Curves on the boundary of handlebodies

José Pedro Alves Quintanilha

Born 17th June 1990 in Vila Nova de Gaia, Portugal

5th October 2017

Master's Thesis Mathematics

Advisor: Dr. Sebastian Hensel

Second Advisor: Prof. Dr. Ursula Hamenstädt

Mathematisches Institut

Mathematisch-Naturwissenschaftliche Fakultät der

Rheinischen Friedrich-Wilhelms-Universität Bonn

Contents

0 Introduction

0.1 Objective and motivation

A handlebody V_g is a certain type of compact 3-manifold, having a closed oriented surface of genus g as its boundary, and whose fundamental group is free on g generators. As is true for all 3-manifolds, any conjugacy class of $\pi_1(V_g)$ can be realized as a curve without self-intersections, but the same is not true of ∂V_g. The map $\pi_1(\partial V_g) \to \pi_1(V_g)$ is well-known to be surjective, but one can ask the natural question: Which homotopy classes of curves in V_g (without basepoint) can be realized in ∂V_g *without self-intersections*?

By choosing an isomorphism $\phi : \pi_1(V_g) \to \mathcal{F}_g$ (where $\mathcal{F}_g$ is the abstract free group on g generators), we define a conjugacy class of $\mathcal{F}_g$ to be geometric if it can be realized as a simple closed curve in ∂V_g. It turns out that the property of being geometric is independent of the choice of ϕ (Proposition 2.2). This tells us that geometricity makes sense as an actual group-theoretical property of conjugacy classes in a free group, rather than being just a geometric aspect of curves in a handlebody. The definition of geometricity can actually be generalized to finite sets of conjugacy classes (multiclasses) of $\mathcal{F}_g$, a multiclass being geometric if there exist pairwise disjoint simple closed curves (a multicurve) on ∂V_g realizing all conjugacy classes.

Knowing that a conjugacy class is geometric allows one to construct 3-manifolds with boundary that have prescribed fundamental groups. For example, if the conjugacy class $w \subseteq \mathcal{F}_g$ is geometric, then the manifold obtained by attaching a "thickened 2-cell" along a regular neighbourhood of a curve on ∂V_g representing w has fundamental group $\mathcal{F}_g/\langle\langle w \rangle\rangle$. Taking this idea a step further, suppose multiclass $w_1, \ldots, w_n$ can be represented by disjoint simple closed curves on ∂V_g cutting ∂V_g into a collection of punctured spheres (that is, the curves form a Heegaard diagram). Then we can obtain a closed 3-manifold with fundamental group $\mathcal{F}_g/\langle\langle w_1, \ldots, w_n \rangle\rangle$ by attaching to V_g a second handlebody of the same genus along a suitable homeomorphism of the boundaries (for an introductory account on Heegaard splittings, see for example [8]).

The main goal of the present thesis is to establish a criterion to determine whether a multiclass is geometric. The answer to this question has been attributed (for instance in [1]) to H. Zieschang, the original article being in Russian [13]. This is used in [1] to characterize the broader concept of virtual geometricity: A multiclass $\{w_1, \ldots, w_n\}$ is virtually geometric if there is a finite index subgroup $G \leq F_g$ such that the maximal cyclic subgroups of G that are conjugate into the w_j define a geometric multiclass A of G. In the case of a single conjugacy class (say of $w \in F_g$), glueing two handlebodies modelling G along a neighbourhood of a multicurve representing A in the boundary yields a manifold whose fundamental group embeds in $D_g(w) = \mathcal{F}_g \underset{\langle w \rangle}{*} \mathcal{F}_g$. In [3], this was exploited for finding surface subgroups of one-ended groups of the form $D_g(w)$, after a question posed by Gromov.

The fundamental result, which we shall prove, is the following.

Theorem 2.31 (Zieschang)**.** *Let A be a disk-busting multiclass of $\mathcal{F}_g$ and $B_\circ$ a free basis for $\mathcal{F}_g$ such that $l_{B_\circ}(A)$ is minimal among all bases. Then A is geometric if and only if $\mathrm{Wh}_{B_\circ}(A)$ has an admissible planar embedding.*

The above concepts shall be rigorously defined in the body of the text, but as an overview:

1. A being disk-busting means that there is no factorization $\mathcal{F}_g = G_1 * G_2$ with each $w \in A$ represented in G_1 or G_2,

2. $l_{B_\circ}(A)$ denotes the length of the reduced multiword representing A in the basis $B_\circ$,

3. $\mathrm{Wh}_{B_\circ}(A)$ is the Whitehead graph, which is constructed by reading off that multiword (see Definition 2.9),

4. Admissibility of a planar embedding of $\mathrm{Wh}_{B_\circ}(A)$ is a compatibility condition between the embedding and an extra piece of structure that Whitehead graphs possess (see Definition 2.13).

We will also present some results relating different isotopy classes of simple closed curves in ∂V representing a fixed conjugacy class w. We construct the induced curve graph $\mathcal{C}_w$, whose vertices are such isotopy classes, with an edge between two classes if they can be made disjoint. When w is the trivial conjugacy class, this is just the disk graph, which has been studied before [4, 5]. We will see how $\mathcal{C}_w$ is related to the handlebody group (see Definition 3.12) and present some results about its connected components.

0.2 Outline

In Part 1, we define and state basic properties of handlebodies V_g, and introduce disk systems, collections of properly embedded disks in V_g that are the geometric incarnation of free bases of $\pi_1(V_g)$. These allow us to represent the class of a curve in V_g by a word that can be read off from the intersection pattern of the curve with the disks. The exposition is loosely based on [8].

Part 2 is when we prove Theorem 2.31. After introducing geometricity of multiclasses of $\mathcal{F}_g$ (Section 2.1), we tackle the problem of characterizing geometricity of a multiword Λ (Section 2.2), that is, determining whether there exists a multicurve on ∂V_g whose intersection pattern with some disk system is exactly the one given by Λ. The key lies in a certain graph, the word graph, whose definition is a direct adaptation of the Whitehead graph to the context of words. We answer this question with a criterion involving planarity of the word graph of Λ, and an extra compatibility condition (Proposition 2.14). This additional requirement is related to the ribbon structure (Definition 2.12) induced on the word graph by such a planar embedding, and prompts the question of determining the minimal genus of a closed oriented surface where a given ribbon graph embeds. As this is a rather self-contained problem, it has been separated into Appendix A.

We then move towards translating a geometricity test for a multiclass A into a geometricity test for some multiword Λ representing A, in an appropriate basis. In Section 2.3, we reduce the problem to the case where A is disk-busting. An algorithm for testing separability of A and decomposing it into disk-busting parts is presented, closely following [9]. In the process, we introduce Whitehead moves (Definition 2.20), a certain type of basis change for $\mathcal{F}_g$, and show how they can be realized geometrically as changes of disk system. Still following [9], we also explain how separability of A can be realized geometrically, and this is used in proving Proposition 2.29, the main result of this section. In Section 2.4, we import an algebraic fact about Whitehead moves (Proposition 2.30) to finally prove Theorem 2.31.

We begin Part 3 with some preliminary concepts about curves on surfaces (Section 3.1), and in Section 3.2 we introduce the induced curve graph $\mathcal{C}_w$ and give an alternative characterization of its edges (Theorem 3.8). In Section 3.3, we explain how $\mathcal{C}_w$ relates to Dehn Twists and the handlebody group, and give a necessary condition for two isotopy classes to be in the same connected component, in the case where w is not in the commutator of $\mathcal{F}_g$ (Theorem 3.17).

0.3 Preliminaries

0.3.1 Geometric topology

Throughout the text, we will shift between the topological and smooth settings as convenient, always assuming submanifolds intersect transversely (which is true up to modification by a small homotopy). We will never work with manifolds or cell complexes of dimension greater than 3, so we can employ the techniques of piecewise-linear topology. In particular, we will make extensive use, often implicitly, of the existence of regular neighbourhoods of cell complexes embedded in manifolds.

Definition 0.1. *Let X be a subset of a manifold M. A (closed)* ***regular neighbourhood*** *of X is a neighbourhood that is homeomorphic to a mapping cylinder*

$$M_f = (Y \times [0,1] \amalg X)/(y,1) \sim f(y),$$

for $Y \subseteq M$ a codimension 1 submanifold and f a map $Y \to X$. Moreover, the inclusion $M_f \hookrightarrow M$ maps X to X and $Y \times \{0\}$ to Y in the obvious manner.

Another construction that will be routinely performed on a manifold is that of **cutting along** a codimension 1 submanifold.

Proposition 0.2. *Let M be a manifold (without boundary) and $Y \subseteq M$ a codimension 1 submanifold. Then there exists a manifold with boundary $M \backslash Y$ and a natural map $M \backslash Y \to M$, whose restriction to the interior of $M \backslash Y$ is a homeomorphism onto the complement of Y in M, and $\partial(M \backslash Y)$ maps onto Y as the sphere bundle of its normal bundle.*

This construction is unique, and generalizes to manifolds with boundary, in which case we need to assume that Y is properly embedded in M.

Definition 0.3. *Given an embedding $f : Y \hookrightarrow M$ between manifolds with boundary, we say that Y is* ***properly embedded*** *in M if $f(Y) \cap \partial M = f(\partial Y)$. In the smooth setting, we further require that this intersection be transverse.*

0.3.2 Low-dimensional topology

Definition 0.4. *A* ***simple closed curve*** *in a manifold is an embedded submanifold that is homeomorphic to S^1.*

Frequently, these will also come equipped with a choice of orientation, or a distinguished basepoint.

Definition 0.5. *A simple closed curve in a closed surface S is called* ***essential*** *if it does not bound an embedded disk in S. For a 3-manifold M with boundary, a properly embedded disk D in M is called essential if ∂D is essential in ∂M.*

Definition 0.6. *Let a, b be isotopy classes of simple closed curves in a surface. Their* ***geometric intersection number*** *$i(a, b)$ is the minimal number of intersection points of a simple closed curve representing a, and one representing b.*

Definition 0.7. *Two (transverse) simple closed curves α, β in a surface are in* ***minimal position*** *if $|\alpha \cap \beta| = i([\alpha], [\beta])$.*

We will make use of the well-known classification of surfaces, and occasionally of the bigon criterion (Proposition 1.7 of [2]):

Definition 0.8. *A* ***bigon*** *in a surface S is a pair of embedded arcs intersecting exactly at their endpoints, whose union bounds an embedded disk in S. Two (transverse) simple closed curves α, β in a surface are said to form a bigon if there is an embedded disk in S whose boundary is the union of an arc form α, and one from β.*

Proposition 0.9 (The bigon criterion)**.** *Two simple closed curves in a surface are in minimal position if and only if they do not form a bigon.*

0.3.3 Graph theory

We will appeal the following well-known result.

Theorem 0.10 (Kuratowski)**.** *A finite graph is planar if and only if it does not contain (as a subgraph) a subdivision of K_5, the complete graph on 5 vertices, or $K_{3,3}$, the complete bipartite graph on 6 vertices partitioned as 3+3.*

At some point, we will need to think of an edge e of a graph Γ as comprised of two segments, each adjacent to one of the endpoints of e. This is formalized as follows.

Definition 0.11. *A* ***half-edge*** *of a graph is an edge of it barycentric subdivision.*

If Γ is a directed graph, then its half-edges are also oriented, and we distinguish between **incoming** and **outgoing** half-edges, depending on whether they are oriented towards or away from their corresponding vertex of Γ. If h is a half-edge, we denote by $\overline{h}$ its **opposite**, that is, the other half-edge coming from the same edge of Γ.

0.4 Acknowledgements

I wish to thank Sebastian Hensel for accepting me as his student, for his insights and guidance, and for his enthusiasm, constantly challenging and motivating me throughout the past year.

My gratitude extends to my friends, particularly Mafalda Santos and João Lourenço, and to my mother and sister for their continued support.

1 Handlebodies

1.1 Definition and basic facts

The starting point of our discussion is a 3-manifold constructed as follows. Begin with a finite disjoint union $Q = \coprod_{j=1}^{n} D_j^3$ of copies of the 3-ball $D^3 := \{x \in \mathbb{R}^3 \mid \|x\| \leq 1\}$, and let $E_1, F_1, \ldots, E_m, F_m$ be pairwise disjoint, embedded disks in ∂Q. Moreover, for $1 \leq i \leq m$, let $f_i : F_i \to E_i$ be orientation-reversing homeomorphisms, where the disks inherit the induced orientation of D^3 on S^2. Finally, define $V = Q/\sim$ to be the quotient space obtained by "glueing the balls along the f_i", that is, by identifying $q \sim f_i(q)$ for all $q \in F_i$. The resulting space V is an oriented 3-manifold with boundary.

Definition 1.1. *A space V constructed in this manner is called a* ***handlebody*** *if it is connected.*

Notice that connectedness of V implies that $m \geq n - 1$.

Since the result of glueing two distinct balls along a pair of disks is homeomorphic to a ball, we may modify our construction of V by replacing two distinct balls in Q, and a pair of disks connecting them, with a single ball. If we do this $n - 1$ times, we get a construction using only one ball. Hence, we could just as well have defined a handlebody by starting with $Q = D^3$. This observation will be used in proving the following result.

Proposition 1.2. *Let V, V' be handlebodies constructed from n, n' balls (respectively), and let m, m' be the number of pairs of disks being glued in each. Then $V \cong V'$ if and only if $m - n = m' - n'$.*

Proof. ($\Leftarrow$) By the preceding remark, we may assume that $n = n' = 1$. Since at each step we decrease both m and n by 1 (and similarly for m', n'), this implies $m = m'$. We then use the following technical lemma, whose proof we omit.

Lemma 1.3. *Let $X, X' \subseteq S^2$ be finite disjoint unions of embedded disks, and $h : X \to X'$ an orientation-preserving homeomorphism. Then there exists a homeomorphism $H : S^2 \to S^2$ extending h.*

Using the notation in the definition of handlebody, this implies our statement by taking $X = \bigcup_i (E_i \cup F_i)$, $X' = \bigcup_i (E_i' \cup F_i')$. Define h by mapping each F_i to F_i' via any orientation-preserving homeomorphism h_i, and map E_i to E_i' via $f_i' \circ h_i \circ f_i^{-1}$. Applying the lemma, we get a map $H : S^2 \to S^2$ that is compatible with the glueing maps. We now extend H to the interior of D^3 by regarding it as the cone of S^2 (that is, $D^3 \cong (S^2 \times I)/(S^2 \times \{1\})$), and sending $(q, t) \mapsto (H(q), t)$. This self-homeomorphism of D^3 descends to the desired homeomorphism $V \to V'$.

($\Rightarrow$) ∂V is an oriented surface with Euler characteristic $\chi(\partial V) = n\chi(S^2) - 2m = 2n - 2m$. Hence, $m - n = -\chi(\partial V)/2$. □

From the classification of closed orientable surfaces, we deduce:

Corollary 1.4 (of proof)**.** *The quantity $g = m - n + 1$ is the genus of ∂V.*

This yields the following restatement of Proposition 1.2.

Corollary 1.5. *Two handlebodies are homeomorphic if and only if their boundaries have the same genus.*

We will thus hereafter denote the handlebody of genus g by V_g.

Proposition 1.6. *V_g is homotopy-equivalent to a wedge of g circles.*

Sketch of proof. Again, we will use a construction of V from a single 3-ball. Choose one point q_i in the interior of each $F_i \subset \partial D^3$, and consider the straight-line segment connecting q_i to the centre of D^3. Do the same for the points of the form $f_i(q_i)$. Now, the map $D^3 \to V_g$ takes the union $\hat{S}$ of all these segments to a wedge S of g circles. With some work, one can find a strong deformation retraction $D^3 \to \hat{S}$ that descends to a strong deformation retraction $V \to S$. $\square$

It follows that the fundamental group $\pi_1(V_g)$ is isomorphic to $\mathcal{F}_g$, the free group on g generators. We further remark that, since the universal cover of a connected 1-dimensional CW-complex is a (possibly infinite) tree, which is contractible, any handlebody is **aspherical**, that is, it has trivial homotopy groups $\pi_i(V_g)$ for $i \geq 2$.

1.2 Disk systems

In this section, we present an explicit, geometric way of constructing free bases for $\pi_1(V_g)$, which will be a cornerstone of most of the text.

Definition 1.7. *An ordered collection $\mathcal{D} = (D_1, \ldots, D_m)$ of pairwise disjoint, oriented, properly embedded disks in V_g is called a **disk system** if cutting V_g along the disks yields a single ball.*[1]

A disk system $\mathcal{D}$ hence provides a way to exhibit V_g as a quotient $Q/\sim$ (with Q a single ball), as in the definition of a handlebody, so the equation $g = m - n + 1$ from before tells us that $m = g$. We thus conclude that any handlebody has a disk system, and a disk system for V_g contains exactly g disks. We will write $V_g \backslash \mathcal{D}$ to denote the ball that results from cutting V_g along the disks in $\mathcal{D}$.

Definition 1.8. *Two disk systems $\mathcal{D} = (D_1, \ldots, D_g), \mathcal{D}' = (D'_1, \ldots, D'_g)$ are **isotopic** if there exists an isotopy of V_g carrying one to the other.*

A typical disk system is illustrated, for $g = 3$. Proposition 1.9 below tells us that this example has full generality.

[1] In the literature, a disk system might refer simply to any collection of properly embedded essential disks that are pairwise disjoint and non-isotopic. A system of disks in this broader sense may very well disconnect V_g, or cut it into handlebodies of positive genus. The ones in our definition are commonly called *minimal* disk systems.

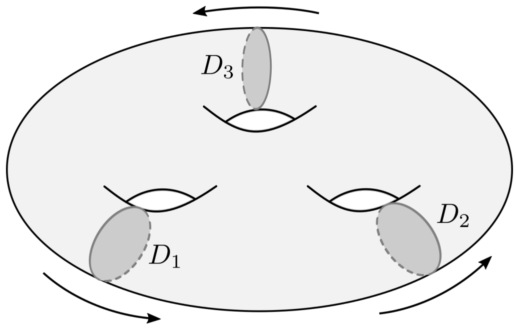


Figure 1.1: A disk system for V_3. Arrows indicate the orientation of the disks.

Since the disks in $\mathcal{D}$ carry orientations, we can distinguish the disks E_i, F_i embedded in the boundary of $Q := V_g \backslash \mathcal{D}$, that result from cutting V_g along the D_i, as follows. The E_i, F_i have natural orientations induced from the orientation of Q, so we declare the homeomorphism $D_i \to E_i$ to be orientation-preserving, and $D_i \to F_i$ orientation-reversing. We will stick to this convention for the remainder of the text.

We also collect the following result.

Proposition 1.9. *If $\mathcal{D} = (D_1, \ldots, D_g), \mathcal{D}' = (D_1', \ldots, D_g')$ are disk systems for V_g, then there exists a self-homeomorphism of V_g carrying each D_i to D_i' in an orientation-preserving manner.*

Proof. We consider the result of cutting V_g along either $\mathcal{D}$ or $\mathcal{D}'$. We get balls Q, Q' with distinguished disks E_i, F_i, E_i', F_i' on their boundaries. As was done in the proof of Proposition 1.2, we choose orientation-preserving homeomorphisms $E_i \to E_i', F_i \to F_i'$, which are then extended to $\partial Q \to \partial Q'$ using Lemma 1.3, and finally to a homeomorphism $Q \to Q'$. This descends to the desired self-homeomorphism of V_g. □

A disk system $\mathcal{D} = (D_1, \ldots, D_g)$ for V_g will be used to define a basis for $\pi_1(V_g, p)$ (for some point p away from the disks), that is, an isomorphism $\phi_{\mathcal{D}} : \pi_1(V_g, p) \to F_g$.[2] The idea is to, given an oriented loop α at p, construct a word on $x_1, x_1^{-1}, \ldots, x_g, x_g^{-1}$ that "tracks the intersections of α with the D_i". More precisely, starting at p with the empty word, we traverse α and append x_i to our word whenever α intersects D_i with positive orientation, and x_i^{-1} when it intersects D_i with negative orientation. We denote the resulting word by ω_α (this notation will also be used when α is an oriented arc), and this word represents the desired element $\phi_{\mathcal{D}}([\alpha]) \in F_g$.

Lemma 1.10. *$\phi_{\mathcal{D}}$ is a well-defined group isomorphism.*

Proof. Checking well-definedness is the hardest part. We first give a description of the universal cover of V_g in terms of $\mathcal{D}$. Let $\tilde{V_g}$ be the quotient of a disjoint union of copies of Q indexed by the elements of F_g:

$$\tilde{V_g} = \left(\coprod_{w \in F_g} Q_w \right) /\sim,$$

[2]Throughout the text, we shall use different notation to distinguish between the abstract free group $\mathcal{F}_g$, most often identified with $\pi_1(V_g)$, and the group F_g of words in the symbols $x_1^{\pm 1}, \ldots, x_g^{\pm 1}$ up to cancellation. A choice of basis for any free group is then exactly an isomorphism onto F_g.

where the equivalence relation is given by $q_w \sim f_i(q)_{wx_i}$, for $q \in F_i$ and $f_i : F_i \to E_i$ the glueing maps. We illustrate (part of) the resulting space, in the case $g = 2$.

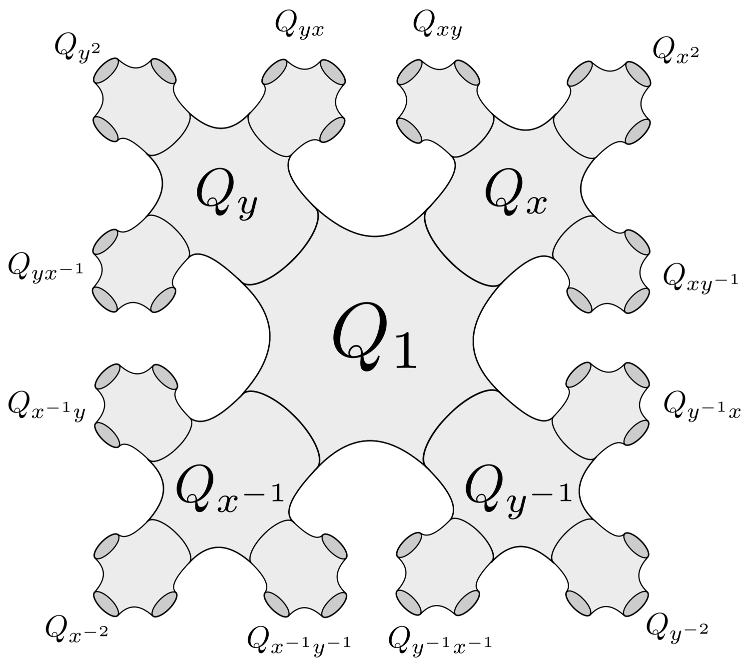


Figure 1.2: A portion of the space $\tilde{V}_2$.

The natural maps $Q_w \to V_g$ descend to a map $\tilde{V}_g \to V_g$, which is actually a covering map. By an argument similar to the one presented in the discussion preceding Proposition 1.6, one can show that $\tilde{V}_g$ is homotopy-equivalent to a tree (an infinite one, if $g \geq 1$), so it is contractible, and thus $\tilde{V}_g$ is indeed the universal cover of V_g.

Now, let α be an oriented loop at p, and denote by p_1 the point in $\tilde{V}_g$ above p in the copy of Q indexed by $1 \in F_g$. The lift $\tilde{\alpha}$ of α starting at p_1 ends at the point $p_w \in Q_w$, where $w \in F_g$ is represented by the word obtained by starting with the empty word, and appending x_i (resp. x_i^{-1}) whenever $\tilde{\alpha}$ goes into a copy of F_i (resp. E_i) and out of E_i (resp. F_i). This corresponds precisely to appending x_i when α intersects D_i positively, and x_i^{-1} when it intersects negatively. In other words, w is represented by ω_α. Therefore, if α' represents the same element as α in $\pi_1(V_g, p)$, in which case $\tilde{\alpha}'$ has the same endpoint as $\tilde{\alpha}$, we conclude that ω_α represents the same element as $\omega_{\alpha'}$ in F_g, whence $\phi_{\mathcal{D}}$ is well-defined. It is also clearly a group homomorphism.

For injectivity, observe that an occurrence of the sub-word $x_i x_i^{-1}$ in ω_α corresponds to a sub-arc of α hitting D_i twice in succession with positive and then negative orientation, but otherwise being disjoint from all disks. In $Q = V_g \backslash \mathcal{D}$, this corresponds to an arc from E_i to E_i. Such an arc can be homotoped in Q to a neighbourhood of E_i, and then through the disk D_i in V_g, thus removing the sequence $x_i x_i^{-1}$ from ω_α. Similarly, we may remove sub-words of the form $x_i^{-1} x_i$. Thus, if ω_α reduces to the empty word, we can modify α enough times, until ω_α is empty. The resulting α intersects none of the D_i, so it lifts to Q and is trivial.

For surjectivity, we need only to see that there exist loops in V_g whose corresponding words are the generators x_i, which is self-evident. □

Note that the isomorphisms $\phi_{\mathcal{D}}, \phi_{\mathcal{D}'}$ obtained from isotopic disk systems $\mathcal{D}, \mathcal{D}'$ are the same. Indeed, an isotopy of V_g carrying $\mathcal{D}$ to $\mathcal{D}'$ transforms any curve α into a curve with the same intersection pattern with the $\mathcal{D}'$ as the one of α with $\mathcal{D}$.

For completeness, we collect one last general result about disk systems, whose proof shall be given much later using tools that are yet to be introduced.

Proposition 1.11. *Any free basis of $\pi_1(V_g, p)$ can be obtained from some disk system.*

We remark that this is not a one-to-one correspondence, as the same basis can be obtained from disk systems that are not isotopic. Consider, for example, the depicted disk systems $\mathcal{D}, \mathcal{D}'$. By the bigon criterion, D_1 and D_1' do not have isotopic boundaries, but one can check on a free basis of $\pi_1(V_g, p)$ that $\phi_{\mathcal{D}} = \phi_{\mathcal{D}'}$.

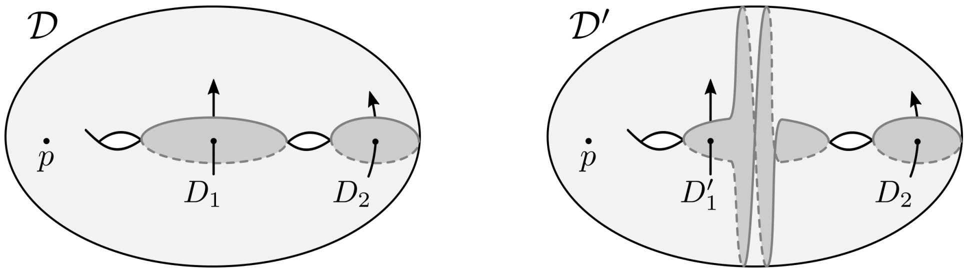


Figure 1.3: Two non-isotopic disk systems for V_2 that give rise to the same basis of $\pi_1(V_2, p)$.

2 A theorem of Zieschang

2.1 Geometric conjugacy classes of $\mathcal{F}_g$

Fix once and for all an isomorphism $\phi : \pi_1(V_g, p) \to \mathcal{F}_g$ (assume $p \in \partial V_g$). We wish to determine what elements of $\mathcal{F}_g$ correspond to homotopy classes that can be represented by simple closed curves on ∂V_g. However, since we do not care so much about curves with basepoints, we shift our attention to conjugacy classes, rather than actual elements of $\mathcal{F}_g$. Indeed, it is well-known that conjugacy classes of the fundamental group are the same as homotopy classes of curves without basepoint. Also, fundamental groups of a topological space X with respect to different basepoints are isomorphic via maps

$$\begin{aligned} \pi_1(X, x) &\to \pi_1(X, x') \\ [\alpha] &\mapsto [\gamma \cdot \alpha \cdot \gamma^{-1}], \end{aligned}$$

where γ is any path from x' to x, and different choices of γ yield isomorphisms that differ only by conjugation with some element in $\pi_1(X, x')$. Therefore, we may unambiguously talk about a conjugacy class in $\pi_1(X)$, without reference to basepoint.

Definition 2.1. *A conjugacy class $w \subseteq \mathcal{F}_g$ is **geometric** if it corresponds to a (free) homotopy class that can be represented as a simple closed curve α in ∂V_g. Such a curve α is called a **geometric realization** of w.*

*More generally, a finite set $A = \{w_1, \ldots, w_n\}$ of conjugacy classes (a **multiclass**) of $\mathcal{F}_g$ is geometric if there exist pairwise disjoint simple closed curves $\alpha_1, \ldots, \alpha_n$ (a **multicurve**) in ∂V_g with each α_i being a geometric realization of w_i. This multicurve is called a geometric realization of A.*

Our main goal is to devise an algorithm to decide whether a multiclass is geometric. The reader may be displeased by the apparent dependence of this definition on the chosen isomorphism ϕ. This choice is however immaterial:

Proposition 2.2. *Let $\phi, \phi' : \pi_1(V_g, p) \to \mathcal{F}_g$ be isomorphisms. If a multiclass A of $\mathcal{F}_g$ is geometric with respect to ϕ, then it is also geometric with respect to ϕ'.*

To see why this is true, we will appeal to the following result, which is an easy consequence of Proposition 1.9 and the yet unproven Proposition 1.11.

Lemma 2.3. *The canonical group homomorphism* $\mathrm{Homeo}(V_g, p) \to \mathrm{Aut}(\pi_1(V_g, p))$ *is surjective.*

Here, $\mathrm{Homeo}(X, x)$ denotes the group of self-homeomorphisms of a space X fixing the point $x \in X$, and $\mathrm{Aut}(G)$ denotes the group of automorphisms of a group G.

Proof of Proposition 2.2. Using Lemma 2.3, let $f : V_g \to V_g$ be a homeomorphism inducing $\phi'^{-1} \circ \phi$ on $\pi_1(V_g, p)$, and suppose the curves $\alpha_1, \ldots, \alpha_n$ are a geometric realization of $A = \{w_1, \ldots, w_n\}$ with respect to ϕ. If $\overline{\phi}$ denotes the map induced by ϕ on the set of conjugacy classes, this can be written as $\overline{\phi}([\alpha_i]) = w_i$. We need only to observe that the curves α_i are taken by f to a geometric realization of A with respect to ϕ':

$$\overline{\phi'}([f(\alpha_i)]) = \overline{\phi'}(\overline{\phi'^{-1} \circ \phi}([\alpha_i])) = w_i.$$

□

2.2 Geometricity of words

2.2.1 Cyclic words vs. conjugacy classes

At this point, it is useful to clarify some terminology, which will be extensively used.

Definition 2.4. *1. L_g shall denote the* ***set of letters*** *$\{x_1, x_1^{-1}, \dots, x_g, x_g^{-1}\}$.*

2. *By a* ***word*** *on L_g, we will strictly mean a (possibly empty) finite sequence of these symbols, without cancellation (so the word $x_1x_1^{-1}$ is different from the empty word). When giving examples with a specific g, we may use simplified notation (say, write x, y, z instead of x_1, x_2, x_3).*

3. *A word is* ***reduced*** *if no two inverse symbols occur consecutively.*

4. *A* ***cyclic word*** *is an equivalence class of words, where two are identified if they can be obtained from one another by a cyclic permutation of the symbols (for example, the words xyz, yzx, zxy are all the representatives of a particular cyclic word).*

5. *A cyclic word is reduced if all its representatives are reduced (so as a cyclic word, xyx^{-1} is not reduced).*

Having fixed a disk system for V_g (hence a free basis for $\mathcal{F}_g$), each element has a unique reduced word representing it. Similarly, a conjugacy class can be represented by a unique reduced cyclic word. Only rarely will we need to refer to elements in $\mathcal{F}_g$ and words representing them – we will instead be focusing on conjugacy classes and cyclic words. For that reason, we will not incorporate into our notation the distinction between elements and conjugacy classes of $\mathcal{F}_g$, or between words and cyclic words. We will however stick to the convention of denoting a (cyclic) word by a Greek letter (most often ω), and an element or conjugacy class by a Latin letter (most often w). This is similar to how we typically name loops of in V_g (with or without basepoint) as $\alpha, \beta, \dots$, but their homotopy classes are $a, b, \dots$. If some free basis is implicit, the element or conjugacy class represented by ω in that basis will be denoted by $[\omega]$. Consistently with notation that was introduced earlier, if a disk system for V_g is implicit and α is a closed curve in V_g, then ω_α shall denote the cyclic word tracking the signed intersections of α with the disks.

Our description of bases of $\pi_1(V_g)$ as disk systems suggests the following adaptation of the definition of geometricity to (sets of) cyclic words.

Definition 2.5. *A cyclic word ω on L_g will be called* ***geometric*** *if for some (hence any) disk system $\mathcal{D}$ for V_g, there exists a simple closed curve α in ∂V_g with $\omega = \omega_\alpha$. α is then a* ***geometric realization*** *of ω with respect to $\mathcal{D}$.*

A finite set $\Lambda = \{\omega_1, \dots, \omega_n\}$ of cyclic words (a ***multiword****) is geometric if for some (hence any) $\mathcal{D}$, there is a multicurve $\{\alpha_1, \dots, \alpha_n\}$ with each α_j a geometric realization of ω_j with respect to $\mathcal{D}$.*

(The statements in parentheses follow from Proposition 1.9.)

Rephrasing the property of a multiclass being geometric in terms of geometricity of multiwords yields the following straightforward statement.

Proposition 2.6. *A multiclass $A = \{w_j\}$ of $\mathcal{F}_g$ is geometric if and only if for some (hence any) free basis B for $\mathcal{F}_g$, there exists a geometric multiword $\Lambda = \{\omega_j\}$ such that $w_j = [\omega_j]$ in B.*

(The statement in parentheses uses Proposition 1.11.)

We emphasize that the "only if" part of this proposition does not guarantee geometricity of *all* multiwords Λ representing A – not even of the set of reduced cyclic words (the reduced multiword).

Proposition 2.7. *Let B be some basis of $\mathcal{F}_4$. There exists a geometric conjugacy class $w \subseteq \mathcal{F}_4$ such that the reduced cyclic word representing w in B is not geometric.*

Proof. Take w to be the conjugacy class represented in B by the reduced cyclic word $\omega = x^2yztz^{-1}xy^{-1}$. In Section 2.2.3, we will see that ω is not geometric, but $\omega' = x^2yztz^{-1}xy^{-1}\boldsymbol{zz^{-1}}$ is. Thus w is geometric. $\square$

The existence of this example is perhaps surprising, as the cyclic word ω' is, in some sense, more complex than ω, so one might think it ought to be harder to draw a geometric realization for it. Another way of phrasing this example is in terms of "waves".

Definition 2.8. *If $\mathcal{D}$ is a disk system for V_g, a* ***wave*** *γ is an embedded arc in V_g that intersects exactly one disk of $\mathcal{D}$, and this intersection is $\partial\gamma$. Moreover, the intersections at each endpoint should have opposite orientations.*

The claim in our previous example is therefore that if $\mathcal{D}$ is a disk system corresponding to B and α is a geometric realization of w, then one of the arcs into which α is cut by $\mathcal{D}$ is a wave. Indeed, such an arc corresponds precisely to the consecutive occurrence of two inverse symbols in ω_α.

Still, we will see that under some additional assumptions on A and B, it is true that geometricity of a multiclass is equivalent to geometricity of the reduced multiword. Proving this reduction will require a substantial amount of work, which we postpone. For the remainder of Section 2.2, we focus instead on understanding when multiwords are geometric, and in so doing we introduce a fundamental tool, to be used extensively throughout this work.

2.2.2 Word graphs and Whitehead graphs

Definition 2.9. *If Λ is a multiword on L_g, then the* ***word graph*** $\mathrm{Wh}(\Lambda)$ *is the directed graph having L_g as its vertex set, and with one edge from y to z^{-1} for each occurrence of the string yz in some cyclic word of Λ.*

If A is a multiclass of $\mathcal{F}_g$, and B is a free basis of $\mathcal{F}_g$, the ***Whitehead graph*** $\mathrm{Wh}_B(A)$ *is the word graph of reduced multiword for A in the basis B.*

The Whitehead graph was first introduced by J. H. C. Whitehead in [11, 12]. Originally, it arose on a different topological setting, where $\mathcal{F}_g$ was being regarded as the fundamental group of a g-fold connected sum of copies of $S^1 \times S^2$.

For our purposes, the motivation lies in the following construction. Fix a disk system $\mathcal{D}$ for V_g, let $\Lambda = \{\omega_1, \ldots, \omega_n\}$ be a multiword in L_g, and let $\{\alpha_1, \ldots, \alpha_n\}$ be a multicurve in V_g with $\omega_{\alpha_j} = \omega_j$ for all j. Cut V_g along $\mathcal{D}$, and then collapse to points the $2g$ disks on the boundary of the resulting ball Q. If none of the ω_j is empty, then the arcs coming from the α_j, together with the collapsed disks, form a directed graph embedded in a 3-ball, with the vertices on the boundary. This graph is precisely the word graph $\mathrm{Wh}(\Lambda)$, where the vertices are labeled as x_i or x_i^{-1} depending respectively on whether they come from collapsing E_i or F_i.

More to our purposes, suppose Λ is geometric and $\{\alpha_j\}$ is a geometric realization of Λ. Then this embedding of $\mathrm{Wh}(\Lambda)$ is entirely contained in a 2-sphere. This observation alone yields a first application of the word graph.

Proposition 2.10. *The word graph of a geometric multiword is planar.*

Sometimes, it will be convenient to refrain from collapsing the disks E_i, F_i, so we may work with embeddings of word graphs or Whitehead graphs that have "thick vertices". We illustrate Proposition 2.10 with $\Lambda = \{\omega\}$, where ω is the (geometric) cyclic word $xyz^{-1}y^{-1}$.

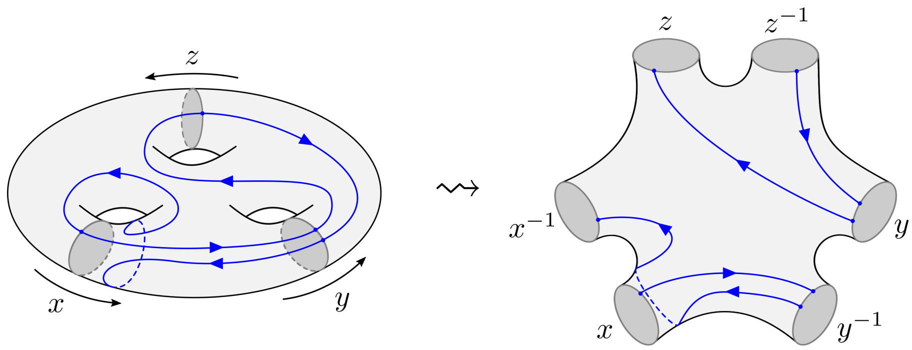


Figure 2.1: A planar "embedding with thick vertices" of $\mathrm{Wh}(\{\omega\})$, obtained by cutting a geometric realization of ω along a suitable disk system.

As an example with $g = 3$, the cyclic word $\omega = x^2y^2z^2xzy$ is not geometric, because the underlying undirected graph of $\mathrm{Wh}(\{\omega\})$ is $K_{3,3}$, which is non-planar.

Since each occurrence of $x_i^{\pm 1}$ in a cyclic word of Λ contributes with one edge out of the vertex labelled $x_i^{\pm 1}$ and one into $x_i^{\mp 1}$, we see that vertices that are inverse to each other have the same degree. Moreover, the construction of $\mathrm{Wh}(\Lambda)$ provides canonical bijections η_i ($1 \le i \le g$) from the set of half-edges incident with x_i^{-1} to the set of half-edges incident with x_i. These η_i map incoming half-edges to outgoing half-edges, and vice-versa. Specifically, if a string of the form yx_iz occurs in a cyclic word of Λ, then η_i maps the half-edge into x_i^{-1} from $y \longrightarrow x_i^{-1}$ to the half-edge out of x_i from $x_i \longrightarrow z^{-1}$ (and similarly for a string $yx_i^{-1}z$). This additional piece of structure will be of paramount importance in characterizing geometric cyclic words.

2.2.3 Testing geometricity of cyclic words

Planarity of the word graph is not a sufficient condition for geometricity of Λ. In this section, we will prove a criterion that is actually necessary and sufficient, by inspecting the embedding given by Proposition 2.10 and its relation with the bijections η_i. This involves an important feature of graphs embedded in oriented surfaces, for which we need the following notion.

Definition 2.11. *A **total cyclic order** C on a set X is a ternary relation $C \subseteq X^3$ satisfying the following axioms, for all $a, b, c, d \in X$:*

1. *Cyclicity: $(a, b, c) \in C$ implies $(b, c, a) \in C$,*
2. *Asymmetry: $(a, b, c) \in C$ implies $(a, c, b) \notin C$,*
3. *Transitivity: If $(a, b, c) \in C$ and $(a, c, d) \in C$, then $(a, b, d) \in C$,*
4. *Totality: If a, b, c are all distinct, then $(a, b, c) \in C$ or $(a, c, b) \in C$.*

We will say that "b lies between a and c" to express the fact that $(a, b, c) \in C$ (this is of course not a synonym to "b lies between c and a"). For any cyclic order C, one can define an **opposite** order $\overline{C}$ by

$$(a, b, c) \in \overline{C} \iff (a, b, c) \notin C.$$

If X is finite, then for each $a \in X$, there exists a unique $b \neq a$ such that for every c distinct from a and b, we have $(a, b, c) \in C$. This b is the **successor** of a. We also define the **predecessor** of a to be its successor in $\overline{C}$.

Importantly, the points in an oriented circle are totally cyclically ordered by setting $(a, b, c) \in C$ whenever the positively oriented arc from a to c contains b in its interior.

Definition 2.12. *A **ribbon graph** is a graph together with total cyclic orders on the sets of half-edges at each vertex.*

Any finite graph embedded in an oriented surface S is naturally endowed with a ribbon structure, by "reading the half-edges at each vertex v counter-clockwise", which we formalize as follows. First, choose a point in the interior of each edge (call it the midpoint), and define the two connected components of its complement in the edge to be the half-edges. Then, pick a closed disk D embedded in S containing v, small enough so that D intersects no other vertex and no midpoint of any edge. Now, ∂D intersects all half-edges incident with v, and for each such half-edge h, there is one intersection point p_h that is closest to v along h. We endow the set of all such points with the total cyclic order inherited from ∂D, and carry this cyclic order to the half-edges themselves. This construction is independent of the choices of midpoints and of D. In particular, D can be made arbitrarily small.

We are now ready to state and prove the main result of this section.

Definition 2.13. *A ribbon structure on a word graph is **admissible** if the cyclic orders are reversed by the bijections η_i. A planar embedding of a word graph is admissible if it induces an admissible ribbon structure.*

Proposition 2.14. *A multiword Λ in L_g is geometric if and only if there exists an admissible planar embedding of* $\mathrm{Wh}(\Lambda)$.

Proof. ($\Rightarrow$) Consider the ribbon structure on $\mathrm{Wh}(\Lambda)$ induced from the planar embedding given by Proposition 2.10. For each $1 \le i \le g$, there are obvious bijections

$$D_i \cap \bigcup_j \alpha_j \longleftrightarrow E_i \cap \left(\bigcup_j \alpha_j \backslash \mathcal{D}\right) \longleftrightarrow \{\text{half-edges of } \mathrm{Wh}(\Lambda) \text{ at } x_i\}.$$

Here, $\left(\bigcup_j \alpha_j \backslash \mathcal{D}\right)$ denotes the collection of arcs obtained from the α_j after cutting V_g along $\mathcal{D}$. The bijection on the left preserves the natural cyclic orders by our convention that $D_i \to E_i$ is orientation-preserving, and one can see by definition of the ribbon structure on $\mathrm{Wh}(\Lambda)$ that the bijection on the right also respects cyclic orders.

Similarly, there are bijections

$$D_i \cap \bigcup_j \alpha_j \longleftrightarrow F_i \cap \left(\bigcup_j \alpha_j \backslash \mathcal{D}\right) \longleftrightarrow \{\text{half-edges of } \mathrm{Wh}(\Lambda) \text{ at } x_i^{-1}\},$$

but this time, the bijection on the left is order-reversing. Since η_i is the composition of all these, we conclude it reverses the cyclic orders at each pair of inverse vertices.

($\Leftarrow$) We need to construct a geometric realization of Λ, given an embedding of $\mathrm{Wh}(\Lambda)$ in S^2 inducing an admissible ribbon structure. At each vertex v of $\mathrm{Wh}(\Lambda)$, find a disk D embedded in S^2 as in the construction of the cyclic ordering of the half-edges. Thus, D contains v and ∂D intersects each half-edge incident with v. We assume that all intersections of ∂D with the edges of $\mathrm{Wh}(\Lambda)$ are transverse, and all $2g$ disks are disjoint. We now modify D so that $\partial D \cap \mathrm{Wh}(\Lambda)$ consists precisely of one point of each half-edge incident with v. We do this by noticing that a point p of intersection of ∂D with another half-edge, or that is not closest to v along the half-edge that contains it, may be removed. Indeed, if we start at such a point p and traverse the edge containing p towards the interior of D, we will cross ∂D again and leave D. Thus, this arc we traversed cuts D into two disks, of which only one contains v. We may therefore modify D so as to discard the other, along with p:

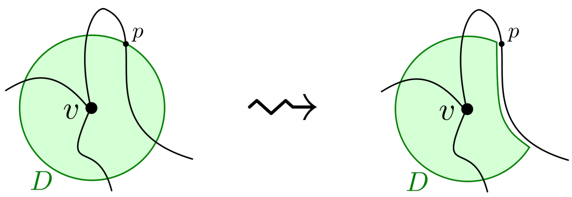


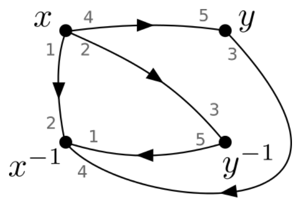
Figure 2.2: Modifying D so that each half-edge at v intersects ∂D exactly once.

After doing this enough times, we obtain D as stated.

The resulting disks will be called E_i, F_i, according to whether they contain the vertices x_i, x_i^{-1}, respectively. The intersection points of $\mathrm{Wh}(\Lambda)$ with their boundaries are naturally identified with the half-edges of $\mathrm{Wh}(\Lambda)$, and thus the bijections η_i can be regarded as maps $\mathrm{Wh}(\Lambda) \cap \partial F_i \to \mathrm{Wh}(\Lambda) \cap \partial E_i$. Now, the assumption that the η_i are order-reversing allows us to construct orientation-reversing homeomorphisms $\partial F_i \to \partial E_i$ that restrict to η_i. These homeomorphisms can in turn be extended to the interior of the disks, yielding glueing maps $f_i : F_i \to E_i$, which we use to construct V_g. After glueing, the portion of

$\mathrm{Wh}(\Lambda)$ away from the interior of the disks has become a multicurve in ∂V_g realizing Λ geometrically. □

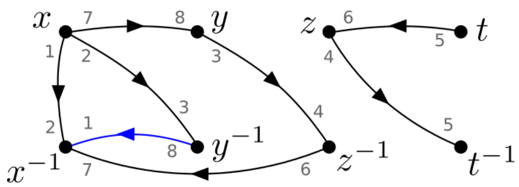
We now provide an example of a multiword Λ that is not geometric, even though its word graph is planar. With $g = 2$, let Λ be comprised solely of the cyclic word x^2yxy^{-1}. $\mathrm{Wh}(\Lambda)$ is depicted below.



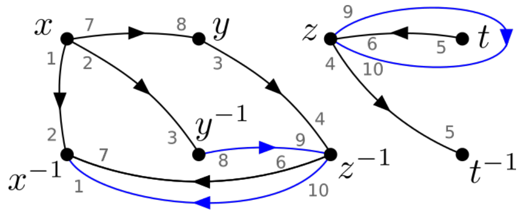
The half-edges have been labelled in order to indicate the η_i. For this specific planar embedding, the bijection between the half-edges at x^{-1} and x does not reverse the induced cyclic order, so this embedding is not admissible. But for any embedding of $\mathrm{Wh}(\Lambda)$ in S^2, the triangle formed by the vertices x, y, x^{-1} separates S^2 into two disks, and the induced ribbon structure on $\mathrm{Wh}(\Lambda)$ is completely determined by which of them contains the remaining vertex y^{-1}. Neither choice results in an admissible ribbon structure, so Λ is not geometric.

In the proof of Proposition 2.7, we stated that the reduced cyclic word $\omega = x^2yztz^{-1}xy^{-1}$ is not geometric, yet any conjugacy class of $\mathcal{F}_g$ represented by ω (in some basis) is geometric, since $\omega' = x^2yztz^{-1}xy^{-1}zz^{-1}$ is geometric. The situation becomes clear once we contemplate the relevant word graphs, shown below with the differences highlighted in blue.

$\mathrm{Wh}(\{\omega\})$:



$\mathrm{Wh}(\{\omega'\})$:



One of the connected components of $\mathrm{Wh}(\{\omega\})$ is a subdivision of the word graph from the previous example, and the same argument shows it does not have an admissible embedding. However, the planar embedding of $\mathrm{Wh}(\{\omega'\})$ above is admissible. The extra syllable zz^{-1} in ω' results in a loop that allows the "last edge" to hit the vertex x^{-1} on the correct side. This loop corresponds to a wave in the geometric realization of w.

As an application of the other direction of Proposition 2.14, we have the following result.

Corollary 2.15. *If each variable occurs in Λ at most twice, then Λ is geometric.*

To clarify, in each cyclic word we count every occurrence of x_i and x_i^{-1}. For example, x occurs twice in each of x^2, xx^{-1} and x^{-2}. We then add up these numbers over all words in Λ.

Proof. $\mathrm{Wh}(\Lambda)$ has all vertices of degree at most 2. Hence, it cannot contain a subdivision of K_5 or $K_{3,3}$, and so by Kuratowski's Theorem, it is planar. But again since there are at most two half-edges at each vertex of $\mathrm{Wh}(\Lambda)$, the ribbon structure will be comprised only of empty cyclic orders, thus being trivially admissible. $\square$

The usefulness of Proposition 2.14 depends on our ability to tell whether a given ribbon graph can be embedded in S^2 in a manner that preserves its ribbon structure. This is related to the following concept.

Definition 2.16. *The* ***genus*** $g(\Gamma)$ *of a finite graph* Γ *is the smallest genus of a closed oriented connected surface where it embeds. If* Γ *is a ribbon graph, we further require that the ribbon structure be induced by the embedding.*

In Appendix A, we show that this notion is well-defined (that is, a minimal surface in this sense always exists), and give a method to compute the genus of any ribbon graph. Having such a procedure in hand, testing geometricity of a multiword Λ is reduced to an algorithmic task, since one needs only to go through all admissible ribbon structures on $\mathrm{Wh}(\Lambda)$ (of which there are only finitely many), and test if any of them has genus 0.

2.3 Separability

2.3.1 Disk-busting conjugacy classes

We now set out to translate our original problem – determining if a multiclass A is geometric – into a geometricity test for some multiword Λ representing A in an appropriate basis. The first step is to reduce the problem to the case where A does not have the following property.

Definition 2.17. *A multiclass* A *of* $\mathcal{F}_g$ *is* ***separable*** *if there exists a non-trivial factorization* $\mathcal{F}_g = G_1 * G_2$ *as a free product, with each* $w \in A$ *having a representative in* G_1 *or* G_2*. If* A *is not separable, it is called* ***disk-busting****.*

The second definition is motivated by the following fact.

Proposition 2.18. *Suppose* A *is a multiclass of* $\mathcal{F}_g$*, and let* $\{\alpha_j\}$ *be a set of closed curves in* V_g *representing* A *(not necessarily simple closed curves in the boundary). If there exists some essential, properly embedded disk* D *in* V_g *disjoint from all* α_j*, then* A *is separable. Moreover, there is a disk system* $\mathcal{D} = (D_1, \dots, D_s, D_{s+1}, \dots, D_g)$ *so that each* α_j *is disjoint from either the* $D_1, \dots, D_s$ *or the* $D_{s+1}, \dots, D_g$*.*

Proof. We construct $\mathcal{D}$ as follows:

Case 1: If D disconnects V_g, then cutting along D produces two handlebodies V_s, V_{g-s} of positive genus (since D is essential), each with a marked disk E, F on its boundary, with E, F disjoint from the α_j. One can then find disk systems $(D_1, \dots, D_s)$ for V_s, $(D_{s+1}, \dots, D_g)$ for V_{g-s}, which can be arranged to miss E, F. Together, these disk systems lift to a disk system $\mathcal{D} = (D_1, \dots, D_s, D_{s+1}, \dots, D_g)$ for V_g.

Case 2: If cutting along D does not disconnect V_g, we complete D to a disk system $\mathcal{D} = (D_1, \ldots, D_{g-1}, D = D_g)$. In this case, $s = g - 1$.

In either case, if $(b_1, \ldots, b_s, b_{s+1}, \ldots, b_g)$ is the basis given by $\mathcal{D}$, then setting $G_1 = \langle b_1, \ldots, b_s \rangle, G_2 = \langle b_{s+1}, \ldots, b_g \rangle$ gives a factorization $\mathcal{F}_g = G_1 * G_2$ showing that A is separable. □

In Section 2 of [9], Stallings describes an algorithm for determining whether a multiclass A is separable. The space used to model $\mathcal{F}_g$ is a the g-fold connected sum $W_g := (S^1 \times S^2)^{\# g}$, which is actually the same as a pair of handlebodies of genus g glued along their boundary via the identity map. The role of our disk systems is, in that case, played by collections of g spheres, and cutting along such systems of spheres yields a space that is homeomorphic to S^3 with $2g$ open balls removed. This is where word graphs are naturally seen, in that context.

Both W_g and the handlebody picture in the previous proof, where we allow curves to be embedded in the interior of V_g, are more flexible models for studying abstract group-theoretical properties of $\mathcal{F}_g$. Indeed, for any system of spheres/disks, we can represent a multiclass A by a multicurve *without waves* (whose definition is adapted in the obvious manner). This means that for any choice of basis B, we can represent A by curves tracking the reduced cyclic words representing A in B, and so after cutting W_g or V_g along the system of spheres/disks, we always recover the picture of $\mathrm{Wh}_B(A)$ (rather than just some word graph). As we saw before, this is radically different from what happens when we insist that the curves lie in ∂V_g. These models are of course not so closely related to our endeavour of characterizing geometricity, but they should nevertheless be kept in mind. Here is one application, which is the first ingredient in Stallings' proof:

Proposition 2.19. *Let A be a multiclass of $\mathcal{F}_g$. If there is a basis B of $\mathcal{F}_g$ such that $\mathrm{Wh}_B(A)$ is disconnected, then A is separable.*

Proof. Write $\mathrm{Wh}_B(A)$ as a non-trivial disjoint union $\Gamma_1 \sqcup \Gamma_2$, and consider the V_g model with A represented by curves in the interior of V_g without waves. For $\mathcal{D}$ a disk system representing B, cutting V_g along $\mathcal{D}$ yields an "embedding" of $\mathrm{Wh}_B(A)$ in $Q = V_g \backslash \mathcal{D}$ (with vertices replaced by disks in ∂V_g). Find a properly embedded disk D in Q disjoint from the $E_i, F_i \subset \partial Q$, so that each ball into which D cuts Q contains the vertex set of one of the Γ_i. Homotope all edges relative endpoints, so that the curves representing A are disjoint from D in V_g. Proposition 2.18 finishes the job. □

2.3.2 Whitehead moves

Definition 2.20. *An automorphism of F_g is called:*

1. ***simple**, if it is induced by a permutation of L_g,*
2. ***of type T**, denoted $T_{v,V}$, if there exist a letter $v \in L_g$ and a subset $V \subset L_g$ containing v but not v^{-1}, such that $T_{v,V}$ fixes v, and for $x_i \neq v, v^{-1}$, it is given by*

$$x_i \mapsto \begin{cases} x_i & \text{if } x_i, x_i^{-1} \notin V \\ x_i v & \text{if } x_i \in V, x_i^{-1} \notin V \\ v^{-1} x_i & \text{if } x_i \notin V, x_i^{-1} \in V \\ v^{-1} x_i v & \text{if } x_i, x_i^{-1} \in V \end{cases},$$

3. a ***Whitehead automorphism****, if it is of either of the previous two types.*

A change of free basis of $\mathcal{F}_g$ from B to B' is called a ***Whitehead move*** *if the basis change map $\phi_{B'} \circ \phi_B^{-1} : F_g \to F_g$ is a Whitehead automorphism.*

Whitehead moves lie at the heart of Stallings' proof, and much like Whitehead graphs, they have a rather vivid geometric interpretation. If B is the basis for $\mathcal{F}_g$ induced by some disk system $\mathcal{D}$, then a simple Whitehead move corresponds to a reordering of the disks in $\mathcal{D}$, possibly with a change of orientation of some of them.

Whitehead moves of type T are more interesting. Consider the picture of $Q = V_g \backslash \mathcal{D}$, with embedded disks E_i, F_i in ∂Q, identified, respectively, with the letters x_i, x_i^{-1}. Let v, V be as in the above definition, and draw a new properly embedded disk D in Q disjoint from the disks E_i, F_i, such that in $Q \backslash D$, the E_i, F_i are partitioned into V and its complement. Moreover, let the basepoint p of V_g lie in the connected component of $V_g \backslash D$ that does not contain the elements of V, and orient D so that its negative side is facing the elements of V.

Since each of v, v^{-1} lies in a different side of D in Q, if we cut Q along D and then glue back the "thick vertices" v, v^{-1}, we again obtain a 3-ball. This shows that if, in $\mathcal{D}$, we replace the disk corresponding to v, v^{-1} with D, we obtain a disk system. One can check, by drawing generators and seeing how they intersect the new disk system, that the corresponding change of basis is precisely the Whitehead move given by $T_{v,V}$. Such a disk D will be said to **model** the Whitehead move, and by "realizing a Whitehead move geometrically", we shall mean performing the change of disk system we just described. Any Whitehead move can thus be specified just by choosing D and selecting a compatible letter $v \in L_g$. As an example with $g = 3$, the disk D pictured below could be used to model the Whitehead moves given by $V = \{x, x^{-1}, y^{-1}, z\}$ and $v = y^{-1}$ or $v = z$.

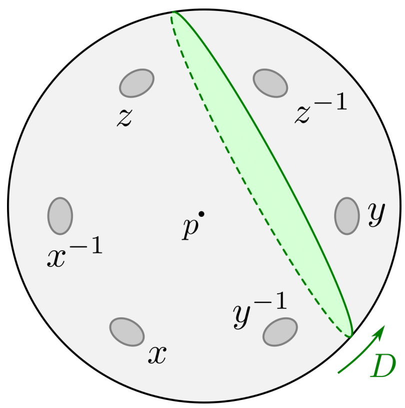


Figure 2.3: A disk D modelling the Whitehead moves given by $T_{y^{-1},V}$ or $T_{z,V}$, where $V = \{x, x^{-1}, y^{-1}, z\}$.

This new language warrants a second inspection of the proof of Proposition 2.19. If we translate the two cases in the proof of Proposition 2.18 into the context of Whitehead graphs, Case 1 corresponds precisely to the situation where the vertex sets of the two subgraphs Γ_1, Γ_2 are closed under inverses. In that case, the induced partition of B gives the desired splitting of $\mathcal{F}_g$. In Case 2, the disk D can be made to correspond to a rank 1 free factor in which none of the conjugacy classes of A is represented. Since we are in the situation where there is a vertex $v \in \Gamma_1$ such that $v^{-1} \in \Gamma_2$, our geometric interpretation of Whitehead moves tells us we can obtain from B a basis compatible with such a splitting by orienting D so that its positive side is facing Γ_2, and applying the Whitehead move given by $T_{v,V}$, where V is the vertex set of Γ_1.

2.3.3 Testing separability

The second ingredient for the separability test, whose proof will be omitted, is a result shown in [9] by employing a method of Whitehead from [11], where the "sphere model" W_g is heavily used. A different proof can be found in [10], using a generalization of Whitehead graphs.

Definition 2.21. *A vertex v of a graph Γ is called a* ***cut vertex*** *if the graph $\Gamma \backslash v$, obtained from Γ by removing v and all edges incident with v, has more connected components than Γ.*

Proposition 2.22. *If a multiclass A is separable and B is a basis for $\mathcal{F}_g$ with $\mathrm{Wh}_B(A)$ connected, then $\mathrm{Wh}_B(A)$ has a cut vertex.*

The last ingredient is the following lemma.

Definition 2.23. *Let A be a multiclass in $\mathcal{F}_g$, and B a free basis. The* ***length*** *of A relative to B, denoted $l_B(A)$, is the sum of the lengths of the reduced cyclic words representing A in B, or, equivalently, the number of edges in $\mathrm{Wh}_B(A)$.*

Lemma 2.24. *Suppose $\mathrm{Wh}_B(A)$ is connected and has a cut vertex v, so $\mathrm{Wh}_B(A)$ is the union of two connected subgraphs Γ_1, Γ_2 whose intersection is exactly v, both containing vertices other than v. Assume Γ_2 contains v^{-1}, and let $V \subset L_g$ be the vertex set of Γ_1. If B' is the basis obtained from B by applying the Whitehead move given by $T_{v,V}$, then $l_{B'}(A) < l_B(A)$.*

Proof. Let $\mathcal{D}$ be a disk system for V_g corresponding to B, and represent A by a multicurve $\{\alpha_j\}$ in the interior of V_g, without waves. Draw a disk D properly embedded in $Q = V_g \backslash \mathcal{D}$ modelling the Whitehead move of interest and homotope the arcs in Q relative endpoints so that all of them intersect D at most once. Thus, in the resulting picture, all intersections of $\mathrm{Wh}_B(A)$ with D correspond to edges connecting Γ_1 to Γ_2, which are a proper subset of the edges incident with v.

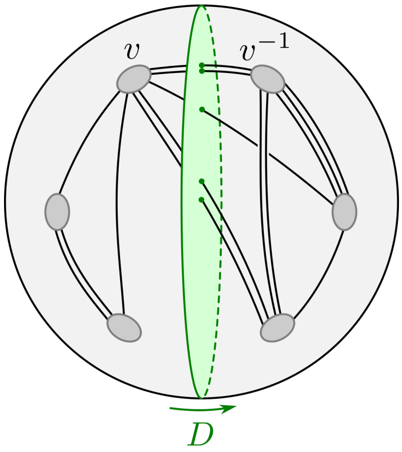


Figure 2.4: A disk D modelling a Whitehead move given by $T_{v,V}$, where v is a cut vertex of $\mathrm{Wh}_B(A)$. On the left of D we see the vertex set of Γ_1, and the vertices on the right, together with v, comprise the vertex set of Γ_2.

Since there are no loops in $\mathrm{Wh}_B(A)$, or arcs from D back to itself, the α_j form no waves with the disk system $\mathcal{D}'$ obtained by realizing our Whitehead move geometrically. Hence, cutting Q along D and glueing the disks labelled v, v^{-1}, we obtain $V_g \backslash \mathcal{D}'$ with the embedded graph $\mathrm{Wh}_{B'}(A)$. Our lemma will follow if we show that the number of edges in the new graph is strictly smaller than in the old one. Since vertices other than v, v^{-1} are not having their degrees changed, this is equivalent to the claim that the degree of v (which is the same as that of v^{-1}) has decreased. But the degree of v in $\mathrm{Wh}_{B'}(A)$ is the number of edges intersecting D in the Q picture, and these are strictly fewer than the edges incident with v. $\square$

Finally, we are ready to state the separability test for a multiclass A: Draw the Whitehead graph $\mathrm{Wh}_B(A)$ for any basis B. If it is disconnected, Proposition 2.19 shows A is separable, and provides a compatible factorization of $\mathcal{F}_g$. Otherwise, look for a cut vertex. If none exists, we conclude by Proposition 2.22 that A is not separable. If there is a cut vertex, perform the Whitehead move dictated by Lemma 2.24, and repeat with the new basis. This step reduces the length of A, which cannot decrease indefinitely, so eventually we reach either a disconnected graph, or a connected graph without cut vertices.

2.3.4 Geometric separability

The goal of this section is to show that geometricity of a separable multiclass A of $\mathcal{F}_g$ is equivalent to geometricity of each of its parts in the free factors of $\mathcal{F}_g$. The main idea we will need is that "algebraic separability implies geometric separability". More precisely, we aim to prove the following result, using the approach in Section 3 of [9].

Proposition 2.25. *If A is a geometric multiclass, with $\{\alpha_j\}$ a geometric realization, and A is separable, then there exists an essential properly embedded disk D in V_g that is disjoint from all α_j.*

The geometric constraints in this setting require us to replace the concept of length from before with a different notion.

Definition 2.26. *Let $\{\alpha_j\}$ be a multicurve in V_g, and $\mathcal{D} = (D_1, \ldots, D_g)$ a disk system for V_g. The **complexity** of the multicurve (relative to $\mathcal{D}$) is either of the following equal numbers:*

1. *the total number of intersections of the α_j with the D_i,*
2. *the sum of the lengths of the cyclic words ω_{α_j},*
3. *the number of edges in $\mathrm{Wh}(\{\omega_{\alpha_j}\})$.*

As we have seen, when given a geometric multiclass A and a disk system $\mathcal{D}$ representing a basis B, it may be impossible to find a geometric realization of A such that cutting V_g along $\mathcal{D}$ yields an embedding of $\mathrm{Wh}_B(A)$, as there may be waves which we cannot get rid of. The following result tells us that if the disk system is chosen carefully, this is not an issue.

Proposition 2.27. *Let $\{\alpha_j\}$ be a multicurve in ∂V_g. Then for any disk system $\mathcal{D}$, there is a system $\mathcal{D}'$ such that the α_j form no waves with $\mathcal{D}'$, and the complexity of $\{\alpha_j\}$ relative to $\mathcal{D}'$ is no more than with respect to $\mathcal{D}$.*

If A is a geometric multiclass, then by taking $\{\alpha_j\}$ to be a geometric realization of A, this result implies that there is some free basis B of $\mathcal{F}_g$ for which $\mathrm{Wh}_B(A)$ has an admissible planar embedding.

Proof. If the curves form no waves with $\mathcal{D}$, then there is nothing to prove. Otherwise, cut V_g along $\mathcal{D}$. Any wave becomes an embedded arc γ in the boundary of $Q = V_g \backslash \mathcal{D}$, connecting two boundary points of one of the E_i, F_i. For concreteness, say it is E_i. A small closed regular neighbourhood of $\partial E_i \cup \gamma$ in ∂Q has three circles as its boundary, only one of which separates E_i from F_i. Let D be an oriented, properly embedded disk in Q having this circle as its boundary.

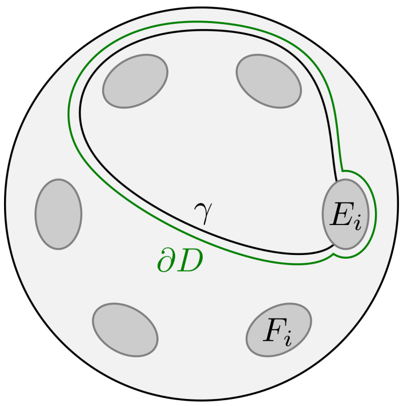


Figure 2.5: Using a wave at a disk D_i of $\mathcal{D}$ to construct a disk D having fewer intersections with the α_j.

Replacing D_i with D in $\mathcal{D}$ yields a new disk system $\mathcal{D}_1$, as D models a Whitehead move of type T. Moreover, this change introduces no new intersections of the α_j with the disk system, and removes the intersections with one of the two arcs into which γ cuts D_i.

Hence, the complexity of $\{\alpha_j\}$ relative to $\mathcal{D}_1$ is strictly less than with respect to $\mathcal{D}$. Since complexity cannot decrease indefinitely, if we iterate this procedure enough times we will eventually reach a disk system for which there are no waves. $\square$

Proof of Proposition 2.25. Start with any disk system $\mathcal{D}$, and apply the following steps:

1. Use Proposition 2.27 to modify $\mathcal{D}$ without increasing complexity, so the α_j form no waves.
2. Cut V_g along $\mathcal{D}$. Since there are no waves, we get an "embedding with thick vertices" of $\mathrm{Wh}_B(A)$ in the boundary of $Q = V_g \backslash \mathcal{D}$, where B is the basis represented by $\mathcal{D}$.
   (a) If $\mathrm{Wh}_B(A)$ is disconnected, one can find a properly embedded disk D in Q disjoint from the Whitehead graph, with connected components of the graph in either side of D. Hence, D is essential in V_g, as desired.
   (b) If $\mathrm{Wh}_B(A)$ is connected, Proposition 2.22 gives a cut vertex, which corresponds to one of the disks E_i, F_i (say E_i, for concreteness). In the disk $\partial Q \backslash \mathring{E}_i$, draw an arc γ connecting two points of ∂E_i, but otherwise disjoint from $\mathrm{Wh}_B(A)$, such that each side of γ contains some connected component of $\mathrm{Wh}_B(A) \backslash E_i$. Then, perform a similar construction as in the proof of Proposition 2.27: Choose a small regular neighbourhood of $\partial E_i \cup \gamma$ in ∂Q and use the boundary component that separates E_i from F_i to produce a properly embedded disk D in Q. Replace D_i in $\mathcal{D}$ with D to obtain a new disk system, strictly reducing complexity. Return to step 1.

Since complexity cannot decrease indefinitely, we eventually land in step 2.(a). $\square$

For our purposes, the usefulness of Proposition 2.25 lies in the following corollary:

Corollary 2.28. *Suppose A is geometric. Then there exits a (possibly trivial) factorization $\mathcal{F}_g = G_1 * \ldots * G_m * H$ and a partition $A = A_1 \sqcup \ldots \sqcup A_m$ such that each A_k is represented in G_k, and A_k is disk-busting and geometric in G_k.*

Proof. The essence of the proof is the following statement.

Claim: Suppose we have a (possibly trivial) factorization $\mathcal{F}_g = G * H$ and a disk system $\mathcal{D} = (D_1, \ldots, D_r, D_{r+1}, \ldots, D_g)$ for V_g giving a compatible basis $(b_1, \ldots, b_r, b_{r+1}, \ldots, b_g)$, that is, $(b_1, \ldots, b_r)$ is a basis for G and $(b_{r+1}, \ldots, b_g)$ is a basis for H. Moreover, let A be a geometric multiclass, and $\{\alpha_j\}$ a geometric realization such that the α_j intersect none of the disks $D_{r+1}, \ldots, D_g$. Then, if A is separable in G, we can replace the disks $D_1, \ldots, D_r$ from $\mathcal{D}$ with $D'_1, \ldots, D'_s, D'_{s+1}, \ldots, D'_r$ such that each α_j intersects only the $D'_1, \ldots, D'_s$, or only the $D'_{s+1}, \ldots, D'_r$.

To justify this fact, we cut V_g along the disks $D_{r+1}, \ldots, D_g$, thus recovering a handlebody V_r in which A is geometrically realized. Moreover, there are marked disks $E_{r+1}, F_{r+1}, \ldots, E_g, F_g$ in ∂V_r that are disjoint from the curves α_j. Since A is separable, by Proposition 2.25 there is an essential properly embedded disk D in V_r disjoint from the α_j. By Proposition 2.18, we can modify $(D_1, \ldots, D_r)$ to a disk system $(D'_1, \ldots, D'_s, D'_{s+1}, \ldots, D'_r)$ for V_r that gives a basis compatible with a non-trivial factorization $G = G_1 * G_2$, so that each curve in A intersects only the $D'_1, \ldots, D'_s$, or only the $D'_{s+1}, \dots, D'_r$. These disks can be modified slightly so as to avoid the $E_{r+1}, F_{r+1}, \dots, E_g, F_g$, and thus lift to V_g, replacing $D_1 \dots, D_r$ in $\mathcal{D}$. This proves the claim.

We now explain how to deduce the corollary. If A is disk-busting, there is nothing to prove. Otherwise, applying our claim with H trivial and $\mathcal{D}$ any disk system for V_g will give a non-trivial factorization of F_g with a new, compatible disk system, together with a (possibly trivial) partition of A. We then ask whether any part A_k of A is separable in the respective factor G_k, and apply the claim again with $A = A_k, G = G_k$. We iterate this argument, each time further refining the partition of A and the factorization of $\mathcal{F}_g$, and we eventually reach a point where all A_k are disk-busting in G_k. Some rank 1 factors of $\mathcal{F}_g$ (coming from applications of Proposition 2.18 where the disk D falls into Case 2 of the proof) will have no conjugacy classes represented in them, and we group them all into the factor H in the statement of the corollary. $\square$

The desired reduction of the geometricity test to disk-busting multiclasses is contained in the following proposition, of which we give only a partial proof.

Proposition 2.29. *Let $A = A_1 \sqcup \dots \sqcup A_m$ be a multiclass of $\mathcal{F}_g = G_1 * \dots * G_m * H$ such that each A_k is non-empty and disk-busting in G_k. Then A is geometric in $\mathcal{F}_g$ if and only if each A_k is geometric in G_k.*

Proof. ($\Leftarrow$) Choose a disk system $\mathcal{D}$ for V_g corresponding to a basis of $\mathcal{F}_g$ that is compatible with the given factorization, and let $Q = V_g \backslash \mathcal{D}$. For each $1 \leq k \leq m$, draw a simple closed curve γ_k on ∂Q separating the $E_i, F_i \subset \partial Q$ coming from disks D_i that correspond to basis elements in G_k, from the other E_i, F_i. Moreover, make it so that the α_k are pairwise disjoint. (One way to obtain such curves would be to, for each k, connect the relevant E_i, F_i by a minimal number of paths in ∂Q, constructing a "tree with thick vertices", and then take the α_k to be boundaries of small regular neighbourhoods in ∂Q of all those trees).

Now, for each k, glue back the E_i, F_i corresponding to the basis elements in G_k, so as to obtain a handlebody V_{g_k} for which the natural map into V_g induces precisely the inclusion $G_k \hookrightarrow F_g$. Then, find a geometric realization of A_k in V_{g_k} such that the curves all lie on the same side of γ_k (this is not hard to arrange, since γ_k is just bounding a disk in ∂V_{g_k}). This guarantees that curves from geometric realizations of different factors are disjoint. Hence, the geometric realizations of the A_k obtained in this manner lift to a geometric realization of A in V_g.

($\Rightarrow$) Corollary 2.28 tells us that the implication is valid for *some* factorization of $\mathcal{F}_g$ and partition of A. It turns out that any other partition of A into disk-busting sets must be the same, and corresponding factorizations $\mathcal{F}_g = G'_1 * \dots * G'_m * H'$ have the G'_k conjugate to G_k. This is proved in [10] using the W_g model. $\square$

2.4 Testing geometricity of conjugacy classes

In this section, we finally provide a test for geometricity of a multiclass A. By the previous results, we need only concern ourselves with the case where A is not separable. We will make heavy usage of the geometric interpretation of Whitehead moves, and also of the following algebraic fact, whose proof is omitted.

Proposition 2.30. *Fix a multiclass A of F_g and let $f \in \mathrm{Aut}(F_g)$ be such that $f(A)$ has minimal length among $\{\phi(A) \mid \phi \in \mathrm{Aut}(F_g)\}$ (in the canonical basis of F_g). Then f factors as a composition $f_m \circ \ldots \circ f_1$ of Whitehead automorphisms with the following property: If we set $A_0 = A$ and $A_k = f_k \circ \ldots \circ f_1(A)$ for $1 \leq k \leq m$, then the length $l(A_k)$ of A_k is always strictly less than $l(A_{k-1})$, unless A_{k-1} is already of minimal length, in which case $l(A_k) = l(A_{k-1})$.*

Whitehead proved this result in [12] using topological arguments, but a purely algebraic proof in a more modern language can be found in [7].

Since this proposition implies that any change of basis can be obtained from a sequence of Whitehead moves, and we have modeled such basis changes geometrically as changes of disk system, we conclude that any basis of $\mathcal{F}_g$ can be realized by some disk system, that is, we have proven Proposition 1.11. This was the missing piece in the proof of Proposition 2.2, so we can finally rest assured that the definition of geometricity of conjugacy classes is independent of the identification $\mathcal{F}_g \cong \pi_1(V_g)$!

We now present the main theorem of Part 2.

Theorem 2.31 (Zieschang)**.** *Let A be a disk-busting multiclass of $\mathcal{F}_g$ and $B_\circ$ a free basis for $\mathcal{F}_g$ such that $l_{B_\circ}(A)$ is minimal among all bases. Then A is geometric if and only if $\mathrm{Wh}_{B_\circ}(A)$ has an admissible planar embedding.*

Proof. ($\Leftarrow$) This is just Proposition 2.6.

($\Rightarrow$) All the hard work is contained in the following lemma.

Definition 2.32. *Let A be a multiclass and B a basis of $\mathcal{F}_g$. A Whitehead move $B \rightsquigarrow B'$ is **optimal** if $l_{B'}(A)$ is minimal among all Whitehead moves.*

Lemma 2.33. *Suppose A is a disk-busting geometric multiclass, $\mathcal{D}$ is a disk system, and $\{\alpha_j\}$ is a geometric realization of A without waves. If B is the basis represented by $\mathcal{D}$ and B' is obtained from B by an optimal Whitehead move, then there exists a disk system $\mathcal{D}'$ corresponding to B' such that the α_j form no waves with $\mathcal{D}'$.*

Let us believe this lemma for the moment. If we start with a geometric realization $\{\alpha_j\}$ of A, we know from Proposition 2.27 that there exists some disk system forming no waves with the α_j. By Proposition 2.30, one can get from the corresponding basis B to $B_\circ$ through a sequence of optimal Whitehead moves. Lemma 2.33 then tells us that we can realize all these Whitehead moves by changes of disk system without ever introducing waves, so in the end we reach a disk system $\mathcal{D}_\circ$ for $B_\circ$ having no waves. Thus, the reduced multiword Λ representing A in $B_\circ$ is geometric, and so Proposition 2.14 gives an admissible planar embedding of $\mathrm{Wh}(\Lambda) = \mathrm{Wh}_{B_\circ}(A)$. ☐

Proof of Lemma 2.33. Draw some properly embedded disk D in $Q = V_g \backslash \mathcal{D}$ modelling the optimal Whitehead move. If the resulting disk system does not form waves with the α_i, we are done. Otherwise, observe that all waves come from edges of the embedded graph $\mathrm{Wh}_B(A) \subset \partial Q$ intersecting ∂D more than once.

Suppose γ is an arc in ∂Q lifting to a wave in V_g, so γ is obtained by cutting an edge of $\mathrm{Wh}_B(A)$ along two consecutive intersections with D. If v is the vertex defining our Whitehead move, we know that v and v^{-1} lie in different components of $Q \backslash D$. Let us say that v^{-1} is in the same side of D as γ (otherwise just interchange v and v^{-1} in what follows). Moreover, suppose γ cuts ∂D into two arcs δ_1, δ_2, such that the region of ∂Q bounded by δ_2 and γ in ∂Q contains v^{-1}, and the region U bounded by δ_1 and γ contains neither v nor v^{-1}. Furthermore, we may assume that γ is innermost with these properties along δ_1, that is, no edge of $\mathrm{Wh}_B(A)$ has two consecutive intersections with D producing an arc contained in U.

We will now modify D to a different disk D' modelling the same Whitehead move, but having fewer intersections with $\mathrm{Wh}_B(A)$. The construction depends on whether U contains any vertex of $\mathrm{Wh}_B(A)$:

Case 1: If there are no vertices of $\mathrm{Wh}_B(A)$ in U, consider a small closed regular neighbourhood of $\partial D \cup \gamma$ in ∂Q. Its boundary has three components, one of which separates v^{-1} from v and U. If D' is a properly embedded disk in Q having this boundary component as $\partial D'$, then the partition of L_g induced by D' is the same as that of D, so D', suitably oriented, models the same Whitehead move.

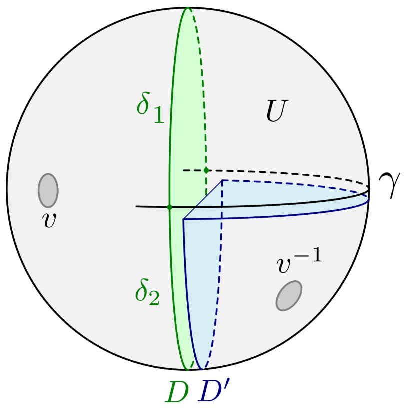


Figure 2.6: Using a wave to modify D and remove intersections with the α_j, in the case where U contains no vertex of $\mathrm{Wh}_B(A)$.

Case 2: If there are vertices of $\mathrm{Wh}_B(A)$ in U, then there are intersections of the edges of $\mathrm{Wh}_B(A)$ with the interior of δ_1, as $\mathrm{Wh}_B(A)$ is connected (from Proposition 2.19 and the fact that A is disk-busting). If we start at any such intersection point and traverse the corresponding edge towards the interior of U, then we will certainly not cross γ, and the assumption that γ is innermost tells us that we will not cross δ_1 either, so we will land in a vertex inside U. On the other hand, if we start at an intersection of $\mathrm{Wh}_B(A)$ with the interior of δ_1 and instead traverse the edge away from U, there are three possibilities:

1. we land in a vertex,
2. we cross δ_2,
3. we cross δ_1.

We shall say that the intersections of $\mathrm{Wh}_B(A)$ with the interior of δ_1 are of type 1, 2 or 3, according to where they fall in this division (note that intersections of type 3 come in pairs).

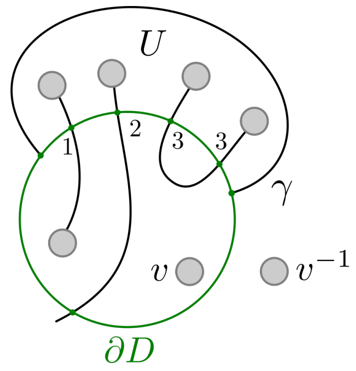


Figure 2.7: The classification of intersections of $\mathrm{Wh}_B(A)$ with δ_1.

We now show that, as a consequence of our Whitehead move being optimal, there is at least one intersection of type 2, which will allow us to construct D'. Let $\tilde{D}$ be the disk whose definition is the same as D' of Case 1. This time, the partition of L_g induced by $\tilde{D}$ is not the same as that of D, although $\tilde{D}$ still separates v from v^{-1}. Let $\tilde{B}$ be the basis obtained from B by the Whitehead move defined by $\tilde{D}$ (together with v).

We wish to compare $l_{B'}(A)$ and $l_{\tilde{B}}(A)$, so set $\Delta := l_{B'}(A) - l_{\tilde{B}}(A)$. The only difference between the cyclic words read off in the disk system representing B', and those read off in the one representing $\tilde{B}$ is the insertion of instances of $v^{\pm 1}$ in some of the former words. These insertions correspond precisely to the intersections we discussed before. Specifically:

1. Each intersection of type 1 corresponds to an insertion of a symbol $v^{\pm 1}$ which is not neighboured on either side by its inverse, so it contributes with $+1$ to Δ.

2. An intersection of type 2 corresponds to an insertion of the form $xv \rightsquigarrow xv^{-1}v$ or $v^{-1}x \rightsquigarrow v^{-1}vx$, with $x \neq v, v^{-1}$. This inserted symbol is thus part of a string of alternating symbols v, v^{-1}, and so the contribution of the intersection to Δ is with $+1$ or -1, depending on whether this string has even or odd length.

3. Intersections of type 3 come in pairs, and each pair corresponds to an insertion of two consecutive, mutually inverse symbols. Hence the overall contribution of all intersections of type 3 to Δ is 0.

The hypothesis of $B \rightsquigarrow B'$ being optimal implies that $\Delta \leq 0$. Therefore, from the above facts we deduce that there are at least as many intersections of type 2 as there are of type 1. Since some intersection exists (our very first remark since studying Case 2), we will be able to conclude that an intersection of type 2 exists if we show that not all intersections are of type 3. But if all intersections were of type 3, we would be able to use the outermost arcs they define to draw a simple closed curve ϵ on ∂Q disjoint from $\mathrm{Wh}_B(A)$, separating it into non-empty disjoint components:

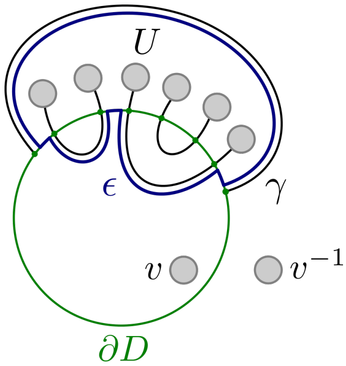


Figure 2.8: Having all intersections of δ_1 and $\mathrm{Wh}_B(A)$ be of type 3 enables the construction of ϵ, contradicting connectedness of $\mathrm{Wh}_B(A)$.

Having shown, at last, that an intersection of type 2 exists, we illustrate the modified disk D'. Again, notice that the partition it induces on L_g is the same as D.

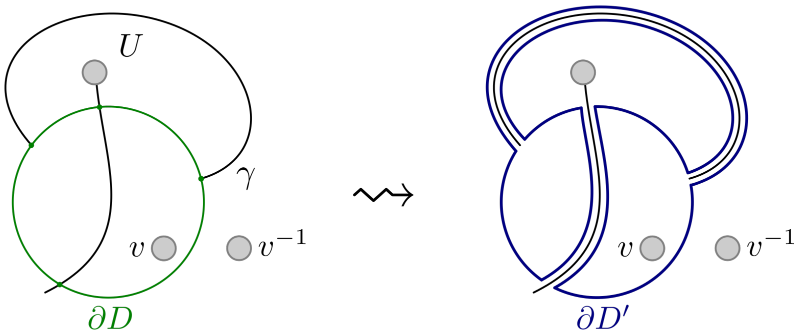


Figure 2.9: Using a wave to modify D and remove intersections with the α_j, in the case where U contains some vertex of $\mathrm{Wh}_B(A)$.

This finishes Case 2.

We can keep modifying our disk as long as it forms waves with the α_j, each time decreasing the number of intersections with the curves, so we will eventually reach a disk modelling our optimal Whitehead move, such that the resulting disk system $\mathcal{D}'$ has no waves, thus proving the lemma. □

Having Theorem 2.31 in hand, we are ready to give an algorithm to test geometricity of a multiclass A, given as a reduced multiword:

1. Using the algorithm at the end of Section 2.3.1, find a factorization $\mathcal{F}_g = G_1 * \ldots * G_m * H$ and a partition $A = A_1 \sqcup \ldots \sqcup A_m$ with each A_k disk-busting in G_k. Perform the following steps for each A_k, and then geometricity of A is equivalent to geometricity of each A_k by Proposition 2.29.

2. For A disk-busting in $\mathcal{F}_g$, find a basis B for which $l_B(A)$ is minimal by searching through all possible Whitehead moves (of which there are only finitely many) for one that decreases the length of A, applying it, and repeating until no Whitehead move decreases length. The resulting basis gives A minimal length as a consequence of Proposition 2.30.

3. If $l_B(A)$ is minimal, test geometricity of the reduced multiword Λ representing A in B, as explained at the end of Section 2.2.3. By Theorem 2.31, geometricity of A is equivalent to that of Λ.

3 The induced curve graph

3.1 Curves on a surface

For this third part, we move towards relating distinct (as in non-isotopic) geometric realizations of the same conjugacy class of $\mathcal{F}_g$. In this section, however, we restrict our attention to curves on a closed orientable surface S, not necessarily regarded as the boundary of a handlebody.

Definition 3.1. *The* ***curve graph*** $\mathcal{C}(S)$ *has as vertices the isotopy classes of (unoriented) essential simple closed in* S*, with an edge between* a, b *whenever* $i(a, b) = 0$*.*

The curve graph has been used in understanding the topology of 3-manifolds and their Heegaard splittings [6]. Here, we present only the following property, whose proof we adapt from [2].

Proposition 3.2. *If* S *is not a torus, then* $\mathcal{C}(S)$ *is connected.*

Proof. Given isotopy classes a, b, we will prove that there exists a path between them by induction on $i(a, b)$. If $i(a, b) = 0$, then they are connected by an edge and there is nothing to show. For $i(a, b) > 0$, it suffices to show that there exists an essential isotopy class c with $i(a, c) < i(a, b)$ and $i(b, c) < i(a, b)$.

Let a, b be represented by simple closed curves α, β with minimal intersection. If $i(a, b) = 1$, then a closed regular neighbourhood of $\alpha \cup \beta$ is homeomorphic to a torus with one open disk removed. Since we are assuming S is not a torus, the boundary component γ cannot bound a disk in S, so we may take c as the isotopy class of γ, for which $i(a, c) = i(b, c) = 0$.

When $i(a, b) \geq 2$, orient α, β and look at two intersections that are consecutive along β. The two situations to consider are when α crosses β twice in the same direction, or in opposite directions.

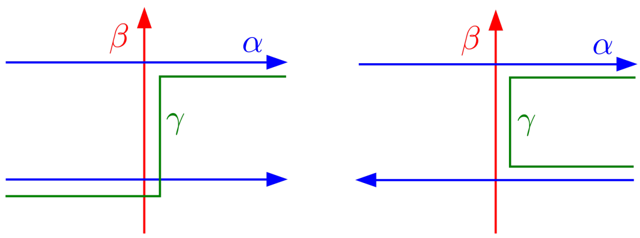


Figure 3.1: The construction of γ, the case where $i(a, b) \geq 2$.

In the first case, we draw a curve γ that always lies to the right of α, except that it skips one of the arcs of α between the two intersection points, as illustrated. Since γ intersects α at a single point, it is essential, and for its isotopy class c we have $i(a, c) = 1$. Our construction also ensures that γ skips one of the intersections with β, so $i(b, c) \leq i(a, b) - 1$, as desired.

In the second case, we construct γ similarly, but justify it being essential by noting that otherwise α and β would form a bigon and not be in minimal position. We get $i(a, c) = 0$ and $i(b, c) \leq i(a, b) - 2$, so we are done. ☐

The following construction will play an important role in the forthcoming.

Definition 3.3. *Let α, β be disjoint simple closed curves in S, and δ an arc connecting some point in α to some point in β, but which does not otherwise intersect α or β. A regular neighbourhood of $\alpha \cup \beta \cup \gamma$ has three simple closed curves as boundary components, one of which is isotopic to α, and another to β. We denote the third one by $\alpha \underset{\delta}{*} \beta$, and call it the **slide** of α over β along δ.*

The curve $\alpha \underset{\delta}{*} \beta$ is only defined up to isotopy. If we have a preferred orientation for α, we may use it to induce an orientation for $\alpha \underset{\delta}{*} \beta$. Moreover, if α has basepoint $p := \alpha \cap \delta$, we may modify $\alpha \underset{\delta}{*} \beta$ slightly in order to make it start and end at p, thus turning it into a representative of $[\alpha] \cdot [\delta \cdot \beta^{\pm 1} \cdot \delta^{-1}] \in \pi_1(S, p)$ (where δ is oriented from α to β, β has $\delta \cap \beta$ as basepoint, and the sign depends on the orientation of β).

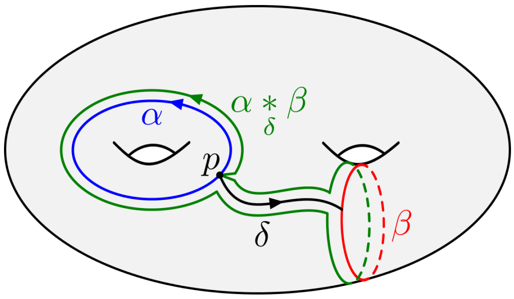


Figure 3.2: The slide construction.

It is interesting to note the following fact.

Proposition 3.4. *In the setting of the previous paragraph, we have $[\alpha] = \left[\left(\alpha \underset{\delta}{*} \beta\right) \underset{\delta}{*} \beta\right]$ in $\pi_1(S, p)$.*

Proof. We only need to observe that, if we traverse the curve on the right hand side, then the second time we track β, we do it with the opposite orientation of the first passing. Hence, the corresponding homotopy class is

$$[\alpha] \cdot [\delta \cdot \beta^{\pm 1} \cdot \delta^{-1}] \cdot [\delta \cdot \beta^{\mp 1} \cdot \delta^{-1}],$$

which yields $[\alpha]$ when all cancellation is done. $\square$

3.2 Edges in $\mathcal{C}_w$

We now return to thinking of the surface of genus g as the boundary of V_g.

Definition 3.5. *Let w be a conjugacy class in $\mathcal{F}_g$. The **induced curve graph** $\mathcal{C}_w$ is the full subgraph of $\mathcal{C}(\partial V_g)$ whose vertices are isotopy classes of curves representing w (when suitably oriented).*

It is immediate from this definition that $\mathcal{C}_w$ is non-empty if and only if w is geometric. In that case, and in contrast to what was seen in the previous section, $\mathcal{C}_w$ is in general not connected. We now seek to better understand its connected components, for which we shall make use of the slide construction from the previous section, in the case where the curve β is a meridian.

Definition 3.6. *A simple closed curve in ∂V_g is a* ***meridian*** *if it bounds a disk in V_g.*

Lemma 3.7. *An oriented simple closed curve α in ∂V_g represents the trivial conjugacy class of $\mathcal{F}_g$ if and only if it is a meridian.*

Proof. The only non-obvious direction is ($\Rightarrow$). Using Proposition 2.27, we get a disk system $\mathcal{D}$ with which α forms no waves, hence the disks in $\mathcal{D}$ are disjoint from α. Cutting V_g along $\mathcal{D}$ makes α into an embedded simple closed curve in the boundary of a 3-ball, where it obviously bounds a properly embedded disk. This disk now lifts to V_g. $\square$

The main result of this section gives an alternative characterization of the edges of $\mathcal{C}_w$:

Theorem 3.8. *Let α, β be disjoint simple closed curves on ∂V_g. Then α, β represent the same conjugacy class of $\mathcal{F}_g$ if and only if β is isotopic (in ∂V_g) to $\alpha \underset{\epsilon}{*} \gamma$, for some meridian γ and arc ϵ.*

It follows from this theorem that the second condition is symmetric with respect to α and β, but that could already be deduced from Proposition 3.4. Another easy consequence is the following:

Corollary 3.9. *For w a geometric conjugacy class, $\mathcal{C}_w$ has edges if and only if w is separable.*

Proof. ($\Rightarrow$) If the isotopy classes of two geometric realizations of w are connected by an edge in $\mathcal{C}_w$, then they are of the form $[\alpha]$, $\left[\alpha \underset{\epsilon}{*} \gamma\right]$ for α a simple closed curve in ∂V_g and γ a meridian. These classes can only be distinct if γ is essential, so by Proposition 2.18 we conclude w is separable.

($\Leftarrow$) If w is separable and α is a geometric realization of w, then Proposition 2.25 gives an essential, properly embedded disk D disjoint from α. If D is part of some disk system (so $V_g \backslash D$ is connected), then ∂D is homologically non-trivial in ∂V_g, and on $H_1(V_g)$ we have, for ϵ an arc connecting α to ∂D as above,

$$\left[\alpha \underset{\epsilon}{*} \partial D\right] = [\alpha] \pm [\partial D] \neq [\alpha].$$

Thus, the isotopy classes of α and $\alpha \underset{\epsilon}{*} \partial D$ are two distinct vertices of $\mathcal{C}_w$ connected by an edge.

In the case where D cuts V_g into two handlebodies of positive genus (so ∂D is homologically trivial in ∂V_g), one can find a properly embedded disk D' as in the first case in the connected component of $V_g \backslash D$ that does not contain α. We then carry out the same argument. $\square$

Our proof of the implication ($\Rightarrow$) in Theorem 3.8 will have a somewhat combinatorial flavor, making use of ideas we have touched on when discussing Whitehead graphs. One can, however, use purely topological methods to prove the following weaker statement: For every two disjoint, non-isotopic simple closed curves $\alpha, \beta \subset \partial V$ representing the same conjugacy class w, there exists a disjoint essential meridian. This would be enough to deduce Corollary 3.9.

This quicker proof would roughly go as follows: We know α, β are homotopic in V_g, and one can use techniques in geometric topology to obtain from one such homotopy a properly embedded annulus $A \in V_g$ whose boundary is $\alpha \cup \beta$. We then consider any properly embedded disk $D \subset V_g$ transverse to A, and modify it so as to eliminate intersections with A: First we remove circles in $D \cap A$ (using that V_g is aspherical), and then intervals whose ends are on the same boundary component of A (surgering D along intervals that are innermost in A). If $D \cap A$ is still non-empty, then all its components are intervals with one end in α and one in β, and there is one such interval δ that is innermost in D. In that case, one can show that $\alpha \underset{\delta}{*} \beta$ is an essential meridian.

Proof of Theorem 3.8. ($\Leftarrow$) Suppose that $\beta \simeq \alpha \underset{\epsilon}{*} \gamma$. Let $p = \alpha \cap \epsilon$ be the basepoint for both α and $\alpha \underset{\epsilon}{*} \gamma$, and give $\alpha \underset{\epsilon}{*} \gamma$ the orientation induced from α. On $\pi_1(\partial V_g, p)$, we have $\left[\alpha \underset{\epsilon}{*} \gamma\right] = [\alpha] \cdot [\epsilon \cdot \gamma \cdot \epsilon^{-1}]$ (for one possible orientation of γ). The second factor becomes trivial in $\pi_1(V_g, p)$, so we conclude β is freely homotopic to α in V_g.

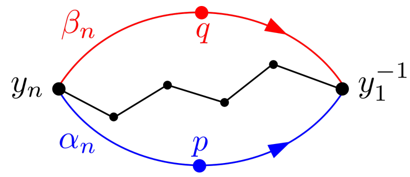
($\Rightarrow$) Using Proposition 2.27, find a disk system $\mathcal{D}$ for V_g such that α, β form no waves with $\mathcal{D}$, so $\omega_\alpha, \omega_\beta$ are reduced. Also, choose basepoints p, q for α, β, respectively, such that the (non-cyclic) words $\omega_\alpha, \omega_\beta$ are the same. We will write $\omega_\alpha = \omega_\beta = y_1 \dots y_n$, where each y_j stands for a symbol $x_i^{\pm 1}$. We need one further assumption, whose justification will take most of the remainder of the proof:

Claim: p, q can be chosen so that there is a path δ from p to q in ∂V_g that is otherwise disjoint from α, β, and from the disks in $\mathcal{D}$.

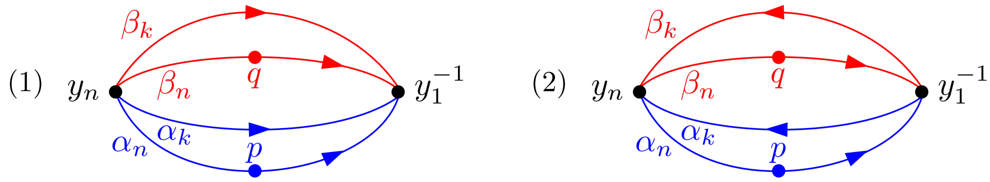
On the boundary of $Q = V_g \backslash \mathcal{D}$, after collapsing the disks to points, we see a planar embedding of a graph Γ that resembles $\mathrm{Wh}(\{\omega_\alpha\})$, except that all edges are doubled. Label the vertices of Γ as $x_i^{\pm 1}$ accordingly. Even though Γ is not exactly a Whitehead graph, much of the theory we developed still holds, namely, there are bijections between the sets of half-edges at inverse vertices, and these bijections reverse the cyclic orders induced from the planar embedding. Also, there are no loops, and two of the edges of Γ have marked points p, q. We colour the edges depending on whether they come from α (blue) or β (red), there being n edges of each colour. We denote by α_j the blue edge from y_j to y_{j+1}^{-1} corresponding to the syllable $y_j y_{j+1}$, and similarly for the red edges (the indices are of course modulo n).

We may assume that one of the bigons formed by α_n, β_n in ∂Q is innermost among all bigons formed by pairs of edges α_j, β_j (note the same index j). Otherwise, let α_k, β_k be innermost in this sense and replace the basepoints p, q with points in α_k, β_k (shifting the labelling of the α_j, β_j accordingly). This induces the same cyclic permutation on both words $\omega_\alpha, \omega_\beta$, so they remain equal to one another.

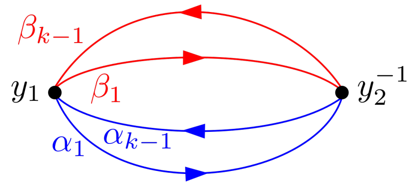
Now, if there is a simple path on ∂Q from p to q that is otherwise disjoint from Γ, we lift it to ∂V_g and take that as δ. If not, consider the disk U bounded by α_n, β_n witnessing that our bigon is innermost. If we have failed to construct δ, there must be a path of Γ inside U connecting the two endpoints y_n, y_1^{-1} of α_n (and β_n). For concreteness, assume the first edge of this path lies between α_n and β_n in the cyclic order at y_n (the other case is analogous):



If the path has length at least 2, then its second vertex is different from y_n, y_1^{-1}. Any edge having that vertex as an endpoint (say, for instance, that it is blue) has its red counterpart also inside U. This contradicts our bigon being innermost, so we conclude that the path is comprised of a single edge (which we assume is blue – the other case is analogous). Hence, Γ contains one of the following two subgraphs:



We proceed to see why situation (2) cannot occur. The existence of this subgraph tells us that $y_k = y_1^{-1}$. In the cyclic order at y_1, we find that α_{k-1} lies between α_1, β_1, and β_{k-1} lies between β_1, α_1. Therefore, α_{k-1} and β_{k-1} are in different sides of the bigon formed by α_1, β_1, so they must both be incident with y_2^{-1}:



This shows $y_{k-1} = y_2^{-1}$. Repeating this argument enough times, we eventually reach one of two contradictions: Either k is even and we obtain $y_{\frac{k}{2}+1} = y_{\frac{k}{2}}^{-1}$ (contradicting ω_α being reduced), or k is odd and we get $y_{\frac{k+1}{2}} = y_{\frac{k+1}{2}}^{-1}$, which is also absurd. This rules out (2).

A similar argument with (1) shows that $y_{j+k} = y_j$ for all j, that is, our word is invariant with respect to shifting by k positions.

At this point, our only obstruction to constructing δ is the existence of edges from y_n to y_1^{-1} between α_n and β_n (with respect to the cyclic ordering around y_n). Some of them might be blue, others red, but there have to exist two consecutive ones of a different colour. By what we just saw, if we replace p, q by points in those consecutive edges, $\omega_\alpha, \omega_\beta$ remain unchanged. With these new basepoints p', q' we are surely able to construct δ:

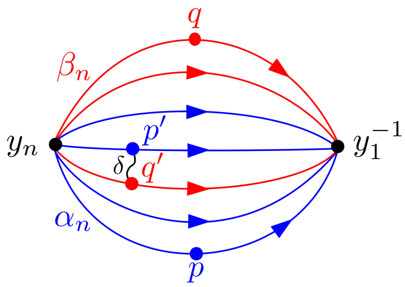


Figure 3.3: Using two edges from y_n to y_1^{-1} of different color, that are adjacent in the cyclic order at y_n, to find new basepoints p', q' for α, β and construct δ.

This finishes the proof of the claim.

Now, let $\gamma = \alpha \underset{\delta}{*} \beta$, with orientation induced from α. A glance at the bigon formed by α_0 and β_0, together with the arc δ, reveals that γ traverses β with opposite orientation:

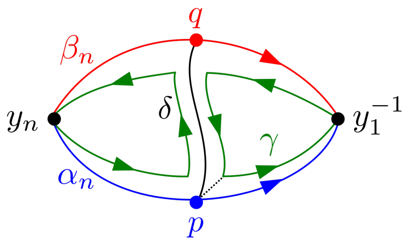


Figure 3.4: The image of γ after cutting V_g along $\mathcal{D}$ and collapsing the E_i, F_i. The dotted path connecting p to γ is the one we will take as ϵ.

Therefore, since we constructed δ so as to not intersect any disk in $\mathcal{D}$, we conclude $\omega_\gamma = \omega_\alpha \omega_\beta^{-1} = \omega_\alpha \omega_\alpha^{-1}$, so γ represents the trivial element in $\mathcal{F}_g$, whence it is a meridian by Lemma 3.7.

A dotted path was added in the above picture, which is the one we shall take as ϵ. On $\pi_1(\partial V_g, p)$, we then have

$$\left[\alpha \underset{\epsilon}{*} \gamma\right] = [\alpha \cdot \epsilon \cdot \gamma^{-1} \cdot \epsilon^{-1}] = [\alpha \cdot \delta \cdot \beta \cdot \delta^{-1} \cdot \alpha^{-1}].$$

The last path is freely homotopic to β, so this finishes the proof. $\square$

We can use Theorem 3.8 to give a homological invariant for the connected components of $\mathcal{C}_w$, in the case where w is the conjugacy class of a non-zero power of a primitive element.

Definition 3.10. *An element of $\mathcal{F}_g$ is **primitive** if it is part of some free basis.*[3]

We will denote the standard generators of $H_1(\partial V_g) \cong \mathbb{Z}^{2g}$ as illustrated for $g = 3$.

[3]This is not widespread terminology. Often, an element of a group is called primitive if it is not a proper power of any other element.

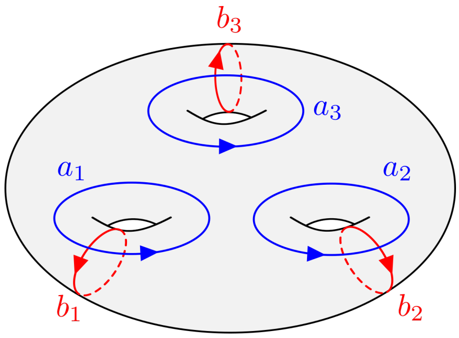


Figure 3.5: The standard generators of $H_1(\partial V_3)$.

Let w be the conjugacy class of some non-zero power of a primitive element, and choose a disk system $\mathcal{D} = (D_1, \ldots, D_g)$ such that w is represented by the cyclic word x_1^k in the corresponding basis ($k \neq 0$). By Proposition 1.9, we may assume that the ∂D_i, with the induced orientations from D_i, are the curves drawn above representing the homology classes b_i.

Corollary 3.11. *Let α, β be geometric realizations of w as above. If their isotopy classes are in the same connected component of $\mathcal{C}_w$, then they represent homology classes whose coefficients in b_1 are the same.*

As an example with $g = 2$, each of the isotopy classes depicted below lies in a distinct connected component of $\mathcal{C}_x$:

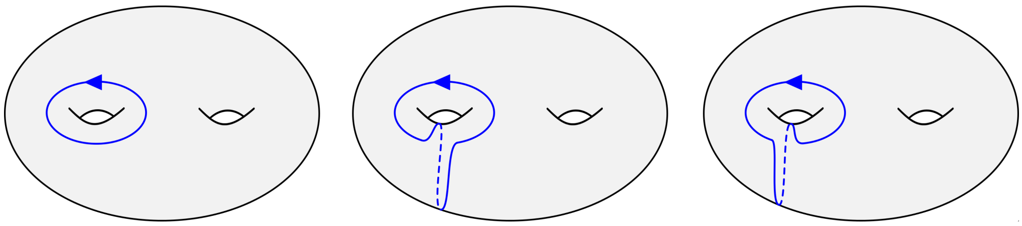


Figure 3.6: Three curves representing the conjugacy class of $x \in \mathcal{F}_2$. Their classes in $H_1(\partial V_2)$ are, from left to right, $a_1, a_1 - b_1, a_1 + b_1$.

Proof. It is enough to prove the result in the case where $[\alpha], [\beta]$ are connected by an edge. By Theorem 3.8, we can write $[\beta] = \left[\alpha \underset{\epsilon}{*} \gamma\right]$ with γ a meridian. Fixing some orientation for γ and using square brackets to denote homology classes, we have $\left[\alpha \underset{\epsilon}{*} \gamma\right] = [\alpha] \pm [\gamma]$. Now, $[\alpha]$ has zero component in the a_i for $i \neq 1$, and the coefficient in a_1 is k. Since γ is a meridian, $[\gamma]$ has zero component in all the a_i. We may therefore write

$$[\alpha] = ka_i + \sum_{j=1}^{g} \lambda_j b_j, \qquad [\gamma] = \sum_{j=1}^{g} \mu_j b_j.$$

Therefore, we need only to show is that $\mu_1 = 0$, which can be seen by looking at intersection products. We know α and γ are disjoint, so $[\alpha] \odot [\gamma] = 0$, but since $a_i \odot b_i = 1$, we have

$$[\alpha] \odot [\gamma] = \left(ka_i + \sum \lambda_j b_j\right) \odot \sum \mu_j b_j = k\mu_i.$$

As $k \neq 0$, we conclude $\mu_i = 0$. $\square$

3.3 $\mathcal{C}_w$ and the handlebody group

One of the reasons to study the induced curve graph of a conjugacy class is its connections to the mapping class group of ∂V_G and the handlebody group. We define these concepts in the present section, and explore some relations with $\mathcal{C}_w$.

Definition 3.12. *The* ***mapping class group*** $\mathrm{Mod}(S)$ *of a closed surface S is the group of orientation-preserving self-homeomorphisms of S, up to homotopy. The* ***handlebody group*** $\mathcal{H}_g < \mathrm{Mod}(\partial V_g)$ *is comprised of the classes of maps that extend to self-homeomorphisms of V_g.*

The following lemma tells us that this notion is well-defined.

Lemma 3.13. *If $f : \partial V_g \to \partial V_g$ is a homeomorphism that extends to V_g and f' is homotopic to f, then f' also extends to V_g.*

Proof. Let $F : V_g \to V_g$ be an extension of f to V_g, and $h : V_g \times [0,1] \to V_g$ be such that $h(-,0) = f, h(-,1) = f'$. We consider a closed regular neighbourhood $N \cong \partial V_g \times [0,1]$ of ∂V_g, with $i : \partial V_g \times [0,1] \to V_g$ taking $\partial V_g \times \{1\}$ to ∂V_g. To obtain an extension F' of f', we define F' as F outside of N, and in N we map

$$i(x,t) \mapsto \begin{cases} F(i(x,2t)) & \text{if } 0 \leq t \leq \frac{1}{2} \\ h(x,2t-1) & \text{if } \frac{1}{2} \leq t \leq 1 \end{cases}.$$

□

To determine whether an element of $\mathrm{Mod}(\partial V_g)$ is in $\mathcal{H}_g$, we use the following proposition.

Proposition 3.14. *The following conditions on a self-homeomorphism $f : \partial V_g \to \partial V_g$ are equivalent:*

1. *f extends to a self-homeomorphism of V_g,*
2. *f takes all meridians to meridians,*
3. *For some disk system $\mathcal{D} = (D_1, \dots, D_g)$, f takes each ∂D_i to a meridian.*

Proof. The implications $(1 \Rightarrow 2)$ and $(2 \Rightarrow 3)$ are trivial.

To prove $(3 \Rightarrow 1)$, we first show that the meridians $f(\partial D_i)$ collectively bound a disk system $\mathcal{D}' = (D'_1, \dots, D'_g)$. By hypothesis, there exist properly embedded disks D'_i having $f(\partial D_i)$ as boundary, however their interiors may intersect. Suppose D'_2 intersects D'_1 non-trivially. One can find a curve of intersection γ that is innermost with respect to D'_1 and, since V_g is aspherical, we can modify D'_2 by homotoping the disk bounded by γ through D'_1, decreasing the number of components of $D'_1 \cap D'_2$. We repeat this until all intersections of D'_1 with D'_2 are removed, and then remove all intersections of D'_1 with the remaining disks. Cutting V_g along D'_1 allows us to now play the same game with D'_2 without creating new intersections with D'_1. If we do this with all disks, in the end all D'_i are disjoint (and the $\partial D'_i$ have not been changed). Since cutting ∂V_g along the ∂D_i yields a sphere with $2g$ open disks removed, and this property is preserved by f, we conclude the D'_i form a disk system.

We now extend f to V_g. First, define it in the interior of the D_i as a collection of homeomorphisms $D_i \to D_i'$ extending the map on the boundary. Then, cut V_g along $\mathcal{D}$ or along $\mathcal{D}'$ in order to obtain balls Q, Q'. We have a homeomorphism $\partial Q \to \partial Q'$, which easily extends to the interior of Q, lifting to a homeomorphism of V_g. $\square$

An important type of mapping classes are **Dehn twists**, constructed in the following manner. Start with a simple closed curve α in the surface S and find a closed regular neighbourhood N of α, which is to be identified with $[0,1]\times S^1$ in an orientation-preserving fashion (where $[0,1] \times S^1$ carries the orientation induced from the standard orientations of $[0,1]$ and S^1). Now, consider the self-homeomorphism ϕ of $N \cong [0,1] \times S^1$ given by

$$(t, e^{i\theta}) \mapsto (t, e^{i(\theta+2\pi t)}),$$

and notice that it fixes $\partial N \cong \{0,1\} \times S^1$ pointwise. This means we can define an orientation-preserving homeomorphism $T_\alpha : S \to S$ by using ϕ on N and extending it to the identity on the rest of S. It turns out that the mapping class of T_α is independent of the choice of N, of the identification with $[0,1] \times S^1$, and of the choice of α within its isotopy class. Dehn twists are extensively discussed in [2], where it is shown that they have infinite order in $\mathrm{Mod}(S)$, and that $\mathrm{Mod}(S)$ is always generated by finitely-many Dehn twists.

We illustrate the effect of T_α on a curve β that intersects it. Explicitly, T_α inserts a "left-handed spiral" about α at every intersection of β with α:

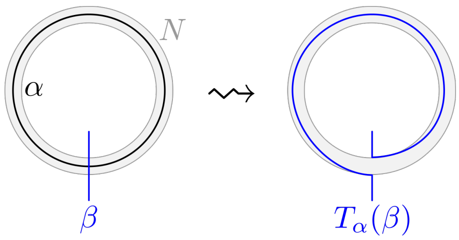


Figure 3.7: For each intersection of α with β, T_α appends to β a left-handed spiral about α.

Similarly, T_α^k would insert a spiral going around α k times, where for negative values of k this would correspond to a "right-handed" spiral. It should be stressed that this does not depend on any sort of orientation of α. With this geometric intuition, we can see the effect of Dehn twists on homology.

Lemma 3.15. *If α, β are two simple closed curves in S, then on $H_1(S)$ we have*

$$T_\alpha([\beta]) = [\beta] + ([\beta] \odot [\alpha])[\alpha].$$

Proof. When traversing β, an intersection where α crosses from right to left contributes with $+1$ to $[\beta] \odot [\alpha]$. From the above picture we see that the spiral that T_α appends has the same orientation as α, so this intersection contributes to $T_\alpha([\beta])$ with $+[\alpha]$. The opposite occurs for crossings from left to right. $\square$

We now present the promised connection between $\mathcal{C}_w$ and $\mathcal{H}_g$.

Proposition 3.16. *If $[\alpha], [\beta]$ are in the same connected component of $\mathcal{C}_w$, then $T_\beta T_\alpha^{-1} \in \mathcal{H}_g$.*

Proof. It is sufficient to show the result in the case where $[\alpha], [\beta]$ are connected by an edge, because then for any path $[\alpha] = [\alpha_0], [\alpha_1], \ldots, [\alpha_k] = [\beta]$ in $\mathcal{C}_w$, we get

$$T_\beta T_\alpha^{-1} = (T_{\alpha_k} T_{\alpha_{k-1}}^{-1}) \ldots (T_{\alpha_2} T_{\alpha_1}^{-1})(T_{\alpha_1} T_{\alpha_0}^{-1}) \in \mathcal{H}_g.$$

If $[\alpha], [\beta]$ are connected by an edge, we can use Theorem 3.8 to write $[\beta] = \left[\alpha \underset{\epsilon}{*} \gamma\right]$. We will check condition 2 of Proposition 3.14 with a disk system $\mathcal{D}$ disjoint from γ (its existence is assured by Proposition 2.27). Moreover, we may assume that $\mathcal{D}$ is disjoint from the arc ϵ as well, by dragging any such intersections along ϵ, towards α, starting with those closest to α:

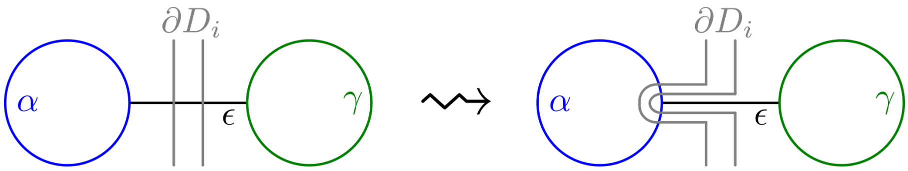


Figure 3.8: Modifying ∂D_i to remove intersections with ϵ.

We now draw $\alpha \underset{\epsilon}{*} \gamma$ and see what happens to ∂D_i at each intersection with α when $T_\beta T_\alpha^{-1}$ is applied. Since γ is a meridian, α and $\alpha \underset{\epsilon}{*} \gamma$ are isotopic (in V_g). Thus, each intersection of ∂D_i with α contributes to a modification of ∂D_i that, from the point of view of $\pi_1(V_g)$, amounts to appending one curve followed by its inverse, as is illustrated:

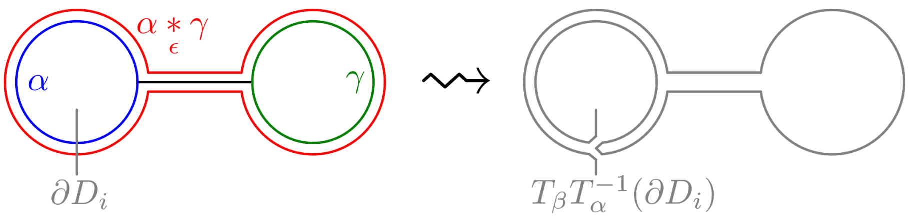


Figure 3.9: The effect of $T_\beta T_\alpha^{-1}$ on ∂D_i.

Hence, $T_\beta T_\alpha^{-1}(\partial D_i)$ remains a meridian, and since this is true for all intersections and all D_i, we conclude $T_\beta T_\alpha^{-1} \in \mathcal{H}_g$. □

We can use this criterion to show that the homological invariant from the previous section is not a perfect invariant. Explicitly, we will give two simple closed curves α, β in ∂V_2, both representing the conjugacy class of a primitive element and, on homology, having the same coefficient on the relevant generator. Yet we will exhibit a meridian γ for which $T_\beta T_\alpha^{-1}(\gamma)$ is not a meridian.

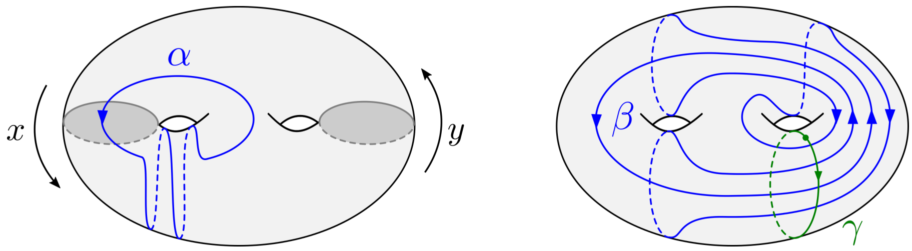


Figure 3.10: Two isotopy classes representing the conjugacy class of x and having the same coefficient for b_1 on $H_1(\partial V)$. Inspecting the effect of $T_\beta T_\alpha^{-1}$ on the meridian γ reveals that they lie in different components of $\mathcal{C}_x$.

For the disk system indicated on the left, we see $\omega_\alpha = x$ and $\omega_\beta = xyy^{-1}y^{-1}y$, so both curves represent the conjugacy class of x in $\mathcal{F}_g$. On $H_1(V_g)$,

$$[\alpha] = a_1 + 2b_1, \qquad [\beta] = a_1 + 2b_1 - b_2,$$

so they do indeed have the same value for our invariant.

However, let us write down the cyclic word for the curve $\gamma' := T_\beta T_\alpha^{-1}(\gamma)$. Since α is disjoint from γ, this is just $T_\beta(\gamma)$. We compute $\omega_{\gamma'}$ by the following recipe:

1. Start at some point in γ (a chosen basepoint was marked on the picture),
2. Traverse γ following its orientation until either:
   (a) an intersection with β is reached, in which case: turn left and traverse β (possibly against its orientation) until back at the same intersection. Return to step 2.
   (b) the basepoint is reached, in which case the trip is finished.

All signed intersections with the D_i along the trip are to be tracked, and the resulting word is the desired $\omega_{\gamma'}$. The result of this computation is

$$\omega_{\gamma'} = (yy^{-1}x^{-1}y^{-1}y)(yxyy^{-1}y^{-1})(yy^{-1}y^{-1}yx)(yyy^{-1}x^{-1}y^{-1}).$$

This does not cancel to the empty word, so γ' is not a meridian. Hence, $T_\beta T_\alpha^{-1} \notin \mathcal{H}_g$, and so by Proposition 3.16 we conclude $[\alpha], [\beta]$ are not in the same connected component of $\mathcal{C}_x$.

We finish with a generalization of the homological criterion of the last section.

Theorem 3.17. *If w is not in the commutator of $\mathcal{F}_g$, and α, β are geometric realizations of w satisfying $T_\beta T_\alpha^{-1} \in \mathcal{H}_g$, then on $H_1(\partial V_g)$ we have $[\alpha] \odot [\beta] = 0$.*

Note that because of Proposition 3.16, this holds in particular if the isotopy classes of α, β are in the same component of $\mathcal{C}_w$. Corollary 3.11 then becomes a particular case of this theorem.

Proof. If we choose any disk system for V_g, the fact that w is not in the commutator of $\mathcal{F}_g$ implies that there is some disk D for which the number of positive and negative intersections of D with α is not the same. In other words, we have $[\partial D] \odot [\alpha] \neq 0$.

Also, notice that since the cyclic words represented by α and β on this disk system are equivalent, we have $[\partial D] \odot [\alpha] = [\partial D] \odot [\beta]$.

We now apply Lemma 3.15 to see the effect of $T_\beta T_\alpha^{-1}$ on $[\partial D] \in H_1(\partial V_g)$. For legibility, square brackets will be omitted in the following three lines, but all curves are to be read as the classes they represent in $H_1(\partial V)$.

$$\begin{aligned} T_\beta T_\alpha^{-1}(\partial D) &= T_\beta(\partial D - (\partial D \odot \alpha)\alpha) \\ &= \partial D + (\partial D \odot \beta)\beta - (\partial D \odot \alpha)(\alpha + (\alpha \odot \beta)\beta) \\ &= \partial D + (\partial D \odot \alpha)(\beta - \alpha - (\alpha \odot \beta)\beta) \end{aligned}$$

If $\iota : \partial V_g \hookrightarrow V_g$ is the inclusion, we have:

1. $\iota T_\beta T_\alpha^{-1}([\partial D]) = 0$, since $T_\beta T_\alpha^{-1}(\partial D)$ is a meridian,
2. $\iota([\partial D]) = 0$,
3. $\iota([\alpha]) = \iota([\beta])$, since these curves are homotopic in V_g.

Therefore, in $H_1(V_g)$ we have

$$([\partial D] \odot [\alpha])([\alpha] \odot [\beta])\iota([\beta]) = 0.$$

Since the conjugacy class w of $\mathcal{F}_g$ represented by β is not contained in the commutator, which is precisely the kernel of $\pi_1(V_g) \to H_1(V_g)$, we conclude $\iota([\beta]) \neq 0$. By choice of D, the first factor of the above equation is not 0 either. As $H_1(V_g)$ is torsion-free, we conclude $[\alpha] \odot [\beta] = 0$. $\square$

As an example, consider the following two simple closed curves α, β in ∂V_2 representing the conjugacy class of x^2y^2 (in the basis given by the usual disk system).

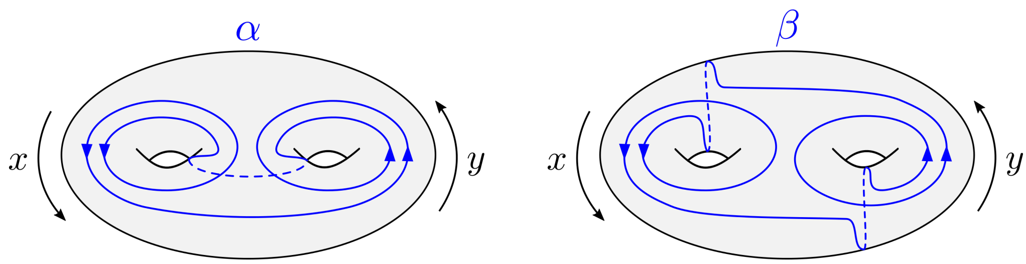


Figure 3.11: Two isotopy classes representing the conjugacy class w of x^2y^2. On $H_1(\partial V_g)$, their intersection product is non-zero, so they lie in different components of $\mathcal{C}_w$.

Since x^2y^2 is not a power of a primitive element (drawing the Whitehead graph immediately reveals that $\{w\}$ is not separable), Corollary 3.11 is not applicable. Still, w is not in the commutator subgroup, and Theorem 3.17 can be used. On $H_1(\partial V_g)$, we have:

$$[\alpha] = 2a_1 + b_1 + 2a_2 - b_2, \qquad [\beta] = 2a_1 + b_1 + 2a_2 + b_2.$$

Computing the intersection product yields $[\alpha] \odot [\beta] = 4$, so $[\alpha]$ and $[\beta]$ are not in the same component of $\mathcal{C}_w$.

Using Theorem 3.17 with this specific example is of course excessive, as we observed $\{w\}$ is disk-busting, and as we saw earlier, that implies $\mathcal{C}_w$ has no edges. However, if we instead regard w as a conjugacy class in, say, $\mathcal{F}_3$, and draw analogous curves α', β' in ∂V_3, this "absence of edges" argument ceases to hold, but Theorem 3.17 nevertheless allows us to conclude $[\alpha'], [\beta']$ are in different components.

A The genus of a ribbon graph

This appendix describes a method to compute the genus of a ribbon graph. All graphs are assumed to be finite.

Proposition A.1. *If Γ is a connected graph embedded in S^2, there is a homeomorphism $S^2 \cong (\bigcup_i D_i) \cup_f \Gamma$ restricting to the identity on Γ, where the D_i are 2-cells and $f : \bigcup_i \partial D_i \to \Gamma$ is an attaching map that depends only on the induced ribbon structure on Γ.*

Proof. A closed regular neighbourhood of Γ in S^2 can be thought of as a mapping cylinder M_f for $f : Y \to \Gamma$, with Y a compact 1-dimensional submanifold of S^2, which we identify with $Y \times \{1/2\} \subseteq M_f$. Hence, Y is a finite disjoint union of circles γ_i, each bounding an embedded disk in S^2 that contains Γ in its interior, and another disk D_i that is disjoint from Γ. Now, M_f is homeomorphic to the space obtained by collapsing $Y \times [1/2, 1] \subseteq M_f$ onto Γ, and this in turn is homeomorphic to the attachment of $Y \times [0, 1/2] \subseteq M_f$ to Γ via $f : Y \times \{1/2\} \to \Gamma$. Both these homeomorphisms fix $Y \times \{0\}$ and Γ pointwise, so we get an extended homeomorphism $S^2 \cong (\bigcup_i D_i) \cup_f \Gamma$ respecting the inclusion of Γ on each side.

We now explain how to read off f from the ribbon structure. Let each circle γ_i carry the orientation induced from D_i, so Γ always lies to the right of γ_i. Each attaching map $f_i : \gamma_i \to \Gamma$ cuts γ_i into a cyclic sequence of arcs that alternate between being mapped onto vertices and onto edges of Γ. These latter maps onto the edges restrict to homeomorphisms in the interior of the arcs. As an example, for the plane graph below, there two 2-cells being attached, with γ_1 being split into 6 arcs, and γ_2 into 10 arcs.

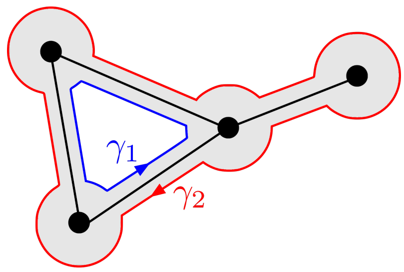


Figure A.1: A plane graph Γ, together with curves γ_i bounding a closed neighbourhood. In this illustration, the circular segments are mapped by f_i to vertices, and the straight ones to edges.

Since the quotient map $D_i \to D_i/\alpha$ that collapses an arc $\alpha \subset \partial D_i$ is a homeomorphism, the previous description of the attaching maps f_i can be made simpler by just ignoring the arcs that are mapped to vertices. It can also be shown that the precise choice of homeomorphisms from each arc of γ_i onto the respective edge of Γ is immaterial, provided that orientations are respected.

We are therefore only concerned with determining the cyclic sequence of edges being traced by each γ_i, together with the orientations in which they are being traversed. That information can be encoded by a set of cyclic sequences s^i of half-edges of Γ, where the occurrence of a half-edge h is to be interpreted as γ_i traversing the edge containing h in the direction away from h, towards its opposite $\overline{h}$. The s^i are completely determined by the following two rules:

1. Each edge of Γ is traversed exactly once in each direction (either by the same or by distinct γ_i),
2. Upon arriving at a vertex v via a half-edge h, γ_i is to traverse the half-edge that precedes h in the cyclic order of the half-edges at v.

Indeed, rule 1 tells us that each half-edge occurs precisely once within the s^i, and rule 2 gives a recurrence relation for constructing the sequences. Thus, one can recover all attaching maps f_i from the ribbon structure on Γ via the following algorithm:

1. Start with a list $\mathcal{L}$ of all half-edges of Γ.
2. Remove one half-edge h from $\mathcal{L}$, and write down a new sequence s^i of half-edges as follows:
   (a) Set $s^i_1 = h$,
   (b) If s^i_k is known, let g be the half-edge that precedes $\overline{s^i_k}$ in the cyclic order at the appropriate vertex. If $g = h$, then the sequence s^i is complete, and it encodes one of the attaching maps f_i, as explained above. Otherwise, g is still in $\mathcal{L}$ (this can be seen by an inductive argument). In this case, remove g from $\mathcal{L}$ and set $s^i_{k+1} = g$.
3. If $\mathcal{L}$ is empty, we are done. Otherwise, return to step 2.

□

Proposition A.1 actually holds for an embedding of a connected graph Γ in any closed oriented surface S, if we further require that S have minimal genus among all closed oriented surfaces that induce the same ribbon structure on Γ. Indeed, the only point of the proof in which we needed $S = S^2$ was in stating that the circles γ_i bound disks disjoint from Γ; in other words, that the complement of an open regular neighbourhood of Γ is a disjoint union of disks. But using the f_i to attach anything other than a union of disks would produce a surface of higher genus. More precisely, if some of the γ_i are used to attach a connected surface of genus ≥ 1 along boundary components, then replacing that surface with a sphere with the same number of boundary components results in a surface of lower genus than S. But attaching a sphere with $k \geq 2$ boundary components also results in higher genus than just attaching k disks (the end result differing by the addition of $k-1$ handles).

Another way of phrasing these observations is to say that a ribbon structure on a connected graph Γ determines at most one embedding of Γ in a minimal (in the above sense) closed orientable surface. It turns out that it *always* determines one such embedding, which is the content of the following result. In particular, the genus of a ribbon graph is well-defined.

Proposition A.2. *If Γ is a ribbon graph and S is obtained from Γ by attaching 2-cells via the maps f_i as in the proof of Proposition A.1, then S is a (not necessarily connected) closed oriented surface.*

Proof. The algorithm for constructing the sequences s^i ensures that each half-edge is used precisely once, so each edge is traversed twice. Hence, the resulting CW-complex is locally euclidean. Since each edge is traversed once in each direction, one can see from cellular homology that $H_2(S)$ is non-trivial, and so the orientations of the 2-cells all match to give a global orientation of S. $\square$

If we drop the connectedness assumption on Γ, we may still embed each of its components in a connected surface, and then take the connected sum of all surfaces. It is however no longer true that the embedding is unique in the sense of Proposition A.1.

The above two propositions allow us to compute the genus of a connected ribbon graph, simply by counting the cyclic sequences s^i from Proposition A.1, each of which will correspond to a 2-cell in S, and then performing an Euler characteristic count. Since the number of possible ribbon structures on a finite graph is finite, it is possible to compute the genus of a graph with no ribbon structure by listing all ribbon structures, and taking the least genus.

For graphs that are not connected, the following result tells us that we can simply compute the genera of all connected components, and then take their sum.

Proposition A.3. *For any graphs* Γ_0, Γ_1 *(with or without ribbon structures), we have*

$$g(\Gamma_0 \amalg \Gamma_1) = g(\Gamma_0) + g(\Gamma_1).$$

Proof. By taking the connected sum of surfaces for Γ_0 and Γ_1, it is clear that

$$g(\Gamma_0 \amalg \Gamma_1) \leq g(\Gamma_0) + g(\Gamma_1).$$

For the other inequality, let $\Gamma_0 \amalg \Gamma_1$ be embedded in a minimal surface S, whose genus we wish to show is at least $g(\Gamma_0) + g(\Gamma_1)$. It is possible to find a 1-dimensional submanifold $Y \subseteq S$ such that cutting S along Y produces two (possibly disconnected) surfaces with boundary S_0, S_1, one containing Γ_0, the other Γ_1 (Y can be obtained by defining a smooth function $S \to \mathbb{R}$ that evaluates to 0 in a neighbourhood of Γ_0, and to 1 in a neighbourhood of Γ_1, and then taking Y to be the pre-image of a regular value in $]0, 1[$). This Y is a disjoint union of circles, each glueing a component of S_0 to one of S_1.

Consider the connected bipartite graph Δ whose vertices are the components of S_0 and S_1, and where the edges between two vertices are the circles in Y connecting them. If we modify S by cutting along one component of Y and capping the two resulting boundary components with a pair of disks, the corresponding change to Δ is the removal of an edge. If this modification leaves S (and thus also Δ) connected, then it decreases its genus by 1, contradicting minimality of S. Hence, the removal of any edge disconnects Δ, in other words, Δ is a tree. Any tree can be inductively constructed by starting with a single vertex, and repeatedly adding a vertex together with an edge connecting it to a pre-existing one. This translates to S being the connected sum of all surfaces obtained by cutting S along Y and capping all boundary circles with disks. If $\hat{S_0}$ is the connected sum of all such components coming from S_0, and $\hat{S_1}$ is defined similarly, then

$$g(S) = g(\hat{S_0}) + g(\hat{S_1}) \geq g(\Gamma_0) + g(\Gamma_1),$$

which proves the result. $\square$

References

[1] Christopher H Cashen. Splitting line patterns in free groups. *Algebraic & Geometric Topology*, 16(2):621–673, 2016.

[2] Benson Farb and Dan Margalit. *A Primer on Mapping Class Groups*. Princeton University Press, 2011.

[3] Cameron Gordon and Henry Wilton. On surface subgroups of doubles of free groups. *Journal of the London Mathematical Society*, 82(1):17–31, 2010.

[4] Ursula Hamenstädt. Geometry of graphs of discs in a handlebody: Surgery and intersection. 2011. arXiv preprint arXiv:1101.1843.

[5] Ursula Hamenstädt. Spotted disk and sphere graphs. 2017. Online preprint at `http://www.math.uni-bonn.de/people/ursula/spotrev.pdf`.

[6] John Hempel. 3-manifolds as viewed from the curve complex. *Topology*, 40(3):631–657, 2001.

[7] Philip J Higgins and Roger C Lyndon. Equivalence of elements under automorphisms of a free group. *Journal of the London Mathematical Society*, 2(2):254–258, 1974.

[8] Jesse Johnson. Notes on heegaard splittings. Online at `users.math.yale.edu/~j327/notes.pdf`.

[9] John R Stallings. Whitehead graphs on handlebodies. *Geometric group theory down under (Canberra, 1996)*, 1999.

[10] Richard Stong. Diskbusting elements of the free group. *Mathematical Research Letters*, 4(2):201–210, 1997.

[11] John HC Whitehead. On certain sets of elements in a free group. *Proceedings of the London Mathematical Society*, 2(1):48–56, 1936.

[12] John HC Whitehead. On equivalent sets of elements in a free group. *Annals of mathematics*, pages 782–800, 1936.

[13] Heiner Zieschang. Simple path systems on full pretzels. *Matematicheskii Sbornik*, 108(2):230–239, 1965.