall in one of the following three disjoint classes:

1. *elliptic* elements, when $\mathrm{Tr}\,A < 2$. Every elliptic element is conjugate to a unique matrix of the form $\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$, for some $\theta \in (0, 2\pi)$.
2. *hyperbolic* elements, if $\mathrm{Tr}\,A > 2$. A hyperbolic element is conjugate to a unique matrix $\begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}$, where $\lambda \neq 1$, $\lambda \in \mathbb{R}$, hence it is diagonalizable over $\mathbb{R}$.

3. *parabolic* elements along with the identity, if $\mathrm{Tr}\, A = 2$. A parabolic element is conjugate to either a positive or a negative translation along the real axis.

This classification is directly related to the dynamics of isometries on the hyperbolic plane. An elementary calculus yields the following characterization in terms of the fixed points solutions:

Proposition 13. *A nontrivial isometry $A \in \mathrm{PSL}(2,\mathbb{R})$ of $\mathcal{D}$ is:*

1. *elliptic iff it has a unique fixed point in $\mathcal{D}$.*
2. *hyperbolic iff it has exactly two fixed points in $\overline{\mathcal{D}}$ which belong to $\partial\overline{\mathcal{D}}$.*
3. *parabolic iff it has a unique fixed point on the circle $\partial\overline{\mathcal{D}}$.*

Conjugacy classes in $\mathrm{SL}(2,\mathbb{R})$ and $\mathrm{PSL}(2,\mathbb{R})$ are essentially determined by the trace and Tr, respectively. This statement is genuinely true for both hyperbolic and elliptic transformations. Moreover, this holds up to a sign ambiguity for the parabolic isometries. Note that parabolics are always conjugate within the larger group $\mathrm{PSL}^*(2,\mathbb{R}) = \mathrm{Isom}(\mathcal{H})$.

The differential of an isometry of $\mathcal{H}$ is a local diffeomorphism of the of the unit tangent bundle $S\mathcal{H}$ of the tangent bundle $T\mathcal{H}$. Therefore, the group of isometries $\mathrm{PSL}(2,\mathbb{R})$ has an induced action on the unit tangent bundle $S\mathcal{H}$, given by the formula:

$$A(z,v) = (Az, dA(v)), \text{ for } z \in \mathcal{H}, v \in T_z\mathcal{H}.$$

Observe that for any couple γ_1, γ_2 of geodesics in $\mathcal{H}$ and points $z_i \in \gamma_i$, $i = 1, 2$, on them there exists an isometry $A \in \mathrm{PSL}(2,\mathbb{R})$ such that $A(\gamma_1) = \gamma_2$ and $Az_1 = z_2$. This action being simply transitive, we can identify $\mathrm{PSL}(2,\mathbb{R})$ and $S\mathcal{H}$. In particular, $\mathrm{PSL}(2,\mathbb{R})$ is diffeomorphic to an open solid torus.

We can identify a geodesic in $\mathcal{H}$ with a pair of disjoint points on the circle $\partial\overline{D}$ corresponding to its intersection with the circle at infinity. These points are also the fixed points of the hyperbolic isometry which translates a nontrivial amount along the geodesic. We then have an identification of the space of geodesics with the space of pairs of distinct points on the circle.

9.2.2 Fuchsian groups

The group $\mathrm{SL}(2,\mathbb{R})$ inherits a topology when seen as a subset of the Euclidean space $\mathrm{SL}(2,\mathbb{R}) \subset \mathbb{R}^4$ by means of the obvious inclusion sending the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ to the point (a, b, c, d). Furthermore, the $\mathbb{Z}/2$-action $(a, b, c, d) \longrightarrow (-a, -b, -c, -d)$ identifies $\mathrm{PSL}(2,\mathbb{R})$ to a quotient of the topological space $\mathrm{SL}(2,\mathbb{R})$, which is then endowed with the quotient topology. Notice that $\mathrm{PSL}(2,\mathbb{R})$ is a topological group since the topology defined above is compatible with the group structure.

Definition 19. *The subgroup* $\Gamma \subset \mathrm{PSL}(2,\mathbb{R})$ *is* discrete *if* Γ *is a discrete set in the topological space* $\mathrm{PSL}(2,\mathbb{R})$. Discrete *subgroups of* $\mathrm{PSL}(2,\mathbb{R})$ *are usually called* Fuchsian *groups.*

Definition 20. *Let* G *be a group of homeomorphisms acting on the topological space* X. *Then* G acts properly discontinuously *an* X *if the* G*-orbit of any* $x \in X$, *i.e.* $Gx = \{gx; g \in G\}$, *is a* locally finite **family***.*

Remark 23. 1. One required that the family of orbits be a locally finite family and not only a locally finite set. Recall that the family $\{M_\alpha\}_{\alpha \in \mathcal{J}}$ is called *locally finite* if for any compact subset $K \subset X$ we have $M_\alpha \cap K \neq \emptyset$ only for finitely many values of $\alpha \in A$.
2. In Gx each point is contained with a *multiplicity* equal to the order of the stabilizer $G_x = \{g \in G; gx = x\}$.
3. In particular G acts properly discontinuously iff each orbit *is discrete* (as a set this time) and the *order of the stabilizer is finite*.
4. G acts properly discontinuously iff for all $x \in X$, there exists a neighborhood $V \ni x$, $V \subset X$ such that $\mathcal{J}(V) \cap V \neq \emptyset$ for only finitely many $g \in G$.

The remarks above imply immediately that:

Proposition 14. *The subgroup* $\Gamma \subset \mathrm{PSL}(2,\mathbb{R})$ *is Fuchsian iff it acts properly discontinuously on* $\mathcal{H}$.

Corollary 17. *1. The set of fixed points of elliptic elements of a Fuchsian group* Γ *form a discrete set in* $\mathcal{H}$.
2. If moreover the Fuchsian group Γ *does not contain elliptic elements, then* $\mathcal{H}/\Gamma$ *is a complete connected orientable Riemannian 2-manifold of curvature* -1.

Remark 24. In general, a discrete group might well act non-discontinuously on a topological space. For example, the action of the discrete subgroup $\mathrm{PSL}(2,\mathbb{Z})$ of $\mathrm{PSL}(2,\mathbb{R})$ on the boundary circle $\partial\overline{\mathcal{D}}$, which is induced by the action of $\mathrm{PSL}(2,\mathbb{R})$ on the boundary, has an orbit equal to $\mathbb{Q} \cup \{\infty\}$ which is dense in $\mathbb{R} \cup \{\infty\}$.

Exercise 18. Show that $\mathrm{PSL}(2,\mathbb{Z})$ is a Fuchsian group, while $\mathrm{PSL}(2,\mathbb{Z}[\sqrt{2}])$ is *not* a Fuchsian group.

9.3 Discreteness of subgroups $\Gamma \subset \mathrm{PSL}(2,\mathbb{R})$

Definition 21. *The limit set* $\lambda(\Gamma) \subset \overline{\mathcal{H}}$ *of the subgroup* $\Gamma \subset \mathrm{PSL}(2,\mathbb{R})$ *is the set ofaccumulation points of* Γ*-orbits* Γz, *for* $z \in \mathcal{H}$.

Definition 22. *The subgroup* $\Gamma \subset \mathrm{PSL}(2,\mathbb{R})$ *is* elementary *if there exists a finite* Γ*-orbit in* $\overline{\mathcal{H}}$. *Equivalently, we have the equality* $\mathrm{Tr}[A,B] = 2$, *whenever* $A, B \in \Gamma$ *have infinite order.*

Elementary Fuchsian groups can be characterized completely. It is known that:

Proposition 15. *An* elementary Fuchsian *group is either cyclic or it is conjugate within* $\mathrm{PSL}(2,\mathbb{R})$ *to a subgroup of the group generated by the two transformations g and j below:*

$$g(z) = \lambda z, \quad \lambda > 1, \quad \text{and} \quad j(z) = -\frac{1}{z}.$$

Proposition 16. *If $\Gamma \subset \mathrm{PSL}(2,\mathbb{R})$ contains only elliptic elements (or the identity $\mathbb{1}$) then Γ is cyclic elementary.*

Proposition 17. *Let Γ be a* non-elementary *subgroup of* $\mathrm{PSL}(2,\mathbb{R})$. *Then Γ is* discrete *iff the following equivalent conditions are satisfied:*

1. *The fixed points of elliptic elements do not accumulate on $\mathcal{H}$.*
2. *The elliptic elements do not accumulate on $\mathbb{1}$.*
3. *Each elliptic element has finite order.*

Proposition 18. *Let $\Gamma \subset \mathrm{PSL}(2,\mathbb{R})$ be non-elementary. Then:*

1. *Γ contains hyperbolic elements.*
2. *If Γ does not contain elliptic elements then it is discrete.*
3. *Γ is discrete iff every cyclic subgroup of Γ is discrete.*

The proof of the third assertion above is due to T.Jorgensen. He used the following inequality valid actually more generally in $\mathrm{PSL}(2,\mathbb{C})$:

$$|\operatorname{tr}^2(A) - 2| + |\operatorname{tr}([A,B]) - 2| \geq 1$$

whenever the subgroup $\langle A, B\rangle \subset \mathrm{PSL}(2,\mathbb{C})$ generated by A and B is discrete. Observe that both $\operatorname{tr}^2(A)$ and $\operatorname{tr}([A,B])$ are well-defined when $A, B \in \mathrm{PSL}(2,\mathbb{C})$.

Notice that a subgroup $\Gamma \subset \mathrm{PSL}(2,\mathbb{C})$ is discrete if and only if every two generators subgroup of Γ is discrete. Moreover, this time the discreteness of its cyclic subgroups is not sufficient.

Another proof of the third part is due to Rosenberger [2], who showed that if we have a discrete subgroup $\Gamma \subset \mathrm{PSL}(2,\mathbb{R})$ which is not isomorphic to $\mathbb{Z}/2 * \mathbb{Z}/2$ then

$$|\operatorname{Tr}([A,B]) - 2| \geq 2 - 2\cos\frac{\pi}{7}$$

for any $A, B \in \Gamma$.

The *discreteness* of a two generator subgroup $\langle A, B\rangle$ can be algorithmically and effectively decided, as it was proved by J.Gilman (see [1]). By instance if A, B are hyperbolic elements with intersecting distinct axes and $\gamma^2 = [A,B]$, $A = E_1E_2$, $B = E_1E_3$, E_i having order 2, $\gamma = E_1E_2E_3$ and G denotes the subgroup $\langle A, B\rangle$, then we have:

- if A, B, γ are primitive (i.e. hyperbolic, parabolic or elliptic of finite order) then G is discrete.
- if $[A, B]$ is elliptic of infinite order then G is not discrete.
- if $\operatorname{tr}[A, B] = -2\cos\frac{2k\pi}{n}$, $1 \leq k < \frac{n}{2}$,
  - if $k \notin \{2, 3\}$ then G is not discrete.
  - if $k \in \{2, 3\}$ then we have special triangular groups G which might be discrete, and the situation is completely understood (see [1]).

Exercise 19. 1. If Γ is the cyclic group generated by the homothety $z \to \lambda z$, with $\lambda > 1$ then $\lambda(\Gamma) = \{0, \infty\}$.
2. $\Gamma = \mathrm{PSL}(2, \mathbb{Z})$ then $\lambda(\Gamma) = \mathbb{R} \cup \{\infty\}$.
3. If Γ is Fuchsian then $\lambda(\Gamma) \subset \mathbb{R} \cup \{\infty\}$.

References

1. J.Gilman, *Two-generator discrete subgroups of PSL$(2, R)$*, Mem. Amer. Math. Soc. 117(1995), no. 561, 204 pp.
2. G.Rosenberger, *Eine Bemerkung zu einer Arbeit von T. Jorgensen*, Math. Zeitschrift 165 (1979), no. 3, 261–266.

9.4 Freeness of subgroups of PSL$(2, \mathbb{R})$ and PSL$(2, \mathbb{C})$

An important problem is to understand the algebraic structure of a group generated by a family of explicit matrices. The simplest instance of this problem is the case of the so-called *real Hecke groups* $\Gamma_\lambda \subset \mathrm{PSL}(2, \mathbb{R})$, which is generated by the two matrices

$$A = \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}, \ B = \begin{pmatrix} 1 & 0 \\ \lambda & 1 \end{pmatrix}, \ \lambda > 1.$$

The first result in this direction was obtained by I.N.Sanov in 1947 and by J.L.Brenner in 1955:

Proposition 19 (Sanov 1947, Brenner 1955). *The real Hecke groups Γ_λ are free, provided that $|\lambda| \geq 2$.*

The definition of Hecke groups extends readily to all complex values of the parameter $\lambda \in \mathbb{C}$. We call λ *free* if the (complex) Hecke group Γ_λ is free. A description of all free λ is still unknown today, although a large number of results are known. The current state of this question is summarized in the following:

Proposition 20. *In the follo wing cases λ is free:*

1. *λ is outside the unit disks centered at -1, 0 and 1 in $\mathbb{C}$.*
2. *λ is outside the disks radius $\frac{1}{2}$ centered at $-\frac{i}{2}$ and $\frac{i}{2}$ and outside the unit open disks centered at -1 and 1.*

3. λ *lies outside the convex hull of the open disk radius* 1 *at origin and the points* ± 2.
4. $|\lambda| > 1$ *and* $|\operatorname{Im}(\lambda)| \geq \frac{1}{2}$.
5. $|\lambda - 1| > \frac{1}{2}$ *and* $1 \leq |\Re\lambda| < \frac{5}{4}$, *where* $\Re$ *denotes the real part and* Im *the imaginary part.*
6. λ *is* transcendental.

The nature of the set of free λ seems to be fractal, according to results obtained by Beardon and Bamberg:

Proposition 21 (Beardon 1993, Bamberg 2000).

1. *The set of algebraic complex numbers* λ *which are free is dense in* $\mathbb{C}$.
2. *There exist sets* S_1, S_2 *consisting of algebraic irrational complex numbers so that* $\overline{S_1} = [-2, 2]$, $\overline{S_2} = [-i, i]$ *and* S_i *consist only of non-free points.*
3. *The points* $\frac{1}{2}e^{2\pi i/k}$, $\frac{1}{k}$, $\frac{k}{k^2+1}$, *for* $k \in \mathbb{Z}$, $\frac{1}{\sqrt{2}}$, $\frac{9}{50}$, $\frac{8}{25}$, $\frac{25}{98}$ *are non-free.*
4. *If* $p^2 - Nq^2 = 1$, *where* p, q *are integers and* N *is a positive integer which is not a perfect square then* $\lambda = \frac{p}{q}$ *is non-free. In fact, we have*
$$W = A^{q^2} B^N A^{-1} = \begin{pmatrix} * & 0 \\ * & * \end{pmatrix}$$
and so there is a nontrivial relation satisfied in Γ_λ *reading:*
$$[W^{-1}BW, B] = 1$$

Remark 25. There exists a family of free subgroups, called Schottky groups, in $\mathrm{PSL}(2, \mathbb{C})$ generated by transformations $\gamma_1 \ldots \gamma_g \in \mathrm{PSL}(2, \mathbb{C})$ with the property that $\gamma_i(\mathrm{int} D_{2i}) = \widehat{\mathbb{C}} \setminus D_{2i+1}$, where D_j are fixed disjoints disks in the plane and $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ is the Riemann sphere compactification of $\mathbb{C}$. Then the ping-pong Lemma shows that this group is free.

Exercise 20. The *Hecke groups* are generated by
$$\Gamma_\lambda = \left\langle \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \right\rangle$$
for the values of the parameter $\lambda = 2\cos\frac{\pi}{2}$ have different behaviour than those for which $|\lambda| \geq 2$ (which are free groups, by the results above). In fact, one has
$$\Gamma_{2\cos\frac{\pi}{q}} = \mathbb{Z}/q\mathbb{Z} * \mathbb{Z}/2\mathbb{Z}$$
and in particular, if $q = 3$ we obtain $\Gamma_3 = \mathrm{PSL}(2, \mathbb{Z})$.

References

1. J.Bamberg, *Non-free points for groups generated by a pair of* 2×2 *matrices* J. London Math. Soc. (2) 62 (2000), no. 3, 795–801.
2. A.F.Beardon, *Pell's equation and two generator free Möbius groups*, Bull. London Math. Soc. 25 (1993), 527–532.

9.5 Arithmetic groups in $\mathbf{PSL(2,\mathbb{R})}$

Definition 23. *The subgroup $\Gamma \subset \mathrm{PSL}(2,\mathbb{R})$ is called* arithmetic *if Γ is a subgroup of finite index in some group $\varphi^{-1}(\mathrm{GL}(n,\mathbb{Z}))$ arising as the inverse image of some finite dimensional representation*

$$\varphi : \Gamma \longrightarrow \mathrm{GL}(n,\mathbb{R})$$

Proposition 22. *An arithmetic Fuchsian group is commensurable with a Fuchsian group determined by a quaternion algebra (which are either 2×2 matrix algebras over some field or else division algebras). Moreover, all these Fuchsian arithmetic groups are either cocompact or commensurable with* $\mathrm{PSL}(2,\mathbb{Z})$.

Remark 26. If $q \notin \{3,4,6\}$ then the Hecke groups are *not* arithmetic. In fact the Hecke groups are of finite covolume, not cocompact and not commensurable with $\mathrm{PSL}(2,\mathbb{Z})$. For instance, if we consider the element $\gamma = \begin{pmatrix} -2\cos\frac{\pi}{5} & 1 \\ -1 & 0 \end{pmatrix} \in \Gamma_{\cos\frac{2\pi}{5}}$ then $\mathrm{Tr}\,\gamma^n \notin \mathbb{Q}$ for any $n \in \mathbb{Z} \setminus \{0\}$. This implies that no finite index subgroup of $\Gamma_{\cos\frac{2\pi}{5}}$ could be contained in $\mathrm{PSL}(2,\mathbb{Z})$. Thus the group is not arithmetic by the criterion provided above.

9.6 Fundamental regions and Siegel's Theorem

9.6.1 Fundamental regions for Fuchsian groups

Definition 24. *We say that a closed region $F \subset \mathcal{H}$ is a fundamental region for the Fuchsian group Γ if*

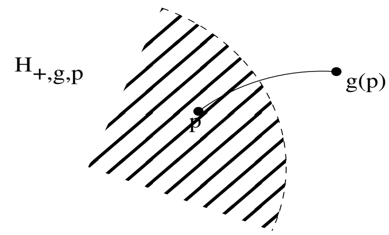
1. $\bigcup_{g\in\Gamma} gF = \mathcal{H}$
2. $\mathrm{int}(X) \cap g(\mathrm{int}(F)) = \emptyset$*, for all $g \in \Gamma$, $g \neq 1$.*

Remark 27. Let μ denote the Lebesque measure on $\mathcal{H}$. If F, F' are both fundamental regions for the group Γ and $\mu(F) < \infty$, then $\mu(F') = \mu(F)$.

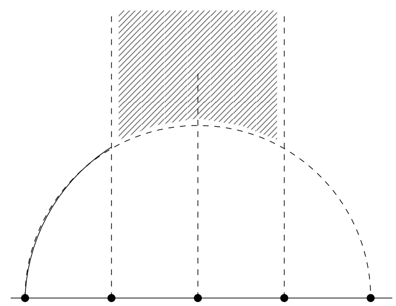
Let us provide examples of fundamental regions. The *Dirichlet fundamental region* associated to a Fuchsian group Γ is constructed as follows:

1. Pick up $p \in \mathcal{H}$ such that p is not fixed by any $g \in \Gamma \setminus \{0\}$.
2. Set then $D(\Gamma,p) = \{z \in \mathcal{H}; d(p,z) \leq d(z,g(p)),\ \text{for all } g \in \Gamma\}$.

Then $D(\Gamma,p)$ is a fundamental region for Γ, called the Dirichlet region. Alternatively let $H_{+,p,g}$ be the half-space containing p and having as boundary the bisector of the segment of geodesic which joins p to $g(p)$. Then $D(\Gamma,p) = \bigcap_{g\in\Gamma\setminus\{0\}} H_{+,p,g}$. In particular, $D(\Gamma,p)$ is geodesically convex, namely it contains the geodesic segments between any two points of it. It is easy to prove that $D(\Gamma,p)$ is a connected fundamental region for Γ.

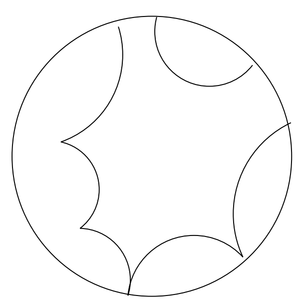


Example 1. Let $\Gamma = \mathrm{PSL}(2,\mathbb{Z})$ and pick-up $p = \lambda i$, where $\lambda > 1$. Thus p is not fixed by any element of Γ. Then $\mathcal{D}(\mathrm{PSL}(2,\mathbb{Z})), \lambda_i) = \{z \in \mathcal{H}; |z| \geq 1, |\Re(z)| \leq \frac{1}{2}\}$



Note that the tesselation $\{g(D(\Gamma,p)),\ g \in \Gamma\}$ is locally finite since Γ acts properly discontinuously. Furthermore, the quotient $\mathcal{H}/\Gamma$ is homeomorphic then to F/Γ for any locally finite fundamental region F of Γ. Moreover F/Γ is obtained from F by using self-identifications of the boundary arcs, hence we can expect a complete topological description of $\mathcal{H}/\Gamma$.

Consider now a Fuchsian group Γ and a fundamental region F for Γ.



Two points of the fundamental region F are said to be *congruent* if they are equivalent under the action of Γ. They should therefore belong to the boundary ∂F of the fundamental region. We have then the following type of vertices of the curved polygon F:

- *elliptic* vertices of F corresponding to fixed points of elliptic elements;
- *parabolic* vertices of F corresponding to fixed points of parabolic elements.

Given a congruence class we consider the Γ-orbit of one vertex in that class and obtain a cycle. Two cycles are then different if they correspond to different classes of vertices modulo Γ. The cycles are said to be elliptic or parabolic according to their vertices.

Example 2. In the picture above x_1, x_2 are elliptic and x_3, x_4 are parabolic.

Remark 28.
1. Elliptic cycles correspond to conjugacy classes of nontrivial maximal finite cyclic subgroups of Γ, while parabolic cycles correspond to maximal (cyclic) parabolic subgroups.
2. If $\theta_1 \dots \theta_t$ are the internal angles of the polygon F at the vertices $x_1 \dots x_t$ which are congruent to each other in F and form a cycle, then $\theta_1 + \cdots + \theta_t = \frac{2\pi}{m}$, where m is the order of the stabilizer in Γ of any of these vertices.

9.6.2 Siegel's Theorem on geometrically finite Fuchsian groups

Definition 25. *The subgroup Γ is* geometrically finite *if there exists a convex fundamental region of Γ with finitely many sides.*

Let μ denote the Lebesgue measure on the quotient $\mathcal{H}/\Gamma$ by a Fuchsian group, induced from that of $\mathcal{H}$. We have the following characterization of geometric finiteness due to Siegel:

Theorem 29 (Siegel). *If $\mu(\mathcal{H}/\Gamma) < \infty$ then Γ is geometrically finite. Actually, for any fundamental region F we have the more precise estimation:*

$$\sum_{v \text{ vertex of } F} (\pi - \theta_v) \leq \mu(F) + 2\pi$$

where θ_v denotes the internal angle of the polygon F at the vertex v.

Remark 29.
1. If Γ has a compact fundamental region then Γ has non parabolic elements.
2. Some (equivalently, any) fundamental region of Γ is noncompact iff the quotient $\mathcal{H}/\Gamma$ is noncompact.
3. Moreover, if $\mu(\mathcal{H}/\Gamma) < \infty$ but $\mathcal{H}/\Gamma$ is noncompact then there exist vertices at infinity in any fundamental region for Γ, which are parabolic vertices.
4. If some fundamental region F for Γ is compact then all fundamental regions of Γ are compact; further, Γ is cocompact iff $\mu(\mathcal{H}/\Gamma) < \infty$ and Γ has no parabolics.

Definition 26. *The orders of elliptic elements fixing the elliptic vertices (in each congruence class) are called the periods of the Fuchsian group. One can add ∞ as period of each parabolic vertex.*

Example 3. $\mathrm{PSL}(2, \mathbb{Z})$ has periods $2, 3$ and ∞.

Remark 30. Let T_i be elements of Γ which are pairing the sides of the fundamental region F. Then the set of elements $\{T_i\}$ generate the group Γ. In particular if F has finitely many sides (e.g. when Γ is geometrically finite) then Γ is finitely generated.

Proposition 23. *1. If the Fuchsian group Γ is finitely generated then it is geometrically finite.*

2. Moreover, if the Fuchsian group Γ is cocompact and $\mathcal{H}/\Gamma$ is a surface of genus g with r singular points corresponding to the periods $m_1 \dots m_r$ then we have the following Siegel formula:

$$\mu(\mathcal{H}/\Gamma) = 2\pi \left(2g - 2 + \sum_{i=1}^{r} \left(1 - \frac{1}{m_i} \right) \right)$$

3. For any $(g, m_1 \dots m_r)$ with $2q - 2t + \sum_{i=1}^{r} 1 - \frac{1}{m_i} > 0$, $g \geq 0$, $r \geq 0$, $m_i \geq 2$, there exists a Fuchsian group of signature $(g, m_1 \dots m_r)$, i.e. such that $\mathcal{H}/\Gamma$ is a surface of genus g with r singular points corresponding to the periods $m_1 \dots m_r$.

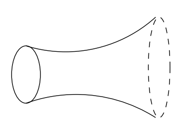
Remark 31. Notice that there exist subgroups of $\mathrm{GL}(2, \mathbb{R})$ which are not finitely generated, e.g.

$$\Gamma = \left\{ \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \;;\quad a = 2^{\alpha},\ b = \frac{p}{2^q},\ \alpha, p, q \in \mathbb{Z} \right\}$$

Remark 32. If Γ is not cocompact then one should consider the Nielsen core $K(\Gamma)$ which is the convex hull of $\lambda(\Gamma)$. Then Γ is finitely generated iff $\mu(K(\Gamma)/\Gamma) < \infty$, which is equivalent to have a Dirichlet region with finitely many sides. Furthermore, Siegel's formula above extends to the finitely generated situation as follows:

$$\mu(K(\Gamma)/\Gamma) = 2\pi \left(2g - 2 + t + \sum_{i=1}^{r} \left(1 - \frac{1}{m_i} \right) \right)$$

where $t = p + b$, where p denotes the number of parabolic cycles and b the number of boundary hyperbolics; one remarks that $K(\Gamma)/\Gamma$ is a surface with p punctures and b boundary curves, which is a core for the non-compact surface $\mathcal{H}/\Gamma$. The surface $\mathcal{H}/\Gamma$ has infinite area, but if we cut open the components corresponding to boundary hyperbolics, which have the shape



we obtain a cusped surface with boundary of finite area.

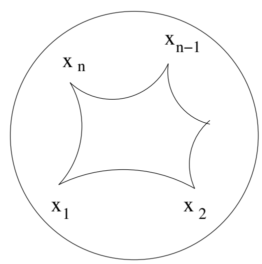
9.7 The Poincaré theorem: construction of Fuchsian groups

9.7.1 The finite case

Polygons in the hyperbolic plane will always have geodesic sides. Consider P a compact convex polygon in $\mathcal{D}$ with $n \geq 3$ vertices $x_1, x_2 \ldots x_n$ such that the interior angle at x_j is $\frac{\pi}{p_j}$, for some $p_j \in \mathbb{Z}_+$. A polygon with these angles exists in the hyperbolic plane if and only if (by the Gauss-Bonnet theorem) the following condition is fulfilled:

$$\sum_{j=1}^{n} \frac{1}{p_j} < n - 2.$$

Remark that there exists a polygon P with these prescribed angles in the Euclidean plane $\mathbb{E}^2$ iff $\sum_{j=1}^{n} \frac{1}{p_j} = n - 2$, while the existence of a spherical polygon P on the sphere S^2 is equivalent to $\sum_{j=1}^{n-2} \frac{1}{p_j} > n - 2$. We will stick to the hyperbolic case.



The reflection σ with respect to a geodesic γ is the unique nontrivial isometry of $\mathcal{D}$ fixing γ pointwise. It is therefore sending $p \in \mathcal{D} \setminus \gamma$ into the point $\sigma(p)$ at the same distance from γ on the other side of the half-plane determined by γ and such that the geodesic $p\sigma(p)$ is orthogonal to γ. Note that σ is orientation reversing.

Set σ_j for the reflection of $\mathcal{D}$ with respect to the side $x_j x_{j+1}$ and $\Gamma \subset \mathrm{PSL}^*(2, \mathbb{R})$ denote the group of isometries generated by $\{\sigma_1, \ldots, \sigma_n\}$.

Proposition 24. *The group Γ admits the following group presentation: by means of the generators $\sigma_1, \ldots, \sigma_n$ and relations:*

$$\Gamma = \langle \sigma_1, \ldots, \sigma_n | \, \sigma_j^2 = 1, \ (\sigma_{j-1}\sigma_j)^{p_j} = 1, \ j = 1, n \rangle$$

Moreover Γ acts properly on $\mathcal{D}$ and P is a fundamental region for Γ.

Remark that Γ is a *hyperbolic Coxeter group.*

Example 4. If $2 \leq a \leq b \leq c \in \mathbb{Z}$ so that $\frac{1}{a} + \frac{1}{b} + \frac{1}{c} < 1$, then the group

$$T^*_{a,b,c} = \langle x, y, z \mid x^2 = y^2 = z^2 = (yz)^a = (zx)^b = (xy)^c = 1 \rangle$$

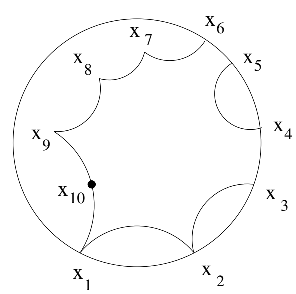
acts on $\mathcal{H}$ and has a fundamental domain given by a triangle whose angles are $\frac{\pi}{a}$, $\frac{\pi}{b}$, $\frac{\pi}{c}$. In particular, the area of the fundamental domain is $\pi(1 - \frac{1}{a} - \frac{1}{b} - \frac{1}{c})$. The index 2 subgroup $T_{a,b,c} \subset T^*_{a,b,c}$ consisting of orientation preserving isometries has the presentation

$$T_{a,b,c} = \langle u, v \mid u^c = v^a = (uv)^b = 1 \rangle, \ u = xy, \ v = yz$$

and a fundamental domain constructed by gluing two adjacent triangles with a common edge. The area of this quadrilateral is then $2\pi(1 - \frac{1}{a} - \frac{1}{b} - \frac{1}{c})$. In particular, its area is bounded from below by $\frac{2\pi}{21}$, with equality for $(a, b, c) = (2, 3, 7)$. It can be shown that $\inf \mu(\mathcal{H}/\Gamma) = \frac{2\pi}{21}$, where the infimum is taken over *all Fuchsian groups* Γ, not only over the triangular groups.

9.7.2 The infinite case

One can also consider noncompact polygons P having a number of proper vertices at infinity (e.g. x_1, x_2 on the picture) but having improper vertices at infinity as well (like x_3, x_4, x_5, x_6 on the figure).



The sides of P are the geodesic segments joining x_k and x_{k+1} where at least one of the vertices x_k, x_{k+1} should be proper. In the picture above we have therefore 8 sides. Notice that we might have vertices whose associated angle is π, like the vertex x_{10} from above.

Consider now that we are given the following data. We have first an involution ι from the set of the sides of P onto itself, and for each side e we are given an isometry $\sigma_e \in \mathrm{PSL}^*(2, \mathbb{R})$ of the hyperbolic plane $\mathcal{D}$, supposed to satisfy the conditions $\sigma_e(e) = \iota(e)$ and $\sigma_{\iota(e)} = \sigma_e^{-1}$.

In general, the group Γ generated by the isometries σ_e, where e belongs to the set of sides of P, is not discrete. However, its discreteness can be decided effectively. Consider the following conditions:

1. **The cycle condition** For *any cycle* $(x_1, x_2 \ldots x_p)$ *of vertices at finite distance there exist* $m \in \mathbb{Z}$, $m \geq 1$ such that

$$\sum_{j=1}^{p} \theta_j = \frac{2\pi}{m}.$$

2. **The cusp condition :** For *any cycle* $(x_1, x_2 \dots x_p)$ *of proper vertices at infinity* the transformation $\sigma_{e_p}\sigma_{e_{p-1}} \cdots e_{e_1}$ is *parabolic.*

The main result of this section is:

Theorem 30 (Poincaré). *If the data* (P, ι, σ_e) *satisfies* the cycle and cusp conditions *above then the group* Γ *generated by the* σ_e *(with* e *belonging to the set of sides of* P*) is finitely presented by means of the generators* σ_e *and relations:*

$$\sigma_e^2 = 1, \text{ whenever } \iota(e) = e$$

and

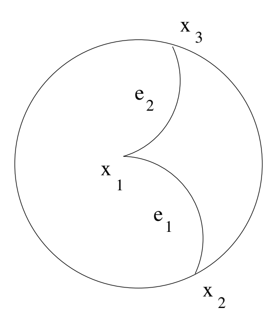
$$(\sigma_{e_p}\sigma_{e_{p-1}} \cdots \sigma_{e_1})^m = 1,$$

for any cycle of vertices at finite distance $(x_1 \dots x_p)$*, where* m *is the positive integer defined by the cycle condition. Moreover* Γ *is acting properly discontinuously on* $\mathcal{D}$ *and* P *is a fundamental domain for* Γ*.*

The vertices at finite distance contain the set of *elliptic fixed points* of Γ and possibly also some regular points.

The cycles above are constructed so that they give elliptic or parabolic cycles. Specifically, we start with x_1 vertex of P, e_1 edge of P incident to x_1. Define next $x_2 = \sigma_{e_1}(x_1)$; let e_2 be the other side, different from $\sigma_{e_1}(e_1)$ which is pending at x_2. Then put $x_3 = \sigma_{e_2}(x_2)$ and e_3 be the new edge pending at x_3, and so on, until we get $(x_{p+1}, e_{p+1}) = (x_1, e_1)$. The sequence $x_1 \dots x_p$ is called a cycle.

Example 5. Take for P the polygon with 2 sides from below where the angle between the sides e_1 and e_2 is $\frac{2\pi}{p}$.

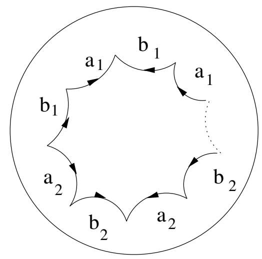


Let ι be the identity $\iota : \{e_1, e_2\} \to \{e_1, e_2\}$. Then Γ is given by

$$\Gamma = \left\langle \sigma_{e_1}, \sigma_{e_2} \mid \sigma_{e_1}^2 = \sigma_{e_2}^2 = (\sigma_{e_1}\sigma_{e_2})^p = 1 \right\rangle$$

and thus it is isomorphic to the dihedral group of order $2p$.

Example 6. Let P be a convex polygon with $4g$ sides in $\mathcal{D}$. The $4g$ edges $e_1, e_2, e_3 \dots e_{4g}$ of P are labeled counterclockwise as $a_1, b_1, a_1^{-1}, b_1^{-1}, a_2, b_2, a_2^{-1}$, $b_2^{-1}, \dots$ by giving them a label and an orientation.



Let ι be the involution which interchanges the edges where the same letter (and opposite exponents) appear. Let σ_{a_i} be the hyperbolic isometry which sends the edge labeled a_i into the edge labeled a_i^{-1} with the reversed orientation. Similarly define the isometries σ_{b_i} associated to the edges labeled b_i.

Assume that the internal angles α_j of the polygon P verify the identity:

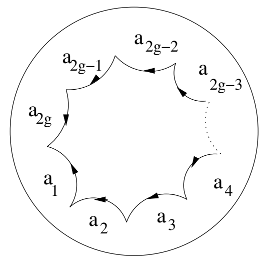
$$\sum_{j=1}^{4g} \alpha_j = 2\pi.$$

Then the cycle conditions are satisfied: there is only one cycle and the corresponding $m = 1$. In this case the vertices are regular points of the quotient (degenerate elliptic fixed points, since $m = 1$). Thus the group Γ is Fuchsian, given by the presentation:

$$\Gamma = \left\langle \sigma_{a_i}, \sigma_{b_j} \mid \sigma_{a_1}\sigma_{b_1}\sigma_{a_1}^{-1}\sigma_{b_1}^{-1}\sigma_{a_2}\sigma_{b_2}\sigma_{a_2}^{-1}\sigma_{b_2}^{-1} \cdots a_{a_g}\sigma_{b_g}\sigma_{a_g}^{-1}\sigma_{b_g}^{-1} = 1 \right\rangle$$

and having P as fundamental domain. In particular, $\mathcal{H}/\Gamma$ is the orientable surface of genus g obtained from the polygon P by identifying the edges according to the labeling.

Example 7. One remarks that different identifications of the sides might yield isomorphic Fuchsian groups with different presentations. Take for instance the pairing of the $4g$-gon above induced by the counterclockwise labeling of the edges $a_1, a_2, \dots, a_{2g}, a_1^{-1}, a_2^{-1}, \dots, a_{2g}^{-1}$.

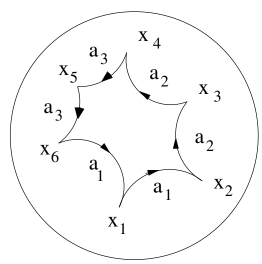


In this case we obtain the presentation

$$\Gamma = \left\langle \sigma_{a_1 \dots \sigma_{a_{2q}}} \mid \sigma_{a_1} \sigma_{a_2} \cdots \sigma_{a_{2q}} \sigma_{a_1}^{-1} \sigma_{a_2}^{-1} \cdots \sigma_{a_{2q}}^{-1} = 1 \right\rangle .$$

In both cases from above $\mathcal{D}/\Gamma$ is a hyperbolic surface of genus g.

Example 8. Let now consider the polygon P with the labeling of edges from the figure, inducing a natural involution, as above.



In this case we want that the isometry σ_{a_1} identifies the side $x_1 x_2$ with the side $x_2 x_3$ and so on. Thus σ_{a_1} does not preserve the orientation of the plane. The group Γ that we obtain is presented by

$$\Gamma = \left\langle \sigma_{a_1}, \sigma_{a_2}, \sigma_{a_3} \mid \sigma_{a_1}^2 \sigma_{a_2}^2 \sigma_{a_3}^2 = 1 \right\rangle \subset \mathrm{PSL}^*(2, \mathbb{R})$$

In fact, the quotient $\mathcal{D}/\Gamma$ is the non-orientable surface of genus 3.

A general finitely generated Fuchsian group has the following presentation:

1. Generators $a_1, b_1, a_2, b_2 \dots a_g, b_g, e_1 \dots e_r, p_1 \dots p_s, h_1 \dots h_i$
2. Relations:
   a) $e_i^{m_i} = 1$
   b) $\left(\prod_{i=1}^{g} [a_i, b_i]\right) e_1 \cdots e_r p_1 \cdots p_s h_1 \cdots h_i = 1$, where e_i are corresponding to the elliptic elements (which have finite orders m_i), p_j to the parabolics and thus in one-to-one correspondence with the cusps, h_j to the boundary hyperbolic, and a_i, b_i are hyperbolic.

9.8 Automorphisms and Fuchsian groups

Let Σ be a hyperbolic surface, namely a surface endowed with a complete Riemannian metric of constant curvature -1. Its universal covering is the hyperbolic plane $\mathcal{D}$ and hence Σ is isometric to $\mathcal{D}/\Gamma$, for a suitable Fuchsian group Γ. The Riemann uniformization theorem identifies hyperbolic structures on a closed surface with the complex structures. Then isometries correspond to biholomorphic maps. We often use interchangeably the terms Riemann surface and hyperbolic surface in the sequel.

Theorem 31 (Hurwitz). *If Σ is a closed Riemann surface of genus $g \geq 2$ then its group $\mathrm{Aut}(\Sigma)$ of isometries is finite and its order is uniformly bounded*

$$\mathrm{card}(\mathrm{Aut}(\Sigma)) \leq 84(g-1)$$

Equality holds for infinitely many g, but there exist infinitely many g for which the inequality is strict.

Proof. If $\Sigma = \mathcal{H}/\Gamma$, Γ Fuchsian and $N(\Gamma)$ is the normalizer of Γ in $\mathrm{PSL}(2,\mathbb{R})$ then the first observation is that $N(\Gamma)$ is also Fuchsian. In fact, if $n_k \in N(\Gamma)$ is a sequence of elements such that $n_k \to 1$, then one knows that $\lim_{k\to\infty} n_k \gamma n_k^{-1} = \gamma$, for all $\gamma \in \Gamma$. Now Γ being discrete implies that $n_k \gamma n_k^{-1} = \gamma$ for large k. Thus n_k and γ commute and thus they have the same fixed points, but there exist two γ_1, $\gamma_2 \in \Gamma$ which have not the same fixed points, contradiction.

Further if we choose $n \in N(\Gamma)$ then we have $n(\Gamma z) = \Gamma \cdot nz$ and thus there is an induced map $n : \mathcal{H}/\Gamma \to \mathcal{H}/\Gamma$, because n sends Γ-orbits into Γ-orbits. Thus we obtained an automorphism $n_* \in \mathrm{Aut}\,\Sigma$. The map $\eta : N(\Gamma) \to \mathrm{Aut}\,\Sigma$ sending n into n_* is a group homomorphism, which is surjective with $\ker \eta = \Gamma$ so that we obtain an isomorphism $\mathrm{Aut}(\Sigma) = N(\Gamma)/\Gamma$.

Moreover $N(\Gamma)$ is Fuchsian so that

$$\mathrm{card}(\mathrm{Aut}\,\Sigma) = \mathrm{card}(N(\Gamma)/\Gamma) = \frac{\mu(\mathcal{H}/\Gamma)}{\mu(\mathcal{H}/N(\Gamma))}$$

By Siegel's formula we have $\mu(\mathcal{H}/\Gamma) = 4\pi(g-1)$ since Γ has neither elliptics nor parabolics. Further

$$\mu(\mathcal{H}/N(\Gamma)) = 2\pi\left(2g' - 2 + t + \sum_{i=1}^{r}\left(1 - \frac{1}{\mathbf{m}_i}\right)\right) = c_{g',\mathbf{m}_i,t}$$

where $N(\Gamma)$ has signature $(g', \mathbf{m}_i, t)$. An elementary computation shows that

$$\mu(\mathcal{H}/N(\Gamma)) \geq c_{g'=0,\mathbf{m}_i=(2,3,7)} = \frac{\pi}{21}$$

with equality iff $N(\Gamma)$ is the triangle group $\langle 2, 3, 7\rangle$.

Moreover, $\langle 2, 3, 7 \rangle$ is a Coxeter group of matrices and hence it is residually finite. Thus there exists finite quotients $\frac{\langle 2,3,7 \rangle}{T_i}$ of arbitrary large index. This proves that there exist infinitely many g for which we have equality.

If $g = p + 1$, with prime $p > 84$ then there is no such quotient of order $84(g-1)$ because first there is no such group of order 84 and second, any Sylow p-subgroup should be normal.

Remark 33. The first part of the theorem is classical and due to Hurwitz and the second part is due to Macbeath. L.Greenberg proved also that for any finite group G there exists a closed Riemann surface Σ such that $\mathrm{Aut}\,\Sigma \simeq G$.

10

Spaces of discrete groups

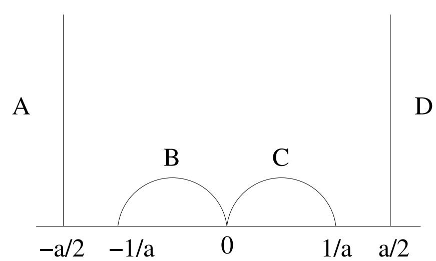
10.1 Non rigidity phenomena for subgroups of $\mathrm{PSL}(2,\mathbb{R})$

We consider first the family of Hecke groups which we already encountered before,

$$\Gamma_a = \left\langle \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ a & 1 \end{pmatrix} \right\rangle \subset \mathrm{PSL}(2,\mathbb{R}).$$

One knows that all Γ_a are free if $|a| \geq 2$ hence isomorphic to each other. If $a \in \mathbb{Z}$ then $\Gamma_a \subset \mathrm{PSL}(2,\mathbb{Z})$ and thus they are also discrete. However, we claim that Γ_2 and Γ_a, $a > 2$ are not conjugate inside $\mathrm{PSL}(2,\mathbb{R})$.

Using the Poincaré theorem we can construct Γ_a by making use of a fundamental regions. It is easy to verify that the domain P_a in the figure below is a fundamental region for Γ_a having the sides A, B, C, D.



Let $\iota : \{A, B, C, D\} \to \{A, B, C, D\}$ be the involution given by $\iota(A) = D$, $\iota(B) = C$, and let us define $\sigma_A = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}$ and $\sigma_B = \begin{pmatrix} 1 & 0 \\ a & 1 \end{pmatrix}$. Then Γ_a is naturally identified with the group generated by σ_A and σ_B.

Now, if $a > 2$ then $\mu(P_a) = +\infty$, while for $a = 2$ we have $\frac{a}{2} = \frac{1}{a}$ and thus the quotient surface is a non-compact cusped surface of finite volume. Thus, $\mathrm{vol}(\mathcal{H}/\Gamma_2) \neq \mathrm{vol}(\mathcal{H}/\Gamma_a)$ and thus Γ_2 and Γ_a cannot be conjugate.

Let now consider the group $\Gamma_2' = \left\langle \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \right\rangle \subset \mathrm{PSL}(2,\mathbb{R})$. Then Γ_2' has the same fundamental region P_2 as Γ_2. Moreover the involution ι :

$\{A, B, C, D\} \to \{A, B, C, D\}$ which yields Γ_2' is different from the previous one, namely $\iota(A) = C$, $\iota(B) = D$. If we consider the matrices ${\sigma'}_A^{-1} = \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}$, $\sigma'_B = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ then one verifies easily that Γ_2' is the group generated by σ'_A and σ'_B. The Poincaré theorem implies that Γ_2' is also free.

However, despite the fact that Γ_1 and Γ_2 share the same fundamental region and thus $\mu(\mathcal{H}/\Gamma_2) = \mu(\mathcal{H}/\Gamma_2')$, these groups are not conjugate within $\mathrm{PSL}(2, \mathbb{R})$. The reason is that $\mathcal{H}/\Gamma_2$ is homeomorphic to a 2-sphere with 3 cusps (i.e. $S^2 - \{0, 1, \infty\}$) while $\mathcal{H}/\Gamma_2'$ is a torus with 1-cusp. This follows immediately by looking at the identifications of sides of the respective fundamental domains induced by the involution.

If one seeks for families of Fuchsian groups then one needs to fix both the isomorphism type of the abstract group Γ as well as the homeomorphism type of the quotient surface $\mathcal{H}/\Gamma$. If Γ has no elliptic points then $\mathcal{H}/\Gamma$ is an orientable surface, with the orientation inherited by taking the quotient.

Definition 27. *The* Teichmüller space $\mathcal{T}(\Sigma)$ *of the oriented surface Σ is the space of marked Fuchsian groups $\Gamma \subset \mathrm{PSL}(2, \mathbb{R})$ such that $\pi_1\Sigma \to \Gamma$ is an isomorphism and Σ is orientation-preserving homeomorphic to $\mathcal{H}/\Gamma$. Notice that a marking of Γ is provided by a system of generators.*

An equivalent definition is to set:

Definition 28. *The Teichmüller space $\mathcal{T}(\Sigma)$ is the set of marked complex structures on Σ up to the equivalence relation below. A marked complex structure is a homotopy equivalence $f : \Sigma \to M$ where M is an arbitrary Riemann surface and two such f and $f' : \Sigma \to M'$ are equivalent $f \sim f'$ if there exist a conformal equivalence $h : M \to M'$ such that $f' \simeq f \circ h$, $\simeq$ denoting homotopy equivalence.*

By Riemann's uniformization theorem we can always write $M = \mathcal{H}/\Gamma$ where $\Gamma \subset \mathrm{PSL}(2, \mathbb{R})$ acts by isometries. In particular, we have an identification:

$$\mathcal{T}(\Sigma) = \mathrm{Hom}^+_{f,d}(\pi_1, \Sigma, \mathrm{PSL}(2, \mathbb{R}))/\text{conjugacy within } \mathrm{PSL}(2, \mathbb{R})$$

where $\mathrm{Hom}^+_{f,d}(\pi_1, \Sigma, \mathrm{PSL}(2, \mathbb{R}))$ denotes the space of *faithful* homomorphisms $\varphi : \pi_1\Sigma \to \mathrm{PSL}(2, \mathbb{R})$ such that $\varphi(\pi_1(\Sigma)) = \Gamma \subset \mathrm{PSL}(2, \mathbb{R})$ is a Fuchsian group and if φ preserves the orientation i.e. the induced homeomorphism $\Sigma \to \mathcal{H}/\Gamma$ is preserving the orientation.

10.2 Thurston-Bonahon-Penner-Fock coordinates on the Teichmüller spaces

10.2.1 Preliminaries on fatgraphs

Let Γ be a finite graph. We denote by V_Γ and E_Γ the set of its vertices and edges respectively.

Definition 29. *An* orientation *at a vertex v is a cyclic ordering of the (half-) edges incident at v. A* fatgraph *(sometimes called* ribbon graph*) is a graph endowed with an orientation at each vertex of Γ. A* left-hand-turn path *in Γ is a directed closed path in Γ such that if e_1, e_2 are successive edges in the path meeting at v, then e_2, e_1 are successive edges with respect to the orientation at v. The ordered pair e_1, e_2 is called a left-turn. We sometimes call* faces *of Γ the left-hand-turn paths and denote them by F_Γ.*

A fatgraph is usually represented in the plane, by assuming that the orientation at each vertex is the counter-clockwise orientation induced by the plane, while the intersections of the edges at points other than the vertices are ignored. There is a natural surface, which we denote by Γ^t obtained by thickening the fatgraph. We usually call Γ^t the ribbon graph associated to Γ. We replace the half-edges around a vertex by thin strips joined at the vertex, whose boundary arcs have natural orientations. For each edge of the graph we connect the thin strips corresponding to the vertices by a ribbon which follows the orientation of their boundaries. We obtain an oriented surface with boundary. The boundary circles are in one-to-one correspondence with the left-hand-turn paths. If one caps each left-hand-turn path by a 2-disk we find a closed surface Γ^c, and this explains why we called these paths faces. The centers of the 2-disks will be called punctures of Γ^c and $\Gamma^o = int(\Gamma^t)$ is homeomorphic to the punctured surface.

There is a canonical embedding $\Gamma \subset \Gamma^t$, and one can associate to each edge e of Γ a properly embedded orthogonal arc $e^\perp$ which joins the two boundary components of the thin strip lying over e. The dual arcs $e^\perp$ divide the ribbon Γ^t into hexagons. When we consider the completion Γ^c, we join the boundary points of these dual arcs to the punctures within each 2-disk face and obtain a set of arcs connecting the punctures, denoted by the same symbols. Then the dual arcs divide Γ^c into triangles. We set $\Delta(\Gamma)$ for the triangulation obtained this way. The vertices of $\Delta(\Gamma)$ are the punctures of Γ^c. Remark that $\Delta(\Gamma)$ is well-defined up to isotopy. Now the fatgraph $\Gamma \subset \Gamma^t$ can be recovered from $\Delta(\Gamma)$ as follows. Mark a point in the interior of each triangle, and connect points corresponding to adjacent triangles. This procedure works for any given triangulation Δ of an oriented surface and produces a fatgraph $\Gamma = \Gamma(\Delta)$ with the property that $\Delta(\Gamma) = \Delta$. The orientation of Γ comes from the surface.

If Γ^o is the surface Σ_g^s of genus g with s punctures then by Euler characteristic reasons we have: $\sharp V_\Gamma = 4g - 4 + 2s$, $\sharp E_\Gamma = 6g - 6 + 3s$, $\sharp F_\Gamma = s$.

10.2.2 Coordinates on Teichmüller spaces

Marked ideal triangles

Let us recall that $\mathcal{D}$ denotes the unit disk, equipped with the hyperbolic metric. Recall that any two ideal triangles are isometric, since we may find a Möbius transformation, which takes one onto the other. Choose a point on each edge of the ideal triangle. The chosen points will be called *tick-marks*.

Definition 30. *A* marked ideal triangle *is an ideal triangle with a tick-mark on each one of its three sides. An isomorphism between two marked ideal triangles is an isomorphism between the ideal triangles which preserves the tick-marks.* A standard marked ideal triangle *is one which is isometric to the marked ideal triangle whose vertices in the disk model are given by* $v_1 = 1,\ v_2 = \omega,\ v_3 = \omega^2$ *and whose tick-marks are* $t_1 = -(2-\sqrt{3}),\ t_2 = -(2-\sqrt{3})\omega,\ t_3 = -(2-\sqrt{3})\omega^2$, *where* $\omega = \mathrm{e}^{2\pi \mathrm{i}/3}$.

The ideal triangle and its tick-marks are pictured in figure 10.1 in both the half-plane model and the disk model; they correspond each other by the map $z \mapsto \frac{z-(\omega+1)}{z-(\bar{\omega}+1)}$.

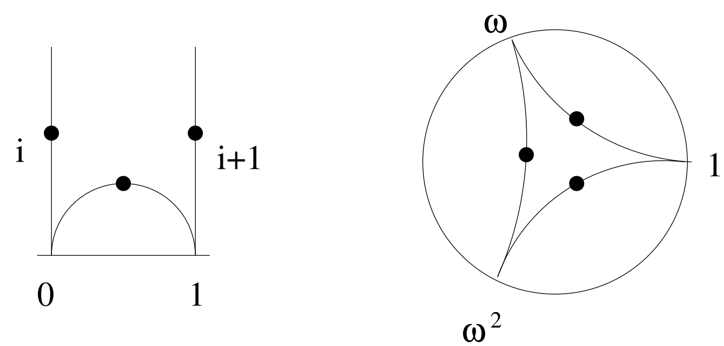


Fig. 10.1. The standard marked ideal triangle

Coordinates on the Teichmüller space of punctured surfaces

Set $\mathcal{T}_g^s$ for the Teichmüller space of the surface of genus g with s punctures. Let Γ be a fatgraph with the property that Γ^c is a surface of genus g with s punctures and let S denote the surface Γ^c endowed with a hyperbolic structure of finite volume, having the cusps at the punctures.

As already explained above we have a triangulation $\Delta(\Gamma)$ associated to Γ. One deforms the arcs of $\Delta(\Gamma)$ within their isotopy class in order to make them geodesic. We shall associate a real number $t_e \in \mathbb{R}$ to each edge of $\Delta(\Gamma)$ (equivalently, to each edge of Γ). Set Δ_v and Δ_w for the two triangles sharing the edge $e^{\perp}$. We consider next two adjacent lifts of these triangles (which we denote by the same symbols) to the hyperbolic space $\mathbb{H}^2$. Then both Δ_v and Δ_w are isometric to the standard ideal triangle of vertices v_1, v_2 and v_3. These two isometries define (by pull-back) canonical tick-marks t_v and respectively t_w

on the geodesic edge shared by Δ_v and Δ_w. Set t_e for the (real) length of the translation along this geodesic needed to shift t_v to t_w. Notice that this geodesic inherits an orientation as the boundary of the ideal triangle Δ_v in $\mathbb{H}^2$ which gives t_e a sign. If we change the role of v and w the number t_e is preserved.

An equivalent way to encode the translation parameters is to use the cross-ratios of the four vertices of the glued quadrilateral $\Delta_v \cup \Delta_w$, which are considered as points of $\mathbb{R}P^1$. It is convenient for us to consider $\mathbb{R}P^1$ as the boundary of the upper half-plane model of $\mathbb{H}^2$, and hence the ideal points have real (or infinite) coordinates. Let assume that Δ_v is the ideal triangle determined by $[p_0 p_{-1} p_\infty]$ and Δ_w is $[p_0 p_\infty p]$. We consider then the following cross-ratios:

$$z_e = [p_{-1}, p_\infty, p, p_0] = [p, p_0, p_{-1}, p_\infty] = \log - \frac{(p_0 - p)(p_{-1} - p_\infty)}{(p_\infty - p)(p_{-1} - p_0)}.$$

This cross-ratio reflects both the quadrilateral geometry and the decomposition into two triangles. In fact the other possible decomposition into two triangle of the same quadrilateral leads to the value z_e.

The relation between the two translation parameters t_e and z_e is immediate. Consider the ideal quadrilateral of vertices $-1, 0, e^z$ and ∞, whose cross-ratio is $z_e = z$, where $e = [0\infty]$. The left triangle tick-mark is located at i, while the right one is located at ie^{-z}, after the homothety sending the triangle into the standard triangle. Taking in account that the orientation of the edge e is up-side one derives that t_e is the signed hyperbolic distance between i and $e^{-z_e}i$, which is z_e.

Proposition 25. *The map* $\mathbf{t}_\Gamma : \mathcal{T}_g^s \to \mathbb{R}^{E_\Gamma}$ *given by* $t_\Gamma(S) = (t_e)_{e \in E_\Gamma}$ *is a homeomorphism onto the linear subspace* $\mathbb{R}^{E_\Gamma / F_\Gamma} \subset \mathbb{R}^{E_\Gamma}$ *given by equations:*

$$t_\gamma := \sum_{k=1}^{n} t_{e_k} = 0,$$

for all left-hand-turn closed paths $\gamma \in F_\Gamma$*, which is expressed as a cyclic chain of edges* $e_1, ..., e_n$*.*

Remark 34. Notice that there are exactly s left-hand-turn closed paths, which lead to s independent equations hence the subspace $\mathbb{R}^{E_\Gamma / F_\Gamma}$ from above is of dimension $6g - 6 + 2s$.

Proof. The map $\mathbf{t}_\Gamma$ is continuous, and it suffices to define an explicit inverse for it. Let Γ be a trivalent fatgraph whose edges are labeled by real numbers $\mathbf{r} = (r_e)_{e \in E_\Gamma}$. We want to paste one copy Δ_v of the standard marked ideal triangle on each vertex v of Γ and glue together by isometries these triangles according to the edges connections. Since the edges of an ideal triangle are of infinite length we have the freedom to use arbitrary translations along these geodesics when gluing together adjacent sides. If $e = [vw]$ is an edge of Γ then one can associate a real number $t_e \in \mathbb{R}$ as follows. There are two tick-marks, namely t_v and t_w on the common side of Δ_v and Δ_w. We denote by t_e the amount needed for translating t_v into t_w according to the orientation inherited as a boundary of Δ_v. Given now the collection of real numbers $\mathbf{r}$ we can construct unambiguously our Riemann surface $S(\Gamma, \mathbf{r})$, which moreover has the property that $t_\Gamma(S(\Gamma, \mathbf{r})) = \mathbf{r}$. Furthermore it is sufficient now to check whenever this constructions yields a complete Riemann surfaces. The completeness at the puncture determined by the left-hand-turn path γ is equivalent to the condition $t_\gamma = 0$, and hence the claim. The cusps of $S(\Gamma)$ are in bijection with the left-hand-turn paths in Γ, and the triangulation of $S(\Gamma)$ obtained by our construction corresponds to Γ.

The Fuchsian group associated to Γ and $\mathbf{r}$

The surface $S(\Gamma, \mathbf{r})$ is uniformized by a Fuchsian group $G = G(\Gamma, \mathbf{r}) \subset \mathrm{PSL}(2, \mathbb{R})$, i.e. $S(\Gamma, \mathbf{r}) = \mathcal{D}/G(\Gamma, \mathbf{r})$. We can explicitly determine the generators of the Fuchsian group, as follows.

We have natural isomorphisms between the fundamental group $\pi_1(S(\Gamma, \mathbf{r})) \cong \pi_1(\Gamma^t) \cong \pi_1(\Gamma)$. Any path γ in Γ is a cyclic sequence of adjacent directed edges $e_1, e_2, e_3, ..., e_n$, where e_i and e_{i+1} have the vertex v_i in common. We insert between e_i and e_{i+1} the symbol lt if e_i, e_{i+1} is a left-hand-turn, the symbol rt if it is a right-hand-turn and no symbol otherwise (i.e. when e_{i+1} is e_i with the opposite orientation). Assume now that we have a Riemann surface whose coordinates are $\mathbf{t}_\Gamma(S) = \mathbf{r}$. We define then a representation $\rho_{\mathbf{r}} : \Pi_1(\Gamma) \to \mathrm{PSL}(2, \mathbb{R})$ of the path groupoid $\Pi_1(\Gamma)$ by the formulas:

$$\rho_{\mathbf{r}}(e) = \begin{pmatrix} 0 & e^{\frac{r_e}{2}} \\ -e^{-\frac{r_e}{2}} & 0 \end{pmatrix}, \text{ and } \rho_{\mathbf{r}}(lt) = \rho_{\mathbf{r}}(rt)^{-1} = \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}.$$

This is indeed well-defined since $\rho_{\mathbf{r}}(e)^2 = -1 = 1 \in \mathrm{PSL}(2, \mathbb{R})$, and hence the orientation of the edge does not matter, and $\rho_{\mathbf{r}}(lt)^3 = \rho_{\mathbf{r}}(rt)^3 = 1$. Furthermore the fundamental group $\pi_1(\Gamma)$ is a subgroup of $\Pi_1(\Gamma)$.

Proposition 26. *The Fuchsian group $G(\Gamma, \mathbf{r})$ is $\rho_{\mathbf{r}}(\pi_1(\Gamma)) \subset \mathrm{PSL}(2, \mathbb{R})$.*

Proof. We can begin doing the pasting without leaving the hyperbolic plane, until we get a polygon P, together with a side pairing. We may think of each triangle as having a white face and a black face, and build the polygon P such that all the triangles have white face up. We attach to each side pairing (s_i, s_j) an orientation preserving isometry A_{ij}, such that $A_{ij}(s_i) = s_j$, A_{ij} sends tick-marks into the tick-marks shifted by r_e, and $P \cap A_{ij}(P) = \emptyset$. Denote by G the subgroup of $\mathrm{Isom}^+(\mathcal{D})$ generated by all the side-pairing transformations. In order to apply the Poincaré Theorem all the vertex-cycle transformations must be parabolic. This amounts to ask that for every left-hand-turn closed path γ we have $t_\gamma = 0$. Then by the Poincaré theorem G is a discrete group of isometries with P as its fundamental domain and $\mathcal{D}/G$ is the complete hyperbolic Riemann surface $S(\Gamma, \mathbf{r})$.

We need now the explicit form of the matrices A_{ij}. We obtain them by composing the isometries sending a marked triangle into the adjacent one, in a suitable chain of triangles, where consecutive ones have a common edge. If e is such an edge we remark that $\rho_{\mathbf{r}}(e)$ do the job we want, because it sends the triangle $[-1, 0, \infty]$ into $[e^{r_e}, \infty, 0]$. Moreover the quadrilateral $[-1, 0, e^{r_e}, \infty]$, with this decomposition into two triangles, has associated the cross-ratio r_e. We need next to use $\rho_{\mathbf{r}}(lt)$ which permutes counter-clockwise the tick-marks and the vertices $-1, 0$ and ∞ of the ideal triangle. Then one identifies the matrices A_{ij} with the images of the closed paths by $\rho_{\mathbf{r}}$.

Remark 35. We observe that the left-hand-turn paths are preserved under an isomorphism of graphs which preserves the cyclic orientation at each vertex. Thus any automorphism of the fatgraph Γ induces an automorphism of $S(\Gamma)$.

10.3 Coordinates on the Teichmüller space of surfaces with geodesic boundary

Set $\mathcal{T}_{g,s;or}$ for the Teichmüller space of surfaces of genus g with s *oriented* boundary components. Here *or* denotes the choice of one orientation for each of the boundary components. Since the surface has a canonical orientation, we can set unambiguously $or : \{1, 2, ..., s\} \to \mathbb{Z}/2\mathbb{Z}$ by assigning $or(j) = +1$ if the orientation of the j-th component agrees with that of the surface and $or(j) = -1$, otherwise. We suppose that each boundary component is a geodesic in the hyperbolic metric, and possibly a cusp (hence in some sense this space is slightly completed). Let Γ be a fatgraph with the property that Γ^t is a surface of genus g with s boundary components and let S denote the surface Γ^t endowed with a hyperbolic structure, for which the boundary is geodesic. Assume that, in this metric, the boundary geodesics b_j have length l_j.

Consider the restriction of the hyperbolic metric to $int(\Gamma^t) = \Gamma^o$. Then Γ^o is canonically homeomorphic to the punctured surface $\Gamma^c - \{p_1, ..., p_s\}$. In particular there is a canonically induced hyperbolic metric on $\Gamma^c - \{p_1, ..., p_s\}$, which we denote by S^*. Moreover this metric is not complete at the punctures p_j. Suppose that the punctures p_j corresponds to the left-hand-turn closed paths γ_j, or equivalently the boundary components geodesics b_j, of length l_j. Assume that we have an ideal triangulation of S^* by geodesic simplices, whose ideal vertices are the punctures p_j. Then the holonomy of the hyperbolic structure around the vertex p_j is a non-trivial, and it can be calculated in the following way (see [26], Prop.3.4.18, p.148). Consider a geodesic edge α entering the puncture and a point $p \in \alpha$. Then the geodesic spinning around p_j in the positive direction (according to the orientation of the boundary circle) is intersecting again α a first time in the point $h_{p_j}(p)$. The hyperbolic distance between the points p and $h_{p_j}(p)$ is the length l_j of the boundary circle in the first metric. Moreover the point $h_{p_j}(p)$ lies in the ray determined by p and the puncture p_j. Notice that if we had chose the loop encircling the puncture to go in opposite direction then the iterations $h_{p_j}(p)$ would have gone faraway from the puncture, and the length would have been given the negative sign. Set therefore l_j^o* for the signed length.

We construct as above the geodesic ideal triangulation $\Delta(\Gamma)$ of the non-complete hyperbolic punctured surface S^*. We can therefore compute the holonomy map using the thick-marks on some edge abutting to the puncture p_j. It is immediately that the the holonomy displacement on this edge is given by t_{γ_j}, where γ_j is the left-hand-turn closed path corresponding to this puncture. In particular we derive that:

$$|t_{\gamma_j}| = l_j, \text{ for all } j \in \{1, 2, ..., s\}.$$

Using the method from the previous section we know how to associate to any edge e of Γ a real number $t_e = t_e(S^*)$ measuring the shift between two ideal triangles in the geodesic triangulation of the surface S^*.

Proposition 27. *The map* $\mathbf{t}_\Gamma : \mathcal{T}_{g,s;or} \to \mathbb{R}^{E_\Gamma}$ *given by* $t_\Gamma(S) = (t_e)_{e \in E_\Gamma}$ *is a homeomorphism.*

Proof. The construction of an inverse map proceeds as above. Given $\mathbf{r} \in \mathbb{R}^{E_\Gamma}$ we construct a non-complete hyperbolic surface S^* with s punctures with the given parameters, by means of gluing ideal triangles. As shown in ([26], Prop. 3.4.21, p.150) we can complete this hyperbolic structure to a surface with geodesic boundary S, such that $int(S) = S^*$. Further if $t_{\gamma_j} > 0$, then we assign the orientation of γ_j for the boundary component b_j, otherwise we assign the reverse orientation. When $t_{\gamma_j} = 0$ it means that we have a cusp at p_j.

Remark 36. The two points of $\mathcal{T}_{g,s;or}$ given by the same hyperbolic structure on the surface $\Sigma_{g,s}$ but with distinct orientations of some boundary components lie in the same connected component. Nevertheless the previous formulas shows that a path connecting them must pass through the points of $\mathcal{T}_{g,s;or}$ corresponding to surfaces having a cusp at the respective puncture.

Set $\mathcal{T}_{g,s}$ for the Teichmüller space of surfaces of genus g with s *non-oriented* boundary components, i.e. hyperbolic metrics for which the boundary components are geodesic. There is a simple way to recover coordinates on $\mathcal{T}_{g,s}$ from its oriented version. Let $\psi : \mathbb{R}^{E_\Gamma} \to \mathbb{R}^{F_\Gamma}$ be the map $\psi(\mathbf{t}) = (t_{\gamma_i})_{\gamma_i \in F_\Gamma}$. Choose a projector $\psi^* : \mathbb{R}^{E_\Gamma} \to \ker\psi = \mathbb{R}^{E_\Gamma / F_\Gamma}$, and set $\iota_{|.|} : \mathbb{R}^{F_\Gamma} \to \mathbb{R}^{F_\Gamma}$ for the map given on coordinates by $\iota_{|.|}(y_j)_{j=1,\sharp F_\Gamma} = (|y_j|)_{j=1,\sharp F_\Gamma}$. Then $\mathcal{T}_{g,s}$ is the quotient by the $(\mathbb{Z}/2\mathbb{Z})^{F_\Gamma}$-action on $\mathcal{T}_{g,s;or}$ which changes the orientation of the boundary components.

Proposition 28. *We have a homeomorphism* $\mathbf{t}_\Gamma : \mathcal{T}_{g,s} \to \mathbb{R}^{6g-6+2s} \oplus \mathbb{R}^s$*, which is induced from the second line of the following commutative diagram:*

$$\begin{array}{ccc}
\mathcal{T}_{g,s;or} & \xrightarrow{(\psi^*\oplus\psi)\circ\mathbf{t}_\Gamma} & \mathbb{R}^{E_\Gamma/F_\Gamma}\oplus\mathbb{R}^{F_\Gamma} \\
\downarrow & & \downarrow id\oplus\iota_{|.|} \\
\mathcal{T}_{g,s} & \longrightarrow & \mathbb{R}^{E_\Gamma/F_\Gamma}\oplus\mathbb{R}_+^{F_\Gamma} \\
\uparrow & & \uparrow id\oplus 0 \\
\mathcal{T}_g^s & \longrightarrow & \mathbb{R}^{E_\Gamma/F_\Gamma}
\end{array}$$

Remark 37. Observe that the embedding $\mathcal{T}_g^s \hookrightarrow \mathcal{T}_{g,s}$ given in terms of coordinates by adding on the right a string of zeroes lifts to an embedding $\mathcal{T}_g^s \hookrightarrow \mathcal{T}_{g,s;or}$.

Putting together the results of the last two sections we derive that:

Proposition 29. *The map $\mathbf{t}_\Gamma : \mathcal{T}_{g,n;or}^s \to \mathbb{R}^{E_\Gamma}$ given by $t_\Gamma(S) = (t_e)_{e\in E_\Gamma}$ is a homeomorphism of the Teichmüller space of surfaces of genus g with n oriented boundary components and s punctures onto the linear subspace $\mathbb{R}^{E_\Gamma/F^*\Gamma}$ of dimension $6g - 6 + 3n + 2s$ given by the equations: $t_{\gamma_j} = 0$, for those left-hand-turn closed paths γ_j corresponding to the punctures, $\gamma_j \in F_\Gamma^* \subset F_\Gamma$.*

Remark 38. W.Thurston associated to an ideal triangulation a system of shearing coordinates for the Teichmüller space in mid eighties (see [27]) and from a slightly different perspective in Penner's treatment of the decorated Teichmüller spaces ([22]). The systematic study of such coordinates appeared later in the papers of F.Bonahon [2] and V.Fock unraveled in [5] the elementary aspects of this theory which lead him further to the quantification of the Teichmüller space.

References

1. F.Bonahon, *Shearing hyperbolic surfaces, bending pleated surfaces and Thurston's symplectic form*, Ann. Fac. Sci. Toulouse Math. (6) 5 (1996), no. 2, 233–297.
2. V.Fock, *Dual Teichmüller spaces*, math.dg-ga/9702018.
3. R.Penner, *The decorated Teichmüller space of punctured surfaces*, Commun. Math. Phys 113(1987), 299-339.
4. W.Thurston, *Three-dimensional geometry and topology*, Vol. 1., Ed. Silvio Levy, Princeton Mathematical Series, 35, 1997.
5. W.Thurston, *Minimal stretch maps between hyperbolic surfaces*, preprint 1986, available at `math.GT/9801039`.

11

The interplay between mapping class groups and Teichmüller spaces

11.1 Mapping class groups acting on Teichmüller spaces

There is a close relation between mapping class groups and Teichmüller spaces. The Dehn-Nielsen-Baer theorem provides an identification between $\mathrm{Mod}(\Sigma)$ and $\mathrm{Out}^+(\pi_1\Sigma)$. Then mapping class group acts by left composition on the space $\mathcal{T}(\Sigma)$ which is a space of group representations, up to conjugacy:

$$\mathcal{T}(\Sigma) = \mathrm{Hom}^+_{f,d}(\pi_1\Sigma, \mathrm{PSL}(2,\mathbb{R})) \,/\, \mathrm{PSL}(2,\mathbb{R})$$

Specifically, this action is given by

$$(\varphi, [\rho]) \longrightarrow [\rho \circ \varphi^{-1}]$$

Moreover, $\mathrm{Mod}(\Sigma)$ acts by real analytic homeomorphisms. This action is important in understanding both the algebraic structure of the mapping class group using the geometry of the Teichmüller space, because of the following basic result going back to F.Klein and R.Fricke.

Proposition 30 (Fricke-Klein 1889, Kravetz 1959). $\mathrm{Mod}(\Sigma)$ *acts properly discontinuously on* $\mathcal{T}(\Sigma)$.

Proof. Let assume that there exist a sequence $\varphi_n \in \mathrm{Mod}\,\Sigma$ so that there exist two compacts C_1, C_2 in the Teichmüller space with the property $\varphi_n(C_1) \cap C_2 \neq \emptyset$ for all n. Then there exists a convergent sequence of points $z_n \to z \in \mathcal{T}(\Sigma)$ so that $\varphi_n(z_n)$ also converges to some point $w \in \mathcal{T}(\Sigma)$. Thus $\varphi_n^{-1}\varphi_{n-1}(z_n) \to z$. We will show that if $\xi_n \in \mathrm{Mod}\,\Sigma$ has the property that $\xi_n z_n \to z$ then $\xi_n = \mathbb{1}$ for large enough n.

This is a consequence of the following facts:

1. If Γ is a Fuchsian group then the set

$$A(\Gamma) = \{\mathrm{Tr}(\gamma); \gamma \in \Gamma \subset \mathrm{PSL}(2,\mathbb{R})\} \subset \mathbb{R}_+$$

is mapped by the function $\cosh \frac{1}{2}(x)$ bijectively into the marked set of lengths of geodesics of the surface $\mathcal{H}/\Gamma$ (indexed by elements of Γ). Moreover, these sets are discrete.
2. If $\xi \in \mathrm{Mod}(\Sigma)$ and $\Gamma = \rho(\pi_1\Sigma)$ is a Fuchsian group then the marked set $A(\xi\Gamma)$ is obtained from the marked set $A(\Gamma)$ by a permutation of its elements.
3. The regular functions $\mathrm{tr}(\rho(\gamma))$, $\gamma \in \pi_1\Sigma$, viewed as functions $\mathcal{T}(\Sigma) \to \mathbb{R}$ are generating a polynomial algebra which is finitely generated. The proof is based on the identity:
$$\mathrm{tr}(x)\,\mathrm{tr}(y) = \mathrm{tr}(xy) + \mathrm{tr}(xy^{-1}).$$
4. If $\xi_n z_n \to z$ then for large n
$$A(\xi_n \Gamma_{zn}) \text{ and } A(\Gamma_s) \text{ agree on their first } N \text{ items}$$
It N is large enough in order that all generators of the algebra above are contained among the first N items then we find that $A(\xi_n\Gamma_{zn}) = A(\Gamma_z)$. Since ξ_n acts as a permutation on the marked sets of geodesics we derive that the permutation is the identity.
5. Two hyperbolic structures on a surface having the same marked lengths of geodesics are isometric. In fact, if the traces of two discrete faithful representations coincide i.e. $\mathrm{tr}(\rho(\gamma) = \mathrm{tr}(\rho'(\gamma))$ for any $\gamma \in \pi_1(\Sigma)$ then the representations are conjugate.

Remark 39. 1. The $\mathrm{Mod}(\Sigma_g)$-action on the Teichmüller space is *effective* if $g \geq 3$. When $g = 1, 2$ the hyperelliptic involution acts trivially on $\mathcal{T}(\Sigma)$.
2. The quotient $\mathcal{T}(\Sigma)/\mathrm{Mod}(\Sigma)$ is naturally a complex space with orbifold singularities (at points where the $\mathrm{Mod}(\Sigma)$ action is not free). However, one knows that all stabilizers should be finite. In this respect the moduli space $\mathcal{M}(\Sigma) = \mathcal{T}(\Sigma)/\mathrm{Mod}(\Sigma)$ plays the role of a *classifying space for the mapping class group*. For instance, we have an isomorphism
$$H^*(\mathcal{M}(\Sigma); \mathbb{Q}) \simeq H^*(\mathrm{Mod}(\Sigma); \mathbb{Q})$$
3. Since $\mathcal{T}(\Sigma)$ is a topological cell each torsion element of $\mathrm{Mod}(\Sigma)$ should fix a non-empty set. In particular, any periodic mapping class contains a periodic homeomorphism which is a conformal homeomorphism for some complex structure on Σ.

References

1. R.Fricke and F.Klein, *Vorlesungen über die Theorie der automorphen Funktionen*, B.G.Teubner, 1889 and 1926.
2. S.Kravetz, *On the geometry of Teichmüller spaces and the structure of their modular groups*, Ann. Acad. Sci. Fenn. 278 (1959), 1–35.

11.2 Stabilizers of the mapping class group action

The action of $\mathrm{Mod}(\Sigma)$ on $\mathcal{T}(\Sigma)$ is properly discontinuous and hence it has finite stabilizers. A point p in $\mathcal{T}(\Sigma)$ corresponds to a class of marked Riemann surface $p = [S]$, and we can identify the stabilizer $\mathrm{Mod}(\Sigma)_p$ of the point p, as follows:

$$\mathrm{Mod}(\Sigma)_p = \{\varphi \in \mathrm{Mod}(\Sigma) \text{ such that } [\varphi S] = [S]\}$$

Moreover, S is defined by the holonomy map $\rho_S : \pi_1\Sigma \to \mathrm{PSL}(2,\mathbb{R})$ and so we have $\rho_{\varphi S} = \varphi \circ \rho_S$, where φ is interpreted now as an element of $\mathrm{Out}^+(\pi_1\Sigma)$. Since the marked surfaces determined by $\rho_{\varphi S}$ and ρ_S are the same they should be obtained by means of a conjugation within $\mathrm{PSL}(2,\mathbb{R})$ i.e. there exists $\lambda = \lambda_\varphi \in \mathrm{PSL}(2,\mathbb{R})$ so that

$$\rho_{\varphi S} = \lambda_\varphi \rho_S \lambda_\varphi^{-1}$$

In particular, λ_φ belongs to the normalizer of the Fuchsian group $\rho_S(\pi_1 S)$ and it is immediate that the map

$$\lambda : \mathrm{Mod}(\Sigma)_p \longrightarrow N\left(\rho_S(\pi_1\Sigma)\right)/\rho_S(\pi_1\Sigma)$$

is a group homomorphism. Actually, we have a more precise result:

Proposition 31. *The stabilizer of the class of the marked Riemann surface* $[S]$ *is given by*

$$\mathrm{Mod}(\Sigma)_{p=[S]} = \mathrm{Aut}(S)$$

where Aut(S) *are the conformal (i.e. holomorphic) automorphism group of* S*.*

In fact, any element of $\mathcal{N}(\Gamma)/\Gamma$, Γ Fuchsian group corresponds to an automorphism of the Riemann surface (see the section 9.8).

Corollary 18. *For a generic Riemann surface* S *we have* $\mathrm{Aut}(S) = \{\mathbb{1}\}$.

Remark 40. It is known that, if the genus of Σ is $g \geq 4$, then the local structure of $\mathcal{T}(\Sigma)/\mathrm{Mod}(\Sigma)$ around $p \in \mathcal{T}(\Sigma)/\mathrm{Mod}(\Sigma)$ is described by the quotient $\mathbb{R}^{6g-6}/F_p$, where $F_p \subset \mathrm{GL}(6g-6)$ is a finite group, which is the image of a faithful linear representation $\mathrm{Aut}(S) \to \mathrm{GL}(6g-6)$ (where $[S] = p$).

In particular, the point p is smooth in the quotient iff S has no automorphisms.

A more elaborate analysis shows that the space $\mathcal{T}(\Sigma)/\mathrm{Mod}(\Sigma)$ is singular at the points when S has automorphisms (as shown by E.Rauch in 1962) for $g \geq 4$. For $g = 2$ there is only one singular point, corresponding to the Riemann surface given by the equation:

$$y^2 = x^5 - 1$$

which has additional symmetries with respect to the rest of Riemann surfaces having only the hyperelliptic involution automorphism. For $g = 3$ the hyperelliptic locus consists of smooth points.

Furthermore it is known that there exists a finite index subgroup of $\mathrm{Mod}(\Sigma)$ which acts freely on $\mathcal{J}(\Sigma)$. A quantitative estimate of the index follows from the following result due to J. P. Serre (1958):

Proposition 32. *If $\varphi \in \mathrm{Mod}(\Sigma)_{[S]}$ is an automorphism of the Riemann surface S and*

$$\varphi_* : H_1(\Sigma, \mathbb{Z}/\ell\mathbb{Z}) \longrightarrow H_1(\Sigma, \mathbb{Z}/\ell\mathbb{Z})$$

is the identity for some $\ell \geq 3$ then $\varphi = \mathbb{1}$. In particular $\ker(\mathrm{Mod}(\Sigma) \to Aut(H_1(\Sigma; \mathbb{Z}/\ell\mathbb{Z})) \cong Sp(2g, \ell))$ *acts freely on $\mathcal{T}(\Sigma)$ for any $\ell \geq 3$.*

11.3 The Ptolemy modular groupoid

The modular groupoid was considered by Mosher in his thesis and further as a key ingredient in [18, 19], it is implicit in Harer's paper on the arc complex (see [12]) and then studied by Penner (see [22, 23]; notice that the correct definition is that from [23]) who introduced also the terminology.

Recall that a groupoid is a category whose morphisms are invertible, such that between any two objects there is at least one morphism. The morphisms from an object to itself form a group (the group associated to the groupoid).

Remark 41. Suppose that we have an action of a group G on a set M. We associate a groupoid $\mathcal{G}(G, M)$ as follows: its objects are the G-orbits on M, and the morphisms are the G-orbits of the diagonal action on $M \times M$. If the initial action was free then G embeds in $\mathcal{G}(G, M)$ as the automorphisms group of any object.

Assume that we have an ideal triangulation $\Delta(\Gamma)$ of a surface Σ_g^s. If e is an edge shared by the triangles Δ_v and Δ_w of the triangulation then we change the triangulation by excising the edge e and replacing it by the other diagonal of the quadrilateral $\Delta_v \cup \Delta_w$, as in figure 11.1. This operation $F[e]$ was called flip in [5] or elementary by Mosher and Penner.

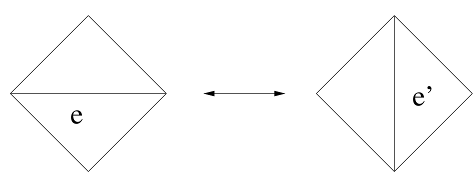


Fig. 11.1. The flip

Let $\mathcal{IT}(\Sigma_g^s))$ denote the set of isotopy classes of ideal triangulations of Σ_g^s. The reduced *Ptolemy groupoid* $\overline{P_g^s}$ is the groupoid generated by the flips action on $\mathcal{IT}(\Sigma_g^s)$). Specifically its elements are classes of sequences $\Delta_0, \Delta_1, ..., \Delta_m$, where Δ_{j+1} is obtained from Δ_j by using a flip. Two sequences $\Delta_0, ..., \Delta_m$ and $\Delta'_0, ..., \Delta'_n$ are equivalent if their initial and final terms coincide i.e. there exists a homeomorphism φ preserving the punctures such that $\varphi(\Delta_0) \cong \Delta'_0$ and $\varphi(\Delta_m) \cong \Delta'_n$, where $\cong$ denotes the isotopy equivalence. Notice that any two (isotopy classes of) ideal triangulations are connected by a chain of flips (see [13] for an elementary proof), and hence $\overline{P_g^s}$ is indeed an groupoid. Moreover $\overline{P_g^s}$ is the groupoid $\mathcal{G}(\mathcal{M}_g^s, \mathcal{IT}(\Sigma_g^s))$ associated to the obvious action of the mapping class group $\mathcal{M}_g^s$ on the set of isotopy classes of ideal triangulations $\mathcal{IT}(\Sigma_g^s)$). One problem in considering $\overline{P_g^s}$ is that the action of $\mathcal{M}_g^s$ on $\mathcal{IT}(\Sigma_g^s)$) is *not free* but there is a simple way to remedy it. For instance in [18, 19] one adds the extra structure coming from fixing an *oriented arc* of the ideal triangulation. A second problem is that we want that the mapping class group action on the Teichmüller space extends to a groupoid action.

Consider now an ideal triangulation $\Delta = \Delta(\Gamma)$, where Γ is its dual fatgraph. A labelling of Δ is a numerotation of its edges $\sigma_\Gamma : E_\Gamma \to \{1, 2, ..., \sharp E_\Gamma\}$. Set now $\mathcal{LIT}(\Sigma_g^s)$) for the set of labeled ideal triangulations. The Ptolemy groupoid P_g^s of the punctured surface Σ_g^s is the groupoid generated by flips on $\mathcal{LIT}(\Sigma_g^s)$). The flip $F[e]$ associated to the edge $e \in E_\Gamma$ acts on the labellings in the obvious way:

$$\sigma_{F[e](\Gamma)}(f) = \begin{cases} \sigma_\Gamma(f), & \text{if } f \neq e' = F[e](e) \\ \sigma_\Gamma(e), & \text{if } f = e', \end{cases}$$

According to ([23] Lemma 1.2.b), if $2g-2+s \geq 2$ then any two labeled ideal triangulations are connected by a chain of flips, and thus P_g^s is indeed a groupoid. Moreover, this allows us to identify P_g^s with $\mathcal{G}(\mathcal{M}_g^s, \mathcal{LIT}(\Sigma_g^s))$.

Remark 42. In the remaining cases, namely Σ_0^3 and Σ_1^1, the flips are not acting transitively on the set of labeled ideal triangulations. In this situation an appropriate labelling consist in an oriented arc, as in [18]. The Ptolemy groupoid associate to this labeling has the right properties, and it acts on the Teichmüller space.

Proposition 33. *We have an exact sequence*

$$1 \to \mathcal{S}_{6g-6+3s} \to P_g^s \to \overline{P_g^s} \to 1,$$

where $\mathcal{S}_n$ denotes the symmetric group on n letters. Notice that $P_1^1 = \overline{P_1}^1$. If $(g, s) \neq (1, 1)$ then $\mathcal{M}_g^s$ naturally embeds in P_g^s as the group associated to the groupoid.

Proof. The first part is obvious. The following result is due to Penner ([23],Thm.1.3):

Lemma 48. *If $(g, s) \neq (1, 1)$ then $\mathcal{M}_g^s$ acts freely on $\mathcal{LIT}(\Sigma_g^s)$.*

Proof. A homeomorphism keeping invariant a labeled ideal triangulation either preserves the orientation of each arc or else it reverses the orientation of all arcs.

In fact once the orientation of an arc lying in some triangle is preserved, the orientation of the other boundary arcs of the triangle must also be preserved. Further in the first situation either the surface is Σ_0^3 (when $\mathcal{M}_0^3 = 1$) or else each triangle is determined by its 1-skeleton, and the Alexander trick shows that the homeomorphism is isotopic to identity. In the second case we have to prove that $(g, s) = (1, 1)$. Since the arcs cannot have distinct endpoints we have $s = 1$. Let Δ_1 be an oriented triangle and $D \subset \Delta_1$ be a 2-disk which is a slight retraction of Δ_1 into its interior. The image D' of D cannot lie within Δ_1 because the homeomorphism is globally orientation preserving while the orientation of the boundary of D' is opposite to that of $\partial\Delta$. Thus D' lies outside Δ_1 and the region between $\partial D'$ and $\partial\Delta_1$ is an annulus, so the complementary of Δ_1 consists of one triangle. Therefore $g = 1$.

Remark 43. The punctured torus Σ_1^1 has an automorphism which reverse the orientation of each of the three ideal arcs.

The case of the punctured torus is settled by the following:

Proposition 34. *Let $\Delta_{st} = \{\alpha_1, \alpha_2, \alpha_3\}$, where $\alpha_1 = (1,0), \alpha_2 = (1,1), \alpha_3 = (0,1)$ be the standard labeled ideal triangulation of the punctured torus $\Sigma_1^1 = \mathbb{R}^2/\mathbb{Z}^2 - \{0\}$.*

1. *If $\Delta = \{\alpha_{\sigma(1)}, \alpha_{\sigma(2)}, \alpha_{\sigma(3)}\}$ is flip equivalent to Δ_{st} then σ is the identity.*
2. *A mapping class which leaves invariant Δ_{st} is either identity or $-id \in SL(2,\mathbb{Z}) = \mathcal{M}_1^1$.*
3. *Let $\Delta = \{\gamma_1, \gamma_2, \gamma_3\}$ be an arbitrary ideal triangulation. Then there exists an unique $\sigma(\Delta) \in \mathcal{S}_3$ such that Δ is flip equivalent with the labeled diagram $\{\alpha_{\sigma(1)}, \alpha_{\sigma(2)}, \alpha_{\sigma(3)}\}$.*
4. *In particular if $\Delta = \varphi(\Delta_{st})$ then we obtain a group homomorphism $\sigma : SL(2,\mathbb{Z}) \to \mathcal{S}_3$, given by $\sigma(\varphi) = \sigma(\varphi(\Delta_{st}))$, whose values can be computed from:*

$$\sigma\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = (23), \quad \sigma\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = (12), \quad \sigma\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = (13).$$

We need therefore another labeling for Σ_1^1, which amounts to fix a *distinguished oriented edge (d.o.e.)* of the triangulation. The objects acted upon flips are therefore pairs (Δ, e), where e is the d.o.e. of Δ. A flip acts on the set of labeled ideal triangulations with d.o.e. as follows. If the flip leaves e invariant then the new d.o.e. is the old one. Otherwise the flip under consideration is $F[e]$, and the new d.o.e. will be the image e' of e, oriented so that the frame (e, e') at their intersection point is positive with respect to the surface orientation. The groupoid Pt_g^s generated by flips on (labeled) ideal triangulations with d.o.e. of is called the *extended Ptolemy groupoid*. Since any edge permutation is a product of flips (when $(g, s) \neq 1$) it follows that any two labeled triangulations with d.o.e. can be connected by a chain of flips.

The case of the punctured torus is subjected to caution again: it is more convenient to define the groupoid Pt_1^1 as that generated by iterated compositions of flips on the standard (labeled or not) ideal triangulation Δ_{st} of Σ_1^1 with a fixed d.o.e., for instance α_1. In fact proposition 34 implies that there are three distinct orbits of the flips on triangulations with d.o.e., according to the the position of the d.o.e. within Δ_{st}.

Remark 44. For all (g, s) we have an exact sequence:

$$1 \to \mathbb{Z}/2\mathbb{Z} \to \mathsf{Pt}_g^s \to P_g^s \to 1.$$

Moreover $\mathcal{M}_g^s \to P_g^s$ lifts to an embedding $\mathcal{M}_g^s \hookrightarrow \mathsf{Pt}_g^s$.

Remark 45. We can define the groupoid $\overline{\mathsf{Pt}_g^s}$ by considering flips on ideal triangulations with d.o.e. without labellings.

Remark 46. The kernel of the map $\mathcal{M}_1^1 \to P_1^1$ is the group of order two generated by $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Therefore any (faithful) representation of P_1^1 induces a (faithful) representation of $\mathrm{PSL}(2, \mathbb{Z})$.

Remark 47. One reason to consider P_g^s instead of $\overline{P_g^s}$ is that P_g^s acts on the Teichmüller space while $\overline{P_g^s}$ does not. The other reason is that $\mathcal{M}_g^s$ injects into P_g^s (if $(g, s) \neq (1, 1)$). The kernel of $\mathcal{M}_g^s \to \overline{P_g^s}$ is the image of the automorphism group $Aut(\Gamma)$ in $\mathcal{M}_g^s$.

Proof. An automorphism of Γ is a combinatorial automorphism which preserves the cyclic orientation at each vertex. Notice that an element of $Aut(\Gamma)$ induces a homeomorphism of Γ^t and hence an element of $\mathcal{M}_g^s$. Now, if φ is in the kernel then φ is described by a permutation of the edges i.e. an element of $\varphi_* \in \mathcal{S}_{\sharp E_\Gamma}$. One can assume that the orientations of all arcs are preserved by φ when $(g, s) \neq (1, 1)$. Then φ_* completely determines φ, by the Alexander trick. Further φ induces an element of $Aut(\Gamma)$ whose image in $\mathcal{S}_{\sharp E_\Gamma}$ is precisely φ_*. This establishes the claim. Notice that the map $Aut(\Gamma) \to \mathcal{S}_{\sharp E_\Gamma}$ is injective for most but not for all fatgraphs Γ. The fatgraphs Γ for which the map $Aut(\Gamma) \to \mathcal{S}_{\sharp E_\Gamma}$ fails to be injective are described in [20].

We can state now a presentation for Pt_g^s which is basically due to Penner ([23]):

Proposition 35. *Pt_g^s is generated by the flips $F[e]$ on the edges. The relations are:*

1. *Set J for the change of orientation of the d.o.e. Then*

$$F[F[e]e]F[e] = \begin{cases} 1, & \text{if } e \text{ is not the d.o.e.} \\ J, & \text{if } e \text{ is the d.o.e.} \end{cases}$$

2. $J^2 = 1$.
3. *Consider the pentagon from picture 11.2, and $F[e_j]$ be the flips on the dotted edges. Let $\tau_{(12)}$ denote the transposition interchanging the labels of the two edges e_1 and f_1 from the initial triangulation. Then we have:*

$$F[e_1]F[e_2]F[e_3]F[e_4]F[e_5] = \begin{cases} J\tau_{(12)}, & \text{if } e_1 \text{ is not the d.o.e.} \\ \tau_{(12)}, & \text{if } e_1 \text{ is the d.o.e.} \end{cases}$$

The action of $\tau_{(12)}$ on triangulations with d.o.e. is at follows: if none of the permuted edges e, f is the d.o.e. then $\tau_{(12)}$ leaves the d.o.e. unchanged. If the d.o.e. is one of the permuted edges, say e, then the new d.o.e. is f oriented such that e (with the former d.o.e. orientation) and f with the given d.o.e. orientation form a positive frame on the surface. Notice that $[F[e_1]F[e_2]F[e_3]F[e_4]F[e_5] = \tau_{(12)}$ even if f_1 is the d.o.e.
4. *If e and f are disjoint edges then $F[e]F[f] = F[f]F[e]$.*
5. *The relations in a $\mathbb{Z}/2\mathbb{Z}$ extension of the symmetric group, expressed in terms of flips. To be more specific, les us assume that the edges are labeled and the d.o.e. is labelled 0. Then we have:*

$$\tau_{(0i)}^2 = J,\ \tau_{(ij)}^2 = 1,\ \text{if } i, j \neq 0,\ \tau_{(st)}\tau_{(mn)} = \tau_{(mn)}\tau_{(st)} \text{ if } \{m,n\} \cap \{s,t\} = \emptyset,$$

$$\tau_{(st)}\tau_{(tv)}\tau_{(st)} = \tau_{(tv)}\tau_{(st)}\tau_{(tv)},\ \text{if } s, t, v \text{ are distinct.}$$

6. *$F[\tau(e)]\tau F[e] = \tau$, for any label transposition τ (expressed as a product of flips as above), which says that the symmetric group is a normal subgroupoid of P_g^s.*

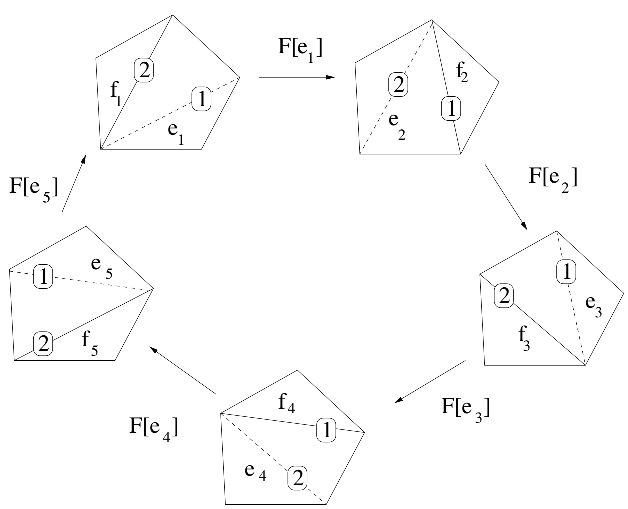


Fig. 11.2. The pentagon relation

Proof. We analyze first the case where labellings are absent:

Lemma 49. $\overline{P_g^s}$ *is generated by the flips on edges* $F[e]$. *The relations are:*

1. $F[e]^2 = 1$, *which is a fancy way to write that the composition of the flip on* $F[e](e)$ *with the flip on* e *is trivial.*
2. $F[e_1]F[e_2]F[e_3]F[e_4]F[e_5] = 1$, *where* $F[e_i]$ *are the flips considered in the picture 11.2.*
3. *Flips on two disjoint edges commute each other.*

Proof. This result is due to Harer (see [12]). It was further exploited by Penner ([22, 23]).

The complete presentation is now a consequence of the two exact sequences from proposition 33 and remark 44, relating $\overline{P_g^s}$, P_g^s and P_g^s.

Remark 48. By setting $J = 1$ above we find the presentation of P_g^s, with which we will be mostly concerned in the sequel.

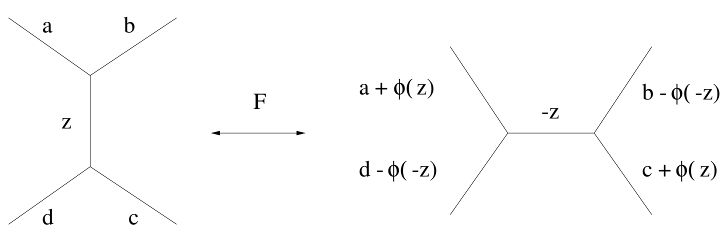
11.4 The mapping class group action on the Teichmüller spaces

In order to understand the action on $\mathcal{T}_g^s$ we to consider also $\mathcal{T}_{g,s;or}$.

The action of $\mathcal{M}_g^s$ on the Teichmüller space extends to an action of P_g^s to $\mathcal{T}_g^s$. Geometrically we can see it as follows. An element of $\mathcal{T}_g^s$ is a marked hyperbolic surface S. The marking comes from an ideal triangulation. If we change the triangulation by a flip, and keep the hyperbolic metric we obtain another element of $\mathcal{T}_g^s$.

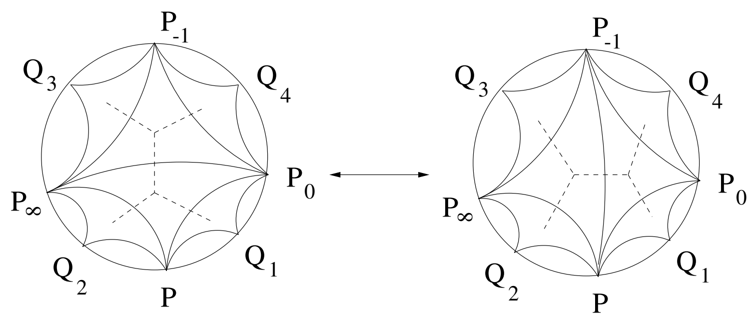
In the same way the $\mathcal{M}_{g,s}$ action on the Teichmüller space $\mathcal{T}_{g,s;or}$ extends to an action of the Ptolemy groupoid $P_{g,s}$. This action is very easy to understand in terms of coordinates. In more specific terms a flip between the graphs Γ and Γ' induces an analytic isomorphism $\mathbb{R}^{E_\Gamma} \to \mathbb{R}^{E_{\Gamma'}}$ by intertwining the coordinate systems t_Γ and $t_{\Gamma'}$. It is more convenient to identify $\mathbb{R}^{E_\Gamma}$ with a fixed Euclidean space, which is done by choosing a labelling $\sigma : E_\Gamma \to \{1, 2, ...\sharp E_\Gamma\}$ of its edges. Thus we have homeomorphism $\mathbf{t}_{\Gamma,\sigma} : \mathcal{T}_{g,s;or} \to \mathbb{R}^{\sharp E_\Gamma}$ given by $(\mathbf{t}_{\Gamma,\sigma}(S))_k = (\mathbf{t}_\Gamma(S))_{\sigma^{-1}(k)\in E_\Gamma}$. Further we can compare the coordinates $\mathbf{t}_{\Gamma,\sigma}$ and $\mathbf{t}_{F(\Gamma,\sigma)}$, for two labelled fatgraphs which are related by a flip. We can state:

Proposition 36. *A flip acts on the edge coordinates of a fatgraph as follows:*



where $\phi(z) = \log(1 + e^z)$. Here it is understood that the coordinates associated to the edges not appearing in the picture remain unchanged.

Proof. The flip on the graph corresponds to the following flip of ideal triangulations:



Then the coordinates a, b, c, d, z using the left-hand-side graph are the following cross-ratios: $a = [Q_3, P_\infty, P_0, P_{-1}]$, $b = [Q_4, P_{-1} P_\infty, P_0]$, $c = [Q_1, P_0, P_\infty, P]$, $d = [Q_2, P, P_0, P_\infty]$, $z = [P_{-1}, P_\infty, P, P_0]$. Let a', b', c', d', z' be the coordinates associated to the respective edges from the right-hand-side graph, which can again be expressed as cross-ratios as follows: $a' = [Q_3, P_\infty, P, P_{-1}]$, $b' = [Q_4, P_{-1}, P, P_0]$, $c = [Q_1, P_0, P_{-1}, P]$, $d = [Q_2, P, P_{-1}, P_\infty]$, $z = [P_\infty, P, P_0, P_{-1}]$. One uses for simplifying computations the half-plane model where, up to a Möbius transformation, the points P_{-1}, P_∞, P, P_0 are sent respectively into $-1, \infty, e^z$ and 0. The flip formulas follow immediately.

Remark 49. Similar computations hold for Penner's λ-coordinates on the decorated Teichmüller spaces. However the transformations of $\mathbf{R}^{6g-6+2s}$ obtained using λ-coordinates are *rational* functions.

Let us denote by $Aut^\omega(\mathbb{R}^m)$ the group of real analytic automorphisms of $\mathbb{R}^m$.

Corollary 19. *1. We have a faithful representation $\rho : \mathcal{M}_{g,s} \to Aut^\omega(\mathbb{R}^{6g-6+3s})$ induced by the $P_{g,s}$ action on the Teichmüller space $\mathcal{T}_{g,s;or}$ if $(g, s) \neq (1, 1)$.*
2. The groupoid $P_g^s \subset P_{g,s}$ leaves invariant the Teichmüller subspace $\mathcal{T}_g^s \subset \mathcal{T}_{g,s;or}$. Therefore the formula given in proposition 36 above for the flip actually yields a representation of P_g^s into $Aut^\omega(\mathbb{R}^{6g-6+2s})$. The restriction to the mapping class groups is a faithful representation $\rho : \mathcal{M}_g^s \to Aut^\omega(\mathbb{R}^{6g-6+2s})$ if $(g, s) \neq (1, 1)$, and a faithful representation of $\mathrm{PSL}(2, \mathbb{R})$ *when $(g, s) = (1, 1)$.*

Proof. The representation of $\mathcal{M}_{g,s}$ (respectively $\mathcal{M}_g^s$) is injective because the mapping class group acts effectively on the Teichmüller space. Therefore if the class of any (marked) Riemann surface is preserved by a homeomorphism then this homeomorphism is isotopic to the identity.

The invariance of the subspace $\mathcal{T}_g^s \subset \mathcal{T}_{g,s;or}$ by flips is geometrically obvious, but we write it down algebraically for further use. This amounts to check that the linear equations $t_\gamma = 0$, for $\gamma \in F_\Gamma$ are preserved. Let γ be a left-hand-turn path, which intersects the part of the graph shown in the picture, say along the edges labeled a, z, b. Then the flip of γ intersects the new graph along the edges labeled by $a + \phi(z)$ and $b - \phi(-z)$. The claim follows from the equality $z = \phi(z) - \phi(-z)$. The remaining three cases reduces to the same equation.

Remark 50. There is a Pt_g^s-action on the Teichmüller space but it is not free, and actually factors through P_g^s.

Remark 51. Assume that there exists an element $\mathbf{r} \in \mathcal{T}_g^s$, which is fixed by *some* $\psi \in \mathcal{M}_g^s$, i.e. $\varphi(\psi)(\mathbf{r}) = \mathbf{r}$. Then $\mathbf{r}$ is contained in some codimension two analytic submanifold $Q_g^s \subset \mathcal{T}_g^s$, and for a given $\mathbf{r}$ its isotropy group is finite. This is a reformulation of the fact that $\mathcal{M}_g^s$ acts properly discontinuously on the Teichmüller space with finite isotropy groups corresponding to the Riemann surfaces with non-trivial automorphism groups (biholomorphic). Moreover the locus of Riemann surfaces with automorphisms is a proper complex subvariety of the Teichmüller space, corresponding to the singular locus of the moduli space of curves.

11.5 Deformations of the mapping class group representations

We want to consider deformations of the tautological representation $\rho = \rho_0$ of $\mathcal{M}_g^s$ obtained in the previous section. We first restrict ourselves to deformations $\rho_h : \mathcal{M}_g^s \to Aut^\omega(\mathbb{R}^{6g-6+2s})$ satisfying the following requirements:

1. The deformation ρ_h extends to the Ptolemy groupoid P_g^s. In particular ρ_h is completely determined by $\rho_h(F)$ and $\rho_h(\tau_{(ij)})$.
2. The image of a permutation $\rho_h(\tau_{(ij)})$ is the automorphism of $\mathbb{R}^{6g-6+2s}$ given by the permutation matrix $P_{(ij)}$, which exchanges the i-th and j-th coordinates.
3. The image $F_h = \rho_h(F)$ of a flip has the same form as for $\rho_0(F)$, namely that given in the picture from proposition 36, but with a deformed function $\phi = \phi_h$, with $\phi_0 = \log(1 + e^z)$.
4. The linear subspace $\mathcal{T}_g^s \subset \mathcal{T}_{g,s;or}$ is invariant by ρ_h.

Proposition 37. *The real function $\phi : \mathbb{R} \to \mathbb{R}$ yield a deformation of the mapping class groups (respectively the Ptolemy groupoids) if and only if it satisfies the following functional equations:*

$$\phi(x) = \phi(-x) + x. \tag{11.1}$$

$$\phi(x + \phi(y)) = \phi(x + y - \phi(x)) + \phi(x). \tag{11.2}$$

$$\phi\left(\phi\left(x + \phi(y)\right) - y\right) = \phi(-y) + \phi(x). \tag{11.3}$$

Proof. The first equation is equivalent to the invariance of the linear equations defining the cusps. The other two equations follow from the cumbersome but straightforward computation of terms involved in the pentagon equation.

11.6 Belyi Surfaces

Let S be a compact Riemann surface. It is well known that there exists a non-constant meromorphic function on S, $\phi : S \to \mathbb{CP}^1$.

Definition 31. *The Riemann surface S is a* Belyi surface *if there exists a ramified covering $\phi : S \to \mathbb{CP}^1$, branched over 0, 1 and ∞.*

A surprising theorem of Belyi ([1]) states that:

Theorem 32. *S is a Belyi surface if and only if it is defined over $\overline{\mathbb{Q}}$ i.e. as a curve in $\mathbb{CP}^2$ its minimal polynomial lies over some number field.*

Following [22, 20] we can characterize Belyi surfaces in terms of fat graphs as follows:

Theorem 33. *A Riemann surface S can be constructed as $S(\Gamma) = S(\Gamma, \mathbf{0})$ for some trivalent fatgraph Γ if and only if S is a Belyi surface.*

Proof. We prove first:

Lemma 50. *Let $G \subset \mathrm{PSL}(2, \mathbb{Z})$ be a finite index torsion-free subgroup. Then $\mathbb{H}^2/G = S(\Gamma)$ for some trivalent fatgraph Γ.*

Proof. Remark that $A = \{z \in \mathbb{H}^2; 0 < \Re(z) < 1, |z| > 1, |z-1| > 1\}$, is a fundamental domain for $\mathrm{PSL}(2, \mathbb{Z})$, with the property that three copies of it around $\omega + 1$ fit together to give the ideal marked triangle. These three copies are equivalent by means of an order three elliptic element γ of $\mathrm{PSL}(2, \mathbb{Z})$.

A fundamental domain for G is composed of copies of A, and since G is torsion free the three copies A, $\gamma(A)$ and $\gamma^2(A)$ are not equivalent under G, thus they can all be included in the fundamental domain for G. In particular it exists a fundamental domain B for G which is made of copies of the ideal triangle I and hence it is naturally triangulated. Consider the graph Γ dual to this triangulation, which takes into account the boundary pairings, and which inherits an orientation from $\mathbb{H}^2/G$. Then $\mathbb{H}^2/G = S(\Gamma)$.

Lemma 51. *S is a Belyi surface if and only if we can find finitely many points on S, $\{p_1, \ldots, p_k\}$, such that $S - \{p_1, \ldots, p_k\}$ is isomorphic to $\mathbb{H}^2/G$, where G is a finite index torsion free subgroup of $\mathrm{PSL}(2, \mathbb{Z})$.*

Proof. Set $\Gamma(2) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PSL}(2, \mathbb{Z}) \mid \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \pmod 2 \right\}$. Then $F = \{z \in \mathbb{H}^2; 0 < \Re(z) < 2, |z - 1/2| > \frac{1}{2}, |z - 3/2| > \frac{1}{2}\}$ is a fundamental domain for $\Gamma(2)$ composed of 2 ideal triangles glued along a common edge. Thus the 3-punctured sphere $\mathbb{CP}^1 - \{0, 1, \infty\}$ is $\mathbb{H}^2/\Gamma(2)$. Moreover each ideal triangle is composed of three copies of the fundamental domain of $\mathrm{PSL}(2, \mathbb{Z})$. Therefore, the 3-punctured sphere is a six-fold branched covering of $\mathbb{H}^2/\mathrm{PSL}(2, \mathbb{Z})$.

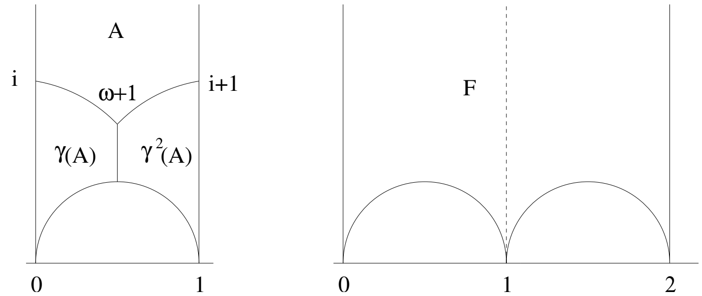


Fig. 11.3. Fundamental domains for $\mathrm{PSL}(2,\mathbb{Z})$ and $\Gamma(2)$

Let S be a compact Riemann surface. If S is a Belyi surface then $S - \{p_1,\ldots,p_k\}$ is a regular smooth finite degree covering of $\mathbb{H}^2/\Gamma(2)$ and thus $S - \{p_1,\ldots,p_k\} = \mathbb{H}^2/G$, where G is a finite-index subgroup of $\Gamma(2)$ (and hence of $\mathrm{PSL}(2,\mathbb{Z})$).

Conversely, if $S - \{p_1,\ldots,p_k\} = \mathbb{H}^2/G$, where G is a finite index torsion free subgroup of $\mathrm{PSL}(2,\mathbb{Z})$, then $S - \{p_1,\ldots,p_k\}$ is a finite-degree branched covering of $\mathbb{H}^2/\mathrm{PSL}(2,\mathbb{Z})$, which is a sphere with one cusp and 2 ramification points. Therefore, if we remove the 2 ramification points and their pre-images, we get that $S - \{p_1,\ldots,p_k,\ldots,p_n\}$ is a regular smooth finite-degree covering of the 3-punctured sphere, i.e. a Belyi surface.

These lemmas show that any Belyi surface can be constructed out of some fatgraph.

Conversely the fundamental polygon constructed for $G(\Gamma)$ is composed of copies of the ideal triangle. By decomposing each ideal triangle into three copies of the fundamental domain for $\mathrm{PSL}(2,\mathbb{Z})$, we see that $G(\Gamma)$ can be embedded as a finite-index torsion free subgroup of $\mathrm{PSL}(2,\mathbb{Z})$.

11.7 The geometry of the Teichmüller space

11.7.1 Symplectic structures for the Teichmüller space of punctured surfaces

The Teichmüller space $\mathcal{T}_g^s$ has a natural structure of complex manifold. Let us recall some of its features. Suppose that the Riemann surface S is uniformized by the Fuchsian group $G \subset \mathrm{PSL}(2,\mathbb{R})$.

One considers first the vector space $Q(S) = Q(G)$ of integrable holomorphic quadratic differentials on S. An element $\varphi \in Q(S)$ is a holomorphic function $\varphi(z)$ on $\mathbb{H}^2$ satisfying $\varphi(\gamma(z))\gamma'(z)^2 = \varphi(z)$ for all $\gamma \in G$, and $\int_F |\varphi|$ is finite, where F is a fundamental domain for G. Then φ induces a symmetric tensor of type $(2,0)$ on S.

Let then $M(S)$ be the space of G-invariant Beltrami differentials. These are measurable, essentially bounded functions $\mu : \mathbb{H}^2 \to \mathbb{C}$ satisfying $\mu(\gamma(z))\frac{\overline{\gamma'(z)}}{\gamma'(z)} = \mu(z)$ for all $\gamma \in G$, and hence define a $(-1,1)$ tensor on S.

There is a natural pairing $(,) : M(S) \times Q(S) \to \mathbb{C}$ given by $(\mu, \varphi) = \int_F \mu\varphi$, with null space $N(S) \subset M(S)$ which induces a duality isomorphism between $M(S)/N(S)$ and $Q(S)$.

The holomorphic cotangent space at the point $[S] \in \mathcal{T}_g^s$ is identified with $Q(S)$ and thus the tangent space is naturally isomorphic to $M(S)/N(S)$. Weil introduced a hermitian product on $Q(S)$ defined in terms of the Petersson product for automorphic forms. This yields the Weil-Petersson (co)metric on $\mathcal{T}_g^s$:

$$< \varphi, \psi >= \frac{1}{2}\mathsf{Re}\int_{\mathbb{H}^2/G} \varphi\overline{\psi}\,(\mathsf{Im} z)^{-2}, \quad \text{for } \varphi, \psi \in Q(S).$$

Remark 52. The Weil-Petersson metric is Kähler, it has negative holomorphic sectional curvature and is invariant under the action of the mapping class group.

The Kähler form of the Weil-Petersson metric is a symplectic form ω_{WP}. In the case of closed surfaces Wolpert ([32]) derived a convenient expression for ω_{WP} in terms of Fenchel-Nielsen coordinates:

$$\omega_{WP} = -\sum_j d\,\tau_j \wedge d\,l_j.$$

Recall that a pair of pants $\Sigma_{0,3}$ has a hyperbolic structure with geodesic boundary. The lengths $l_j \in \mathbb{R}_+$ of the boundary circles can be arbitrarily prescribed. To each decomposition of S into pairs of pants $P_1, ..., P_{2g-2}$ we have therefore associated the lengths of their boundary geodesics $l_1, ..., l_{3g-3}$. In fact given pairs of pants, not necessarily distinct, P_1 and P_2 with boundary circles c_1 on P_1 and c_2 on P_2, of the same length we can glue the pants by identifying c_1 with c_2 by an isometry. The hyperbolic metric extends over the connected sum. Therefore we can glue together the pants $P_1, ..., P_{2g-2}$ to obtain the Riemann surface S. If a length $l = 0$ then this corresponds to the situation where the surface has a cusps. We can therefore extend this description to punctured surfaces Σ_g^n with cusps at punctures. The pants decomposition is specified therefore by $3g-3+n$ geodesics on S. Each boundary circle c belongs to two pairs of pants P_j and P_k. The geodesics joining the circles of P_j to the circles of P_k intersect c into two points. The parameter τ_j is the (signed) hyperbolic distance between these two points. The parameters (τ_j, l_j) are the Fenchel Nielsen coordinates on $\mathcal{T}_g^s$.

Fricke and Klein established that, if one carefully choose the curves $\gamma_1, ..., \gamma_{6g-6+2n}$ then the associated lengths l_j can also give local coordinates on $\mathcal{T}_g^s$. A typical example is to pick up first the curves $\gamma_1, ..., \gamma_{3g-3+n}$ arising from a pants decomposition, and then a dual pants decomposition obtained as follows. Consider the pieces of geodesics which yield the canonical points on the circles, and then identify combinatorially the canonical points. We obtained this way a family of closed loops $\gamma_{3g-3+n+1}, ..., \gamma_{6g-6+2n}$. Wolpert ([33],Lemma 4.2, 4.5) expressed the Kähler form in these coordinates:

Lemma 52. *Assume that $l_1, ..., l_{6g-6+2n}$ provide local coordinates on $\mathcal{T}_g^s$ and denote:*

$$\alpha^{jk} = \sum_{p \in \gamma_j \cap \gamma_k} \cos\theta_p,$$

where θ_p is the angle between the geodesic γ_j and γ_k at the point p. Let $W = (w_{jk})_{j,k}$ be the inverse of the matrix $A = (\alpha^{jk})_{j,k}$. Then the Weil-Petersson form is:

$$\omega_{WP} = -\sum_{j<k} w_{jk}\, d\, l_j \wedge d\, l_k.$$

11.7.2 Poisson structure for the Teichmüller space of surfaces with boundary

Let G be a connected Lie group, which will be most of the time $\mathrm{PSL}(2, \mathbb{R})$ in this section. Set $M(\Sigma, G) = \mathrm{Hom}(\pi_1(\Sigma), G)/G$ for the moduli space of representations of the fundamental groups.

Goldman ([8]) proved that $M(\Sigma, G)$ is endowed with a natural symplectic structure, whenever Σ is a closed oriented surface. Moreover Fock and Rosly ([6]) was able to show more generally that there is a Poisson structure on $M(\Sigma, G)$, even in the case when Σ is a surface with boundary. Furthermore the symplectic leaves of this structures are precisely the singular submanifolds $M(\Sigma, G)_{\lambda_1,...,\lambda_s}$, where λ_j is the conjugacy class of the holonomy around the j-th boundary component.

Notice also that Zocca have shown that $M(\Sigma, G)$ has a pre-symplectic structure, whose restriction to the symplectic leaves is the symplectic form.

11.7.3 Penner's decorated Teichmüller space

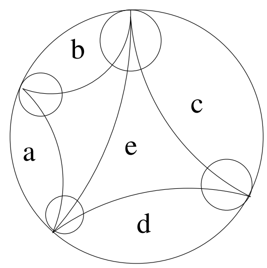
Penner ([22, 23] considered the space $\widetilde{\mathcal{T}}_g^s$ of cusped Riemann surfaces endowed with a horocycle around each puncture, and called it the decorated Teichmüller space. There is a natural family of coordinates (called lambda lengths), associated to the edges of an ideal triangulation $\Delta = \Delta(\Gamma)$ of the surface. For each such edge e one puts $\lambda_e = \sqrt{2exp(\delta)}$, where δ is the signed hyperbolic distance between the two horocycles centered at the two endpoints of the edge e. The sign convention is that $\lambda_e > 0$ if the horocycles are disjoint. It is not difficult to see that these coordinates give a homeomorphism $\widetilde{\mathcal{T}}_g^s \to \mathbb{R}^{6g-6+3s}$. The map which forgets the horocycles $\pi : \widetilde{\mathcal{T}}_g^s \to \mathcal{T}_g^s$ is a fibration having $\mathbb{R}_+^s$ as fibers. Moreover:

Lemma 53. *The projection π is expressed in terms of Penner and Fock coordinates as follows:*

$$\pi\left((\lambda_e)_{e\in\Delta(\Gamma)}\right) = \left(\log\frac{\lambda_a\lambda_c}{\lambda_b\lambda_d}\right)_{e\in\Delta(\Gamma)},$$

where, for each edge e we considered the quadrilateral of edges a, b, c, d, uniquely determined by the following properties:

- *the cyclic order a, b, c, d is consistent with the orientation of Σ_g^s.*
- *e is the diagonal separating a, b from c, d (see the figure 53).*
- *each triangle of Δ has an orientation inherited from Σ_g^s, in particular the edge e is naturally oriented. We ask that a (and d) be adjacent to the start-point of e, while b and c is adjacent to the endpoint of e.*



Proof. The proof is a mere calculation.

Proposition 38. *The pull-back $\pi^*\omega_{WP}$ of the Weil-Petersson form on the decorated Teichmüller space $\widetilde{\mathcal{T}}_g^s$ is given in Penner's coordinates as:*

$$\pi^*\omega_{WP} = -2\sum_{T\subset\Delta} d\log\lambda_a\wedge d\log\lambda_b + d\log\lambda_b\wedge d\log\lambda_c + d\log\lambda_c\wedge d\log\lambda_a,$$

where the sum is over all triangles T in Δ whose edges have lambda lengths a, b, c in the cyclic order determined by the orientation of Σ_g^s.

Proof. See [24], Appendix A.

Remark 53. For dimensional reasons the pre-symplectic form $\pi^*\omega_{WP}$ is degenerate.

Proposition 39. *The Poisson structure on $\mathcal{T}_{g,s;or}$ is given by the following formula in the Fock coordinates (t_e):*

$$P_{WP} = \sum_{T\subset\Delta} dt_a\wedge dt_b + dt_b\wedge dt_c + dt_c\wedge dt_a,$$

where the sum is over all triangles T in Δ whose edges are a, b, c in the cyclic order determined by the orientation of Σ_g^s. This Poisson structure is degenerate. Moreover $\mathcal{T}_g^s\subset\mathcal{T}_{g,s;or}$ is a symplectic leaf and hence the restriction of P_{WP} is the Poisson structure dual to the Weil-Petersson symplectic form ω_{WP}.

References

1. G. V. Belyi, *On Galois Extensions of a Maximal Cyclotomic Polynomial*, Math. USSR Izvestia 14 (1980), 247-256.
2. F.Bonahon, *Shearing hyperbolic surfaces, bending pleated surfaces and Thurston's symplectic form*, Ann. Fac. Sci. Toulouse Math. (6) 5 (1996), no. 2, 233–297.
3. R. Brooks, H. M. Farkas, I. Kra, *Number Theory, theta identities, and modular curves*, Contemp. Math. 201(1997), pp.125-154.
4. L.Chekhov, V.Fock, *Quantum Teichmüller spaces*, Theoret. Math. Phys. 120 (1999), no. 3, 1245–1259.
5. V.Fock, *Dual Teichmüller spaces*, math.dg-ga/9702018.
6. V.Fock, A.Rosly *Poisson structure on moduli of flat connections on Riemann surfaces and the r-matrix*, Moscow Seminar in Mathematical Physics, 67–86, Amer. Math. Soc. Transl. Ser. 2, 191, 1999.
7. J.Gilman, *Two-generator discrete subgroups of $PSL(2, R)$*, Mem. Amer. Math. Soc. 117(1995), no. 561, 204 pp.
8. W.Goldman, *The symplectic nature of fundamental groups of surfaces*, Advances in Math. 54 (1984), 200–225.
9. W.Goldman, *Topological components of spaces of representations*, Invent. Math. 93 (1988), no. 3, 557–607.
10. W.Goldman, *Invariant functions on Lie groups and Hamiltonian flows of surface group representations*, Invent. Math. 85 (1986), no. 2, 263–302.
11. W.Goldman, *Ergodic theory on moduli spaces*, Ann. of Math. (2) 146 (1997), no. 3, 475–507.
12. J.Harer *The virtual cohomological dimension of the mapping class group of an orientable surface*, Invent.Math. 84(1986), 157-176.
13. A.Hatcher, *On triangulations of surfaces*, Topology Appl. 40(1991), 189-194.
14. R. Kashaev, *Quantization of Teichmüller spaces and the quantum dilogarithm*, Lett. Math. Phys. 43 (1998), no. 2, 105–115.
15. R. Kashaev, *The Liouville central charge in quantum Teichmüller theory*, Proc. Steklov Inst. Math. 1999, no. 3 (226), 63–71.
16. R. Kashaev, *The pentagon equation and mapping class groups of surfaces with marked points*, Theoret. and Math. Phys. 123 (2000), no. 2, 576–581.
17. R.Kashaev, *On the spectrum of Dehn twists in quantum Teichmüller theory*, Physics and combinatorics, 2000 (Nagoya), 63–81, World Sci. Publishing, River Edge, NJ, 2001.
18. L. Mosher, *Mapping class groups are automatic*, Ann. of Math. 142(1995), 303–384.
19. L. Mosher, *A user's guide to the mapping class group: once punctured surfaces*, Geometric and computational perspectives on infinite groups, 101–174, DIMACS, 25, A.M.S., 1996.
20. M.Mulase, M.Penkava, *Ribbon graphs, quadratic differentials on Riemann surfaces, and algebraic curves defined over $\overline{\mathbf{Q}}$*, Asian J. Math. 2 (1998), no. 4, 875–919.

21. G.Moore, N.Seiberg, *Classical and quantum conformal field theory*, Comm. Math. Phys. 123 (1989), no. 2, 177–254.
22. R.Penner, *The decorated Teichmüller space of punctured surfaces*, Commun. Math. Phys 113(1987), 299-339.
23. R.Penner, *Universal constructions in Teichmüller theory*, Adv.Math 98(1993), 143-215.
24. R.Penner, *Weil-Petersson volumes*, J. Diff. Geometry 35 (1992), 559–608.
25. G.Rosenberger, *Eine Bemerkung zu einer Arbeit von T. Jorgensen*, Math. Zeitschrift 165 (1979), no. 3, 261–266.
26. W.Thurston, *Three-dimensional geometry and topology*, Vol. 1., Ed. Silvio Levy, Princeton Mathematical Series, 35, 1997.
27. W.Thurston, *Minimal stretch maps between hyperbolic surfaces*, preprint 1986, available at `math.GT/9801039`.
28. H.Verlinde, *Conformal field thoery, two-dimensional gravity and quantization of Teichmüller space*, Nuclear Phys. B, 337(1990), 652-680.
29. E.Witten, $2+1$-*dimensional gravity as an exactly soluble system*, Nuclear Phys. B 311 (1988/89), no. 1, 46–78.
30. E.Witten, *Topology-changing amplitudes in* $(2{+}1)$-*dimensional gravity*, Nuclear Phys. B 323 (1989), no. 1, 113–140.
31. E.Witten, *Quantum field theory and the Jones polynomial*, Comm. Math. Phys. 121 (1989), no. 3, 351–399
32. S.Wolpert, *On the Weil-Petersson geometry of the moduli space of curves*, Amer.J.Math. 107 (1985), 969–997.
33. S.Wolpert, *On the symplectic geometry of deformations of a hyperbolic surface*, Ann. of Math. 117 (1983), 207–234.

12

Riemann surfaces

12.1 Generalities

Note that Radó proved in 1925 that all Riemann surfaces are second-countable. This cannot be extended to higher dimensions, as Prüfer's complex surface is not second-countable.