Combinatorial Graphs and Surfaces from the Computational and Topological Viewpoint

Followed by some notes on

The Isometric Embedding of the square Flat Torus

par

FRANCIS LAZARUS

Mémoire présenté en vue de l'obtention de

l'Habilitation à Diriger des Recherches
de
l'Université de Grenoble

Spécialité Informatique et Mathématiques Appliquées

Soutenue le 16 septembre 2014 devant le jury composé de

Jean-Marc Chassery,	*Directeur de Recherche au CNRS*
Damien Gayet,	*Professeur à l'Université Joseph Fourier*
Michel Pocchiola,	*Professeur à l'Université Pierre et Marie Curie*
Monique Teillaud,	*Directrice de Recherche à l'INRIA*
Uli Wagner,	*Professeur à l'Institute of Science and Technology Austria*

au vu des rapports de

Dan Archdeacon,	*Professeur à l'University of Vermont*
Victor Chepoi,	*Professeur à la Faculté des Sciences de Luminy*
John Sullivan,	*Professeur à la Technische Universität de Berlin*

Remerciements

Je tiens tout d'abord à remercier Victor Chepoi, Dan Archdeacon et John Sullivan d'avoir accepté de rapporter sur ce document malgré leurs emplois du temps chargés. Leur lecture minutieuse, attentive, critique mais bienveillante m'est particulièrement précieuse compte tenu de mon relatif isolement thématique au quotidien. Je suis également très reconnaissant à Jean-Marc Chassery, Damien Gayet, Michel Pocchiola, Monique Teillaud et Uli Wagner d'avoir accepté de participer au jury.

Une bonne partie de mon travail résulte d'une longue collaboration avec Éric Colin de Verdière initiée il y a une quinzaine d'années. Son esprit critique et synthétique, sa droiture intellectuelle ont toujours été pour une moi une référence. Qu'il en soit ici remercié.

J'ai découvert grâce aux immenses qualités pédagogiques de Vincent Borrelli et à sa patience la beauté des mathématiques des plongements isométriques. Le projet Hevea qui, à l'initiative de Vincent, nous a réuni avec Boris Thibert et Saïd Jabrane, et plus tard Damien Rohmer, a été une expérience extraordinaire, une aventure scientifique et humaine hors du commun. Je souhaite à tout scientifique de vivre pareille aventure et je remercie du fond du coeur mes collaborateurs.

Ma gratitude va également à mes autres co-auteurs et à toutes celles ou ceux avec qui j'ai eu le plaisir de travailler et surtout d'apprendre. Je pense en particulier à Sergio Cabello, Jeff Erickson, André Guéziec, William Horn, Julien Rivaud, Gabriel Taubin, Gert Vegter, Erin Wolf Chambers,... mais aussi à mes collègues grenoblois des laboratoires LJK, G-SCOP ou de l'Institut Fourier.

Je suis aussi reconnaissant envers mes nombreux collègues du laboratoire GIPSA et en particulier à celles et ceux dans les services administratifs et techniques qui nous permettent de travailler au quotidien dans les meilleures conditions malgré les demandes de financement variées, les réformes et évaluations incessantes, pas toujours appropriées et souvent chronophages.

Je remercie enfin Pascale et nos enfants Mehdi, Pablo et Ruben pour leur soutien sans faille et leur présence inspiratrice.

Contents

Preface

I have always been fascinated by shapes. For a while I wanted to be a designer, a professional of the form and function of objects. I thought it could be some ideal synthesis of aesthetic and mathematics. I eventually started a PhD at INRIA (Institut National de Recherche en Informatique et Automatique) in computer graphics, studying deformations and metamorphoses of three dimensional shapes. There, I met Arghyro Paouri, a multimedia artist, also working for the communication service at INRIA. This was an exciting period with real interaction between art and technology. I was supplying computer graphic tools for Arghyro and in exchange she would exploit the potential of the tools to produce artistic objects. Thanks to the emerging technology of stereolithography she created a series of sculpture, Metadata, based on the software I had developed and evoking the ancient Greek theme of metamorphosis. After completing my PhD I spend some time as a post-doctoral researcher at the IBM T.J. Watson Center in the New York area. There, I joined the 3D graphics and interactions team working on algorithms for the compression of three dimensional models over the Internet. Little by little, I started to realize that Computer Graphics might not give me the satisfaction I was expecting. Although Computer Graphics was more and more utilised by artists, I was not an artist myself and I missed the creativity and the questioning proper to artistic or scientific activities. Back to France, I thus started to concentrate on more theoretical aspects of computational topology. Topology is obviously a fundamental approach to shapes and the computational part allows to anchor the research in a concrete perspective.

The first part of this document is an attempt for an elementary combinatorial theory of surfaces. The subject is quite old and standard but I tried to give it a modern treatment as well as a global shape and unity not excluding some aesthetic criteria. All the results presented in this part (except for the last chapter on the homotopy test) are proved in the simple framework of combinatorial surfaces without ever resorting to topology. I nevertheless gave topological intuitions in order to avoid the aridity of purely combinatorial arguments. This first part is more akin to a survey as none of the presented results is formally new. However, I claim some originality in the exposition. In particular, the document contains

- a formal correspondence between the two main encodings of non-orientable surfaces: signed rotation systems and triplets of involutions,
- a full description of combinatorial surfaces with boundaries,
- a formal description of surgery (cutting and stitching) of combinatorial surfaces,

- a simple proof that a contractible cycle cuts a disc in a combinatorial surface (this includes the Jordan curve theorem),
- a simple computation of the homology class of a cycle on a combinatorial surface,
- a combinatorial framework for curves and graphs immersions and embeddings,
- a careful construction of canonical systems of loops.

The presentation is relatively formal in the spirit of a reference document; I also hope it will be useful for the implementation of algorithms related to the computational topology of surfaces. Indeed, most of the abundant literature in the domain has been published without implementation. One reason could be the apparent hiatus between the combinatorial aspect of the data-structures and the non-discrete nature of topology. Still, many of the published algorithms could be implemented with a reasonable effort and could find practical applications if some implemented library were available along the line of the existing CGAL C++ library for computational geometry.

The second part of the document is the result of a beautiful scientific and human adventure. The space devoted to this part in the document is in no way proportional to its importance. However, a 90 page paper recently appeared in the Brazilian journal *Ensaios Matemáticos* describing this project in great details. I thus chose to explain the main ideas in a rather informal manner, referring the reader to the long paper for the details. Together with Vincent Borrelli, Saïd Jabrane and Boris Thibert, we have succeeded to compute the first isometric embedding of a flat torus and we have discovered a new geometric structure: the C^1 fractals. Our work is based on an amazing result of John Nash later generalised by Misha Gromov; any n-dimensional Riemannian manifold that embeds into $\mathbb{R}^k$ with $k \geq n+1$ can be isometrically embedded in the Euclidean space of the same dimension k in a smooth (C^1) manner. It follows that any Riemannian orientable surface can be embedded isometrically into the three dimensional Euclidean space. In particular, a flat torus (a locally Euclidean torus whose Gaussian curvature thus vanishes everywhere) has such an embedding. An easy argument using Gaussian curvature shows that this is actually impossible if the embedding is required to be C^2. Nash's exploit was to prove that this argument could be bypassed with C^1 embeddings. Our work had quite a success; It was on the cover of the Proceedings of the National Academy of Science (PNAS), the cover of Pour La Science (the French edition of the Scientific American) and was selected among the ten most beautiful discovery of the year 2012 by the French journal La Recherche. It was also covered by many medias all over the world. Thanks to Damien Rohmer who joined our team in 2012, we were able to produce splendid pictures of the flat torus embedding (some of them were exposed in the subway in Paris). I could not hope for a better incarnation of the interaction between Science and Aesthetics.

I must admit that the present document may appear unusual for a mémoire d'habilitation. Rather than describing my actual few results, the first and main part is more akin to a thesis where I essentially defend the opinion that a combinatorial theory of surfaces can be effective when dealing with topological properties on surfaces. As explained above, this deliberate choice imposed upon me as a necessary and hopefully useful task to transform the abundant literature on the subject into real implementations on computers (that are still to be done). I have nonetheless included an appendix where I give a quick survey of my published works, emphasizing my thematic evolution.

One final remark; following the third edition of Ian Stewart's Galois theory, I shall not use punctuation when displaying formulas.

Part I

Combinatorial Graphs and Surfaces

Introduction

Writing on topological algorithms for graphs on surfaces was not an easy task, especially after my colleague and long time collaborator Éric Colin de Verdière defended his own *habilitation* on the subject [CdV12]. Many of my results in the domain were obtained with Éric and the risk was to produce at best a redundant document. While all the results obtained in the domain of Computational Topology are undoubtedly of a discrete nature, after all they can *theoretically* be implemented on a computer, the proofs and arguments are most often appealing to results from pointset topology. This is for example true concerning the use of the Jordan-Schoenflies theorem or the introduction of the cross-metric surface model proposed by Éric and Jeff Erickson [CE10]. It appeared to me that a self-contained combinatorial presentation was missing not only for aesthetic reasons but also because the proofs are generally much simpler and because the combinatorial viewpoint is necessary when it comes to implementation. Other attempts of such a purely combinatorial approach already exists (see e.g. [GT87] or [BL95]), each with a slightly different objective so that I still believe that there is room for a treatment as in the present document. I tried to write the elements of a self-contained combinatorial theory of surfaces, always keeping the topological intuition in mind.

Adding topological properties to combinatorial objects is not new. Indeed, the systematic study of topological invariants of combinatorial objects was initiated at the end of the nineteenth century with the birth of algebraic topology. The main contributor to this new area, at the time part of *Analysis Situs*, was Henri Poincaré (1854–1912) and according to Agoston [Ago76]: “Poincaré’s 1899 paper quite definitely can be said to have founded combinatorial topology. In this and the second supplement Poincaré also described a method for computing the Betti numbers and torsion coefficients in terms of "incidence matrices" wich are naturally associated to complexes.” On the other hand, the computational aspects of combinatorial topology is the subject of *Computational Topology*. This branch of Computer Science studies effective computations of topological problems and their complexity, although their *computability* is often regarded as part of mathematics. Gert Vegter, thus starts his survey [Veg97] on Computational Topology in the Handbook of Discrete and Computational Geometry: “Topology studies point sets and their invariants under continuous transformations, like the number of connected components ... Computational Topology deals with the complexity of such problems, and with the design of efficient algorithms ... These algorithms can only deal with spaces and maps that have finite representation.”

The computational topologist is thus interested in combinatorial structures likely to represent topological spaces. This ability to represent topologies is usually recorded in functors from the various combinatorial categories (like simplicial complexes, combinatorial surfaces,...) towards the categories of topological spaces. However, while the categories are well established on the topological side, it is not so clear for combinatorial categories. Indeed, many combinatorial structures are used in practice to represent essentially the same things: simplicial complexes, simplicial sets, polyhedral complexes, triangulations, etc. This diversity could be related to the negative answer to the *Hauptvermutung* problem asking whether homeomorphic polyhedra (of simplicial complexes) are combinatorially equivalent [Ran96]. Even though the Hauptvermutung is true in dimension two and for three dimensional manifolds, there does not seem to be a fixed combinatorial definition of two-dimensional spaces. For instance, Zieschang et al. [ZVC80] have a definition of combinatorial 2-complexes that is generally more concise than simplicial complexes:

Definition 0.0.1 ([ZVC80, p. 37])**.** A *2-dimensional complex or 2-complex* is a system of vertices, edges and (oriented) faces with the following properties:

(a) The vertices and edges constitute a graph C.

(b) For each face ϕ there is a closed path ω in C. The set of paths ω' which result from ω by cyclic interchange is called the class of *positive boundary paths* of ϕ. The path ω^{-1} bounds ϕ negatively. The positive boundary of ϕ is often denoted by $\partial\phi$.

(c) For each face ϕ there is an oppositely oriented face ϕ^{-1}, the positive boundary path of which are the negative boundary path of ϕ. A pair $\{\phi, \phi^{-1}\}$ is called a *geometric face*.

The corresponding definition of morphism (or mapping), the combinatorial counterpart of a continuous map, remains relatively complicated:

Definition 0.0.2 ([ZVC80, p. 45])**.** By a *mapping of a 2-complex F' into a 2-complex F* we mean a correspondance $f : F' \to F$ which sends faces of F' to faces, edges or vertices of F; edges of F' to edges or vertices of F; and vertices of F' finally to vertices of F, in such a way that boundary relations are preserved. More precisely, this means:

(a) The boundary objects of an element are carried into the image element or its boundary objects. Inverse elements are carried to inverse elements or else to the same vertex.

(b) If the face ϕ' is mapped to ϕ, then after removal of spurs the image $\partial\phi'$ is a positive power of $\partial\phi$.

(c) If the image of ϕ' is the edge σ, then boundary edges and vertices of ϕ' are associated with the edges σ and σ^{-1} or their endpoints. As a result $f(\partial\phi')$ will be a closed path.

(d) If a face or an edge is mapped to a vertex, so are all its boundary objects.

This is not the only possible definition and some years later, Collins and Zieschang simplify their definition:

Definition 0.0.3 ([vCGKZ98, p. 17])**.** A *mapping* or *homomorphism* $f : C \to D$ between two 2-complexes C, D assigns to each vertex of C a vertex of D, to each edge of C an edge or vertex of D, and to each face of C a face of path of D preserving the boundary behavior:

- if $v \in C$ is the initial vertex of the oriented edge $\sigma \in C$ then $f(v)$ is the initial vertex of $f(\sigma)$ if this is an edge, and otherwise $f(\sigma) = f(v)$;
- if $\prod_{i=1}^{k} \sigma_i$ is the boundary of the disc $\psi \in C$ then $\prod_{i=1}^{k} f(\sigma_i)$ bounds $f(\psi)$ if it is a face; otherwise $f(\psi)$ must be a path nullhomotopic in D^1;

- $f(\sigma^{-1}) = (f(\sigma))^{-1}$ and $f(\psi^{-1}) = (f(\psi))^{-1}$ for every edge σ and face ψ of C.

One may compare point (b) in the first definition with the second item in the second definition. In fact, the two definitions are not equivalent; a morphism according to the first definition may require a subdivision of complexes to provide a morphism according to the second definition. This would be the case of a morphism from a sphere subdivided with one vertex, one edge and two faces onto a projective plane, composed of a single vertex, edge and face.

As far as graphs and surfaces are concerned, the debate seems closed. There is essentially one admitted combinatorial structure for representing graphs and the same is true for surfaces. This document describes these structure in details, keeping the implementation aspect in mind. The content is well-known and appears in various texts. The books of Gross and Tucker [GT87], Mohar and Thomassen [MT01] and parts of the book edited by Beineke and Wilson [BW09] are the most adequate references. Those books concentrate on combinatorial and algorithmic problems related to combinatorial graphs and surfaces, which is obviously the most interesting part. However, as claimed above, I think that a methodological study of the data-structures should find its place on the shelf. This document is primarily for those wishing to implement these data-structures. As commonly assumed in computational topology, we shall analyze the complexity of our algorithms with the uniform cost RAM model of computation [AHU74]. A notable feature of this model is the ability to manipulate arbitrary integers in constant time per operation and to access an arbitrary memory register in constant time.

For completeness, I included in Appendix B a short introduction to the few properties of combinatorial group theory used in this document.

Chapter 1

Combinatorial Graphs

Contents

1.1 What Is a Graph?

Graphs are among the most ubiquitous objects in Computer Science. Still, there might be as many formal definitions of a graph as there are books on the subject. This is even the case in the more formalized subfield of algebraic graph theory. For instance, Biggs starts his book on algebraic graph theory [Big94] with

> **Basic definitions and notations**
>
> Formally, a *general graph* Γ consists of three things: a set $V\Gamma$, a set $E\Gamma$ and an incidence relation, that is, a subset of $V\Gamma \times E\Gamma$. An element of $V\Gamma$ is called a *vertex*, an element of $E\Gamma$ is called an *edge*, and the incidence relation is required to be such that an edge is incident with either one vertex (in which case it is a *loop*) or two vertices.

While Godsil and Royle [GR01] begin with

1.1 Graphs

A *graph* X consists of a *vertex* set $V(X)$ and an edge set $E(X)$, where an edge is an unordered pair of distinct vertices of X.

We advocate the following universal definition (see e.g. [Ser77, Sec. 2.1])

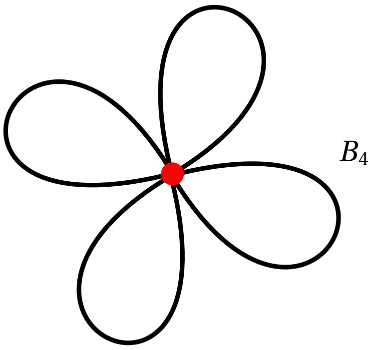
Definition 1.1.1. A **graph** is a quadruple $G=(V,A,o,\iota)$ where

- V is a set whose elements are called **vertices**,
- A is a set whose elements are called (oriented) **arcs**,
- $o: A \to V$ is a map that sends each arc a to its **origin** vertex $o(a)$.
- $\iota: A \to A$ is a fixed point free involution that sends each arc to is **inverse** arc. We usually write a^{-1} for $\iota(a)$.

The origin is also called the **tail** of an arc and the inverse is called the **opposite**.

A (non-oriented) **edge** is a pair $\{a, a^{-1}\}$. The origin of a^{-1} is the **destination**, or **head**, of a. The tail and head of an edge are its two **endpoints** to which the edge is **incident**. Because ι has no fixed point the set of arcs is the disjoint union $A = A_+ \cup \iota(A_+)$ for some $A_+ \subset A$. The set of edges is thus in bijection with A_+. Fixing A_+ defines a **default orientation** of the edges. *We will assume this default orientation given once for all for each graph in this document.*

Example 1.1.2. A graph with a single vertex is called a **bouquet of circles**. The bouquet of circles with n edges is denoted by B_n.



Following Serre [Ser77] "there is an evident notion of morphisms for graphs". Serre defines a morphism as two maps, one between the vertex sets and one between the arc sets, that "commute" with the origin and inverse maps. For this definition, a non-loop edge contraction would not be a morphism. We thus find more convenient the following slightly modified definition.

Definition 1.1.3. A **morphism** from a graph (V,A,o,ι) to a graph (V',A',o',ι') is a map $f: V \cup A \to V' \cup A'$ such that $f(V) \subset V'$ and f commutes with the origin and inverse maps, i.e., $f \circ o = o' \circ f$ and $f \circ \iota = \iota' \circ f$, where by convention the origin and inverse maps are the identity on the vertex sets.

We will often denote by $V(G)$ and $E(G)$ the respective sets of vertices and edges of a graph G. Note that the number $|E(G)|$ of edges is half the number of arcs of G.

1.1.1 Basic operations on graphs

Let $G = (V, A, o, \iota)$ be a graph, and let $e = \{a, a^{-1}\}$ be an edge of G.

Definition 1.1.4. The **contraction** of e in G transforms G to the graph $G/e = (V', A', o', \iota')$ where $V' = V/\{o(a) = o(a^{-1})\}$, $A' = A \setminus e$ and o' and ι' are the obvious restrictions of o and ι with the identification of $o(a)$ and $o(a^{-1})$. If e has a degree one endpoint, the contraction is called an **elementary retraction**.

It is an easy exercise to check that the edge contraction is a graph morphism.

Definition 1.1.5. The **deletion** of the edge e of G transforms G to the graph $G - e = (V, A', o', \iota')$ where $A' = A \setminus e$ and o' and ι' are the obvious restrictions of o and ι. Similarly, we define the deletion of a vertex v as the graph $G - v$ with v and all the incident edges removed.

Definition 1.1.6. The **elementary subdivision** of e in G transforms G to the graph $S_e G = (V', A', o', \iota')$ where $V' = V \cup \{x\}$, $A' = A \cup \{a', a'^{-1}\}$ for some new elements x, a', a'^{-1} not in $V \cup A$. The maps o' and ι' are defined in the obvious way. In particular, $S_e G/\{a', a'^{-1}\} = G$. A **subdivision** of G is the result of a finite sequence of elementary subdivisions.

Two graphs are **combinatorially equivalent** if they have isomorphic subdivisions. Intuitively, two graphs are combinatorially equivalent if they have homeomorphic realizations (to be defined). An **invariant** for graphs is a property, usually a functor, that is invariant under combinatorial equivalence. This will be the case for the fundamental group, the homology or cohomology groups as defined in the next sections.

Exercise 1.1.7. Prove that the Euler characteristic of finite graphs, i.e., the number of vertices minus the number of edges, is an invariant. What could be the adequate functor?

As far as topological properties are concerned the most fundamental properties of a graph are recorded in the next two lemmas.

Lemma 1.1.8. *Any subtree of a graph G can be extended to a maximal subtree under inclusion.*

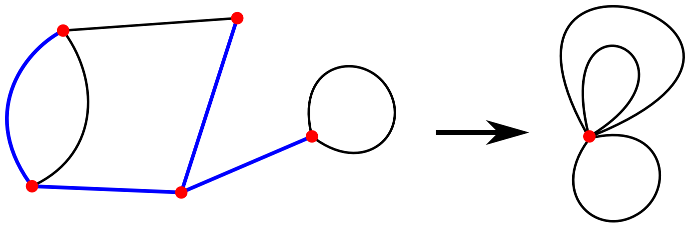
PROOF. This is clear if G is finite as it contains a finite number of subtrees. Otherwise, we can apply Zorn's lemma to the set of subtrees extending the given subtree. This set is indeed inductive since the union of an increasing sequence of subtrees is itself a subtree, being connected and acyclic. A more "constructive" proof makes use of the axiom of choice to build a shortest path tree inductively. □

Lemma 1.1.9. *If G is a connected graph, every maximal subtree is a spanning tree (and vice versa).*

If T is a spanning tree of the connected graph G, any edge of G that is not in T is a **chord**. Note that a chord may or may not be a loop edge.

Example 1.1.10. If T is a subtree of a G, then there is a morphism $c_T : G \to G/T$ contracting the edges of T. If T is finite, c_T is a composition of edge contractions. Otherwise, G/T is a direct limit (in the sense of categories) of such contractions.

Example 1.1.11. When T is a spanning tree, G/T is a bouquet of circles whose edges correspond to the chords of T in G.



1.2 Paths and Homotopy

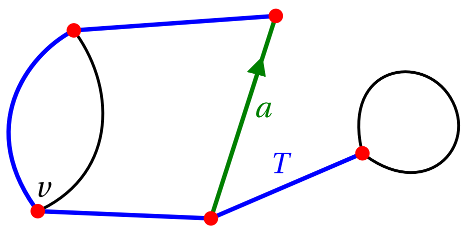
Definition 1.2.1. A **path** in G is a finite alternating sequence of vertices and arcs of G of the form $(v_0, a_1, v_1, a_2, \dots, a_k, v_k)$ such that the tail and head of a_i are respectively v_{i-1} and v_i. The integer k is the **length** (or size) of the path. When $k = 0$, the path is said to be **constant**. It is denoted by 1 when the vertex v_0 is implicit. The **inverse path** of $(v_0, a_1, v_1, a_2, \dots, a_k, v_k)$ is the path $(v_k, a_k^{-1}, v_{k-1}, a_{k-1}^{-1}, \dots, a_1^{-1}, v_0)$. A path is **simple** if is has no repeated vertices. It is **closed** if its first and last vertex coincide. For non-constant paths, the vertices are redundant and we usually write $(a_1, a_2, \dots, a_k)$ for $(v_0, a_1, v_1, a_2, \dots, a_k, v_k)$. A path is also called a **walk** and a closed path is also called a **loop**. The first vertex of a loop is its **basepoint**. We reserve the term **circuit** to a closed path without fixing its basepoint. Formally, a circuit is a class of closed paths related by circular permutations of their arcs. A circuit is **simple** if all its paths are simple.

Example 1.2.2. If T is a spanning tree of the connected graph G, we denote by $T[v, w]$ the unique shortest path in T from v to w. Any arc a of G determines a loop with basepoint v:

$$T[v, a] := T[v, o(a)] \cdot a \cdot T[o(a^{-1}), v]$$

and a circuit

$$T[a] := a \cdot T[o(a^{-1}), o(a)]$$



Here, $T[v, a] = T[a]$.

Definition 1.2.3. A **spur** is a subsequence of the form (a, a^{-1}) in a path. Adding or removing a spur in a path is called an **elementary homotopy**. A **free elementary homotopy** is an elementary homotopy applied to any of the path representatives of a circuit. The **homotopy** relation is the transitive closure of elementary homotopies. Likewise, **free homotopy** is the transitive closure of free elementary homotopies. We write $\gamma \sim \lambda$ if γ and λ are homotopic paths and $\gamma \overset{free}{\sim} \lambda$ when they are freely homotopic circuits. A **reduced path** or **cyclically reduced circuit** is a path or circuit without spur. A path or circuit (freely) homotopic to a constant path is said to be **contractible**. If the last vertex of a path γ coincides with the first vertex of a path λ, their **concatenation** is the path $\gamma \cdot \lambda$ whose arc sequence is the the arc sequences of γ followed by the arc sequence of λ.

Proposition 1.2.4. *Let v be a vertex of G. The set of homotopy classes of loops with basepoint v is a group for the law of path concatenation. It is called the* ***fundamental group*** *of G based at v and denoted by $\pi_1(G, v)$. The free homotopy classes are the conjugacy classes in this group.*

Exercise 1.2.5. Let γ be a path from a vertex v to a vertex w in G. Prove that the map $\lambda \mapsto \gamma \cdot \lambda \cdot \gamma^{-1}$ taking a loop with basepoint w to a loop with basepoint v induces an isomorphism from $\pi_1(G, w)$ to $\pi_1(G, v)$. What is this isomorphism when γ is a loop?

When G is connected we can thus speak of its fundamental group up to isomorphism without referring to its basepoint.

Proposition 1.2.6. *The fundamental group of a bouquet of circles is a free group over its edges.*

Proof. Let B be bouquet of circles with arc set A. Recall that we can write $A = A_+ \cup \iota(A_+)$, so that the set of edges can be identified with A_+. We denote by $F(A_+)$ the free group generated by A_+. Since B has a single vertex, each arc a is a loop (a). Obviously, the set of loops $\{(a)\}_{a\in A_+}$ generates $\pi_1(B, \bullet)$. The map $a \mapsto (a)$ extends uniquely, by the universal property of free groups, to a group morphism $F(A_+) \to \pi_1(B, \bullet)$ that is onto by the preceding remark. Since elementary homotopies correspond to free elementary reductions of words (of the type $u a a^{-1} v \mapsto u v$), the kernel of this morphism is trivial and $\pi_1(B, \bullet) \simeq F(A_+)$. □

Remark 1.2.7. To keep the notations light, we will often identify a loop with its homotopy class.

A graph morphism $f : G \to G'$ sends a path, loop or circuit $(a_1, a_2, \ldots, a_k)$ of G to a path, loop or circuit $(f(a_1), f(a_2), \ldots, f(a_k))$ of G' (formally, one should remove the $f(a_i)$ that are vertices). Homotopic paths and freely homotopic circuits are sent to homotopic paths and freely homotopic circuits respectively. It follows that

Lemma 1.2.8. *A graph morphism $f : (G, v) \to (G', f(v))$ induces in a* natural way *a group morphism $f_* : \pi_1(G, v) \to \pi_1(G', f(v))$, i.e., if $G \xrightarrow{f} G' \xrightarrow{g} G''$ are two morphisms, then $(fg)_* = f_* g_*$.*

Exercise 1.2.9. Prove that an edge contraction of a connected graph induces an isomorphism of fundamental groups if and only if its endpoints are distinct.

Theorem 1.2.10. *Let T be a spanning tree of a connected graph G. For any vertex v of G, $\pi_1(G,v)$ is isomorphic to the free group on the set of chords of T in G.*

PROOF. We give two proofs of this basic fact. We again write $A = A_+ \cup \iota(A_+)$ for the set of arcs of G and we will freely identify a subset of edges with a subset of A_+ when convenient. We denote by C the set of chords of T in G.

Combinatorial group flavoured proof We observe that any loop $(a_1, a_2, \dots, a_k)$ is homotopic to the concatenation of loops $T[v,a_1] \cdot T[v,a_2] \cdots T[v,a_k]$. Since $T[v,a]$ is contractible whenever a is in T, we see that the family $\Gamma = \{T[v,a]\}_{a\in C}$ generates $\pi_1(G,v)$. Each arc of C appears exactly once in one loop of this family. It follows that Γ only satisfies trivial relations (of the type $T[v,a]\cdot T[v,a]^{-1} = 1$) and is thus a free generating set. Said differently, the map $a \mapsto T[v,a]$ extends uniquely to a morphism $F(A_+) \to \pi_1(G,v)$ whose kernel is the subgroup spanned by the edges of T. We conclude that $\pi_1(G,v) \simeq F(A_+)/F(T) \simeq F(C)$.

Homomorphic flavoured proof Following Example 1.1.10, we have a graph morphism $G \to G/T$ to a bouquet of circles with edge set C. When T is finite this morphism is a finite product of edge contractions and thus induces an isomorphism of fundamental groups (see Exercise 1.2.9). We then conclude with Proposition 1.2.6. When T is infinite, we can view G/T as a direct limit as noted in Example 1.1.11. By the functoriality of π_1, and making the functor and direct limit commute, the fundamental group of G/T is the direct limit of fundamental groups of graphs obtained by finite sequences of edge contractions. These fundamental groups are all isomorphic by the finite case and their direct limit is thus an isomorphic group. In particular, $\pi_1(G,v) \simeq \pi_1(G/T,\bullet)$. □

Remark 1.2.11. From the first proof, we note that a basis of $\pi_1(G,v)$ is given by the loops $T[v,a]$ when a runs through the chords of T in G. The expression in this basis of the homotopy class of a loop ℓ is obtained as follows. We first take the trace of ℓ over C, i.e., we discard the arcs of T in ℓ. We then freely reduce the resulting word on $C \cup \iota(C)$, and finally replace each occurrence of c (resp. c^{-1}) by $T[v,c]$ (resp. $T[v,c]^{-1}$).

Corollary 1.2.12. *If G is a finite connected graph, its fundamental group is a free group of rank*

$$1-\chi(G) = 1-|V(G)|+|E(G)|$$

PROOF. From the preceding theorem $r := \mathrm{rank}\pi_1(G,v)$ is the number of chords of a spanning tree T, so that $r = |E(G)|-|E(T)|$. But T being a tree we have $|E(T)| = |V(T)|-1$ and T being spanning we have $|V(T)| = |V(G)|$. Whence $r = |E(G)|-(|V(G)|-1)$. □

Exercise 1.2.13. Let H a connected subgraph of a connected graph G and let v be vertex of H. Prove that the inclusion $H \hookrightarrow G$ induces a monomorphism $\pi_1(H,v) \hookrightarrow \pi_1(G,v)$. (Hint. You may use Lemma 1.1.8.)

1.3 Some Elementary Algorithms Related to Homotopy

Here, we examine how to compute a basis of the fundamental group of a graph in practice, and how to decide whether a loop is contractible or whether two loops are homotopic. We assume given a finite connected graph $G = (V, A, o, \iota)$ with a default orientation A_+ and a vertex $v \in V$.

1.3.1 Computing a basis of $\pi_1(G, v)$

By a **basis** we mean a minimal size set of loops whose homotopy classes generate $\pi_1(G, v)$. Note that a reduced loop (without spur) has minimal length among all its homotopic loops.

Lemma 1.3.1. *We can compute a basis of $\pi_1(G, v)$ in time $O(|A_+| + r|V|)$ where $r = 1 - |V| + |A_+|$.*

PROOF. We already know from Corollary 1.2.12 that any basis has r elements. By the remark following Theorem 1.2.10, such a basis is provided by the loops $\{T[v, a]\}$, for a a chord of a spanning tree T. The spanning tree can be computed using a graph traversal such as depth-first search or breadth-first search in $O(|A_+|)$ time. Each of the r loops $\{T[v, a]\}$ can be written down in time proportional to its length $O(|V|)$. $\square$

It should be noted that a basis of $\pi_1(G, v)$ does not necessarily arise from the chords of a spanning tree. However, a shortest basis – that is a basis minimizing the total length of its loops – indeed arises this way. To see this we state a preliminary lemma.

Lemma 1.3.2. *Let F be a free group over $(x_1, x_2, \ldots, x_n)$. For any base $(u_1, u_2, \ldots, u_n)$ of F there exists a permutation σ of $[1, n]$ such that each x_i appears in the reduced expression of $u_{\sigma(i)}$ in terms of the x_j.*

PROOF. The automorphism of L defined by $x_i \mapsto u_i$, $i \in [1, n]$, quotients to an automorphism f of its abelianized group $L/[L, L]$ which is a free abelian group of rank n. The map f can thus be seen as an automorphism of $\mathbb{Z}$-module whose matrix (c_{ij}) with respect to the basis formed by the cosets of the x_i – so that c_{ij} is the cumulated exponents of x_i in u_j – has a non-zero determinant. It follows that at least one term $\prod_{i \in [1,n]} c_{i\sigma(i)}$ of the usual Leibnitz expansion of the determinant must be non-zero. This implies the lemma. $\square$

Proposition 1.3.3. *The basis of $\pi_1(G, v)$ associated to a breadth-first-search tree from v is a shortest basis.*

PROOF. Let T be a breadth-first-search tree from v. In particular, for any arc a, $T[v, a]$ is a shortest loop with base v through a in G. We denote by $(c_1, c_2, \ldots, c_r)$ the chords of T in G. Let B be another basis. According to the previous lemma, the elements of B can be ordered in a such a way that its ith element b_i contains $T[v, c_i]$ in its reduced expression in terms of the $T[v, c_j]$. It follows that b_i goes through c_i, hence is longer than $T[v, c_i]$. $\square$

Exercise 1.3.4. Show, using graph distances but no algebra, that among all the bases associated to spanning trees the shortest ones correspond to breadth-first-search trees.

Remark 1.3.5. If the edges of G are positively weighted, a shortest basis is a basis whose total weight is minimal. Then, Proposition 1.3.3 still holds if we replace the breadth-first-search tree by a shortest path tree.

Open question: Can we characterize which source vertices v in G lead to the shortest of the shortest bases?

1.3.2 Homotopy test

Proposition 1.3.6. *After $O(|A|)$ time preprocessing we can test whether any two loops ℓ and ℓ' are homotopic in $O(|\ell|+|\ell'|)$ time.*

PROOF. Compute a spanning tree T of G in $O(|A|)$ time. Now, ℓ and ℓ' are homotopic if and only if $\ell \cdot \ell'^{-1}$ is contractible. From the remark following Theorem 1.2.10, this can be easily checked in time proportional to the length of the loop $\ell \cdot \ell'^{-1}$. □

Every circuit $c = (a_1, a_2, \ldots, a_k)$ is freely homotopic to the loop $\ell(c) := T[v, a_1] \cdot T[v, a_2] \cdots T[v, a_k]$. It ensues that two circuits c, c' are freely homotopic if and only if $\ell(c)$ and $\ell(c')$ are conjugates in $\pi_1(G, v)$. On the other hand, the conjugacy class of any loop in a free group contains a unique cyclically reduced representative that can be computed in time proportional to the length of the loop. It thus remains to compare the cyclically reduced representatives, say ℓ and ℓ', of $\ell(c)$ and $\ell(c')$. These representatives are cyclic permutation of each other if and only if they have the same length and ℓ' is a subword of $\ell \cdot \ell$. This last condition can be tested in $O(|\ell| + (|\ell'|)$ time using the Knuth-Morris-Pratt algorithm [CLRS09]. Summarizing, we have

Proposition 1.3.7. *After $O(|A|)$ time preprocessing we can test whether any two circuits ℓ and ℓ' are freely homotopic in $O(|\ell|+|\ell'|)$ time.*

1.4 Homology

We now define the cycle group of a graph $G = (V, A, o, \iota)$ with $A = A_+ \cup \iota(A_+)$. The homology of graphs appears in a 1847 paper by Kirchhoff [BLW98, p. 133] concerning electric networks. Those are modeled as graphs whose edges represent electrical connections each having a resistance r_j and a voltage source E_j. Kirchhoff's voltage law states that the directed sum of the electrical potential differences around a cycle must be zero. Applied to a cycle of a graph, this leads to equations of the form

$$\sum_j r_j I_j = \sum_j E_j$$

where j runs through the arcs of the cycle, the arc j being traversed by the current I_j. If the resistances and sources are known, Kirchhoff explains how to find the minimum number of equations as above, hence the minimum number of cycles, necessary to determine the currents (assuming the Kirchhoff's current law). The answer is given by the *cyclomatic number* of the graph, which is also the rank of its *cycle space*.

Let $C_0(G)$ and $C_1(G)$ be the free abelian groups with basis V and A_+ respectively. The elements of $C_i(G)$, $i = 0, 1$ are called **i-chains**. The **support** of a chain is the set of vertices or arcs with nonzero coefficients; its elements are **contained** in the chain. We also consider the **boundary operator** $\partial : C_1(G) \to C_0(G)$ defined by $\partial a = o(a^{-1}) - o(a)$. The **homology group of dimension zero** is the quotient

$$H_0(G) = C_0(G)/\operatorname{Im}\partial$$

We simply write C_i for $C_i(G)$ when there is no ambiguity on the graph G.

Proposition 1.4.1. *$H_0(G)$ is isomorphic to the free abelian group over the set of connected components of G.*

PROOF. Let K be the set of connected components of G. Consider the *augmentation map* $\varepsilon : C_0 \to \oplus_K \mathbb{Z}, c \mapsto \sum_{\kappa \in K} \alpha_\kappa \kappa$ where α_κ is the sum of the coefficients in c of the vertices belonging to κ. We claim that $\ker\varepsilon = \operatorname{Im}\partial$. Indeed, for any arc a we obviously have $\varepsilon(\partial a) = 0$, whence $\operatorname{Im}\partial \subset \ker\varepsilon$. On the other hand, if $c = \sum_i \alpha_i v_i \in \ker\varepsilon$ has all its vertices v_i in a single component, we can join some fixed vertex in this component to each v_i with a path γ_i and we easily check that $c = \partial(\sum_i \alpha_i \gamma_i)$. It follows that $c \in \operatorname{Im}\partial$ thus proving the claim. We know conclude by the surjectivity of the augmentation map that

$$\oplus_K \mathbb{Z} \simeq C_0/\ker\varepsilon = C_0/\operatorname{Im}\partial$$

□

We put $Z_1 := \ker\partial$ and call its elements **cycles**. Z_1 is the **cycle group** of G. This group is also called the **first homology group** and denoted by $H_1(G)$. As for homotopy, a graph morphism $f : G \to H$ induces group morphisms between homology groups. Indeed, the map $\sum_i n_a a \mapsto \sum_i n_a f(a)$ commutes with boundary operator.

A **simple cycle** is the sum of the signed arcs of a simple circuit in G, where an arc of A_+ has a positive sign and a negative sign otherwise. Any path or circuit of G can be considered as a chain or cycle by considering the signed sum of its arcs.

Lemma 1.4.2. *Every cycle is a combination of simple cycles.*

PROOF. Let $c = \sum_{a \in A_+} \alpha_a a$ be a cycle of G. The proof is by induction on the size of the support of c. Let H be the subgraph induced by the support edges of c. A vertex of H cannot have degree one. Otherwise, such a vertex would be contained in the support of ∂c, contradicting $\partial c = 0$. But H being a finite graph, it must contain a simple circuit γ. Let α be the coefficient in c of some chosen arc in γ. We conclude by applying the induction hypothesis to the cycle $c - \alpha\gamma$. □

Corollary 1.4.3. *A tree is* ***acyclic****, i.e., its cycle space is trivial.*

PROOF. From the previous lemma, since a tree has no simple cycle. □

Proposition 1.4.4. *Suppose G connected and let T be a spanning tree of G. Then, $H_1(G)$ is isomorphic to the free abelian group generated by the chords of T in G.*

PROOF. Let C be the set of chords and let $B = \{T[a]\}_{a\in C}$ be the set of simple cycles associated to the chords. The cycles in B are independent as each chord appears in the support of exactly one of them. It remains to prove that B is generating. Let $c = \sum_{a\in T} \alpha_a a + \sum_{e\in C} \beta_e e$ be a cycle in G. Then, $c - \sum_{e\in C} \beta_e T[e]$ is a cycle whose support lies in T. It must be null by the above corollary, i.e., $c = \sum_{e\in C} \beta_e T[e]$. □

The rank of $H_1(G)$ is called the **cyclomatic number** or **first Betti number** and denoted by $\beta_1(G)$. When G is connected and has a finite number of chords, we observe that $\beta_1(G) = 1 - \chi(G)$ is also the rank of the fundamental group. In fact

Proposition 1.4.5. *$H_1(G)$ is isomorphic to the abelianization of the fundamental group of G.*

PROOF. Denote by $\mathscr{L}(G, v)$ the set of loops of G with basepoint v. The map $\mathscr{L}(G, v) \to H_1(G)$ defined by $(a_1, a_2, \ldots, a_k) \mapsto \sum_{i=1}^k a_i$ is compatible with elementary homotopies. It thus defines a morphism $\varphi : \pi_1(G, v) \to H_1(G)$. We just saw that any cycle can be written as a combination of cycles of the form $T[a]$. Noting that $\varphi(T[v, a]) = T[a]$ in $H_1(G)$, it follows that φ is onto. Let $\gamma = T[v, a_1] \cdot T[v, a_2] \cdots T[v, a_k]$ be any element of $\pi_1(G, v)$, written over the basis $\{T[v, a]\}_{a\in C}$. Then $\varphi(\gamma) = \sum_{a\in C} n_a T[a]$ where n_a is the cumulated exponent of $T[v, a]$ in γ. Hence, $\gamma \in \ker\varphi$ if and only if all the n_a cancels. This is exactly saying that γ belongs to the **derived subgroup** $[\pi_1(G, v), \pi_1(G, v)]$ of $\pi_1(G, v)$. We thus have

$$H_1(G) \simeq \pi_1(G, v)/\ker\varphi = \pi_1(G, v)/[\pi_1(G, v), \pi_1(G, v)]$$

□

Exercise 1.4.6. Show that the one dimensional homology of G is the direct sum of the one dimensional homology of its 2-connected components (the blocks of G). (Hint: consider the map sending a cycle to its traces over the 2-connected components of the graph.)

The homology functor Let $f : G \to G'$ be a graph morphism. f induces a **chain morphism** $f_\# : C_i(G) \to C_i(G')$ by setting for $v \in V(G)$ and $a \in A(G)$:

$$f_\#(v) = f(v) \quad \text{and} \quad f_\#(a) = \begin{cases} 0 & \text{if } f(a) \in V(G') \\ f(a) & \text{otherwise} \end{cases}$$

and by linear extension to chains.

Proposition 1.4.7. *The chain morphism commutes with the boundary operator, i.e.,*

$$f_\# \circ \partial = \partial' \circ f_\#$$

(We use a prime to denote the boundary operator for G'.) Hence, f induces a morphism of homology groups $f_* : H_i(G) \to H_i(G')$, $i = 0, 1$.

Proof. The commutativity of $f_\#$ with the boundary operator is trivial. It follows that $f_\#$ sends the kernel and image of ∂ to the kernel and image of ∂'. Hence, $f_\#$ descends to a quotient $f_* : C_0/\mathrm{Im}\,\partial \to C_0'/\mathrm{Im}\,\partial'$ and restricts to a morphism $f_* : \ker\partial \to \ker\partial'$. □

It is easily checked that the composition of two graph morphisms $f \circ g$ satisfies $(f \circ g)_* = f_* \circ g_*$ and that the identity of a graph induces the identity of its homology group. In other words the association of graphs and morphisms to the corresponding homology groups and group morphisms is a functor.

Homology with other coefficients We can define homology relatively to any abelian coefficient group Γ. To this end, we define a chain of vertices or arcs as a formal combination with coefficients in Γ. The set of chains is equipped with a group structure induced by the law of Γ. Alternately, these chain groups could be defined by tensoring C_0 and C_1 with Γ. The boundary operator and the homology groups are then defined as for integer coefficients taking into account the new definition of chain groups. The homology with integer coefficients is the most general in the sense that it determines homology over any other group. This is the content of the *universal coefficient theorem* for homology [Hat02]. However, it is often convenient to restrict to other coefficients for computational reasons or to concentrate on specific properties of homology. Common choices for the coefficients include the field of rationals $\mathbb{Q}$ and the finite cyclic groups $\mathbb{Z}/p\mathbb{Z}$. A specific case occurs for $\Gamma = \mathbb{Z}/2\mathbb{Z} = \{\bar{0}, \bar{1}\}$. A chain with $\mathbb{Z}/2\mathbb{Z}$ coefficients can be interpreted as a subset of vertices or edges and the sum of two chains becomes their symmetric difference. A cycle is just a subgraph of G, each vertex of which has even degree. Such subgraphs are sometimes called *Eulerian*[1], or *even subgraphs*.

Exercise 1.4.8. Show that for a field Γ, $H_1(G, \Gamma)$ is a vector space of dimension $\beta_1(G)$ (in general a direct sum of copies of Γ) and that the set B of cycles in the proof of Proposition 1.4.4 is also a basis of $H_1(G, \Gamma)$.

1.5 Cohomology

The cohomology is defined dually to homology. We again consider a graph G and its chain groups $C_0(G)$ and $C_1(G)$. The **cochain groups** are the dual groups $C^0(G) = \hom(C_0(G), \mathbb{Z})$ and $C^1(G) = \hom(C_1(G), \mathbb{Z})$ of linear maps $C_0(G) \to \mathbb{Z}$ and $C_1(G) \to \mathbb{Z}$ respectively. We simply write C^i for $C^i(G)$ when there is no ambiguity on the underlying graph G. The elements of C^1 are also called **cocycles**[2]. The dual of the boundary

[1] The terminology can be confusing as it is also used to mean a connected Eulerian subgraph

[2] For higher dimensional complexes, the cocycle group is the kernel of the coboundary operator $C^1 \to C^2$.

operator is the **coboundary operator** $\delta : C^0 \to C^1, f \mapsto f \circ \partial$. The group $\mathrm{Im}\,\delta$ is the coboundary group whose elements are called coboundaries. The **cohomology groups** of G are

$$H^0(G) = \ker \delta \qquad \text{and} \qquad H^1(G) = C^1/\mathrm{Im}\,\delta$$

When G is not connected the cohomology groups are products of the cohomology groups of each component. This follows from the fact that the cochain groups are products of the cochain groups of the components of G and that the coboundary operator is a product of componentwise coboundaries. We can thus restrict ourselves to connected graphs.

Lemma 1.5.1. *If G is connected $H^0(G)$ is infinite cyclic (isomorphic to $\mathbb{Z}$).*

PROOF. An element f of $\ker\delta$ is such that $f(\partial a) = 0$ for any arc a, i.e. $f(o(a^{-1})) = f(o(a))$. By connectivity of G it follows that f takes the same value for all the vertices. The kernel of δ is thus the set of multiples of the map sending each vertex to one. □

Lemma 1.5.2. *The first homology group of a tree is trivial.*

PROOF. Let v be a vertex of a tree T. We consider the map

$$\sigma_T : C^1(T) \to C^0(T), \qquad g \mapsto \sigma_T(g) : w \mapsto \sum_{a \in T[v,w]} f(a)$$

We easily check that for any $g \in C^1(T)$ we have $\delta\sigma_T(g) = g$. It follows that $\mathrm{Im}\,\delta = C^1(T)$, i.e., $H^1(T)$ is trivial. □

Proposition 1.5.3. *Let T be spanning tree of a connected graph G. Then $H^1(G)$ is isomorphic to the product of copies of $\mathbb{Z}$, with one copy per chord of T in G.*

PROOF. Let C be the set of chords of T in G. We view elements of $\Pi_C\mathbb{Z}$ as functions $C \to \mathbb{Z}$. We consider the group morphism $\pi : \Pi_C\mathbb{Z} \to C^1/\mathrm{Im}\,\delta$ that maps a function $\phi : C \to \mathbb{Z}$ to the class of the cocycle $\pi(\phi) : C^1 \to \mathbb{Z}$ defined for all $a \in A_+$ by:

$$\pi(\phi)(a) = \begin{cases} \phi(a) & \text{if } a \in C \\ 0 & \text{if } a \in T. \end{cases}$$

If g is a cocycle of $C^1(G)$, we can apply the morphism σ_T of Lemma 1.5.2 to its restriction on T and extend $\delta\sigma_T(g)$ to the null function on C. Note that g and $\delta\sigma_T(g)$ restrict to the same map on T, so $g - \delta\sigma_T(g)$ cancels over T. It follows that the class of any cocycle g in $C^1/\mathrm{Im}\,\delta$ contains a cocycle that cancels over T, namely $g - \delta\sigma_T(g)$. This last cocycle restricts in turn to a function $\phi : C \to \mathbb{Z}$ with $\pi(\phi) = g - \delta\sigma_T(g)$, showing that π is onto. On the other hand, if $\pi(\phi) \in \mathrm{Im}\,\delta$ then we must have $\pi(\phi) = \delta f$ for some $f \in C^0(G)$. Because T is connected and $\pi(\phi)$ cancels over T, the cochain f must be constant on the vertices of T. Because T is spanning, δf is also null on C, whence $\phi = 0$. It follows that π is injective, hence an isomorphism. □

1.6 Some Elementary Algorithms Related to Homology

As for the fundamental group, we examine how to compute a basis of the first homology group of a connected graph G. Following the proof of Proposition 1.4.4, or by applying Proposition 1.4.5, the cycles $T[a]$ when a runs over the chords of a spanning tree T of G constitute a basis of $H_1(G)$. Such a basis is called a **fundamental cycle basis** or a **Kirchhoff basis**. When G is not connected, we can work independently on each connected component of G since homology is the direct sum of the component homologies. We can even refine this decomposition into 2-connected components (cf. Exercise 1.4.6).

We will thus assume that G is connected. When the edges of G are positively weighted, we can search for a basis that minimizes the sum of the length of its cycles. Such a basis is called a **minimum weight (cycle) basis**. Here, the length of a cycle $c = \sum_a n_a a$ is $|c|_w := \sum_a |n_a| w(a)$ where $w : C \to \mathbb{Q}_+^*$ is the weight function.

Exercise 1.6.1. Suppose that we allow a non-negative weight function to cancel on some of the edges. By introducing a new weight function with two components, show how to reduce the minimum weight basis computation to the case of a strictly positive weight function.

A simple adaptation of the Corollary 1.6.4 below to integer coefficients shows that a minimum cycle bases is made of simple cycles. We could therefore define the length as $|c|_w := \sum_a w(a)$. However, as opposed to Proposition 1.3.3, a minimal weight basis is not always a fundamental cycle basis. The counterexample in Figure 1.1 was found by Hartvigsen and Mardon [HM93]. In fact, it seems that little is known concerning the

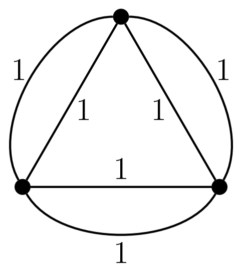


Figure 1.1: Each spanning tree in this graph is a path of length 2. The corresponding fundamental basis is composed of two cycles of length 2 and two cycles of length 3 leading to a fundamental cycle basis of total weight 10. However a minimum weight basis of total weight 9 is given by the three outer cycles of length 2 and the central triangle.

minimum weight bases of the *integer* homology of a graph. Most of the literature on the subject has been concentrated on homology with $\mathbb{Z}/2\mathbb{Z}$ coefficients. Even in this case, the same counterexample as above shows that a minimum weight basis is not always a fundamental cycle basis. Hartvigsen and Mardon [HM93] characterize the graphs possessing a minimum weight basis that is also a fundamental cycle basis, independently of the weight function. In general, looking for the minimum weight *fundamental* basis is NP-hard [DPeK82]. However, Horton [Hor87] proved that computing a minimum weight basis with $\mathbb{Z}/2\mathbb{Z}$ coefficients can be done in polynomial time. We now present his algorithm.

If B is a family of cycles of G, we denote by $\ell(B)$ the list of the cycle lengths in B in increasing order. We first observe that

Lemma 1.6.2. *A basis B of $H_1(G,\mathbb{Z}/2\mathbb{Z})$ has minimum weight if and only if $\ell(B)$ is minimal for the lexicographic order. The following algorithm thus returns a minimum weight basis.*

1. *Initialize B to the empty set.*
2. *Scan the cycles in $H_1(G,\mathbb{Z}/2\mathbb{Z})$ in an increasing order of their length. At each step, add the scanned cycle c to B if $B\cup\{c\}$ is an independent family.*
3. *return B.*

PROOF. Using the coefficient field $\mathbb{Z}/2\mathbb{Z}$ provides the homology group $H_1(G,\mathbb{Z}/2\mathbb{Z})$ with a vector space structure. It is thus a matroid to which we can apply the classical greedy algorithm. $\square$

Since $H_1(G,\mathbb{Z}/2\mathbb{Z})$ contains $2^{\beta_1(G)}$ cycles, this algorithm is not very efficient. In order to restrict the search, Horton characterizes the cycles that may belong to a minimum weight basis.

Lemma 1.6.3. *Suppose $b = c + d$ is a cycle of a basis B of $H_1(G,\mathbb{Z}/2\mathbb{Z})$. Then either $B\setminus\{b\}\cup\{c\}$ or $B\setminus\{b\}\cup\{d\}$ is a basis.*

PROOF. If c and d were both in the linear span of $B\setminus\{b\}$, then so would b. $\square$

Corollary 1.6.4. *The cycles of a minimum weight basis are simple.*

PROOF. Suppose that b is a non-simple cycle of a minimum weight basis B. Then b can be written as the sum $b = c + d$ of two edge disjoint cycles. In particular, b is longer than c or d. By the preceding lemma, we can replace b by c or d in B to get a shorter basis, contradicting the minimality of B. $\square$

Lemma 1.6.5. *Let b be a cycle of a minimum weight basis. Let p and q be two edge disjoint paths such that $b = p\cdot q^{-1}$. Then p or q is a shortest path for $|\cdot|_w$.*

PROOF. Let r be a shortest path from the common initial vertex of p and q to their common last vertex. With a little abuse of notation, we can write $b = p\cdot r^{-1} + r\cdot q^{-1}$. By Lemma 1.6.3, b must be no longer than $p\cdot r^{-1}$ or $r\cdot q^{-1}$, implying that either q or p is a shortest path. $\square$

Corollary 1.6.6. *Let v be a vertex of a simple cycle b of a minimum weight basis. Then $b = p\cdot a\cdot q^{-1}$ where a is an arc and p, q are two shortest paths with v as initial vertex.*

PROOF. We write $b = (a_1, a_2, \ldots, a_k)$ with $v = o(a_1) = o(a_k^{-1})$. Let i be the maximal index such that $(a_1, a_2, \ldots, a_i)$ is a (simple) shortest path. Then $b = (a_1, a_2, \ldots, a_i) \cdot a_{i+1} \cdot (a_{i+2}, \ldots, a_k)$ and the previous lemma implies that $(a_{i+2}, \ldots, a_k)$ is a (possibly empty) shortest path □

When there is a unique shortest path between every pair of vertices, this corollary allows us to reduce the scan of step (2) in Lemma 1.6.2 to $|V||A_+|$ cycles, one for each (vertex,edge) pair. In general this cannot be assumed[3] and there still might be too many cycles to test. Suppose that for every vertex v we choose a shortest path tree T_v with root v. For every pair $(v, a) \in V \times A_+$, let $c(v, a) = T_v[v, a]$ be the loop with basepoint v through a relatively to T_v.

Lemma 1.6.7. *The scan of step (2) in the greedy algorithm of Lemma 1.6.2 can be restricted to the loops $c(v, a)$ for $(v, a) \in V \times A_+$.*

PROOF. It is enough to prove (cf. Exercise 1.6.8 below) that there exists a minimum weight basis composed of cycles of the form $c(v, a)$ only. Let B be a minimum weight basis with the minimal number of cycles not of the form $c(v, a)$. If this minimal number is zero, we are done. Otherwise, consider a cycle $b \in B$ that is not equal to any $c(v, a)$. For a decomposition $b = p \cdot e \cdot q^{-1}$ such as in Corollary 1.6.6, let d_e be the number of arcs in b that are not in $c(w, e)$ with w the starting vertex of p and q. Define the *default value* $d(b)$ as the minimum of d_e taken over all such decompositions of b. Let $b = p_0 \cdot a \cdot q_0^{-1}$ be the decomposition for which this minimum occurs. Denote by x and y the endpoints of a and by v the starting vertex of p, so that $c(v, a) = T_v[v, x] \cdot a \cdot T_v[y, v]$. We can write

$$b = p_0 \cdot T_v[x, v] + c(v, a) + T_v[v, y] \cdot q_0^{-1}$$

Applying Lemma 1.6.3 twice, we see that b can be replaced in B by at least one of the three cycles in the above sum. The cycle $p_0 \cdot T_v[x, v]$ is either shorter than b, in which case it cannot replace b my minimality of B, or its default value is strictly less than $d(b)$ (to see this, write $p_0 \cdot T_v[x, v]$ as $p_0' \cdot e \cdot T_v[x, v]$ with $p_0 = p_0' \cdot e$). The same is true for $T_v[v, y] \cdot q_0^{-1}$. In any case, b can be replaced by a cycle whose default value is strictly less than $d(b)$. By induction on $d(b)$ we can thus assume $d(b) = 0$, i.e. $b = c(v, a)$. This is in contradiction with the assumption on B. □

Exercise 1.6.8. Suppose that the cycles of B all belong to a subset $C \subset H_1(G)$. Check that the scan of step (2) in the greedy algorithm of Lemma 1.6.2 can be restricted to C.

Exercise 1.6.9. In full generality it is not necessary to define the cycles $c(v, a)$ out of shortest path trees. For each couple (x, y) of vertices, fix any shortest path $p(x, y)$ from x to y and put $c'(v, a) = p(v, o(a)) \cdot a \cdot p(o(a^{-1}), v)^{-1}$. Show that the step (2) in Lemma 1.6.2 can be restricted to the cycles $c'(v, a)$ for all $(v, a) \in V \times A_+$ (cf. [Hor87]).

[3] A probabilistic perturbation [CCE13] technique allows to enforce this assumption, at the price of loosing determinism.

Proposition 1.6.10. *A minimum weight basis of G can be computed in $O(|V|^2\log|V| + \beta_1^2(G)|V||A|) = O(|V||A|^3)$ time.*

PROOF. By Lemma 1.6.7, we restrict the scan step of the greedy algorithm to the cycles $c(v,a)$ with $(v,a) \in V\times A_+$. For each vertex v, we compute a shortest path tree T_v in $O(|V|\log|V|+|A|)$ time using Dijkstra's algorithm. There are $\beta_1(G)$ cycles of the form $c(v,a)$, each of size $O(|V|)$. Their computation and storage for all the vertices v thus requires $O\big(|V|(|V|\log|V|+|A|+\beta_1(G)|V|)\big)$ time. They can be sorted according to their length in $O(\beta_1(G)|V|\log(\beta_1(G)|V|))$ time. In order to check if a cycle is independent of the current family of basis elements, we view a cycle as a vector in $(\mathbb{Z}/2\mathbb{Z})^{A_+}$. We use Gauss elimination to maintain the current family in row echelon form. This family has at most $\beta_1(G)$ vectors and testing a new vector against this family by Gauss elimination needs $O(\beta_1(G)|A|)$ time. The cumulated time for testing independence is thus $O(\beta_1^2(G)|A||V|)$. The whole greedy algorithm finally takes

$$O\big(|V|(|V|\log|V|+|A|+\beta_1(G)|V|)+\beta_1(G)|V|\log(\beta_1(G)|V|)+\beta_1^2(G)|A||V|\big)$$

time which reduces to $O(|V|^2\log|V|+\beta_1^2(G)|V||A|)$ after simplification. □

Note that the above scan can be further reduced by discarding the loops $c(v,a)$ that are not simple. We can also decompose a cycle into a combination of a fixed fundamental basis associated to a tree. The decomposition of a cycle is just given by its trace over the chords of that tree. This allows to represent the current family of basis elements by a matrix of size $\beta_1(G)\times\beta_1(G)$ instead of $\beta_1(G)\times A_+$.

The computation of a minimal weight basis is often designated by MCB (Minimum Cycle Basis problem). Many properties of minimum weight bases and other short cycles are discussed in Gleiss's thesis [Gle01]. This minimal weight basis problem can be recast in the more formal language of matroids, see Golinski and Horton [GH02]. The greedy algorithm as analysed in Proposition 1.6.10 is not optimal. Further improvements were proposed [KMMP04, KMMP08, MM09]. For integer coefficients the set of $\mathbb{Z}$-homology classes do not form a matroid in general. The greedy algorithm cannot be applied anymore. To my knowledge, the status of the computation of a minimal weight $\mathbb{Z}$-homology basis is still unknown.

Open problem: Decide if a minimal weight $\mathbb{Z}$-homology basis is the same as a minimal weight $\mathbb{Z}/2\mathbb{Z}$-homology basis or if its computation is a NP-hard problem.

1.7 Coverings, Actions and Voltages

Covering projections are among the most fruitful morphisms when associated to homotopy. They allow to translate topological properties into group properties, leading to surprisingly simple proofs in one of the two fields. Intuitively, a covering of a graph G is a morphism $H\to G$ that is locally an isomorphism. The graphs G and H are respectively called the **base** and the **total space** of the covering. Looking from the base or from the total space provides different ways of describing coverings. This section details those point of views, leading to a classification of coverings. All the material covered here is classical and can be found in textbooks on algebraic topology such as [Mas77]. It was later recast in the realm of graph theory [GT87, BW09].

1.7.1 Coverings

Definition 1.7.1. The **star** of a vertex v in a graph G is the set of arcs with origin v. It is denoted $\mathrm{Star}(v) = \{a \in A(G) \mid o(a) = v\}$.

Definition 1.7.2. A **graph covering** is a graph morphism $p : H \to G$ such that the restriction $p : \mathrm{Star}(w) \to \mathrm{Star}(p(w))$ is bijective for all vertex w of H. For x a vertex or arc of the base graph G, the set $p^{-1}(x)$ is called the **fiber** above x.

Figure 1.2 depicts a graph covering. If $p : H \to G$ is a covering and γ is a path in G, then

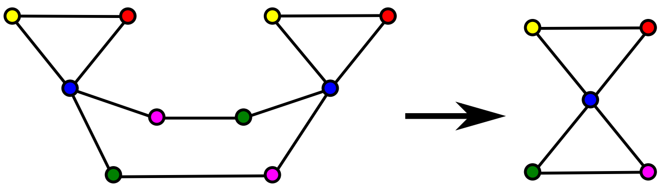


Figure 1.2: Each vertex of the left graph is sent to the vertex of the same color in the right graph. Arcs are mapped accordingly.

a path δ in H that projects to γ, i.e., such that $p(\delta) = \gamma$, is called a **lift** of γ.

Lemma 1.7.3 (Unique lift property)**.** *Let $w \in V(H)$ with $p(w) = o(\gamma)$. There exists a* unique *lift of γ with origin w.*

Figure 1.3 illustrates the property.

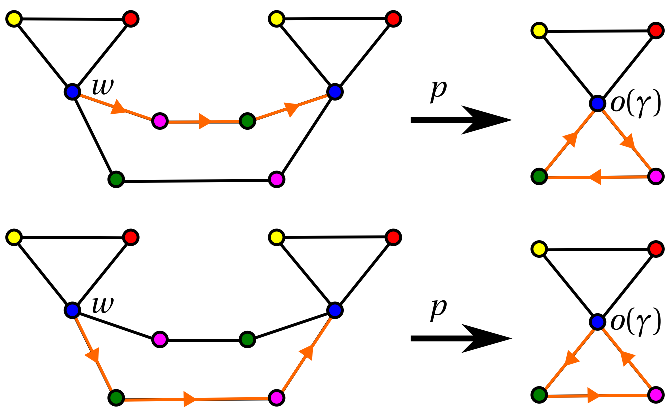


Figure 1.3: Top, the right closed path with origin $o(\gamma)$ has a unique lift starting at w. Bottom, reverting the orientation of the path changes its lift accordingly.

PROOF. Since Star(w) is sent bijectively to Star($o(\gamma)$) by p, there exists a unique arc in Star(w) sent to the first arc of γ. We can continue this way, lifting the arcs of γ one after the other. □

Lemma 1.7.4. *Let $p: H \to G$ be covering. Consider two homotopic paths α, β in G and two respective lifts $\tilde{\alpha}$ and $\tilde{\beta}$ with the same origin. Then $\tilde{\alpha}$ and $\tilde{\beta}$ are homotopic in H.*

PROOF. If α and β are related by one elementary homotopy, then so are $\tilde{\alpha}$ and $\tilde{\beta}$ since a spur $a \cdot a^{-1}$ lifts to a spur. In the general case, the lemma follows by induction on the number of elementary homotopies relating α to β. □

Hence, the final endpoint of the lift of a path α from a given vertex w only depends on the homotopy class of α. We denote by $w.[\alpha]$ this final endpoint. We trivially check that for any path β starting at the end of α:

$$w.[\alpha \cdot \beta] = (w.[\alpha]).[\beta]$$

Corollary 1.7.5. *If $p: H \to G$ is a covering, then the induced morphism $p_*: \pi_1(H, w) \to \pi_1(G, p(w))$ is one-to-one.*

PROOF. Denote by $[\alpha]$ the homotopy class of a loop α. By definition $p_*[\alpha] = p_*[\beta]$ means $p(\alpha) \sim p(\beta)$. By the preceding lemma this implies $\alpha \sim \beta$, i.e., $[\alpha] = [\beta]$. In other words p_* is one-to-one. □

A direct application of this corollary to the graph coverings of Figure 1.4 shows that a free group over a countable set of elements embeds as a subgroup of the free group over two elements!

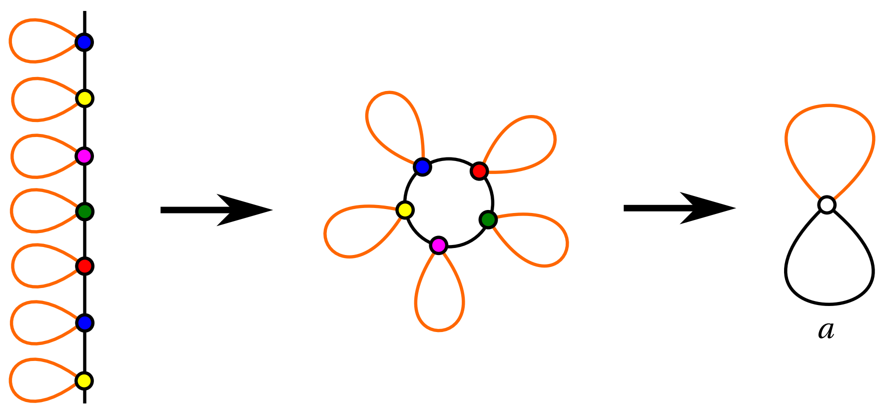


Figure 1.4: Left, an infinite graph with a countable set of generators. This graph covers the middle graph by mapping vertices and edges according to their colors. The middle graph covers the bouquet B_2 to the right. It follows that the fundamental group of the left graph, a free group over an infinite countable set of elements, embeds into the free group with $n > 2$ elements which itself embeds into $F(2)$.

Exercise 1.7.6. Let $p : Flower_5 \to B_2$ be the right covering in Figure 1.4. Call a the lower loop edge of B_2. On what condition related to a does a loop of B_2 lift to a loop in $Flower_5$? Deduce that for any vertex v of $Flower_5$, we have $p_*\pi_1(Flower_n, v) \lhd \pi_1(B_2)$, i.e. $\pi_1(Flower_n, v)$ is normal in $\pi_1(B_2)$. What is the quotient group? (Hint: you may read the rest of the section to answer this last question.)

Corollary 1.7.5 tells that the fundamental group of the total space can be seen as a subgroup of the fundamental group of the base. The reciprocal is also true.

Proposition 1.7.7. *Let v be a vertex of the connected graph G. For every subgroup $U < \pi_1(G, v)$, there exists a connected covering $p_U : (G_U, w) \to (G, v)$ with $p_{U*}\pi_1(G_U, w) = U$.*

Proof. Fix a spanning tree T of G. We write γ_a for the loop $T[v, a]$. Define G_U by

- $V(G_U) = V(G) \times \{Ug\}_{g\in\pi_1(G,v)}$,
- $A(G_U) = A(G) \times \{Ug\}_{g\in\pi_1(G,v)}$,
- $o(a, Ug) = (o(a), Ug)$ and $(a, Ug)^{-1} = (a^{-1}, Ug[\gamma_a])$,

where Ug denotes the right coset representative in $\pi_1(G, v)$ of g with respect to U. Schematically, the typical edge of G_U is

$$(o(a), Ug)\bullet \underset{(a^{-1},Ug[\gamma_a])}{\overset{(a,Ug)}{\rightleftarrows}} \bullet(o(a^{-1}), Ug[\gamma_a])$$

and let p_U be the projection on first component. Note that for a vertex x of G, $\mathrm{Star}(x, Ug) = \mathrm{Star}(x) \times \{Ug\}$. It follows that $p_U : \mathrm{Star}(x, Ug) \to \mathrm{Star}(x)$ is a bijection and that p_U is indeed a covering.

Let $\lambda = (a_1, a_2, \dots, a_k)$ be a path from v to a vertex x in G. Setting $w = (v, U)$, a simple induction on k shows that the lift of λ from w has destination $w.[\lambda] = (x, U[\gamma_{a_1}] \cdot [\gamma_{a_2}] \cdots [\gamma_{a_k}])$. In particular, this destination is (x, U) when λ is contained in T (see Figure 1.5, Left) and $(x, U[\lambda])$ when λ is a loop with homotopy class $[\lambda] \in \pi_1(G, v)$. Now, for a vertex (x, Ug) of G_U with $g = [\lambda]$, we have $w.[\lambda \cdot T[v, x]] = (x, Ug)$ (see Figure 1.5, Right). It ensues that G_U is connected. Finally, a loop λ with basepoint v satisfies

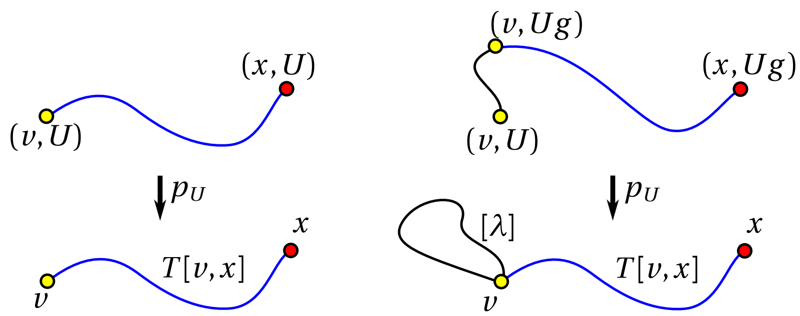


Figure 1.5: Left and right: two lifts in G_U of paths in G.

$[\lambda] \in \mathrm{Im}\, p_*$ if and only its lift starting from w is closed, i.e., $(v, U[\lambda]) = (v, U)$. In turns, this means $[\lambda] \in U$. $\square$

Example 1.7.8. If G is a 2-circuit and $U = 2\mathbb{Z} < \mathbb{Z} \simeq \pi_1(G, v)$, we obtain a covering by a 4-circuit as on Figure 1.6.

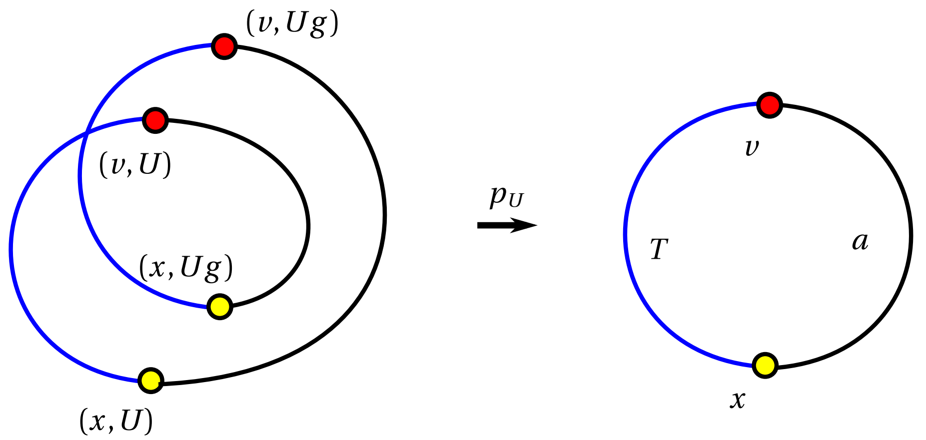


Figure 1.6: A covering of a 2-circuit. The spanning tree T is composed of a single edge. The fundamental group $\pi_1(G, v)$ is generated by $g = [\gamma_a]$, so that $U =< g^2 >$.

As an immediate application, we get

Theorem 1.7.9 (Nielsen-Schreier, mid 1920's)**.** *Every subgroup of a free group is free.*

Proof. Let $F(S)$ be a free group over S. We realize $F(S)$ as the fundamental group of the bouquet of circles with edge set S. By Proposition 1.7.7, every subgroup of $F(S)$ is the fundamental group of a graph (covering) which we know to be free[4]. □

Exercise 1.7.10. Let $p : H \to G$ be a graph covering and let α be a path from a vertex v of H to a vertex w in the same fiber as v. Show that $p_*\pi_1(H, w) = [p(\alpha)]^{-1} \cdot p_*\pi_1(H, v) \cdot [p(\alpha)]$. In particular, $p_*\pi_1(H, w)$ and $p_*\pi_1(H, v)$ are conjugate subgroups in $\pi_1(G, p(v))$.

Covering morphisms

We now consider the set of all the coverings of a given connected graph G. They can be considered as the objects of a category whose morphisms are defined as follows.

Definition 1.7.11. A **morphism** between coverings $p : H \to G$ and $q : K \to G$ is a graph morphism $f : H \to K$ that sends fibers to fibers in such a way that the diagram

$$\begin{tikzcd} H \arrow[rr, "f"] \arrow[dr, "p"'] & & K \arrow[dl, "q"] \\ & G & \end{tikzcd}$$

is commutative.

Since the restrictions of p and q to stars are bijective it must be the case for f. It follows that f is a covering. Hence, a covering morphism is a covering of (the total space of) a covering.

[4] Another quick proof uses the fact that a group is free if and only if it acts freely on a tree. But any subgroup obviously acts freely on the same tree, so that it must be free [Ser77].

Exercise 1.7.12. Let f be a morphism from the covering $p : H \to G$ to the covering $q : K \to G$. Consider a vertex v in H and a path α in G with initial vertex $p(v)$. Show the identity

$$f(v).\alpha = f(v.\alpha)$$

Exercise 1.7.13. Show that a covering morphism

$$\begin{tikzcd} (H,v) \arrow[rr] \arrow[dr, "p"'] & & (H,v) \arrow[dl, "p"] \\ & (G,u) & \end{tikzcd}$$

must be the identity.

Lemma 1.7.14. *There is a morphism between the coverings $p : (H, v) \to (G, u)$ and $q : (K, w) \to (G, u)$ if and only if $p_*\pi_1(H, v)$ is a subgroup of $q_*\pi_1(K, w)$ in $\pi_1(G, u)$.*

Proof. The condition is clearly necessary. Indeed, if f is a covering morphism as in the lemma, then by functoriality it satisfies $p_* = q_* \circ f_*$, implying $p_*\pi_1(H, v) < q_*\pi_1(K, w)$. It remains to prove that the condition is sufficient. So, we suppose $p_*\pi_1(H, v) < q_*\pi_1(K, w)$. We shall construct a covering morphism $f : H \to K$. Let x be a vertex of H and let γ be a path from the basepoint v of H to x. We put

$$f(x) = w.[\gamma]$$

If a is an arc with origin x, we set $f(a)$ to the unique edge with origin $f(x)$ that projects to $p(a)$ (see Figure 1.7). We claim that f is a well-defined map: if λ is another path

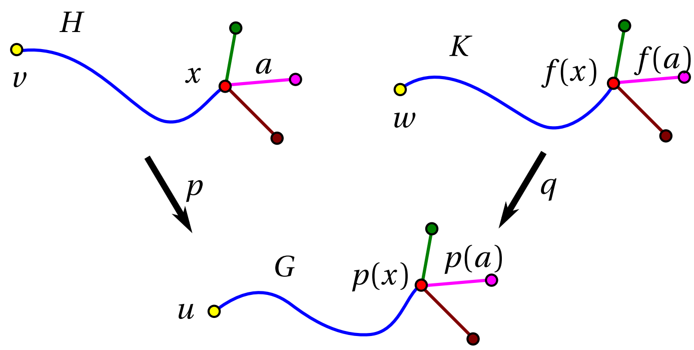


Figure 1.7: The image of $x \in V(H)$ is a vertex $f(x) \in V(K)$ obtained by lifting in K the projection in G of a path from v to x in H. Arcs are mapped accordingly.

from v to x then $w.p_*[\lambda] = (w.p_*[\lambda.\gamma^{-1}]).p_*[\gamma]$. By assumption, $p_*[\lambda.\gamma^{-1}] \in q_*\pi_1(K, w)$. This means that the lift of $p_*[\lambda.\gamma^{-1}]$ from w is closed, or equivalently: $w.p_*[\lambda.\gamma^{-1}] = w$. Whence $w.p_*[\lambda] = p_*[\gamma]$ as claimed. The map f so defined is clearly a graph morphism: it commutes with the origin and inverse operators. Finally, we have $q(f(x)) = q(w.p_*[\gamma])$ which is the final endpoint $p(x)$ of the path $p \circ \gamma$. Moreover, $q(f(a)) = p(a)$ by construction, so that $p = q \circ f$ as required. $\square$

Corollary 1.7.15. *The coverings $p : H \to G$ and $q : K \to G$ are isomorphic if and only if $p_*\pi_1(H, v)$ and $q_*\pi_1(K, w)$ are in the same conjugacy class in $\pi_1(G, u)$ for $p(v) = q(w) = u$.*

PROOF. The condition is necessary by the previous lemma. So, we suppose that $p_*\pi_1(H,v) = g^{-1}\cdot q_*\pi_1(K,w)\cdot g$ for some $g \in \pi_1(G,u)$. We easily check that $g^{-1}\cdot q_*\pi_1(K,w)\cdot g = q_*\pi_1(K,w.g)$ (see Exercise 1.7.10). It follows that $p_*\pi_1(H,v) = q_*\pi_1(K,w.g)$, and by two applications of the previous lemma, we get covering morphisms $(H,w)\to(K,v.g)$ and $(K,v.g)\to(H,w)$. By Exercise 1.7.13 those morphisms are inverse isomorphisms. □

The corollary reformulates as follows.

Theorem 1.7.16. *The set of isomorphism classes of coverings of a graph G corresponds to the set of conjugacy classes of subgroups of the fundamental group of G. The preorder relation given by the existence of a covering morphism $H\to K$ corresponds to the inclusion $g^{-1}\cdot\pi_1(H,v)\cdot g \subset \pi_1(K,w)$ for some $g\in\pi_1(G,u)$.*

The trivial group $\{1\}\subset\pi_1(G,u)$ is obviously the maximal element for this preorder. The corresponding covering is called the **universal cover**. Since its fundamental group is trivial, the universal cover is a tree by Theorem 1.2.10. Figure 1.8 shows the universal cover of the Bouquet B_2.

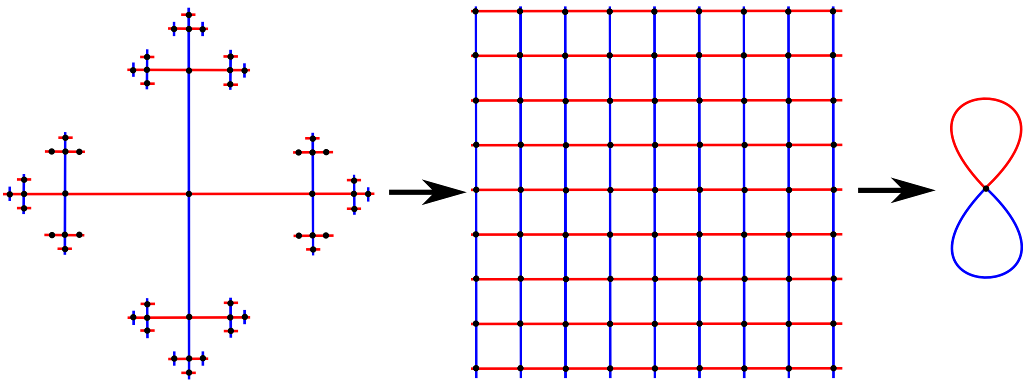


Figure 1.8: The universal cover of the $\mathbb{Z}^2$ grid is also the universal cover of B_2.

1.7.2 Actions and quotients

We denote by $Aut(G)$ the group of automorphisms of a graph G. The **orbit** of a vertex or arc x of G by a subgroup Γ of automorphisms is denoted by $\Gamma\cdot x = \{g(x)\mid g\in\Gamma\}$.

Definition 1.7.17. The subgroup $\Gamma < Aut(G)$ **acts without arc inversion** if for any arc a of G and any automorphism g in Γ, we have $g(a)\neq a^{-1}$. In other words $a^{-1}\notin\Gamma\cdot a$. If Γ acts without arc inversion, we can define the **quotient graph** G/Γ by

- $V(G/\Gamma) = \{\Gamma\cdot v\}_{v\in V(G)}$,
- $A(G/\Gamma) = \{\Gamma\cdot a\}_{a\in A(G)}$,
- $o(\Gamma\cdot a) = \Gamma\cdot o(a)$ and $(\Gamma\cdot a)^{-1} = \Gamma\cdot a^{-1}$

Note that Γ acting without inversions, we have $(\Gamma \cdot a)^{-1} \neq \Gamma \cdot a$, i.e., the arc inversion is fixed point free. The **quotient map** $p_\Gamma : G \to G/\Gamma$ sending a vertex or arc to its orbit is obviously a graph morphism.

Although the quotient map is onto, it is generally not a covering as illustrated on Figure 1.9.

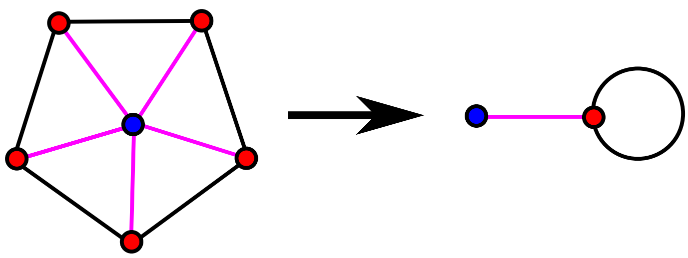


Figure 1.9: The quotient of the wheel graph with 5 spokes by the subgroup of automorphism generated by the rotation with angle $2\pi/5$ about the center (blue) vertex. Note that the quotient map is *not* a covering.

Definition 1.7.18. A group of automorphisms $\Gamma < Aut(G)$ **acts freely** on G if it acts without arc inversion and each automorphism in Γ that is not the identity is fixed vertex free (i.e., does not fix any vertex). Intuitively, this means that the corresponding topological (PL) automorphisms (extending the vertex maps to the edges in the obvious way) are fixed point free. Indeed, acting without inversion prevents the automorphisms from fixing the middlepoint of edges and being fixed vertex free prevents them from fixing the edge endpoints.

Proposition 1.7.19. *If Γ acts without arc inversion on G, then $p_\Gamma : G \to G/\Gamma$ is a covering if and only if Γ acts freely on G.*

PROOF. Since p_Γ is onto, it is a covering if and only if its restriction to stars is one-to-one. This is equivalent to say that whenever a, b are two distinct arcs with common origin then $\Gamma.a \neq \Gamma.b$. To prove the proposition, we rather show the contrapositive: there exists two distinct arcs a, b of common origin with the same orbit if and only if there exists a vertex v fixed by some automorphism $g \in \Gamma \setminus \{Id\}$. Indeed, if $\Gamma.a = \Gamma.b$ then $a = g(b)$ for some $g \in \Gamma \setminus \{Id\}$. This implies $v = g(v)$ for $v = o(a)$. On the other hand, if $g \neq Id$ fixes a vertex v, we consider the set of arcs fixed by g. This set induces a subgraph H fixed by g. Since $g \neq Id$ we have $H \subsetneq G$ and there must be an arc a whose origin is in H but that is not fixed by g. Then a and $b = g(a)$ are two distinct arcs with common origin in the same orbit (see Figure 1.10). $\square$

Lemma 1.7.20. *If Γ acts freely on G then $(p_\Gamma)_* \pi_1(G, v) \lhd \pi_1(G/\Gamma, \Gamma \cdot v)$.*

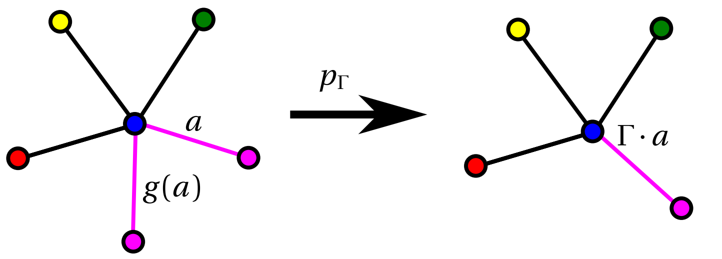


Figure 1.10: If a and $g(a)$ are two distinct arcs with common origin for some $g \in \Gamma \setminus \{Id\}$ then p_Γ is not a covering.

Proof. Let $p_\Gamma(\alpha)$ be a representative of an element in $(p_\Gamma)_*\pi_1(G,v)$ and let β be a loop with basepoint $\Gamma.v$ in G/Γ. We just need to show that the conjugate $\beta \cdot p_\Gamma(\alpha)\cdot\beta^{-1}$ represents a class in $(p_\Gamma)_*\pi_1(G,v)$, or equivalently that the lift of $\beta\cdot p_\Gamma(\alpha)\cdot\beta^{-1}$ starting from v is a (closed) loop.

Since the lift of β in G/Γ trivially projects to β, we have $p_\Gamma(v.\beta) = p_\Gamma(v)$. This means that $v.\beta \in \Gamma.v$, i.e., that there exists $g \in \Gamma$ with $g(v) = v.\beta$. Hence,

$$v.(\beta\cdot p_\Gamma(\alpha)\cdot\beta^{-1}) = g(v).(p_\Gamma(\alpha)\cdot\beta^{-1}) = (g(v).p_\Gamma(\alpha)).\beta^{-1}$$

On the other hand, the lift of $p_\Gamma(\alpha)$ from $g(v)$ is $g(\alpha)$ (see Figure 1.11) and is thus closed. It follows that $(g(v).p_\Gamma(\alpha)).\beta^{-1} = g(v).\beta^{-1} = v$, which was to be proved. □

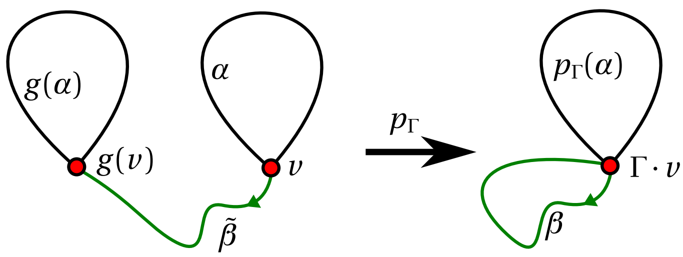


Figure 1.11: The lift of $p_\Gamma(\alpha)$ from $g(v)$ is $g(\alpha)$.

Definition 1.7.21. If $p : H \to G$ is a graph covering, we denote by $Aut(p)$ the group of automorphisms of p. This is the subgroup of $Aut(H)$ composed of the automorphisms f of H preserving the fibers of p, i.e., such that the diagram
$$\begin{tikzcd} H \arrow[rr, "f"] \arrow[dr, "p"'] & & H \arrow[dl, "p"] \\ & G & \end{tikzcd}$$
commutes. Automorphisms in $Aut(p)$ shuffle the vertices in each fiber and are sometimes called **deck transformations** by analogy with the shuffling of a deck of playing cards.

Lemma 1.7.22. *$Aut(p)$ acts freely on H.*

Proof. Let $f \in Aut(p)$. Since $p(f(a)) = p(a)$ for all arcs a, we cannot have $f(a) = a^{-1}$. For the arc $p(a)$ would be equal to its inverse $p(a)^{-1}$, a contradiction. It follows that $Aut(p)$ acts without arc inversion. On the other hand, suppose that f fixes some vertex v. Consider any other vertex w of H and a path α from v to w. We compute

$$f(w) = f(v).p(f(\alpha)) = v.p(\alpha) = w$$

So that f fixes all the vertices. Moreover, the restriction of p to stars being bijective, the commutation equation $p(f(a)) = p(a)$ together with $o(f(a)) = o(a)$ implies $f(a) = a$. Consequently, f must be the identity morphism and the action of $Aut(p)$ is fixed vertex free. □

In conjunction with Proposition 1.7.19 this lemma implies that the quotient projection $H \to H/Aut(p)$ is a covering. It is natural to ask whether this covering is isomorphic to p. In particular, when arises p as a quotient, we have

Lemma 1.7.23. *If $\Gamma < Aut(G)$ acts freely on G then $Aut(p_\Gamma) = \Gamma$.*

Proof. Obviously, $\Gamma \subset Aut(p_\Gamma)$ and Γ acts transitively on the fibers of p_Γ. Since $Aut(p_\Gamma)$ acts freely by the previous lemma, this implies $Aut(p_\Gamma) \subset \Gamma$. Indeed, fix a vertex v in H and let $f \in Aut(p_\Gamma)$. Then, $f(v)$ being in the fiber of v, the transitive action of Γ implies the existence of $g \in \Gamma$ with $g(v) = f(v)$. But $f \circ g^{-1}$ is an automorphism of $Aut(p_\Gamma)$ fixing v, so it must be the identity. Whence $f = g \in \Gamma$. □

Lemma 1.7.24. *If $p : (H, v) \to (G, u)$ is a covering with $p_*\pi_1(H, v) \lhd \pi_1(G, u)$ then $Aut(p)$ acts transitively on the fiber of v.*

Proof. Let w be a vertex in the fiber of v. We first remark that $p_*\pi_1(H, v)$ being normal in $\pi_1(G, u)$, we have $p_*\pi_1(H, w) = p_*\pi_1(H, v)$ (see Exercise 1.7.10). We shall construct an automorphism $f \in Aut(p)$ such that $f(v) = w$. For a vertex x of H and a path α from v to x, we set

$$f(x) = w.[p(\alpha)]$$

See Figure 1.12. f is well-defined. Indeed, if β is another path from v to x then $\beta \cdot \alpha^{-1}$ is a

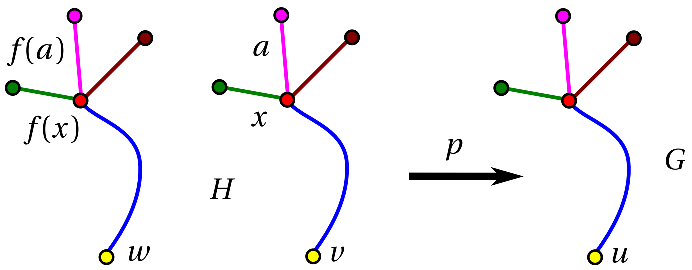


Figure 1.12: To define $f(x)$, we "translate" to w the origin of a path from v to x.

loop with basepoint v. By the preceding remark, it ensues that $[p(\beta \cdot \alpha^{-1})] \in p_*\pi_1(H, w)$. It follows that the lift of $p(\beta \cdot \alpha^{-1})$ from w is closed. We can thus write

$$w.[p(\beta)] = w.[p(\beta \cdot \alpha^{-1})][p(\alpha)] = w.[p(\alpha)]$$

We can easily extend f to arcs in order to define a p-automorphism. The details are left to the reader. □

Exercise 1.7.25. With the assumptions of the lemma show that $Aut(p)$ acts transitively on any fiber, not just the fiber of v.

Proposition 1.7.26. *Let $p:(H,v)\to(G,p(v))$ be a covering and let $\Gamma < Aut(H)$ be a subgroup of automorphisms of H. Then, p and p_Γ are isomorphic, i.e., there is an isomorphism $H/\Gamma\to G$ making the following diagram commutative*
$$\begin{tikzcd} & H \arrow[dl, "p_\Gamma"'] \arrow[dr, "p"] & \\ H/\Gamma \arrow[rr, "\simeq"] & & G \end{tikzcd}$$
if and only if

1. *$\Gamma = Aut(p)$, and*
2. *$p_*\pi_1(H,v) \lhd \pi_1(G,p(v))$*

PROOF. Condition (1) is necessary: if p_Γ is a covering then Γ acts freely on H by Lemma 1.7.19. Lemma 1.7.23 then states that $\Gamma = Aut(p_\Gamma)$. In turn, we have $Aut(p_\Gamma) = Aut(p)$ by the commutativity of the diagram in the lemma. So that $\Gamma = Aut(p)$ as claimed. Condition (2) is also necessary: by Lemma 1.7.20, we also have $p_{\Gamma*}\pi_1(H,v)\lhd \pi_1(H/\Gamma,p_\Gamma(v))$ whence $p_*\pi_1(H,v)\lhd\pi_1(G,p(v))$, again by the commutativity of the diagram in the lemma.

It remains to prove that conditions (1) and (2) are sufficient. By Exercise 1.7.25 those conditions imply that $Aut(p)$ acts transitively on each fiber of $p:H\to G$. Since this action is free by Lemma 1.7.19, it follows that $H/Aut(p)\simeq G$. □

Definition 1.7.27. A covering as in the proposition, i.e., such that the fundamental group of the total space is normal in the fundamental group of the base, is called **normal** or **regular** or **Galois**.

Proposition 1.7.28. *Let $p:(H,v)\to(G,p(v))$ be a covering. Then*

$$Aut(p)\simeq N\left(p_*\pi_1(H,v)\right)/p_*\pi_1(H,v),$$

where $N\left(p_*\pi_1(H,v)\right)$ is the ***normalizer*** *of $p_*\pi_1(H,v)$, i.e., the largest subgroup of $\pi_1(G,p(v))$ containing $p_*\pi_1(H,v)$ as a normal subgroup. In particular, if p is a normal covering then $Aut(p)\simeq\pi_1(G,p(v))/p_*\pi_1(H,v)$.*

PROOF. Let $\lambda\in N\left(p_*\pi_1(H,v)\right)$. We claim that there exists an automorphism $f_\lambda\in Aut(p)$ such that $f_\lambda(v)=v.\lambda$. Indeed, we have from Exercise 1.7.10 that $p_*\pi_1(H,v.\lambda)=\lambda^{-1}\cdot p_*\pi_1(H,v)\cdot\lambda = p_*\pi_1(H,v)$. We can thus construct the desired automorphism as in the proof of Lemma 1.7.24. By Lemma 1.7.22 this automorphism is unique and we have a well-defined map $\varphi: N\left(p_*\pi_1(H,v)\right)\to Aut(p), \lambda\mapsto f_\lambda$. By connectivity of G, this map is onto. We next compute $f_{v.(\alpha\cdot\beta)}(v) = v.(\alpha\cdot\beta) = (v.\alpha).\beta = f_{v.\alpha}(v).\beta$. But $f_{v.\alpha}(v).\beta = f_{v.\alpha}(v.\beta) = f_{v.\alpha}\circ f_{v.\beta}(v)$ (see Exercise 1.7.12). It follows that $\varphi(\alpha\cdot\beta)=\varphi(\alpha)\circ\varphi(\beta)$ showing that φ is a group morphism. We finally note that

$$\ker\varphi = \{\alpha\in N\left(p_*\pi_1(H,v)\right) \mid v.\alpha = v\} = p_*\pi_1(H,v)$$

We conclude as desired that $Aut(p)$ is isomorphic to the quotient $N\left(p_*\pi_1(H,v)\right)/p_*\pi_1(H,v)$. □

Exercise 1.7.29. What is the automorphism group of the covering in the right Figure 1.8? Is this a regular covering? Describe the fundamental group of the total space as a subgroup of the fundamental group of the base. Use Proposition 1.7.28 to justify your answer to the first question.

1.7.3 Voltage Graphs

Voltage graphs provide a concise way to encode a graph covering by labelling the arcs of the base graph. They were introduced by Gross and Tucker (see [BW09, Ch. 1] for references).

Definition 1.7.30. A **voltage** on a graph G with values in a group Γ is a map $\kappa : A(G) \to \Gamma$ that commutes with the relevant inverse operations:

$$\forall a \in A(G), \quad \kappa(a^{-1}) = \kappa(a)^{-1}$$

When Γ acts on the right on a set F, the voltage κ induces a covering $p_\kappa : G_\kappa \to G$ where G_κ is the graph defined by

- $V(G_\kappa) = V(G) \times F$,
- $A(G_\kappa) = A(G) \times F$,
- $o(a,s) := (o(a), s)$ and $(a,s)^{-1} := (a^{-1}, s.\kappa(a))$ for all $(a,s) \in A(G) \times F$,

and p_κ is the projection on the first component, $(x,s) \mapsto x$. Schematically, the typical edge of G_κ is

$$(o(a),s)\bullet \underset{(a^{-1},s.\kappa(a))}{\overset{(a,s)}{\rightleftarrows}} \bullet(o(a^{-1}), s.\kappa(a))$$

It is a simple matter of definition to check that p_κ is indeed a covering.

Exercise 1.7.31. Give a necessary and sufficient condition on κ and Γ for G_κ to be connected.

In fact, every covering arises this way.

Lemma 1.7.32. *Every covering $p : H \to G$ is isomorphic to a covering induced by some voltage on G.*

Proof. Let T be a spanning tree of G. We set $\Gamma = \pi_1(G, v)$ for some fixed vertex v. The group Γ acts on the right on the fiber $F = p^{-1}(v)$ in the usual way, letting $w.\lambda$ be the final vertex of the lift of λ starting from w. We now define $\kappa(a)$ as the homotopy class of the loop $T[v,a]$. We thus have an induced covering $p_\kappa : G_\kappa \to G$. We shall prove that there exists an isomorphism $\varphi : G_\kappa \to H$ making the following diagram commutative:

$$\begin{tikzcd} G_\kappa \arrow[rr, "\varphi"] \arrow[dr, "p_\kappa"'] & & H \arrow[dl, "p"] \\ & G & \end{tikzcd}$$

To this end, for any two vertices x, y of G we introduce a map $f_x^y : p^{-1}(x) \to p^{-1}(y)$ between their fibers:

$$\begin{array}{ccc} p^{-1}(x) & \xrightarrow{f_x^y} & p^{-1}(y) \\ u & \mapsto & u.[T[x,y]] \end{array}$$

Note that f_x^y and f_y^x are inverse to each other. We next define $\varphi : G_\kappa \to H$ by

$$\begin{cases} \forall (w,x) \in V(G) \times F : \varphi(w,x) = f_v^w(x) \\ \forall (a,x) \in A(G) \times F : \varphi(a,x) \text{ is the unique arc with origin } f_v^{o(e)}(x) \text{ above } a \end{cases}$$

and $\psi : H \to G_\kappa$ by

$$\begin{cases} \forall u \in V(H) : \psi(u) = (p(u), f_{p(u)}^v(u)) \\ \forall e \in E(H) : \psi(e) = (p(e), f_{p(o(e))}^v(o(e))) \end{cases}$$

It is an easy exercise to check that φ and ψ are inverse morphisms making the above diagram commute. □

Fix a vertex v in G. A voltage $\kappa : A(G) \to \Gamma$ extends to the loops with basepoint v by defining

$$\kappa(a_1, a_2, \ldots, a_k) = \kappa(a_1)\kappa(a_2)\cdots\kappa(a_k)$$

An elementary homotopy on the loop $(a_1, a_2, \ldots, a_k)$ leaves this value unchanged, so that κ induces a group morphism $\bar{\kappa} : \pi_1(G, v) \to \Gamma$.

Exercise 1.7.33. Let $p_\kappa : G_\kappa \to G$ be the induced covering of a voltage $\kappa : A(G) \to \Gamma$ on the graph G with Γ acting on a set F. Fix a vertex v of G and an element $s \in F$. Show that $(p_\kappa)_* \pi_1(G_\kappa, (v,s)) = (\bar{\kappa})^{-1}(S_s)$ where $S_s \subset \Gamma$ is the stabilizer of s.

Proposition 1.7.34. *A covering $p : H \to G$ is normal if and only if it is induced by a voltage κ on G with values in a group Γ acting on itself by right translations. Here, it is assumed that the induced morphism $\bar{\kappa} : \pi_1(G, v) \to \Gamma$ is onto for some fixed vertex v of G. Otherwise we can still replace Γ by the range of $\bar{\kappa}$.*

Note that requiring Γ to act on itself is equivalent to require that Γ acts freely and transitively.

PROOF. We first assume that we are given a voltage as in the proposition. Considering the basepoint $(v, 1_\Gamma)$ in G_κ we easily check that

$$(v, 1_\Gamma).\lambda = (v, \bar{\kappa}(\lambda)) \tag{1.1}$$

It follows that $p_{\kappa *}\pi_1(G_\kappa, (v, 1_\Gamma)) = \ker \bar{\kappa}$ (the set of homotopy classes with closed lift). It ensues that $p_{\kappa *}\pi_1(G_\kappa, (v, 1_\Gamma))$ is normal in $\pi_1(G, v)$, i.e., that p_κ is a normal covering. Remark that $\bar{\kappa}$ being surjective implies with (1.1) that G_κ is connected.

We now assume given a normal covering $p : (H, w) \to (G, v)$. Let T be a spanning tree of G. For every arc a of G, lemmas 1.7.22 and 1.7.24 imply the existence of a unique automorphism $f_a \in Aut(p)$ with $f_a(w) = w.[T[v, a]]$. We put $\Gamma = Aut(p)$ and $\kappa(a) = f_a$

and let Γ acts on itself on the right. It remains to check that p and p_κ are isomorphic coverings. We define $\varphi : G_\kappa \to H$ by

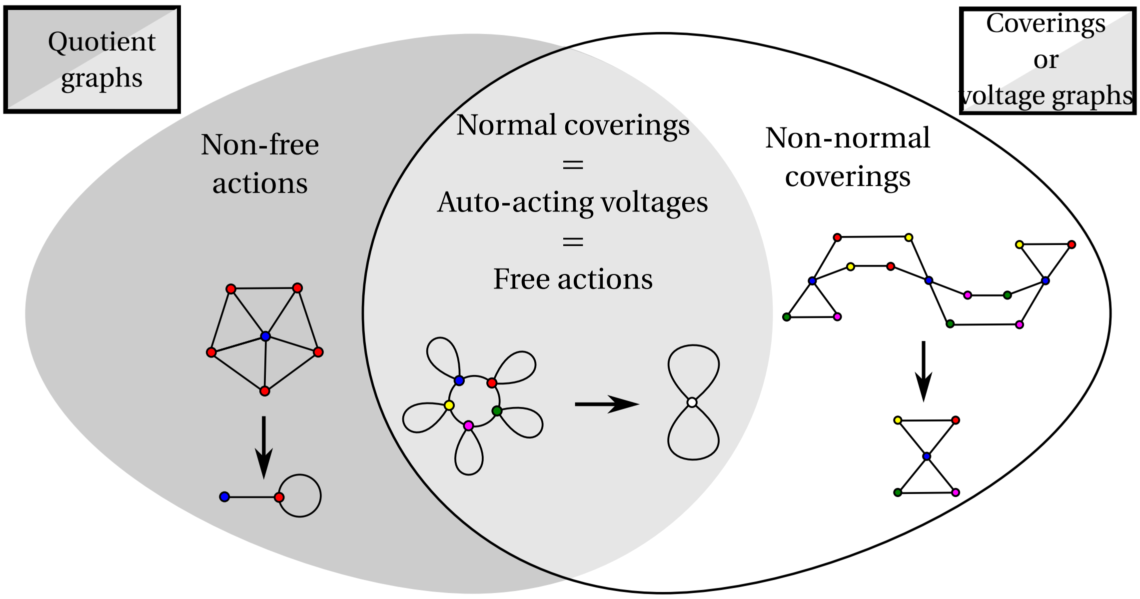
$$\varphi(x, f) = f(w.[T[v, x]])$$

and by extending φ to arcs in the unique way to make it a covering morphism. We also define $\psi : H \to G_\kappa$ by

$$\psi(y) = (p(y), f^y_{w.[T[v,p(y)]]})$$

extending it to arcs. We trivially check that φ and ψ are inverse morphisms. $\square$

We end this section on graph coverings with a graphics representing the different types of quotients and coverings.



Chapter 2

Combinatorial Surfaces

Contents

Topologically, a map is a 2-cellular embedding of a graph in a 2-dimensional manifold. This is a drawing of a graph in a topological surface without crossing of the edges such that the embedded graph dissects the surface into topological open discs. Figure 2.1 shows a cellular embedding in a genus two surface. Up to homeomorphism,

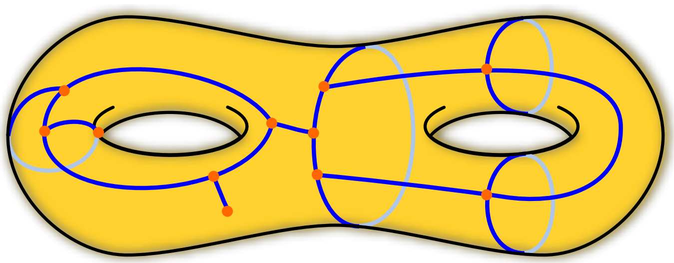


Figure 2.1: The complement of the graph in the surface is a disjoint union of open discs.

such a cellular embedding can be described by the graph together with the circular ordering of the edges incident to each vertex. These are purely combinatorial data referred to as a **combinatorial map**, a **combinatorial surface**, a **cellular embedding of a graph**, or just a **map**.

The theory of combinatorial maps was developed from the early 1970's in two parallel and independent directions. Both developments acknowledge the original works of Heffter [Hef91, Hef98] and Edmonds [Edm60] for the notion of combinatorial description of a graph embedded on a surface. On the more abstract side, mathematicians have succeeded to make beautiful connections between analysis, topology and algebra, going from Riemann surfaces and their coverings to algebraic curves and Galois theory of field extensions. Those connections were crystallised by Grothendieck through the notion of *dessins d'enfants* thanks to the Belyĭ's theorem (see the gentle introduction by Zvonkin [Zvo]).

On the combinatorial side, maps appeared as the adequate formalism for topological graph theory such as exposed in a dedicated volume of the Cambridge Encyclopedia of mathematics [BW09]. Applications range from colouring problems, such as the four colour theorem and its generalization to higher genus surfaces, to embedding characterizations generalizing Kuratowski's theorem, up to the modern structural graph theory of Robertson and Seymour. The monograph by Mohar and Thomassen [MT01] is another important reference representing this trend. Pushing the combinatorial aspect to its limit, Tutte [Tut73, Tut79] was among the first to develop an axiomatic theory of combinatorial surfaces. His aim was to banish any reference to topology while getting equivalent results such as the Jordan's curve theorem [Tut79, Sta83, VL89], using combinatorial properties only. This point of view lead Tutte [Tut79] "... to eschew diagrams ... because of their topological flavour". This might appear as a rather extreme attitude, although necessary when it comes to implementing algorithms.

A third development appeared in the early 1990's concerning curves on surfaces with a strong algorithmic objective [VY90, DS95]. Those works were recognized as part of *Computational topology* [Veg97, DEG98], a branch of Computational geometry focusing on algorithmic problems related to the topology of discrete structures. The point of view of Tutte is especially well suited to the computational aspects. This is the finality I have in mind while writing these notes. In particular, a special treatment for curves drawn on surfaces is needed to cover the most basic problems in the field. Such curves may be used for cutting surfaces into pieces or for optimization purposes (e.g., find the shortest curve homotopic to a given curve).

In the framework of combinatorial maps, a curve is just a path[1] in the associated embedded graph. Quite often, a path will have to go several times along a same edge and still should be considered as simple.

2.1 Oriented Maps

We start with the description of combinatorial orientable surfaces. Although they can be considered as special cases of general surfaces, orientable or not, they deserve their own treatment as a simpler introduction to combinatorial surfaces. Their connection with Riemann surfaces through the theory of *dessins d'enfants* also provides them with a well established status. Indeed, Riemann surfaces are naturally oriented: such a surface is defined by a complex analytic atlas whose transition maps have positive Jacobians by the Cauchy-Riemann equations.

Definition 2.1.1. An **oriented map** is a triple $M = (A, \rho, \iota)$ where

- A is a set whose elements are called **arcs**,

[1] I use the word *path* as a synonym for *walk* : a sequence of arcs connected by their ends with possible repetitions.

- $\rho : A \to A$ is a permutation of A,
- $\iota : A \to A$ is a fixed point free involution.

The permutations ρ and ι generate a subgroup of the permutation group of A called the **cartographic group** or the **monodromy group** of M (see below for an explanation of the terminology). The oriented map M has an associated **graph** $G(M) = (A/\langle\rho\rangle, A, o, \iota)$ whose vertices are the cycles of ρ, i.e., the orbits of the cyclic group of permutations $\langle\rho\rangle$ generated by ρ. The origin of an arc a is defined as the orbit $o(a) = \langle\rho\rangle a$. We will equally refer to a **vertex** of $G(M)$ as a vertex of M and denote the set of vertices of M by $V(M)$. A map is **connected** if its graph is connected, or equivalently, if its monodromy group acts transitively on its arcs. All the surfaces will be assumed connected in this section.

A **face** of M is a cycle of the permutation $\rho \circ \iota$. The face of an arc a is denoted $F(a)$ and the set of faces of M is denoted by $F(M)$. The **star** of a vertex or face x, denoted $\mathrm{Star}(x)$, is the set of arcs in the corresponding cycle. In particular, $\mathrm{Star}(x) = F(a)$ for $x = \langle\rho \circ \iota\rangle a$. Since vertices and faces are defined as orbits, they are formally the same as their star. We will nonetheless avoid to say that an arc belongs to a vertex x and rather say that it belongs to $\mathrm{Star}(x)$, or is **incident** to x. The size of $\mathrm{Star}(x)$ is the **degree** of x.

The permutation ρ is sometimes designated as a **rotation system** as it encodes the cyclic ordering of the arcs incident to a vertex. An oriented map can equivalently be described as a pair (G, ρ) where G is a graph in the sense of Definition 1.1.1 and ρ is permutation on the arcs of G whose cycles are the stars of the vertices of G.

Proposition 2.1.2. *To every cellular embedding $\eta : G \to S$ of a graph G in a topological oriented surface S we can associate an oriented map $M(\eta) = (G, \rho)$ where ρ is the rotation system corresponding to the oriented cyclic orderings of the vertex stars induced by the embedding. Conversely, every oriented map M can be realized as a cellular embedding η of its graph $G(M)$ such that $M(\eta)$ is isomorphic to M (see below for the definition of map isomorphisms).*

This proposition is essentially stated here to guide the intuition of the reader that would encounter maps for the first time. Its presence somehow contradicts the implicit credo that a purely combinatorial theory of surface can be developed without reference to topology. But, possibly in contradiction with Tutte, we strongly believe in the benefit of diagrams and topological intuition. A proof of the proposition can be found in Mohar and Thomassen's book [MT01] or Bryant and Singerman's foundational paper [BS85] for topological surfaces and in [GGD12] for the complex analytic case. Given an oriented map M there are two basic ways of visualizing the corresponding cellular embedding. One way is to consider for each face of M an oriented polygon with one side per arc in the corresponding face cycle. These polygons are further glued so that the sides corresponding to an arc and to its opposite are identified (see Figure 2.2). Another way, consists in thickening the graph of the map to transform it into a *ribbon graph*. We obtain a surface with boundaries that we can close with discs (see Figure 2.3). Guided by the topological realization of a map, we have

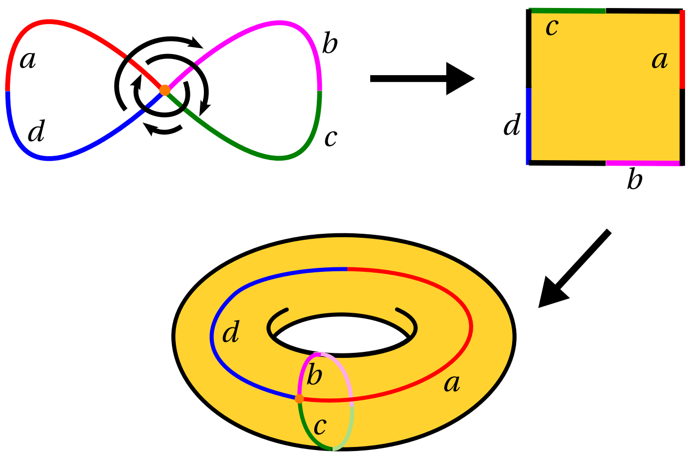


Figure 2.2: A cellular embedding associated to the map (A,ρ,ι) with $A=\{a,b,c,d\}$, $\rho=(a,c,d,b)$ and $\iota=(a,d)(c,b)$. The arcs are represented as colored half edges.

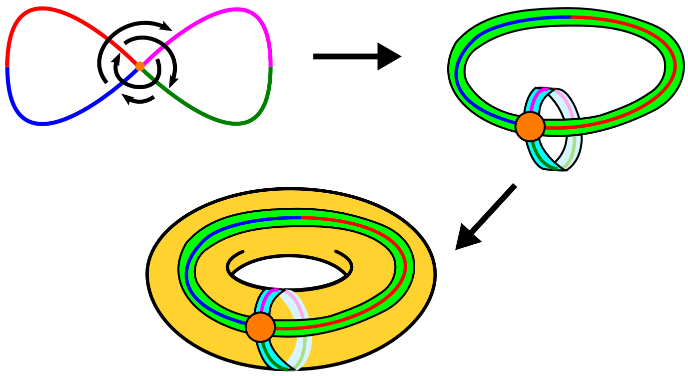


Figure 2.3: The same map as above. The unique vertex of the corresponding graph is replaced by a disc and each edge is replaced by a strip attached to the disc in the cyclic order of ρ. The resulting surface with boundary is closed with a single disc corresponding to the unique face of the map.

Definition 2.1.3. The **Euler characteristic** of a finite oriented map is the integer

$$\chi(M)=|V(M)|-|A|/2+|F(M)|$$

Its **genus** is the non-negative integer $g(M)=1-\chi(M)/2$.

Exercise 2.1.4. Show that $g(M)$ is indeed a non-negative integer.

We now define the morphisms between oriented maps. Intuitively, a morphism of combinatorial surfaces correspond to a branched covering of their topological realizations. The intuition will be made more precise below.

Definition 2.1.5. A **morphism** of oriented maps $(A,\rho,\iota)\to(B,\sigma,\jmath)$ is a function $f:A\to B$ that commutes with the rotation systems and with the opposite operators, i.e., such that

- $f\circ\rho=\sigma\circ f$, and

- $f \circ \iota = \jmath \circ f$.

Lemma 2.1.6. *Any morphism $f : (A, \rho, \iota) \to (B, \sigma, \jmath)$ is onto and sends stars to stars surjectively. Moreover, for any vertex or face x of finite degree of the map (A, ρ, ι), the restriction of f to* Star(x) *is isomorphic to the quotient*

$$\begin{array}{rcl} \mathbb{Z}/(e_x d)\mathbb{Z} & \to & \mathbb{Z}/d\mathbb{Z} \\ i \mod e_x d & \mapsto & i \mod d \end{array}$$

*where d is the size of $f(\text{Star}(x))$ and e_x is a positive integer called the **ramification index** of f at x.*

Proof. Let $a \in A$ be an arc. By connectedness, its orbit by the monodromy group satisfies $\langle \rho, \iota \rangle a = A$. Since f commutes with the rotation systems and the opposite operators, we have $f(\langle \rho, \iota \rangle a) = \langle \sigma, \jmath \rangle f(a) = B$. Thus f is onto. We also have for any integer n that $f \circ \rho^n(a) = \sigma^n \circ f(a)$. It follows that the size of the orbit $\langle \rho \rangle a$, i.e., the degree of the vertex $x = o(a)$, is a multiple of the degree d of the vertex $o(f(a))$. Whence $\deg(x) = e_x d$ for some positive integer e_x and the lemma follows for x a vertex. An analogous property holds when replacing ρ by $\rho \circ \iota$ and σ by $\sigma \circ \jmath$ proving the lemma when x is a face. □

Thanks to the lemma we can define the image of a vertex or face x of the map $M = (A, \rho, \iota)$ as the vertex or face of $N = (B, \sigma, \jmath)$ whose star is $f(\text{Star}(x))$. In particular, we can associate to f a graph morphism $f : G(M) \to G(N)$. Note that this graph morphism is dimension preserving: a vertex or arc is mapped to a vertex or arc, respectively.

Exercise 2.1.7. Prove that a morphism $f : (A, \rho, \iota) \to (B, \sigma, \jmath)$ induces a group epimorphism $\hat{f} : \langle \rho, \iota \rangle \twoheadrightarrow \langle \sigma, \jmath \rangle$ between the corresponding monodromy groups such that $f \circ \theta = \hat{f}(\theta) \circ f$ for all $\theta \in \langle \rho, \iota \rangle$.

Lemma 2.1.8. *All the edge fibers of a morphism $f : (A, \rho, \iota) \to (B, \sigma, \jmath)$ have the same size called the **degree** of f, and denoted* $\deg(f)$.

Proof. Let $b, b' \in B$. Since $\langle \sigma, \jmath \rangle$ acts transitively, there is some $\tau \in \langle \sigma, \jmath \rangle$ such that $b' = \tau(b)$. By the above exercise we can write $\tau = \hat{f}(\theta)$ for some $\theta \in \langle \rho, \iota \rangle$. Now, the equation $f(a) = b$ is equivalent to $\tau(f(a)) = \tau(b)$, i.e., $f(\theta(a)) = b'$. It follows that θ establishes a bijection from $f^{-1}(b)$ to $f^{-1}(b')$. □

A morphism of oriented maps $f : M \to N$ can be realized as a branched covering of degree $\deg(f)$, preserving the orientation, between the corresponding topological surfaces $S(M)$ and $S(N)$. The branch points of the covering are vertices and face centers of M with ramification index as defined above. Intuitively, we can view the branched covering as a projection where each point of $S(N)$ has $\deg(f)$ preimages in $S(M)$ except for the images of the branch points. In the neighbourhood of a branch point, the projection looks like the map $z \mapsto z^k$ in the complex plane $\mathbb{C}$ with k equals to the ramification index of the branch point. Moreover, the branched covering restricts to a surjection $G(M) \to G(N)$. This correspondence between morphisms and branched coverings can be performed in the realm of Riemann surfaces [GGD12], providing adequate functors.

Proposition 2.1.9 (Index formula). *Let $f:(A,\rho,\imath)\to(B,\sigma,\jmath)$ be a morphism of finite oriented maps. For any vertex or face w of $(B,\sigma,\jmath)$, we have*

$$\sum_{f(v)=w} e_v = \deg(f)$$

PROOF. Suppose that w is a vertex and consider an arc $a \in \mathrm{Star}(w)$. We partition $f^{-1}(a)$ according to vertex stars: $f^{-1}(a) = \bigcup_{f(v)=w}(f^{-1}(a)\cap \mathrm{Star}(v))$. Because $\mathrm{Star}(v)$ wraps around $\mathrm{Star}(w)$ exactly e_v times, each intersection $f^{-1}(a)\cap\mathrm{Star}(v)$ contains e_v arcs (see Figure 2.4). The proposition then follows from Lemma 2.1.8. Replacing vertices

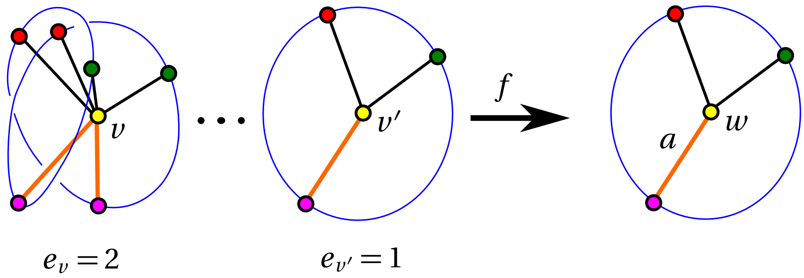


Figure 2.4: The preimage of the star of w can be decomposed into stars.

by faces gives the formula when w is a face. □

Theorem 2.1.10 (Riemann-Hurwitz Formula). *For a morphism $f: M \to N$ of degree n of finite oriented maps we have*

$$\chi(M) = n.\chi(N) + \sum_{v\in V(M)\cup F(M)} (e_v - 1)$$

PROOF. We know from Lemma 2.1.8 that $|A(M)| = n|A(N)|$. Also, by the Index formula, we have for every vertex w of N that $n = \sum_{f(v)=w} e_v = \sum_{f(v)=w}(e_v-1)+|f^{-1}(w)|$. So,

$$\begin{aligned}
\chi(M) &= |V(M)| - |A(M)| + |F(M)| \\
&= \sum_{w\in V(N)} |f^{-1}(w)| - n|A(N)| + \sum_{w\in F(N)} |f^{-1}(w)| \\
&= \sum_{w\in V(N)\cup F(N)} \left(n - \sum_{f(v)=w} (e_v - 1) \right) - n|A(N)| \\
&= n(|V(N)| + |F(N)|) - n|A(N)| + \sum_{v\in V(M)\cup F(M)} (e_v - 1)
\end{aligned}$$

□

The canonical morphism of a map and its monodromy group. In order to complete the parallel between combinatorial maps and Riemann surfaces we shall define the combinatorial counterpart of a Belyi function, that is of a function to the Riemann sphere with at most three ramification values. However, to be defined properly this combinatorial counterpart requires to allow the opposite operator to have fixed points. Hence, an oriented map becomes a triple $M=(A,\rho,\iota)$ where, as before, ρ and ι are permutations of A with ι an involution, but ι may now fix some arcs. If an arc $a\in A$ is fixed by ι it corresponds in the graph $G(M)$ to an edge having only one of its two endpoints considered as a vertex. The other endpoint is called a *free end*. The simplest map shown on Figure 2.5 has one arc and corresponds to the Riemann sphere. It now

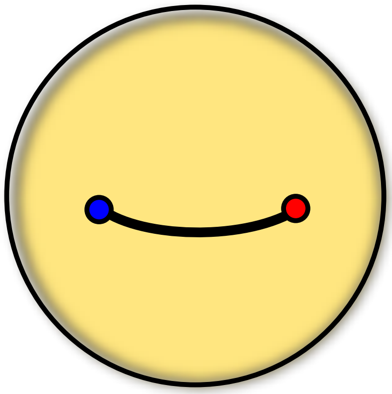


Figure 2.5: The trivial map $(\{a\},Id,Id)$ has one edge, one face, one (blue) vertex and one (red) free end.

appears that for any map $M=(A,\rho,\iota)$ there is a **canonical morphism** to the **trivial map** $(\{a\},Id,Id)$ given by the constant function $A\to\{a\}$. It corresponds to a branched covering of the sphere that ramifies above its vertex, the center of its face *and* its free end. Identifying free ends with their arc, we see that the above correspondence between topological and combinatorial maps must take into account ramifications at edges in addition to vertices and faces [JS78]. The formalisms of constellations and *hypermaps* permit us to avoid those singular free ends. Those formalisms are sketched in the next section for completeness. However, as far as algorithms on curves on surfaces are concerned the point of view of rotation systems seems more adequate and more intuitive.

Given a topological covering, $f:(S,y)\to(B,x)$ there is a right action of $\pi_1(B,x)$ on the fiber $f^{-1}(x)$ obtained in the same way as for graph coverings in Section 1.7.1 by lifting a loop representative of a homotopy class. The representation of $\pi_1(B,x)$ as a subgroup of permutations of $f^{-1}(x)$ is called the *monodromy group* of the covering. Changing the basepoint produces an isomorphic action, so that the monodromy group is well-defined up to isomorphism. When f is a branched covering, we can still define its monodromy group by considering the restriction $f:S\setminus f^{-1}(C)\to B\setminus C$, where C is the set of branch values (also called critical values or ramification points) of f. This restriction is indeed a (unbranched) covering on which acts the fundamental group of $B\setminus C$.

We can now consider the monodromy group of the branched covering corresponding to the topological realization of the canonical morphism of a combinatorial surface

M. This branched covering has the form $S(M) \to \mathbb{S}^2$ whose set C of branch values contains the two endpoints of (the embedding of) the unique edge of the trivial map and the center of its unique face. Hence, $\mathbb{S}^2 \setminus C$ is a sphere with three punctured, i.e., a **pair of pants**. Its fundamental group is a free group of rank 2 generated by two loops λ, μ, each surrounding one of the edge endpoints (see Figure 2.6). If we connect the chosen

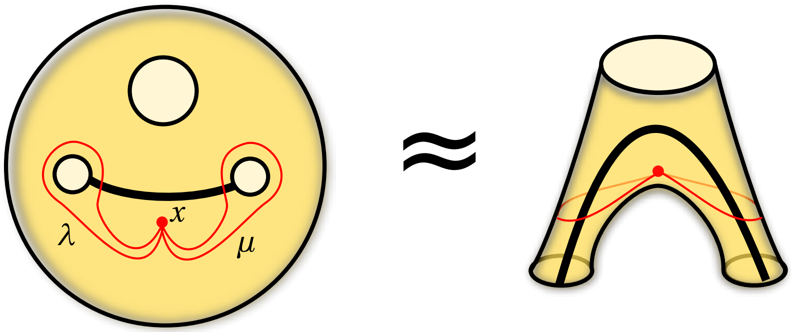


Figure 2.6: Two views of a pair of pants with two loops generating its fundamental group. The loop λ surrounds the vertex of the embedded edge and the loop μ surrounds its free end.

basepoint x to a point p on the unique edge of the punctured sphere by a path, then the lifts of this path establish a correspondence between the fiber of x and the fiber of p. In turn, the points of this fiber can be identified with the set of arcs of $M = (A, \rho, \iota)$ they lie on. Indeed, the fiber of the unique arc a of the trivial map by the canonical morphism is the whole set A. Because each lift of λ crosses exactly one arc, the action of λ on a point in the fiber of x, identified with an arc e of M, corresponds to a rotation of e about its origin. This action thus corresponds to the rotation system ρ. Similarly, the action of μ corresponds to the involution ι. The action of the whole group $\pi_1(\mathbb{S}^2 \setminus C, x) = \langle \lambda, \mu \rangle$ thus corresponds to the monodromy group $\langle \rho, \iota \rangle$, whence the terminology.

2.1.1 Constellations and hypermaps

We briefly mention the notion of constellation as presented by Lando and Zvonkin [LZ04]. This is definitely not the subject of my notes and far beyond my capacities. But it might be interesting for the reader to make the connection with the usual notion of map.

Definition 2.1.11. A **constellation** is a finite sequence of permutations $(g_1, \ldots, g_k)$ acting transitively on a finite set $\{1, \ldots, n\}$ and such that the product $g_1 \cdots g_k$ is the identity permutation.

This algebraico-combinatorial object can be interpreted as an n-fold branched covering of the Riemann sphere with k branch values indexed by $1, \ldots, k$. The ramification indices of the branch points above the branch value of index i are given by the length of the cycles of the permutation g_i. Given a constellation $(g_1, \ldots, g_k)$ the construction of this branched covering can be performed as follows. We consider a set C of k punctures in the oriented sphere $\mathbb{S}^2$ and a basepoint $x \in \mathbb{S}^2 \setminus C$. We draw a star graph in $\mathbb{S}^2$ connecting x to each point in C (see Figure 2.7). We obtain a generating

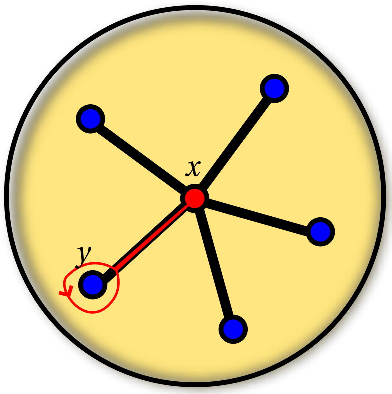


Figure 2.7: A sphere with five (blue) punctures. The loop γ_y with basepoint x surrounds the puncture y.

set for $\pi_1(\mathbb{S}^2 \setminus C, x)$ by forming a loop γ_y for each $y \in C$; this loop follows the edge xy in the star graph and stops just before reaching y, goes around y in the counterclockwise direction and travels back to x. The product of the γ_y in the counterclockwise order of the star edges is clearly contractible. Letting $y_1, \ldots, y_k$ be the points of C in *clockwise* order, with thus have a presentation $\langle [\gamma_{y_1}], \ldots, [\gamma_{y_k}]; [\gamma_{y_1}]^{-1} \cdots [\gamma_{y_k}]^{-1} = 1\rangle$ for $\pi_1(\mathbb{S}^2 \setminus C, x)$. Since the unique relations of the $[\gamma_{y_i}]$ is satisfied by the g_i^{-1}, the map $[\gamma_{y_i}] \mapsto g_i^{-1}$ induces a group morphism $\phi : \pi_1(\mathbb{S}^2 \setminus C, x) \to G$ where $G = \langle g_1, g_2, \ldots, g_k\rangle$ is the monodromy group of the constellation. Let $U = \{g \in G; g(1) = 1\}$ be the stabilizer of 1 (recall that G acts on $\{1, \ldots, n\}$). Similarly to Proposition 1.7.7 for graphs, the preimage $\phi^{-1}(U) < \pi_1(\mathbb{S}^2 \setminus C, x)$ determines a covering $p_U : S_U \to \mathbb{S}^2 \setminus C$ whose fiber elements are the right cosets of $\phi^{-1}(U)$ in $\pi_1(\mathbb{S}^2 \setminus C, x)$. The set of right cosets, as preimages of the map $\pi_1(\mathbb{S}^2 \setminus C, x) \to \{1, \ldots, n\}, \alpha \mapsto \phi(\alpha^{-1})(1)$, are in bijection with $\{1, \ldots, n\}$; note that this map is onto because G acts transitively on $\{1, \ldots, n\}$. We can check that the monodromy group of p_U is given by the constellation: as in Proposition 1.7.7, the action of $\pi_1(\mathbb{S}^2 \setminus C, x)$ on the fiber above x is given by $(x, \phi^{-1}(U)\alpha).[\gamma_i] = (x, \phi^{-1}(U)\alpha[\gamma_i])$. Using the correspondence $(x, \phi^{-1}(U)\alpha) \mapsto \phi(\alpha^{-1})(1)$, the action of $[\gamma_i]$ on $p_U^{-1}(x)$ transforms to the permutation g_i.

We can finally compactify S_U and $\mathbb{S}^2 \setminus C$ to extend p_u to a branched covering $\bar{S}_U \to \mathbb{S}^2$. To this end, we consider small punctured discs D_y^* centered at each puncture $y \in C$ and note that the restriction $p_U : p_U^{-1}(D_y^*) \to D_y^*$ being a covering of finite degree, $p_U^{-1}(D_y^*)$ must be a disjoint union of punctured discs. We formally add a center to those punctured discs and extend p_U trivially by sending the added disc centers to y. We obtain this way a branched covering of compact surfaces whose ramification indices are the cycle lengths of the g_i. Note that the genus of this covering is given by an adapted version of the Riemann-Hurwitz formula of theorem 2.1.10. This branched covering comes with a cellular embedding of a graph obtained by lifting the star graph on $\mathbb{S}^2$. This graph is bipartite, the partition being given by the fiber of the basepoint on the one side and the union of the fibers of the branch values on the other side. It can be seen as a union of stars of degree k centered at the vertices in the basepoint fiber. This explains the name of constellations.

Hypermaps. Other cellular embeddings can be obtained starting with a different graph on the sphere. One possibility is to start with a chain graph going trough the branch values $y_1, y_2, \dots, y_{k-1}$, leaving the point y_k aside. In this case, the covering has branch points at vertices and at the center of faces (the points above y_k). When $k=3$, that is when we start with a 3-constellation (g_1, g_2, g_3), we obtain a single edge on the sphere as for the trivial map on Figure 2.5. The partition corresponding to the two fibers of y_1 and y_2 again make the lifted graph bipartite. If we further impose that g_2 is a fixed point free involution, then y_2 lifts to degree two vertices. Viewing the edges as arcs oriented from (lifts of) y_1 to y_2 we get exactly the same picture as for the canonical morphism of a map, where the lifts of y_2 become the middle-point of the edges. Formally we have an identification of the combinatorial maps as 3-constellations given by the correspondence $(A, \rho, \iota) \mapsto (\rho, \iota, (\rho \circ \iota)^{-1})$, where the set A of arcs should be identified with $\{1, \dots, n\}$. We can thus identify the fibers of y_1, y_2 and y_3 as vertices, edges and faces respectively. If we still keep this terminology for general 3-constellations, where g_2 may be any permutation, then each "edge" in the fiber of y_2 becomes incident to possibly more than two vertices. The lifted graph can thus be interpreted as a hypergraph (by the way, a hypergraph, or set system, is just another name for a bipartite graph). This is why 3-constellations are usually called **hypermaps**.

Intrinsic algebraic formalism Given a connected combinatorial map (A, ρ, ι), its monodromy group $\langle \rho, \iota \rangle$ acts transitively on the set A of arcs that can thus be identified with the left cosets of the stabilizer $S_a = \{\tau \in \langle \rho, \iota \rangle \mid \tau(a) = a\}$ of some fixed arc $a \in A$. Indeed, it is easily seen that the correspondence $\langle \rho, \iota \rangle / S_a \to A$ given by $\tau S_a \mapsto \tau(a)$ is well defined and one-to-one. To obtain an isomorphic action of the monodromy group, one should consider its left action on the left cosets $\langle \rho, \iota \rangle / S_a$. A map can thus be represented by a (monodromy) group Γ, a (stabilizer) subgroup S, and two generators ρ, ι of Γ such that $\iota^2 = 1$. However, to represent a map, one should make sure that Γ acts faithfully on Γ/S. Indeed, since the monodromy group was originally defined as a subgroup of permutations of A it acts faithfully, meaning that each element of the monodromy group is uniquely determined by its action. In general, if we are given (Γ, S, ρ, ι) as above there is no reason why Γ should act faithfully on Γ/S. Note that for $h, g \in \Gamma$, having $h(gS) = gS$ is equivalent to $h \in gSg^{-1}$. So, any h in the intersection of the conjugate subgroups of S acts as the identity on Γ/S. This intersection is the largest normal subgroup of Γ contained in S and is usually denoted by $core_\Gamma(S)$. Hence, $core_\Gamma(S)$ should be trivial if we want Γ to act faithfully. When this is not the case this can be enforced by considering the action of $\Gamma/core_\Gamma(S)$ on Γ/S with the condition $\iota^2 \in core_\Gamma(S)$. See [BS85] and [BW09, Ch. 10] for further details.

We close this section by noting that constellations should not be compared with combinatorial maps but rather with map morphisms to maps of genus zero endowed with fixed embedded graphs (like star graphs). The two formalisms are equivalent but thanks to its symmetry the formalism of constellations is much more powerful when dealing with algebraic properties. However, the graph embedding we can associate to a constellation is not really encoded in the constellation as it depends on the graph drawn on the base sphere. When dealing with combinatorial curves on surfaces, the embedded graph itself becomes the main object of study and combinatorial maps provide this graph more directly.

2.1.2 Basic operations on oriented maps

One advantage of the map formalism is the ability to modify the embedded graph rather easily. A notion of elementary modification gives rise to combinatorial equivalence between maps that shall replace topological homeomorphisms and allows for a classification of surfaces. This classification is detailed in the more general framework of non necessarily oriented maps. We continue to assume that all maps are oriented in this section.

Dual maps

Intuitively the dual of a map is obtained by inverting the roles of vertices and faces. The topological counterpart of the dual map is obtained by placing a (dual) vertex at the center of each face of the (primal) map, adding an edge between two dual vertices if their corresponding face share an edge. Figure 2.8 illustrates the dual of a spherical map.

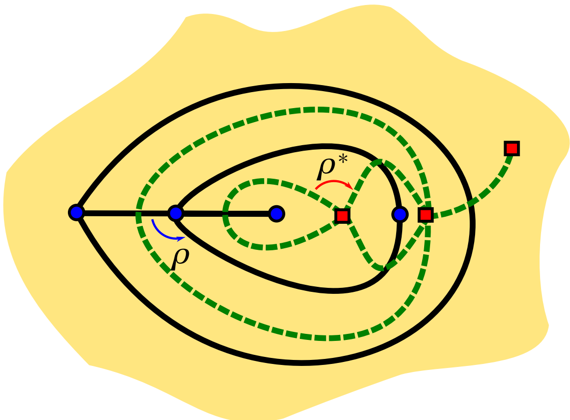


Figure 2.8: A map on the sphere (with plain line edges) with rotation system ρ and its dual map (with dashed line edges).

Definition 2.1.12. The **dual** of the map $M = (A, \rho, \iota)$ is the map $M^* = (A, \rho \circ \iota, \iota)$. The **dual graph** of M is the graph $G^*(M) = G(M^*)$ of the dual map. The vertices of the dual graph are the cycles of $\rho \circ \iota$, i.e., the faces of M. More precisely, $G^*(M) = (F(M), A, o^*, \iota)$ where $o^*(a) = F(a)$.

It is immediate that

Lemma 2.1.13. *M and M^* have the same monodromy group. In particular, M is connected if and only if M^* is connected.*

Lemma 2.1.14. $(M^*)^* = M$

Edge contraction

Definition 2.1.15. Let $M = (A, \rho, \iota)$ be a map with at least two edges. If $e = \{a, a^{-1}\}$ is an edge of M, the **contraction** of e in M transforms M to a map $M/e = (A \setminus e, \rho', \iota')$ where ι' is the restriction of ι to $A \setminus e$ and ρ' is obtained by merging the cycles of a and a^{-1}, i.e.,

- if $o(a) \neq o(a^{-1})$

$$\forall b \in A \setminus e, \rho'(b) = \begin{cases} \rho(b) & \text{if } \rho(b) \notin e, \\ \rho \circ \iota(\rho(b)) & \text{if } \rho(b) \in e \text{ and } \rho \circ \iota(\rho(b)) \notin e, \\ (\rho \circ \iota)^2(\rho(b)) & \text{otherwise.} \end{cases}$$

- if $o(a) = o(a^{-1})$

$$\forall b \in A \setminus e, \rho'(b) = \begin{cases} \rho(b) & \text{if } \rho(b) \notin e, \\ \rho^2(b) & \text{if } \rho(b) \in e \text{ and } \rho^2(b) \notin e, \\ \rho^3(b) & \text{otherwise.} \end{cases}$$

Figure 2.9 shows the effect of an edge contraction in the simple case where e has distinct endpoints of degree at least two. The contraction of a loop edge is illustrated on

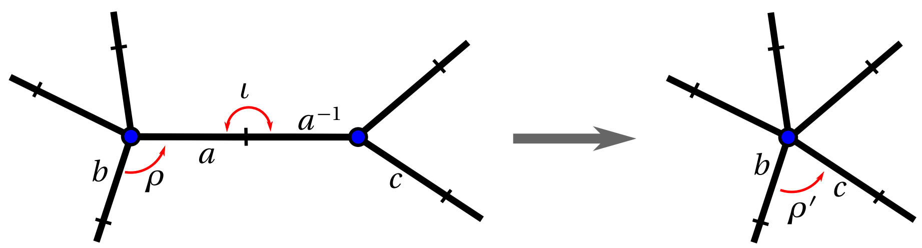


Figure 2.9: The contraction of a non-loop edge. $\rho(b) = a \implies \rho'(b) = \rho \circ \iota(\rho(b)) = c$.

Figure 2.10. Note that the contraction of a loop edge is the same as its removal.

Exercise 2.1.16. Show that $G(M/e) = G(M)/e$ (see Definition 1.1.4).

Lemma 2.1.17. *If M is a connected map with at least two edges and $e = \{a, a^{-1}\}$ is an edge of M then M/e is connected and*

$$\chi(M/e) = \begin{cases} \chi(M) & \textit{if } e \textit{ has distinct endpoints, or if } F(a) \neq F(a^{-1}) \\ \chi(M) + 2 & \textit{otherwise.} \end{cases}$$

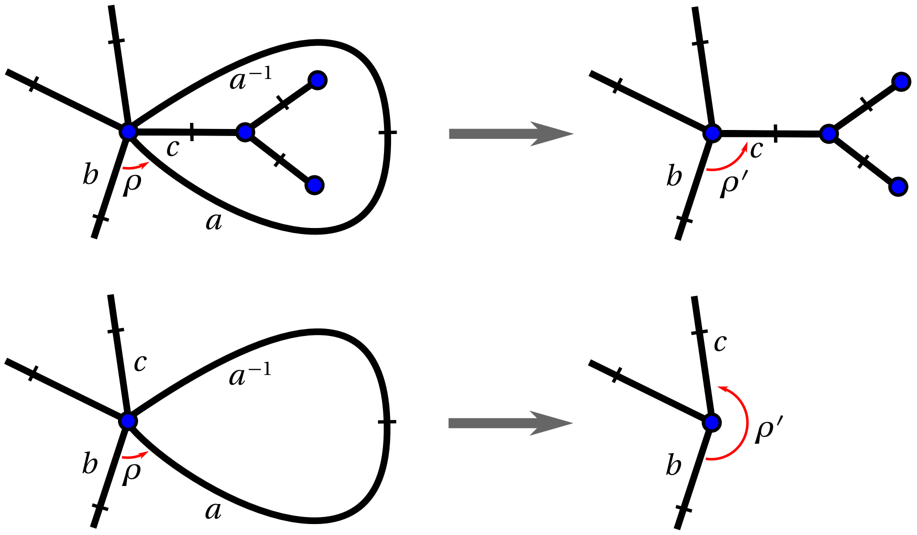


Figure 2.10: The contraction of a loop edge with an arc b such that $\rho(b) \in \{a, a^{-1}\}$. Above, We have $\rho^2(b) \notin \{a, a^{-1}\}$ implying $\rho'(b) = \rho^2(b) = c$. Below, $\rho^2(b) \in \{a, a^{-1}\}$ so that $\rho'(b) = \rho^3(b) = c$.

Proof. Put $M = (A, \rho, \iota)$ and $M/e = (A', \rho', \iota)$. If $o(a) \neq o(a^{-1})$, then we have by definition that

$$\rho' \circ \iota(b) = \begin{cases} \rho \circ \iota(b) & \text{if } \rho \circ \iota(b) \notin e, \\ (\rho \circ \iota)^2(b) & \text{if } \rho \circ \iota(b) \in e \text{ and } (\rho \circ \iota)^2(b) \notin e, \\ (\rho \circ \iota)^3(b) = & \text{otherwise.} \end{cases}$$

The faces of M/e are thus obtained by deleting a and a^{-1} from the faces of M. Since no face is reduced to the singleton a or a^{-1}, as e would be a loop edge otherwise, it follows that $|F(M/e)| = |F(M)|$. On the other hand, we have $|V(M/e)| = |V(M)| - 1$ by Exercise 2.1.16. We thus have

$$\chi(M/e) = |V(M/e)| - (|A|/2 - 1) + |F(M/e)| = |V(M)| - |A|/2 + |F(M)| = \chi(M)$$

If on the contrary $o(a) = o(a^{-1})$, then $|V(M/e)| = |V(M)|$ by Exercise 2.1.16. Moreover, assuming $F(a) \neq F(a^{-1})$, the two corresponding cycles $(a, \rho \circ \iota(a), \ldots, b)$ and $(a^{-1}, \rho \circ \iota(a^{-1}), \ldots, d)$ are merged by $\rho' \circ \iota$ into a single cycle $(\rho \circ \iota(a), \ldots, b, \rho \circ \iota(a^{-1}), \ldots, d)$. For instance, we check that for $b \neq a^{-1}$, we have $\rho' \circ \iota(b) = \rho(\rho(\iota(b)) = \rho(a) = \rho \circ \iota(a^{-1})$ and for $d \neq a$ we have $\rho' \circ \iota(d) = \rho(a^{-1})$. The other faces are left unchanged so that $|F(M/e)| = |F(M)| - 1$ and $\chi(M/e) = |V(M)| - (|A|/2 - 1) + |F(M)| - 1 = \chi(M)$. Finally, if $o(a) = o(a^{-1})$ and $F(a) = F(a^{-1})$, we check that the corresponding face cycle is split in M/e while the other are left unchanged (see Figure 2.11). We conclude that $|F(M/e)| = |F(M)| + 1$ and $\chi(M/e) = \chi(M) + 2$. $\square$

Edge deletion

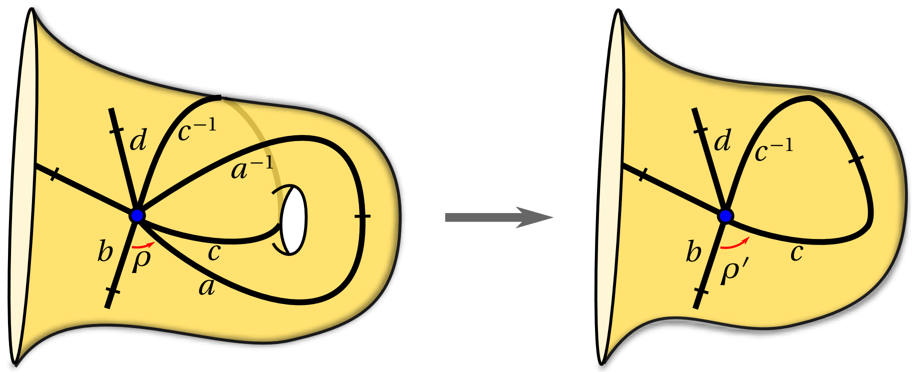


Figure 2.11: The contraction of the loop edge $\{a, a^{-1}\}$ replaces the facial cycle $(\dots, b^{-1}, a, c^{-1}, a^{-1}, c, d, \dots)$ by the two cycles $(\dots, b^{-1}, c, d, \dots)$ and (c^{-1}).

Definition 2.1.18. Let $M = (A, \rho, \iota)$ be a map with at least two edges. If $e = \{a, a^{-1}\}$ is an edge of M, the **deletion** of e in M transforms M to a map $M - e = (A \setminus e, \rho', \iota')$ where ι' is the restriction of ι to $A \setminus e$ and ρ' is obtained by deleting a and a^{-1} in the cycles of ρ, i.e.,

$$\forall b \in A \setminus e, \rho'(b) = \begin{cases} \rho(b) & \text{if } \rho(b) \notin e, \\ \rho^2(b) & \text{if } \rho(b) \in e \text{ and } \rho^2(b) \notin e, \\ \rho^3(b) & \text{otherwise.} \end{cases}$$

Observe that $G(M - e) = G(M) - e$ (see Definition 1.1.5).

Lemma 2.1.19. *If M is a connected map with at least two edges and $e = \{a, a^{-1}\}$ is an edge of M, then*

$$\chi(M - e) = \begin{cases} \chi(M) & \textit{if } e \textit{ has a degree one endpoint, or if } F(a) \neq F(a^{-1}) \\ \chi(M) + 2 & \textit{else.} \end{cases}$$

Note that the deletion of e may disconnect the map.

PROOF. First suppose that e has no degree one endpoint. Then $|V(M - e)| = |V(M)|$. We note that ρ' is defined the same way as for a loop edge contraction. Following the proof of Lemma 2.1.17, we thus have $|F(M - e)| = |F(M)| - 1$ if $F(a) \neq F(a^{-1})$ and $|F(M - e)| = |F(M)| + 1$ otherwise. It easily follows that $\chi(M - e) = \chi(M)$ in the first case while $\chi(M - e) = \chi(M)$ in the second case.

If e has a degree one endpoint, then $F(a) = F(a^{-1})$ and a belongs to a cycle $(a, a^{-1}, b, \dots)$ giving the cycle $(b, \dots)$ after the deletion of e. The other cycles are unchanged so that $|F(M - e)| = |F(M)|$. Since $|V(M - e)| = |V(M)| - 1$, we conclude $\chi(M - e) = \chi(M)$. □

We leave as an exercise, the following link between edge contraction and deletion.

Lemma 2.1.20. *Let e be an edge of a connected map M with at least two edges. Then, $(M/e)^* = M^* - e$ and $(M-e)^* = M^*/e$.*

Edge subdivision

Definition 2.1.21. Let $e = \{a, a^{-1}\}$ be an edge of a map $M = (A, \rho, \iota)$. The **subdivision** of e in M transforms M to a map $S_e M = (A', \rho', \iota')$ where

- $A' = A \cup \{b, b'\}$, where b, b' are new arcs not in A,
- the restriction of ι' to A is equal to ι and $\iota'(b) = b'$,
- ρ' is defined by

$$\forall c \in A',\ \rho'(c) = \begin{cases} b & \text{if } c = a^{-1} \\ a^{-1} & \text{if } c = b \\ \rho(a^{-1}) & \text{if } c = b' \\ b' & \text{if } c = \rho^{-1}(a^{-1}) \\ \rho(c) & \text{otherwise.} \end{cases}$$

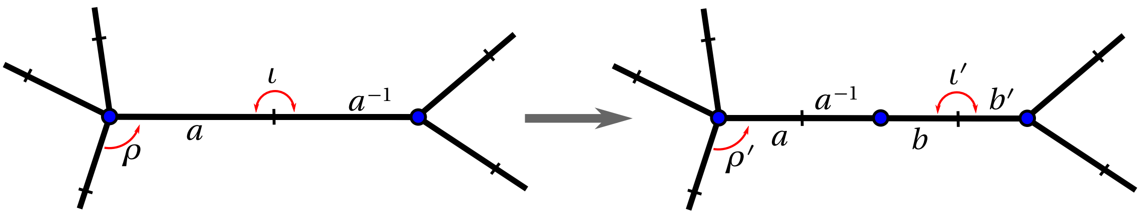


Figure 2.12: The subdivision of an edge splits that edge, introducing a new vertex on the edge.

We observe that $G(S_e M) = S_e G(M)$ (see Definition 1.1.6) and we trivially check that the edge subdivision preserves the number of connected components and the Euler characteristic.

Face subdivision

Definition 2.1.22. Let $M = (A, \rho, \iota)$ be a map and let a, b be two arcs, possibly equal, belonging to a same face $F(a) = F(b)$. The **subdivision** of $F(a)$ from a to b transforms M to a map $S_{(a,b)} M = (A \cup \{c, c^{-1}\}, \rho', \iota)$ obtained by adding a new edge $\{c, c^{-1}\}$ in $F(a)$ between the heads of a and b (see Figure 2.13). When $a \neq b$ the new rotation system ρ' is given by

$$\forall d \in A \cup \{c, c^{-1}\},\ \rho'(d) = \begin{cases} c & \text{if } d = a^{-1} \\ c^{-1} & \text{if } d = b^{-1} \\ \rho(a^{-1}) & \text{if } d = c \\ \rho(b^{-1}) & \text{if } d = c^{-1} \\ \rho(d) & \text{otherwise.} \end{cases}$$

When $a = b$, ρ' is given by

$$\forall d \in A \cup \{c, c^{-1}\}, \rho'(d) = \begin{cases} c & \text{if } d = a^{-1} \\ c^{-1} & \text{if } d = c \\ \rho(a^{-1}) & \text{if } d = c^{-1} \\ \rho(d) & \text{otherwise.} \end{cases}$$

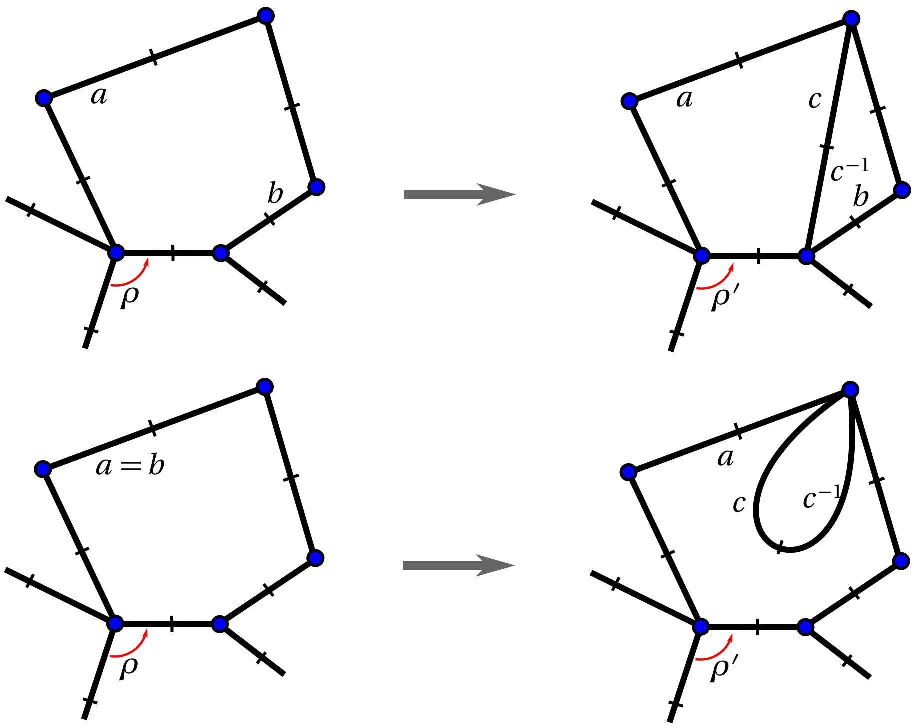


Figure 2.13: The subdivision of face $F(a)$ between the heads of a and b. The case $a = b$ is shown below.

We trivially check that the edge subdivision preserves the number of connected components and the Euler characteristic.

Remark 2.1.23. The inverse of an edge subdivision amounts to "remove" a degree two vertex, merging its incident edges. The inverse of a face subdivision is an edge deletion, where the edge is incident to two distinct faces. Zieschang et al. [ZVC80, p. 67] observe that the contraction of a non-loop edge can be obtained from a sequence of edge or face subdivisions and their inverses. Figure 2.14 illustrates the process of contracting an edge in this way.

Exercise 2.1.24. The sequence of operations in Figure 2.14 is still valid when the right endpoint on the figure has degree one, i.e., when the edge is a pendant edge. Propose a simpler sequence of face or edge contractions (and their inverses) equivalent to an edge contraction in that case.

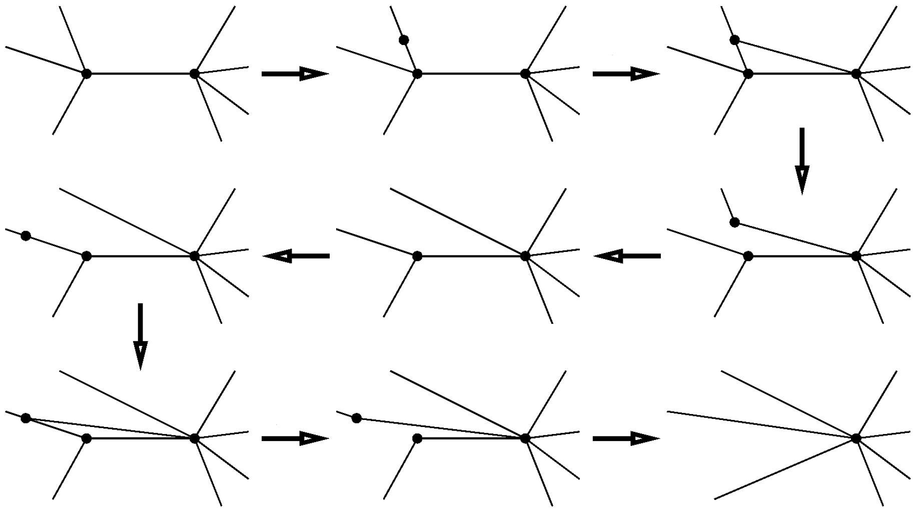


Figure 2.14: Let us denote by S_e and S_f respectively, an edge and face subdivision and by S_e^{-1} and S_f^{-1} the corresponding inverse operations. The contraction of the upper left edge is the result of the sequence of operations: $S_e, S_f, S_f^{-1}, S_a^{-1}, S_a, S_f, S_f^{-1}, S_a^{-1}$.

2.2 General Maps

We now describe the notion of combinatorial surface, orientable or not. We follow the formalism of Mohar and Thomassen [MT01, Ch.4.] [Moh01].

Definition 2.2.1. A **combinatorial map**, or simply a **map**, is a quadruple $M = (A, \rho, \iota, s)$ where

- A is a non-empty set whose elements are called **arcs**,
- $\rho : A \to A$ is a permutation of A,
- $\iota : A \to A$ is a fixed point free involution,
- $s : A \to \{-1, 1\}$ is a **signature** satisfying $s(a) = s(a^{-1})$. Equivalently, s is defined over the set of edges $\{a, a^{-1}\}$.

A **connected component** of M is the restriction (ρ, ι, s) to an orbit of $\langle \rho, \iota \rangle$ acting on A. The associated **graph** $G(M)$ is defined in the same way as for oriented maps, its vertices being the cycles of ρ. A map is **connected** if its graph is.

A **flag**, or **dart**, of M is a signed arc, i.e., an element of $A \times \{-1, 1\}$. A flag (a, ϵ) has **arc component** $p_A(a, \epsilon) := a$ and **sign** $sign(a, \epsilon) := \epsilon$. A **facial permutation** φ and an involution α_0 are defined over the set of flags by

$$\forall (a, \epsilon) \in A \times \{-1, 1\}, \qquad \varphi(a, \epsilon) = (\rho^{\epsilon s(a)}(a^{-1}), \epsilon s(a)), \quad \text{and} \tag{2.1}$$

$$\alpha_0(a, \epsilon) = (a^{-1}, -\epsilon s(a)) \tag{2.2}$$

Notice that every oriented map (A, ρ, ι) identifies with the combinatorial map $(A, \rho, \iota, 1)$.

Remark 2.2.2. The condition that the set of arcs should be non-empty is essentially a matter of convention. We could as well consider the empty combinatorial map to be a sphere tessellated with a single vertex. However, since there cannot be any mapping to the empty set, the definition would be of little use.

Exercise 2.2.3. Check that φ is indeed a permutation of $A \times \{-1, 1\}$.

Map realization

To every cellular embedding of a graph in a topological surface, we can associate a combinatorial map (A, ρ, ι, s) as follows. We first choose a local surface orientation at each vertex of the graph and mark each half-edge (cutting each edge in the middle) with a distinct label. We let A be the set of labels. The circular orders of the half-edge labels around each vertex of the graph, *turning in the direction of the chosen orientation*, determine the cycles of ρ. The involution ι exchanges the two half-edges of an edge and the signature s of an edge is chosen positive whenever the orientation at one of its endpoints coincides with the orientation at the other endpoint when transported along that edge. It is negative otherwise. This signature applies to non-loop as well as loop edges.

Conversely, we can construct a cellular embedding from a combinatorial map $M = (A, \rho, \iota, s)$ such that the map induced by the embedding coincides with M. The construction starts with a set of positively oriented discs in the oriented xy-plane of $\mathbb{R}^3$. Those discs are pairwise disjoint, with one disc per vertex of M. We then attach rectangular strips to the discs, with one strip per edge. The strips expand in $\mathbb{R}^3$ so that they do not intersect. The direct ordering of the strips attached to a discs should coincide with the cycle of ρ defining the corresponding vertex. Moreover, each strip is applied a half-twist whenever its signature is negative. See Figure 2.15 for an illustration. The boundaries of the resulting thickened graph are finally closed with discs. The graph

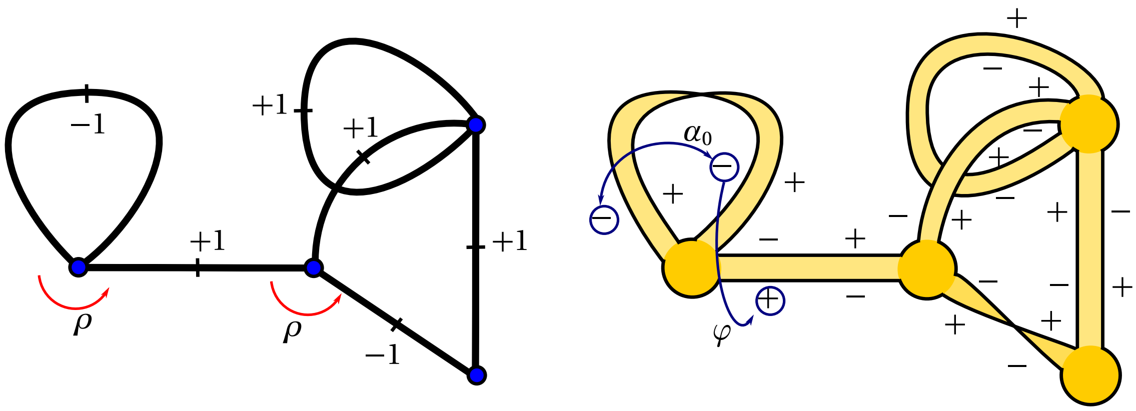


Figure 2.15: Left, a schematic representation of a map. Right, Its topological realization starts with a graph thickening. The strips are labelled with the signs of the corresponding flags.

$G(M)$ embeds in this surface by placing a vertex at the center of each disc and drawing each edge between such centers in its corresponding strip.

Using this correspondence, a flag corresponds to a signed half-side of a strip. The facial permutation transforms a flag (a, ϵ) into the next half-side on the next edge, following the boundary of (a,ϵ) in the direction of a, assuming that a points towards $o(a^{-1})$ (see Figure 2.15). The involution α_0 exchange a flag with the other half-side of its side.

Map faces

Definition 2.2.4. An **oriented face** of a map is a cycle of its facial permutation φ. We denote by $F(a,\epsilon)$ the φ-cycle of the flag (a,ϵ). The **degree** $\deg(f)$ of an oriented face f is its cycle length. We denote the **facial circuit** of f by ∂f; this is the circuit of $G(M)$ obtained by listing the arc components of the flags in f in cyclic order. Any arc of ∂f is said **incident** to f.

Lemma 2.2.5. *The correspondence $F(a,\epsilon)\mapsto F(\alpha_0(a,\epsilon))$ defines a fixed point free involution on the set of oriented faces of a map. The (oriented) face $F(\alpha_0(a,\epsilon))$ is called the* ***opposite*** *of the face $F(a,\epsilon)$ and is denoted by $F^{-1}(a,\epsilon)$. The facial circuits of a face and of its opposite face are inverse circuits: $\partial F^{-1}(a,\epsilon)=(\partial F(a,\epsilon))^{-1}$.*

Proof. Because α_0 is an involution, the opposite of a face defines an involution. We should verify that for any flag (a,ϵ), we have $F(\alpha_0(a,\epsilon))\neq F(a,\epsilon)$. In other words, the orbits of (a,ϵ) and $(a^{-1},-\epsilon s(a))$ under the action of $\langle\varphi\rangle$ must be disjoint. Suppose, on the contrary, that $\varphi^k(a,\epsilon)=(a^{-1},-\epsilon s(a))$ for some $k\geq 0$. Let us choose (a,ϵ) and k such that k is minimal with this property. Since $a\neq a^{-1}$, we cannot have $k=0$. We can neither have $k=1$ since $\varphi(a,\epsilon)$ and $(a^{-1},-\epsilon s(a))$ have opposite signs. Hence $k-2\geq 0$. Put $(b,\eta)=\varphi(a,\epsilon)$. We have

$$\varphi^{k-2}(b,\eta)=\varphi^{k-1}(a,\epsilon)=\varphi^{-1}(a^{-1},-\epsilon s(a))=(b^{-1},-\eta s(b)),$$

where we used that $\varphi^{-1}\circ\alpha_0=\alpha_0\circ\varphi$ for the last equality. But this contradicts the minimality of k.

From the equality $\varphi^{-1}\circ\alpha_0=\alpha_0\circ\varphi$ we deduce $\phi^j\circ\alpha_0=\alpha_0\circ\phi^{-j}$ for all j and conclude that $\partial F(\alpha_0(a,\epsilon))=(\partial F(a,\epsilon))^{-1}$ after projecting the orbit $\langle\varphi\rangle(\alpha_0(a,\epsilon))$ on its first coordinate. $\square$

Definition 2.2.6. A **face** is a pair of opposite oriented faces. We say that an edge $\{a,a^{-1}\}$ is **incident** to the possibly equal faces $\{F(a,1),F^{-1}(a,1)\}$ and $\{F(a,-1),F^{-1}(a,-1)\}$. An edge that is incident to two distinct faces is said **regular** and **singular** otherwise.

Example 2.2.7. There are two maps with one edge and one vertex as depicted on Figure 2.16.

The **degree of a face** f, denoted $\deg(f)$ is the common size of its two oriented faces or, equivalently of their facial circuit.

Exercise 2.2.8. A **bridge** of a graph is an edge whose deletion increases the number of connected components of the graph. Show that a bridge of the graph $G(M)$ of a map M must be singular.

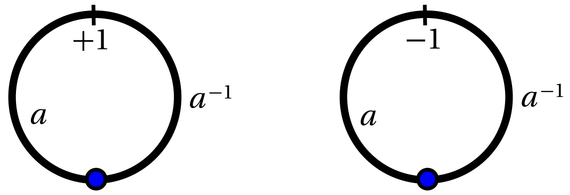


Figure 2.16: Left, The oriented map with one loop edge embeds cellularly into a sphere with two faces $\{F(a,1), F(a^{-1},-1)\} \neq \{F(a,-1), F(a^{-1},-1)\}$. Right, The non-orientable map with one loop edge embeds cellularly into a projective plane with one face $\{F(a,1), F(a^{-1},-1)\} = \{F(a,-1), F(a^{-1},1)\}$.

2.2.1 Orientation

The formalism of rotation systems for unoriented surfaces has a little drawback. A rotation system relies on a local orientation of every vertex star but for non-orientable surfaces such local orientations cannot be defined in a canonical way. If we refer to the graph thickening stage in the above map realization, we see that we could flip any of the discs and modify the rotation system and the signature accordingly to obtain another equivalent map as on Figure 2.17. Depending on the chosen orientations we

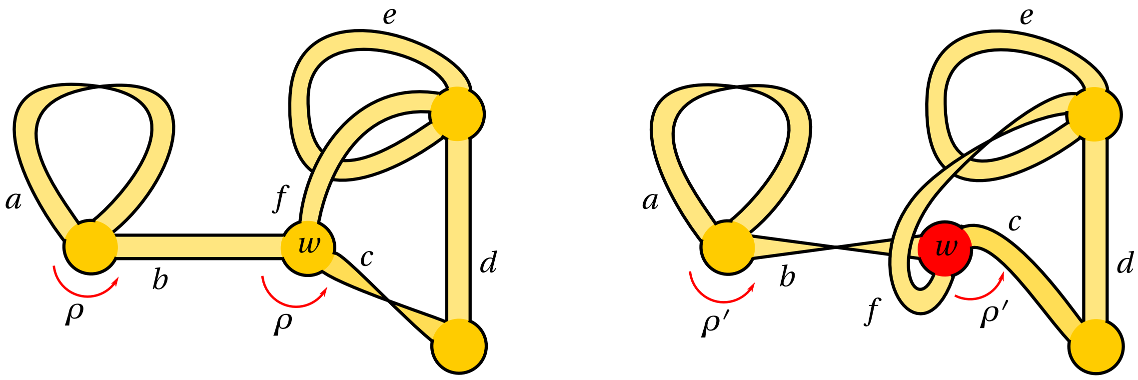


Figure 2.17: Flipping the disc of the vertex w changes the twist of the incident edges, hence their signature, and the direction of the corresponding cycle of ρ. Left, a graph thickening of the map with $\rho = (a,b,a^{-1})(b^{-1},c,f)(c^{-1},d)(d^{-1},e,f^{-1},e^{-1})$, $s(a,c) = -1$ and $s(b,d,e,f) = 1$. Right, After flipping w we get $\rho' = (a,b,a^{-1})(f,c,b^{-1})(c^{-1},d)(d^{-1},e,f^{-1},e^{-1})$, $s(a,b,f) = -1$ and $s(c,d,e) = 1$.

thus get several combinatorial representations of the same cellular graph embedding. We are thus led to consider maps *up to reorientation* obtained by flipping any subset of vertices.

Definition 2.2.9. Let $M = (A, \rho, \iota, s)$ be a map with vertex set V and let $\omega : V \to \{-1, 1\}$ be a “flip” function. The **reorientation** of M induced by ω is the map (A, ρ', ι, s') where for every arc a:

$$\rho'(a) = \rho^{\omega(o(a))}(a) \qquad \text{and} \qquad s'(a) = \omega(o(a))\omega(o(a^{-1}))s(a) \tag{2.3}$$

Exercise 2.2.10. Check that every connected map M with n vertices has 2^n distinct reorientations.

Exercise 2.2.11. Let M' be the map obtained by the reorientation of M induced by a flipping ω. Put $\lambda : (a,\epsilon) \mapsto (a, \epsilon\omega(o(a)))$. Show that the facial permutations of M and M' are conjugate by the involution λ.

It follows directly from the exercise that

Lemma 2.2.12. *The oriented faces of any reorientation of a map M are in correspondence with the oriented faces of M and this correspondence preserves facial circuits.*

Definition 2.2.13. A map is **orientable** if one of its reorientations has a constant positive signature. A map is **oriented** when its signature is 1.

In order to test if a map M is orientable we could check if one of the $2^{|V(M)|}$ reorientations has positive signature, but there is a more efficient test.

Lemma 2.2.14. *The orientability of a map $M = (A, \rho, \iota, s)$ can be determined in $O(|A|)$ time.*

Proof. Let T be a spanning tree of $G(M)$ with a chosen root vertex v. If M is orientable there exists a flipping $\omega : V(M) \to \{-1, 1\}$ such that the induced reoriented map has positive signature. Replacing ω by $-\omega$ if necessary we can assume that $\omega(v) = 1$. By induction on the distance to v in T, we see that ω is entirely determined by s: if e is the edge of T linking x to its parent y in T we must have $\omega(x) = \omega(y)s(e)$. It remains to check if the reoriented map induced by ω has positive signature, i.e., if all chords of T have signature one. In the affirmative we conclude that M is orientable, otherwise M is non-orientable. The whole computation trivially takes $O(|A|)$ time. $\square$

There are other characterizations of orientability.

Definition 2.2.15. A circuit of $G(M)$ is **two-sided** if the number of its arcs with negative signature is even. It is **one-sided** otherwise.

Exercise 2.2.16. Let M be a map, orientable or not. Show that the facial circuits of the oriented faces of M are two-sided.

Lemma 2.2.17. *A map is orientable if and only if all of its circuits are two-sided.*

Proof. Let M be a map. It is easily seen that a reorientation of M does not change the property of a circuit to be one or two-sided. It follows that the circuits of an orientable map must all be two-sided. Conversely, suppose that every circuit in M is two-sided. Choose a spanning tree T of M and a flipping inducing a positive signature on T. Then any chord a of T must also have positive signature since otherwise the circuit $T[a]$ would be one-sided. $\square$

Lemma 2.2.18. *A map M is orientable if and only if we can choose an orientation for each face of M so that each arc of M appears exactly once in the facial circuits of the chosen oriented faces.*

PROOF. Suppose that $M=(A,\rho,\iota,s)$ is orientable. By Lemma 2.2.12, we can reorient M so that its signature s is positive. The facial permutation then leaves invariant the sign of a flag. It follows that each face has one orientation whose flags all have positive signs while the flags of the other orientation have negative signs. Choosing all the oriented faces with a positive sign thus gives the direct implication in the lemma.

For the reverse implication, we now suppose that we can choose an orientation for each face with the property in the lemma. Hence, for each arc $a\in A$, there is exactly one flag d_a of the form (a,ϵ_a) such that $F(d_a)$ is a chosen oriented face. We define the *turning sense* $\tau(a)\in\{-1,1\}$ of a by

$$\varphi(d_a)=(\rho^{\tau(a)}(a^{-1}),\tau(a))$$

where φ is the facial permutation. In other words $\tau(a)=sign(d_a)s(a)$. We note that $\alpha_0(d_a)=(a^{-1},-\tau(a))$. We claim that all the arcs pointing to a same vertex have the same turning sense. Otherwise, there would be two arcs $a,b\in A$ satisfying

$$o(a^{-1})=o(b^{-1}),\quad \rho(a^{-1})=b^{-1},\quad \tau(a)=1\quad\text{and}\quad \tau(b)=-1$$

This would imply $\varphi(d_a)=(\rho(a^{-1}),1)=(b^{-1},1)=\alpha_0(d_b)$. Then $F(d_a)=F(\phi(d_a))=F(\alpha_0(d_b))$ contradicting that $F(d_b)$ is the chosen oriented face. Furthermore, the relation $\alpha_0(d_a)=(a^{-1},-\tau(a))$ implies $d_{a^{-1}}=(a^{-1},\tau(a))$ so that $\tau(a^{-1})=\tau(a)s(a)$. Together with the above claim this last equality implies that for any circuit $(a_0,a_1,\dots,a_{k-1})$ of M, we have $\tau(a_{i+1})=\tau(a_i)s(a_{i+1})$ where indices are taken modulo k. Accumulating along the circuit we deduce that $\prod_{i=0}^{k-1}s(a_i)=1$. But this precisely means that the circuit is two-sided and we conclude from Lemma 2.2.17 that M is orientable. □

Definition 2.2.19. Consider a connected map $M=(A,\rho,\iota,s)$ and the associated flag permutations φ and α_0 as in 2.1 and 2.2. We introduce another flag permutation r defined by

$$r(a,\epsilon)=(\rho^{\epsilon}(a),\epsilon)$$

and call it the **rotational permutation**. The group of permutations generated by φ,r and α_0 is called the **monodromy group** of M and is denoted by $\mathrm{Mon}(M)=\langle\varphi,r,\alpha_0\rangle$.

Lemma 2.2.20. *M is connected if and only if* $\mathrm{Mon}(M)$ *acts transitively on the set of flags.*

PROOF. Writing compositions as products, we compute from the relation $\alpha_0\varphi=\varphi^{-1}\alpha_0$:

$$\begin{aligned}\varphi^{-1}r(a,\epsilon)=\varphi^{-1}\alpha_0^2(\rho^{\epsilon}(a),\epsilon)=\alpha_0\varphi\alpha_0(\rho^{\epsilon}(a),\epsilon) &= \alpha_0\varphi\left((\rho^{\epsilon}(a))^{-1},-\epsilon s(\rho^{\epsilon}(a))\right)\\ &= \alpha_0(\rho^{-\epsilon}(\rho^{\epsilon}(a)),-\epsilon)\\ &= (a^{-1},\epsilon s(a)) \end{aligned}\tag{2.4}$$

Whence,

$$\varphi^{-1} r\alpha_0(a,\epsilon) = \varphi^{-1} r(a^{-1}, -\epsilon s(a)) = (a, -\epsilon)$$

It follows that an orbit of $\mathrm{Mon}(M)$ contains $(a,1)$ if and only if it contains $(a,-1)$. Denoting by $p_A(a,\epsilon) = a$ the arc component of a flag this implies that

$$\mathrm{Mon}(M)(a,\epsilon) = p_A\big(\mathrm{Mon}(M)(a,\epsilon)\big) \times \{-1,1\} = \langle\iota,\rho\rangle a \times \{-1,1\}$$

Recall that M is connected if its graph is connected, i.e. if $\langle\iota,\rho\rangle b = A$. Hence, M is connected if and only if $\mathrm{Mon}(M)(b,1) = A \times \{-1,1\}$. □

Lemma 2.2.21. *A connected map M is orientable if and only if the subgroup $\langle\varphi, r\rangle$ of its monodromy group does not act transitively on the set of flags. In such a case $\langle\varphi, r\rangle$ has precisely two orbits.*

Proof. Let M' be a reorientation of M induced by a vertex flipping ω. We consider the involution $\lambda : (a,\epsilon) \mapsto (a, \epsilon\omega(o(a)))$ as in Exercise 2.2.11. It is easily checked from (2.3) that $\varphi' = \lambda^{-1}\varphi\lambda$ and $r' = \lambda^{-1} r\lambda$ where φ' and r' are the facial and rotational permutations of M'. It follows that $\langle\varphi', r'\rangle$ and $\langle\varphi, r\rangle$ are conjugate subgroups of flag permutations. In particular, they have the same number of orbits. If M is orientable we can thus assume after reorientation that its signature is positive. In any case, we deduce from (2.4), that

$$p_A\big(\langle\varphi, r\rangle(a,\epsilon)\big) = p_A\big(\langle\varphi^{-1} r, r\rangle(a,\epsilon)\big) = \langle\iota,\rho\rangle a \tag{2.5}$$

In particular, when M is orientable, φ and r do not change the sign of any flag and the orbits of $\langle\varphi, r\rangle$ must be $A \times \{1\}$ and $A \times \{-1\}$ by connectivity of M.

If M is non-orientable, it contains a one-sided circuit $(a_0, a_1, \ldots, a_{k-1})$ by Lemma 2.2.17. We recursively define a permutation $g_i \in \langle\varphi, r\rangle$ with $g_0 = Id$ and $g_i = r^{n_i}\varphi g_{i-1}$, for $i = 1\ldots k$, where n_i is chosen so that $g_i(a_0,\epsilon) = (a_{i \bmod k}, \epsilon\prod_{j=0}^{i-1} s(a_j))$. This is indeed possible since by induction $g_{i-1}(a_0,\epsilon) = (a_{i-1}, \epsilon\prod_{j=0}^{i-2} s(a_j))$, implying that $\varphi g_{i-1}(a_0,\epsilon)) = (a, \epsilon\prod_{j=0}^{i-1} s(a_j))$ for some arc a with the same origin as a_i. We can then apply r as many times as needed to reach a_i. See Figure 2.18. Because r does not modify the sign of flags, the resulting flag must be $(a_{i \bmod k}, \epsilon\prod_{j=1}^{i} s(a_j))$. In particular, $g_k(a_0,\epsilon) = (a_0, \epsilon\prod_{j=0}^{k-1} s(a_j))$. But the circuit being one-sided, we have $\prod_{j=0}^{k-1} s(a_j) = -1$, whence $g_k(a_0,\epsilon) = (a_0, -\epsilon)$. Since every arc belongs to a one-sided circuit (you may use an approach path to an existing one-sided), it follows from (2.5) that $\langle\varphi, r\rangle$ is transitive.
□

We keep the following definition for later use.

Definition 2.2.22. If two flags (a,ϵ) and (b,η) of an oriented map belong to the same orbit of its monodromy subgroup $\langle\varphi, r\rangle$ the oriented faces $F(a,\epsilon)$ and $F(b,\eta)$ are said to have a **consistent orientation**.

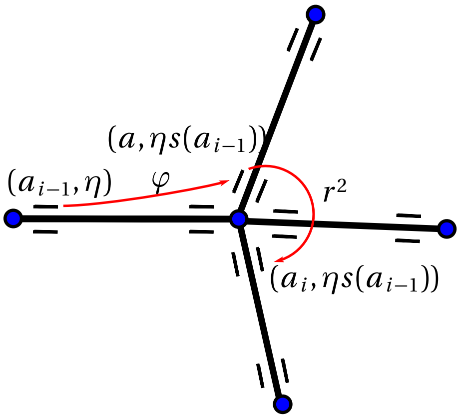


Figure 2.18: Flags are represented as dashes. The facial permutation φ transforms (a_{i-1}, η) into a flag with the same origin as a_i. One can rotate with r about that origin until reaching a flag with a_i as a first component.

2.2.2 Euler characteristic

We recall from Definitions 2.2.1 and 2.2.6 that for a map M, its vertices and edges are those of its graph and that its faces are the pair of opposite oriented faces.

Definition 2.2.23. The **Euler characteristic** of a finite map M is

$$\chi(M) = |V(M)| - |E(M)| + |\mathbf{F}(M)|$$

where $V(M)$, $E(M)$, and $\mathbf{F}(M)$ are the respective sets of vertices, edges and faces of M. Its **genus** $g(M)$ is defined by

$$\chi(M) = \begin{cases} 2 - 2g(M) & \text{if } M \text{ is orientable, and} \\ 2 - g(M) & \text{otherwise.} \end{cases}$$

2.2.3 Map morphisms

As noted at the beginning of the Orientation section maps are defined up to reorientation. Said differently, a reorientation should be considered as an isomorphism. It is thus not surprising that one should take reorientation into account to define morphisms.

Definition 2.2.24. Let $M = (A, \rho, \iota, s)$ be a map and let $N = (B, \sigma, \jmath, t)$ be a connected map. A **morphism** $(f, \omega) : M \to N$ is composed of an arc function $f : A \to B$ and a flipping $\omega : V(M) \to \{-1, 1\}$ satisfying for all arcs $a \in A$:

$$(f(a))^{-1} = f(a^{-1}) \tag{2.6}$$
$$\sigma(f(a)) = f(\rho^{\omega(o(a))}(a)) \tag{2.7}$$
$$t(f(a)) = \omega(o(a))\omega(o(a^{-1}))s(a) \tag{2.8}$$

Example 2.2.25. It directly follows from Equations (2.3) that the reorientation M' of a map M induced by a flipping $\omega : V(M) \to \{-1, 1\}$ defines a morphism $(Id, \omega) : M \to M'$.

Exercise 2.2.26. Check that the condition (2.7) can be equivalently replaced by the condition $\sigma^{\omega(o(a))}(f(a)) = f(\rho(a))$. Deduce that arcs with a common origin are mapped by f to arcs with a common origin, so that f induces a map $V(M) \to V(N)$ and a graph morphism $G(M) \to G(N)$.

Exercise 2.2.27. Let $(f,\omega): M \to N$ be a morphism and let N' be the reorientation of N induced by a flipping $\xi : V(N) \to \{-1,1\}$. We set $\omega' : V(M) \to \{-1,1\}, v \mapsto \omega(v)\xi(f(v))$. Check that $(f,\omega'): M \to N'$ is a morphism.

This last exercise leads to following composition of morphisms.

Definition 2.2.28. The composite of two morphisms $(f,\omega): M \to N$ and $(g,\xi): N \to P$ is the morphism $(g \circ f, \omega') : M \to P$ with $\omega' : V(M) \to \{-1,1\}, v \mapsto \omega(v)\xi(f(v))$. Iso-, mono-, epi-, endo-, auto-, morphisms are defined as usual.

Example 2.2.29. The reorientation M' of a map M induced by a flipping ω defines an isomorphism $(Id,\omega): M \to M'$ with inverse $(Id,\omega): M' \to M$.

For any morphism $(f,\omega) : M \to N$, we have $(f,\omega) \circ (Id,\omega) = (f,1)$. So that after reorientation of M induced by ω, Equations (2.6-2.8) are replaced by the simpler commutation relations:

$$\jmath \circ f = f \circ \iota, \quad \sigma \circ f = f \circ \rho, \quad \text{and} \quad t \circ f = s \tag{2.9}$$

Definition 2.2.30. Let $(f,\omega): M \to N$ be a morphism. The **flag extension** $\bar{f}$ of (f,ω) is defined by

$$\bar{f}(a,\epsilon) = (f(a), \epsilon\,\omega(o(a)))$$

Lemma 2.2.31. *The flag extension $\bar{f}$ of a morphism $(f,\omega): M \to N$ commutes with the three generators φ, r, α_0 of the monodromy groups of M and N respectively. Moreover, $\bar{f}$ induces a group epimorphism $\hat{f} : \mathrm{Mon}(M) \twoheadrightarrow \mathrm{Mon}(N)$ such that $\bar{f} \circ \theta = \hat{f}(\theta) \circ \bar{f}$ for all $\theta \in \mathrm{Mon}(M)$.*

PROOF. The commutation with each generator φ, r, α_0 is immediate after applying the relevant definitions. The existence of $\hat{f}$ can be shown as in Exercise 2.1.7 for the oriented case. □

As an immediate consequence, we can extend f to vertices, edges, oriented faces and faces:

Corollary 2.2.32. *$\bar{f}$ sends opposite oriented faces onto opposite oriented faces and vertex stars onto vertex stars.*

Note that the transitivity of the monodromy group of N implies that $\bar{f}$ (and f) is onto. In particular, f induces a graph epimorphism $G(M) \to G(N)$. Using the last lemma, we obtain an analogue of Lemma 2.1.6 for general maps.

Lemma 2.2.33. *Let $(f,\omega): M \to N$ be a morphism. For any vertex or face x of M of finite degree, the restriction of f to* $\mathrm{Star}(x)$*, respectively ∂x, is isomorphic to the map*

$$\begin{array}{rcl} \mathbb{Z}/(e_x d)\mathbb{Z} & \to & \mathbb{Z}/d\mathbb{Z} \\ i & \mapsto & i \mod d \end{array}$$

where d is the size of $f(\mathrm{Star}(x))$, respectively $f(\partial x)$, and e_x is a positive integer called the ***ramification index*** *of (f,ω) at x.*

Corresponding to Lemma 2.1.8, we have

Lemma 2.2.34. *All the edge fibers of a morphism $(f,\omega): M=(A,\rho,\iota,s)\to(B,\sigma,\jmath,t)$ have the same size called the* ***degree*** *of (f,ω), and denoted* $\deg(f,\omega)$ *or simply* $\deg(f)$*. Similarly, all flag fibers have size* $\deg(f)$*.*

Proof. Assuming that M has been reoriented with ω, we observe that f defines a morphism of *oriented maps* $(A,\rho,\iota)\to(B,\sigma,\jmath)$. We can thus apply Lemma 2.1.8. The proof concerning flag fibers is analogous, replacing (f,ω) by its flag extension. □

Proposition 2.2.35 (Index formula)**.** *Let $f: M \to N$ be a morphism of finite maps. For any vertex or face w of $(B,\sigma,\jmath)$, we have*

$$\sum_{f(v)=w} e_v = \deg(f)$$

Theorem 2.2.36 (Riemann-Hurwitz Formula)**.** *For a morphism $M \to N$ of degree n of finite maps we have*

$$\chi(M) = n.\chi(N) + \sum_{v\in V(M)\cup \mathbf{F}(M)} (e_v - 1)$$

Definition 2.2.37. A morphism $(f,\omega): M \to N$ is a **covering** if the restrictions of f to vertex stars and to facial circuits are bijective. Equivalently, the restrictions of the flag extension $\bar{f}$ to cycles of the facial and rotational permutations are bijective.

Example 2.2.38. To any connected map $M=(A,\rho,\iota,s)$ we can associate the map $(A',\rho',\iota',s') = (A\times\{-1,1\},\rho\times 1,\iota\times s\circ p_A, s\circ p_A)$, where

$$\rho'(a,\epsilon) = (\rho(a),\epsilon), \quad \iota'(a,\epsilon) = (\iota(a),\epsilon s(a)) \quad \text{and} \quad s'(a,\epsilon) = s(a)$$

It is easily checked that the projection $p_A: A\times\{-1,1\}\to A$ on the arc component defines a covering $(p_A,1):(A',\rho',\iota',s')\to M$. The morphism $(p_A,1)$ is a two-fold covering sometimes called the **orientation covering**. A topological construction can be obtained as follows: we first take two parallel copies of the thickened graph corresponding to the realization of M as described in the *Map realization* paragraph. The two copies of each twisted edge (i.e., with negative signature) are then modified by cutting the corresponding two strips and exchanging the pieces of the two copies before gluing them back as exampled on Figure 2.19.

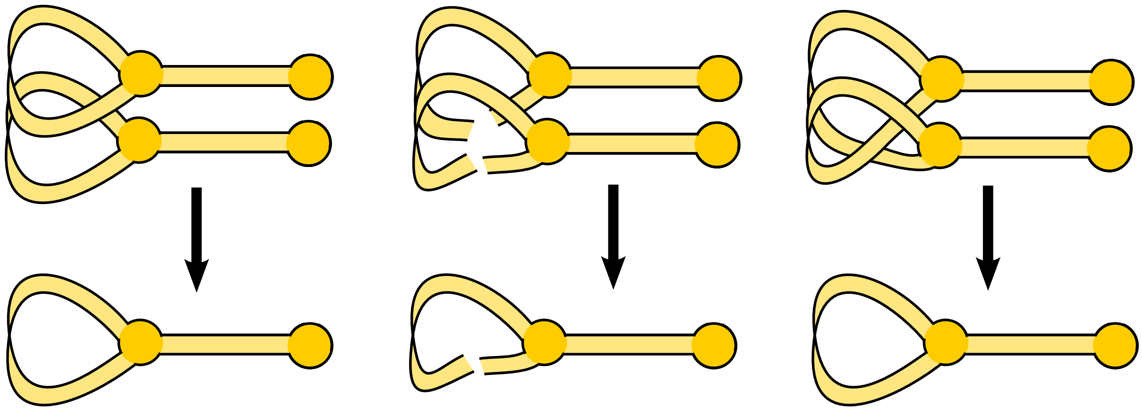


Figure 2.19: Left, The bottom map has one twisted loop edge and another edge with a degree one vertex. It is covered by two identical copies. Middle, The loop edge and its copies are cut in two pieces. Right, The cut pieces are glued back, exchanging pieces of the two copies.

Exercise 2.2.39. Show that the map (A', ρ', ι', s') in the previous example is orientable and show that it is connected if and only if (A, ρ, ι, s) in non-orientable.

2.2.4 δ-Maps

Viewing a cellular embedding of a graph as a cell complex with 0,1 and 2 dimensional cells, we may consider the maximal chains of incident cells. Those are triples vertex/edge/face such that the vertex is an endpoint of the edge which is itself bounding the face in a triple. Some triples should have multiplicities to handle multi-incidences occurring with loop edges or self-adjacent faces. The maximal chains are called flags[2] or darts and satisfy the *diamond property*: for each $i = 0,1,2$ there are exactly two flags sharing all their cells except the i-dimensional one. We thus have three fixed point free involutions acting on the set of flags and corresponding to the unique possible exchange of the i-cells. Clearly, exchanging the vertex and then the face of a flag gives the same result as exchanging the face and then the vertex. In other words, the 0 and 2-cell exchanges commute. This leads to the representation of a cellular embedding of a graph as a triple $(\alpha_0, \alpha_1, \alpha_2)$ of fixed point free involutions[3] of a set D of flags and such that $\alpha_0\alpha_2 = \alpha_2\alpha_0$. Tutte [Tut73, Tut01] calls this representation a premap and a map in the connected case. We will call it a δ-map, where δ stands for *dart*, to differentiate with the rotation system formalism.

Definition 2.2.40. A δ**-map** is a quadruple $M = (D, \alpha_0, \alpha_1, \alpha_2)$ where D is a non-empty set of **darts** and the α_i are three fixed point free involutions of D, such that α_0 and α_2 commute. A **connected component** of M is the restriction of the α_i to an orbit of the action of its **monodromy group** $\langle \alpha_0, \alpha_1, \alpha_2 \rangle$. Hence, M is connected if its monodromy group acts transitively on D. An i-**cell** of M, for $i = 0,1,2$, is an orbit of $\langle \alpha_j, \alpha_k \rangle$ with

[2] Flags can also be identified with the triangles of a barycentric triangulation of the cell complex [LZ04, Sec. 1.5.4.2]. They are called *blades* by Bryant ans Singerman [BS85].

[3] $(\alpha_0, \alpha_1, \alpha_2)$ is denoted by (L, R, T) in [BS85]. The letters L, R, T respectively stands for *longitudinal, rotational* and *transverse* reflexions.

$\{j,k\} = \{0,1,2\} \setminus \{i\}$. The **graph** of M is given by $G(M) = (V, A, \iota, o)$ where V is the set of 0-cells, $A = D/\langle\alpha_2\rangle$, $\iota = \alpha_0$ and $o = \langle\alpha_1, \alpha_2\rangle$. Here ι and o are seen as sets of permutations acting on sets of darts.

Note that α_0 and α_2 commute if and only if $\alpha_0\alpha_2$ is an involution. The subgroups $\langle\alpha_1,\alpha_2\rangle$, $\langle\alpha_2,\alpha_0\rangle$, $\langle\alpha_0,\alpha_1\rangle$ of the monodromy group are called the **vertex, edge and face group** respectively. Hence, a vertex is an orbit of the vertex group and similarly for edges and faces. By the commutation of α_0 and α_2, the edge group has at most four elements.

Similarly to the realization of a map, we can realize a δ-map $M = (D, \alpha_0, \alpha_1, \alpha_2)$ as a cellular embedding of its graph $G(M)$ so that the above δ-map representation of this cellular embedding coincides with M. To this end we consider a set of triangles indexed by the set of darts. The three vertices of each triangle are given three distinct i-cell type for $i = 0,1,2$. If two darts are related by α_i, we glue the corresponding triangles along the sides whose endpoints have the $i+1$ and $i+2$-cell type, taking indices modulo 3. See Figure 2.20. Since α_0 and α_2 commute each 1-cell type vertex is incident, after the gluing, to exactly four triangles. Those triangles correspond to an orbit of $\langle\alpha_0,\alpha_2\rangle$, i.e., an edge of M. This edge is embedded as the union of the two segments connecting the 1-cell type vertex to its 0-cell type neighbours.

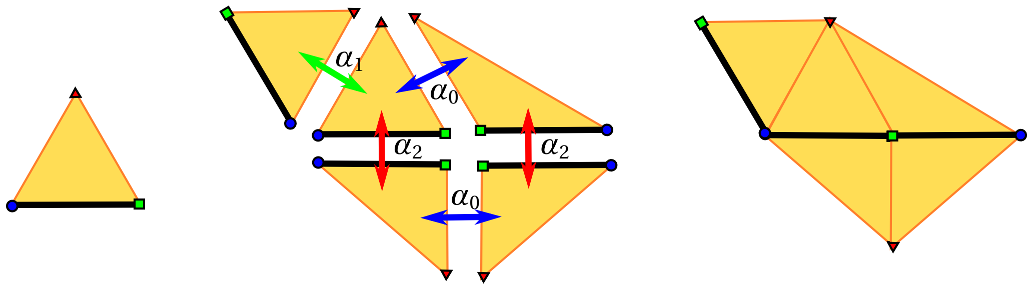


Figure 2.20: Left, a triangle whose vertices are labelled as a 0-cell (a small blue disc), a 1-cell (a small green square) and a 2-cell (a small red triangle) respectively. The side opposite to the 1-cell vertex is drawn thicker and corresponds to half an edge. Middle, The triangles are glued along their sides opposite to the i-cell type vertex whenever their corresponding darts are related by α_i. Right, The resulting piece of surface. The graph embedding is obtained as the union of the thick sides.

Exercise 2.2.41. Let α, β be two fixed point free involutions of a set D. Show that for each $d \in D$ the orbits $\langle\alpha\beta\rangle d$ and $\langle\alpha\beta\rangle\beta(d)$ are disjoint. (Hint: see the proof of Lemma 2.2.5.) Deduce that $\langle\alpha,\beta\rangle$ is isomorphic to a dihedral group $D_n = \langle a, b \mid a^2 = b^n = (ab)^2 = 1\rangle$ where n is the order of $\alpha\beta$. By convention we set $D_\infty = \langle a, b \mid a^2 = (ab)^2 = 1\rangle$.

Definition 2.2.42. Let $M = (D, \alpha_0, \alpha_1, \alpha_2)$ and $N = (B, \beta_0, \beta_1, \beta_2)$ be two δ-maps with N connected. A **morphism** $M \to N$ is a surjective dart function $f : D \to B$ that commutes with the monodromy actions, i.e., such that $f \circ \alpha_i = \beta_i \circ f$ for $i = 0,1,2$.

Exercise 2.2.43. Check that the dart function f of a morphism is onto and that the correspondence $\alpha_i \to \beta_i$, $i = 0,1,2$, induces a group morphism of the respective monodromy groups. Check that a bijective dart function defines a δ-map isomorphism.

The **centralizer** of a subset of a group is the set of group elements that commute with all the subset elements. It is immediate that the set of automorphisms of a δ-map $M=(D,\alpha_0,\alpha_1,\alpha_2)$ is the centralizer of its monodromy group considered as a subgroup of the symmetric group on its darts. If H is a subgroup of automorphisms, we have a natural action of the monodromy group on the set of orbits D/H defined for every $\alpha\in\langle\alpha_0,\alpha_1,\alpha_2\rangle$ by $\alpha(Hd):=H\alpha(d)$ for any $d\in D$. Let $C=\{\alpha\in\langle\alpha_0,\alpha_1,\alpha_2\rangle\mid\forall d\in D: \alpha(Hd)=Hd\}$ be the kernel of this action.

Definition 2.2.44. The **quotient** of the δ-map M by the subgroup H of automorphisms is the δ-map $M/H=(D/H,[\alpha_0],[\alpha_1],[\alpha_2])$ where $[\alpha_i]$ is the class of α_i in the quotient group $\langle\alpha_0,\alpha_1,\alpha_2\rangle/C$. The projection $D\to D/H$ is the dart function of the quotient morphism $M\to M/H$.

Let $\Gamma=\langle a_0,a_1,a_2\mid a_0^2=a_1^2=a_2^2=(a_0a_2)^2=1\rangle$. Considering the left action of Γ onto itself, we have the **universal δ-map** $U=(\Gamma,a_0,a_1,a_2)$.

Proposition 2.2.45. *Every connected δ-map is a quotient of the universal δ-map.*

Proof. Let $M=(D,\alpha_0,\alpha_1,\alpha_2)$ be a connected δ-map. We fix $d\in D$ and denote by $S_d=\{\alpha\in\langle\alpha_0,\alpha_1,\alpha_2\rangle\mid\alpha(d)=d\}$ the stabilizer of d for the monodromy group action. We consider the group morphism $\mu:\Gamma\to\langle\alpha_0,\alpha_1,\alpha_2\rangle$, $a_i\mapsto\alpha_i$ and the subgroup $H=\mu^{-1}(S_d)<\Gamma$. Each $h\in H$ defines an automorphism of U by the action $g\overset{h}{\mapsto}gh^{-1}$ (see Exercise 2.2.46). An H-orbit is thus a left H-coset, i.e. $\Gamma/H=\{gH\}_{g\in\Gamma}$. The kernel of the natural action of Γ on Γ/H is

$$C=\{h\in\Gamma\mid\forall g\in\Gamma: hgH=gH\}=\{h\in\Gamma\mid\forall g\in\Gamma: h\in gHg^{-1}\}=core_\Gamma(H)$$

(see the paragraph entitled 'Intrinsic algebraic formalism' in Section 2.1.1.) We thus have a quotient δ-map $U/H=(\Gamma/H,[a_0],[a_1],[a_2])$ where $[a_i]$ is the class of a_i in the quotient group Γ/C. We shall prove that M is isomorphic to U/H. For this, we consider the dart function $f:\Gamma/H\to D$, $gH\mapsto\mu(g)(d)$ where d is fixed as above. This is well defined since $\mu(H)=S_d$ fixes d. It remains to check that f induces an isomorphism of δ-maps. We first claim that f is bijective. Let $d'\in D$. By transitivity of $Mon(M)$ on D, we have $d'=\alpha(d)$ for some $\alpha\in Mon(M)$, whence $f(gH)=d'$ for any g with $\mu(g)=\alpha$. Note that μ being onto, such a g exists. So f is onto. But $f(gH)=f(g'H)$ means $\mu(g)(d)=\mu(g')(d)$, i.e., $\mu(g^{-1}g')\in S_d$. This in turn implies $g^{-1}g'\in H$, or equivalently $gH=g'H$. We conclude that f is one-to-one and onto. We finally claim that f commutes with the monodromy actions. In fact, we have

$$f([a_i](gH))=f(a_igH)=\mu(a_ig)(d)=\alpha_i(\mu(g)(d))=\alpha_i(f(gH))$$

□

Exercise 2.2.46. Check that the above action of $h\in H$ given by $g\overset{h}{\mapsto}gh^{-1}$ defines an automorphism of U.

Exercise 2.2.47. With the notations in the Proposition, prove that $core_\Gamma(H)=\ker\mu$.

The universal δ-map has faces and vertices with infinite degrees. There is actually a family of universal δ-maps $U_{m,n}$ of type (m,n) with degree m vertices and degree n faces. The δ-map $U_{m,n}$ is defined by the left action of the triangle group

$$\Gamma_{m,n} = \langle a_0, a_1, a_2 \mid a_0^2 = a_1^2 = a_2^2 = (a_0a_2)^2 = (a_2a_1)^m = (a_1a_0)^n = 1\rangle$$

on itself. This group acts as isometries of a simply connected Riemann surface $\mathscr{U}$. This surface is the sphere, the Euclidean plane or the hyperbolic plane according to whether $1/m+1/n$ is respectively larger, equal or smaller than $1/2$. A fundamental domain of $\Gamma_{m,n}$ is given by a right-angled triangle whose other two angles are π/m and π/n. The orbit of this triangle thus provides a triangulation of $\mathscr{U}$ on which $\Gamma_{m,n}$ acts transitively and freely (a triangle can only be fixed by the identity). This allows us to identify the elements of $\Gamma_{m,n}$ with the triangles of this triangulation. It can be shown [BS85] that the action on triangles of the a_i acting as isometries is isomorphic to their (algebraic) action on $\Gamma_{m,n}$ by left multiplication[4]. The four triangles incident to a degree four vertex then correspond to the four darts of an edge of $U_{m,n}$ and we can realize that edge as the union of the two sides of those triangles incident to that vertex and to the vertices of degree $2m$. This gives a geometric realization of the universal δ-map $U_{m,n}$ of type (m,n) as a regular tessellation of $\mathscr{U}$ with degree m vertices and degree n faces. See Figure 2.21.

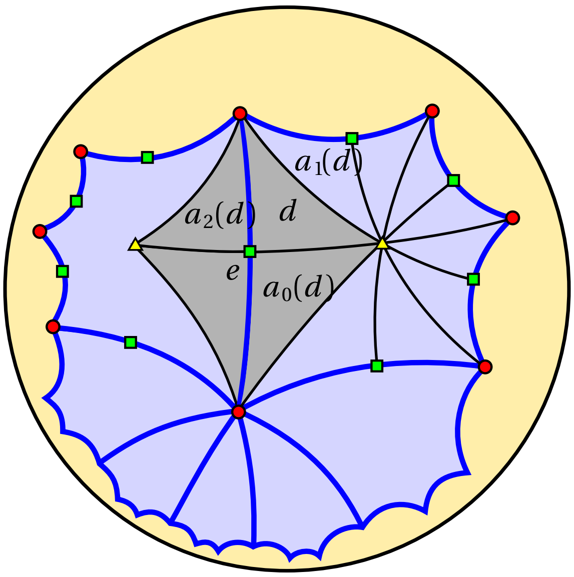


Figure 2.21: Part of the geometric realization of $U_{6,5}$ in the Poincaré disc model of the hyperbolic plane. The four grey triangles correspond to the four darts of the (blue) edge e.

Similarly to Proposition 2.2.45 we can show that a δ-map M whose vertex and face degrees have least common multiples m and n respectively is a quotient of $U_{m,n}$. This gives another way to realize M topologically. Indeed, this quotient corresponds to a subgroup of automorphisms of $U_{m,n}$ which in turn is identified with a subgroup

[4] Bryant and Singerman [BS85] use a right action, so that left and right are reversed compared to the present notes.

$H < \Gamma_{m,n}$. Then we may as well quotient the geometric realization of $U_{m,n}$ by H, acting as isometries, to obtain a topological – in fact geometric – realization. See Jones and Singerman [JS78] in the oriented case and Bryant and Singerman [BS85] for the general case. In particular, Jones and Singerman show that every finite δ-map is the quotient of a finite regular δ-map!

Equivalence between maps and δ-maps

It should be clear that maps and δ-maps are essentially the same (see Figure 2.22). Here, we make this similarity explicit with the help of an equivalence of categories.

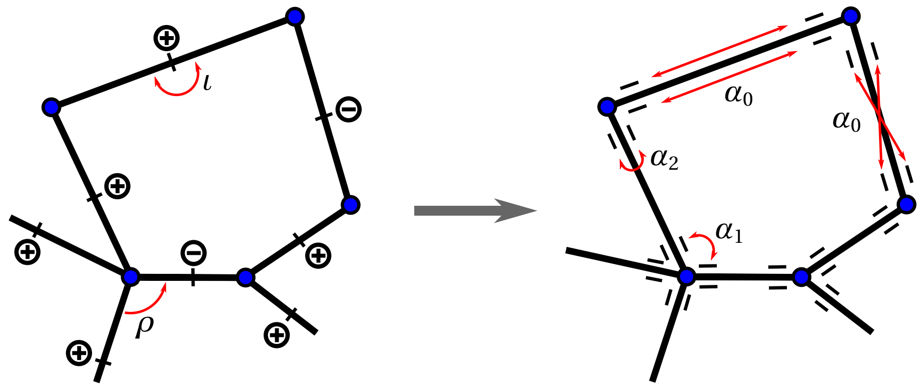


Figure 2.22: A map defined by its rotation system ρ, its involution ι and the signature of the arcs (circled signs). The corresponding δ-map with its three involutions α_0, α_1 and α_2.

Theorem 2.2.48. *The category $\mathcal{M}$ of maps and the category $\mathcal{D}$ of δ-maps are equivalent.*

Proof. We consider the functor $\delta : \mathcal{M} \to \mathcal{D}$ defined as follows. For a map $M = (A, \rho, \iota, s)$, we set $\delta(M) = (D, \alpha_0, \alpha_1, \alpha_2)$ where $D = A \times \{-1, 1\}$, and for all $(a, \epsilon) \in D$:

$$\alpha_0(a,\epsilon) = (a^{-1}, -\epsilon s(a)) \tag{2.10}$$
$$\alpha_1(a,\epsilon) = (\rho^{-\epsilon}(a), -\epsilon) \tag{2.11}$$
$$\alpha_2(a,\epsilon) = (a, -\epsilon) \tag{2.12}$$

For a morphism $(f, \omega) : M \to N = (B, \sigma, \jmath, t)$, we set $\delta(f, \omega) : F(M) \to F(N)$ with dart function $(a, \epsilon) \mapsto (f(a), \epsilon\omega(o(a))$. We leave as an exercise the verification that δ is indeed a functor. It is a theorem that δ defines an equivalence of categories if and only if it is an *essentially surjective, full and faithful* functor [Bil13]. By definition, δ is essentially surjective if every δ-map is isomorphic to some $\delta(M)$. It is full if the restriction $\delta : \mathrm{Hom}(M, N) \to \mathrm{Hom}(\delta(M), \delta(N))$ is onto and it is faithful if this restriction is one-to-one. We now check that δ satisfies those three properties.

δ is essentially surjective: Let $C = (D, \alpha_0, \alpha_1, \alpha_2)$ be a δ-map. By Exercise 2.2.41, the involution α_2 induces a fixed point free involution on the set of orbits $D/\langle\alpha_1\alpha_2\rangle$. We choose one orbit in each pair $\{\langle\alpha_1\alpha_2\rangle d, \langle\alpha_1\alpha_2\rangle\alpha_2(d)\}$ of orbits and denote by D_+ the union of the chosen orbits. We thus have a partition $D = D_+ \cup \alpha_2(D_+)$ and $d \in D_+$ implies $\langle\alpha_1\alpha_2\rangle d \subset D_+$. But $d \in D_+$ also implies $\alpha_2(d) \notin D_+$, whence $\alpha_1(d) \notin D_+$. We set for all $d \in D$

$$\varpi(d) = \begin{cases} 1 & \text{if } d \in D_+ \\ -1 & \text{otherwise.} \end{cases}$$

We now consider the map $M = (A, \rho, \iota, s)$ with

- $A = D/\langle\alpha_2\rangle$; we note $\bar{d} = \{d, \alpha_2(d)\} \in A$ the orbit of d,
- $\rho(\bar{d}) = \begin{cases} \overline{\alpha_1\alpha_2(d)} & \text{if } d \in D_+ \\ \overline{\alpha_1(d)} & \text{otherwise.} \end{cases}$
- $\iota(\bar{d}) = \overline{\alpha_0(d)}$, and
- $s(\bar{d}) = -\varpi(d)\varpi(\alpha_0(d))$.

We claim that

$$\begin{array}{ccc} D & \xrightarrow{\theta} & A \times \{-1, 1\} \\ d & \mapsto & (\bar{d}, \varpi(d)) \end{array}$$

induces an isomorphism $C \to \delta(M)$. We write $\delta(M) = (A \times \{-1, 1\}, \beta_0, \beta_1, \beta_2)$. Since θ is trivially bijective we just need to check that it commutes with the actions of the monodromy groups. We have

$$\beta_0\theta(d) = \beta_0(\bar{d}, \varpi(d)) = (\overline{\alpha_0(d)}, \varpi(\alpha_0(d))) = \theta\alpha_0(d).$$

Moreover, $\beta_1\theta(d) = (\rho^{-\varpi(d)}(\bar{d}, -\varpi(d)))$. If $\varpi(d) = 1$ then $\varpi(\alpha_2\alpha_1(d)) = 1$, whence

$$\rho^{-1}(\bar{d}) = \overline{\alpha_2\alpha_1(d)} \text{ and } \beta_1\theta(d) = (\overline{\alpha_2\alpha_1(d)}, -1) = \theta\alpha_1(d).$$

If $\varpi(d) = -1$, then we also get $\beta_1\theta(d) = (\overline{\alpha_1(d)}, 1) = \theta\alpha_1(d)$. We finally have

$$\beta_2\theta(d) = (\bar{d}, -\varpi(d)) = \theta\alpha_2(d)$$

δ is full: Let $M = (A, \rho, \iota, s)$ and $N = (B, \sigma, \jmath, t)$. We need to prove that any morphism $\delta(M) \to \delta(N)$ with dart function $g : A \times \{-1, 1\} \to B \times \{-1, 1\}$ has the form $\delta(f, \omega)$ for some morphism $(f, \omega) : M \to N$. We write $\delta(M) = (A \times \{-1, 1\}, \alpha_0, \alpha_1, \alpha_2)$, $\delta(N) = (B \times \{-1, 1\}, \beta_0, \beta_1, \beta_2)$ and $g = (g_B, g_\epsilon)$ with $g_B : A \times \{-1, 1\} \to B$ and $g_\epsilon : A \times \{-1, 1\} \to \{-1, 1\}$. The commutation conditions $g \circ \alpha_i = \beta_i \circ g$, $i = 0, 1, 2$, for g give

$$g \circ \alpha_2(a, \epsilon) = (g_B(a, -\epsilon), g_\epsilon(a, -\epsilon)) = \beta_2 \circ g(a, \epsilon) = (g_B(a, \epsilon), -g_\epsilon(a, \epsilon))$$

whence $g_B(a, -\epsilon) = g_B(a, \epsilon)$ and $g_\epsilon(a, -\epsilon) = -g_\epsilon(a, \epsilon)$. It follows that $g_B(a, \epsilon) = g_B(a)$ only depends on a and that $g_\epsilon(a, \epsilon) = \epsilon\upsilon(a)$ for some function $\upsilon : A \to \{-1, 1\}$. For $i = 1$ we get

$$\begin{aligned} g \circ \alpha_1(a, \epsilon) &= g(\rho^{-\epsilon}(a), -\epsilon) = (g_B(\rho^{-\epsilon}(a)), -\epsilon\upsilon(\rho^{-\epsilon}(a))) \quad \text{and} \\ \beta_1 \circ g(a, \epsilon) &= \beta_1(g_B(a), \epsilon\upsilon(a)) = (\sigma^{-\epsilon\upsilon(a)}(g_B(a)), -\epsilon\upsilon(a)) \end{aligned}$$

and we infer that

$$g_B(\rho^{-\epsilon}(a)) = \sigma^{-\epsilon\upsilon(a)}(g_B(a)) \quad \text{and} \quad \upsilon(\rho^{-\epsilon}(a)) = \upsilon(a)$$

In particular $\upsilon(a)$ only depends on $o(a)$. We finally get for $i = 0$

$$\begin{aligned} g \circ \alpha_0(a,\epsilon) &= g(\iota(a), -\epsilon s(a)) = (g_B(\iota(a)), -\epsilon s(a)\upsilon(\iota(a))) \quad \text{and} \\ \beta_0 \circ g(a,\epsilon) &= \beta_0(g_B(a), \epsilon\upsilon(a)) = (\jmath(g_B(a)), -\epsilon\upsilon(a)t(g_B(a))) \end{aligned}$$

whence

$$g_B(\iota(a)) = \jmath(g_B(a)) \quad \text{and} \quad t(g_B(a)) = \upsilon(a)\upsilon(\iota(a))s(a)$$

Setting $f = g_B$ and $\omega = \upsilon$, it follows directly from Definition 2.2.24 that $\delta(f,\omega) = g$.

δ is faithful: This is trivial from the definition of δ. $\square$

Exercise 2.2.49. Using the notations as in the above proof that δ is essentially surjective prove that $\theta^{-1}\phi\theta = \alpha_1\alpha_0$ where φ is the facial permutation of M.

Exercise 2.2.50. Show that the graphs of M and $\delta(M)$ as defined in Definitions 2.2.1 and 2.2.40 are indeed isomorphic.

Remark 2.2.51. Let $M = (A,\rho,\iota,s)$ and $\delta(M) = (D,\alpha_0,\alpha_1,\alpha_2)$ where δ is the functor introduced in the proof of the theorem. It is easily checked that

$$\begin{aligned} \varphi &= \alpha_1\alpha_0 \qquad \text{and} \\ r &= \alpha_1\alpha_2 \end{aligned}$$

where φ and r are the facial and rotational permutations of M introduced in Definitions 2.2.1 and 2.2.19 respectively. In particular, the monodromy group as defined in 2.2.19 and 2.2.40 are the same, i.e., $Mon(M) = Mon(\delta(M))$.

2.2.5 Basic operations on maps

As for the oriented case, we review some basic operations on maps.

Dual maps

Exchanging the roles of vertices and faces is rather trivial for δ-maps. It amounts to swap the involutions α_0 and α_2, so that the **dual** of the δ-map $(D,\alpha_0,\alpha_1,\alpha_2)$ is the δ-map $(D,\alpha_2,\alpha_1,\alpha_0)$. In terms of maps, the duality is made more complicated by the fact that the inverse of the above category equivalence δ is not canonical. We can only define a dual map up to reorientation. Rewording the same construction as in the proof of Theorem 2.2.48, we obtain

Definition 2.2.52. Let $M = (A,\rho,\iota,s)$ be a map with facial permutation φ. We chose an orientation for each face of M, i.e., a function $\eta : A \times \{-1,1\} \to \{-1,1\}$ such that

$$\eta(\varphi(a,\epsilon)) = \eta(a,\epsilon) \quad \text{and} \quad \eta(\alpha_0(a,\epsilon)) = -\eta(a,\epsilon)$$

A **dual map** is any reorientation of the map $M^* = (A^*,\rho^*,\iota^*,s^*)$ where

- $A^* = (A \times \{-1,1\})/\langle \alpha_0 \rangle$, and denoting by $[a,\epsilon]$ the orbit $\{(a,\epsilon), \alpha_0(a,\epsilon)\}$,
- $\iota^*[a,\epsilon] = [a,-\epsilon]$
- $\rho^*[a,\epsilon] = [\varphi^{\eta(a,\epsilon)}(a,\epsilon)]$
- $s^*[a,\epsilon] = -\eta(a,1)\eta(a,-1)$

We leave to the reader the verification that M^* is indeed a map. We shall speak of *the* dual of a map albeit it is formally defined up to reorientation.

Lemma 2.2.53. *The vertices, edges and faces of M^* are in one-to-one correspondence with the faces, edges and vertices of M, respectively. A vertex is incident to an edge in M^* if and only if the corresponding face and edge in M are incident.*

PROOF. The property is trivial for δ-maps when viewing vertices, edges and faces as orbits of $\langle \alpha_1, \alpha_2 \rangle$, $\langle \alpha_0, \alpha_2 \rangle$ and $\langle \alpha_0, \alpha_1 \rangle$ respectively. It remains true for maps since by construction $\delta(M^*)$ is isomorphic to the dual of the δ-map δM. □

Proposition 2.2.54. *The dual M^* of a connected map M is connected and has the same orientability as M. Furthermore, if M is finite then M^* has the same Euler characteristic as M, hence the same genus.*

PROOF. From the equivalence between maps and δ-maps, it is sufficient to prove the proposition for δ-maps. Let $M = (D, \alpha_0, \alpha_1, \alpha_2)$ be a connected δ-map and let $M^* = (D, \alpha_2, \alpha_1, \alpha_0)$ be its dual. M and M^* trivially have the same monodromy group and the same subgroup $\langle \alpha_1\alpha_0, \alpha_1\alpha_2 \rangle$. So, M^* is connected and from the characterisation of Lemma 2.2.21 and the discussion after Exercise 2.2.49, M^* is orientable if and only if M is orientable. The second part of the proposition is then a direct consequence of the previous lemma. □

The fact that $(M^*)^* = M$ up to reorientation is also a trivial consequence of the equivalence with δ-maps.

Exercise 2.2.55. Assuming that M is oriented, hence orientable, define a canonical dual map.

Edge contraction

Similarly to the definition 2.1.15 of an edge contraction in the oriented case, we have

Definition 2.2.56. Let $M = (A, \rho, \iota, s)$ be a connected map with at least two edges and let $e = \{a, a^{-1}\}$ be an edge of M. If e has two distinct endpoints, we further assume that e has positive signature. The **contraction** of e in M transforms M to a map $M/e =$

$(A\setminus e, \rho', \iota', s')$ where ι' and s' are the restrictions to $A\setminus e$ of ι and s respectively and ρ' is obtained by merging the cycles of a and a^{-1}, i.e.,

$$\forall b \in A\setminus e,\ \rho'(b) = \begin{cases} \rho(b) & \text{if } \rho(b) \notin e, \\ \rho\circ\iota(\rho(b)) & \text{if } \rho(b)\in e \text{ and } \rho\circ\iota(\rho(b))\notin e, \\ (\rho\circ\iota)^2(\rho(b)) & \text{otherwise.} \end{cases}$$

We remark that the condition of positive signature is not restrictive as any non-loop edge with negative signature appears with positive signature after flipping the orientation of one of its endpoints. Moreover, the contraction of a loop edge corresponds to the deletion of this edge and is also treated thereafter as an edge deletion. Finally, the condition that M contains at least two edges ensures that M/e is not the empty map (see Remark 2.2.2).

As in the oriented case, we can check that the graph of M/e is obtained by contracting e in the graph of M, i.e., $G(M/e) = G(M)/e$.

Exercise 2.2.57. Show that the facial permutation φ' of M/e satisfies

$$\forall (b,\epsilon)\in (A\setminus e)\times\{-1,1\},\quad \varphi'(b,\epsilon) = \begin{cases} \varphi(b,\epsilon) & \text{if } \varphi(b,\epsilon)\notin e\times\{1,-1\} \\ \varphi^2(b,\epsilon) & \text{if } \varphi(b,\epsilon)\in e\times\{1,-1\} \text{ and} \\ & \varphi^2(b,\epsilon)\notin e\times\{1,-1\} \\ \varphi^3(b,\epsilon) & \text{otherwise.} \end{cases}$$

where φ is the facial permutation of M.

Proposition 2.2.58. *If e is a non-loop edge with positive signature in a finite connected map M having at least two edges, then M/e is connected and has the same orientability and Euler characteristic as M, hence the same genus.*

Proof. The connectedness of M/e results from the identity $G(M/e) = G(M)/e$. It easily follows from the characterization of Lemma 2.2.17 that M and M/e have the same orientability. Now, it results from the previous exercise that the cycles of the facial permutation φ' of M/e are deduced from the cycles of the facial permutation φ of M by the removal of the flags in $e\times\{-1,1\}$. Because M has at least two edges and because e is not a loop edge, none of the cycles of φ is reduced to flags in $e\times\{-1,1\}$. So, φ' and φ have the same number of cycles, whence M and M/e have the same number of faces. On the other hand, M/e has one edge less and one vertex less than M. We conclude that M and M/e have the same Euler characteristic. □

Edge deletion

Definition 2.2.59. Let $M = (A, \rho, \iota, s)$ be a connected map with at least two edges and let $e = \{a, a^{-1}\}$ be an edge of M. The **deletion** of e in M transforms M into a map

$M-e=(A\setminus e,\rho',\iota',s')$ where ι' and s' are the restrictions to $A\setminus e$ of ι and s respectively, and ρ' is obtained by deleting a and a^{-1} in the cycles of ρ, i.e.,

$$\forall b\in A\setminus e,\ \rho'(b)=\begin{cases}\rho(b) & \text{if } \rho(b)\notin e,\\ \rho^2(b) & \text{if } \rho(b)\in e \text{ and } \rho^2(b)\notin e,\\ \rho^3(b) & \text{otherwise.}\end{cases}$$

As in the oriented case, we observe that $G(M-e)=G(M)-e$. The facial permutation φ' of $M-e$ can be made explicit:

Exercise 2.2.60. Show that

$$\forall (b,\epsilon)\in(A\setminus e)\times\{-1,1\},\ \varphi'(b,\epsilon)=\begin{cases}\varphi(b,\epsilon) & \text{if } \varphi(b,\epsilon)\notin e\times\{-1,1\},\\ r\varphi(b,\epsilon) & \text{if } \varphi(b,\epsilon)\in e\times\{-1,1\} \text{ and}\\ & r\varphi(b,\epsilon)\notin e\times\{-1,1\},\\ r^2\varphi(b,\epsilon) & \text{otherwise,}\end{cases}$$

where r is the rotational permutation given in Definition 2.2.19.

Recall that an edge incident to two distinct faces is said regular and singular otherwise.

Proposition 2.2.61. *Let $e=\{a,a^{-1}\}$ be an edge without degree one endpoint of a connected map M with at least two edges. Setting $d=(a,1)$, we have*

$$\chi(M-e)=\begin{cases}\chi(M) & \textit{if } e \textit{ is regular}\\ \chi(M)+1 & \textit{if } F(d)=F(\alpha_2(d)),\\ \chi(M)+2 & \textit{otherwise, i.e., if } F(d)=F(\alpha_0\alpha_2(d)).\end{cases}$$

Moreover, $M-e$ is orientable if and only if every one-sided circuit of M contains e. (This last condition is trivially verified when M is orientable.)

The different cases are illustrated in Figure 2.23. If the edge e of the proposition had a degree one endpoint, its deletion would leave an isolated vertex corresponding to the empty map. The formula for the case $F(d)=F(\alpha_2\alpha_0(d))$ would become wrong unless this empty map is considered as a sphere component as in Remark 2.2.2. However, we can avoid such considerations replacing the deletion of e by its contraction according to Definition 2.2.56.

PROOF. The last equivalence on orientability is a direct consequence of Lemma 2.2.17. For the first part of the proposition, there are several cases to consider depending upon whether $F(d)$, $F(\alpha_0(d))$, $F(\alpha_2(d))$ and $F(\alpha_0\alpha_2(d))$ are pairwise distinct or not. Note that $F(d)\neq F(\alpha_0(d))$ and $F(\alpha_2(d))\neq F(\alpha_0\alpha_2(d))$ by Lemma 2.2.5. In any case, $M-e$ has one edge less than M and the same number of vertices as M. Moreover, from the previous exercise, only the faces containing flags in $e\times\{-1,1\}$ are modified by the deletion of e.

If e is regular: Then $F(d)$, $F(\alpha_0(d))$, $F(\alpha_2(d))$ and $F(\alpha_0\alpha_2(d))$ are pairwise distinct. By Remark 2.2.51, the rotational permutation can be expressed as $r=\varphi\alpha_0\alpha_2$. The

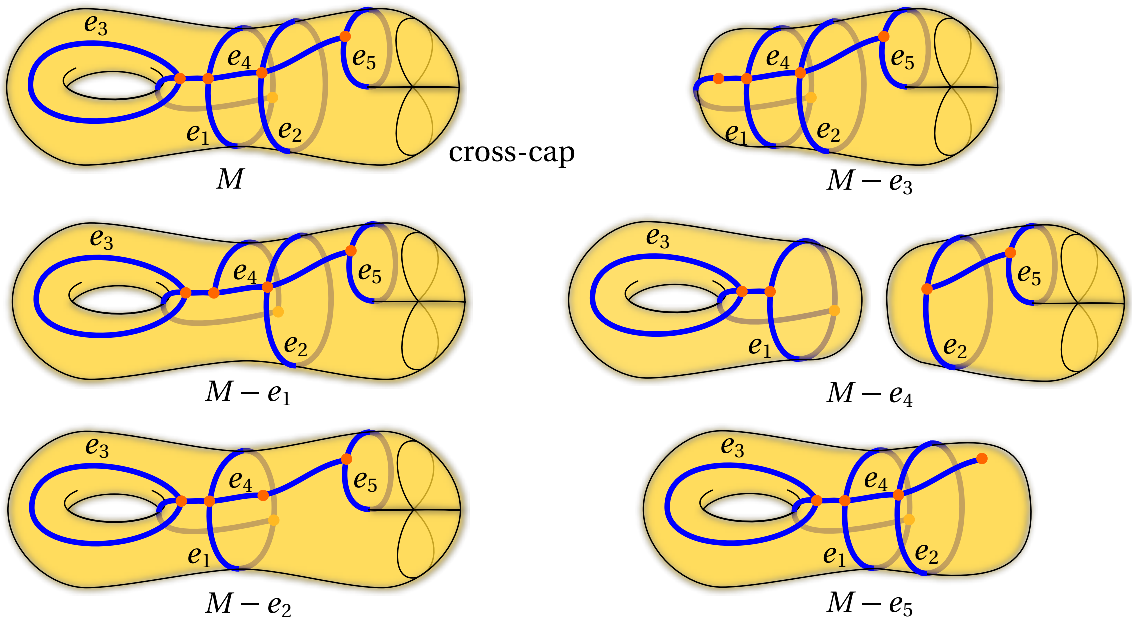


Figure 2.23: Deletion of an edge in the map M. Since e_1 and e_2 are regular $\chi(M-e_1)=\chi(M-e_2)=\chi(M)=-1$. On the other hand, e_3 and e_4 are singular and $F(e,1)=F(\alpha_0\alpha_2(e,1))$ for $e=e_3,e_4$, whence $\chi(M-e_3)=\chi(M-e_4)=\chi(M)+2=1$. The edge e_5 is also singular but $F(e_5,1)=F(\alpha_2(e_5,1))$ so that $\chi(M-e_5)=\chi(M)+1=0$.

expression of the facial permutation φ' of $M-e$ given in the above exercise then shows that $F(d)$ is merged with $F(\alpha_0\alpha_2(d))$ and the darts in $e\times\{-1,1\}$ are removed to form an oriented face F of $M-e$. Note that d and $\alpha_0\alpha_2(d)$ cannot both be fixed points of φ so that F is not empty. Likewise, $F(\alpha_0(d))$, $F(\alpha_2(d))$ are merged to give the face opposite to F. The number of (unoriented) faces is therefore reduced by one by the deletion of e. It follows that the characteristic of $M-e$ and M are equal.

If $F(d)=F(\alpha_2(d))$: Then, either $\varphi^{\epsilon}(d)=\alpha_2(d)$ for some $\epsilon\in\{-1,1\}$ or d and $\alpha_2(d)$ are not consecutive in $F(d)$. Suppose $\varphi(d)=\alpha_2(d)$ so that e must be a loop edge with negative signature. Then $F(d)=(d,\alpha_2(d),L)$ for some non-empty sequence L of flags and $F(\alpha_0(d))=(\overline{\alpha_0(L)},\alpha_0\alpha_2(d),\alpha_0(d))$ where $\overline{\alpha_0(L)}$ is the reversed of the sequence L to which α_0 is applied. From the expression of φ', we get two new faces (L) and $(\overline{\alpha_0(L)})$. It follows that the number of faces in $M-e$ is the same as in M. The case where $\varphi^{-1}(d)=\alpha_2(d)$ can be treated the same way. Now, if d and $\alpha_2(d)$ are not consecutive in $F(d)$, then $F(d)=(d,L_1,\alpha_2(d),L_2)$ for some non-empty sequences of flags L_1,L_2 and $F(\alpha_0(d))=(\overline{\alpha_0(L_2)},\alpha_0\alpha_2(d),\overline{\alpha_0(L_1)},\alpha_0(d))$. From the expression of φ', we get two new opposite faces $(L_1,\overline{\alpha_0(L_2)})$ and $(L_2,\overline{\alpha_0(L_1)})$. It again follows that $|\mathbf{F}(M-e)|=|\mathbf{F}(M)|$, and we conclude in both cases that $\chi(M-e)=\chi(M)+1$.

If $F(d)=F(\alpha_0\alpha_2(d))$: Then $F(d)=(d,L_1,\alpha_2(d),L_2)$ for some non-empty sequences of flags L_1,L_2. Indeed, if L_1 or L_2 were empty, then e would have a degree one vertex in contradiction with the hypotheses in the proposition. From the expression of φ', the opposite faces $F(d)$ and $F(\alpha_0(d))$ give rise to four cycles of φ', namely (L_1), (L_2) and their opposite faces. The number of faces is therefore increased by one and the characteristic is increased by two. □

Edge deletion and contraction commutes with reorientation by vertex flipping so that these operations are defined up to map isomorphism. This is best seen by the fact that edge contraction and deletion can be expressed in terms of δ-maps.

Definition 2.2.62. Let $S = (D, \alpha_0, \alpha_1, \alpha_2)$ be a connected δ-map with at least two edges and let $e = \langle \alpha_0, \alpha_2 \rangle d$ be an edge of S for some $d \in D$.

- The **contraction** of e in S transforms S to $S/e = (D \setminus e, \beta_0, \beta_1, \beta_2)$ where β_0, β_2 are the restrictions to $D \setminus e$ of α_0, α_2 respectively and β_1 is given by

$$\forall x \in D \setminus e, \quad \beta_1(x) = \begin{cases} \alpha_1(x) & \text{if } \alpha_1(x) \notin e \\ \alpha_1\alpha_0\alpha_1(x) & \text{if } \alpha_1(x) \in e \text{ and } \alpha_1\alpha_0\alpha_1(x) \notin e \\ (\alpha_1\alpha_0)^2\alpha_1(x) & \text{otherwise.} \end{cases}$$

- The **deletion** of e in S transforms S to $S - e = (D \setminus e, \gamma_0\gamma_1, \gamma_2)$ where γ_0, γ_2 are the restrictions to $D \setminus e$ of α_0, α_2 respectively and γ_1 is given by

$$\forall x \in D \setminus e, \quad \gamma_1(x) = \begin{cases} \alpha_1(x) & \text{if } \alpha_1(x) \notin e \\ \alpha_1\alpha_2\alpha_1(x) & \text{if } \alpha_1(x) \in e \text{ and } \alpha_1\alpha_2\alpha_1(x) \notin e \\ (\alpha_1\alpha_2)^2\alpha_1(x) & \text{otherwise.} \end{cases}$$

Exercise 2.2.63. Check that S/e and $S - e$ are well-defined δ-maps. Check that the edge deletion and contraction of δ-maps and of maps are in correspondence. In other words, if M is a map, check that $\delta(M - e) = \delta(M) - e$ and $\delta(M/e) = \delta(M)/e$ where δ is the equivalence between maps and δ-maps.

Lemma 2.2.64. *Let e be an edge of a connected map M with at least two edges. Then, $(M/e)^* = M^* - e$ and $(M - e)^* = M^*/e$.*

Proof. By the previous exercise, it is sufficient to prove the property for δ-maps. Since the dual of a δ-map is obtained by exchanging the first and last of its three involutions, the lemma is trivial in view of Definition 2.2.62. □

Let e be an edge of a map M. We may consider the darts of $M - e$ as a subset of the darts of M.

Lemma 2.2.65. *Suppose that e is singular. The face of a dart d of $M - e$, viewed as an orbit of the face group of $M - e$, is included in the face of d in M. Moreover, an edge of $M - e$ is regular if and only if it is regular in M.*

Proof. Denote by $\gamma_0, \gamma_1, \gamma_2$ the dart involutions of $M - e$ and by $\alpha_0, \alpha_1, \alpha_2$ those of M. We write $e = \langle \alpha_2, \alpha_0 \rangle d_e$ and denote by $F = \langle \alpha_0, \alpha_1 \rangle d_e$ the face of d_e in M. By assumption, $e \subset F$. If the face F' of d is distinct from F, then the γ_i and α_i have the same restrictions on F'. If $F' = F$ and $u \in F \setminus e$ we claim that $\gamma_1(u) \in F \setminus e$. Indeed, if $\alpha_1(u) \notin e$ then $\gamma_1(u) = \alpha_1(u) \in F \setminus e$. If $\alpha_1(u) \in e$ and $\alpha_1\alpha_2\alpha_1(u) \notin e$ then $\gamma_1(u) = \alpha_1\alpha_2\alpha_1(u) \in F \setminus e$. We also conclude that $\gamma_1(u) \in F \setminus e$ when $\alpha_1(u) \in e$ and $\alpha_1\alpha_2\alpha_1(u) \in e$. In either case, γ_0 and γ_1 are stable on $F \setminus e$, implying that $\langle\gamma_0, \gamma_1\rangle d \subset F'$.

Suppose that the edge e' of d is regular in M, i.e., that $\alpha_2(d) \notin F'$. By the first part of the lemma we have $\gamma_2(d) = \alpha_2(d) \notin \langle\gamma_0, \gamma_1\rangle d$ implying that e' is regular in $M - e$. On the contrary, if e' is singular in M we have $\alpha_2(d) \in F'$. In other words there exists a sequence of darts $\mathscr{D} = (d_1 = d, d_2, \ldots, d_k = \alpha_2(d))$ with $d_{j+1} = \alpha_{i_j}(d_j)$ where $i_j \in \{0, 1\}$ for $j = 1, \ldots, k-1$. We may assume that the sequence has no repeated dart. If no dart of e appears in $\mathscr{D}$, then $\alpha_{i_j}(d_j) = \gamma_{i_j}(d_j)$ for all j so that $\gamma_2(d) \in \langle\gamma_0, \gamma_1\rangle d$. Otherwise, let j be the smallest index such that $d_j \in e$. We must have $d_j = \alpha_1(d_{j-1})$ because $d_{j-1} \notin e$. If d_j is the only dart of e in $\mathscr{D}$ then $d_{j+1} = d_{j-1}$ and we can short cut d_j to get a sequence a darts from d to $\gamma_2(d)$, not in e and related by α_0 and α_1, hence by γ_0 and γ_1. We conclude that $\gamma_2(d) \in \langle\gamma_0, \gamma_1\rangle d$. If d_j is not the only dart of e, then there is exactly one other dart $d_\ell = \alpha_2(d_j)$ in $\mathscr{D}$ and we easily compute from the definition of γ_1 that $\gamma_1(d_{j-1}) = d_{\ell+1}$. We can again extract from $\mathscr{D}$ a subsequence of darts from d to $\gamma_2(d)$, not in e and related by γ_0 and γ_1. We conclude again that $\gamma_2(d) \in \langle\gamma_0, \gamma_1\rangle d$, i.e., that e' is singular in $M - e$.
□

Proposition 2.2.66. *Let e be a regular edge of a map M. For each face incident to e there is a simple circuit of regular edges through e that is included in the support of this face. In particular, the regular edge-induced subgraph of $G(M)$ is bridgeless.*

Intuitively, we can consider the topological realization of M and erase all the singular edges incident to the chosen face, say F, to obtain a non cellular graph embedding on the same surface. The closure of the components of the graph complement are left unchanged. In particular, regular edges remain regular and incident to the same "face components". We next delete the component corresponding to F to obtain a surface with boundaries that we can recap with discs. Compare M and $M - e_4$ on Figure 2.23. The boundaries of those discs are simple circuits of regular edges. This is the idea of the following proof.

Proof. Let a, a^{-1} be the two arcs of e and let $F(a, \epsilon)$ be an oriented face incident to e where $\epsilon \in \{-1, 1\}$. As long as $\partial F(a, \epsilon)$ contains an arc of a singular edge, we delete that edge from M. By the previous lemma, we obtain a new map whose graph is a subgraph of $G(M)$ and such that the face of (a, ϵ) contains regular edges only that are also regular edges of $F(a, \epsilon)$ in M. We conclude the lemma by noting that the facial circuit of this face is a simple circuit in $G(M)$. □

Edge subdivision

Definition 2.2.67. Let e be an edge of a map $M = (A, \rho, \iota, s)$. The **subdivision** of e in M transforms M to a map $S_e M = (A', \rho', \iota', s)$ where (A', ρ', ι') is given as in Definition 2.1.21 for the oriented case, and s' extend s by assigning 1 to the two newly introduced arcs.

Using the orientability criterion of Lemma 2.2.17, we easily check that

Proposition 2.2.68. *$S_e M$ and M have the same orientability, the same number of connected components and the same Euler characteristic.*

Face subdivision

A face subdivision of a map consists in adding an edge inside a face of the map. The endpoints of this edge are specified between two flags of the (oriented) face. Figure 2.24 depicts a face subdivision when these flags are either distinct or equal.

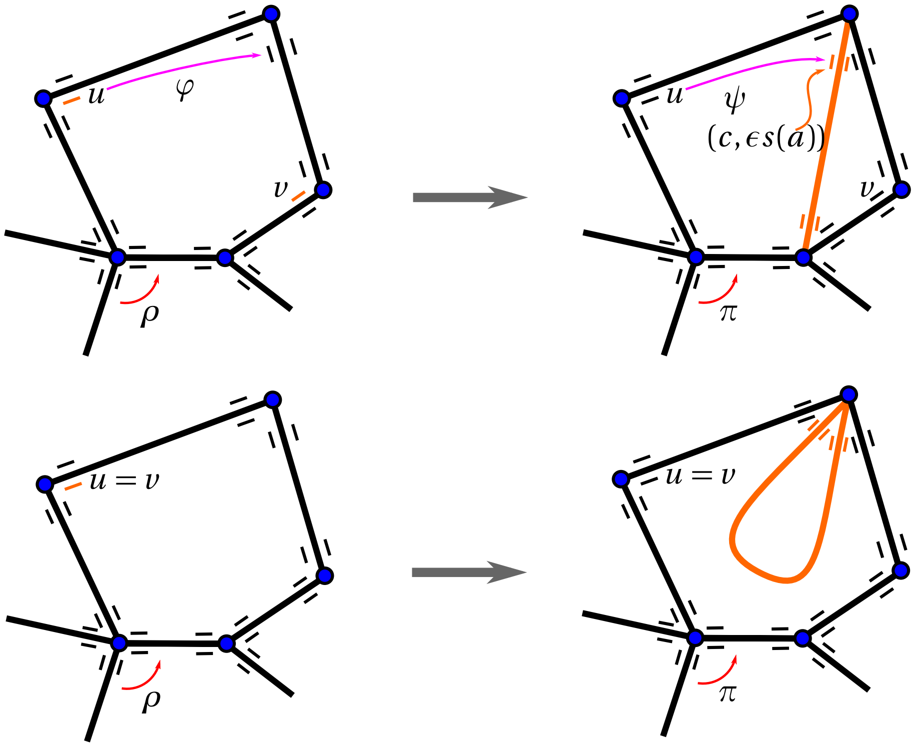


Figure 2.24: Up, a face subdivision between two distinct flags u and v. The facial permutations before and after the subdivision are denoted by φ and ψ respectively. Down, the case where $u = v$.

Definition 2.2.69. Let $M = (A, \rho, \iota, s)$ be a map and let $u = (a, \epsilon)$, $v = (b, \eta) \in A \times \{-1, 1\}$ be two flags of M, possibly equal, belonging to a same face $F(u) = F(v)$. The **subdivision** of $F(u)$ from u to v transforms M to a map $S_{(u,v)}M = (A \cup \{c, c^{-1}\}, \pi, \iota', s')$ obtained by adding a new edge $e = \{c, c^{-1}\}$ in $F(u)$ between the heads of a and b (see Figure 2.24). The involution ι' extends ι to $A \cup e$ by setting $\iota(c) = c^{-1}$ and the signature s' extends s by setting $s(e) = \epsilon\eta s(a)s(b)$ (see Exercise 2.2.16).

- When $u \neq v$ the rotation system π is determined by the relations:

$$\begin{aligned} \pi^{\epsilon s(a)}(c) &= \rho^{\epsilon s(a)}(a^{-1}), & \pi^{\eta s(b)}(c^{-1}) &= \rho^{\eta s(b)}(b^{-1}), \\ \pi^{-\epsilon s(a)}(c) &= a^{-1}, & \pi^{-\eta s(b)}(c^{-1}) &= b^{-1} \end{aligned}$$

  These four relations provide the π-image of c, c^{-1} and of either a^{-1} or $\rho^{-1}(a^{-1})$ and of either b^{-1} or $\rho^{-1}(b^{-1})$ depending upon the signs of $\epsilon s(a)$ and $\eta s(b)$. For instance, the first relation gives $\pi(\rho^{-1}(a^{-1}))=c$ when $\epsilon s(a)=-1$. For every other $x\in A$, we set $\pi(x)=\rho(x)$.

- When $u=v$ the rotation system π is determined by

$$\pi^{\epsilon s(a)}(c)=c^{-1},\quad \pi^{-\epsilon s(a)}(c)=a^{-1},\quad \pi^{\epsilon s(a)}(c^{-1})=\rho^{\epsilon s(a)}(a^{-1})$$

For every $x\in A$ such that $\pi(x)$ is not determined by these three relations, we set $\pi(x)=\rho(x)$.

Denoting by φ and ψ the facial permutations of M and $S_{(u,v)}M$ respectively, we easily compute (tacking that $\psi\alpha_0=\alpha_0\psi^{-1}$):

if $u\neq v$:			if $u=v$:		
$\psi(u)$	$=$	$(c,\epsilon s(a))$,	$\psi(u)$	$=$	$(c,\epsilon s(a))$,
$\psi(v)$	$=$	$(c^{-1},\eta s(b))$,			
$\psi(c,\epsilon s(a))$	$=$	$\phi(v)$,	$\psi(c,\epsilon s(a))$	$=$	$\phi(u)$,
$\psi(c,-\epsilon s(a))$	$=$	$\alpha_0(v)$,	$\psi(c,-\epsilon s(a))$	$=$	$(c,-\epsilon s(a))$,
$\psi(c^{-1},\eta s(b))$	$=$	$\phi(u)$,	$\psi(c^{-1},\epsilon s(a))$	$=$	$(c^{-1},\epsilon s(a))$,
$\psi(c^{-1},-\eta s(b))$	$=$	$\alpha_0(u)$,	$\psi(c^{-1},-\epsilon s(a))$	$=$	$\alpha_0(u)$.

For all other flags, ψ coincides with ϕ. Since $S_{(u,v)}M-e=M$ and since e is regular in $S_{(u,v)}M$, we deduce from Lemma 2.2.17 and Proposition 2.2.61 that

Proposition 2.2.70. *$S_{(u,v)}M$ and M have the same orientability, the same number of connected components and the same Euler characteristic.*

Map restriction

Considering the cellular embedding of a map $M=(A,\rho,\iota,s)$, we can remove some of the edges of this embedding and its isolated vertices to obtain a graph embedding which might not be cellular anymore. Some components of the complement of the remaining subgraph might indeed be non trivial. Removing each non disc component, we obtain a surface with boundaries (a first thickening of the graph prevents from singularities) that we can fill in with discs to get a cellular embedding. This leads to the notion of restricted map.

Definition 2.2.71. Let $B\subset A$ be a subset of arcs stable by the inversion ι. The **restriction** of the map M to B is the map $M_B=(B,\rho_B,\iota_B,s_B)$ where ι_B and s_B are the respective restrictions of ι and s to B and ρ_B is given for any $b\in B$ by

$$\rho_B(b)=\rho^k(b) \text{ where } k=\min\{i\geq 1 \mid \rho^i(b)\in B\}$$

If α_i, $i=0,1,2$ are the three dart involutions of M and α_i^B are the corresponding involutions for M_B, we remark that $\alpha_0^B=\alpha_0$.

Note that the restricted map M_B is the result of the deletion of all the edges in B.

2.3 Maps with Boundary

2.3.1 δ-maps with boundary

Bryant and Singerman [BS85] have developed a theory of combinatorial surfaces with boundary. Formally, a combinatorial surface with boundary is just a δ-map whose three involutions may have fixed points. The topological counterpart is a cellular embedding of a graph in a surface with boundary. The embedding is required to be such that each component in the complement of the graph is homeomorphic to either an open disc $\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1\}$ or a half-disc $\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1, y \geq 0\}$. The graph is allowed to have specified degree one vertices called *free ends* (see Figure 2.5) either interior to the surface or on its boundary. From this embedding, we construct a δ-map with boundary as follows. We perform a barycentric subdivision by first inserting a vertex

- at the center of each disc-face,
- at the midpoint of the boundary of each half-disc-face,
- at the midpoint of each edge without free end, and
- in place of free ends (see Figure 2.25).

We call a vertex of the graph a 0-vertex, while the midpoints and free-ends of the edges are called 1-vertices, and the vertices added in faces, either of type disc or half-disc, are called 2-vertices. A *corner* of a i-vertex is a component of the complement of the embedded graph in a small surface disc centered at that vertex. We then join the 2-vertex of each face to the corners of the 0- and 1-vertices incident to this face unless they are already connected by a boundary segment. We obtain this way a set of triangular faces, each one being incident to a 0-vertex, a 1-vertex and a 2-vertex. We define the set of darts of the δ-map as the set of triangular faces. For $i = 0, 1, 2$, the involution α_i maps a triangular face to the adjacent triangular face sharing the side opposite to its i-vertex or to itself if there is no such adjacent face. For the cellular embedding on the upper Figure 2.25, we obtain this way:

$$\begin{aligned} \alpha_0 &= (1,2)(3,4)(5,7)(6,8) \\ \alpha_1 &= (1,5)(2,6)(3,7)(4)(8) \\ \alpha_2 &= (1)(2)(3)(4)(5,6)(7,8) \end{aligned}$$

(This is Example 8.2 in [BS85].) For the cellular embedding on the lower Figure 2.25, we obtain:

$$\begin{aligned} \alpha_0 &= (1)(2)(3)(4)(5,6)(7,8) \\ \alpha_1 &= (1,5)(2,6)(3,7)(4)(8) \\ \alpha_2 &= (1,2)(3,4)(5,7)(6,8) \end{aligned}$$

The process can be reversed. Given a δ-map whose three involutions α_i, $i = 0, 1, 2$, may have fixed points, we can construct a cellular embedding of a graph in a surface with

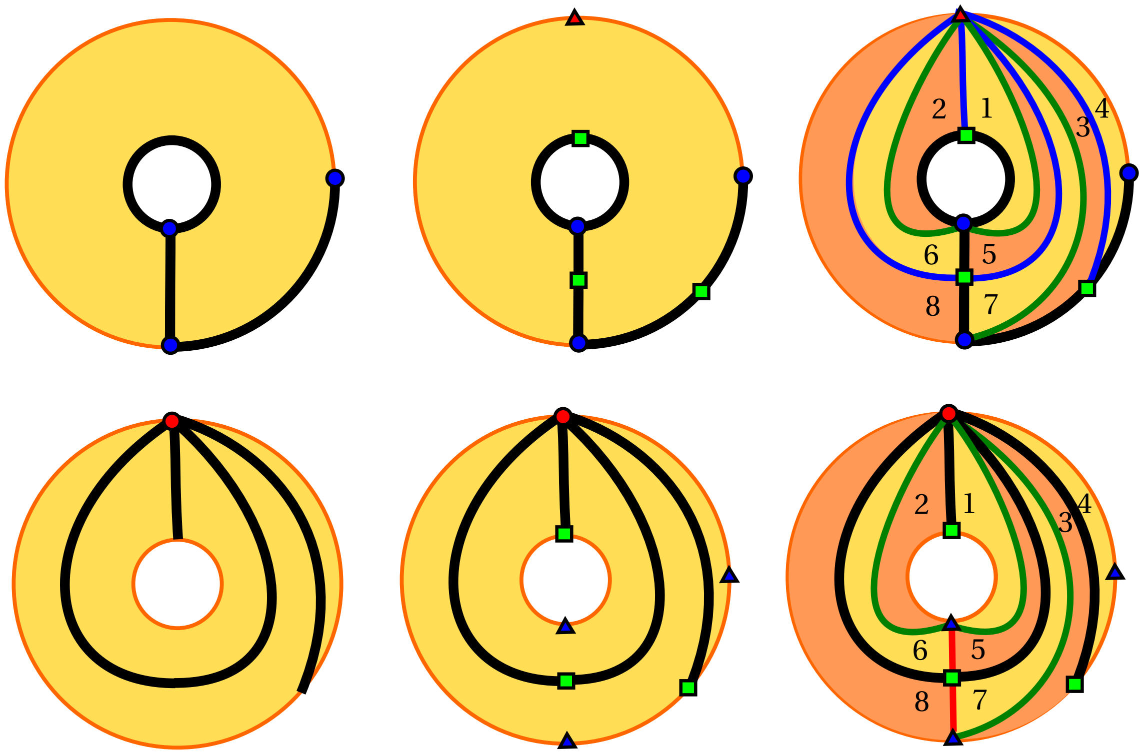


Figure 2.25: Left column, two graphs (with thick lines for the edges and fat dots for the vertices) cellularly embedded in an annulus. Both graphs have three edges, including one loop edge. The lower graph has one vertex and two free ends on the annulus boundary. Middle column, a 2-vertex (small triangle) is introduced in each face and a 1-vertex (small square) is introduced at the midpoint or free-end of each edge. Right column, each face is triangulated by the insertion of edges from its 2-vertex to (the corners of) its incident 0- and 1-vertices.

boundary whose associated δ-map is the given one. Exactly as in Section 2.2.4, we consider a set of triangles, one for each dart, each of whose vertices are labelled in $\mathbb{Z}/3\mathbb{Z}$ with $0,1,2$ respectively. We then glue two triangles along their edge $(i,i+1)$ whenever their corresponding darts are related by α_{i+2}. We obtain this way a topological surface in which the union of the $(0,1)$ edges defines a cellularly embedded graph. This gives yet another equivalent realization of combinatorial maps!

The Euler characteristic of a finite δ-map $M=(D,\alpha_0,\alpha_1,\alpha_2)$ with boundary is defined as

$$\chi(M)=|D|-\sum_{i\in\mathbb{Z}/3\mathbb{Z}}|D/\langle\alpha_i\rangle|+\sum_{i\in\mathbb{Z}/3\mathbb{Z}}|D/\langle\alpha_i,\alpha_{i+1}\rangle|$$

This is indeed the characteristic of the above topological realization: the triangular faces induce a subdivision of the surface with one (triangular) face per dart in D, one (half) edge per cycle of α_0, α_1 and α_2 and one i-vertex per orbit of the subgroup $\langle\alpha_{i+1},\alpha_{i+2}\rangle$. One can also count the number of boundary components of a δ-map with boundary by a relatively simple combinatorial procedure (see [BS85, Th. 8.3]). Defining the orientability of a δ-map with boundary is a bit trickier, though. The characterization in Lemma 2.2.21 does not work anymore. A simple counterexample is the trivial δ-map with boundary $(\{d\}, Id, Id, Id)$ whose realization has a single triangle. It corresponds to a free edge embedded on the boundary of a disc. See Figure 2.26. The subgroup $\langle \varphi, r\rangle = \langle \alpha_1\alpha_0, \alpha_1\alpha_2\rangle$ of the monodromy group (see Remark 2.2.51) trivially acts transitively whilst the map is clearly orientable. Bryant and Singerman defines the orientability of a δ-map

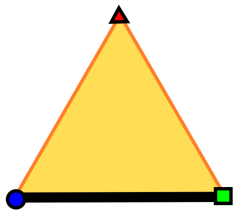


Figure 2.26: The trivial map $(\{d\}, Id, Id, Id)$ has a single free edge, one vertex, one free end and one face.

M with non-empty boundary thanks to a *canonical double* obtained by identifying the boundary of a copy $M' = (D', \alpha'_0, \alpha'_1, \alpha'_2)$ of $M = (D, \alpha_0, \alpha_1, \alpha_2)$ with the boundary of M. The canonical double is the δ-map $(D \cup D', \alpha''_0, \alpha''_1, \alpha''_2)$, where α''_i restricts to α_i and α'_i except when d is a fixed point of α_i – and d' is the corresponding fixed point of α'_i – in which case we set $\alpha''_i(d) = d'$ and $\alpha''_i(d') = d$. This canonical double is a surface without boundary having the same orientability as M. We can now check its orientability by the characterisation of Lemma 2.2.21. Having defined the number B of boundary components, the characteristic and the orientability, we can set the $genus$ of M to

$$g(M) = \begin{cases} (2 - \chi(M) - |B|)/2 & \text{if } M \text{ is orientable,} \\ 2 - \chi(M) - |B| & \text{otherwise.} \end{cases}$$

The basic operations of Section 2.2.5 can be performed on δ-maps with boundary. Duality is the result of exchanging the role of α_0 and α_2 and edge contraction and deletion are defined as in Definition 2.2.62. The δ-maps of Figure 2.25 are dual to one another.

Exercise 2.3.1. Perform the contraction of the edge $\langle \alpha_0, \alpha_2\rangle 1 = \{1, 2\}$ in the δ-map of the upper Figure 2.25. Do you preserve the topology?

2.3.2 Maps with boundary

It is not clear how to built a pertinent equivalent of δ-maps with boundary that would lead to a nice notion of maps with boundary. In particular, trying to extend or invert the functor $\delta : \mathscr{M} \to \mathscr{D}$ to obtain a reasonable notion of maps with boundary seems to fail. We rather view a map with boundary as an embedding of a graph in a topological surface with boundary such that each component of the complement of the graph is either an open disc or an annulus with one boundary component. We can describe such an embedding as a cellular embedding in a surface without boundary to which we remove an open disc in the interior of some of the faces of the embedding. We thus obtain the following

Definition 2.3.2. A **map with boundary** is a pair (M,b) where M is a map and b : $\mathbf{F}(M) \to \{0,1\}$ is a **boundary indicator** defined over the set of faces of M. A face F of M such that $b(F)=0$ is called a **face** of (M,b) and a **perforated** (or **punctured**) **face** otherwise. The set of faces and perforated faces of (M,b) are denoted $\mathbf{F}(M,b)$ and $B(M,b)$ respectively. The **Euler characteristic** of a finite map with boundary (M,b) is by definition $\chi(M)=|V(M)|-|E(M)|+|\mathbf{F}(M,b)|$. The graph, **orientability** and **connectedness** of (M,b) are defined similarly to the corresponding properties for M. The **genus** $g(M,b)$ of (M,b) is defined by

$$\chi(M,b)+|B(M,b)|=\begin{cases} 2-2g(M,b) & \text{if } M \text{ is orientable,}\\ 2-g(M,b) & \text{otherwise.}\end{cases}$$

The above boundary indicator $b:\mathbf{F}(M)\to\{0,1\}$ can be equivalently defined over oriented faces as long as each oriented face gets the same value as its opposite face, i.e., $b(F(d))=b(F(\alpha_0(d)))$ for each flag d of M.

Example 2.3.3. Any map M is identified with the map with (empty) boundary $(M,0)$.

Example 2.3.4. A map with boundary of the form $(M,1)$ has perforated faces only. It corresponds to the intuitive notion of a ribbon graph as on Figure 2.27. There is no

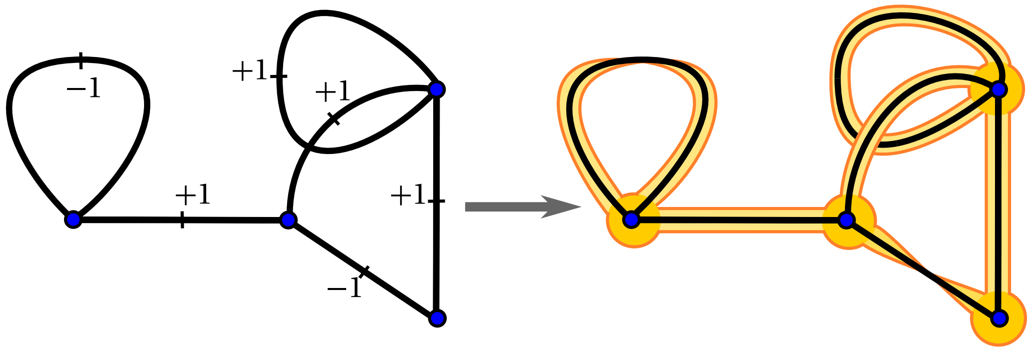


Figure 2.27: A map with perforated faces only.

equivalent in the above formalism of δ-maps with boundary. However, while duality is easily defined in this last formalism, it is not clear how to define a general notion of duality for maps with boundary, especially for maps of the form $(M,1)$. On the other hand, when duality is required we can still restrict ourselves to maps with boundary that can be interpreted as δ-maps with boundary. This simply means that the perforated faces should be pairwise vertex disjoint. With this condition, the corresponding graph embedding can be isotoped so that the surface boundary is included in the graph. The components of the graph complement are now (open) discs as required for the topological counterpart of a δ-map with boundary.

Taking into account the presence of perforated faces, we get a natural notion of morphism for maps with boundary.

Definition 2.3.5. A **morphism** $(f,\omega):(M,b)\to(M',b')$ from a map with boundary (M,b) to a map with boundary (M',b') is a map morphism $(f,\omega):M\to M'$ in the sens of Definition 2.2.24 such that a face F of M is perforated in (M,b) if and only if its image by the flag extension $\bar{f}$ of (f,ω) (see Corollary 2.2.32) is a perforated face, i.e., $b(F)=b'(\bar{f}(F))$.

The operations of edge contraction, deletion, subdivision and face subdivision as exposed in Section 2.2.5 easily extend to maps with boundary once the modification of the boundary indicator is defined properly.

Definition 2.3.6. Let (M,b) be a map with boundary and let $e=\{a,a^{-1}\}$ be an edge of M. We assume that the considered operations on M do not leave an empty map. For instance, the contraction of e in (M,b) is only defined when the component of e in M has at least two edges. We use a prime to denote the maps or objects relative to the map resulting from the considered operations.

- If e is a non-loop edge with positive signature, the **contraction** of e in (M,b) yields the map with boundary $(M,b)/e=(M/e,b')$ where M/e is given by Definition 2.2.56. The boundary indicator b' is trivially deduced from b using that the faces of M and M/e are in bijection (see the proof of Proposition 2.2.58. Formally, $b'(F'(d))=b(F(d))$ for every flag d of M/e (also considered as a flag of M).

- If e is incident to two distinct faces f_1 and f_2 of M and f_1 is a face of (M,b), then the **deletion** of e yields the map with boundary $(M,b)-e=(M-e,b')$ where $M-e$ is given by Definition 2.2.59. The faces f_1 and f_2 are merged in $M-e$ (see the proof of Proposition 2.2.61) into a single face f and we set $b'(f)=b(f_2)$. Any other face f of M gets the same value in $M-e$ and M, i.e., we set $b'(f)=b(f)$.

- The **subdivision** of e in (M,b) yields the map with boundary $S_e(M,b)=(S_eM,b')$ where S_eM is given by Definition 2.2.67. The boundary indicator b' is trivially deduced from b using that the faces of M and M/e are in bijection.

- If u and v are two flags of M with $F(u)=F(v)$ then the **subdivision** of $F(u)$ from u to v yields a map $S_{(u,v)}(M,b)=(S_{(u,v)}M,b')$ where $S_{(u,v)}M$ is given by Definition 2.2.69. The face $F(u)$ is split into two faces in $S_{(u,v)}M$, namely $F'(u)$ and another face $f'=F'(c,-\epsilon(a))$ according to the notations of Definition 2.2.69. The boundary indicator b' is defined by

$$b'(f)=\begin{cases} b(f) & \text{if } f \text{ is a face of } M \text{ distinct from } F(u) \text{ or from its opposite,}\\ 0 & \text{if } f=F(u) \text{ or its opposite,}\\ b(F(u)) & \text{if } f=f'.\end{cases}$$

  In the case where $F(u)$ is perforated, as illustrated on Figure 2.28, we note that $S_{(v,u)}(M,b)$ only differs from $S_{(u,v)}(M,b)$ in that the perforation of $F(u)$ is placed in $F'(u)$ instead of f'.

It is easily verified that

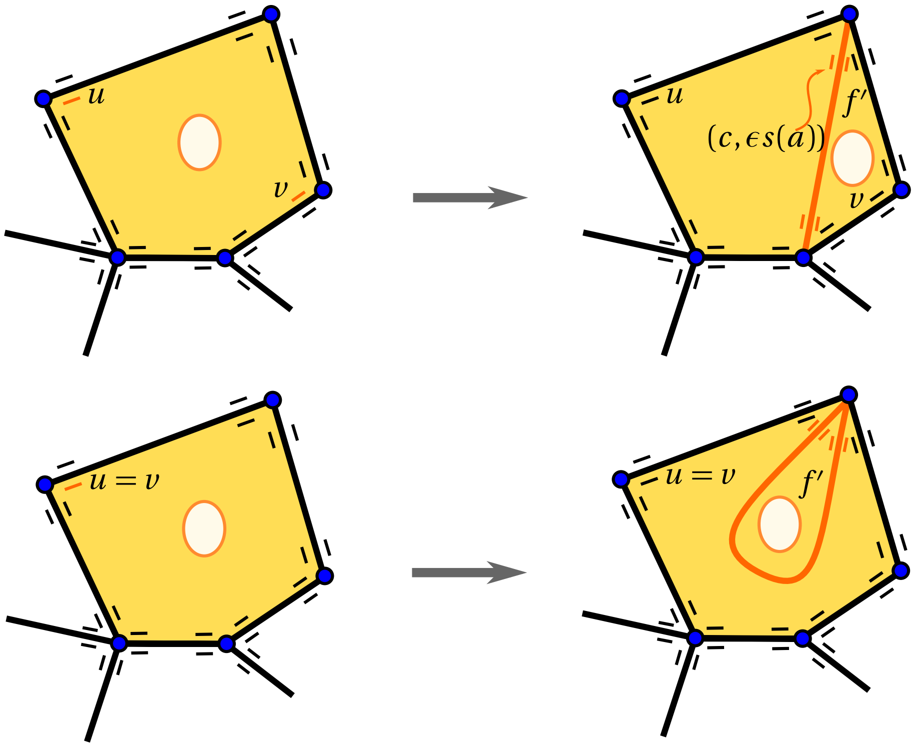


Figure 2.28: Subdivision of a perforated face. Top, The subdivision flags u and v are distinct. Bottom, the case $u = v$.

Proposition 2.3.7. *All the above operations on maps with boundary preserve the connectedness, the orientability, the number of perforated faces and the Euler characteristic.*

Chapter 3

Topology of Combinatorial Surfaces

Contents

Compared to topological surfaces, combinatorial surfaces have the advantage of being of a discrete nature. Combinatorial surfaces such as maps are easier to manipulate, to encode and naturally lead to computations. On the other hand, the topological surface encoded by a map is less apparent and non-isomorphic maps may encode the same topological surface. In other words, the realization functor $\mathscr{M} \to \mathbf{Top}$ is many to one and is not an equivalence. Map morphisms are too rigid to allow for a full recording of topology. The usual way to circumvent this rigidity is to introduce a combinatorial equivalence.

3.1 Classification of Maps

Here and below, we use the term map for a map with (possibly empty) boundary.

Definition 3.1.1. Combinatorial equivalence of maps is the equivalence relation generated by edge and face subdivisions as specified in Definition 2.3.6.

Following Remark 2.1.23, two maps are combinatorially equivalent if and only if any of the two maps can be obtained from the other one by a finite sequence of the operations described in Definition 2.3.6: edge contraction, edge deletion, edge subdivision or face subdivision. By Proposition 2.3.7, equivalent maps share the same connectedness, orientability, number of perforated faces and Euler characteristic. We shall often say that two maps are equivalent when they are combinatorially equivalent.

Example 3.1.2. The **normal form of a sphere** is the map with empty boundary and with a unique loop edge of positive signature. The normal form of a **sphere with** k **boundaries**, or k punctures, $k > 0$, is the map with k edges $c_1, c_2, \dots, c_k$ of positive signature, whose rotation system has a single cycle

$$(c_1, c_1^{-1}, \dots, c_k, c_k^{-1})$$

and whose boundary indicator b is given by

$$b(F(c_1, 1)) = 0, \text{ and } \forall i \in [1, k] : b(F(c_i^{-1}, 1)) = 1$$

See Figure 3.1 for an illustration. A sphere with one boundary, two or three boundaries

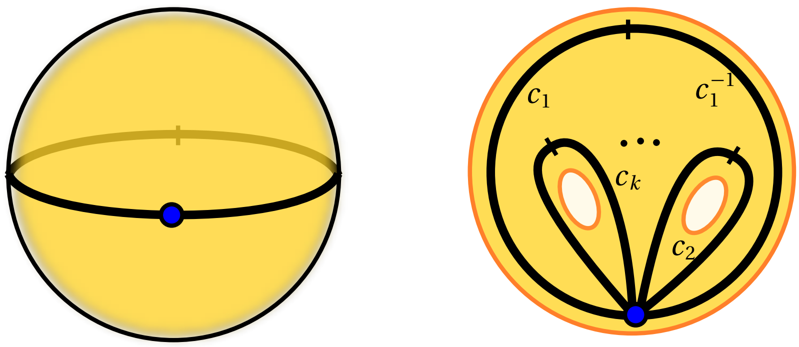


Figure 3.1: Graph embedding corresponding to the canonical form of a sphere (left) and the canonical form of a perforated sphere (right).

is respectively called a **disc**, an **annulus** (or a **cylinder**), or a **pair of pants**.

Example 3.1.3. The normal form of a **connected sum of** g **tori** with k boundaries, $g > 0$ and $k \geq 0$, is the map with $2g + k$ edges $a_1, b_1, a_2, b_2, \dots, a_g, b_g, c_1, \dots, c_k$ of positive signature whose rotation system has a unique cycle

$$\rho_{g,k} = (a_1, b_1^{-1}, a_1^{-1}, b_1, \dots, a_g, b_g^{-1}, a_g^{-1}, b_g, c_1, c_1^{-1}, \dots, c_k, c_k^{-1}) \tag{3.1}$$

and whose boundary indicator b is given by

$$b(F(a_1, 1)) = 0, \text{ and } \forall i \in [1, k] : b(F(c_i^{-1}, 1)) = 1$$

See Figure 3.2 for an illustration.

Example 3.1.4. The normal form of a **connected sum of** g **projective planes** with k boundaries, $g > 0$ and $k \geq 0$, is the map with g edges $a_1, \dots, a_g$ of negative signature and k edges $c_1, \dots, c_k$ of positive signature, whose rotation system has a unique cycle

$$\rho'_{g,k} = (a_1, a_1^{-1}, \dots, a_g, a_g^{-1}, c_1, c_1^{-1}, \dots, c_k, c_k^{-1}) \tag{3.2}$$

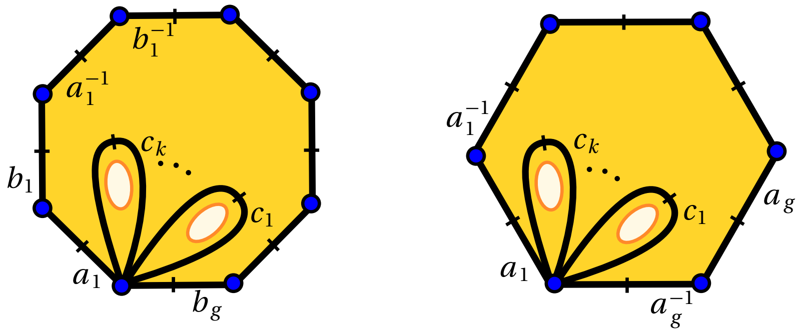


Figure 3.2: A shattered view of the unique face of the canonical form of a connected sum of tori (left) or projective planes (right). All the vertices are identified to a single vertex. Some of them are already identified in order to highlight the perforated faces.

and whose boundary indicator b is given by

$$b(F(a_1,1)) = 0, \text{ and } \forall i \in [1,k] : b(F(c_i^{-1},1)) = 1$$

A (connected sum of one) projective plane with one boundary is also called a **Möbius band**. A connected sum of three projective planes without boundary is also called a **Klein bottle**.

Definition 3.1.5. A sphere, annulus, connected sum of tori, etc, is a map combinatorially equivalent to its corresponding normal form.

Exercise 3.1.6. Show that the map with a single non-loop edge is a sphere or a disc.

Exercise 3.1.7. Show that the normal forms with $k > 0$ boundaries in the above examples 3.1.2 to 3.1.4 are equivalent to maps with perforated faces only.

We observe that spheres and connected sums of tori are orientable while connected sums of projective planes are non-orientable.

Exercise 3.1.8. Compute the Euler characteristic and the genus of a connected sum of g tori with k boundaries and of a connected sum of g projective planes with k boundaries.

We shall prove that every finite map is equivalent to exactly one of the normal forms in Examples 3.1.2, 3.1.3 and 3.1.4.

Lemma 3.1.9. *Every finite connected map is equivalent to a map with a single vertex.*

PROOF. Let M be a finite connected map and let T be a spanning tree of the graph $G(M)$ of M. If M has at least one edge not in T, we may contract the edges of T one after the other, in any order. We obtain this way an equivalent map with a single vertex. If $G(M)$ is a tree, we may contract all of its edges but one and conclude with Exercise 3.1.6
□

The normal form of a sphere, and the maps with a single vertex and a single face, but possibly several perforated faces, are said **reduced**.

Lemma 3.1.10. *Every finite connected map is equivalent to a reduced map.*

PROOF. Let (M,b) be a finite connected map. By the previous lemma we may assume that (M,b) has a single vertex. If every face of M is a perforated face of (M,b), we may subdivide a face to obtain an equivalent reduced map. Otherwise, we note that every face of M corresponds to a vertex of the dual graph $G(M^*)$ which is connected by Proposition 2.2.54. As long as (M,b) has more than one face, the connectivity of the dual graph implies that there exists an edge incident to two distinct faces of M, at least one of which is a face of (M,b). We can delete this edge, thus reducing the number of faces of (M,b). By induction, we obtain an equivalent reduced map. □

The classification of maps relies on a repeated use of face subdivisions and edge deletions. We introduce some concise notations to describe these operations. The sequence of flags of an oriented face will be denoted by a cyclic word on the flags, as in $F(d) = d\varphi(d)\varphi^2(d)\ldots\varphi^{-1}(d)$. We shall not distinguish the notation for a cyclic or a linear word. The distinction should be clear from the context. Let u,v be two flags of an oriented face $F(u) = uYvX$ of a map M, where X,Y are flag words. Following Definitions 2.3.6 and 2.2.69, the subdivision of $F(u)$ from u to v splits $F(u)$ by the introduction of an edge between the heads of u and v. This edge corresponds to an orbit $\langle\alpha_0,\alpha_2\rangle d$ (see Equations (2.10)-(2.12)) of four flags in $S_{(u,v)}M$, where d is the flag following u in its oriented face *after* the face subdivision (see Figure 3.3). The subdivision of

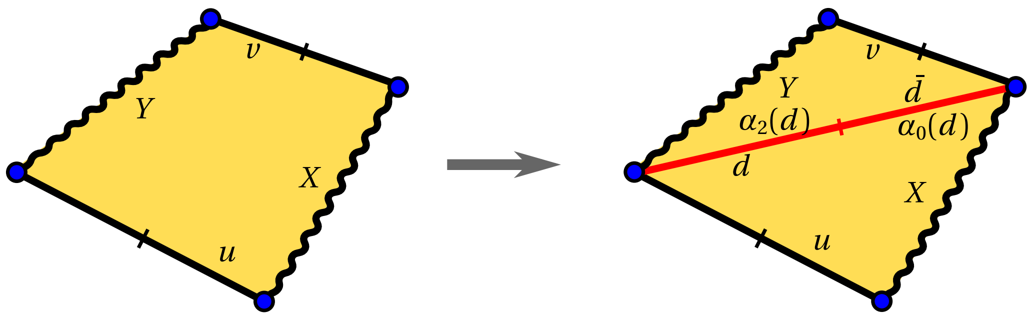


Figure 3.3: Subdivision of the face $XuYv$ from u to v. Four flags $d,\alpha_0(d),\alpha_2(d)$ and $\bar{d} = \alpha_0\alpha_2(d)$ are introduced for the insertion of a new edge splitting $XuYv$.

$F(u) = uYvX$ from u to v is then depicted by the following diagram

$$XuYv \xrightarrow{d=(u,v)} Xud + \bar{d}Yv$$

where we have put $\bar{d} = \alpha_0\alpha_2(d)$. This diagram tells that the oriented face $XuYv$ has been split into two faces, one orientation of which are Xud and $\bar{d}Yv$ respectively. Conversely, if the edge $e = \langle\alpha_0,\alpha_2\rangle d$ is incident to two distinct faces $X'd$ and $\bar{d}Y'$, then the deletion of e is depicted by the diagram

$$X'd + \bar{d}Y' \xrightarrow{-d} X'Y'$$

Finally, if $W = d_1d_2\ldots d_k$ is a word on flags, then W^{-1} shall designate the word $\alpha_0(d_k)\ldots\alpha_0(d_2)\alpha_0(d_1)$. In particular, for a flag u, the face $F(\alpha_0(u))$ can also be denoted $F(u)^{-1}$.

We first state an auxiliary lemma.

Lemma 3.1.11. *Let M be a finite reduced map without boundary. Suppose that $F(u) = F(\bar{u})$ for some flags u and $\bar{u} = \alpha_0\alpha_2(u)$ of M. Then, there exists a flag $v \notin \langle\alpha_0, \alpha_2\rangle u$ such that either*

$$F(u) = u \dots v \dots \bar{u} \dots \bar{v} \dots \tag{3.3}$$

or

$$F(u) = u \dots v \dots \bar{u} \dots \alpha_2(v) \dots \tag{3.4}$$

where each $\dots$ is a possibly empty flag word. Moreover, if M is orientable only 3.3 may occur.

Proof. By hypothesis, we have $F(u) = uX\bar{u}Y$ for some flag words X, Y. Assume for a contradiction that for every flag v in X, either $\bar{v}$ or $\alpha_2(v)$ belongs to X. We consider the set U of flags comprising $\alpha_0(u)$, $\bar{u}$, the flags in X and X^{-1}. We claim that U is stable by α_2; if w occurs in X, then either $\alpha_2(w)$ or $\bar{w}$ occurs in X. In the latter case $\alpha_0(\bar{w}) = \alpha_2(w)$ occurs in X^{-1}, so that in both cases $\alpha_2(w) \in U$. Likewise, w occurs in X^{-1} implies $\alpha_2(w) \in U$. Furthermore, $\alpha_2(\alpha_0(u)) = \bar{u}$ and U is indeed stable by α_2. We also claim that U is stable by α_1. Recall from Remark 2.2.51 that $\varphi = \alpha_1\alpha_0$. If w occurs in X, then $\alpha_1(w) = \alpha_0\alpha_0\alpha_1(w) = \alpha_0\varphi^{-1}(w)$ so that $\alpha_1(w)$ occurs in $(uX)^{-1} = X^{-1}\alpha_0(u)$. If w occurs in X^{-1}, then $\alpha_1(w) = \alpha_1\alpha_0\alpha_0(w) = \varphi\alpha_0(w)$ implying that $\alpha_1(w)$ occurs in $X\bar{u}$. Furthermore $\alpha_1(\alpha_0(u)) = \varphi(u)$ occurs in $X\bar{u}$ and $\alpha_1(\bar{u}) = \varphi\alpha_2(u)$ occurs in $X^{-1}\alpha_0(u)$. In any case, $\alpha_1(w) \in U$.

We conclude from these two claims that U is a union of orbits of $\langle\alpha_1, \alpha_2\rangle$, whilst $u \notin U$. Since the vertices of M are identified with the orbits of $\langle\alpha_1, \alpha_2\rangle$, this contradicts the fact that M has a single vertex. It ensues that X cannot be empty and that for some flag v in X, none of $\bar{v}$ and $\alpha_2(v)$ belongs to X. Since every flag must occur in exactly one of $F(u)$ and $F(\alpha_0(u))$, either $\bar{v}$ and $\alpha_2(v)$ occurs in Y. When M is orientable, its signature must be positive and Equation (2.1) shows that φ preserves the sign of flags. On the contrary, Equation (2.12) shows that α_2 inverts the sign of flags, so that (3.4) cannot occur. □

Exercise 3.1.12. Let φ and β be two permutations of $[1, n]$ such that φ is cyclic, β is a fixed point free involution and $\varphi\beta$ is cyclic. Let $k \in [1, n-1]$ such that $\varphi^k(1) = \beta(1)$. Show that there exists j, l with $1 \le j < k < \ell < n$ satisfying $\beta\phi^j(1) = \phi^\ell(1)$.

Exercise 3.1.13. Let φ and β be as in the previous exercise. Show that n must be a multiple of 4.

Exercise 3.1.14. Using the result of Exercise 3.1.12, give another proof of Lemma 3.1.11 when M is orientable.

3.1.1 Classification of orientable maps

Theorem 3.1.15. *Every finite orientable map without boundary is either a sphere or a connected sum of tori.*

PROOF. Let M be a finite orientable map without boundary. We need to prove that M is equivalent to the normal form of a sphere or of a connected sum of tori. By Lemma 3.1.10, we may assume that M is reduced. The theorem is trivially true if M has the normal form of a sphere, so that we can further assume that M has a single vertex and a single face. Let $F(u)$ be one orientation of this face, where u is a flag of M. According to Lemma 3.1.11, we may write

$$F(u) = uXvY\bar{u}Z\bar{v}WT_k$$

where each of X, Y, Z and W is a possibly empty flag word and where

$$T_k = u_1v_1\bar{u}_1\bar{v}_1u_2v_2\bar{u}_2\bar{v}_2\dots u_kv_k\bar{u}_k\bar{v}_k$$

for some $k \geq 0$ (by convention T_0 is the empty word) and some pairwise distinct flags u_i, v_i. We apply the following sequence of operations as illustrated on Figure 3.4.

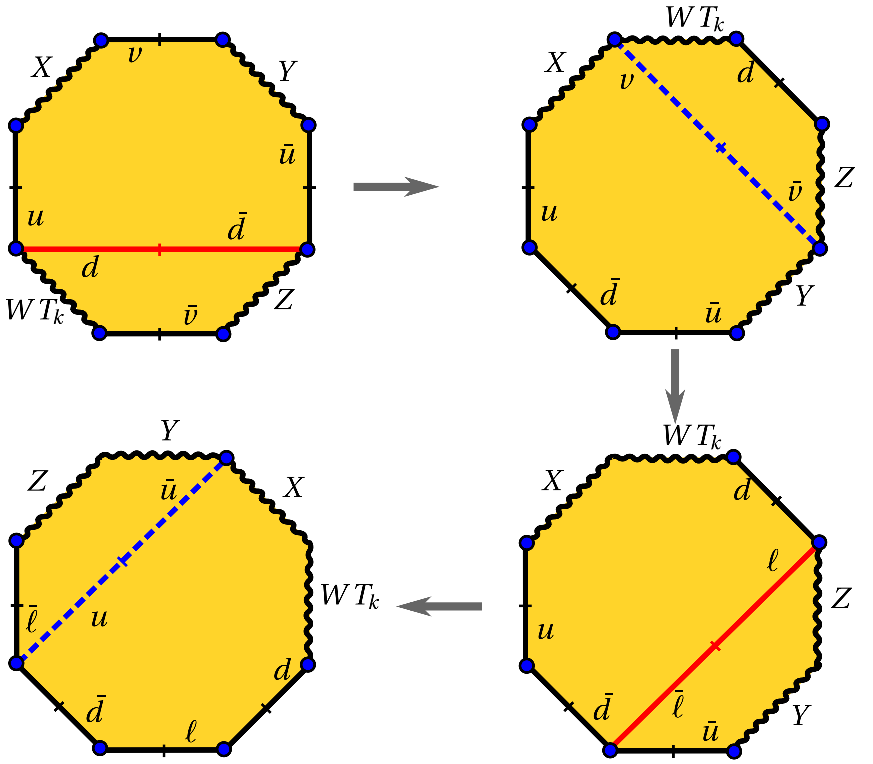


Figure 3.4: Upper left, A schematic view of the unique face of M. The map M is applied a face subdivision by inserting a (red plain) edge. Upper right, An equivalent view of the resulting two faces. The deletion of the dashed (blue) edge merges the two faces in a different way. We further apply a face subdivision (lower right figure) and an edge deletion (lower left figure) to obtain an equivalent reduced map.

$$Z\bar{v}WT_kuXvY\bar{u} \xrightarrow{d=(\varphi^{-1}(u),\bar{u})} Z\bar{v}WT_kd + \bar{d}uXvY\bar{u} \tag{3.5}$$

$$WT_kdZ\bar{v} + vY\bar{u}\bar{d}uX \xrightarrow{-v} WT_kdZY\bar{u}\bar{d}uX \tag{3.6}$$

$$dZY\bar{u}\bar{d}uXWT_k \xrightarrow{\ell=(d,\bar{u})} \bar{d}uXWT_kd\ell + \bar{\ell}ZY\bar{u} \tag{3.7}$$

$$XWT_kd\ell\bar{d}u + \bar{u}\bar{\ell}ZY \xrightarrow{-u} XWT_kd\ell\bar{d}\bar{\ell}ZY \tag{3.8}$$

We obtain this way a reduced equivalent map M' whose unique face has facial circuit $ZYXWT_{k+1}$, where $T_{k+1} = T_k d\ell\bar{d}\bar{\ell}$. Analogous operations may be repeated if $ZYXW$ is not the empty word. However the face of M' and the face of M have the same degree while $|T_{k+1}| > |T_k|$. It follows by induction on k that M must be equivalent to a reduced map whose unique face has the form T_g for $g = |A|/4$ (cf. Exercise 3.1.13). The corresponding rotation system ρ can be extracted from the arc component of the rotational permutation $r : x \mapsto \varphi(\bar{x})$ since $r(a,\epsilon) = (\rho^\epsilon(a),\epsilon)$. We easily conclude that up to a renaming of the arcs we have $\rho = \rho_{g,0}$ as in (3.1). □

We now turn to the non-orientable case.

3.1.2 Classification of non-orientable maps

Theorem 3.1.16. *Every finite non-orientable map without boundary is a connected sum of projective planes.*

Proof. Let M be a finite non-orientable map without boundary. We need to prove that M is equivalent to the normal form of a connected sum of projective planes. By Lemma 3.1.10, we may assume that M is reduced. The characterisation of orientable surfaces in Lemma 2.2.18 implies that we must have

$$F(x) = xU\alpha_2(x)V$$

for some flag x of M and some flag words U, V. We perform the following face subdivision:

$$VxU\alpha_2(x) \xrightarrow{u=(x,\alpha_2(x))} Vxu + \bar{u}U\alpha_2(x)$$

and, using that $(\bar{u}U\alpha_2(x))^{-1} = \bar{x}U^{-1}\alpha_2(u)$, we next perform the following edge deletion:

$$uVx + \bar{x}U^{-1}\alpha_2(u) \xrightarrow{-x} uVU^{-1}\alpha_2(u)$$

Setting $u_1 = \alpha_2(u)$, we have obtained an equivalent reduced map M_1 whose unique face has facial circuit $X_1u_1\alpha_2(u_1)$ for some flag word $X_1 = VU^{-1}$. We note that this face has the same degree as the face of M. We now assume by induction on k that M is equivalent to a reduced map M_k whose unique face is given by

$$X_kP_k \quad \text{where} \quad P_k = u_1\alpha_2(u_1)u_2\alpha_2(u_2)\dots u_{i_k}\alpha_2(u_{i_k}) \tag{3.9}$$

for some flags u_i, some flag word X_k and where $i_k \geq k \geq 1$. We also require that the degree of the above face is the same as the degree of the face of M. We shall prove that the above form can be chosen such that X_k is the empty word. If this is not the case, let x be the first flag in X_k. Since M_k is reduced, either $\bar{x}$ or $\alpha_2(x)$ occurs in X_k.

I) If $\alpha_2(x)$ occurs in X_k then X_k has the form $xU\alpha_2(x)V$ and we perform the face subdivision:

$$\alpha_2(x)VP_kxU \xrightarrow{u=(\varphi^{-1}(x),\varphi^{-1}\alpha_2(x))} \alpha_2(x)VP_ku + \bar{u}xU$$

followed by the edge deletion:

$$VP_k u\alpha_2(x)+\alpha_0(x)\alpha_2(u)U^{-1} \xrightarrow{-\alpha_2(x)} VP_k u\alpha_2(u)U^{-1}$$

Here we have used that $\alpha_0(\bar{u})=\alpha_2(u)$ and $\overline{\alpha_2(x)}=\alpha_0(x)$. Setting $i_{k+1}=i_k+1$, $u_{i_{k+1}}=u$ and $X_{k+1}=U^{-1}V$, we remark that we have obtained an equivalent map M_{k+1} whose unique face is given by (3.9) at the order $k+1$.

II) If $\bar{x}$ occurs in X_k, i.e., $X_k=x\cdots\bar{x}\cdots$ then by Lemma 3.1.11 there is a flag y such that the pair $(x,\bar{x})$ is crossed with either $(y,\bar{y})$ or $(y,\alpha_2(y))$.

i) In the latter case we can write $X_k=xAyB\bar{x}C\alpha_2(y)D$ for some flag words A,B,C,D. We perform the following face subdivisions and edge deletions.

$$\begin{array}{rcl}
C\alpha_2(y)DP_kxAyB\bar{x} & \xrightarrow{d=(\varphi^{-1}(x),\bar{x})} & C\alpha_2(y)DP_kd+\bar{d}xAyB\bar{x}\\
DP_kdC\alpha_2(y)+\alpha_0(y)A^{-1}\alpha_0(x)\alpha_2(d)\alpha_2(x)B^{-1} & \xrightarrow{-\alpha_2(y)} & DP_kdCA^{-1}\alpha_0(x)\alpha_2(d)\alpha_2(x)B^{-1}\\
\alpha_2(d)\alpha_2(x)B^{-1}DP_kdCA^{-1}\alpha_0(x) & \xrightarrow{u=(\varphi^{-1}(d),\alpha_0(x))} & \alpha_2(d)\alpha_2(x)B^{-1}DP_ku+\bar{u}dCA^{-1}\alpha_0(x)\\
\alpha_2(x)B^{-1}DP_ku\alpha_2(d)+\alpha_0(d)\alpha_2(u)xAC^{-1} & \xrightarrow{-d} & \alpha_2(x)B^{-1}DP_ku\alpha_2(u)xAC^{-1}
\end{array}$$

Setting $i_{k+1}=i_k+1$, $u_{i_{k+1}}=u$ and $X_{k+1}=xAC^{-1}\alpha_2(x)B^{-1}D$, we are brought back to the form (3.9) at the order $k+1$.

ii) If $(x,\bar{x})$ is crossed with $(y,\bar{y})$, we have $X_k=xAyB\bar{x}C\bar{y}D$ for some flag words A,B,C,D. We write $DP_k=Pz\alpha_2(z)$, so that $z=u_{i_k}$. We then apply the following alternate sequence of face subdivisions and edge deletions.

$$\begin{array}{rcl}
AyB\bar{x}C\bar{y}Pz\alpha_2(z)x & \xrightarrow{d=(z,x)} & AyB\bar{x}C\bar{y}Pzd+\bar{d}\alpha_2(z)x\\
C\bar{y}PzdAyB\bar{x}+x\bar{d}\alpha_2(z) & \xrightarrow{-\bar{x}} & C\bar{y}PzdAyB\bar{d}\alpha_2(z)
\end{array}$$

$$\begin{array}{rcl}
\alpha_2(z)C\bar{y}PzdAyB\bar{d} & \xrightarrow{u=(\varphi^{-1}(z),\bar{d})} & \alpha_2(z)C\bar{y}Pu+\bar{u}zdAyB\bar{d}\\
C\bar{y}Pu\alpha_2(z)+\alpha_0(z)\alpha_2(u)\alpha_2(d)B^{-1}\alpha_0(y)A^{-1}\alpha_0(d) & \xrightarrow{-\alpha_2(z)} & \\
 & & C\bar{y}Pu\alpha_2(u)\alpha_2(d)B^{-1}\alpha_0(y)A^{-1}\alpha_0(d)
\end{array}$$

$$\begin{array}{rcl}
C\bar{y}Pu\alpha_2(u)\alpha_2(d)B^{-1}\alpha_0(y)A^{-1}\alpha_0(d) & \xrightarrow{t=(\varphi^{-1}\alpha_2(d),\alpha_0(d))} & \\
 & & C\bar{y}Pu\alpha_2(u)\alpha_2(d)t+\bar{t}B^{-1}\alpha_0(y)A^{-1}\alpha_0(d)\\
Pu\alpha_2(u)\alpha_2(d)tC\bar{y}+yB\bar{d}\alpha_2tdA & \xrightarrow{-y} & Pu\alpha_2(u)\alpha_2(d)tCB\bar{d}\alpha_2tdA
\end{array}$$

$$\begin{array}{rcl}
\alpha_2(t)dAPu\alpha_2(u)\alpha_2(d)tCB\bar{d} & \xrightarrow{v=(\alpha_2(u),\bar{d})} & \alpha_2(t)dAPu\alpha_2(u)v+\bar{v}\alpha_2(d)tCB\bar{d}\\
dAPu\alpha_2(u)v\alpha_2(t)+\alpha_0(t)\alpha_2(v)\alpha_2(d)B^{-1}C^{-1} & \xrightarrow{-\alpha_2(t)} & dAPu\alpha_2(u)v\alpha_2(v)\alpha_2(d)B^{-1}C^{-1}
\end{array}$$

$$\begin{array}{rcl}
dAPu\alpha_2(u)v\alpha_2(v)\alpha_2(d)B^{-1}C^{-1} & \xrightarrow{w=(\alpha_2(v),\varphi^{-1}(d))} & dAPu\alpha_2(u)v\alpha_2(v)w+\bar{w}\alpha_2(d)B^{-1}C^{-1}\\
APu\alpha_2(u)v\alpha_2(v)wd+\bar{d}\alpha_2(w)CB & \xrightarrow{-d} & APu\alpha_2(u)v\alpha_2(v)w\alpha_2(w)CB
\end{array}$$

We now set $u_{i_k}=u$, $u_{i_k+1}=v$, $u_{i_k+2}=w$ and $i_{k+1}=i_k+2$. We have thus obtained an equivalent reduced map whose face is given by the form (3.9) at the order $k+1$, where $X_{k+1}=CBAD$. Note that the u_{i_k} may change from step k to step $k+1$.

In any case i_k is strictly increasing and we conclude by induction on k that M is equivalent to a reduced map whose unique face is given by the form $u_1\alpha_2(u_1)u_2\alpha_2(u_2)\dots u_g\alpha_2(u_g)$, where g is the number of edges of M. Since $\varphi(u_i)=\alpha_2(u_i)$, it follows from (2.1) and (2.12) that the signature of the arc component of u_i is negative and that all the u_i have the same sign. Considering the opposite face $F(u_1)^{-1}$ (and re-indexing) if necessary, we may assume that the u_i have positive sign. We then infer from $\varphi(u_i)=\alpha_2(u_i)$ and $\varphi\alpha_2(u_i)=u_{i+1}$, that the corresponding rotation system ρ satisfies $\rho(a_i)=a_i^{-1}$ and $\rho(a_i^{-1})=a_{i+1}$ where a_i is the arc component of u_i. In other words, we have $\rho=\rho'_{g,0}$ as in (3.2) and we recognize the normal form of a connected sum of g projective planes.
□

3.1.3 Classification of maps with boundary

Lemma 3.1.17. *Every finite reduced map (M,b) with $k>0$ perforated faces is equivalent to a reduced map whose rotation system satisfies for all $i\in[1,k-1]$:*

$$\forall i\in[1,k]:\rho(c_i)=c_i^{-1} \qquad \textit{and} \qquad \forall i\in[1,k-1]:\rho(c_i^{-1})=c_{i+1}$$

where $c_1,\dots,c_k$ are arcs with positive signature. If (M,b) is non-orientable we can further impose that $\rho(c_k^{-1})=a$ for some arc a such that $F(a,1)=F(a,-1)$. In particular, setting $d_i=(c_i,1)$ and $x=(a,1)$, the unique non-boundary face of (M,b) is given in the non-orientable case by

$$F(d_1)=d_1d_2\dots d_k xY\alpha_2(x)Z$$

for some flag words Y,Z.

Proof. By connectedness of the dual map, there is an edge e incident to the unique face of (M,b) and to a perforated face. Deleting that edge provides an equivalent map (M_0,b_0) with perforated faces only. Let $F(u)$ be an oriented face of (M_0,b_0). If $k=1$, the equivalent map $S_{(u,u)}(M_0,b_0)$ satisfies the conditions in the first part of the lemma. If $k>1$, by connectedness of the dual map, there is an edge e_0 incident to $F(u)$ and to another distinct perforated face. We consider the equivalent map $(M_1,b_1)=S_{(u,u)}(M_0,b_0)-e_0$. The face subdivision $S_{(u,u)}$ introduces a new edge $\{c_1,c_1^{-1}\}$ with positive signature. Let $d_1=\varphi(u)$, where φ is the facial permutation of M_1, so that $d_1=(c_1,\epsilon)$ for some $\epsilon\in\{-1,1\}$. From the definition of a face subdivision 2.3.6, we have $\rho_1^{\epsilon}(c_1)=c_1^{-1}$ where ρ_1 is the rotation system of M_1. Changing the roles of c_1 and c_1^{-1} if necessary, we may assume that $\rho_1(c_1)=c_1^{-1}$. We note that (M_1,b_1) has perforated faces only. We can now assume that for some $1\le i<k$ the map (M,b) is equivalent to a reduced map (M_i,b_i) whose k faces are perforated faces and whose rotation system ρ_i satisfies

$$\forall j\in[1,i]:\rho_i(c_j)=c_j^{-1} \qquad \text{and} \qquad \forall j\in[1,i-1]:\rho_i(c_j^{-1})=c_{j+1} \tag{3.10}$$

where the c_js have positive signature. In particular, $(c_j,-1)$ is the only flag of $F(c_j,-1)$ and $F(c_1,1)=F(c_2,1)=\dots=F(c_i,1)$. By connectedness of the dual map, there is an edge $e_i\neq\{c_j,c_j^{-1}\}$ for $j\in[1,i]$ such that e_i is incident to $F(c_1,1)$ and to another distinct face. We then consider the equivalent map $(M_{i+1},b_{i+1})=S_{(d_i,d_i)}(M_i,b_i)-e_i$ where $d_i=(c_i,1)$. It is easily seen that the rotation system of M_{i+1} satisfies (3.10) at the order $i+1$. By

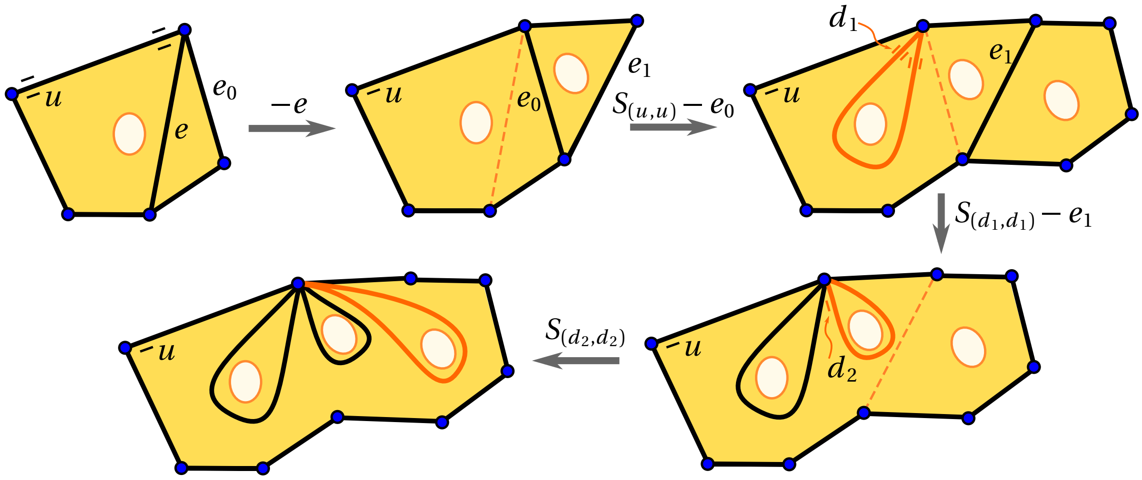


Figure 3.5: Upper left, The unique face of (M, b) shares an edge e with a perforated face. The deletion of this edge gives a map with perforated faces only. Then, each perforated face is replaced in turn by a degree one face which is the result of a face subdivision introducing a loop edge. This subdivision also introduces a non-perforated face. If this face is adjacent to another boundary face, the two are merged by the deletion of a common edge. The process is repeated until all the perforated faces have been transformed into a sequence of degree one perforated faces.

induction on i we infer that (M, b) is equivalent to a reduced map (M_{k-1}, b_{k-1}) satisfying (3.10) at the order $i = k-1$. The face subdivision $S_{(v,v)}$, with $v = (c_{k-1}, 1)$ introduces a new edge $\{c_k, c_k^{-1}\}$ with positive signature. We eventually set $(M_k, b_k) = S_{(w,w)}(M_{k-1}, b_{k-1})$ where $w = (c_k, 1)$. It is easily seen that (M_k, b_k) allows to conclude the first part of the lemma.

When M is non-orientable, we consider a spanning tree T of the dual map of the above map M_0. We delete in M_0 the primal of the edges of T to obtain a reduced map N without boundary. The characterisation of orientable surfaces in Lemma 2.2.18 implies that the unique face of N has an orientation of the form $xX\alpha_2(x)Y$. We consider x as a flag in M_0 and set $u = \varphi_0^{-1}(x)$, where φ_0 is the facial permutation of M_0. The same proof as above, starting with this specific u leads to the second part of the lemma. In particular, since at each step $F(x) = F(\alpha_2(x))$, the deleted edge e_i never contains x that thus subsists until the last step. □

Proposition 3.1.18. *Every finite map with with $k \geq 0$ perforated faces is either a sphere, a connected sum of tori or a connected sum of projective planes, each with k boundaries.*

Proof. Let (M, b) be a finite map with k boundaries. We may suppose by Lemma 3.1.17 that the unique non-perforated face of (M, b) has the form

$$F(u) = d_1 d_2 \dots d_k X$$

where each $F(\bar{d}_i)$ is a degree one perforated face and where X is a flag word. If M is orientable, we can further assume that for any flag x occurring in X, the flag $\bar{x}$ also occurs in X. The proof of Theorem 3.1.15 can be repeated verbatim if we replace T_i by

$T_i' = d_1 d_2 \ldots d_k T_i$. We thus conclude that (M, b) is equivalent to the normal form of a connected sum of tori with k boundaries.

If M is non-orientable, we can suppose by Lemma 3.1.17 that $X = xY\alpha_2(x)Z$ for some flag words Y, Z. The equivalent map $(M', b') = S_{(x,\alpha_2(x))}(M, b) - x$ has a unique non-perforated face of the form $d_1 d_2 \ldots d_k u_1 \alpha_2(u_1) X'$. We may now apply the proof of Theorem 3.1.16, replacing P_k in (3.9) by $P_k' = d_1 d_2 \ldots d_k P_k$, to show that (M', b') is equivalent to the normal form of a connected sum of projective planes with k boundaries. □

Corollary 3.1.19. *Two finite maps are combinatorially equivalent if and only if they have the same orientability, the same Euler characteristic and the same number of perforated faces.*

PROOF. The condition is necessary by Proposition 2.3.7. On the other hand, it is immediate that each normal form in Examples 3.1.2, 3.1.3 and 3.1.4 is uniquely determined by its orientability, Euler characteristic and number of perforated faces. The condition is thus sufficient by Proposition 3.1.18. □

3.1.4 Bibliographical notes

A classification of surfaces based on normal forms appears in a paper by H. R. Brahana [Bra21]. Such normal forms were already known to M. Dehn and P. Heegaard in 1907 [DH07] [Sti87, p. 52]. Brahana's paper deals with system of curves cutting a surface into a polygon, called a polygon schema. A sequence of transformations is proposed to reduce any system of curves, or the corresponding polygonal schema, to a normal form. Brahana's transformations can be interpreted as a sequence of cut and paste operations on the polygonal schema. The first operations provides a reduced polygonal schema in which all the vertices get identified to a single vertex on the surface. The reduced polygonal schema is further rearranged to correspond to a sequence of handles and cross-caps. In the non-orientable case, a last step is necessary to make the cross-caps absorb the handles. This last step is sometimes referred to as Dyck's theorem and denoted by $P\#T = P\#P\#P$, expressing a relation in the monoid structure on surfaces.

This is the usual way surface classification is presented nowadays [Mas77, Sti93]. Although considered as a combinatorial approach at the time of Brahana, this classification relies on the existence of a polygonal schema, a question related to point-set topology. A formal proof for the existence of a polygonal schema had to wait for the fact that any compact surface can be triangulated, a result due to T. Radó [Rad25, DM68, Moi77, Tho92].

A purely combinatorial approach was developed by Tutte [Tut73, Tut01, Ch. X]. Tutte's formalism is equivalent to δ-maps except that a δ-map $(D, \alpha_0, \alpha_1, \alpha_2)$ is encoded by Tutte as the quadruplet (D, θ, ϕ, P) where $\theta = \alpha_2$, $\phi = \alpha_0$ and $P = \alpha_1\alpha_2$. The classification also results from the existence of canonical forms and Tutte uses the same sequence of operations as Brahana's. This is not the sequence that we have used in these notes. The reason for our specific sequence will be more apparent when dealing with the computation of canonical systems of loops in Section 4.3. C. Bonnington and C. Little [BL92] also propose a proof of the classification in the formalism of cubic graphs endowed with a three coloring of their edges. Some of those graphs are called *gems* and correspond to δ-maps. In this correspondence the vertices of a gem are the darts of the δ-map and each color $i \in \{0,1,2\}$ encodes the involution α_i. The cut and paste operations are replaced in this framework by cancellation and creation of *dipoles*. However, the dipole operations are only stable on the set of three colored cubic graphs, not on the set of gems. The meaning of the classification is thus less transparent in this formalism.

Other proofs of the classification of surfaces can be found in the books of Gross and Tucker [GT87, Sec. 3.3], Mohar and Thomassen [MT01, Ch. 3] or Zieschang et al. [ZVC80, Sec. 3.2].

3.2 Homotopy

A **path**, **loop** or **circuit** of a map is a path, loop or circuit of its graph as in Definition 1.2.1. The notion of path deformation via elementary homotopies should now take into account that a path can be deformed inside a face since a face has the topology of a disc. This leads to the following definition:

Let f be an oriented face of a map (M,b). Recall from Definition 2.2.4 that the facial circuit ∂f is the circuit of arcs around f. If $\partial f = u \cdot v^{-1}$, where u is a possibly constant subpath of ∂f, then u and v are said *complementary subpaths*.

Definition 3.2.1. Let (M,b) be a map. An **elementary homotopy** in a path γ of (M,b) consists either in adding or removing a spur in γ, or in replacing a subpath of γ that is also a subpath of a facial circuit by its complementary subpath. In other words, if $\gamma = \lambda \cdot u \cdot \mu$ and $\partial f = u \cdot v^{-1}$ then γ is transformed into $\lambda \cdot v \cdot \mu$ by elementary homotopy. A **free elementary homotopy** is an elementary homotopy applied to any of the path representatives of a circuit. The **homotopy** relation is the transitive closure of elementary homotopies. Likewise, **free homotopy** is the transitive closure of free elementary homotopies. We write $\gamma \sim \lambda$ if γ and λ are homotopic paths and $\gamma \overset{free}{\sim} \lambda$ when they are freely homotopic circuits. A loop or circuit (freely) homotopic to a constant path is said **contractible**. If the last vertex of a path γ coincides with the first vertex of a path λ, their **concatenation** is the path $\gamma \cdot \lambda$ whose arc sequence is the the arc sequences of γ followed by the arc sequence of λ.

Remark 3.2.2. For maps without boundary the homotopy relation is generated by the second type of homotopies, replacing a piece of facial walk by a complementary subpath. Indeed, the addition of a spur $u \cdot v \sim u \cdot a \cdot a^{-1} \cdot v$ can be obtained via elementary homotopies of the above type: if $\partial f = a \cdot w$, then $u \cdot v \sim u \cdot a \cdot w \cdot v \sim u \cdot a \cdot a^{-1} \cdot v$. Here, the first elementary homotopy replaces the constant path $(o(a))$ by $a \cdot w$ and the second elementary homotopy replaces w by the complementary subpath a^{-1}.

Homotopy versus free homotopy

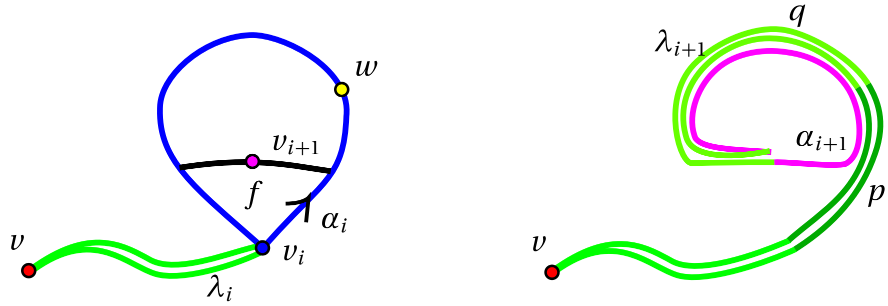


Figure 3.6: Left, v is the common basepoint of α and β while v_i and v_{i+1} are the basepoints of α_i and α_{i+1} respectively, and w is the common basepoint of α'_i and α'_{i+1}. The elementary homotopy $\alpha'_i \to \alpha'_{i+1}$ is supposed to use complementary subpaths of ∂f. Right, The loop $\lambda_{i+1}\alpha_{i+1}\lambda_{i+1}^{-1}$.

Lemma 3.2.3. *Two loops α and β with a common basepoint on a map (M, b) are* freely *homotopic if and only if there exists a loop ℓ such that α and $\ell \cdot \beta \cdot \ell^{-1}$ are homotopic.*

Proof. Let n be the number of free elementary homotopies separating α from β. We thus have for some circuits α_i, $i = 1, n-1$, that $\alpha = \alpha_0 \overset{free}{\to} \alpha_1 \overset{free}{\to} \ldots \overset{free}{\to} \alpha_n = \beta$ where each arrow is a free elementary homotopy. We claim that there exist paths $\lambda_1, \ldots, \lambda_n$ such that $\alpha_0 {\sim} \lambda_1\alpha_1\lambda_1^{-1} {\sim} \ldots {\sim} \lambda_n\alpha_n\lambda_n^{-1}$. Indeed, setting λ_0 to the constant path, we can recursively define λ_i as follows. Since $\alpha_i \overset{free}{\to} \alpha_{i+1}$ we have an elementary homotopy $\alpha'_i \to \alpha'_{i+1}$ for some cyclic permutations of α_i and α_{i+1} respectively. If p is the subpath of α_i from its basepoint to the basepoint of α'_i, we have $\alpha_i \sim p \cdot \alpha'_i \cdot p^{-1}$. Similarly, $\alpha_{i+1} \sim q \cdot \alpha'_{i+1} \cdot q^{-1}$ for some subpath q of α'_i. Assuming $\alpha_0 \sim \lambda_i\alpha_i\lambda_i^{-1}$ we can thus write $\alpha_0 \sim \lambda_{i+1}\alpha_{i+1}\lambda_{i+1}^{-1}$ with $\lambda_{i+1} = \lambda_i.p \cdot q^{-1}$. See Figure 3.6. Choosing $\ell = \lambda_n$ we may concludes the lemma. $\square$

Corollary 3.2.4. *Let α and β be two circuits and let v be a vertex of a map (M, b). The circuits α and β are freely homotopic if and only if for any paths p and q from v to the basepoints of α and β respectively, there exists a loop ℓ such that the loops $p \cdot \alpha \cdot p^{-1}$ and $\ell \cdot q \cdot \beta \cdot q^{-1} \cdot \ell^{-1}$ are homotopic.*

Lemma 3.2.5. *Let α and β be two paths on a map (M, b). We have the three equivalences*

$$\alpha \sim \beta \quad \Leftrightarrow \quad \alpha \cdot \beta^{-1} {\sim} 1 \quad \Leftrightarrow \quad \alpha \cdot \beta^{-1} \overset{free}{\sim} 1$$

Proof. By Lemma 3.2.3, $\alpha \cdot \beta^{-1} \overset{free}{\sim} 1 \Leftrightarrow \alpha \cdot \beta^{-1} \sim 1$. We also have $\alpha \sim \beta \implies \alpha \cdot \beta^{-1} \sim \beta \cdot \beta^{-1} \sim 1$. It remains to prove the reverse implication. Let $\alpha \cdot \beta^{-1} \to \gamma_1 \to \ldots \to \gamma_n = 1$ be a sequence of elementary homotopies transforming $\alpha \cdot \beta^{-1}$ to a constant path. We claim that for each $i = 1, \ldots, n$ there exist paths λ_i, α_i and β_i such that $\alpha_i \cdot \lambda_i \sim \alpha$, $\beta_i \cdot \lambda_i \sim \beta$ and $\alpha_i \cdot \beta_i^{-1} = \gamma_i$. In particular, $\alpha_n = \beta_n = \gamma_n = 1$ and $\alpha \sim \lambda_n \sim \beta$. We prove the claim by induction on i. By the induction hypothesis, we thus have $\alpha_i \cdot \beta_i^{-1} = \gamma_i$. If the elementary homotopy $\gamma_i \to \gamma_{i+1}$ consists in inserting or removing a spur, or in replacing a subpath

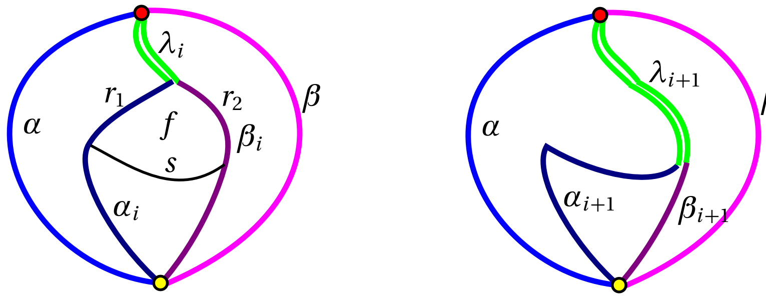


Figure 3.7: The face f is bounded by the circuit $r_1 \cdot r_2 \cdot s^{-1}$.

of α_i to obtain a homotopic path α'_i , we can just set $\alpha_{i+1} = \alpha'_i$, $\beta_{i+1} = \beta_i$ and $\lambda_{i+1} = \lambda_i$. We proceed similarly if $\gamma_i \to \gamma_{i+1}$ consists in modifying β_i only. Otherwise, we can write

$$\alpha_i = p \cdot r_1, \quad \beta_i^{-1} = r_2 \cdot q^{-1}, \quad \gamma_{i+1} = p \cdot s \cdot q^{-1}$$

for some paths p, q, r_1, r_2, s such that $r_1 \cdot r_2 \cdot s^{-1}$ bounds a face (see Figure 3.7). Setting

$$\alpha_{i+1} = p \cdot s, \quad \beta_{i+1} = q, \quad \text{and } \lambda_{i+1} = r_2^{-1} \cdot \lambda_i$$

we obtain $\alpha_{i+1} \cdot \beta_{i+1}^{-1} = \gamma_{i+1}$, $\alpha_{i+1} \cdot \lambda_{i+1} = p \cdot s \cdot r_2^{-1} \cdot \lambda_i \sim p \cdot r_1 \cdot \lambda_i \sim \alpha$ and $\beta_{i+1} \cdot \lambda_{i+1} = q \cdot r_2^{-1} \cdot \lambda_i \sim \beta$ as desired for the induction step. $\square$

3.2.1 The fundamental group of maps

Let v be a vertex of a map (M, b). It is easily checked that the path concatenation $\lambda \cdot \mu$ is homotopic to the path concatenation $\lambda' \cdot \mu'$ whenever $\lambda \sim \lambda'$ and $\mu \sim \mu'$. Hence, similarly to Proposition 1.2.4, the set of homotopy classes of loops with basepoint v is a group for the law of path concatenation. It is called the **fundamental group** of (M, b) based at v and denoted by $\pi_1((M, b), v)$. The free homotopy classes correspond to the conjugacy classes in this group.

Lemma 3.2.6. *Let T be a spanning tree of a connected map (M, b). Then $\pi_1((M, b), v)$ is isomorphic to the group with combinatorial presentation*

$$\Pi = \langle E(M) \mid E(T), \{\partial F\}_{F \in \mathbf{F}(M,b)} \rangle$$

If we denote by C the set of chords of T in $G(M)$ and by r_F the sequence of arcs not in T in the facial circuit of a face F, this group is also isomorphic to

$$\langle C \mid \{r_F\}_{F \in \mathbf{F}(M,b)} \rangle$$

Proof. Recall from Example 1.2.2 the notations $T[v, w]$ for the vw-path in T and $T[v, a]$ for the loop with basepoint v obtained by joining the endpoints of the arc a to v by paths in T. With a little abuse of notation we shall use $T[v, a]$ for the loop or for its homotopy class. If $s = (a_1, a_2, \ldots, a_k)$ is a sequence of arcs, we write $T[v, s]$ for the concatenation $T[v, a_1] \cdot T[v, a_2] \cdots T[v, a_k]$. When s is the sequence of arcs of a loop with basepoint w, we remark that the loop $T[v, s]$ is homotopic in $G(M)$ (by removing spurs) to the loop $T[v,w]\cdot s\cdot T[w,v]$. We assume that each edge has a default orientation so that it can be identified with an arc. We consider the map $\pi : E(M)\to \pi_1((M,b),v), a \mapsto T[v,a]$. If $a\in E(T)$ we know that $T[v,a]$ is contractible. By the preceding remark $T[v,\partial F]\sim T[v,w]\cdot\partial F\cdot T[w,v]$, where ∂F is any path representative of the corresponding circuit. An elementary homotopy replaces ∂F by the empty path in this loop, giving $T[v,\partial F]\sim T[v,w]\cdot T[w,v]\sim 1$. The relations $E(T)\cup\{\partial F\}_{F\in\mathbf{F}(M,b)}$ in the above presentation of Π are thus satisfied in $\pi_1((M,b),v)$. It follows that the map π extends to a group morphism $\pi:\Pi\to\pi_1((M,b),v)$. By the above remark any loop ℓ with basepoint v is homotopic to $T[v,\ell]$, implying that π is onto. It remains to prove that π is one-to-one. For two words w,z in the free group $\langle E(M)\mid -\rangle$ we write $w=_\Pi z$ if w and z are equal as elements in Π. Let $w\in\langle E(M)\mid -\rangle$ such that $\pi(w)$ is contractible. In particular, $T[v,w]$ can be reduced to the constant path by elementary homotopies. Considering $T[v,w]$ as a word in $\langle E(M)\mid -\rangle$, we thus have $T[v,w]=_\Pi 1$. On the other hand, since the edges of T are relations in Π, we have $w\sim_\Pi T[v,w]$, whence $w=_\Pi 1$. □

Example 3.2.7. Since each normal map in examples 3.1.2, 3.1.3 and 3.1.4 has a single vertex and a single (non-perforated) face we easily deduce that the fundamental group of a sphere is trivial while the fundamental group of a sphere with $k>0$ punctures is isomorphic to

$$\langle c_1,\dots,c_k \mid c_1\cdots c_k\rangle \cong \langle c_1,\dots c_{k-1}\mid -\rangle$$

The fundamental group of a connected sum of g tori with $k\geq 0$ punctures is

$$\langle a_1,b_1,\dots,a_g,b_g,c_1,\dots,c_k \mid [a_1,b_1]\cdots[a_g,b_g]c_1\cdots c_k\rangle$$

where $[a,b]=aba^{-1}b^{-1}$. For $k>0$ this is the free group $\langle a_1,b_1,\dots,a_g,b_g,c_1,\dots c_{k-1}\mid -\rangle$. The fundamental group of a connected sum of g projective planes with $k\geq 0$ punctures is

$$\langle a_1,\dots,a_g,c_1,\dots,c_k \mid a_1a_1\cdots a_ga_gc_1\cdots c_k\rangle$$

For $k>0$ this is the free group $\langle a_1,\dots,a_g,c_1,\dots,c_{k-1}\mid -\rangle$.

The homotopy functor From Definition 2.3.5, it is seen that the arc function of a map morphism $\phi:(M,b)\to(M',b')$ extends to a map from the loops of (M,b) to the loops of (M',b'). Moreover, the property that ϕ sends faces onto faces and preserves the boundary indicator implies that homotopic loops are sent to homotopic loops. Indeed, if two loops ℓ,ℓ' are related by an elementary homotopy that replaces a subpath of ℓ by its complementary subpath in a face of (M,b) with ramification index e, then $\phi(\ell)$ is related to $\phi(\ell')$ by a sequence of e elementary homotopies. It ensues that ϕ *induces a group morphism*

$$\phi_*:\pi_1((M,b),v)\to\pi_1((M',b'),\phi(v))$$

The following lemma is immediate.

Lemma 3.2.8. *The association of a map morphism to its induced group morphism is functorial. In other words, the induced group morphism of a composition of map morphisms is the composition of the induced group morphisms.*

Example 3.2.9. Since the elementary homotopies in the graph $G(M)$ of a map (M,b) are homotopies in the map (M,b), we have an epimorphism $\jmath : \pi_1(G(M),v) \twoheadrightarrow \pi_1((M,b),v)$. More generally if $f : H \to G(M)$ is a graph morphism, we have a morphism $\jmath \circ f_* : \pi_1(H,w) \to \pi_1((M,b), f(w))$.

The induced morphisms of basic operations

Although the basic operations of Definition 2.3.6 are not morphisms (they could however be interpreted as such in a more relaxed definition of morphisms), we can define an induced morphism for each basic operation.

- Let e be a non-loop edge of a map (M,b) with basepoint v. The contraction of e yields a map $(M,b)/e$ with an evident mapping C_e from the vertices of (M,b) to the vertices of $(M,b)/e$. This mapping extends to loops as follows. If ℓ is a loop of (M,b) with basepoint v, we define $C_e(\ell)$ as the loop of $(M,b)/e$ with basepoint $C_e(v)$ obtained by removing the occurrences of e in ℓ. Since the edge contraction sends faces to faces, two loops related by an elementary homotopy are send by C_e to homotopic loops. As C_e trivially commutes with path concatenation, we conclude that C_e induces a morphism $(C_e)_* : \pi_1((M,b),v) \to \pi_1((M,b)/e, C_e(v))$.

- The subdivision of any edge e of (M,b) yields a map $S_e(M,b)$. Every loop ℓ of (M,b) maps to a loop $S_e(\ell)$ of $S_e(M,b)$ obtained by replacing each occurrence of e with the sequence of two edges resulting from its subdivision. Similarly to the edge contraction, the mapping S_e induces a morphism $(S_e)_* : \pi_1((M,b),v) \to \pi_1(S_e(M,b),v)$.

- The subdivision of a face in (M,b) from a flag u to a flag v induces an inclusion $S_{(u,v)}$ of the loops of (M,b) into the set of loops of $S_{(u,v)}(M,b)$. In turn, $S_{(u,v)}$ induces a morphism $(S_{(u,v)})_* : \pi_1((M,b),v) \to \pi_1(S_{(u,v)}(M,b),v)$.

We note that the morphisms $(C_e)_*, (S_e)_*$ and $(S_{(u,v)})_*$ are obtained from the mappings C_e, S_e and $S_{(u,v)}$ as quotients by the homotopy relation. It follows that the induced morphism of a composition of such mappings is the composition of the induced morphism. For an appropriate notion of morphisms of map, this just means that the association $f \mapsto f_*$ is functorial.

Lemma 3.2.10. *The above group morphisms $(C_e)_*, (S_e)_*$ and $(S_{(u,v)})_*$ are isomorphisms.*

Proof. The subdivision of an edge e of (M,b) replaces this edge by two edges e_1, e_2. Using the above notations and identifying e_1 with e, we get that S_e inserts occurrences of e_2 in a loop while C_{e_2} removes such occurrences. It follows that $C_{e_2} \circ S_e$ is the identity on loops, whence $(C_{e_2})_* \circ (S_e)_* = Id$. On the other hand, for each loop ℓ of $S_e(M,b)$ the loop $S_e \circ C_{e_2}(\ell)$ is obtained from ℓ by possible insertions of the spur (e_2^{-1}, e_2). It follows that $(S_e)_* \circ (C_{e_2})_* = Id$, implying that $(S_e)_*$ is an isomorphism.

Let e be the edge added by the face subdivision $S_{(u,v)}$. To every loop ℓ of $S_{(u,v)}(M,b)$ we associate the loop $D_e(\ell)$ obtained by substituting each occurrence of e with the complementary subpath in one of the two incident faces. In particular the $D_e \circ S_{(u,v)}$ does not modify any loop of (M,b) while $S_{(u,v)} \circ D_e(\ell)$ is homotopic to ℓ. The mapping

D_e induces a group morphism $(D_e)_*$ and we conclude that $(S_{(u,v)})_*$ and $(D_e)_*$ are inverse to each other.

The proof that $(C_e)_*$ is an isomorphism can be done analogously by providing an adequate mapping from the set of loops of $C_e(M,b)$ to the set of loops of (M,b). A less direct proof follows from Remark 2.1.23. □

As combinatorial equivalence is generated by edge and face subdivisions we conclude that

Corollary 3.2.11. *Combinatorially equivalent maps have isomorphic fundamental groups.*

Exercise 3.2.12. Let M be a map with signature function s and let v be a vertex of M. We extend s to paths in M in the natural way by $s(a_1,a_2,\dots,a_k) = \prod_{i=1}^k s(a_i)$. Show that s quotient to a group morphism $\pi_1(M,v) \to \mathbb{Z}/2\mathbb{Z}$. Describe $s^{-1}(1)$ and $s^{-1}(-1)$.

3.3 Coverings

Recall from Definition 2.2.37 that a covering is a map morphism $\phi : M \to N$ whose restrictions to vertex stars and to facial circuits are bijective. The definition extends to maps with boundary by requiring that perforated faces are sent to perforated faces. Most of the properties proved for graph coverings in Section 1.7 remains valid for map coverings. The definition of a lift for graphs applies verbatim to maps: a lift of a path γ in M is a path δ in N such that $\phi(\delta) = \gamma$. The unique lift property of Lemma 1.7.3 remains valid for maps. We also have

Lemma 3.3.1. *Let $\tilde{\alpha}$ and $\tilde{\beta}$ be the respective lifts in N of two homotopic paths α,β in M. If $\tilde{\alpha}$ and $\tilde{\beta}$ share the same origin then they are homotopic in N.*

Proof. If α and β are related by one elementary homotopy, then so are $\tilde{\alpha}$ and $\tilde{\beta}$. This is obvious for the insertion or deletion of a spur. Otherwise, the replacement in α of a subpath of the facial circuit of a face f by its complementary subpath lifts to the replacement in $\tilde{\alpha}$ of complementary subpaths in a face above f. The lemma now follows by induction on the number of elementary homotopies relating α to β. □

We thus have a right action of $\pi_1(M,v)$ on the fiber above $v \in V(M)$ given by the final endpoint $w.[\alpha]$ of the lift with origin w of a loop α, where $\phi(w) = v$. Similarly to Corollary 1.7.5, we get

Corollary 3.3.2. *If $\phi : (M',b') \to (M,b)$ is a covering, then the induced morphism ϕ_* : $\pi_1((M',b'),w) \to \pi_1((M,b),\phi(w))$ is one-to-one.*

The fundamental group of (M',b') can thus be considered as a subgroup of the fundamental group of (M,b). We also have that every subgroup of $\pi_1((M,b),v)$ can be realized as the fundamental group of a covering.

Proposition 3.3.3. *Let v be a vertex of the connected map (M,b). For every subgroup $U<\pi_1((M,b),v)$ there exists a connected covering $p_U:((M_U,b_U),w)\to((M,b),v)$ with $p_{U*}\pi_1((M_U,b_U),w)=U$.*

Proof. Fix a spanning tree T of $M=(A,\rho,\iota,s)$. We write γ_a for the loop $T[v,a]$ (see the notations in Example 1.2.2). Define $M_U=(A_U,\rho_U,\iota_U,s_U)$ by

- $A_U=A\times\{Ug\}_{g\in\pi_1(G,v)}$,
- $\rho_U(a,Ug)=(\rho(a),Ug)$,
- $\iota_U(a,Ug)=(\iota(a),Ug[\gamma_a])$ and
- $s_U(a,Ug)=s(a)$,

where Ug denotes the right coset representative in $\pi_1(G,v)$ of g with respect to U. M_U is indeed a map: we trivially check that ι_U is a fixed point free involution and that ρ_U is a permutation of A_U. We consider the projection on the first component $p_A:A_U\to A$. The following relations are immediate

$$\iota\circ p_A=p_A\circ\iota_U,\quad \rho\circ p_A=p_A\circ\rho_U,\quad s\circ p_A=s_U$$

Hence, $p_U:=(p_A,\omega=1):M_U\to M$ is a map morphism in the sens of 2.2.24. Let φ and φ_U be the facial permutation of M and M_U respectively. We have

$$\varphi_U((a,Ug),\varepsilon)=\big(\rho_U^{\varepsilon s_U(a,Ug)}(\iota(a),Ug[\gamma_a]),\varepsilon s_U(a,Ug)\big)=\big((\rho^{\varepsilon s(a)}(\iota(a),Ug[\gamma_a]),\varepsilon s(a)\big)$$

Denoting a dart of M_U by (a,ε,Ug), rather than $((a,Ug),\varepsilon)$, we thus write

$$\varphi_U(a,\varepsilon,Ug)=(\varphi(a,\varepsilon),Ug[\gamma_a])$$

As a consequence the cycles of φ_U projects onto cycles of φ. We can now define the boundary indicator b_U by declaring a face of M_U to be perforated if its projection by p_U is a perforated face of (M,b). We obtain this way a morphism $p_U:(M_U,b_U)\to(M,b)$. We claim that p_U is a covering. Since $\langle\rho_U\rangle(a,Ug)=\langle\rho\rangle a\times\{Ug\}$, the restrictions of p_A to vertex stars are bijective. We also check that each cycle of φ_U corresponding to a non-perforated face is sent bijectively to a cycle of φ. For if $(a_1,a_2,\dots,a_k)$ is the facial circuit of a (non-perforated) face of M corresponding to a cycle $\langle\varphi\rangle d$ of length k, and if (d,Ug) is a dart above d, then $\varphi_U^{k+1}(d,Ug)=(d,Ug[\gamma_{a_1}]\dots[\gamma_{a_k}])=(d,Ug)$. (The last equality comes from the fact that $\gamma_{a_1}\cdots\gamma_{a_k}$ is contractible.) It remains to prove that M_U is connected and that $p_{U*}\pi_1((M_U,b_U),w)=U$. The proof is formally identical to the proof of Proposition 1.7.7. □

Example 3.3.4. When $U=\{1\}$ is the trivial group, p_U is the **universal cover** of (M,b). Up to isomorphism, this is the unique simply connected covering of (M,b) (see Corollary 1.7.15).

Example 3.3.5. Let $s:\pi_1((M,b),v)$ be the signature morphism as in Exercise 3.2.12. When (M,b) is non-orientable and $U=s^{-1}(1)$, we obtain the **orientation covering** $p_U:(M_U,b_U)\to(M,b)$ as in example 2.2.38.

Example 3.3.6. When $U=[\pi_1((M,b),v),\pi_1((M,b),v)]$ is the derived subgroup of $\pi_1((M,b),v)$, p_U is the **homology covering**. Choosing for U the subgroup of $\pi_1((M,b),v)$ generated by the squares, we get the $\mathbb{Z}/2\mathbb{Z}$-homology covering. See the next section for the description of homology of maps.

3.3.1 A note on Dehn diagrams

Disc diagrams, also called van Kampen diagrams or Dehn diagrams, are extensively used in the geometric approach of combinatorial group theory. They are especially effective in the theory of small cancellations for the word and conjugacy problems. Intuitively, a disc diagram is the combinatorial counterpart of the following characterization of contractible loops in a topological space X: a loop $\partial D^1 \to X$ is contractible if and only if it extends to a continuous map $D^1 \to X$, where D^1 is the unit disc. To my knowledge discs diagrams are either presented as topological objects [LS77, Chap.V], [MW02] or manipulated somewhat informally [Str90]. It appears that the language of combinatorial maps is perfectly suited for diagrams.

Definition 3.3.7. A **disc diagram** in a map (M, b) with arc set A is a disc map D with arc set B together with two mappings $f : B \to A$, called a *labelling*, and $g : F(D) \to F(M, b)$ such that

- for every arc $a \in B$, $f(a^{-1}) = f(a)^{-1}$,
- for every oriented face $F \in F(D)$, $g(F^{-1}) = g(F)^{-1}$ and the facial circuit of F is labelled with the facial circuit of its image: $f(\partial F) = \partial g(F)$.

It is easily checked that f, g induce a vertex mapping $V(D) \to V(M, b)$ that commutes with the arc origin.

Definition 3.3.8. A diagram is **reduced** if the following situation does *not* occur: There exists two loops ℓ, ℓ' in D with the same basepoint such that

- ℓ and ℓ' are the facial circuits of two distinct faces $F, F' \in F(D)$ consistently oriented,
- $f(\ell) = f(\ell')^{-1}$,
- $g(F) = g(F')^{-1}$.

There is an analogous notion of **annular diagrams**, replacing the disc D by an annulus. The main properties of diagrams are

Proposition 3.3.9 (van Kampen, 1933)**.** *A loop ℓ is contractible in (M, b) if and only if there exists a reduced disc diagram in (M, b) such that the facial circuit of its perforated face is labelled with ℓ.*

PROOF. We first prove the existence of a disc diagram. Let $\ell_0 = 1 \to \ell_1 \to \cdots \to \ell_k = \ell$ be a sequence of k elementary homotopies attesting the contractibility of ℓ. By induction on k, we may assume the existence of a disc diagram bounded by ℓ_{k-1} and defined by two mappings (f, g). There are three cases to consider.

- If $\ell_{k-1} \to \ell_k$ consists in adding a spur aa^{-1}, then we can form a disc diagram for ℓ_k by attaching a pendant edge (formally, the inverse of an edge contraction) to the boundary of D and extend f by sending that edge to $\{a, a^{-1}\}$.

- If $\ell_{k-1} \to \ell_k$ consists in removing a spur, then either this spur corresponds to two consecutive arcs of ∂D with distinct edge support or it corresponds to the two arcs of a single pendant edge. In the former case, we form a disc diagram for ℓ_k by gluing the two arcs along ∂D. In the latter case, we contract the pendant edge. f is restricted accordingly to the set of remaining arcs.

- Otherwise $\ell_{k-1} \to \ell_k$ consists in the replacement of a facial subpath u by a facial subpath v such that uv^{-1} is the facial circuit of some face $F \in F(M,b)$. We then perform a face subdivision of the perforated face of D, inserting a new edge between the extremities of u. We next subdivide the new edge $k-1$ times, where k is the number of arcs of v. We finally extend f trivially by sending the subdivided edge to the edges of u and we extend g by sending the new (unperforated) face in D to F.

If the resulting diagram is not reduced, there are two loops ℓ, ℓ' bounding two faces as in Definition 3.3.8. We open D at the common basepoint of ℓ and ℓ' and identify each arc of ℓ with the corresponding arc of ℓ'. This produces a new diagram with two faces less. We repeat the procedure as long as the diagram is not reduced. By induction on the number of faces this procedure must end. Note that the final diagram may have no face, in which case its graph must be a tree corresponding to a loop that is "freely" contractible. □

Proposition 3.3.10 (Schupp, 1968)**.** *Two loops ℓ and ℓ' in (M,b) are freely homotopic if and only if there exists a reduced annular diagram in (M,b) such that the facial circuits of its two perforated faces (oriented consistently) are labelled with ℓ and ℓ' respectively.*

PROOF. By Corollary 3.2.4 there exists a path p such that $\ell \cdot p \cdot \ell'^{-1} \cdot p^{-1}$ is contractible. By the previous Proposition, there exists a disc diagram in (M,b) whose boundary is labelled with $\ell \cdot p \cdot \ell'^{-1} \cdot p^{-1}$. We may identify the subpaths corresponding to p and p^{-1} respectively and get an annular diagram whose perforated faces are labelled with ℓ and ℓ'. If the diagram is not reduced, we proceed as in the proof of Proposition 3.3.9. □

3.4 Homology

The homotopy functor identifies loops that can be continuously deformed one into the other. Intuitively, the one-dimensional homology functor identifies set of loops that can be transformed one into the other using continuous deformations as well as splittings and mergings of loops. In particular homotopic loops are homologous.

Chain groups and boundary operators We consider a map (M,b) with vertex set V, arc set A, oriented face set F and associated graph $G = (V, A, o, \iota)$. We form the set A_+ of edges with a default orientation by choosing an arc in each edge $\{a, a^{-1}\}$. Likewise, we choose an oriented face in each pair of oppositely oriented faces to form the set F_+ of faces with a default orientation. If a is an arc, we set $\partial_1 a = o(a^{-1}) - o(a)$. If F is an oriented face, we define $\partial_2 F$ as the formal sum of the arcs occurring in the facial circuit ∂F.

Definition 3.4.1. The group of i**-chains**, for $i = 0,1,2$, is the free abelian group generated by V, A and F respectively and quotiented by the relations $a^{-1} = -a$ and $F^{-1} = -F$ for every arc and oriented face. It is denoted by $C_i(M,b)$. The **boundary operators** $\partial_1 : C_1(M,b) \to C_0(M,b)$ and $\partial_2 : C_2(M,b) \to C_1(M,b)$ are the linear extensions of their value on arcs and oriented faces respectively. For notational convenience, we also let $\partial_0 : C_0(M,b) \to 0$ be the zero morphism on 0-chains and $\partial_3 : 0 \to C_2(M,b)$ be the zero morphism into the 2-chains.

The chain groups can equally be defined as the free abelian groups over V, A_+ and F_+ respectively. However, the definition of the boundary operator should be adapted to take the default orientations into account. For instance, if f is a face with default orientation F, we let $\partial_2 f$ be $\partial_2 F$ where we must replace each arc by the corresponding edge with a plus or minus sign according to whether or not the arc coincides with the default orientation of its edge. We observe that

Lemma 3.4.2. $\partial_{i-1} \circ \partial_i = 0$ *for* $i = 1,2,3$. *Equivalently,* $\operatorname{Im} \partial_i < \ker \partial_{i-1}$.

We put $Z_i(M,b) = \ker \partial_i$ and call its elements i-**cycles**. Remark that $Z_1(M,b) = Z_1(G) = H_1(G)$ so that we have a quotient $H_1(G) \twoheadrightarrow H_1(M,b)$.

The homology groups We define the ith homology group for $i = 0,1,2$ by

$$H_i(M,b) = \ker \partial_i / \operatorname{Im} \partial_{i+1}$$

In particular, $H_0(M,b) = C_0(M,b)/\operatorname{Im} \partial_1 = H_0(G)$ and $H_2(M,b) = \ker \partial_2$.

Lemma 3.4.3. *Let* (M,b) *be a connected map. If* (M,b) *is a finite orientable map without boundary then* $H_2(M,b)$ *is isomorphic to* $\mathbb{Z}$. *Otherwise,* $H_2(M,b)$ *is trivial.*

Proof. Let $\sigma = \sum n_F F$ be a 2-cycle where n_F is non-zero for a finite number of oriented faces. We consider the graph Γ whose vertices are the oriented faces of M with an edge connecting two oriented faces F and F' whenever there is an arc a such that a occurs in ∂F and a^{-1} occurs in $\partial F'$. We also consider the subgraph Γ_b of Γ obtained by deleting the oriented perforated faces (for which $b = 1$) of (M,b). Since $\partial_2 \sigma = \sum n_F \partial_2 F = 0$, each component of Γ_b must share the same coefficient in σ for all its oriented faces.

- If (M,b) has at least one perforated face then, by connectedness of M, each component of Γ must have a degree one vertex. The coefficient of the corresponding oriented face must be zero, implying that all the coefficients in σ are zero. It follows that $H_2(M,b)$ is trivial.
- If $M = (M,b)$ is a map without boundary then $\Gamma_b = \Gamma$. On the one hand, if M is non-orientable it follows from Lemma 2.2.18 that Γ is connected. As a result, any oriented face has the same coefficient has its opposite face in σ, so that σ is actually the trivial cycle in $C_2(M)$. On the other hand, if M is orientable, the same

lemma implies that Γ has two components separating each pair of oppositely oriented face. If M is infinite, the coefficient corresponding to each component must cancel as σ has finite support, whence $H_2(M)=0$. Finally, if M is orientable and finite, we may choose any coefficient for each component of Γ. The 2-cycle σ is thus an integer multiple of the sum of all the oriented face in a single component. We conclude that $H_2(M)\simeq\mathbb{Z}$ in this last case.

□

In analogy with Proposition 1.4.5, we have

Proposition 3.4.4. *$H_1(M,b)$ is isomorphic to the abelianization of the fundamental group of (M,b).*

PROOF. Fix a basepoint v in (M,b) and denote by $\mathscr{L}$ the set of loops of (M,b) with basepoint v. The map $\phi:\mathscr{L}\to H_1(M,b)$ defined by $\phi(a_1,a_2,\dots,a_k)=\sum_{i=1}^k a_i$ is compatible with elementary homotopies. This is obvious for the addition or removal of a spur. If $\lambda\cdot p\cdot\mu\mapsto\lambda\cdot q\cdot\mu$ is an elementary homotopy with $\partial F=p\cdot q^{-1}$ then $\phi(\lambda\cdot p\cdot\mu)-\phi(\lambda\cdot q\cdot\mu)=\phi(\partial F)\in\operatorname{Im}\partial_2$. Whence $\phi(\lambda\cdot p\cdot\mu)=\phi(\lambda\cdot q\cdot\mu)$ in $H_1(M,b)$. It follows that ϕ descend to the quotient $\bar\phi:\pi_1((M,b),v)\to H_1(M,b)$. On the other hand, homotopic loops in G are homotopic in (M,b) so that we have an epimorphism $\pi_1(G,v)\twoheadrightarrow\pi_1((M,b),v)$. We also know from the proof of Proposition 1.4.5 that the map $\pi_1(G,v)\twoheadrightarrow H_1(G)$ is onto. We thus have two equal compositions $\pi_1(G,v)\twoheadrightarrow H_1(G)\twoheadrightarrow H_1(M,b)$ and $\pi_1(G,v)\twoheadrightarrow\pi_1((M,b),v)\to H_1(M,b)$ implying that $\pi_1((M,b),v)\to H_1(M,b)$ is onto.

It remains to prove that $\ker\bar\phi$ is the derived subgroup of $\pi_1((M,b),v)$ to conclude that $H_1(M,b)\simeq\pi_1((M,b),v)/\ker\bar\phi$ is the abelianization of $\pi_1((M,b),v)$. Since $H_1(M,b)$ is abelian, we have $[\pi_1((M,b),v),\pi_1((M,b),v)]\subset\ker\bar\phi$. For the reverse inclusion we shall show that $[\gamma]\in\ker\bar\phi$ implies that $[\gamma]$ belongs to the kernel of the quotient

$$\pi_1((M,b),v)\to\pi_1((M,b),v)_{ab}:=\pi_1((M,b),v)/[\pi_1((M,b),v),\pi_1((M,b),v)]$$

Let T be a spanning tree of G. If a is an arc we write as usual γ_a for the loop $T[v,a]$. Considering γ as a loop in G we can write $[\gamma]=[\gamma_{a_1}]\cdot[\gamma_{a_2}]\cdots[\gamma_{a_k}]$ for some chords $a_1,a_2,\dots,a_k$ of T in G (see Theorem 1.2.10). Because $[\gamma]\in\ker\bar\phi$, there must be a 2-chain $\sum_j F_j$ such that $\gamma=\sum_j\partial_2 F_j$. For each j, we can choose a vertex v_j incident to F_j and write $T[v,v_j]\cdot[\partial_2 F_j]\cdot T[v_j,v]=\prod_k[\gamma_{a_{jk}}]$ for some chords a_{jk}. The previous identity now writes

$$\sum_i\gamma_{a_i}=\sum_{j,k}\gamma_{a_{jk}} \tag{3.11}$$

Since $\{\gamma_a\}_a$, for a running over the chords, is a free basis for $H_1(G)$, each chord appears the same number of times on both sides of (3.11). It follows that $[\gamma]$ and $\prod_j T[v,v_j]\cdot[\partial_2 F_j]\cdot T[v_j,v]$ are equal in $\pi_1((M,b),v)_{ab}$. But each loop $T[v,v_j]\cdot[\partial_2 F_j]\cdot T[v_j,v]$ is contractible, so that $[\gamma]=1$ in $\pi_1((M,b),v)_{ab}$. □

Definition 3.4.5. With the notations of the previous proof, two loops ℓ, ℓ' in $\mathscr{L}$ such that $\bar{\phi}(\ell) = \bar{\phi}(\ell')$ are said **homologous**.

Corollary 3.4.6. *Combinatorially equivalent maps have isomorphic homology groups.*

PROOF. Since H_0 is the same for a map and its graph, its invariance by combinatorial equivalence results from Proposition 1.4.1. For H_1 this is a direct consequence of the previous proposition and of Corollary 3.2.11. Finally, the invariance of H_2 results from Lemma 3.4.3. $\square$

Example 3.4.7. Denote by $M_{g,k}$ and $N_{g,k}$ an orientable, respectively non-orientable, finite map of genus g with k boundaries. From the previous proposition and corollary, and from Example 3.2.7, we compute:

$$H_1(M_{g,k}) \simeq \begin{cases} \mathbb{Z}^{2g} & \text{if } k = 0 \\ \mathbb{Z}^{2g+k-1} & \text{otherwise} \end{cases} \quad \text{and} \quad H_1(N_{g,k}) \simeq \begin{cases} \mathbb{Z}^{g-1} \times \mathbb{Z}/2\mathbb{Z} & \text{if } k = 0 \\ \mathbb{Z}^{g+k-1} & \text{otherwise.} \end{cases}$$

Note that the first homology group of a sphere or a disc is trivial.

The homology functor Let $\mu : (M', b') \to (M, b)$ be a map morphism. Recall from Corollary 2.2.32 that μ extends to vertices and oriented faces via its flag extension. We define the induced **chain morphism** $\mu_\# : C_i(M', b') \to C_i(M, b)$ by setting for a vertex v, an arc a and an oriented face F:

$$\mu_\#(v) = \mu(v), \quad \mu_\#(a) = \mu(a), \text{ and } \quad \mu_\#(F) = e_F \mu(F)$$

where e_F is the ramification index of F as in Lemma 2.2.33, that is the number of times the facial circuit of F wraps around its image by μ. We then extend $\mu_\#$ linearly to obtain a chain morphism.

Proposition 3.4.8. *The chain morphism commutes with the boundary operator, i.e.,*

$$\mu_\# \circ \partial_i' = \partial_i \circ \mu_\#$$

for $i = 0, 1, 2, 3$. (We use a prime to denote the boundary operators for (M', b').) Hence, μ induces a morphism of homology groups $\mu_* : H_i(M', b') \to H_i(M, b)$.

PROOF. The commutativity of $\mu_\#$ with the boundary operators is trivial. It implies that $\mu_\#$ sends the kernel and image of ∂_i' to the kernel and image of ∂_i. Hence, $\mu_\#$ descends to a quotient $\mu_* : H_i(M', b') \to H_i(M, b)$. $\square$

Exercise 3.4.9. Given two map morphisms $(M'', b'') \xrightarrow{\nu} (M', b') \xrightarrow{\mu} (M, b)$, prove that $(\mu \circ \nu) = \mu_* \circ \nu_*$. Prove that the identity map morphism induces the identity morphism at the level of homology.

The last exercise says that the association of maps and morphisms to the corresponding homology groups and group morphisms is a functor. For one dimensional homology, this functor can be obtained by abelianization of the homotopy functor.

Homology with other coefficients In Section 1.4 we saw that the homology of a graph could be defined over any abelian group Γ by considering chains as formal linear combinations of vertices or arcs with coefficients in Γ. We can do the same for the vertices, edges and faces of a map M to define the homology groups $H_i(M,\Gamma)$, $i=0,1,2$ with coefficients in Γ. When Γ is a field, it follows from the universal coefficient theorem for homology [Hat02, Sec. 3.A] that $H_i(M,\Gamma)$ is a vector space isomorphic to $H_i(M)\otimes\Gamma$, where $\otimes$ is the tensor product of $\mathbb{Z}$-modules. More simply, one can check that equivalent maps have isomorphic homology groups, for any chosen coefficients, and make the explicit computation for the maps in canonical form. It is also trivial to check from the definitions that homotopic loops are homologous for any chosen coefficients.

Exercise 3.4.10. Show that for any map (M,b) with basepoint v there is a group epimorphism $\pi_1((M,b),v)\twoheadrightarrow H_1(M,\mathbb{Z}/2\mathbb{Z})$. Deduce that for finite maps the homology classes of the loops in any basis of $\pi_1((M,b),v)$ form a (vector space) basis for $H_1(M,\mathbb{Z}/2\mathbb{Z})$. Is it still true if the coefficient field $\mathbb{Z}/2\mathbb{Z}$ is changed for any other coefficient field?

3.5 Cutting and Stitching

3.5.1 Cutting a map

Apart from the basic operations of Definition 2.3.6 such as edge contraction or deletion, cutting a map along a subgraph, sometimes referred to as surgery, is a common operation. On the topological side, cutting a surface S through a graph H drawn on the surface is defined as the completion with respect to some adequate metric of the complement $S\setminus H$ of H in S. On the combinatorial side, it essentially amounts to double the edges in H. We first consider the case where $H=G(M)$ is the whole graph of a map M. Cutting the map through its graph disconnects the map into a set of discs corresponding to the faces of M. We can view this cutting in the ribbon graph representation as on Figure 3.8. Each arc in the cut map can be identified with a dart of M. Letting $M=(A,\rho,\iota,s)$, we

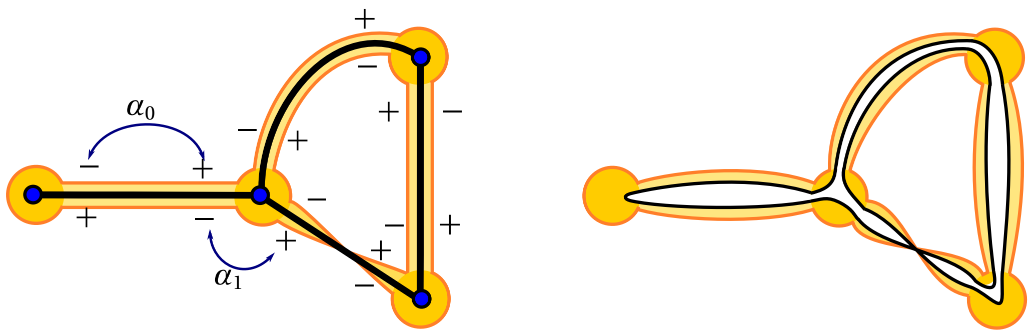


Figure 3.8: Left, the ribbon graph of a map. Right, the ribbon is cut in its middle spine. The open part (black sides) of the cut ribbon corresponds to darts of the form $(a,\epsilon,-\epsilon)$.

recall from (2.10) and (2.11) that α_0 and α_1 are involutions on the set of darts $A\times\{-1,1\}$. The previous discussion leads to define the cut map as a map with boundary (M',b') given by

$$M'=(A\times\{-1,1\},\alpha_1,\alpha_0,s\circ p_A)$$

where p_A is the projection of darts on their arc component. In order to define the boundary indicator b' we note that a dart of M' is an element of $A \times \{-1,1\} \times \{-1,1\}$ and that the cut *side* of a ribbon arc (a,ϵ) corresponds to the dart $(a,\epsilon,-\epsilon)$. This side becomes incident to a new perforated face while the other side is incident to a formerly existing face. For an arc $a \in A$ and $\epsilon \in \{-1,1\}$ we thus set

$$b'(F(a,\epsilon,-\epsilon)) = 1 \quad \text{and} \quad b'(F(a,\epsilon,\epsilon)) = 0$$

More generally, when H is an edge induced subgraph, i.e., without isolated vertices, of $G(M)$ we obtain the following

Definition 3.5.1. Let $C \subset A$ be the set of arcs of H and let $\overline{C} = A \setminus C$. Denote by α_0^C and α_1^C the first two involutions on the darts of the restriction M_C of M to C (see Definition 2.2.71). The **cut map** (M', b') is given by $M' = (A', \rho', \iota', s')$ where

- $A' = \overline{C} \cup (C \times \{-1,1\})$,
- if $a \in \overline{C}$ and $\rho(a) \in \overline{C}$ then $\rho'(a) = \rho(a)$
  if $a \in \overline{C}$ and $\rho(a) \in C$ then $\rho'(a) = (\rho(a), 1)$
  if $a \in C$ and $\rho(a) \in \overline{C}$ then $\rho'(a,1) = \alpha_1^C(a,1)$ and $\rho'(a,-1) = \rho(a)$
  if $a \in C$ and $\rho(a) \in C$ then $\rho'(a,\epsilon) = \alpha_1^C(a,\epsilon)$ for $\epsilon \in \{-1,1\}$.
- if $a \in \overline{C}$ then $\iota'(a) = \iota(a)$, otherwise $\iota'(a,\epsilon) = \alpha_0^C(a,\epsilon)$ for $\epsilon \in \{-1,1\}$.
- if $a \in \overline{C}$ then $s'(a) = s(a)$ else $s'(a,\epsilon) = s(a)$ for $\epsilon \in \{-1,1\}$

and the boundary indicator b' is given for $a \in A$ and $\epsilon \in \{-1,1\}$ by

$$\begin{array}{lll} b'(F(a,\epsilon)) = 0 & \text{if} & a \in \overline{C} \\ b'(F(a,\epsilon,\epsilon)) = 0 & \text{if} & a \in C \\ b'(F(a,\epsilon,-\epsilon)) = 1 & \text{if} & a \in C \end{array}$$

The cut map (M', b') is denoted by $M \backslash\backslash H$. The arcs in $C \times \{-1,1\} \subset A'$ are called the **cut arcs** of $M \backslash\backslash H$. For a cut arc (a,ϵ), the perforated face $F(a,\epsilon,-\epsilon)$ is called a **cut face** and the dart $(a,\epsilon,-\epsilon)$ is called a **cut dart**.

The above definition of the boundary indicator b' is justified by the next

Lemma 3.5.2. *Let φ' be the facial permutation of $M \backslash\backslash H$. The φ'-orbit of any cut dart is composed of cut darts only.*

Proof. We denote by φ_C the facial permutations of M_C. For a cut arc (a,ϵ) we compute

$$\varphi'(a,\epsilon,-\epsilon) = (\rho'^{-\epsilon s(a)}(\alpha_0^C(a,\epsilon)), -\epsilon s(a)) = (\alpha_1^C \alpha_0^C(a,\epsilon), -\epsilon s(a)) = (\varphi_C(a,\epsilon), -\epsilon s(a)) \quad (3.12)$$

We note that the sign of $\varphi_C(a,\epsilon)$ is $\eta = \epsilon s(a)$ whence $\varphi'(a,\epsilon,-\epsilon) = (b,\eta,-\eta)$ for some cut arc (b,η). The lemma follows. □

Exercise 3.5.3. Show that the faces of M are in one-to-one correspondence with the (unperforated) faces of $M \backslash\backslash H$.

Exercise 3.5.4. Ignoring the boundary indicator, show that the δ-map associated to $M\backslash\backslash H$ is given by $\delta(M\backslash\backslash H) = (D', \alpha'_0, \alpha'_1, \alpha'_2)$ where

- $D' = (\overline{C} \times \{-1, 1\}) \cup (C \times \{-1, 1\} \times \{-1, 1\})$
- for $\epsilon, \eta \in \{-1, 1\}$:
  - If $a \in \overline{C}$:
    * $\alpha'_0(a, \epsilon) = \alpha_0(a, \epsilon)$,
    * $\alpha'_1(a, \epsilon) = \begin{cases} \alpha_1(a, \epsilon) & \text{if } \rho^{-\epsilon}(a) \in \overline{C} \\ (\alpha_1(a, \epsilon), -\epsilon) & \text{otherwise.} \end{cases}$
    * $\alpha'_2(a, \epsilon) = \alpha_2(a, \epsilon)$
  - If $a \in C$:
    * $\alpha'_0(a, \epsilon, \eta) = (\alpha_0(a, \epsilon), -\eta s(a))$
    * $\alpha'_1(a, \epsilon, \eta) = \begin{cases} \alpha_1(a, \epsilon) & \text{if } \rho^{-\epsilon}(a) \in \overline{C} \text{ and } \epsilon = \eta \\ (\alpha_1^C(a, \epsilon), -\eta) & \text{otherwise.} \end{cases}$
    * $\alpha'_2(a, \epsilon, \eta) = (a, \epsilon, -\eta)$

Figure 3.9 illustrates the effect of cutting a map on the corresponding δ-map.

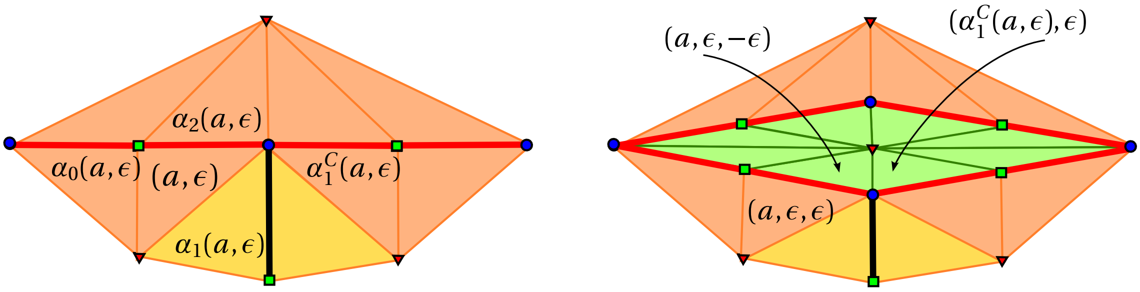


Figure 3.9: Left, A piece of δ-map. The cutting graph H is composed of the two horizontal (red) edges. Right, The cut face is made of cut darts.

As a first intuitive property, we note that the basic operations of Definition 2.3.6 commute with the cutting of a map. The omitted proof, though cumbersome, is a simple matter of applying the relevant definitions and verifying the commutation.

Proposition 3.5.5. *Let H be a subraph of (M, b) and let e be an edge of M not in H.*

- *the contraction of e and the cutting along H commute: $S_e((M, b)\backslash\backslash H) = S_e(M, b)\backslash\backslash H$,*
- *if e is a non-loop edge (with positive signature) then $((M, b)\backslash\backslash H)/e = ((M, b)/e)\backslash\backslash H$.*
- *if e is a regular edge incident to at least one face, then $((M, b)\backslash\backslash H) - e = ((M, b) - e)\backslash\backslash H$,*
- *the subdivision of a (possibly perforated) face in $(M, b)\backslash\backslash H$ yields the same map as the subdivision of the corresponding face in (M, b) followed by the cutting along H.*

We next describe the effect of cutting.

Proposition 3.5.6. *With the notations of Definition 3.5.1, we have the following properties.*

a) The cut faces have simple and vertex disjoint facial circuits.

b) The cut faces of $M\backslash\backslash H$ are in one-to-one correspondence with the faces of M_C.

c) If M is connected each component of $M\backslash\backslash H$ contains at least one cut face.

d) $\chi(M\backslash\backslash H) = \chi(M) - \chi(H)$.

Proof. a) is a consequence of the claims that (1) for any cut arc (a,ϵ) all the arcs of $\partial F(a,\epsilon,-\epsilon)$ are cut arcs and (2) the subgraph of the graph of $M\backslash\backslash H$ induced by the cut arcs has degree two vertices only. This second claim follows from the fact that each cycle of ρ' contains either zero or two cut arcs by construction. The first claim is a direct consequence of the previous lemma. We also deduce from (3.12) that the arc projection of a cycle of φ' is a cycle of φ_C, thus proving b). Point c) is trivial. It remains to prove d).

We denote by V, E, F the respective number of vertices, edges and faces of M. We denote with a prime the corresponding quantities for $M\backslash\backslash H$. We also let V_H and E_H be the number of vertices and edges of H. From Definition 3.5.1, every edge of H is doubled in $M\backslash\backslash H$. We thus have $E' = 2E_H + (E - E_H) = E + E_H$. We also have $F' = F$ (see Exercise 3.5.3). Since the cut faces have simple and disjoint facial circuits in correspondence with the faces of M_C by a) and b) we can write $V' - 2E_H = V - V_H$, i.e., $V' = V - V_H + 2E_H$. We finally compute

$$\chi(M\backslash\backslash H) = V' - E' + F' = V - V_H + 2E_H - E - E_H + F = \chi(M) - \chi(H)$$

□

Let $(\alpha'_0, \alpha'_1, \alpha'_2)$ be the involutions of $M\backslash\backslash H$ as in Exercise 3.5.4. We consider the graph K whose vertices are the set D' of darts of $M\backslash\backslash H$ and whose edges are the orbits of the α'_i. We also denote by $D_C = \{(a,\epsilon,-\epsilon) \mid a \in C, \epsilon \in \{-1,1\}\}$ the set of cut darts.

Lemma 3.5.7. *The components of K are in correspondence with the components of $K - D_c$, i.e., the inclusion $K - D_c \hookrightarrow K$ induces an isomorphism $H_0(K - D_c) \simeq H_0(K)$.*

Proof. Since $d \in D_C$ implies $\alpha'_2(d) \notin D_C$, each component of K contains a non cut dart. Moreover, if $d, d' \in D' \setminus D_C$ are in a same component of K, we claim that exists a path in K connecting d to d' and avoiding D_C, thus proving the lemma. We prove the claim using an induction on the number of cut darts in a path π from d to d'. Let t be the first occurrence of a cut dart in π. Denote by u and v the vertex that precedes, respectively follows, t along π. Since α'_0 and α'_1 leaves D_C invariant, we must have $u = \alpha'_2(t)$. In each of the three possibilities for v we construct a (d,d')-path π' with one cut dart less than π:

- If $v = \alpha'_2(t)$ then $u = v$ and we can short-cut t in π to obtain a shorter (d,d')-path π'.

- If $v = \alpha'_0(t)$, then $v = \alpha'_2\alpha'_0(u) = \alpha'_0\alpha'_2(u)$ and we replace t by $\alpha'_0(u)$ in π to get π'.
- Finally, if $v = \alpha'_1(t)$, then writing $t = (a, \epsilon, -\epsilon)$, we have $u = (a, \epsilon, \epsilon)$ and $v = (\alpha_1^C(a, \epsilon), \epsilon)$. By definition of the restricted map M_C there exists a sequence of arcs $(a_0 = a, a_1, a_2, \dots, a_k)$ of M, with $k \geq 1$, such that $a_{i+1} = \rho^{-\epsilon}(a_i)$, $v = (a_k, -\epsilon, \epsilon)$ and $a_j \in \overline{C}$ for $j < k$. We construct from this sequence a (u, v)-path in $K - D_C$:
$$\gamma = (u_0 = u, u_1, u'_1, u_2, u'_2, \dots, u_{k-1}, u'_{k-1}, v = \alpha'_1(u'_{k-1})),$$
where for $j = 1, \dots, k-1$, $u_j = (a_j, \epsilon) = \alpha'_1(u'_{j-1})$ and $u'_j = (a_j, -\epsilon) = \alpha'_2(u_j)$. We can now define π' by substituting γ to (u, t, v) in π.

By induction, we can thus assume that π does not contain any cut dart. □

Proposition 3.5.8. *Let $G(M^*)$ be the graph of the dual map of M. We denote by $E^*(H)$ the set of edges dual to the edges of H. The connected components of $M\backslash\backslash H$ are in one-to-one correspondence with the connected components of $G(M^*) - E^*(H)$*

Proof. By definition, the components of $M\backslash\backslash H$ correspond to the orbits of its monodromy group $\mathrm{Mon}(M\backslash\backslash H) = (\alpha'_0, \alpha'_1, \alpha'_2)$, where the α'_i are given in Exercise 3.5.4. Such orbits correspond to the connected components of the above graph K. The previous lemma shows in turn that those components correspond to the components of $K - D_C$. It remains to prove that those last components are in bijection with the components of $G(M^*) - E^*(H)$. Letting D be the set of darts of M, we remark that there is a bijection $\theta : D \to D' \setminus D_C$ such that $\theta \circ \alpha_0 = \alpha'_0 \circ \theta$ and $\theta \circ \alpha_1 = \alpha'_1 \circ \theta$. The vertices of the dual map M^* seen as the orbits of $\langle \alpha_0, \alpha_1 \rangle$ are thus in bijection with the orbits of $\langle \alpha'_0, \alpha'_1 \rangle$ that do not contain cut darts. Let $d, d' \in D' \setminus D_C$ and let p, q be the corresponding vertices of M^*. We claim that d and d' are in a same component of $K - D_C$ if and only if p and q are in a same component of $G(M^*) - E^*(H)$. Indeed, let $\pi = \pi_1 \cdot \pi_2 \cdots \pi_k$ be a (d, d')-path in $K - D_C$, where any two consecutive darts in each subpath π_i is related by α'_0 or α'_1. We can construct a (p, q)-path in $G(M^*) - E^*(H)$ by replacing each π_i with the dual vertex p_i of M^* corresponding to the orbit $\langle \alpha'_0, \alpha'_1 \rangle(\pi_i)$. Moreover if d_i is the last dart of π_i then, since $\alpha'_2(d_i)$ is not a cut dart (being the first dart of π_{i+1}), the orbit $\langle \alpha'_0, \alpha'_2 \rangle(d_i)$ corresponds to an edge of M between p_i and p_{i+1} that is not in H. Conversely, the fact that $D' \setminus D_C$ is stable by α'_0 and α'_1 allows to construct a (d, d')-path in $K - D_C$ from any (p, q)-path in $G(M^*) - E^*(H)$. By sending each dart in $D' \setminus D_C$ to the corresponding vertex of M^*, we thus establish a bijection between the connected components of $K - D_C$ and of $G(M^*) - E^*(H)$. □

Cutting maps with boundary

We can easily extend the definition of a cut map for a subgraph H of a map (M, b) with boundary. Only the boundary indicator of $(M, b)\backslash\backslash H := (M\backslash\backslash H, b')$ needs to be specified. For this we declare as perforated the faces of $M\backslash\backslash H$ that are already cut faces or that are in correspondence with faces of M perforated in (M, b) (see Exercise 3.5.3). All the results of this section apply to maps with boundary in a straightforward manner.

Exercise 3.5.9. Show that a map (M, b) cut along its graph has disc and annulus components only that are in correspondence with the set of faces, respectively perforated faces, of the map (M, b).

The stitching morphism The map $(M,b)\backslash\backslash H$ naturally projects to (M,b): with the notations of Definition 3.5.1, we have a **stitching** mapping $\sigma : A' \to A$ defined by $\sigma(a) = a$ for $a \in \overline{C}$ and $\sigma(a,\epsilon) = a$ for $a \in C \times \{-1,1\}$. Although σ defines a graph morphism $G((M,b)\backslash\backslash H) \to G(M,b)$, it does *not* extend to a map morphism in the sens of Definition 2.3.5. However, similarly to the basic operations (see Lemma 3.2.10), σ induces morphisms of fundamental and homology groups. Indeed, it is easily checked that the stitching mapping sends the facial circuits of $(M,b)\backslash\backslash H$ to the facial circuits of (M,b). Homotopic loops in $(M,b)\backslash\backslash H$ are thus sent to homotopic loops in (M,b) and for any vertex v of $(M,b)\backslash\backslash H$ the induced graph morphism $\sigma_* : \pi_1(G((M,b)\backslash\backslash H), v) \to \pi_1(G(M,b),\sigma(v))$ quotients to a morphism $\pi_1((M,b)\backslash\backslash H, v) \to \pi_1((M,b),\sigma(v))$. A similar statement holds for the homology. With a little abuse of notation we write $\sigma : (M,b)\backslash\backslash H \to (M,b)$ for the stitching mapping. When H is the union of two edge induced subgraphs H_1, H_2, it is straightforward that

Lemma 3.5.10. *The stitching $\sigma : (M,b)\backslash\backslash H \to (M,b)$ factors as*

$$(M,b)\backslash\backslash H \to (M,b)\backslash\backslash H_1 \to (M,b)$$

The left arrow can be seen as a partial stitching of H_2. The composition extends to the various induced morphisms.

Beware that the intuitive reverse statement is false: in general, $M\backslash\backslash H$ is *not* isomorphic to $(M\backslash\backslash H_1)\backslash\backslash H_2$.

Exercise 3.5.11. Compare the result of cutting a map along its graph and cutting the same map along each of its edges, one after the other.

3.5.2 The Seifert-van Kampen theorem

When (M,b) is connected and H is a cycle of $G(M)$, the fundamental groups of (M,b) and of the components of $(M,b)\backslash\backslash H$ are algebraically related. This is the content of the following two versions of the Seifert-van Kampen's theorem. In these two versions, we fix a vertex v of the cycle H and a generator h of $\pi_1(H,v)$.

Theorem 3.5.12 (Seifert-van Kampen's theorem, version I)**.** *Suppose that $(M,b)\backslash\backslash H$ has two components M_1 and M_2 (with boundaries). Let (h_1,v_1) and (h_2,v_2) correspond to (h,v) in M_1 and M_2, respectively, via the stitching morphism. The fundamental group of (M,b) is isomorphic to the* free product with amalgamation $\pi_1(M_1,v_1) *_{h_1=h_2} \pi_1(M_2,v_2)$. *In other words, if $\langle X_i \mid R_i\rangle$ is a presentation of $\pi_1(M_i,v_i)$ and x_i is an expression for h_i in $\langle X_i \mid R_i\rangle$, we have*

$$\pi_1((M,b),v) \simeq \langle X_1 \cup X_2 \mid R_1 \cup R_2 \cup \{x_1 x_2^{-1}\}\rangle$$

Proof. Let e be an edge of H. The path $H - e$ correspond to a path in M_1 that we can extend to a spanning tree T_1 of M_1. We similarly get a spanning tree T_2 of M_2. The union $T = \sigma(T_1) \cup \sigma(T_2)$, where σ is the stitching mapping, is a spanning tree of M. Let C_i be the set of chords of T_i in $G(M_i)$ for $i = 1,2$. By Lemma 3.2.6, $\pi_1(M_i,v_i)$ is isomorphic to $\langle C_i \mid \{r_F\}_{F\in F(M_i)}\rangle$, while $\pi_1((M,b),v)$ is isomorphic to $\langle (C_1 - e_1) \cup (C_2 - e_2) \cup \{e\} \mid \{r_F\}_{F\in F(M,b)}\rangle$, where $\sigma(e_i) = e$. Applying some elementary Tietze transformations we get that $\pi_1((M,b), v)$ is isomorphic to

$$\langle C_1 \cup C_2 \mid \{r_F\}_{F\in F(M,b)} \cup \{e_1 e_2^{-1}\}\rangle = \pi_1(M_1, v_1) *_{e_1=e_2} \pi_1(M_2, v_2)$$

□

Theorem 3.5.13 (Seifert-van Kampen's theorem, version II)**.** *Suppose that $(M,b)\backslash\backslash H$ is connected. Let (h_1, v_1) and (h_2, v_2) correspond to (h, v) in $(M,b)\backslash\backslash H$ via the stitching morphism. The fundamental group of (M,b) is isomorphic to the* HNN extension *$\pi_1((M,b)\backslash\backslash H, v_1)*_h$. In other words, if $\langle X \mid R\rangle$ is a presentation of $\pi_1((M,b)\backslash\backslash H, v_1)$ and x_i is an expression for h_i in $\langle X \mid R\rangle$, we have (using a new symbol t):*

$$\pi_1((M,b), v) \simeq \langle X \cup \{t\} \mid R \cup \{x_1 t x_2^{-1} t^{-1}\}\rangle$$

Proof. Let e be an edge of H. Denote by e_i and H_i, $i = 1,2$, the reciprocal images of e and H by the stitching mapping σ. Proposition 3.5.6.a)) ensures that the paths $H_1 - e_1$ and $H_2 - e_2$ are disjoint in $(M,b)\backslash\backslash H$ and can be extended to a spanning tree T' of $(M,b)\backslash\backslash H$. Consider a simple path in T' from a vertex of H_1 to a vertex of H_2 and let t' be an edge of this path not in $H_1 \cup H_2$. The graph $T = \sigma(T') - \sigma(t')$ is a spanning tree for (M,b). Let C be the set of chords of T in $G(M)$. By Lemma 3.2.6, $\pi_1((M,b), v)$ is isomorphic to $\langle C \mid \{r_F\}_{F\in F(M,b)}\rangle$. By some elementary Tietze transformations, this last group is isomorphic to

$$\langle C - \{e,t\} \cup \{e_1, e_2\} \mid \{r_F\}_{F\in F((M,b)\backslash\backslash H)} \cup \{e_1 t e_2^{-1} t^{-1}\}\rangle = \pi_1((M,b)\backslash\backslash H, v_1)*_e$$

□

Corollary 3.5.14. *Let (M,b) be a connected map and let H be a cycle of $G(M)$ with generator h. If none of the components of $(M,b)\backslash\backslash H$ is a disc, then h is non-contractible.*

Proof. By Proposition 3.5.6.b) h has two reciprocal images, h_1 and h_2, by the stitching mapping σ. We claim that h_1 and h_2 are non-contractible in $(M,b)\backslash\backslash H$. Indeed, the perforated faces of combinatorially equivalent maps are in one-to-one correspondence and it is easily seen from Section 3.2.1 that the induced morphisms of the basic operations send the homotopy class of the facial circuit of a perforated face to the homotopy class of the facial circuit of the corresponding perforated face. We can thus assume that $(M,b)\backslash\backslash H$ is in canonical form. On the other hand, it is seen from Example 3.2.7, that the facial circuit of a perforated face in a canonical map that is not a disc is part of a basis of a non-trivial free group, hence non-contractible. Since by assumption neither h_1 nor h_2 bounds a disc, we conclude that none of them is contractible.

Now, it follows from the normal form theorems for amalgamated products and HNN extensions (see [LS77, Sec. IV.2]) that the groups embed into their amalgamated product and that a group embeds into its HNN extension via the obvious identifications of the generators. In particular, the homotopy class of h_i gets identified to the homotopy class of h, which is thus non-contractible. □

Theorem 3.5.15. *Let (M,b) be a connected map and let H be a simple cycle in $G(M)$. The cycle H is the support of a simple circuit that is contractible in (M,b) if and only if $(M,b)\backslash\backslash H$ has a disc component bounded by H.*

PROOF. The condition is necessary by the previous corollary. For the reverse implication, we assume that $(M,b)\backslash\backslash H$ has a disc component. Let c be one of the two oriented simple circuits in H. By Proposition 3.5.6.b) the boundary c' of the disc projects to c. Since a disc has a trivial fundamental group, c' is contractible and the existence of the induced stitching morphism implies that c is also contractible in (M,b). □

Notes: The above proof of the Seifert-van Kampen theorem is from [vCGKZ98, Sec. 1.2]. The combinatorial framework makes this theorem almost a tautology as opposed to the usual and relatively involved proof for topological spaces [Hat02, Sec. 1.2],[Mas91, chap. IV]. The above extension of the Jordan-Schoenflies theorem is classical in the realm of topological surfaces [Eps66, Th. 1.7]. The present proof is purely combinatorial and relies on standard properties of combinatorial group theory. Needless to say that those properties have short algebraic proofs [Ser77, Chap. I] or [LS77, Sec. IV.2] and do not require topological arguments.

3.5.3 The Jordan curve theorem

The following theorem is a direct consequence of Theorem 3.5.15. We nonetheless provide a simpler proof for this specific case.

Theorem 3.5.16 (Combinatorial Jordan-Schoenflies theorem)**.** *If M is a sphere map and H is a simple circuit of $G(M)$, then $M\backslash\backslash H$ is composed of two disc components. In particular, all the edges of H are regular.*

PROOF. By point d) of Proposition 3.5.6, we have $\chi(M\backslash\backslash H) = \chi(M) = 2$. By b) $M\backslash\backslash H$ has two cut boundaries, while c) implies that $M\backslash\backslash H$ has at most two components. It is clear from the definition of cut maps that these components must be orientable. Since the Euler characteristic of a map with $k \geq 1$ boundaries is at most $2-k$ (see Exercise 3.1.8), the only possibility is that $M\backslash\backslash H$ has two disc components. □

Exercise 3.5.17. Prove the following combinatorial θ's lemma. If M is a sphere map and p,q,r are three simple paths in M sharing exactly their endpoints then $M\backslash\backslash(p\cup q\cup r)$ is composed of three disc components such that the facial circuit of their cut faces project on $p\cdot q^{-1}$, $q\cdot r^{-1}$ and $r\cdot p^{-1}$ respectively.

Bibliographical notes

The Jordan curve theorem is one of the most emblematic results in topology. Its statement is intuitively obvious but it is rather difficult to prove, unless more advanced arguments of algebraic topology are used. One has to deal with the fact that a continuous curve can be quite wild, e.g. fractal, which explains the difficulty of the proof. A rather accessible proof was proposed by Helge Tverberg [Tve80] (see my course notes [Laz12a] for a gentle introduction). Eventually, a formal proof, automatically checked by a computer, was recently given by Thomas Hales and other mathematicians [Hal07b, Hal07a]. Concerning the Jordan-Schoenflies theorem, the situation is even worth. This stronger version of the Jordan curve theorem asserts that a simple curve does not only cut a sphere into two pieces but that each piece is actually a topological disc. A nice proof by elementary means – but far from simple – and resorting to the fact that $K_{3,3}$ is not planar is due to Carsten Thomassen [Tho92].

As one might expect, the analogous result for combinatorial maps is much simpler to prove. Several versions and proofs of the combinatorial Jordan(-Schoenflies) theorem were proposed. Tutte [Tut79][Tut01, Sec. XI.3] uses a notion of *premaps* equivalent to δ-maps and defines the cutting [Tut79] of a premap along a discrete cut that corresponds to a continuous curve drawn on the topological surface and intersecting the embedded graph at vertices only. This curve is required to traverse each face at most once but can traverse a vertex several times as long as some natural notion of crossing is avoided. A combinatorial and rather abstract form of Jordan's theorem is proved using these notions and some amount of specific terminology. Stahl [Sta83] later gave his own version and proof of the theorem. This time, a combinatorial surface is encoded as a pair (P,Q) of permutations to which is associated a cellular embedding of a digraph defined as a the union of the cycles of P and Q. This abstract framework can actually be interpreted as follows. We consider the branched covering associated to the 3-constellation $[P,Q,Q^{-1}P^{-1}]$ (see Section 2.1.1). The digraph is then the lift of the graph composed of the loops λ,μ as on Figure 2.6. Stahl defines an ad hoc notion of cycle and separation for the combinatorial surface defined by (P,Q) and obtains another form of the Jordan's theorem. Vince and Little [VL89] provide yet another presentation and make the connection between the approaches of Stahl and Little [Lit88]. This work was extended by Bonnington and Little [BL94] to show that the genus of a surface is the maximal number of disjoint closed curves that do not disconnect the surface. Noteworthy, all the above mentioned works are quite formal with a very few geometric intuition. They seem to be independent of the concurrent point of view of Jones and Singerman [JS78] or Bryant and Singerman [BS85].

3.5.4 Cut graphs

Definition 3.5.18. Let (M,b) be a connected map. A subgraph H of $G(M)$ **disconnects** (M,b) if $(M,b)\backslash\backslash H$ is not connected. H is called a **cut graph** of the map M without boundary if $M\backslash\backslash H$ is a disc. If (M,b) has at least one perforated face, H is called a cut graph of (M,b) when every component of $(M,b)\backslash\backslash H$ is an annulus with exactly one cut face.

If M has a single face then its graph $G(M)$ is a cut graph. Figure 3.10 shows a cut graph of a map with boundary.

Proposition 3.5.19. *Any cut graph H of M is connected with cyclomatic number*

$$\beta_1(H) = 2 - \chi(M)$$

Moreover, for any vertex v of H, the morphism $\pi_1(H,v) \to \pi_1(M,v)$ induced by the inclusion $H \hookrightarrow M$ is onto.

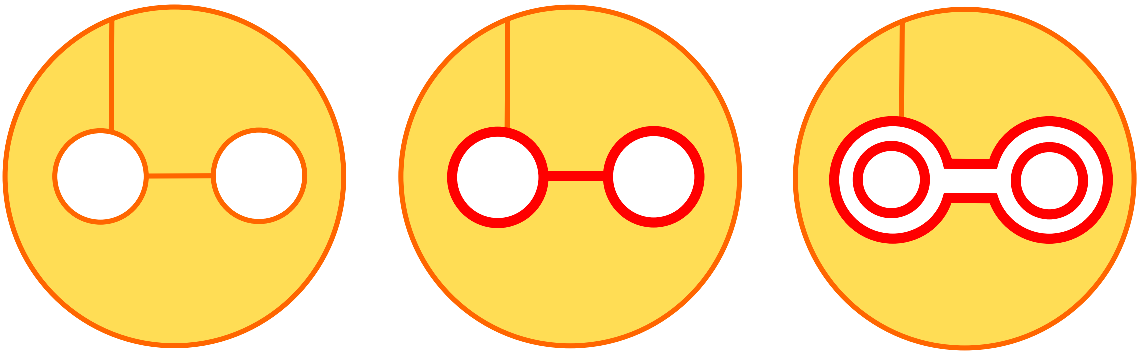


Figure 3.10: Left, A sphere with three perforated faces (a pair of pants). Middle, A cut graph. Right, The left map cut through the cut graph has three components. Each component is a cylinder with one cut face (only one of the two perforated faces of each of the two "circle" maps is a cut face).

PROOF. Since H is the image by the stitching mapping of the facial circuit of the unique cut face of $M\backslash\backslash H$, it is connected. The formula for the cyclomatic number directly follows from Proposition 3.5.6-d) noting that $\chi(M\backslash\backslash H) = 1$ because H is a cut graph. To prove that the morphism $\pi_1(H, v) \to \pi_1(M, v)$ is onto it is sufficient to show that every loop γ in M is homotopic to a loop in H. Thanks to Lemma 3.2.5 and the fact that the fundamental group of a disc is trivial, we can indeed replace each maximal sequence of arcs of γ not in H by a homotopic sequence of arcs in H. □

Proposition 3.5.20. *The cyclomatic number of a cut graph H of a connected map M of genus g without boundary is $2g$ if M is orientable and g otherwise. Moreover, any basis of $\pi_1(H, v)$ is a basis for $\pi_1(M, v)$.*

PROOF. Recall from Section 1.4 that the cyclomatic number of H satisfies $\beta_1(H) = 1 - \chi(H)$. We deduce from the previous lemma that $\beta_1(H) = 2g$ when M is a genus g orientable map, i.e., when $\chi(M) = 2 - 2g$ and that $\beta_1(H) = g$ when M is a genus g non-orientable map, i.e., when $\chi(M) = 2 - g$. The free group $\pi_1(H, v)$ has rank $\beta_1(H) = 1 - \chi(G)$ (Corollary 1.2.12). By the previous lemma, any basis of $\pi_1(H, v)$ is a generating set for $\pi_1(M, v)$. We call a generating set of $\pi_1(M, v)$ of minimal size a **basis**. It is easily seen from Example 3.2.7 that the abelianization of $\pi_1(M, v)$ is a free abelian group of rank $\beta_1(H)$. It follows that the minimal size of any generating set is indeed $\beta_1(H)$. □

We now turn to cut graphs of maps with nonempty boundary.

Lemma 3.5.21. *Let (M, b) be an annulus and let B_1 be one of the two perforated faces of (M, b). If γ is a path with endpoints in ∂B_1, then γ is homotopic to a path whose support is in ∂B_1.*

PROOF. The deletion of a regular edge e of (M, b) incident to at least one (unperforated) face is the inverse of a face subdivision and thus leads to a combinatorially equivalent map. It follows from the proof of Lemma 3.2.10 that the inclusion of $G(M, b) - e$

into $G(M,b)$ induces an isomorphism between the fundamental groups of $(M,b)-e$ and of (M,b). Moreover, the perforated faces of $(M,b)-e$ naturally correspond to those of (M,b). We shall first reduce (M,b) into a combinatorially equivalent map whose graph is naturally identified with the subgraph of $G(M)$ induced by support edges of ∂B_1. The reduction is performed in two steps.

1. Starting with (M,b), we recursively delete regular edges *not* incident to B_1, as long as we can find one such edge. We obtain an equivalent map (M',b') with a bijection between the edges of B_1 and the edges of the corresponding perforated face in (M',b'). We still denote by B_1 this perforated face. All the regular edges of (M',b') are now incident to B_1.

2. By Lemma 3.2.10 we can contract any *pendant edge* in (M',b'), i.e., an edge with a degree one vertex, to obtain a combinatorially equivalent map with an inclusion of its graph in $G(M')$, hence in $G(M)$. As long as there exists a pendant, and thus singular, edge that is *not* incident to B_1, we contract that edge. We obtain this way an equivalent map (M'',b'') with an inclusion of its graph in $G(M)$ and with a bijection between the edges of B_1 and the edges of the corresponding perforated face in (M'',b'').

We claim that all the edges of (M'',b'') must be incident to B_1. Suppose not. Then there must be a singular edge e incident to a possibly perforated face F distinct from B_1. By connectivity of the dual map M''^*, the face F must be incident to a regular edge which is also incident to B_1 by the first reduction. By Proposition 2.2.66 this edge belongs to a simple circuit c of edges incident to F. By the Jordan curve theorem 3.5.16, c cuts M'' into two discs, one of which contains F and the other one contains B_1. The piece containing F, say D_1, cannot contain any other face or perforated face since D_1 would then contain a regular edge not incident to B_1, in contradiction with the first step. Using the Euler characteristic, we get that c is the only cycle of $G(D_1)$. Since the assumed edge e is not in c by hypothesis, it follows that $G(D_1)$ contains a pendant edge incident to F, which is also pendant in M''. This contradicts the second step and prove the claim.

Being on ∂B_1, the endpoints of the given path γ correspond to two vertices in (M'',b''). Since (M,b) and (M'',b'') have isomorphic fundamental groups, we may choose in (M'',b'') a path λ representing the homotopy class of γ. Since the facial circuit of B_1 covers all the edges of (M'',b''), the path λ corresponds to a path in (M,b) whose support is in ∂B_1. □

Analogously to Proposition 3.5.19, we have

Proposition 3.5.22. *Let (M,b) be a connected map with at least one perforated face. Any cut graph H of (M,b) is connected with cyclomatic number*

$$\beta_1(H) = 1 - \chi(M,b)$$

Moreover, for any vertex v of H, the morphism $\pi_1(H,v) \to \pi_1((M,b),v)$ is an isomorphism.

PROOF. Suppose for a contradiction that H is not connected. Consider a path in $G(M) - E(H)$ connecting two distinct components of H. This path corresponds to a path contained in a single component of $(M,b)\backslash\backslash H$ and joining two distinct cut faces. This contradicts the hypothesis that each component of $(M,b)\backslash\backslash H$ should contain exactly one cut face. The formula for the cyclomatic number of H results from Proposition 3.5.6-d) and the fact that the characteristic of a cylinder is null.

Let γ be a loop of (M,b) with basepoint v. Every maximal subpath q of γ included in $G(M) - E(H)$ corresponds to a path q' of some component of $(M,b)\backslash\backslash H$. The endpoints of q' are on the cut face of this component and by the previous lemma it is homotopic to a path contained in this cut face. It follows that q is homotopic to a path in H. In turns, this implies that the morphism $\pi_1(H,v) \to \pi_1((M,b),v)$ induced by the inclusion $H \hookrightarrow M$ is onto. This concludes the proof of the proposition by noting from Example 3.2.7 that $\pi_1(H,v)$ and $\pi_1((M,b),v)$ are free groups of the same rank. □

Given a cut graph H of a map (M,b) with or without boundary, we can easily obtain a combinatorial presentation of its fundamental group. We consider a spanning tree T of H and we let C be the set of chords of T in H.

Theorem 3.5.23. *If (M,b) has at least one boundary, its fundamental group is isomorphic to the free group over C. Otherwise, the fundamental group of M is isomorphic to the group*

$$\langle C \mid r \rangle$$

where r is the trace over C of the facial circuit of the perforated face of $M\backslash\backslash H$.

PROOF. When (M,b) has at least one boundary, the lemma follows from Proposition 3.5.22 and Theorem 1.2.10. Otherwise, $M\backslash\backslash H$ is a disc whose perforated face B has a simple facial circuit by Proposition 3.5.6. a). We shall first reduce M to an equivalent map whose graph is exactly H. Suppose that M, hence $M\backslash\backslash H$, has at least two faces. We claim that $M\backslash\backslash H$ has a regular edge that is not a cut edge. Indeed, let F be a face of $M\backslash\backslash H$. If F does not share any edge with B, or if F is the only face adjacent with B, then F must share an edge with some other face by connectivity of the map dual to $M\backslash\backslash H$ (forgetting about its boundary indicator). Otherwise, let a, b be two consecutive arcs of ∂B such that a is incident to F and b is incident to another face distinct from F. Then one of the edges incident to the common vertex of a and b is incident to two distinct faces, thus proving the claim. We can delete that edge to get a map with one face less. By induction on the number of faces of M, we can remove a set of regular edges $E \subset E(M) \setminus E(H)$ to obtain another map $M' := (M\backslash\backslash H) - E$ with a single face. We now claim that every edge of M' not contained in ∂B is a bridge of $G(M')$. Otherwise, there would be an edge e not incident to ∂B and contained in a simple cycle. The Jordan curve theorem would imply that e is regular, a contradiction. Thus, unless $G(M')$ is reduced to the edges of ∂B there must be a pendant edge that we can contract to obtain an equivalent map. By induction on their number, we can contract all the edges E' of $G(M')$ not in ∂B to obtain a disc M'/E' with a single face sharing its whole facial circuit with B. By Proposition 3.5.5, $(M - E)/E'$ is equivalent to M and has a single face whose facial circuit coincides with ∂B. We may now conclude with Lemma 3.2.6. □

3.6 Some Elementary Algorithms Related to Homotopy

3.6.1 Computing a basis of the fundamental group

As for graphs in Section 1.3, we can easily determine a basis of the fundamental group of a map.

Theorem 3.6.1. *Let (M,b) be a finite map of genus g with n arcs and k perforated faces. We can compute a set of representatives of a basis of the fundamental group of (M,b) in $O((g+k)n)$ time.*

PROOF. We first assume that (M,b) has no perforated faces. Let E^* be the set of edges of a spanning tree of the dual graph $G(M^*)$. We delete the corresponding set E of edges in M one at a time in any order. Since each edge in E is regular, its deletion only merges the two incident faces and leaves the other faces unchanged so that the edges left in E remain regular. By induction on the number of edges we obtain a combinatorially equivalent map $M' := M - E$ with a single face and whose graph is a subgraph of $G(M)$. In particular, the fundamental group of $G(M')$ generates $\pi_1(M', v)$, hence $\pi_1(M, v)$. (Here v is a vertex of M fixed once for all.) It follows from 1.2.12 and Example 3.2.7 that any basis for $\pi_1(G(M'), v)$ contains the minimal number of generators for $\pi_1(M, v)$.

If (M,b) has at least one perforated face, we let E^* be the set of edges of a spanning forest of the dual graph $G(M^*)$ with the property that every tree in the dual forest contains exactly one vertex dual to a perforated face of (M,b). Analogously to the previous case we delete the corresponding set E of edges in (M,b) to obtain a map (M'',b'') with perforated faces only. This time $\pi_1(G(M''), v) = \pi_1((M'',b'), v)$ because there is no face relation. A basis for $\pi_1(G(M''), v)$ thus gives a basis for $\pi_1((M,b), v)$.

By Lemma 1.3.1 such bases can be computed in $O(\beta_1(G)n)$ time, where $G = G(M')$ or $G = G(M'')$. It results from Propositions 3.5.19 and 3.5.22 that $\beta_1(G) = O(g+k)$. □

Incidentally, it results from Exercise 3.5.9 that the above graphs $G(M')$ and $G(M'')$ are cut graphs of M' and (M'',b'') respectively. Noting that $(M-E)\backslash\backslash G(M') = (M\backslash\backslash G(M')) - E$ and that $(M\backslash\backslash G(M')) - E$ is combinatorially equivalent to $M\backslash\backslash G(M')$, we conclude that each component of $M\backslash\backslash H$ is a disc, i.e. that $G(M')$ is also a cut graph for M. Similarly, $G(M'')$ is a cut graph of (M,b). In particular, this proves the existence of cut graphs. If T is the set of edges of a spanning tree of $G(M')$, and C is the complementary set of edges in $G(M')$, we have a partition $T \cup E \cup C$ of the edges of M such that $T \cup C$ are the edges of the cut graph $G(M')$ of M, and $E^* \cup C^*$ are the edges of a cut graph of M^*, where C^* are the edges dual to C. The partition (T, E, C) is sometimes called a **tree-cotree decomposition** [Epp03]. Likewise, in the case that (M,b) has at least one perforated face and E^* are the edges of a spanning forest of $G(M^*)$ as described above, we let T be the edges of a spanning tree of $G(M'')$ and C be the complementary set of edges in $G(M'')$. This time we call (T, E, C) a **tree-coforest decomposition**.

Computation of the homotopy class of a loop

We suppose given a tree-cotree decomposition (T, E, C) of a map M with basepoint v. In particular, the graph H induced by $T \cup C$ is a cut graph. We let B be the unique perforated face of $M \backslash\backslash H$. According to Theorem 3.5.23, $\pi_1(M, v)$ is isomorphic to $\langle C, r\rangle$ where r is the trace of ∂B over C. For $c \in C$ this isomorphism sends the loop $T[v, c]$ to c. More generally, the image of a homotopy class of a loop by this isomorphism is an element of $\langle C, r\rangle$ that can be represented by a word on $C \cup C^{-1}$. By the computation of a homotopy class we mean the computation of such a word. For an arc a of M, the homotopy class of $T[v, a]$ can be computed as follows: since H is spanning, all the vertices of $M \backslash\backslash H$ are incident to B. In particular, a cuts ∂B into two subpaths each homotopic to a. Let p_a be the image in M of one of these two paths. Then $T[v, a]$ is homotopic to $T[v, c_1] \cdot T[v, c_2] \cdots T[v, c_k]$ where $c_1, c_2, \ldots, c_k$ is the sequence of arcs in C along p_a (see 1.3.2). It follows that the homotopy class of $T[v, a]$ can be encoded by the word

$$W(a) := c_1 c_2 \ldots c_k \tag{3.13}$$

A loop $(a_1, a_2, \ldots, a_j)$ is thus encoded by the product of words $W(a_1)W(a_2)\ldots W(a_j)$. By Proposition 3.5.19 C contains $1-\chi(M) = O(g)$ edges. Since each edge in C appears twice in ∂B, we have that r, hence each $W(a_i)$, has length $O(g)$. In practice, we can encode each $W(a_i)$ implicitly by two pointers in r, viewed as a circular sequence, pointing to the occurrences in r of the first and last arc of $W(a_i)$. The encoding of the word $W(a_1)W(a_2)\ldots W(a_j)$ by a sequence of such pairs of pointers is called a **term product representation**. The proof of the next lemma is left to the reader; it follows easily from the above description.

Lemma 3.6.2. *Let M be a finite map with n edges and let v be a vertex of M. We can preprocess M in $O(n)$ time such that for any loop of j arcs in M its term product representation can be computed in $O(j)$ time. Moreover this representation is composed of at most j terms.*

Beware that this representation is not unique. For instance the trivial homotopy class is represented by the empty word as well as r, or any word in the normal subgroup of $\langle C \mid -\rangle$ generated by r. Deciding whether two loops are homotopic by comparing their word representations in $\langle C, r\rangle$ is not immediate. This last problem is referred to as the *word problem* in combinatorial group theory. See the last chapter of this part on the homotopy test.

A statement analogous to Lemma 3.6.2 holds for maps with at least one perforated face. This time we start with a tree-coforest decomposition (T, E, C) of a map (M, b) with basepoint v. Again, $T \cup C$ defines a cut graph H and according to Theorem 3.5.23, $\pi_1((M, b), v)$ is isomorphic to the free group over C. We let $B_1, B_2, \ldots, B_k$ be the cut faces of $(M, b)\backslash\backslash H$, where k is the number of perforated faces of (M, b). Since H is spanning every arc a of $(M, b)\backslash\backslash H$ has its endpoints on some cut face B_i. Because each component of $(M, b)\backslash\backslash H$ is an annulus, a is homotopic to one p_a of the two paths of ∂B_i sharing their endpoints with a. See Lemma 3.5.21 or Exercise 3.5.17. We let $W(a)$ be the concatenation of the arcs of C along p_a. Then $W(a)$ represents the homotopy class of $T[v, a]$ in $\langle C, -\rangle$. A loop $(a_1, a_2, \ldots, a_j)$ is thus encoded by the product of words $W(a_1)W(a_2)\ldots W(a_j)$, which we call again a term product representation. By Proposition 3.5.20 C contains $2-\chi(M) = O(g+k)$ edges, so that each $W(a_i)$ has length $O(g)$. In practice, we can encode each $W(a_i)$ implicitly by two pointers in ∂B, where B is the cut face of a component of $(M, b)\backslash\backslash H$ containing a_i.

Lemma 3.6.3. *Let (M,b) be a finite map with n edges and $k > 0$ perforated faces. Let v be a vertex of M. We can preprocess (M,b) in $O(n)$ time such that for any loop of j arcs in M its term product representation can be computed in $O(j)$ time. Moreover this representation is composed of at most j terms.*

This time the term product representation has a unique reduced representative in the free group $\langle C,-\rangle$. It is possible to obtain this reduced form in time proportional to the number of terms without expending each word. After $O(n)$ time preprocessing it is thus possible to test if two loops are homotopic in time proportional to their length [RL12].

Minimal weight homotopy basis

As for graphs in Section 1.6, we now assume that the edges of the map M are positively weighted. The weights trivially extend to paths by summing the weights of each occurrence of the edges in a path. We write $|\ell|$ for the weight of a path ℓ. Given a vertex v of M, a minimal weight basis of $\pi_1(M,v)$ is a set of homotopy class representatives that minimizes the total weight of each representative. Erickson and Whittlesey have proposed a simple algorithm to compute a minimal weight basis. The original description [EW05] relies on a discrete notion of cut locus and was further simplified [CdV10, Eri12]. The following presentation essentially relies on the same ideas recast in the combinatorial framework of maps. As a benefit, we get almost for free the extension of [EW05] to non-orientable maps (this is also the case of [CdV10]) and to maps with boundary.

Let v be a vertex of a finite connected map M and let T be a shortest path tree with root v in $G(M)$. If H is a subgraph of $G(M)$, we denote by E_H the set of edges of H. By convention, the set of dual edges in M^* of an edge set in M is denoted with a star superscript. Hence, E_T^* is the set of edges of M^* dual to the edges of T. We remark from Proposition 2.2.58 that the contraction of T in M gives a connected map with a one-to-one correspondence between the faces of M and M/T. It ensues from Lemma 2.2.64 that $G(M^* - E_T^*) = G((M/T)^*)$ is connected and spanning in $G(M^*)$. For each chord e of T, the loop $T[v,e]$ is a shortest loop through e with basepoint v and we define the weight of the edge e^* dual to e as

$$w(e^*) = |T[v,e]|$$

Finally, we consider a maximum weight spanning tree K^* of $G(M^* - E_T^*)$ and we let C be the set of edges primal to the chords of K^* in $G(M^*) - E_T^*$. Following the proof and discussion of Theorem 3.6.1, we have a tree-cotree decomposition (E_T, E_K, C) and the set of loops

$$\Gamma := \{T[v,e] \mid e \in C\}$$

is a basis of $\pi_1(M,v)$. Following [EW05], we call Γ a **greedy homotopy basis**. The name comes from a greedy computation of the maximum spanning tree K^* which makes the loops in Γ appear in a greedy fashion. It results from Exercise 3.4.10 that the set of homology classes of the loops in Γ is a basis of $H_1(M,\mathbb{Z}/2\mathbb{Z})$. A **greedy factor** of a loop ℓ with basepoint v is any loop in Γ which appears with a non-zero coefficient in the decomposition of ℓ in this homology basis.

Lemma 3.6.4. *For any chord e of T in $G(M)$, the weight $w(e^*) = |T[v,e]|$ is larger or equal to the weight of any of the greedy factors of $T[v,e]$.*

PROOF. The set of chords of T is the disjoint union $E_K \cup C$. If $e \in C$, then $T[v,e]$ is its own and unique greedy factor and the result is trivial. We now assume that $e \in E_K$. We put $C_1 := \{c \in C \mid w(c^*) \leq w(e^*)\}$ and $C_2 := C \setminus C_1 = \{c \in C \mid w(c^*) > w(e^*)\}$. We consider the connected subgraph $K_e^* := G(M^*) - (E_T^* \cup C_1^*) = K^* + C_2^*$ of $G(M^*) - E_T^*$. We claim that e^* is a bridge of K_e^*. Otherwise, e^* would belong to a cycle of K_e^*. This cycle would contain an edge c^* in C_2^* and exchanging e^* with c^* in K^* would give a spanning tree with strictly larger weight, contradicting the maximality of K^*. It follows that $K_e^* - e^* = G(M^*) - (E_T^* \cup C_1^* \cup \{e^*\})$ is not connected. It ensues by Proposition 3.5.8 that $M \backslash\backslash (T \cup C_1)$ is connected while $M \backslash\backslash (T \cup C_1 \cup \{e\})$ is not connected. Hence by Lemma 3.5.10 e appears exactly once in the boundary of each component of $M \backslash\backslash (T \cup C_1 \cup \{e\})$. Considering the formal sum of the faces of one component and its image by the boundary operator, we obtain that $e + \kappa$ is 0-homologous for some chain κ with support in $T \cup C_1$. We conclude that the greedy factors of $T[v,e]$ are contained in $\{T[v,c] \mid c \in C_1\}$, as desired. □

Lemma 3.6.5. *Let ℓ be a loop with basepoint v in $G(M)$. Any greedy factor of ℓ has weight at most $|\ell|$.*

PROOF. We consider ℓ as a loop of $G(M)$ and express its homotopy class in the free basis of $\pi(G(M), v)$ given by the chords of T in $G(M)$: $\ell \sim T[v,e_1] \cdot T[v,e_2] \cdots T[v,e_k]$. We assume this expression reduced, so that each e_i, $1 \leq i \leq k$, occurs at least once in ℓ. In particular, $|\ell| \geq w(e_i^*)$. Since any greedy factor of ℓ must occur as a greedy factor of some $T[v,e_i]$, we can apply Lemma 3.6.4 to e_i and conclude. □

We denote by $\gamma_1, \dots \gamma_{|C|}$ the loops in the greedy homology basis Γ. Similarly to Lemma 1.3.2, we can easily show that

Lemma 3.6.6. *For any basis $\{\ell_i\}_{1 \leq i \leq |C|}$ of $\pi(M,v)$, there exists a permutation τ of $\{1 \dots |C|\}$ such that for each $i \in \{1 \dots |C|\}$, the loop γ_i is a greedy factor of $\ell_{\tau(i)}$.*

Exercise 3.6.7. Prove the following generalization of the lemma. For any finite map with boundary (M,b), for any basis $\{\gamma_i\}$ of $\pi((M,b),v)$, and for any basis $\{h_i\}$ of $H_1((M,b),\mathbb{Z}/2\mathbb{Z})$, we can reorder the γ_i so that each h_j appears in the decomposition in $\{h_i\}$ of the homology class of γ_j.

It directly follows from the two preceding lemmas that

Proposition 3.6.8. *Any greedy homotopy basis is a minimal weight basis.*

In order to compute a greedy homotopy basis one needs to compute a shortest path tree and a maximum weight spanning tree. A shortest path tree of a graph with n edges can be computed in $O(n \log n)$ time using Dijkstra's algorithm. Classical maximum (minimum) weight spanning tree algorithms run in $O(n \log n)$ time [Tar83]. Since a homotopy basis of a map of genus g has $O(g)$ loops, and since each loop of a greedy basis may have size $O(n)$ we obtain

Theorem 3.6.9 ([EW05])**.** *Let M be a finite connected map of genus g without boundary with n weighted edges. Given a vertex v of M, a minimal weight basis of $\pi_1(M, v)$ can be computed in $O(n \log n + gn)$ time.*

Minimal weight basis of maps with boundary We can extend the computation of a minimal weight basis to maps with boundary. We consider a vertex v of a map (M, b) with at least one perforated face and a shortest path tree T of $G(M)$. We use the same notational conventions as for maps without boundary, noting that duality refers to the underlying map M without boundary. We denote by B^* be the set of vertices of M^* dual to the perforated faces of (M, b). Let K^* be a maximum weight spanning forest in $G(M^* - E_T^*)$ with the property that each tree component contains exactly one vertex of B^*. We now have a decomposition (E_T, E_K, C) of $E(M)$, where $C := E(M) \setminus (E_T \cup E_K)$. Following the proof of Theorem 3.6.1, the set of loops

$$\Gamma := \{T[v, e] \mid e \in C\}$$

is a basis of $\pi_1((M, b), v)$. We claim that Lemma 3.6.4 remains valid in this context. To see this, it is enough to consider $e \in E_K$. We let C_1, C_2, K_e^* be literally as in the proof of Lemma 3.6.4. Note that every component of $K_e^* = K^* + C_2^*$ contains at least one vertex of B^*. We claim that one component of $K_e^* - e^*$ does not intersect B^*. Otherwise, e^* would either belong to a cycle of K_e^* or lie on a simple path of K_e^* joining two vertices of B^*. In both case we could exchange e^* with an edge of C_2^* to obtain a spanning forest with strictly larger weight, contradicting the maximality of K^*. It follows that e^* disconnect a component of K_e^* and that one of the resulting two components has no vertex in B^*. Analogously to the proof of Lemma 3.6.4, we infer that $e + \kappa$ is 0-homologous for some chain κ with support in $T \cup C_1$ and conclude that $w(e^*)$ is larger or equal to the greedy factors of $T[v, e]$. Lemmas 3.6.5 and 3.6.6 remains true if M is replaced by (M, b). We can finally assert as in Proposition 3.6.8 that Γ is a minimal weight basis.

The computation of Γ only differs from the case of maps without boundary in the fact that K^* is a maximum spanning forest instead of a maximum spanning tree. However, it is a simple exercise to check that such a forest corresponds to a maximum spanning tree in the graph obtained from $G(M^* - E_T^*)$ by identifying all the vertices in B^* into a single vertex. The computation of K^* thus takes $O(n \log n)$ time and we may now generalized Theorem 3.6.9 to any finite map.

Theorem 3.6.10. *Let (M, b) be a finite connected map of genus g with n weighted edges and k perforated faces. Given a vertex v of M, a minimal weight basis of $\pi_1((M, b), v)$ can be computed in $O(n \log n + (g + k)n)$ time.*

Further improvements on the complexity in Theorem 3.6.9 are possible. For instance, using a more restricted notion of RAM, or restricting to minor closed graph families, a maximum weight spanning tree can be computed in linear time [FW94, Mar04]. See [EW05] for further details. Surprisingly, despite the close connection between cut graphs and homotopy basis, it was shown by Erickson and Har-Peled that computing a minimum weight cut graph is NP-hard [EHP04]. In a more relaxed version, a cut graph is a union of paths between a prescribed set of vertices. The paths may traverse an edge several times as long as an infinitesimal perturbation of the paths provides a topological cut graph. Colin de Verdière [CdV10] shows that a minimum weight cut graph can be computed in polynomial time for this notion of cut graph.

3.7 Some Elementary Algorithms Related to Homology

3.7.1 Computing a basis of the first homology group

Thanks to Proposition 3.4.4 we know that a homotopy basis provides a generating set for the first homology group of a map (M,b). When (M,b) is orientable or has at least one perforated face this generating set is a free basis of the free abelian group $H_1(M,b)$. This generating set becomes a vector space basis when considering homology with coefficients in a field. However, some care must be taken for non-orientable surfaces without boundary. We still get a basis with $\mathbb{Z}/2\mathbb{Z}$ coefficients (see Exercise 3.4.10) but this is no more the case for rational coefficients.

Theorem 3.7.1. *Let (M,b) be a finite connected map of genus g with n edges and k perforated faces. A basis of the first homology group of (M,b) can be computed in $O((g+k)n)$ time. The homology coefficients should be $\mathbb{Z}/2\mathbb{Z}$ for a non-orientable map without boundary. We can take the integers or any field of coefficients in the other cases.*

Computation of the homology class of a 1-cycle

We first consider a map M without boundary. As in Section 3.6.1 we suppose given a tree-cotree decomposition (T,E,C). Considering the appropriate homology coefficients, $\{T[c] \mid c \in C\}$ is a basis of the first homology group. For an arc a of M we can encode the homology class of $T[a]$ with the same word $W(a)$ as for the homotopy class of $T[v,a]$ (cf. (3.13)). This time however, one should consider $W(a)$ in the abelianization of the fundamental group, taking coefficients modulo 2 for $\mathbb{Z}/2\mathbb{Z}$ coefficients. Equivalently, one can consider $W(a)$ as a vector of $d = 2-\chi(M)$ coefficients (see Propositions 3.5.19). The coordinates of the homology class of any cycle $\sum_a \alpha_a a$ is now given by $\sum_a \alpha_a W(a)$. For a loop $\ell = (a_1, a_2, \dots, a_j)$ we can compute its homology class in $O(gj)$ time by adding the j vectors $W(a_1), W(a_2), \dots, W(a_j)$. It is possible to reduce this computation to $O(g+j)$ time as follows. We again consider the relation r defined as the trace over C of the facial circuit of the cut face of $M\backslash\backslash(T\cup C)$. We view r as a circular sequence of $O(g)$ arcs where each edge of C occurs twice, possibly with the same orientation when M is non-orientable. Each $W(a_i)$ corresponds to an interval defined by two pointers to its first ans last arc in the relation r. Note that the edges of M define $t := O(g)$ distinct intervals because their respective intervals do not cross. This is obvious for the edges in $T\cup C$ and results from the fact that $M\backslash\backslash(T\cup C)$ is a disc for the edges in E (or the fact that E^* is a tree). We let $W_1, \dots W_t$ be those intervals. The homology class of ℓ is thus equal to the weighted sum $\sum_i n_i W_i$ where n_i is the number of occurrences of W_i in the sequence $W(a_1), W(a_2), \dots, W(a_j)$. The coefficient of an arc occurrence in r is the sum of the weighs of the intervals W_i covering this arc. In a precomputation, we record for each arc occurrence in r the set of intervals W_i ending or starting at this occurrence. We can now obtain the coefficient of each arc in r by a simple sweep: after computing the coefficient of an arc occurrence, we deduce the coefficient of the next swept arc by adding the weights of the intervals starting at the current occurrence and subtracting the weights of the intervals ending at this occurrence. Since every edge $c \in C$ appears twice in r, we finally add or subtract the coefficients of the two occurrences of c, taking their orientation into account. (A more detailed implementation is given in my course notes [Laz12b].)

Proposition 3.7.2. *Let M be a finite map of genus g with n edges. We can preprocess M in $O(n)$ time so that for any loop of j arcs in M the coefficients of its homology class (in the appropriate basis) can be computed in $O(j+g)$ time.*

In order to test if two loops are homologous it is sufficient to compare their vectors of coefficients. Hence, the homology test can be performed in time proportional to the length of the loops plus a term proportional to the genus g of the map. It is somehow surprising that the complexity of the homotopy test does not contain this last $O(g)$ term [LR12, EW13]. One would expect that the homology test is easier than the homotopy test. It is possible though that the present homology test is not optimal.

Open problem: Design a more efficient homology test or prove that the present test is optimal.

Exercise 3.7.3. We consider $\mathbb{Z}/2\mathbb{Z}$ coefficients. What is the complexity of constructing the homology covering as in Example 3.3.6? Describe a homology test based on this covering. What is its complexity?

For maps with boundary, we can proceed much the same way as for maps without boundary. Let k be the number of perforated faces of a map (M,b) of genus g. We start with a tree-coforest decomposition (T,E,C) of (M,b) and denote by $B_1, B_2, \ldots, B_k$ the cut faces of $(M,b)\backslash\backslash(T\cup C)$. The trace over C of each facial circuit ∂B_i provides a cyclic sequence of arcs r_i. Every arc a of (M,b) determines a subword $W(a)$ of some r_i representing the homology class of $T[a]$. Considering these subwords as intervals in the r_i, we can accumulate the total number of times each arc occurrence is covered by the intervals of the arcs of a given loop ℓ. Using a sweep algorithm as above, this can be done in $O(|\ell|+g+k)$ time. We eventually group the coefficients of the two occurrences of each $c \in C$ to get the vector of coefficients of the homology class of ℓ with respect to the basis $\{T[c] \mid c \in C\}$.

Proposition 3.7.4. *Let (M,b) be a finite map of genus g with n edges and k perforated faces. We can preprocess (M,b) in $O(n)$ time so that for any loop of j arcs in (M,b) the coefficients of its homology class (in the appropriate basis) can be computed in $O(j+g+k)$ time.*

Minimal weight homology basis

We assume given a positive weight function $w : E(M) \to \mathbb{Q}_+^*$ defined over the edges of a finite map M. With a little abuse of terminology, a set of 1-cycles whose homology classes form a basis of $H_1(M,\mathbb{Z}/2\mathbb{Z})$ is still called a basis. Analogously to Section 1.6, we look for a basis of $H_1(M,\mathbb{Z}/2\mathbb{Z})$ such that the sum of the weights of the cycles in the basis is minimal. Since $H_1(M,\mathbb{Z}/2\mathbb{Z})$ is a vector space, the greedy matroidal algorithm of Lemma 1.6.2 remains valid. For each vertex of M, $v \in V(M)$, we let T_v be a shortest path tree rooted at v. Following Lemmas 1.6.2 and 1.6.7, we can restrict the greedy algorithm to simple cycles of the form $T_v[v,e] = T_v[e]$ for $(v,e) \in V(M) \times E(M)$, where each edge is assumed a default orientation. In fact, we can further restrict this set of candidate cycles to a subset of $O(g)$ cycles.

Lemma 3.7.5. *The set of loops $\mathcal{L}_v = \{T_v[e] \mid e \in E(M)\setminus E(T)\}$ contains at most $3(1-\chi(M))$ distinct homology classes. Moreover, we can select in $O(n \log n)$ time a subset $\mathcal{S}_v \subset \mathcal{L}_v$ of at most $3(1-\chi(M))$ loops that contains the homologous loop of minimal weight of each of the distinct homology classes in $\mathcal{L}_v$.*

Proof. We denote by $e^* \in E(M^*)$ the edge dual to $e \in E(M)$ and we let $E^*(H)$ be the dual set of the edges of a subgraph H of $G(M)$. We know from the description of tree-cotree decompositions in Section 3.6.1 that $K^* := G(M^*) - E^*(T_v)$ is a cut graph of M^*. We thus have $\beta_1(K^*) = 2 - \chi(M)$ by Proposition 3.5.19. If $e_1^*, \dots, e_k^*$ are the edges incident to a vertex of K^* dual to a face F of M then $\partial_2 F = \sum_i T_v[e_i]$ (see the proof of Proposition 1.4.4), so that $\sum_i T_v[e_i]$ is null-homologous. It follows that $T_v(e)$ is null-homologous whenever e^* is a pendant edge in K^*. We can further delete recursively all the pendant edges in K^* since their corresponding cycle is null-homologous. We are left with a subgraph K_1^* without degree one vertex and with the same cyclomatic number as K^*. If two edges e^* and e'^* share a degree two vertex in K_1^* we also have that $T_v[e]$ and $T_v[e']$ are homologous. It follows that the number of distinct homology classes is at most the number m of branches, i.e., of maximal chains of edges linked by degree two vertices in K_1^*. Contracting the edges incident to a degree two vertex in K_1^* we obtain a graph K_2^* with n_e edges combinatorially equivalent to K_1^*. Because each of the n_v vertices of K_2^* has degree three or more, we have $2n_e \geq 3n_v$ by double counting of the vertex-edge incidences. On the other hand,

$$2 - \chi(M) = \beta_1(K^*) = \beta_1(K_1^*) = \beta_1(K_2^*) = 1 - (n_e - n_v)$$

It ensues that $n_e \leq 3(n_e - n_v) = 3(1-\chi(M))$ as desired. In practice, we first compute T_v and the distance of each vertex to the root v in $O(n \log n)$ time using Dijkstra's algorithm. For any arc a of M, the length of $T_v[v,a]$ can then be computed in constant time. We recursively remove the pendant edges of K^* and traverse the resulting graph K_1^* in linear time, only keeping in $\mathcal{S}_v$ the loop $T_v[v,e]$ corresponding to the traversed edge e^* if the loop has minimal weight in its branch. □

The greedy matroidal algorithm requires to test if a loop is homologically independent of the already selected loops. To this end we consider a fixed homology basis $\mathcal{B} := \{T[c] \mid c \in C\}$ associated to a tree-cotree decomposition (T,E,C) as explained for the above *Computation of the homology class of a 1-cycle.*

Lemma 3.7.6. *We can compute the homology coordinates with respect to $\mathcal{B}$ of each of the loops in $\mathcal{S}_v$ in $O(g\,n)$ total time.*

PROOF. We first compute for each edge e of M, the coordinates of $T[e]$ with respect to $\mathscr{B}$. This can be done in $O(gn)$ time for all the edges in E by a simple traversal of the dual tree E^*. In fact, only $O(n+g^2)$ time is needed since the edges of M determine $O(g)$ distinct homology classes (see the above paragraph on the computation of homology classes). We then traverse T_v from its root v in order to compute for each vertex x of T_v the homology coordinates with respect to $\mathscr{B}$ of the loop $\gamma_v(x) := T_v[v.x] \cdot T[x,v]$ composed of the two $(x-v)$-paths in T_v and T respectively. The traversal needs $O(gn)$ time, spending $O(g)$ time per vertex to update the coordinates of $\gamma_v(x)$ for the successor of x in T_v. The coordinates of a loop $T_v[e]$ in $\mathscr{S}_v$ can now be decomposed into the sum of the coordinates of $\gamma_v(x)$, $T[e]$ and $\gamma_v(y)$ where x, y are the endpoints of e. It thus takes $O(g)$ time to compute the coordinates of any loop in $\mathscr{S}_v$ and the whole computation needs to $O(gn)$ time. □

Theorem 3.7.7 ([EW05]). *Let M be a finite connected map of genus g with n weighted edges. A minimal weight basis of $H_1(M, \mathbb{Z}/2\mathbb{Z})$ can be computed in $O(n^2 \log n + gn^2 + g^3 n)$ time.*

PROOF. We can select $O(gn)$ loop candidates for the minimal weight basis and compute their weight in $O(n^2 \log n)$ time according to Lemma 3.7.5. Their homology coordinates with respect to $\mathscr{B}$ is computed in $O(gn^2)$ time following Lemma 3.7.6. After sorting the $O(gn)$ candidate loops according to their weight, the greedy algorithm consists in scanning the candidate loops in increasing order, keeping the scanned loop in the minimal basis if it is homologically independent of the previously selected loops. This last test can be answered in $O(g^2)$ time using Gauss elimination to maintain the $O(g)$ selected loops in row echelon form. The whole scan thus takes $O(g^3 n)$ time. Summing up all the steps and noting that $g = O(n)$ we may conclude the theorem. □

Minimal weight basis of maps with boundary The previous algorithm easily extends to a map with k perforated faces if we replace g by $g+k$ in the above description. The reader is invited to fill in the algorithm details.

Theorem 3.7.8. *Let (M, b) be a finite connected map of genus g with n weighted edges and k perforated faces. A minimal weight basis of $H_1((M,b), \mathbb{Z}/2\mathbb{Z})$ can be computed in $O(n^2 \log n + (g+k)n^2 + (g+k)^3 n)$ time.*

Chapter 4

Curves on Surfaces

Contents

Here, we discuss curves on combinatorial surfaces from a computational point of view. More generally, we provide a formal framework to represent immersions of graphs in a map. In the preceding chapters, curves on surfaces were implicitly identified as paths in the vertex-edge graph of maps. When dealing with curve intersections, this representation becomes ambiguous. If a path goes twice through a vertex or an edge, shall we consider that the path self-intersects? Looking for a faithful image of the topology of curves we must consider such type of paths simply because some homotopy classes do not have simple closed paths as a representative (think of the powers of a non-trivial homotopy class). In fact, even the homotopy class of a simple continuous curve may not have a simple combinatorial counterpart. For a positive genus surface has an infinite number of simple continuous loops with distinct homotopy classes, while a finite map has a finite number of loops without repeated vertices. It seems difficult to be categorical on what should be the best way to represent curves on surfaces. It depends on the kind of curve properties we are interested in since the datastruture should obviously encode those properties. It also depends on the type of computations we want to perform as some datastructures might lead to more efficient algorithms than others. For instance, simple curves on triangulated surfaces have a very concise representation known as *normal coordinates*. Those type of coordinates were introduced by Kneser [Kne29] and Haken [Hak61] for studying three-manifold topology. Using quite intricate arguments, Agol et al. [AHT02] and Schaefer et al [SSv02, SSS08] were able to answer queries such as the number of isotopy classes among the components of a (non necessarily connected) normal curve. Their approach was greatly simplified by Erickson and Nayyeri who also describe very nice algorithms to convert normal curve coordinates into a piecewise-linear representation [EN13]. However, all the known algorithms for processing normal coordinates on triangulated surfaces run in at least quadratic time in the number of triangles.

Another curve representation was proposed by Colin de Verdìere and Erickson [CE10]. They start with *cross-metric surface*, that is a topological realization of a combinatorial surface together with an embedding of its dual graph. They only consider continuous curves in "generic positions". In particular, they require the curves to intersect the dual graph transversely, away from its vertices and a finite number of times. On the computational side, one should maintain an arrangement of the curves and the dual graph. For simple curves, this is actually what was implemented in [LPVV01]. Most of the topological arguments on cross-metric surfaces relies on topological properties of the underlying continuous surface. In the spirit of this document I have tried to avoid the recourse to continuous arguments and I develop the purely combinatorial approach proposed in [CdVL05]. It is not clear how it compares with the cross-metric surface framework as far as implementation is concerned. It seems adequate at least for the computation of a canonical system of loops.

4.1 Drawing Graphs on Maps

4.1.1 Combinatorial graph drawing

Intuitively, a drawing of a graph on a map consists in representing each edge of the graph by a path in the 1-skeleton of the map. In this section M will represent a map, possibly with boundary.

Definition 4.1.1. A **drawing** of a graph H in a map M is a graph morphism $f: H' \to G(M)$ from a subdivision H' of H to the graph of M.

Every arc h of H corresponds to a path p_h in H'. With a little abuse of notation, we write $f(h)$ for the image of this path. In particular, $f(h^{-1})$ is the reverse path $f(h)^{-1} = f(p_h^{-1})$. Let $f(h) = (a_1, a_2, \ldots, a_k)$ be composed of $k = |f(h)|$ arcs. If e is an arc of M, an **arc occurrence** of e in $f(h)$ is a pair (h,i), $i \le k$, such that $a_i = e$. We denote by $\mathcal{O}_f(e)$, or $\mathcal{O}_H(e)$ when the drawing f is implicit, the set of arc occurrences of e:

$$\mathcal{O}_f(e) := \{(h,i) \mid h \in A(H) \text{ and the } i\text{th arc of } f(h) \text{ is } e\}$$

The set of arc occurrences is denoted by $\mathcal{O}_f := \bigcup_{e \in A(M)} \mathcal{O}_f(e)$, or by $\mathcal{O}_H$.

Remark 4.1.2. A path p in M may be considered as a drawing of a graph reduced to an edge. Similarly, a circuit c may considered as a drawing of a loop-edge. We will thus write $\mathcal{O}_p$ and $\mathcal{O}_c$ for the set of arc occurrences of p and c respectively.

If (h,i) is an arc occurrence of e, the mapping $(h,i) \mapsto (h^{-1}, |f(h)|+1-i)$ is a bijection between $\mathcal{O}_f(e)$ and $\mathcal{O}_f(e^{-1})$. We denote by u^{-1} the image by this mapping of an occurrence $u \in \mathcal{O}_f(e)$.

Definition 4.1.3. The first arc occurrence $(h,1)$ of $f(h)$ is said **extremal**. Its other arc occurrences are said **internal**. The **binding vertex** of the extremal occurrence $(h,1)$ is the vertex $o(h) \in V(H)$. We define a binary relation over $\mathcal{O}_f$ by

$$u \,\mathrm{R}_f\, w$$

if and only if one of the following two situations holds:

- u and w are distinct extremal arc occurrences sharing a common binding vertex,
- $u = (h,i)$ is internal and w is internal with $w^{-1} = (h, i-1)$.

The binary relation R_f is symmetric and restricts to a fixed point free involution over the internal arc occurrences. For such occurrences we can thus write $w = \mathrm{R}_f(u)$ instead of $u \,\mathrm{R}_f\, w$. We shall drop the f subscript in R_f when there is no ambiguity on the drawing.

4.1.2 Graph immersions

As discussed above, one needs extra information to get a proper notion of crossing. For this purpose, we shall define an analog of curve immersions and embeddings.

Definition 4.1.4. An **immersion** I of a graph H in a map M is a drawing f of H in M together with the data for each arc e of M of a total ordering $\preceq_e^I$ over $\mathcal{O}_f(e)$. We require that the order for e^{-1} corresponds to the order for e or its reverse according to whether the signature $s(e)$ of e is negative or positive. In other words,

$$\forall u, w \in \mathcal{O}_f(e), \qquad u \preceq_e^I w \Leftrightarrow w^{-1} (\preceq_{e^{-1}}^I)^{s(e)} u^{-1}$$

Remark 4.1.5. An immersion of a graph H in M induces an immersion of any subgraph. Moreover, if the drawing of H does not use an edge e of M, then any immersion of H in M restricts to immersion of H in $M - e$.

When there is no ambiguity on the immersion we may write $\preceq_e$ for $\preceq_e^I$. We also denote by $\lessdot_e$ the corresponding covering relation ($u \lessdot_e w$ if w is the direct successor of u for $\preceq_e$). Let $\mathcal{O}_f(v) := \bigcup_{o(e)=v} \mathcal{O}_f(e)$ be the set of arc occurrences incident to a vertex v of M. The immersion I induces a cyclic permutation π_v^I of $\mathcal{O}_f(v)$ obtained by concatenating the orderings $\preceq_e$ of the arcs e incident to v according to the rotation system ρ of M. More precisely, denoting by $h(i)$ the ith arc of $f(h)$, we have

$$\pi_v^I(h,i) = (k,j) \Leftrightarrow \begin{cases} \text{either } h(i) = k(j) \text{ and } (h,i) \lessdot_{h(i)} (k,j), \\ \text{or } \exists k > 0 : k(j) = \rho^k(h(i)),\ (h,i) \text{ is maximal for } \preceq_{h(i)}, \\ (k,j) \text{ is minimal for } \preceq_{k(j)}, \text{ and } \forall n \in [1,k) : \mathcal{O}_f(\rho^n(h(i))) = \emptyset. \end{cases}$$

Restricting the cycle of π_v^I to the internal arc occurrences, we get the **word of internal occurrences** of I at v, that we denote by W_v^I. It is a cyclic word over the internal occurrences of $\mathscr{O}_f(v)$ defined up to cyclic permutation. The relation R defines a pairing in W_v^I. A **crossed pair** is a subword of W_v^I of the form

$$u\,w\,\mathrm{R}(u)\,\mathrm{R}(w)$$

We are now ready to define a crossing of an immersion.

Definition 4.1.6. Let I be an immersion of H. A **crossing**, or **intersection** of I at $v \in V(M)$ is quadruplet of arc occurences $(u_1, u_2, u_3, u_4) \in (\mathscr{O}_H(v))^4$ satisfying the following conditions:

- Not all of u_1, u_2, u_3, u_4 are extremal with a common binding vertex,
- $u_1 \,\mathrm{R}\, u_3$ and $u_2 \,\mathrm{R}\, u_4$,
- u_1, u_2, u_3, u_4 appears in this cyclic order for π_v^I, i.e., there exist three positive integers $k_1 < k_2 < k_3 < |\mathscr{O}_H(v)|$ such that

$$u_i = (\pi_v^I)^{k_i}(u_1), \quad i = 2, 3, 4$$

Exercise 4.1.7. Let c be a circuit without repeated vertices in M. Consider the circuit c^2 obtained by concatenating c with a copy of itself. Assuming that c is two-sided, show that any immersion of c^2 has a crossing. What can you say if c is one-sided? Answer the same questions replacing c^2 by any power c^k of c, $k \geq 2$.

Graph embeddings

Definition 4.1.8. An **embedding** of a graph H in a map M is a graph immersion without crossing whose corresponding drawing is dimension preserving (an arc of H is drawn as a non-constant path in M).

If I is an embedding, the word of internal occurrences W_v^I at each vertex v has no crossed pair. Viewing paired occurrences as unoriented parentheses, W_v^I is well-parenthesized and can be represented by a labelled plane tree such that the labels of an arc and of its opposite are related by R. More precisely, W_v^I corresponds to an oriented sphere map whose graph is a labelled tree. The word reads as the labels of the facial circuit with positive arc signature of the unique face of the sphere map. We denote by T_v^I the tree associated to W_v^I.

Under some mild hypothesis, a drawing of a path or circuit has a unique embedding, if any.

Lemma 4.1.9. *Let c be a path or circuit of M such that*

1. *c has no spur,*
2. *if c is a path, it can not be factored as $c = p \cdot q \cdot p$ where p is a non-empty path,*
3. *if c is a circuit, it can not be written as a power $c = p^k$ of any proper subpath p.*

Then c has at most one embedding.

PROOF. We assume that c has at least two arcs since the result is trivial for a single arc. If c is a circuit we extend the pairing of internal arc occurrences to its two external occurrences. Let I and J be two embeddings of c. Suppose for a contradiction that $I \neq J$ and let u, v be two arc occurrences whose order is opposite in I and J. If c is a path, we cannot have u and v^{-1} both external by condition 2. Considering u^{-1}, v^{-1} instead of u, v if necessary, we can thus assume that u and v are internal. In this case, or if c is a circuit, $\mathrm{R}_c(u)$ and $\mathrm{R}_c(v)$ must be occurrences of a same arc and must have opposite order for I and J. For otherwise I or J would have a crossing. Now, if u and v traverse c in opposite directions, then it follows by induction on the length of the path between u and v^{-1} along c that c has a spur, contradicting condition 1. On the other hand, if u and v traverse c in the same direction we conclude this time that condition 2 or 3 is not satisfied depending on whether c is a path or a circuit. □

Lemma 4.1.10. *Let c be a path or circuit of M. We can decide in $O(|c|)$ time if c satisfies the conditions of Lemma 4.1.9. If so, we can decide in $O(|c|\log|c|)$ time if c has an embedding and in this case, we can construct the embedding in the same amount of time, storing for each arc a of M the ordered list $\mathscr{O}_c(a)$.*

PROOF. We can check that a circuit c is a power by verifying that c is a factor of cc that is not a prefix nor a suffix. This can be done in linear time with the Knuth-Morris-Pratt algorithm [CLRS09]. This algorithm starts computing the longest proper prefix p of c that is also a suffix of c. Hence, condition 2 of Lemma 4.1.9 is satisfied if and only if p is the empty word. We now assume that c satisfies the conditions of Lemma 4.1.9. We recursively compute an immersion of subpaths of c of increasing lengths. We initialize for every arc a an empty search structure $\mathscr{O}_c(a)$. We start inserting the occurrence $(c,1)$ into $\mathscr{O}_c(c(1))$ as well as $(c,1)^{-1}$ into $\mathscr{O}_c(c(1)^{-1})$. We assume computed an immersion of $c(1\ldots k-1)$ for some $k>1$. In order to extend the immersion we need to insert (c,k) into $\mathscr{O}_c(c(k))$. If $\mathscr{O}_c(c(k))$ is empty there is nothing to do, otherwise we need to decide for each $u \in \mathscr{O}_c(c(k))$ how it compares with (c,k) if the immersion was an embedding. There are two possibilities.

- Either u is internal, in which case $\mathrm{R}_c(u)$ exists in the current immersion. In particular, $\mathrm{R}_c(u) \neq (c,k)$ by condition 1. If $\mathrm{R}_c(u)$ and $(c,k-1)^{-1}$ are occurrences of a same arc, we may utilise their relative ordering to infer the ordering of u and (c,k). Otherwise we can use the circular ordering of the arcs of $\mathrm{R}_c(u)$ and $(c,k-1)^{-1}$ to compare u and (c,k).

- Or, $u=(c,1)$. We then consider the largest j_k such that $i \le j_k \implies c(k+i)=c(i)$. If c is a path, condition 2 implies $k+j_k<|c|$, while if c is a circuit, condition 3 implies $j_k<|c|$. The relative order of u and (c,k) must be the same, up to the accumulation of the arc signatures, as the relative order of (c,j_k) and $(c,k+j_k)$. This last ordering is inferred from their paired occurrences.

From the next exercise 4.1.11, j_k can be precomputed for each k in amortized constant time. We finally insert $(c,k)^{-1}$ into $\mathscr{O}_c(c(k)^{-1})$ taking into account the ordering of u and (c,k) and the signature of $c(k)$. Assuming that the arcs are indexed around each vertex, we can compare arc indices in constant time when necessary. The insertion of (c,k)

thus takes $O(\log|c|+j_k)$ time and the overall immersion is computed in $O(|c|\log|c|)$ time.

It is easily seen that the computed immersion coincide with the unique embedding of c if it exists. It remains to decide if this immersion is indeed an embedding. To this end we can check for each vertex that no crossing occurs at that vertex, i.e., that the word of internal occurrences is balanced for the pairing. This takes $O(|c|)$ times in total. □

Exercise 4.1.11. Let $W = W_1, W_2, \dots, W_n$ be a word of length n over a fixed alphabet. Denote by $W^i = W_i W_{i+1} \dots W_n$ the ith suffix of W. Let p be the prefix function that maps each $i \in [1,n]$ to the length of the longest prefix common to W and W^i. Hence, $p(1) = n$. Propose a linear time algorithm to compute p.

4.2 Complexity of Drawings and Immersions

We measure the complexity of a path p by its number $|p|$ of arcs. More generally, if f is a drawing of a graph H, its complexity is defined as

$$|f| = \frac{1}{2}\sum_{h\in A(H)} |f(h)|$$

We define two multiplicity parameters that allow to relate the complexity of a drawing with the complexity of the underlying map.

Definition 4.2.1. Let f be a drawing of a graph H in M. The **multiplicity** of an arc e of M with respect to f is the number of internal arc occurrences in $\mathcal{O}_f(e)$. It is denoted by $\mu_a(f,e)$ or $\mu_a(H,e)$ when there is no ambiguity on the drawing. The **arc multiplicity**, denoted by $\mu_a(f)$ (or $\mu_a(H)$) is the maximal multiplicity of any arc of M.

Note that $\mu_a(f,e) = \mu_a(f,e^{-1})$ when neither $\mathcal{O}_f(e)$ nor $\mathcal{O}_f(e^{-1})$ contain extremal occurrences. Let $E(M)$ and $E(H)$ be the set of edges of M and H respectively.

Lemma 4.2.2. *For f a drawing of H in M, we have*

$$|f| \le \mu_a(f)|E(M)| + |E(H)|$$

Proof. Let $\eta(f,e)$ be the number of extremal occurrences in $\mathcal{O}_f(e)$. We have

$$2|f| = \sum_{e\in A(M)} (\mu_a(f,e)+\eta(f,e)) = \sum_{e\in A(M)} \mu_a(f,e) + |A(H)| \le \mu_a(f)|A(M)| + |A(H)|$$

□

Definition 4.2.3. Let I be an embedding of H in M. The **nesting multiplicity** at a vertex v of M with respect to I is the maximal number of nested pairings in (any cyclic permutation of) W_v^I, i.e. the largest k such that W_v^I has a subword of the form

$$u_1, u_2, \dots, u_k\, \mathrm{R}(u_k) \dots \mathrm{R}(u_2)\mathrm{R}(u_1)$$

We denote by $\mu_n(I,v)$ the nesting multiplicity at v and by $\mu_n(I)$ the maximum of $\mu_n(I,v)$ over all the vertices of M.

It is an easy exercise to check that $\mu_n(I,v)$ is the diameter of the tree T_v^I associated to W_v^I.

Definition 4.2.4. Let I be an embedding of H in M an let e be an arc of M. The embedding is said **tight** along e if for any u,w in $\mathcal{O}_H(e)$ such that $w = \mathrm{R}(u)$ and $u \preceq_e^I w$ there exists an extremal arc occurrence $z \in \mathcal{O}_H(e)$ such that $u \preceq_e^I z \preceq_e^I w$. An embedding is tight if it is tight along every arc.

If I is not tight along e, there is a spur u,w such that $w = \mathrm{R}(u)$ and $u \preceq_e^I w$. Removing that spur leaves an embedding whose arc and nesting multiplicities can only decrease. Recursively removing such spurs, we get a tight embedding with lower multiplicities. we can thus assume that an embedding is tight when discussing upper bounds on multiplicities.

Proposition 4.2.5. *Let I be an embedding of H in M that is tight along e and let k be the number of distinct binding vertices of the extremal arc occurrences of e. Then, if $k = 0$*

$$\mu_a(H,e) \leq \mu_n(I,o(e))$$

else,

$$\mu_a(H,e) \leq \frac{k^2+2k-1}{k}\mu_n(I,o(e))$$

Proof. We denote by $\mathcal{O}_H^{int}(e) \subset \mathcal{O}_H(e)$ the set of internal occurrences of e. We also set

$$\mathcal{O} = \{u \in \mathcal{O}_H^{int}(e) \mid \mathrm{R}(u) \in \mathcal{O}_H^{int}(e)\}$$

If k is null then $\mathcal{O}$ must be empty because I is tight along e. It follows that all internal occurrences of e together with their paired occurrences constitute as many nested pairs, so that

$$\mu_n(I,o(e)) \geq |\mathcal{O}| = \mu_a(H,e)$$

If $k \neq 0$ we consider the restriction W to $\mathcal{O}$ of the occurrence word $W_{o(e)}^I$. We denote by T the plane labelled tree representing W. Because I is tight along e, for each factor $u\,\mathrm{R}(u)$ in W there must be an extremal occurrence w such that $u \preceq_e^I w \preceq_e^I \mathrm{R}(u)$. Moreover, if z is another extremal occurrence with the same binding vertex as w then $u \preceq_e^I z \preceq_e^I \mathrm{R}(u)$. It ensues that T has at most $k+1$ leaves. Because T has $|\mathcal{O}|/2$ edges, its diameter is bounded below by $|\mathcal{O}|/(k+1)$ (see Exercise 4.2.7 below).

On the other hand, by the pigeonhole principle, one of the extremal occurrences of e must be surrounded by at least $|\mathcal{O}|/(2k)$ internal occurrence pairs. Those pairs are nested with at least half of the remaining occurrences in $\mathcal{O}_H^{int}(e) \setminus \mathcal{O}$. We infer that

$$\mu_n(I,o(e)) \geq \max\{\frac{|\mathcal{O}|}{k+1}, \frac{|\mathcal{O}|}{2k} + \frac{\mu_a(H,e) - |\mathcal{O}|}{2}\}$$

The minimum of the right hand side is reached when its two expressions are equal. This allows to conclude that

$$\mu_n(I,o(e)) \geq \frac{k}{k^2+2k-1}\mu_a(H,e)$$

□

Exercise 4.2.6. Let I be an embedding of a drawing f of H in M. Show how to compute a tight embedding J of a drawing g of H homotopic to f ($g(h) \sim f(h)$ for each arc h of H) in $O(|f|+|AM|$ time. Show that the nesting multiplicity of the vertices do not increase: $\mu_n(J,v) \leq \mu_n(I,v)$ for each vertex v. It will be assumed that the following operations can be performed in constant time.

- given an arc e of M, return the next arc turning about its origin vertex,
- access the next to a an arc occurrence along the image path of an arc of H,
- return the successor of predecessor of an arc occurrence of e for the ordering $\preceq_e^I$.

Exercise 4.2.7. Show that the diameter of a tree with k leaves and n edges is at least $2n/k$.

Exercise 4.2.8. Let M be a map each of whose vertices have degree at least 3. Let p be a subpath of a facial circuit of M. Show that p has a unique embedding and that its nesting multiplicity is at most 2. (Hint: you may first show that any two paired internal occurrences in $\mathcal{O}_p$ are neighbours in the occurrence word of its origin vertex.)

Exercise 4.2.9. Let $\mathscr{M}$ be a piecewise linear surface with n edges and let M be the corresponding combinatorial map. Any drawing f of a graph H in M can be realized trivially by a piecewise linear function sending each edge e of H to the piecewise linear path in $\mathscr{M}$ corresponding to $f(e)$. Let I be an embedding of H in M. Show how to construct a piecewise linear embedding of H in $\mathscr{M}$ with $O(\mu_n(I)n)$ segments that is homotopic to the above trivial realization of the drawing of I. For each n, give an example of a surface $\mathscr{M}$ with a graph embedding I in M whose arc multiplicity $\mu_a(I)$ equals 1 and such that any infinitesimal perturbation of its trivial realization has $\Omega(n^2)$ segments. What can you tell about the nesting multiplicity of your embeddings?

4.2.1 Multiplicity and basic operations

Let M' be the result of one of the basic operations of Definition 2.3.6 applied to M. Given a drawing f of H in M we obtain a drawing of H in M' by composing f with a mapping t defined as follows.

- If M' is obtained from M by a face subdivision, we may take for t the inclusion of the set of the arcs of M into the arcs of M',
- if $M' = M - e$, where e is regular, we take for t the mapping D_e as defined in Lemma 3.2.10,
- if $M' = M/e$, where e is a non-loop edge, we take for t the mapping C_e of Lemma 3.2.10.

In each of these cases, an embedding I of the drawing f of H in M gives rise to an embedding of the drawing $t \circ f$ in M'. (In the case of the contraction of e, it is assumed that the image of an edge of H is not reduced to e.) The embedding resulting from an edge deletion or contraction is denoted by $I - e$ or I/e respectively. When M' is obtained by the reverse of an edge contraction, the construction of an embedding of

H in M' needs more care. We consider an embedding I of H in M corresponding to a drawing f and we assume that $M = M'/e$ where e is a non-loop edge of M'. We also consider a vertex mapping $f_0 : V(H) \to V(M')$ such that $f_{|V(H)} = C_e \circ f_0$.

Lemma 4.2.10. *There exists an embedding J of H in M' corresponding to a drawing g such that $g_{|V(H)} = f_0$ and $C_e \circ g = f$. Moreover, viewing the arc occurrences of f as arc occurrences of g, we can further impose that $J_{|\mathcal{O}_f} = I$. If I is tight, so is J.*

PROOF. The proof reduces to a simple construction with some case analysis. We first fix some notations: $x := o(e)$, $y = o(e^{-1})$, and $z := C_e(x) = C_e(y)$. We also consider the subsets of arcs of M' given by $A_x := o^{-1}(x) \setminus \{e\}$ and $A_y := o^{-1}(y) \setminus \{e^{-1}\}$. From the inclusion $A(M) \subset A(M')$ we may as well view A_x and A_y as subsets of $A(M)$. We designate by A_x^{-1} and A_y^{-1} the set of opposite arcs of A_x and A_y. Putting $\mathcal{O}_x = \bigcup_{a \in A_x} \mathcal{O}_f(a)$ and $\mathcal{O}_y = \bigcup_{a \in A_y} \mathcal{O}_f(a)$, we finally denote by $\preceq_x$ and $\preceq_y$ the orderings induced by the restriction of $\pi_f(z)$ to $\mathcal{O}_x$ and $\mathcal{O}_y$, respectively.

Let $h \in A(H)$, we define $g(h)$ by applying substitutions in $f(h) = (a_1, a_2, \ldots, a_k)$. In each of the following cases substitute (a_1) with

i. (e^{-1}, a_1) if $a_1 \in A_x$ and $f_0(o(h)) = y$,

ii. (e, a_1) if $a_1 \in A_y$ and $f_0(o(h)) = x$,

iii. (e, e^{-1}, a_1) if $a_1 \in A_x$ and $f_0(o(h)) = x$ and $(h,1) \preceq_x u \preceq_x v$ for some extremal occurrences $u, v \in \mathcal{O}_x$ such that $(h,1)\,\mathrm{R}_f\, v$ and such that the binding vertex s of u satisfies $f_0(s) = y$,

iv. (e^{-1}, e, a_1) if we are in the previous situation after exchanging the role of x and y.

Then for $i \in [1, k-1]$ substitute (a_i, a_{i+1}) with

v. (a_i, e, a_{i+1}) if $a_i \in A_x^{-1}$ et $a_{i+1} \in A_y$,

vi. (a_i, e^{-1}, a_{i+1}) if $a_i \in A_y^{-1}$ et $a_{i+1} \in A_x$,

vii. $(a_i, e, e^{-1}, a_{i+1})$ if $a_i \in A_x^{-1}$ and $a_{i+1} \in A_x$ and if there exists an extremal occurrence $u \in \mathcal{O}_x$ with binding vertex s such that $f_0(s) = y$ and such that u is in-between $(h,i)^{-1}$ and $(h, i+1)$ for $\preceq_x$,

viii. $(a_i, e^{-1}, e, a_{i+1})$ if we are in the previous situation after exchanging the role of x and y.

Finally, substitute (a_k) with

ix. (a_k, e) if $a_k \in A_x^{-1}$ and $f_0(o(h^{-1})) = y$,

x. (a_k, e^{-1}) if $a_k \in A_y^{-1}$ and $f_0(o(h^{-1})) = x$,

xi. (a_k, e, e^{-1}) if $a_k \in A_x^{-1}$ et $f_0(o(h)) = x$ and $(h,k)^{-1} \preceq_x u \preceq_x v$ for some extremal occurrences $u, v \in \mathcal{O}_x$ such that $(h,k)^{-1}\,\mathrm{R}_f\, v$ and such that the binding vertex s of u satisfies $f(s) = y$,

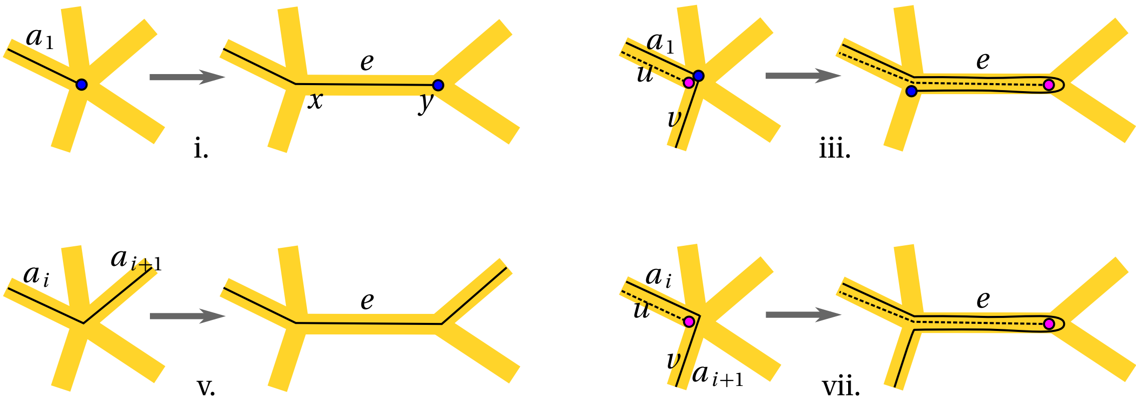


Figure 4.1: The inverse of an edge contraction, sometimes called a *vertex splitting*, may cause some modifications in a graph drawing.

xii. (a_k, e^{-1}, e) if we are in the previous situation after exchanging the role of x and y.

Figure 4.1 illustrates some of the cases. It is easily checked that the mapping g thus constructed defines a drawing of H in M' with the properties of the lemma. We now extend I to an embedding J of g. Since for each $h \in A(H)$, $g(h)$ and $f(h)$ only differs by the insertion of occurrences of e or e^{-1}, we may identify $\mathcal{O}_g(a)$ with $\mathcal{O}_f(a)$ for each arc a of M. Thanks to this identification, we let $\preceq_a^J$ equal $\preceq_a^I$. It remains to define $\preceq_e^J$ over $\mathcal{O}_g(e)$. Let $u, v \in \mathcal{O}_g(e)$. We declare $u \preceq_e^J v$ in either one of the following situations as illustrated on Figure 4.2.

i. There exist $u', v' \in A_x$ such that $u \,\mathrm{R}_g\, u'$ and $v \,\mathrm{R}_g\, v'$ and $v' \preceq_x u'$,

ii. there exist $u', v' \in A_y$ such that $u^{-1} \,\mathrm{R}_g\, u'$ and $v^{-1} \,\mathrm{R}_g\, v'$ and $u' \preceq_y v'$,

iii. there exist $u', u'' \in A_x$, $v', v'' \in A_y$ such that $u \,\mathrm{R}_g\, u'$, $v^{-1} \,\mathrm{R}_g\, v'$ and $u'' \,\mathrm{R}_f\, v''$ with $u'' \preceq_x u'$ and $v'' \preceq_y v'$.

It is easily checked from the fact that I is an embedding that we can not have both $u \preceq_e^J v$ and $v \preceq_e^J u$. The resulting partial order can be extended arbitrarily to a total order. We can finally claim that J is indeed an embedding by verifying that there is no crossing at x or y. This again results from the fact that I has no crossing at z. Moreover, if I is tight, the construction of J is clearly tight. $\square$

We now relate the nesting multiplicities of an embedding of a graph H after and before an edge contraction. A **zigzag** along an arc e in a drawing of H is a subpath (e, e^{-1}, e) or (e^{-1}, e, e^{-1}) of an image path of an arc of H.

Lemma 4.2.11. *Let I be an embedding of a drawing f of H in M that is tight and without zigzag along a non-loop arc e with endpoints $x = o(e)$ and $y = o(e^{-1})$. Let η_x and η_y be the number of extremal occurrences of e and e^{-1} respectively. We also set $k_y := |f^{-1}(y)|$ and suppose that the drawing of the edges of H are not reduced to e. The nesting multiplicities of I and I/e at every vertex of M other than x and y are the same in M and M/e. Moreover, if z is the identification of x and y in M/e,*

$$\mu_n(I, x) \leq \begin{cases} \mu_n(I/e, z) & \text{if } k_y = 0 \\ (2 + k_y)\mu_n(I/e, z) + \eta_x + \eta_y & \text{otherwise.} \end{cases}$$

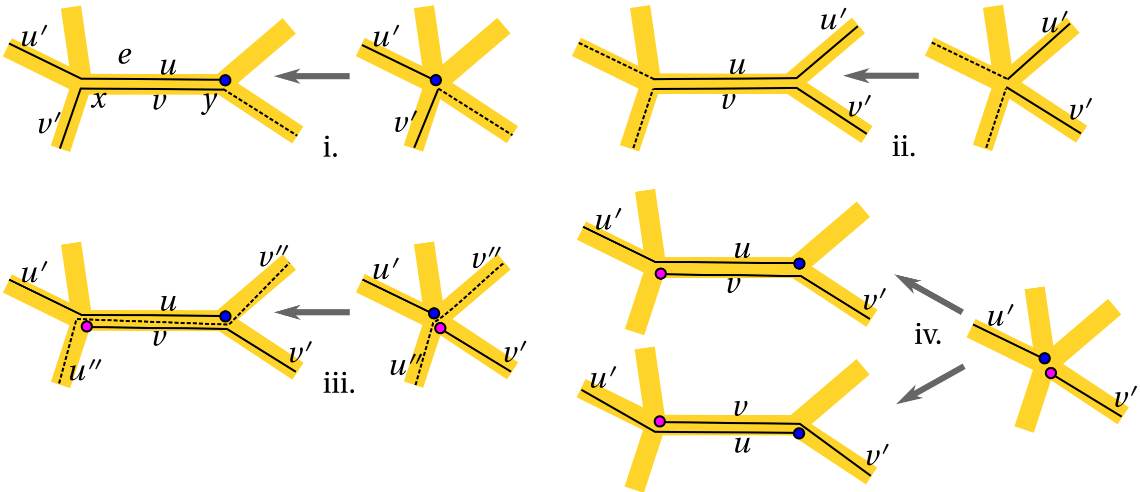


Figure 4.2: The ordering of the arc occurrences of e obtained by "stretching" the embedding I along e. (iv), A case where two occurrences u and v may be ordered in both directions without creating crossings.

Proof. We use the same notations $\mathcal{O}_x$ and $\mathcal{O}_y$ as in Lemma 4.2.10. We also use the upper-scripts int and ext to denote internal external arc occurrences. Let W^I_x and $W^{I/e}_z$ be the words of internal occurrences of I at x and of I/e at v respectively. We consider the following subsets of $\mathcal{O}^{int}_x$. See Figure 4.3.

- $\mathcal{O}_X = \{u \in \mathcal{O}^{int}_x \mid (\mathrm{R}_f(u))^{-1} \in \mathcal{O}^{int}_f(e^{-1}) \text{ and } (\mathrm{R}_f(\mathrm{R}_f(u)^{-1}))^{-1} \in \mathcal{O}^{int}_f(e)\}$
- $\mathcal{O}_{X'} = \{u \in \mathcal{O}^{int}_x \mid (\mathrm{R}_f(u))^{-1} \in \mathcal{O}^{int}_f(e^{-1}) \text{ and } (\mathrm{R}_f(\mathrm{R}_f(u)^{-1}))^{-1} \in \mathcal{O}^{ext}_f(e)\}$,
- $\mathcal{O}_Y = \{u \in \mathcal{O}^{int}_x \mid (\mathrm{R}_f(u))^{-1} \in \mathcal{O}^{ext}_f(e^{-1})\}$.

For $A \in \{X, X', Y\}$ we let A be the trace of W^I_x over $\mathcal{O}_A \cup \mathrm{R}_f(\mathcal{O}_A)$, where $\mathrm{R}_f(\mathcal{O}_A) := \{\mathrm{R}_f(u) \mid u \in \mathcal{O}_A\}$. We finally define Z as the remaining subword of W^I_x after deleting X, X' and Y. We analyse the nesting multiplicity of each subword separately. We trivially have $\mu_n(X') \le \eta_x$ and $\mu_n(Y) \le \eta_y$. Let $u \in \mathcal{O}_X$. Setting $v := \mathrm{R}_f((\mathrm{R}_f(u))^{-1})$ we have $v \in \mathcal{O}^{int}_f(e^{-1})$ and we must have $u' := \mathrm{R}_f(v^{-1}) \in \mathcal{O}^{int}_x$ because I has no zigzag. Note that $u' = \mathrm{R}_{f/e}(u)$, where f/e is the drawing of H in M/e associated to I/e. Since I is tight, there must be $w \in \mathcal{O}^{ext}_f(e^{-1})$ such that w^{-1} lies between v^{-1} and $\mathrm{R}_f(u)$. We define $b(u) \in V(H)$ as the binding vertex of w and $\mathcal{O}_X(b) := \{u \in \mathcal{O}_X \mid (b(u) = b\}$. For $b \neq b'$, the u in $\mathcal{O}_X(b) \cup \mathcal{O}_X(b')$ correspond to nested pairs $(u, \mathrm{R}_{f/e}(u))$ at z. This implies $(|\mathcal{O}_X(b)| + |\mathcal{O}_X(b')|)/2 \le \mu_n(I/e, z)$. It ensues that

$$\mu_n(X) = |\mathcal{O}_X| = \sum_{b \in f^{-1}(y)} |\mathcal{O}_X(b)| \le \lceil \frac{k_y}{2} \rceil 2\mu_n(I/e, z) \le (k_y + 1)\mu_n(I/e, z)$$

By substituting each $u \in Z \cap \mathcal{O}^{int}_f(e)$ with $\mathrm{R}_f(u^{-1})$ we obtain a subword of $W^{I/e}_z$, whence $\mu_n(Z) \le \mu_n(I/e, z)$. We finally obtain (see next Exercise 4.2.13)

$$\mu_n(I, x) \le \mu_n(X) + \mu_n(X') + \mu_n(Y) + \mu_n(Z) \le (2 + k_y)\mu_n(I/e, z) + \eta_x + \eta_y$$

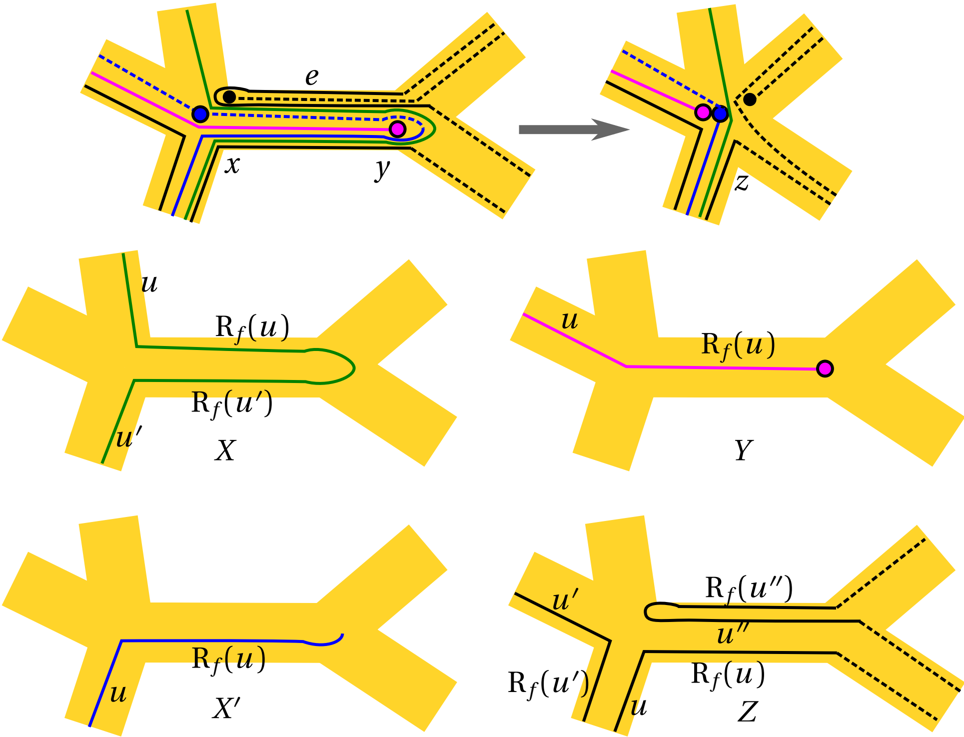


Figure 4.3: Decomposition of W_x^I into the subwords X, X', Y and Z.

When k_y is null we remark that X' and Y are empty, whence

$$\mu_n(I,x) \le \mu_n(I/e,z)$$

□

Corollary 4.2.12. *Let I be a tight embedding of H in M/e where e is an edge of M with distinct endpoints $x := o(e)$ and $y := o(e^{-1})$. Let J be the embedding of the drawing g of H in M as constructed in Lemma 4.2.10. We put $k_y = |g^{-1}(y)|$. The nesting multiplicities of J and I at every vertex of M other than x and y are the same in M and M/e. Moreover, if z is the identification of x and y in M/e and η_z is the number of extremal arc occurrences incident to z:*

$$\mu_n(J,x) \le \begin{cases} \mu_n(I,z) & \textit{if } k_y = 0 \\ (2+k_y)\mu_n(I,z) + \eta_z & \textit{otherwise.} \end{cases}$$

PROOF. Note that by construction, J is tight and has no zigzag along e. We can thus apply the previous lemma, noting that $\eta_x + \eta_y \le \eta_z$. □

Exercise 4.2.13. Let W be a well (unoriented) parenthesized word relatively to some pairing R (i.e., a fixed point free involution on W's letters). Let U be a subword of W which is stable by R and let V be the complementary subword. Show that

$$\mu_n(W) \le \mu_n(U) + \mu_n(V)$$

4.3 Canonical Systems of loops

The computation of a canonical systems of loops was studied by Vegter and Yap [VY90] for triangulated orientable surfaces. The paper seems to have a problematic claim in it (see Appendix A) and two other methods, together with implementations, were later proposed [LPVV01]. The first method relies on a traversal of the triangulation corresponding intuitively to the sublevel sets of a Morse function. When the boundary of the current sublevel splits, two new generators are added to the current system. This method does not seem to extend to non-orientable surfaces. The other method is based on a work of Brahana [Bra21] for the mathematical construction of canonical systems. This second method is only sketched in the original paper [LPVV01]. I present the details below and give extensions to general maps, not necessarily triangulated nor orientable.

Definition 4.3.1. A **fundamental system of loops** with basepoint x on a map M is an embedding of a bouquet of circles in M such that the homotopy classes of the circle embeddings of the bouquet form a basis for $\pi_1(M,x)$. Such a system is **canonical** if

- either M is a genus g orientable map and the loops $a_1, b_1, a_2, b_2, \dots, a_g, b_g$ of the bouquet satisfies the canonical relation

  $$[a_1,b_1][a_2,b_2]\dots[a_g,b_g] = 1$$

  in $\pi_1(M,x)$,

- or M is a genus g non-orientable map and the loops $a_1, a_2, \dots, a_g$ of the bouquet satisfies the canonical relation

  $$a_1a_1a_2a_2\dots a_ga_g = 1$$

Theorem 4.3.2. *Let M be an orientable map of genus g with n edges and let x be a vertex of M. We can compute a canonical system of loops with basepoint x in $O(gn)$ time where each loop has at most $8n-2$ arcs.*

We start with the case of a reduced map.

Lemma 4.3.3. *Assuming the hypotheses of the theorem and assuming that M is reduced, we can compute a canonical system of loops in $O(g^2)$ time such that each loop in the system has an embedding with nesting multiplicity at most two.*

Proof. From the proof of Theorem 3.1.15, we can put M into normal form by applying at most g times the sequence of 4 basic operations described by equations (3.5)-(3.8): a face subdivision, call it A_k, followed by an edge deletion B_k, another face subdivision C_k and another edge deletion D_k. We let $M_0 = M$ and for $k \in [0, g-1]$: $M_{k+1} = D_kC_kB_kA_k(M_k)$. The maps M_k are thus reduced and M_g is in normal form. Let b_k and d_k be the edges deleted by B_k and D_k respectively, and let a_k and c_k be the edges introduced by the face subdivisions A_k and C_k respectively (they were denoted v, u, d, ℓ in (3.5)-(3.8)). An essential property of the sequence of operations proposed in

Theorem 3.1.15 is that a_k and c_k are not deleted by later operations and thus persist in the normal form M_g. Remark that this property also implies that the edges of M_0 are precisely the b_k's and d_k's. For one can only delete something that was previously there! In M_g we choose the canonical system composed of the unique embedding I_g of its $2g$ loop-edges $a_1, c_1, \dots, a_g, c_g$ with their default orientation.

We next apply the reverse sequence of operations to this embedding in order to get an embedding in M_0. Assume that we have computed an embedding I_{k+1} in M_{k+1}. The inverse of D_k is a face subdivision, introducing d_k. We can keep the embedding I_{k+1} of the same system of loops. The inverse of C_k is the deletion of c_k. We modify I_{k+1} as described at the beginning of Section 4.2.1; this replaces the loop c_k by a complementary path γ_k in one of the two incident faces. Likewise, the inverse of B_k only adds the edge b_k while the inverse of A_k implies the replacement of a_k by a complementary path α_k in one of its incident faces. This defines the embedding I_k. Looking at (3.7) and (3.5) we can choose γ_k and α_k to be composed of b_i's and d_i's only (γ_k and α_k correspond respectively to $\bar{\ell}ZY\bar{u}$ in (3.7) and $\bar{d}uXvY\bar{u}$ in (3.5)). By the above remark, this second essential property implies that γ_k and α_k will not be modified by the next inverse operations and will thus be part of our final system of loops in M_0. From Lemma 3.2.10, this system satisfies the same canonical relation as $(a_1, c_1, \dots, a_g, c_g)$. Moreover, because each γ_k and α_k appears as a subpath of a facial circuit, it follows from Exercise 4.2.8 that $\mu_n(\gamma_k, x) \leq 2$ and $\mu_n(\alpha_k, x) \leq 2$, where x is the unique vertex of M. □

Exercise 4.3.4. Propose a $O(gn)$ time algorithm to compute a fundamental system of loops on a genus g map with fixed basepoint , orientable or not, such that each loop has at most $2n$ arcs. Show how to reduce the size of each loop to at most n if the basepoint is not fixed in advance.

PROOF OF THEOREM 4.3.2. Following Lemma 3.1.10, we apply to M a sequence of edge contractions followed by a sequence of edge deletions to obtain a reduced map M'. By Lemma 4.3.3, we can construct in $O(g^2)$ time a canonical system of loops on M' whose loops have nesting multiplicity at most two. It is easily verified that the system constructed in Lemma 4.3.3 is tight without zigzag. We now transform M' back to M, first applying the face subdivisions inverse to the above edge deletions. Since this only adds edges to M', we can keep the same embedding in the resulting map. We next apply the vertex splittings inverse to the above edge contractions. At each vertex splitting, we transform the current system embedding as in Lemma 4.2.10. Note that the restriction of the transformed embedding to a single loop in the system coincides with the transformation applied to this single loop. Using corollary 4.2.12 inductively, we conclude that each loop in the resulting canonical system on M has an embedding I satisfying

$$\mu_n(I, x) \leq 2 \quad \text{and} \quad \forall y \neq x, \quad \mu_n(I, y) \leq 8$$

By Proposition 4.2.5, we get an arc multiplicity at most 4 for the arcs with origin x, and at most 8 for the other arcs. It follows from Lemma 4.2.2 that each loop has size at most $8n+2$. We can reduce this size to $8n-2$ taking into account the multiplicity at most 4 of the arcs with origin x. □

Open question: A simple example in [LPVV01] shows that the complexity of the canonical system in Theorem 4.3.2 is asymptotically tight. However, this does not imply that the $O(g^2)$ system constructed for a reduced map in Lemma 3.1.10 is optimal. Rewording in the language of combinatorial groups, this suggests the following question. Given a basis $(b_1, b_2, \dots, b_{2g})$ of a surface group, is there a canonical basis whose members, as words over the b_i and b_i^{-1}, have total length $o(g^2)$?

Exercise 4.3.5. We may slightly refine Corollary 4.2.12 for the drawing of a bouquet of m circles. We use the notations as in the corollary. Assuming that z is the image in M/e of the bouquet basepoint, show that we can choose for basepoint u of M either x or y so as to obtain

$$\mu_n(J,u) \leq \mu_n(I,z) \quad \text{and} \quad \mu_n(J,v) \leq \mu_n(I,z) + 2m$$

where $\{u,v\} = \{x,y\}$. Deduce a construction of a canonical system on M whose loops have at most $4n$ arcs each.

Exercise 4.3.6. With the notations of Exercise 4.2.9, prove that we can construct a piecewise linear canonical system of loops on $\mathscr{M}$ whose loops have $O(n)$ segments each.

We now turn to non-orientable maps.

Lemma 4.3.7. *Let M be a non-orientable reduced map of genus g. We can compute a canonical system of loops in $O(g^3)$ time such that each loop in the system has an embedding with nesting multiplicity at most* $6g+3$.

Proof. The case $g = 1$ is trivial and we now assume $g \geq 2$. Following the proof of the classification Theorem 3.1.16, we may put M into normal form by a sequence of transformations, say $T_0, T_1, \dots, T_{n-1}$, where each T_i is of type I, II.i or II.ii. We set $M_0 = M$ and $M_{k+1} = T_k(M_k)$ for $k \in [0, n-1]$. Hence, each M_k is reduced and M_n is in normal form. Referring to Equation (3.9), the facial circuit of the unique face of M_k is $X_k P_k$. Looking at the specific form of X_k and X_{k+1} for each type of transformation, we remark that the edges of the flags in X_k are edges of M_0. Each T_k adds some edges to M_k and delete some other edges to form M_{k+1}. In particular, if T_k has type I or II.i, it adds one edge u_k corresponding to the flag u in the proof of Theorem 3.1.16. if T_k has type II.ii, it adds three edges u_k, v_k, w_k corresponding to the flags u, v, w. The edges of M_n are thus among the added edges u_k, v_k or w_k. However, as opposed to the orientable case, some of these edges may be further deleted by transformations of type II.ii. In M_n we choose the canonical system composed of the unique embedding of its g loop-edges. We now apply the reverse sequence of inverse transformations, T_k^{-1}, applying the corresponding modification to the current canonical system. According to Section 4.2.1, the embedding is left unchanged when an edge is added and each occurrence of a deleted edge is replaced by a complementary subpath in an incident face. If p is a path in some M_{k+1} we denote by $\mu(p)$ the nesting multiplicity at the unique vertex of M_0 of the embedding of the path $T_0^{-1} \circ T_1^{-1} \circ \cdots \circ T_k^{-1}(p)$.

- If T_k is of type I, the reverse of T_k deletes u_k. Referring to the proof of Theorem 3.1.16, u_k bounds a face with facial circuit $\bar{u} x U$ and we choose to replace each occurrence of u_k in the current canonical system by the complementary path corresponding to xU. Since xU is a subword of X_k, the above remark implies that this path is also a path in M_0. It follows from Exercise 4.2.8 that $\mu(u_k) \leq 2$.

- If T_k is of type II.i, the reverse of T_k introduces an edge d_k with flag d, deletes u_k and finally deletes d_k (see the proof of Theorem 3.1.16). Just before its deletion u_k bounds a face with facial circuit $\bar{u}\, d\, CA^{-1}\alpha_0(x)$ and similarly d_k bounds the facial circuit $\bar{d}xAyB\bar{x}$. We choose to perform the associated subpath substitutions, so that u_k is replaced by the path with flags $xAyB\bar{x}CA^{-1}\alpha_0(x)$. This path is contained in M_0 and we can write

$$\mu(u_k) = \mu(xAyB\bar{x}CA^{-1}\alpha_0(x)) \leq \mu(xAyB\bar{x}) + \mu(CA^{-1}\alpha_0(x)) + 1 \leq 5$$

For the first inequality we used that the nesting multiplicity of the concatenation of two paths is bounded by one plus the sum of their multiplicities (see Exercise 4.3.8). For the second inequality, we used that the two paths are facial subpaths.

- If T_k is of type II.ii, the reverse of T_k deletes u_k, v_k and w_k and uses some auxiliary edges. We shall refer to the transformations in case II.ii of Theorem 3.1.16 and adjoin a subscript k to denote the arc components of a flag sequence. By arguments similar as above, we choose to replace w_k by $d_k B_k^{-1} C_k^{-1}$, v_k by $d_k t_k C_k B_k d_k^{-1}$, u_k by $z_k d_k A_k y_k B_k d_k^{-1}$, d_k by $z_k x_k$, and t_k by $B_k^{-1} y_k^{-1} A_k^{-1} d_k^{-1}$. Here, x_k, y_k, A_k, B_k, C_k are composed of edges of M_0 but z_k may not appear in M_0. From Exercises 4.3.8 and 4.2.8 and the fact that the nesting multiplicity of a one-edge path in M_0 is null we derive that

$$\begin{array}{rcl}
\mu(d_k) & = & \mu(z_k x_k) \leq \mu(z_k) + 1 \\
\mu(d_k t_k) & = & \mu(z_k x_k B_k^{-1} y_k^{-1} A_k^{-1} x_k^{-1} z_k^{-1}) \leq 2\mu(z_k) + \mu(x_k B_k^{-1} y_k^{-1} A_k^{-1} x_k^{-1}) + 2 \\
 & \leq & 2\mu(z_k) + 4 \\
\mu(u_k) & = & \mu(z_k d_k A_k y_k B_k d_k^{-1}) = \mu(z_k z_k x_k A_k y_k B_k x_k^{-1} z_k^{-1}) \\
 & \leq & 3\mu(z_k) + \mu(x_k A_k y_k B_k x_k^{-1}) + 3 \leq 3\mu(z_k) + 5 \\
\mu(v_k) & = & \mu(d_k t_k C_k B_k d_k^{-1}) \leq \mu(d_k t_k) + \mu(C_k B_k) + \mu(d_k) + 2 \leq 3\mu(z_k) + 9 \\
\mu(w_k) & = & \mu(d_k B_k^{-1} C_k^{-1}) \leq \mu(d_k) + \mu(B_k^{-1} C_k^{-1}) + 1 \leq \mu(z_k) + 4
\end{array}$$

Note that z_k is either equal to u_{k-1} or to w_{k-1} depending on the type of T_{k-1}. Hence, $\mu(z_k) \leq 2$ or $\mu(z_k) \leq 5$ if T_{k-1} is of type I or II.i, respectively. Otherwise, we get the recursion formula $\mu(w_k) \leq \mu(w_{k-1}) + 4$. Whence $\mu(w_k) \leq 4k + 2$, since T_0 is of type I. Because each type II.ii transformation increases the number of canonical loops by two, the depth of the recursion is at most $(g-1)/2$ and

$$\mu(w_k) \leq 4\frac{g-1}{2} + 2 = 2g$$

We also deduce that

$$\mu(u_k) \leq 3\mu(w_{k-1}) + 9 \leq 6g + 3$$

and

$$\mu(v_k) \leq 3\mu(w_{k-1}) + 5 \leq 6g - 1$$

In any case, each loop in the embedding of the canonical system in M_0 has nesting multiplicity at most $6g + 3$. □

Exercise 4.3.8. Let I be an embedding in M of a path which is the concatenation of two subpaths. Let I' and I'' be the restrictions of I to these subpaths. Show that

$$\mu_n(I,x) \le \mu_n(I,x) + \mu_n(I,x) + 1$$

for every vertex x on M.

Theorem 4.3.9. *Let M be a non-orientable map of genus g with n edges and let x be a vertex of M. We can compute a canonical system of loops with basepoint x in $O(g^2n)$ time where each loop has at most $(18g+11)n$ arcs.*

Proof. As for the orientable case we start with a sequence of edge contractions and deletions to obtain a reduced map M'. We construct a canonical system in M' as given by the preceding lemma. Each loop in this system has nesting multiplicity at most $6g+3$. We next transform M' back to M by face subdivisions and vertex splittings inverse to the above sequence of operations. Only the vertex splitting imply changes in the canonical system. Using corollary 4.2.12 inductively, we conclude that each loop in the resulting canonical system on M has an embedding I satisfying

$$\mu_n(I,x) \le 6g+3 \quad \text{and} \quad \forall y \ne x, \quad \mu_n(I,y) \le 18g+11$$

By Proposition 4.2.5, we get an arc multiplicity at most $12g+6$ for the arcs with origin x, and at most $18g+11$ for the other arcs. It follows from Lemma 4.2.2 that each loop has size at most $(18g+11)n+2$. We can reduce this size to $(18g+11)n-6g-5$ taking into account the multiplicity at most $12g+6$ of the arcs with origin x. $\square$

Chapter 5

The Homotopy Test

Contents

This last chapter of the first part is essentially the arxiv version [LR11] of a conference paper on the homotopy test [LR12]. It was written before the rest of this document and most of the arguments rely upon classical topology as opposed to purely combinatorial arguments. In particular, it uses the language of embedded graphs rather than combinatorial maps. Some definitions and concepts are thus redundant with the first chapters but expressed in the classical framework of surface topology.

5.1 Main Result

In their 1999 paper [DG99], Dey and Guha announced a linear time algorithm for testing whether two curves on a triangulated surface are freely homotopic. This appeared as a major breakthrough for one of the most basic problem in computational topology. Dey and Guha's approach relies on results by Greendlinger [Gre60] for the conjugacy problem in one relator groups satisfying some small cancellation condition. In Appendix C, we show several subtle flaws in the paper of Dey and Guha [DG99] that invalidate their approach and leave little hope for repair. Inspired by the recent work of Colin de Verdière and Erickson [CE10] for computing a shortest cycle in a free homotopy class, we propose a different geometric approach and confirm the results of Dey and Guha. In addition, our free homotopy test covers the cases of orientable surfaces of genus 2 or non orientable surfaces of genus 3 and 4 which are not addressed in Dey and Guha's approach.

In a first part we consider the homotopy test for curves with fixed endpoints drawn in a graph cellularly embedded in a surface $\mathscr{S}$. This test reduces to decide if a loop is contractible in $\mathscr{S}$, *i.e.*, null-homotopic, since a curve c is homotopic to a curve d with fixed endpoints if and only if the concatenation $c \cdot d^{-1}$ is contractible. The contractibility test was already considered by Dey and Schipper [DS95] using a partial and implicit construction of the universal cover of $\mathscr{S}$. Indeed, a curve is null-homotopic in $\mathscr{S}$ if and only if its lift is closed in the universal cover of $\mathscr{S}$. Given a closed curve c, Dey and Schipper detect if c is null-homotopic in $O(|c|\log g)$ time, where g is the genus of $\mathscr{S}$. Their implicit construction is relatively complex and does not seem to extend to handle the free homotopy test. Our solution to the contractibility test also relies on a partial construction of the universal cover. We use the more explicit construction of Colin de Verdière and Erickson [CE10, Sec. 3.3 and 4] for tightening paths. It amounts to build a convex region of the universal cover (with respect to some hyperbolic metric) large enough to contain a lift of c. An argument *à la Dehn* shows that this region can be chosen to have size $O(|c|)$, leading to our first theorem:

Theorem 5.1.1 (Contractibility test)**.** *Let G be a graph of complexity n cellularly embedded in a surface $\mathscr{S}$, not necessarily orientable. We can preprocess G in $O(n)$ time, so that for any loop c on $\mathscr{S}$ represented as a closed walk of k edges in G, we can decide whether c is contractible or not in $O(k)$ time.*

We next study the free homotopy test, that is deciding if two cycles c and d drawn in a graph G cellularly embedded in $\mathscr{S}$ can be continuously deformed one to the other. By theorem 5.1.1, we may assume that none of c and d is contractible. Our strategy is the following. We first build (part of) the cyclic covering of $\mathscr{S}$ induced by the cyclic subgroup generated by c in the fundamental group of $\mathscr{S}$. We denote by $\mathscr{S}_c$ this covering. Assuming that $\mathscr{S}$ is orientable (the non-orientable case is discussed in Section 5.4.4), $\mathscr{S}_c$ is a topological cylinder[1], and we call any of its non-contractible simple cycles a *generator*. Since the generators of $\mathscr{S}_c$ are freely homotopic, their projection on $\mathscr{S}$ are freely homotopic to c. Our next task is to extract from $\mathscr{S}_c$ a *canonical* generator γ_R whose definition only depends on the isomorphism class of $\mathscr{S}_c$. To this end, we lift in $\mathscr{S}_c$ the graph G of $\mathscr{S}$ and we endow $\mathscr{S}_c$ with the corresponding cross-metric introduced by Colin de Verdière and Erickson [CE10]. The set of generators that are minimal for this metric form a compact annulus in $\mathscr{S}_c$. We eventually define γ_R as the "right" boundary of this annulus. We perform the same operations starting with d instead of c to extract a canonical generator δ_R of $\mathscr{S}_d$. From standard results on covering spaces [Mas91, §V.6], we know that $\mathscr{S}_c$ and $\mathscr{S}_d$ are isomorphic covering spaces if c and d are freely homotopic. It follows that c and d are freely homotopic if and only if γ_R and δ_R have equal projections on $\mathscr{S}$. Proving that γ_R and δ_R can be constructed in time proportional to $|c|$ and $|d|$ respectively, we finally obtain:

Theorem 5.1.2 (Free homotopy test)**.** *Let G be a graph of complexity n cellularly embedded in a surface $\mathscr{S}$. We can preprocess G in $O(n)$ time, so that for any cycles c and d on $\mathscr{S}$ represented as closed walks with a total number of k edges in G, we can decide if c and d are freely homotopic in $O(k)$ time.*

[1] If $\mathscr{S}$ is non orientable, its cyclic coverings can be either cylinders or Möbius rings.

The *word problem* in a group presented by generators and relators is to decide if a product of generators and their inverses, called a *word*, is the unit in the group. The *conjugacy problem* is to decide if two words represent conjugate elements in the group. The length of a word is its number of factors. As an immediate consequence of our two theorems, we can solve the word problem and the conjugacy problem in surface groups in optimal linear time. More precisely, suppose we are given a presentation by generators and by a single relator of the fundamental group of a compact surface $\mathscr{S}$ of genus g without boundary. After $O(g)$ time preprocessing, we can solve the word problem in time proportional to the length of the word. Moreover, we can report if two words are conjugate in time proportional to their total length. The preprocessing reduces to build a cellular embedding of a wedge of loops in $\mathscr{S}$, with one loop for each generator; the rotation system of this embedding (see the Background Section for a definition) is easily deduced from the relator of the group. Any word can then be interpreted as a walk in this cellular embedding so that we can directly apply the previous theorems. The word and conjugacy problems have a long standing history starting with Dehn's seminal papers [Sti87]. Recent developments include linear time solutions to the word problem in much larger classes of groups comprising hyperbolic groups [DA85, Hol00, HR01]. We emphasize that such developments assume a multi-tape Turing machine as a model of computation and, most importantly, that the size of the group presentation is considered as a constant. In our case, the group itself is part of the input and, after the preprocessing phase, the decision problems have linear time solutions *independent of the genus of the surface*.

A short time after our paper appeared, Jeff Erickson and Kim Whittlesey, aware of our work, have proposed a simpler approach to the homotopy test based on van Kampen diagrams and small cancellation theory [EW13]. Their algorithm starts with a simplified graph, the *radial graph*, that we had previously introduced (see Section 5.3.1) in our method. However, Erickson and Whittlesey have succeeded to avoid the use of a cyclic covering to work directly at the level of the surface. They apply combinatorial homotopies, pushing the cycle always in the same direction, until a canonical position is reached. A compact run-length encoding allows them to perform the homotopies in linear time. Their approach is justified by combinatorial versions of the Gauss-Bonnet formula and of the four vertex theorem [GS90]. It appears that their canonical representative is the same as the projection of the above canonical generator γ_R. I still think our method has some interest, even if it may appear more complicated. Our construction directly provides the canonical cycle, with a geometric interpretation, without performing the homotopy. This is justified by the combinatorial structure of the cyclic covering. Moreover, we do not need any specific compact encoding, keeping an explicit encoding of the cycles, and still get the same optimal complexity.

5.2 Topological Background

We review some basic definitions and properties of surfaces and their covering spaces, as well as combinatorial embeddings of graphs. We refer the reader to Massey [Mas91] or Stillwell [Sti93] for further details on covering spaces.

Surfaces. We only consider surfaces without boundary. A **surface** (or 2-manifold) $\mathscr{S}$ is a connected, Hausdorff topological space where each point has a neighborhood homeomorphic to the plane. A compact surface is homeomorphic to a sphere where either:

- $g \geq 0$ open disks are removed and a handle (*i.e.*, a perforated torus with one boundary component) is attached to each resulting circle, or
- $g \geq 1$ open disks are removed and a Möbius band is attached to each resulting circle.

The surface is called **orientable** in the former case and **non-orientable** in the latter case. In both cases, g is the **genus** of the surface.

A **path** in a surface $\mathscr{S}$, is a continuous map $p : [0,1] \to \mathscr{S}$. A **loop** is a path p whose endpoints $p(0)$ and $p(1)$ coincide. This common endpoint is called the **basepoint** of the loop.

Homotopy and fundamental group. Two paths p, q in $\mathscr{S}$ are **homotopic** (with fixed endpoints) if there is a continuous map $h : [0,1] \times [0,1] \to \mathscr{S}$ such that $h(0,t) = p(t)$ and $h(1,t) = q(t)$ for all t, and $h(\cdot,0)$ and $h(\cdot,1)$ are constant maps. Being homotopic is an equivalence relation. The set of homotopy classes of loops with given basepoint $x \in \mathscr{S}$ forms a group where the operation in the group corresponds to the concatenation of the loops. This group is called the **fundamental group** of $\mathscr{S}$ and denoted by $\pi_1(\mathscr{S}, x)$. The homotopy class of a loop c is denoted by $[c]$. The loop c is said **contractible**, or null-homotopic, if c is homotopic to the constant loop, *i.e.*, if $[c]$ is the identity of $\pi_1(\mathscr{S}, x)$. The fundamental group of the orientable surface $\mathscr{S}$ of genus $g \geq 1$ admits finite minimal presentations composed of $2g$ generators and one relator expressed as the product of $4g$ generators and their inverses. A group defined by a set A of generators and a set R of relators is denoted $\langle A\, ; R\rangle$. In particular, $\pi_1(\mathscr{S}, x)$ is isomorphic to the **canonical** presentation $\langle a_1, b_1, \dots, a_g, b_g\, ; a_1 b_1 a_1^{-1} b_1^{-1} \cdots a_g b_g a_g^{-1} b_g^{-1}\rangle$ but not all minimal presentations of $\pi_1(\mathscr{S}, x)$ are canonical. Likewise, the fundamental group of the non-orientable surface of genus g has a presentation with g generators and one relator of length $2g$.

Two loops c, d in $\mathscr{S}$ are **freely homotopic** if there is a continuous map $h : [0,1] \times [0,1] \to \mathscr{S}$ such that $h(0,\cdot) = c$, $h(1,\cdot) = d$ and $h(s,0) = h(s,1)$ for all s. The loops c and d of respective basepoints x and y are freely homotopic if and only if $[c]$ and $[u \cdot d \cdot u^{-1}]$ are conjugate in $\pi_1(\mathscr{S}, x)$ for any path u linking x to y.

Covering spaces. A **covering space** of the surface $\mathscr{S}$ is a surface $\mathscr{S}'$ together with a continuous surjective map $\pi : \mathscr{S}' \to \mathscr{S}$ such that every $x \in \mathscr{S}$ lies in an open neighborhood U such that $\pi^{-1}(U)$ is a disjoint union of open sets in $\mathscr{S}'$, each of which is mapped homeomorphically onto U by π. The map π is called a **covering map**; it induces a monomorphism $\pi_* : \pi_1(\mathscr{S}', y) \to \pi_1(\mathscr{S}, \pi(y))$, so that $\pi_1(S', y)$ can be considered as a subgroup of $\pi_1(S, \pi(y))$. If p is a path in $\mathscr{S}$ and $y \in \mathscr{S}'$ with $\pi(y) = p(0)$, then there exists a unique path $q : [0,1] \to \mathscr{S}'$, called a **lift** of p, such that $\pi \circ q = p$ and $q(0) = y$.

A **morphism** between the covering spaces $(\mathscr{S}', \pi)$ and $(\mathscr{S}'', \pi')$ of $\mathscr{S}$ is a continuous map $\varphi : \mathscr{S}' \to \mathscr{S}''$ such that $\pi' \circ \varphi = \pi$. Up to isomorphism, each surface $\mathscr{S}$ has a unique simply connected covering space, called its **universal cover** and denoted $\tilde{\mathscr{S}}$. Unless $\mathscr{S}$ is a sphere or a projective plane $\tilde{\mathscr{S}}$ has the topology of a plane. More generally, for every subgroup $\mathscr{G}$ of $\pi_1(\mathscr{S}, x)$ there is a covering space of $\mathscr{S}$, unique up to isomorphism, whose fundamental group is conjugate to $\mathscr{G}$ in $\pi_1(\mathscr{S}, x)$. If $\mathscr{G}$ is cyclic and generated by the homotopy class of a non-contractible loop c in $\mathscr{S}$, this covering space is called the c-**cyclic cover** and denoted by $\mathscr{S}_c$. The c-cyclic cover can be constructed as follows: take a lift $\tilde{c}$ of c in the universal cover $\tilde{\mathscr{S}}$ and let τ be the unique automorphism of $\tilde{\mathscr{S}}$ sending $\tilde{c}(0)$ to $\tilde{c}(1)$; then $\mathscr{S}_c$ is the quotient of the action of $\langle\tau\rangle$ on $\tilde{\mathscr{S}}$. When $\mathscr{S}$ is orientable, $\mathscr{S}_c$ has the topology of a cylinder whose generators project on $\mathscr{S}$ to loops that are freely homotopic to c or its inverse. Conversely, every loop freely homotopic to c has a closed lift generating the fundamental group of $\mathscr{S}_c$. In Section 5.4 we shall use the following properties of curves on cylinders. We assume that the considered curves are in general position: all (self-)intersections are transverse and have multiplicity two, *i.e.*, exactly two curve pieces cross at an intersection.

Lemma 5.2.1. *Let c be a loop obtained as the concatenation of k simple paths on a cylinder. Then c is freely homotopic to the i-th power of a generator of the cylinder with $|i| < k$.*

Proof. We view the cylinder as a punctured plane; so that the i-th power of a generator has winding number i with respect to the puncture. The winding number of c is the sum of the angular extend of each of its subpaths divided by 2π. But the angular extend of a simple path has absolute value strictly smaller than 2π. □

Lemma 5.2.2. *A self-intersecting generator of a cylinder has a contractible closed subpath.*

Proof. Consider a self-intersecting generator γ and define a bigon of γ as a disk bounded by two subpaths of γ. Applying local homotopies to γ we can remove all of its bigons one by one. If γ is still self-intersecting, it must have a contractible subpath by [HS85, Lemma 1.4]. If γ is simple, we consider the loop just before we remove the last bigon. It is easily seen that this loop has a contractible subpath. In both cases we have found a contractible subpath that corresponds to a contractible subpath of the initial loop if we undo the local homotopies.

For completeness, we also give a self-contained proof. We again view the cylinder as a punctured plane. Consider the domain of γ as the unit circle $\mathbb{R}/\mathbb{Z}$. A *loop-segment* is a closed connected part $[x, y] \subset \mathbb{R}/\mathbb{Z}$ whose image by γ is a simple loop, *i.e.*, such that $\gamma(x) = \gamma(y)$ and the restriction of γ to $[x, y)$ is one-to-one. Let $[x_1, y_1], \ldots, [x_k, y_k]$ be a maximal set of pairwise interior disjoint loop-segments. By the general position assumption, k is finite. We denote by γ_i the restriction of γ to $[x_i, y_i]$. It is easily seen that $\mathbb{R}/\mathbb{Z} \setminus \cup_{i=1}^{k}[x_i, y_i]$ has at most k connected components (exactly k if we take into account that each self-intersection has multiplicity two) whose images are simple paths composing a loop c. If some loop γ_i is contractible, then we are done. Otherwise, being simple, each γ_i has winding number $\varepsilon_i = \pm 1$. Let $w(c)$ denote the winding number of c. By additivity of the winding number, we have

$$\sum_{i=1}^{k} \varepsilon_i + w(c) = 1.$$

By the preceding lemma we have $|w(c)| < k$, implying that $\varepsilon_j = 1$ for some $j \in [1, k]$. It ensues that the complementary part of γ_j in γ is a loop with winding number $1 - \varepsilon_j = 0$, hence contractible. □

Cellular embeddings of graphs. All the considered graphs may have loop edges and multiple edges. We denote by $V(G)$ and $E(G)$ the set of vertices and edges of a graph G, respectively. A graph G is **cellularly embedded** on a surface $\mathscr{S}$ if every open face of (the embedding of) G on $\mathscr{S}$ is a disk. The combinatorial description of the embedding of G is a combinatorial map in the sense of Definition 2.2.1. For a finite graph G, storing a cellular embedding takes a space linear in the **complexity** of G, that is, in its total number of vertices and edges. Combinatorial maps can be implemented efficiently [Epp03, Lin82] so that we can traverse the neighbors of a vertex in time proportional to its degree or obtain a facial walk in time proportional to its length.

Any graph G cellularly embedded in $\mathscr{S}$ has a **dual graph** denoted G^* whose vertices and edges are in one-to-one correspondence with the faces and edges of G respectively; if two faces of G share an edge $e \in E(G)$ its **dual edge** $e^* \in E(G^*)$ links the corresponding vertices of G^*. This dual graph can be cellularly embedded on $\mathscr{S}$ so that each face of G contains the matching vertex of G^* and each edge e^* dual of $e \in E(G)$ crosses only e, only once.

Let $(\mathscr{S}', \pi)$ be a covering space of $\mathscr{S}$. The **lifted graph** $G' = \pi^{-1}(G)$ is cellularly embedded in $\mathscr{S}'$. The restriction of π from the star of each vertex $x \in V(G')$ to the star of $\pi(x) \in V(G)$ is an isomorphism. Note that $\mathscr{S}'$ can be non compact even though $\mathscr{S}$ is compact. In that case, G' has an infinite number of edges and vertices.

Regular paths and crossing weights. For the free homotopy test in Section 5.4, we make use of the cross metric surface model [CE10]. Let G be a graph cellularly embedded in $\mathscr{S}$. A path p in $\mathscr{S}$ is **regular** for G if every intersection point of p and G is an endpoint of p or a transverse crossing, *i.e.*, has a neighbourhood in which $p \cup G$ is homeomorphic to two perpendicular line segments intersecting at their midpoint. The **crossing weight** with respect to G of a regular path p is the number $\left|p\right|$ of its transverse crossings and is always finite. In particular, if p is a path of the dual graph G^* then p is regular for G and $\left|p\right|$ is the number of edges in p.

Homotopy encoding in cellular graph embeddings. Consider a graph G cellularly embedded in $\mathscr{S}$. If H is a subgraph of G, we will denote by $\mathscr{S} \backslash\backslash H$ the surface obtained after cutting $\mathscr{S}$ along H. If $\mathscr{S} \backslash\backslash H$ is a topological disk, then H is called a **cut graph**. A cut graph can be computed in linear time [CdVL10, Epp03]. Note that a cut graph defines a cellular embedding in $\mathscr{S}$ with a unique face. Let T be a spanning tree of a cut graph H and consider the set of edges $A := E(H) \setminus E(T)$. See Figure 5.1. If $\mathscr{S}$ is a compact surface of genus g, Euler's formula easily implies that A contains either $2g$ or g edges depending on whether $\mathscr{S}$ is respectively orientable or non-orientable. For each vertex $s \in H$, we have $\pi_1(\mathscr{S}, s) \cong \langle A \,;\, f_H|_A \rangle$, where f_H is the facial walk of the unique face of H and $f_H|_A$ denotes its restriction to the edges in A. Indeed, if we contract T to the vertex s in $\mathscr{S}$, the graph H becomes a bouquet of circles whose complementary set in $\mathscr{S}$ is a disk bounded by the facial walk $f_H|_A$. The above group presentation then follows from the classical Seifert and Van Kampen Theorem [Mas91, Chap. IV].

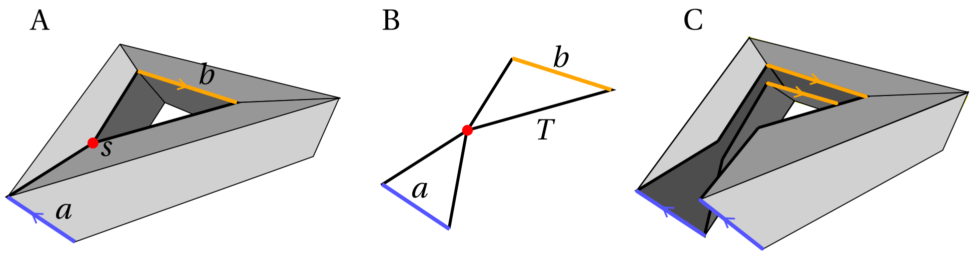


Figure 5.1: (A) A quadrangulation of a torus ($g = 1$). (B) A cut graph of the torus with a spanning tree T. The set $A = \{a, b\}$ of non-tree edges contains $2g = 2$ elements. (C) The torus cut through the cut graph can be flattened into the plane. Here $f_H|_A = aba^{-1}b^{-1}$.

Let c be a closed walk in G with basepoint $x \in V(H)$. Denote by $T(s,x)$ the unique simple path in T from s to x. We can express the homotopy class $[c']$ of the closed walk $c' := T(s,x) \cdot c \cdot T(x,s)$ as follows. Let $x = u_0, u_1, \ldots, u_k = x$ be the sequence of vertices that belong to H, while walking along c'. The subpath of c' between u_{i-1} and u_i is homotopic to a subpath w_i of the facial walk f_H between occurrences of u_{i-1} and u_i. Denote by $w_i|_A$ the restriction of w_i to the edges in A. We have, in the above presentation of $\pi_1(\mathscr{S}, s)$:

$$[c'] = (w_1|_A) \cdot (w_2|_A) \cdots (w_k|_A)$$

We call such a product a **term product** representation of $[c']$ of **height** k. If we encode each term $w_i|_A$ implicitly by two pointers, in $f_H|_A$, corresponding to its first and last edge, the above representation can be stored as a list of k pairs of pointers. We repeat Lemma 3.6.2 in the language of embedded graphs.

Lemma 5.2.3 (see [DG99, Sec. 3.1]). *Let G be a graph of complexity n cellularly embedded on a surface $\mathscr{S}$. We can preprocess G and its embedding in $O(n)$ time such that the following holds. For any closed walk c in G with k edges, we can compute in $O(k)$ time a term product representation of height at most k of some closed walk freely homotopic to c.*

5.3 The Contractibility Test

After reduction to a simplified framework, our test for the contractibility of a closed curve p drawn on $\mathscr{S}$ relies on the construction of a *relevant region* Π_p in a specific tiling of the universal cover $\tilde{\mathscr{S}}$ of $\mathscr{S}$. This relevant region, introduced by Colin de Verdière and Erickson [CE10], contains a lift of p and we shall compute that lift as we build Π_p. We can now decide whether p is contractible by just checking if its lift is a closed curve in $\tilde{\mathscr{S}}$. We detail our simplified framework and tiling of $\tilde{\mathscr{S}}$, as well as some of its combinatorial properties, in Section 5.3.1. We define the relevant region and its construction in Section 5.3.2.

5.3.1 A simplified framework

Graphic interpretation of term products. Following Lemma 5.2.3, we can assume that G is in reduced form, with a single vertex s and a single face, and that the input

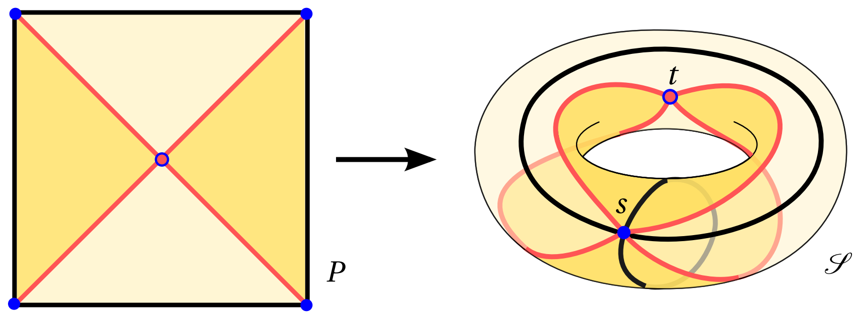


Figure 5.2: When $\mathscr{S}$ is a torus, P is a 4-gon and $r = 4$. The black edges of ∂P projects to G in $\mathscr{S}$. Here G is composed of two loops with basepoint s. The four radial edges in P projects to the $s - t$ edges of the radial graph which possesses two faces.

closed walks are given by their term product representations in $\langle E(G); f_G\rangle$. We denote by r the size of the facial walk f_G. Hence, $r = 4g$ if $\mathscr{S}$ is orientable with genus g and $r = 2g$ if $\mathscr{S}$ is a non-orientable surface of genus g. We can view the cut surface $\mathscr{S} \backslash\backslash G$ as an open regular r-gon P whose boundary sides are labelled by the edges in f_G. Obviously, the boundary ∂P of P maps to G after gluing back the sides of P, and its vertices all map to s. Each subpath of f_G labels one (or more) oriented subpath of ∂P which we can associate with the chord between its endpoints. A term product representation can thus be seen as a sequence of chords inside P, where each chord stands for a term. In order to represent these chords as walks of constant complexity, we introduce an embedded graph H obtained as follows. Consider the radial graph in P linking the center of P to each vertex of ∂P along a straight segment; this graph has $r + 1$ vertices and r edges. After gluing back the boundary of P we obtain a bipartite graph H cellularly embedded on $\mathscr{S}$ with two vertices $\{s, t\}$ and r edges. Both vertices of H are r-valent; each of the $r/2$ faces of H is of length 4 and is cut in two "triangles" by the unique edge of G it contains — see figure 5.2. We call H the **radial graph** of G. A rotation system of H can be computed from that of G in $O(r)$ time. Any chord in P is now homotopic to the 2-walk with the same extremities and passing through t. Consequently, if $[c] = w_1 \cdots w_k$ is a term product representation of height k stored as a list of pointers then in $O(k)$ time we can obtain a closed walk of length $2k$ in H, homotopic to c.

Tiling of the universal cover. Let $(\tilde{\mathscr{S}}, \pi)$ be the universal cover of $\mathscr{S}$. A loop of H is contractible if and only if its lift in $\tilde{\mathscr{S}}$ is a loop and we want to construct a finite part of $\tilde{\mathscr{S}}$ large enough to contain that lift. To this end we rely on a decomposition of $\tilde{\mathscr{S}}$ similar to the octagonal decomposition of [CE10]. Due to our complexity constraints we cannot afford the cost of computing a tight octagonal decomposition of $\mathscr{S}$. We nevertheless present a tiling of $\tilde{\mathscr{S}}$ with similar properties.

The lifted graph $\tilde{H} := \pi^{-1}(H)$ of H is an infinite regular bipartite graph with r-valent vertices and 4-valent faces. A vertex of $\tilde{H}$ is said of s**-type** or t**-type** if it respectively projects to s or t. Let H^* be the dual of H on $\mathscr{S}$ chosen so that its vertices lie in the middle of the edges of G. The lifted graph $\tilde{H}^*$ is the dual of $\tilde{H}$; it is an infinite graph whose faces form a $\{r, 4\}$-tiling of the universal cover, *i.e.*, four r-gon faces meet at every vertex of $\tilde{H}^*$. We say that two edges of H^* are **facing** each other if they share an endpoint x and are not consecutive in the circular order around x. The four edges meeting at any vertex of H^* form two pairs of facing edges. Facing edges are defined similarly in $\tilde{H}^*$. The **line** induced by $e^* \in E(\tilde{H}^*)$ is the smallest set $\ell_{e^*} \subset E(\tilde{H}^*)$ containing e^* and the facing edge of any of its edges. Every line is an infinite lift of some cycle of facing edges of H^*, but contrary to the tight cycles of the octagonal decomposition in [CE10] this cycle may self-intersect. In fact, it can happen that every line projects onto the *same* self-intersecting cycle of length r. (As an example, one may consider the case where the reduced graph G is embedded on the genus two orientable surface with $f_G = aba^{-1}cb^{-1}dc^{-1}d^{-1}$.) To begin with, we state a Dehn-like result about tilings. It will be used in Section 5.3.2 to bound the size of the relevant region.

Lemma 5.3.1. *Let Γ be a regular $\{r,4\}$-tiling of the plane (each face has r edges and each vertex has 4 neighbours) with $r \geq 4$. Every finite non-empty union R of faces of Γ contains at least one face sharing $r-2$ consecutive edges with the boundary of R.*

Proof. We follow the proof in [Sti93, p. 188] and gather the faces of Γ in concentric rings at increasing distance to some root face. Let f be a face of R in the outermost ring of R. At least $r-3$ consecutive edges of f lie on the boundary of R. Indeed, f shares (at most) two edges with other faces of the same ring, and at most one edge with faces of the ring just beneath it. One more edge of f lies on the boundary of R if f shares no edge with the inner ring, or if at least one of the neighbors of f in the outermost ring is not in R. Else, we can replace f by any of its two ring neighbors, which do not share edges with the inner ring, to obtain a face with the required property. □

Let again Γ be a regular $\{r,4\}$-tiling of the plane with $r \geq 4$. By the Jordan curve theorem a closed simple path c in the vertex-edge graph of Γ bounds a finite union R of faces. A vertex v of c is called **convex**, **flat**, or **reflex** if it is incident to respectively 0, 1, or 2 edges interior to R.

Lemma 5.3.2. *A closed simple path c has at least r convex vertices.*

Proof. We denote by V_c, V_f and V_r the respective number of convex, flat and reflex vertices of c. If R is the region bounded by c, we also denote by F its number of faces and by V_i its number of interior vertices. We finally set $V := V_i + V_c + V_f + V_r$ and call E the total number of edges of the closed region R. By double-counting of the vertex-edge incidences, we get:

$$2E = 4V_i + 2V_c + 3V_f + 4V_r.$$

By double-counting of the face-edge incidences, we get:

$$rF = 2E - (V_c + V_f + V_r).$$

Euler's formula then implies $V - E + F = 1$. Multiplying by r we obtain:

$$r = rV - rE + rF = rV - (r-2)E - (V_c + V_f + V_r),$$

and expending the values of V and E:

$$r = V_c - (r-4)V_i - \frac{r-4}{2}V_f - (r-3)V_r.$$

Since $r \geq 4$, we conclude that $V_c \geq r$. □

We can now state the main properties of lines in $\tilde{H}^*$.

Proposition 5.3.3. *Suppose that $\mathscr{S}$ is orientable with genus ≥ 2 or non-orientable with genus ≥ 3. Then lines of $\tilde{H}^*$ do not self-intersect nor are cycles and two distinct lines intersect at most once. Moreover, the following properties hold.*

- ***(triangle-free Property)*** *Three pairwise distinct lines cannot pairwise intersect.*
- ***(quad-free Property)*** *Four pairwise distinct lines can neither form a quadrilateral,* i.e.*, include a possibly self-intersecting 4-gon.*

PROOF. With the above hypothesis $\tilde{H}^*$ defines a regular $\{r,4\}$-tiling of the plane with $r \geq 6$. A contradiction of any of the above properties would imply the existence of a region bounded by at most four lines. Its boundary would contain a simple closed curve with at most four convex vertices, which is forbidden by the previous lemma. □

Lemma 5.3.4. *Every line cuts $\tilde{\mathscr{S}}$ in two infinite connected components.*

PROOF. Let ℓ be a line induced by an edge e^*. We first show that $\tilde{\mathscr{S}} \setminus \ell$ has at most two connected components. Consider a small disk D centered at the midpoint of e^*, so that $D \setminus \ell$ has two half-disk components D_1 and D_2. Any point x of $\tilde{\mathscr{S}} \setminus \ell$ is in the same component as one of D_1 or D_2. Indeed, we can use an approach path from x to a point close to ℓ and then follow a path along ℓ on the same side as x until we meet D, hence falling into D_1 or D_2. It remains to prove that $\tilde{\mathscr{S}} \setminus \ell$ has at least two components. Consider the two endpoints y and z of the edge e dual to e^*. Let p be any path joining y and z in $\tilde{\mathscr{S}}$. We just need to show that p and ℓ intersect. By pushing p along $\tilde{H}$, it is easily seen that p is homotopic to a path q in $\tilde{H}$. Moreover, p cuts ℓ if and only if q does so. Since e and q share the same endpoints, their projection $\pi(e)$ and $\pi(q)$ are homotopic paths of H in $\mathscr{S}$. It is part of the folklore that homotopy in a cellularly embedded graph can be realized by combinatorial homotopies. Such homotopies are finite sequences of elementary moves obtained by replacing a subpath contained in a facial walk by the complementary subpath in that facial walk. Any combinatorial homotopy between $\pi(e)$ and $\pi(q)$ lifts to a combinatorial homotopy between e and q. Since the lift of an elementary move preserve the parity of the number of intersections with ℓ, we conclude that q has an odd number of intersections with ℓ. It follows that q, whence p, must cross ℓ. □

5.3.2 Building the relevant region

The relevant region. Let p be a path in $\tilde{H}$. Following [CE10] we denote by Π_p, and call the **relevant region** with respect to p, the union of closed faces of $\tilde{H}^*$ reachable from $p(0)$ by crossing only lines crossed by p, in any order. In other words, if for any line ℓ we denote by ℓ^+ the component of $\tilde{\mathscr{S}} \setminus \ell$ that contains $p(0)$, then Π_p is the "convex polygon" of $\tilde{\mathscr{S}}$ formed by intersection of the sets $\ell \cup \ell^+$ for all ℓ *not* crossed by p. See Figure 5.3 for an illustration. This region has interesting properties which make it easy and efficient to build:

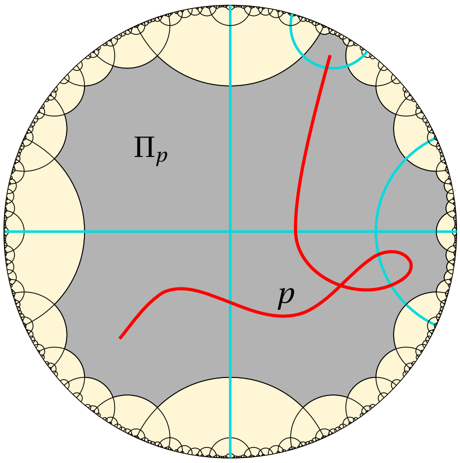


Figure 5.3: The Poincaré disk model of $\tilde{\mathscr{S}}$ with lines represented as hyperbolic geodesics. Remark that the union of (light blue) lines crossed by p is not necessarily connected.

Lemma 5.3.5 ([CE10, lemma 4.1])**.** *For any path $p \subset \tilde{H}$ and any line ℓ the intersection $\ell \cap \Pi_p$ is either empty or a segment of ℓ whose relative interior is entirely included in either the interior or the boundary of Π_p.*

PROOF. If $\ell \cap \Pi_p$ is not empty, consider two points $x, y \in \ell \cap \Pi_p$. By definition of the relevant region, there is a path u joining x to y (through $p(0)$) whose relative interior is only crossed by lines also crossing p. Any line crossing ℓ between x and y separates these two points, hence crosses u, hence crosses p. It easily follows that the open segment of ℓ between x and y is included in Π_p, either along its boundary or in its interior. □

Lemma 5.3.6 ([CE10, lemma 4.3])**.** *Π_p contains at most $\max(5|p|, 1)$ faces of $\tilde{H}$.*

PROOF. We recall that a vertex v of the boundary of Π_p is **convex** if no line passing through v is crossed by p. It is **flat** if on the contrary one line through v is crossed by p. By Lemma 5.3.4, a line that intersects the interior of Π_p must cross p. By Lemma 5.3.5, such a line intersect the boundary of Π_p in at most two flat vertices. It follows that the number of flat vertices is at most $2|p|$. If there is no flat vertex then Π_p is a single face, $|p| = 0$ and the lemma holds. Else, note that between two consecutive flat vertices there are at most $r-2$ convex vertices, all on the boundary of a single face. Accordingly the boundary of Π_p has at most $2(r-1)|p|$ edges.

On the other hand, any union R of $k \geq 1$ faces has at least $(r-4)k$ edges on its boundary. Indeed, thanks to Lemma 5.3.1 we can recursively remove a face with at least $r-2$ boundary edges to decrease the perimeter by at least $r-4$, until R is empty. Applying this last result to Π_p we get $(r-4)k \leq 2(r-1)|p|$, whence:

$$k \leq \frac{2(r-1)}{r-4}|p| \leq 5|p|$$

the latter inequality stemming from the hypothesis $r \geq 6$. □

The next result is crucial for building the relevant region:

Lemma 5.3.7 ([CE10, lemma 4.2])**.** *Let p be a path of $\tilde{H}$ and let e be an edge with $p(1)$ as an endpoint. Suppose e crosses a line ℓ not already crossed by p. Then $\Pi_p \cap \ell$ is a segment of the boundary of Π_p along a connected set of faces $V \subset \Pi_p$. Moreover $\Pi_{p\cdot e} = \Pi_p \cup \Lambda$ where Λ is the reflection of V across ℓ.*

Proof. The intersection point of e and ℓ belong to the same (closed) face of Π_p as $p(1)$. The set $X := \Pi_p \cap \ell$ is thus non-empty and, by lemma 5.3.5, connected. Moreover, as ℓ does not cross p, the line segment X is part of the boundary of Π_p. In particular, each edge in X bounds some face of Π_p. These faces form a strip V where two consecutive faces have an interior edge of Π_p in common.

Each face in V has a bounding edge in X that also bounds a symmetric face outside Π_p. The union of these symmetric faces is the reflection Λ of V. Furthermore $\Pi_p \cup \Lambda \subset \Pi_{p\cdot e}$ since one can reach any face of Λ from the symmetrical face of V just by crossing ℓ. To prove the reverse inclusion it is sufficient to show that none of the lines bounding Λ are crossed by p. A single face f of Λ is bounded by ℓ, two *inner* lines crossing ℓ at vertices of f, and $r-3$ other *outer* lines. Since these two inner lines both cross ℓ, the triangle-free Property ensures that the two are disjoint and bound a *band* of $\tilde{\mathscr{S}}$ that contains f. No line crosses ℓ in this band, and a line crossing ℓ cannot cross an inner line as the three would pairwise intersect — which is again prohibited by the triangle-free Property. Since the outer lines bounding f have an edge in the band they cannot cross ℓ, hence cannot cross p. Consider now an inner line ℓ'. It contributes to the boundary of Λ only when f is an extremal face of Λ. In that case, the symmetrical face of f in V is also extremal and ℓ' is bounding Π_p. So that ℓ' does not cross p. We conclude the proof by noting that ℓ is the last line to consider on the boundary of Λ. $\square$

Computing the relevant region. Following lemma 5.3.7 we will build the relevant region of the lift $\tilde{c}$ of a loop c incrementally as we lift its edges one at a time. If p is the subpath of the lift $\tilde{c}$ already traversed and $\tilde{e}$ is the edge following p in $\tilde{c}$ we extend Π_p to $\Pi_{p\cdot\tilde{e}}$ whenever $\tilde{e}$ crosses the boundary of Π_p. From Lemma 5.3.6, we know that Π_p contains $O(\left|p\right|)$ faces. However, a naive representation of Π_p as a subgraph of $\tilde{H}^*$ with its faces and edges would require $O(r\left|p\right|)$ space, which we cannot afford to obtain a linear time contractibility test. We rather store the interior of Π_p by its dual graph, *i.e.*, by the subgraph of $\tilde{H}$ induced by the vertices dual to the faces of Π_p. More precisely, to represent such a finite subgraph Γ of $\tilde{H}$ we use an abstract data structure with the following operations:

- new(x), which returns and adds to $V(\Gamma)$ a new vertex v with no neighbour, where $x \in \{s,t\}$ represents the projection $\pi(v)$.
- type(v), which yields the projection $\pi(v)$ for any vertex $v \in V(\Gamma)$ — that is type(new(x)) $= x$.
- join(v,w,e), which adds a new edge $\tilde{e} \in E(\Gamma)$ with endpoints v and w, where v and w are existing vertices of Γ of different type and e is an edge of H representing $\pi(\tilde{e})$. Calling join(v,w,e) and join(w,v,e) is equivalent, and using any of those more than once for the same v and e is unsupported.

- next(v, e), which finds the other endpoint of the lift of e in Γ, if it exists. More formally, it returns the unique vertex $w \in \Gamma$ such that join(v, w, e) has been called, if it exists, and a special value NONE otherwise.

Given a loop c in H, the following procedure then constructs $\Pi_{\tilde{c}}$:

1. Create a variable v which will point to the *current endpoint* of the partial lift of c. With the operation $v \leftarrow$ new(s), create a single vertex corresponding to the relevant region with respect to a null path.

2. Call e the next edge to process in c; if there is none left, exit.

3. If next$(v, e) \neq$ NONE, set $v \leftarrow$ next(v, e) and return to step 2.

4. If next$(v, e) =$ NONE, the partial lift is exiting the current relevant region. In other words the lift $\tilde{e}$ of e crosses a line ℓ not crossed until now and we are in the situation of lemma 5.3.7. We need to enlarge Γ by performing a *mirror operation*, creating and attaching vertices mirrored across ℓ as suggested by the lemma. When done, next$(v, e) \neq$ NONE and we can return to step 3.

When we reach the end of c we have computed the relevant region of its lift. Moreover, c is contractible if and only if its lift is closed, which reduces to the equality of the current endpoint v with the first created endpoint. It remains to detail the mirror operation.

The mirror operation. Let $v \in V(\Gamma)$ and $e \in E(H)$ such that next$(v, e) =$ NONE. Denote by $\tilde{e}$ the lift of e from v in $\tilde{H}$ and by ℓ the line crossed by $\tilde{e}$. Let also p be the partial lift of c already processed; in particular Γ is the subgraph of $\tilde{H}$ included in Π_p. We begin the mirror sub-procedure by creating a vertex $w =$ new(x), where $x \in \{s, t\} \setminus \{\text{type}(v)\}$, and calling join$(v, w, e)$ so that $\tilde{e}$ is now represented in Γ.

We say that two edges of $\tilde{H}$ are **siblings** if their dual edges are facing each other in $\tilde{H}^*$. Obviously two siblings cross the same line and have no common endpoint. The four edges bounding any face of $\tilde{H}$ form two pairs of siblings matching the two pairs of facing edges around the corresponding vertex of $\tilde{H}^*$. Let $\tilde{e}_1$ be one of the two siblings of $\tilde{e}$. Denote by $\tilde{e}_0$ and $\tilde{e}_2$ the other sibling pair bounding the same face as $\tilde{e}$ and $\tilde{e}_1$, so that v is an endpoint of $\tilde{e}_0$ — see Figure 5.4. If $\tilde{e}_1$ has an endpoint in Π_p, or equivalently if $\tilde{e}_0 \in E(\Gamma)$, then by lemma 5.3.7 $\tilde{e}$, $\tilde{e}_0$, $\tilde{e}_1$ and $\tilde{e}_2$ are all included in $\Pi_{p\cdot\tilde{e}}$. In that case $\tilde{e}_1$ and $\tilde{e}_2$ need to be added to Γ. If on the contrary $\tilde{e}_0 \notin E(\Gamma)$ then v lies in a face of $\tilde{H}^*$ that is extremal in the chain V of lemma 5.3.7, and the sibling $\tilde{e}_1$ should not be added to Γ. We proceed as follows: by walking from $\pi(v)$ around one face bounded by e in H we figure out the projections e_0, e_1 and e_2 of $\tilde{e}_0$, $\tilde{e}_1$ and $\tilde{e}_2$ respectively. If $v_1 =$ next$(v, e_0) \neq$ NONE then we create $w_1 \leftarrow$ new(type(v)), and call join(v_1, w_1, e_1) and join(v_1, w, e_2). We handle similarly the sibling of $\tilde{e}_1$ which is not $\tilde{e}$, walking around the other face bounded by e_1 and checking whether v_1 is extremal in the chain V, and so on until we reach the end of V. There remains to do the mirror in the other direction, starting from the still unprocessed face of H bounded by e. Eventually, every vertex dual to a face in the chain Λ of Lemma 5.3.7 has been created, and that lemma ensures that we missed no vertex. Since each time we add a vertex we also add all its edges in $\tilde{H}$ linking to existing vertices of Γ, we now have that Γ is the dual graph of $\Pi_{p\cdot\tilde{e}}$.

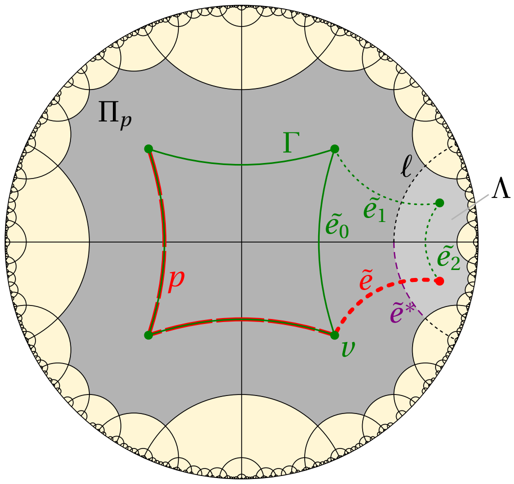


Figure 5.4: The path p is composed of two edges. The graph Γ corresponding to its relevant region Π_p has four edges bounding a face of $\tilde{H}$.

Data structure. We present an implementation of the abstract graph structure that only needs *constant time* to perform any of its operations. We use a technique inspired from [AHU74, exercise 2.12 p. 71] taking advantage of the RAM model to allocate in $O(1)$ time an r-sized segment of memory without initializing it. We begin by giving integer indices between 1 and r to edges of H. The index of e will be denoted id(e), and tables will be indexed from 1. Then:

- new(x) creates a vertex structure v with a field `type` pointing to x, an integer field `count` with value 0, two uninitialized tables `index` and `rev` of r integers, and an uninitialized table `neighbor` of r pointers;

- type(v) returns the value of `type`;

- join(v, w, e) increments `count` in v, points `neighbor[count]` towards w, sets `index[count]` $\leftarrow$ id(e) and `rev`[id(e)] $\leftarrow$ `count`, and affects w similarly;

- next(v, e) returns `neighbor[rev`[id(e)]] if we have $1 \leq$ `rev`[id(e)] $\leq$ `count` together with `index[rev`[id(e)]] $=$ id(e), and returns NONE otherwise.

Of course if join(v, w, e) has been called for some w then next(v, e) will indeed return a pointer to w. If not, then `rev`[id(e)] will still be uninitialized, and even if by chance $1 \leq$ `rev`[id(e)] $\leq$ `count` the corresponding cell of `index` will have been filled by another join operation and the *round-trip* check will fail.

Complexity. Checking and adding a sibling in the mirror subprocedure needs one face traversal in H, one call to next and if needed one call to new and two to join. The initialisation of the mirror operation reduces to one call to new and another one to join. In the end all mirrors will have used at most one next, one new and two join operations for each vertex of Γ — that is each face of $\Pi_{\tilde{c}}$. Moreover, step 3 in the construction of $\Pi_{\tilde{c}}$ uses one next operation per edge of c. Because every operation takes $O(1)$ time, and thanks to lemma 5.3.6 our algorithm takes $O(|c|)$ time. Taking into account the precomputation of Lemma 5.2.3 we have proved Theorem 5.1.1 when

$r \geq 6$. This is the case when $\mathscr{S}$ is an orientable surface of genus at least 2 or when $\mathscr{S}$ is non-orientable with genus at least 3. In the remaining cases we can expand the term product representations in $O(|c|)$ time to obtain a word in the computed presentation of $\pi_1(\mathscr{S}, s)$. The word problem, that is testing if a word represents the unity in a group presentation, is trivial in those cases. This is clear for the torus or the projective plane since their fundamental group is commutative. When $\mathscr{S}$ is the Klein bottle, we can assume that the computed presentation is $\langle a, b\,;\, abab^{-1}\rangle$ by applying an easy change of generators if necessary. The relator $abab^{-1}$ allows us to commute a and b up to an inverse and to get a canonical form $a^u b^v$ in $O(|c|)$ time. This in turn solves the word problem in linear time.

5.4 The Free Homotopy Test

We now tackle the free homotopy test. We restrict to the case where $\mathscr{S}$ is an orientable surface of genus at least two. We can thus orient the cellular embedding. This amounts to give a preferred traversal direction to each facial walk. In particular, every oriented edge e belongs to exactly one such facial walk, which we designate as the left face of e. This allows us in turn to associate with e a dual edge oriented from the left face to the right face of e. This correspondence between the oriented edges of the embedded graph and its dual will be implicit in the sequel.

We want to decide if two cycles c and d on $\mathscr{S}$ are freely homotopic. After running our contractibility test on c and d, we can assume that none of these two cycles is contractible. From Lemma 5.2.3 and the discussion of the simplified framework in Section 5.3.1, we can also assume that c and d are given as closed walks in the radial graph H. Let $(\mathscr{S}_c, \pi_c)$ be the c-cyclic cover of $\mathscr{S}$. Recall that $\mathscr{S}_c$ can be viewed as the orbit space of the action of $\langle\tau\rangle$ where τ is the unique automorphism of $(\tilde{\mathscr{S}}, \pi)$ sending $\tilde{c}(0)$ on $\tilde{c}(1)$ for a given lift $\tilde{c}$ of c. We will refer to τ as a *translation* of $\tilde{\mathscr{S}}$, as it can indeed be realized as a translation of the hyperbolic plane. Notice that $\mathscr{S}$ being orientable, τ is orientation preserving. The projection φ_c sending a point of $\tilde{\mathscr{S}}$ to its orbit makes $(\tilde{\mathscr{S}}, \varphi_c)$ a covering space of $\mathscr{S}_c$ with $\pi = \pi_c \circ \varphi_c$. We denote by H_c and H_c^* the respective lifts of H and H^* in $(\mathscr{S}_c, \pi_c)$. In the sequel, regularity of paths in $\mathscr{S}$, $\tilde{\mathscr{S}}$ or $\mathscr{S}_c$ is considered with respect to H^*, $\tilde{H}^*$ and H_c^* respectively, and so are the crossing weights of regular paths.

5.4.1 Structure of the cyclic cover

Recall that a line in $\tilde{\mathscr{S}}$ is an infinite sequence of facing edges in $\tilde{H}^*$. We start by stating some structural properties of lines. Most of these properties appear in [CE10] in one form or another. However, due to our different notion of lines, we cannot rely on the proofs therein. For instance, a line as in [CE10] projects to a simple curve in the c-cyclic cover $\mathscr{S}_c$. In our case, though, a line may project to a self-intersecting curve in $\mathscr{S}_c$. In fact, as previously noted, it may be the case that all of our lines have the same projection on $\mathscr{S}$.

τ-transversal lines. Let ℓ be a line such that $\ell \cap \tau(\ell) = \emptyset$ and denote by $\mathring{B}_\ell$ the open band of $\tilde{\mathscr{S}}$ bounded by ℓ and $\tau(\ell)$. The line ℓ is said **τ-transversal** if $B_\ell := \mathring{B}_\ell \cup \ell$ is a

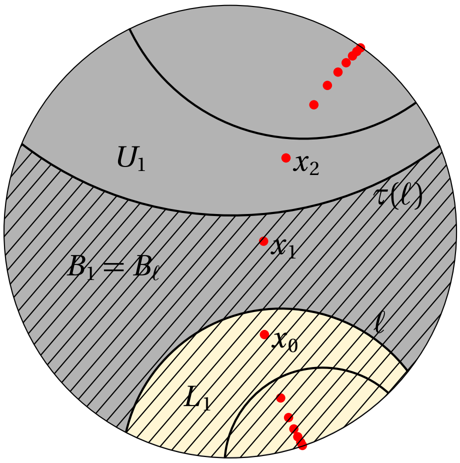


Figure 5.5: The fundamental domain B_1 is the intersection of the grey upper region U_1 with the hatched lower region L_1.

fundamental domain[2] for the action of $\langle\tau\rangle$ over $\tilde{\mathscr{S}}$. Equivalently, ℓ is τ-transversal if $\tilde{\mathscr{S}}$ is the disjoint union of all the translates of B_ℓ. In such a case we can obtain $\mathscr{S}_c$ by pointwise identification of the boundaries ℓ and $\tau(\ell)$ of B_ℓ. The following proposition gives a characterisation τ-transversal lines whose existence are stated in Proposition 5.4.6.

Proposition 5.4.1. *Let ℓ be a line such that $\ell \cap \tau(\ell) = \emptyset$. Then, ℓ is τ-transversal if and only if there exists $x \in \tilde{\mathscr{S}}$ such that ℓ separates x from $\tau(x)$ but $\tau^{-1}(\ell)$ and $\tau(\ell)$ do not.*

Proof. If B_ℓ is a fundamental domain, then $\mathring{B}_\ell$ and $\tau^{-1}(\mathring{B}_\ell)$ are disjoint and separated by ℓ. The direct implication in the equivalence of the lemma is thus trivial. For the reverse implication consider an x as in the proposition. Put $x_i := \tau^i(x)$ and $\ell_i := \tau^i(\ell)$. By assumption, for all i, ℓ_i separates x_i from x_{i+1} while ℓ_{i-1} and ℓ_{i+1} do not. Let U_i and L_i be the two components of $\tilde{\mathscr{S}} \setminus \ell_i$ with $x_{i+1} \in U_i$ and $x_i \in L_i$. From the assumption, we also have for all i that $x_i \in L_{i+1}$ and $x_{i+1} \in U_{i-1}$. In particular, L_i contains both x_i and x_{i-1}. Since ℓ_{i-1} separates these two points it must be that $\ell_{i-1} \subset L_i$. Likewise, we have $\ell_{i+1} \subset U_i$. In other words, we have $L_{i-1} = \tau^{-1}(L_i) \subset L_i$ and $U_{i+1} = \tau(U_i) \subset U_i$. We finally set $B_i := (L_i \cap U_{i-1}) \cup \ell_{i-1}$. See Figure 5.5.

We claim that the B_i's are pairwise disjoint. Indeed, for $i < j$, the inclusions $\ell_{i-1} \subset L_i \subset L_{j-1}$ imply $\ell_{i-1} \cap B_j = \emptyset$ since L_{j-1} and U_{j-1} are disjoint. Likewise, the inclusions $\ell_{j-1} \subset U_{j-2} \subset U_{i-1}$ imply $\ell_{j-1} \cap B_i = \emptyset$. We thus obtain

$$B_i \cap B_j = (L_i \cap U_{i-1}) \cap (L_j \cap U_{j-1}) = L_i \cap U_{j-1} \subset L_i \cap U_i = \emptyset.$$

Moreover, the union of the B_i's cover $\tilde{\mathscr{S}}$. To see this, consider any point y in $\tilde{\mathscr{S}}$ contained in the closure of some face f_y of $\tilde{H}^*$. Let p be a path in $\tilde{H}$ between the vertex dual to f_y and the vertex dual to a face that contains x_1. If p does not cross any ℓ_i, then p, hence y, must be contained in B_1. Otherwise, let z be the first intersection point along p with $\cup_i \ell_i$. By considering the appropriate translate of the subpath of p between y and z, we see that some translate of y is contained in B_1. We conclude the lemma by noting that $B_{i+1} = \tau(B_i)$, whence $\tilde{\mathscr{S}}$ is the disjoint union of the translates of $B_1 = B_\ell$. □

[2] *i.e.*, every orbit of $\langle\tau\rangle$ has a unique representative in B_ℓ.

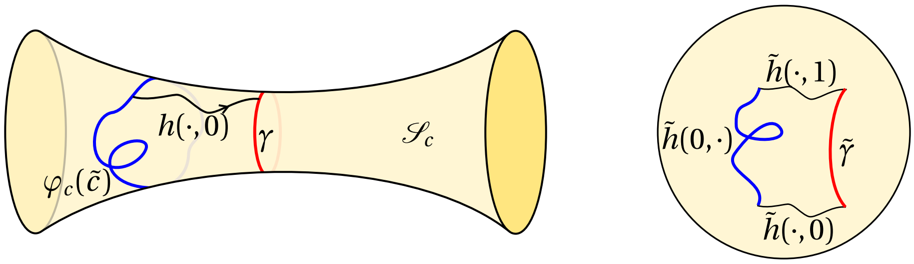


Figure 5.6: (Left) $\varphi_c(\tilde{c})$ and γ are homotopic as generators of $\mathscr{S}_c$. (Right) Any homotopy h lifts to a homotopy in $\tilde{\mathscr{S}}$.

Generators of the cyclic cover. A loop of $\mathscr{S}_c$ that is regular (for H_c^*) and freely homotopic to $\varphi_c(\tilde{c})$ is called a **generator** of $\mathscr{S}_c$. Since $\pi_c(\varphi_c(\tilde{c})) = \pi(\tilde{c}) = c$, every generator projects on $\mathscr{S}$ to a loop that is freely homotopic to c. Conversely, every regular loop freely homotopic to c has a lift in $\mathscr{S}_c$ which is a generator. A **minimal generator** is a generator whose crossing weight (with respect to H_c^*) is minimal among generators; it projects to a regular loop of minimal crossing weight in the free homotopy class of c.

Lemma 5.4.2. *Let $\tilde{\gamma} \subset \tilde{\mathscr{S}}$ be any lift of a generator γ of $\mathscr{S}_c$. Then, τ sends $\tilde{\gamma}(0)$ on $\tilde{\gamma}(1)$.*

Proof. By definition of a generator there exists a homotopy $h : [0,1]^2 \to \mathscr{S}_c$ from $\varphi_c(\tilde{c})$ to γ. Consider in $\tilde{\mathscr{S}}$ the unique lift $\tilde{h}$ of h with $\tilde{h}(1,0) = \tilde{\gamma}(0)$ (see Figure 5.6). Then $\tilde{h}(1,\cdot) = \tilde{\gamma}$ and $\tilde{h}(0,\cdot)$ is some lift of $\varphi_c(\tilde{c})$. Since $\mathscr{S}_c$ is a cylinder, the automorphism group of $(\tilde{\mathscr{S}}, \varphi_c)$ is the cyclic group generated by τ. It follows that $\tilde{h}(0,\cdot) = \tau^i(\varphi_c(\tilde{c}))$ for some integer i. In particular, $\tilde{h}(0,1) = \tau(\tilde{h}(0,0))$. Since $h(\cdot,0) = h(\cdot,1)$, it ensues by uniqueness of lifts that $\tilde{h}(.,1) = \tau(\tilde{h}(\cdot,0))$. Whence $\tilde{h}(1,1) = \tau(\tilde{h}(1,0))$ which is exactly $\tilde{\gamma}(1) = \tau(\tilde{\gamma}(0))$. □

It follows that a regular path joining two points x and y in $\tilde{\mathscr{S}}$ is a lift of a generator if and only if $y = \tau(x)$. Consider a minimal generator γ and one of its lifts $\tilde{\gamma} \subset \tilde{\mathscr{S}}$. Its reciprocal image $\ell_\gamma := \varphi_c^{-1}(\gamma)$ is the curve obtained by concatenation of all the translates $\tau^i(\tilde{\gamma})$, $i \in \mathbb{Z}$, of $\tilde{\gamma}$. Remark that, as far as intersections of $\tilde{\gamma}$ with lines is concerned, we can assume that γ, hence ℓ_γ, is simple. Indeed, if γ self-intersects it must contain by Lemma 5.2.2 a subpath forming a contractible loop; by minimality of γ this subpath does not intersect any line and we can cut it off without changing the intersections of γ with any line. By induction on the number of self-intersections, we can thus make γ simple. The simple curve ℓ_γ actually behaves like a line in $\tilde{\mathscr{S}}$:

Lemma 5.4.3. *ℓ_γ is separating in $\tilde{\mathscr{S}}$.*

Proof. Since γ is already separating on $\mathscr{S}_c$, this is certainly the case for its reciprocal image ℓ_γ. □

Lemma 5.4.4. *A line can intersect ℓ_γ at most once.*

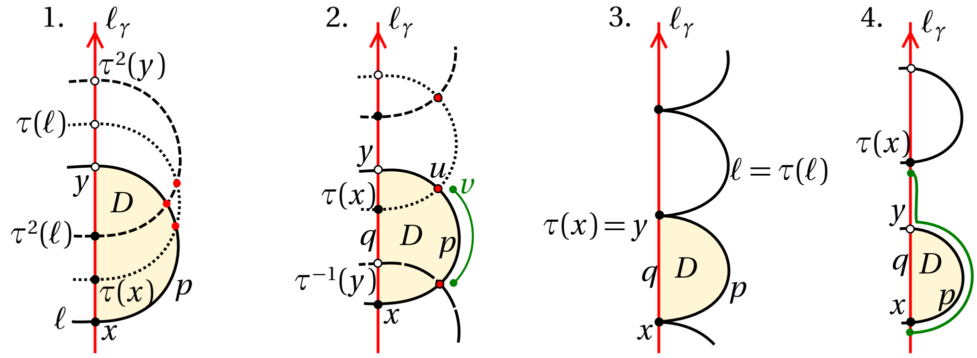


Figure 5.7: Along ℓ_γ, the point y lies (1.) beyond $\tau^2(x)$, (2.) strictly between $\tau(x)$ and $\tau^2(x)$, (3.) on $\tau(x)$ and (4.) strictly between x and $\tau(x)$.

Proof. Suppose that a line ℓ crosses ℓ_γ at least twice. Then, there exists a disk bounded by a subpath of ℓ and a subpath of ℓ_γ that only intersect at their common extremities. Let D be an inclusion-wise minimal such disk. Call p the subpath of ℓ bounding D and call q the subpath of ℓ_γ bounding D. Let x and y be the common extremities of p and q so that x occurs before y along ℓ_γ (oriented from x to $\tau(x)$). Note that ℓ cannot cross q as this would otherwise contradict the minimality of D. Since τ is orientation preserving, p and its translates $\tau^i(p)$ occur on the same side of ℓ_γ (see previous lemma). We shall consider all the relative positions of y with respect to x, $\tau(x)$ and $\tau^2(x)$ along ℓ_γ and reach a contradiction in each case. Figure 5.7 illustrates each of the cases.

1. If $y = \tau^2(x)$, or if $\tau^2(x)$ separates x from y on ℓ_γ, then $\tau(\ell)$ and $\tau^2(\ell)$ cross q and also cross p by minimality of D. It follows that ℓ, $\tau(\ell)$ and $\tau^2(\ell)$ would pairwise intersect, in contradiction with the triangle-free Property.

2. If y strictly lies between $\tau(x)$ and $\tau^2(x)$, then p and $\tau(p)$ must intersect. Let u be their intersection point (there is a unique one since we cannot have $\ell = \tau(\ell)$). We introduce the notation $\alpha[u, v]$ for the subpath of a simple path α between two of its points u and v. The concatenation

$$p[\tau^{-1}(u), u] \cdot \tau(p)[u, \tau(x)] \cdot q[\tau(x), \tau^{-1}(y)] \cdot \tau^{-1}(p)[\tau^{-1}(y), \tau^{-1}(u)]$$

is a simple closed path. It follows from the triangle-free Property that any line intersecting $p[\tau^{-1}(u), u]$ must also intersect $q[\tau(x), \tau^{-1}(y)]$. Consider a point v near u and outside $D \cup \tau(D)$ and consider a regular path p' between $\tau^{-1}(v)$ and v, close to $p[\tau^{-1}(u), u]$ but disjoint from p; we may choose p' in such a way that it crosses the same line as the relative interior of $p[\tau^{-1}(u), u]$. It follows that the crossing weight of p' is strictly less than the crossing weight of the regular path $q[x', \tau(x')]$, where $x' \in q$ is a point close to x. This contradicts the fact that γ is a minimal generator since $q[x', \tau(x')]$ projects to γ and p' projects to a shorter generator of $\mathscr{S}_c$.

3. If $y = \tau(x)$ then γ and $\pi_c(\ell) = \pi_c(p)$ are homotopic and intersect once. Since $\mathscr{S}_c$ is a cylinder, this implies that γ and $\pi_c(\ell)$ touch but do not cross transversely. This in turn contradicts the regularity of γ.

4. Finally, if y strictly separates x from $\tau(x)$ then the path obtained by following ℓ between x and y and then ℓ_γ between y and $\tau(x)$ provides, after infinitesimal perturbation, a lift of a generator with crossing weight strictly less than γ. This contradicts the minimality of γ.

□

Lemma 5.4.5. *If a line ℓ intersects ℓ_γ, then $\ell \cap \tau(\ell) = \emptyset$.*

PROOF. Suppose that ℓ intersects both ℓ_γ and $\tau(\ell)$. From Lemma 5.4.4, ℓ and ℓ_γ cross exactly once. In particular, ℓ and $\tau(\ell)$ cannot coincide, hence must cross exactly once. Let $x := \ell \cap \ell_\gamma$ and let $y := \ell \cap \tau(\ell)$. Since τ preserve the orientation, y and $\tau(y)$ must be on the same side of ℓ_γ. The point $\tau(y)$ cannot lie on the same side of ℓ as $\tau(x)$. Indeed, $\tau^2(\ell)$ would then enter the triangle $(x, \tau(x), y)$ formed by ℓ_γ, $\tau(\ell)$ and ℓ. From the preceding Lemma 5.4.4 and from the triangle-free Property, $\tau^2(\ell)$ could not exit that triangle, leading to a contradiction. But $\tau(y)$ can neither lie on the other side of ℓ as $\tau^2(\ell)$ would have to cross both ℓ and $\tau(\ell)$, contradicting the triangle-free Property. □

Proposition 5.4.6. *Let γ be a minimal generator whose basepoint is not on any line. Let $\tilde\gamma$ be a lift of γ in $\tilde{\mathscr{S}}$. Any line ℓ crossed by $\tilde\gamma$ is τ-transversal. In particular, there exists a τ-transversal line.*

PROOF. Put $x := \tilde\gamma(0)$. By Lemma 5.4.2, $\tilde\gamma(1) = \tau(x)$. By Lemma 5.4.4, any line ℓ crossing $\tilde\gamma$ separates x from $\tau(x)$ while $\tau^{-1}(\ell)$ and $\tau(\ell)$ do not. Together with Lemma 5.4.5, this allows us to apply Proposition 5.4.1 and to conclude that ℓ is τ-transversal. Since $\tilde\gamma$ must be crossed by at least one line, this finally prove the existence of a transversal line. Indeed, if $\tilde\gamma$ was not crossed by any line it would lie inside a single face of $\tilde H^*$: its projection $\pi(\tilde\gamma) \subset \mathscr{S}$ would also stay in a single face of H^* and be trivially contractible, contradicting the fact that $\pi(\tilde\gamma)$ is freely homotopic to c. □

Corollary 5.4.7. *There exists a τ-transversal line that separates the endpoints of $\tilde c$ from each other.*

PROOF. By Proposition 5.4.6, there exists a τ-transversal line ℓ. Since $\tilde c(1)$ is a translate of $\tilde c(0)$, these two endpoints are included in two successive fundamental domains determined by ℓ and its translates. Since the basepoint s of c is not on the projection of any line, the translate of ℓ that separates the interior of these two fundamental domains also separates $\tilde c(1)$ from $\tilde c(0)$. □

τ-invariant line. A line ℓ such that $\tau(\ell) = \ell$ is said **τ-invariant**. Note that this equality does not hold pointwise, but globally.

Lemma 5.4.8. *There is at most one τ-invariant line.*

PROOF. Let ℓ and ℓ' be two τ-invariant lines. Since a τ-invariant line cannot lie in a single fundamental domain, any τ-transversal line λ must intersect ℓ and ℓ'. If ℓ and ℓ' where distinct then they would form a quadrilateral with λ and $\tau(\lambda)$, in contradiction with the quad-free Property. $\square$

Lemma 5.4.9. *A line that intersects three consecutive translates of a τ-transversal is τ-invariant.*

PROOF. Suppose that a line ℓ intersects a τ-transversal line λ as well as $\tau(\lambda)$ and $\tau^2(\lambda)$. If $\ell \neq \tau(\ell)$ then ℓ, $\tau(\ell)$, $\tau(\lambda)$ and $\tau^2(\lambda)$ form a quadrilateral, in contradiction with the quad-free Property. $\square$

5.4.2 The canonical generator

Since $\mathscr{S}$ is oriented we can speak of the left or right side of a minimal generator. Our aim is to prove that the set of minimal generators of $\mathscr{S}_c$ covers a bounded cylinder allowing us to define its right boundary as a canonical representative of the free homotopy class of c. By definition, a τ-transversal line projects in $\mathscr{S}_c$ to a simple curve. We call this projection a c-**transversal**. A c-transversal ℓ crosses exactly once every minimal generator γ. Indeed, if ℓ and γ had two intersections x and y, the subpath of ℓ between x and y would be homotopic to one of the two paths cut by x and y along γ (see Lemma 5.2.1). These two homotopic subpaths would lift in $\tilde{\mathscr{S}}$ to a closed path, implying that a lift of ℓ cuts ℓ_γ twice; a contradiction with Lemma 5.4.4. Moreover, Proposition 5.4.6 implies that any minimal generator is crossed by c-transversals only. The number of c-transversals in $\mathscr{S}_c$ is thus equal to the length of the minimal generators which is in turn no larger than $|c|$. Notice that the orientation of $\mathscr{S}$ and of the minimal generators induce a left-to-right orientation of the c-transversals.

Lemma 5.4.10. *Let μ and ν be two minimal generators. There exist two disjoint and simple minimal generators γ and σ such that the set of edges of H_c^* crossed by $\gamma \cup \sigma$ is the same as the set of edges crossed by $\mu \cup \nu$.*

PROOF. As remarked above Lemma 5.4.3, we can assume that both μ and ν are simple. Suppose that they intersect. If necessary, we can modify μ and ν inside the interior of each face and edge so that they cross only inside faces and only transversely. Since μ is separating, it must intersect ν in at least two points. These two points, say x and y, cut each generator in two pieces, say $\mu = a \cdot b$ and $\nu = c \cdot d$. By Lemma 5.2.2 we may assume without loss of generality, that a is homotopic to c and that $a \cdot d$ is a generator (see Figure 5.8). By minimality of μ and ν, the paths a and c must have the same crossing weight and similarly for b and d. We can thus replace μ and ν by the minimal generators obtained by slightly perturbing the concatenations $a \cdot d$ and $c \cdot b$ so as to remove the two intersections at x and y. We claim that $a \cdot d$ and $c \cdot b$ have fewer intersections than μ and ν. Indeed, denoting by $\mathring{e}$ the relative interior of a path e, we have

$$|\mu \cap \nu| = |\mathring{a} \cap \mathring{c}| + |\mathring{a} \cap \mathring{d}| + |\mathring{b} \cap \mathring{c}| + |\mathring{b} \cap \mathring{d}| + 2$$

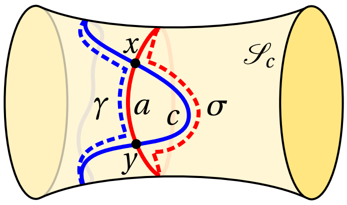


Figure 5.8: The intersecting generators μ and ν (thick plain line) may be replaced by the disjoint generators γ and σ (thick dashed line).

while

$$|a \cdot d \cap c \cdot b| = |\mathring{a} \cap \mathring{c}| + |\mathring{a} \cap \mathring{b}| + |\mathring{d} \cap \mathring{c}| + |\mathring{d} \cap \mathring{b}| = |\mathring{a} \cap \mathring{c}| + |\mathring{d} \cap \mathring{b}|,$$

since, by simplicity of μ and ν: $|\mathring{a} \cap \mathring{b}| = |\mathring{d} \cap \mathring{c}| = 0$. The new minimal generators $a \cdot d$ and $c \cdot b$ obviously cross the same edges as μ and ν. They may now self-intersect, but as initially noted, we can remove contractible loops until they become simple. We conclude the proof with a simple recursion on $|\mu \cap \nu|$. □

We now consider two disjoint and simple minimal generators γ and σ. They bound an annulus $\mathscr{A}$ in $\mathscr{S}_c$. Since γ and σ are crossed by c-transversal curves only, a line ℓ of $\tilde{\mathscr{S}}$ whose projection $\varphi_c(\ell)$ intersects $\mathscr{A}$ is either τ-transversal or τ-invariant. Indeed, if ℓ is not τ-transversal, $\varphi_c(\ell)$ must stay in the finite subgraph of H_c^* interior to $\mathscr{A}$; it follows that $\varphi_c(\ell)$ uses some edge twice, which can only happen if ℓ is τ-invariant by Proposition 5.3.3. In this latter case, $\varphi_c(\ell)$ is a simple generator and by Lemma 5.4.8, there is only one such curve in $\mathscr{S}_c$. Note that this curve is crossed once by every c-transversal, hence is composed of $|\gamma|$ edges. We first bound the complexity of $\mathscr{A}$.

Lemma 5.4.11. *Let V_I, E_I and F be the respective numbers of vertices, edges and faces of H_c^* intersected by $\mathscr{A}$. Then $V_I \le |\gamma|$, $E_I \le 3|\gamma|$ and $F \le 2|\gamma|$.*

Proof. If $\mathscr{A}$ contains the projection $\varphi_c(\ell)$ of a τ-invariant line ℓ, we consider two minimal generators λ and ρ running parallel to $\varphi_c(\ell)$ and respectively to the left and to the right of $\varphi_c(\ell)$. Assuming that γ is to the left of $\varphi_c(\ell)$, we can bound the complexity of H_c^* in $\mathscr{A}$ by the complexity of $\varphi_c(\ell)$ plus the complexity of H_c^* inside the two annuli bounded by γ and λ, and by ρ and σ respectively.

We can now assume that $\mathscr{A}$ is crossed by c-transversals only. These c-transversals form an arrangement of $|\gamma|$ curves in $\mathscr{A}$ where two c-transversals can cross at most twice by Lemma 5.4.9. Together with γ and σ, this arrangement defines a subdivision of $\mathscr{A}$ whose number of boundary vertices is $2|\gamma|$. We distinguish two cases according to whether c-transversals pairwise intersect at most once or twice.

Case where each pair of c-transversals intersects at most once. If the c-transversals are pairwise disjoint in $\mathscr{A}$, the lemma is trivial: $\mathscr{A}$ contains no vertex and intersects $|\gamma|$ edges and as many faces. Otherwise, we first establish a correspondence between faces and interior vertices.

Consider an interior vertex x of the subdivision; it is the intersection of two c-transversals u and v (refer to Figure 5.9.A). Call $t_\gamma(x)$ the triangle formed by u, v and γ. No

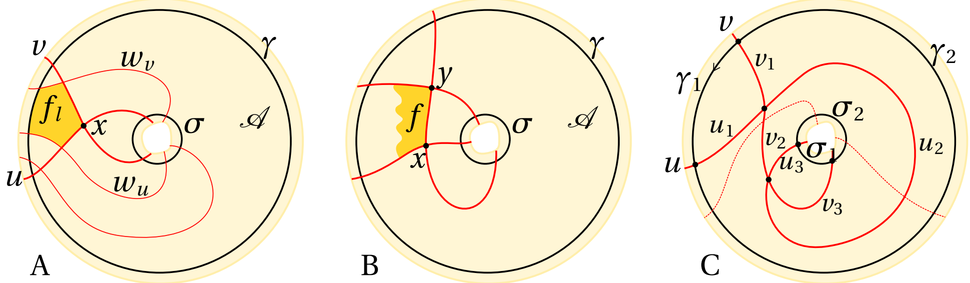


Figure 5.9: (A) f_l is the left face of x in the annulus $\mathscr{A}$. (B) The three supporting line of f passing through x or y. (C) The lines u and v cut $\mathscr{A}$ into simply connected faces.

c-transversal can join u and v inside $t_\gamma(x)$. Otherwise u, v and this c-transversal would form a triangle that would lift into a triangle in $\tilde{\mathscr{S}}$, in contradiction with the triangle-free Property. Moreover, a c-transversal w that crosses the u-side of $t_\gamma(x)$ cannot be crossed inside $t_\gamma(x)$ by any w', as w, w', u and v would then form a quadrilateral. If the u-side of $t_\gamma(x)$ is indeed crossed, we let w_u be the crossing curve closer to x along the u-side. We define w_v similarly for the v-side of $t_\gamma(x)$. The c-transversal curves w_u and w_v, if any, together with u, v and γ bound a face of the subdivision of $\mathscr{A}$. This is the only face incident to x in $t_\gamma(x)$. We call it the *left face* of x; it has one side along γ and no side along σ. Conversely, we claim that every face f of the subdivision of $\mathscr{A}$ with no side along σ is the left-face of a unique interior vertex. The unicity is clear since a left face has a unique side on γ. In fact, by minimality, γ crosses every face of H_c^* at most once; so that f has at most two (consecutive) vertices on γ. If f has a unique interior vertex on its boundary, then f is precisely the triangle t_γ for that vertex. Otherwise, consider two interior vertices x and y that are consecutive along the boundary of f as on Figure 5.9.B. By considering the arrangement made by the three c-transversals defining x and y, it is easily seen that f is included in exactly one of $t_\gamma(x)$ or $t_\gamma(y)$, thus proving the claim. We define *right faces* analogously and remark that a face that is neither a left nor a right face must have one side along γ and one side along σ.

We can now determine the complexity of the subdivision of $\mathscr{A}$. Denote by V and E its respective numbers of vertices and edges. With the notations in the lemma, we have $V = V_I + 2|\gamma|$ and $E = E_I + 2|\gamma|$. By the preceding remark, every face has a side on either γ or σ. It ensues that $F \le 2|\gamma|$. Euler's formula then implies $0 = V - E + F = V_I - E_I + F$, whence $E_I \le V_I + 2|\gamma|$. Since interior and boundary vertices have respective degree 4 and 3, we get $E_I = 2V_I + |\gamma|$ by double-counting of the vertex-edge incidences. Combining with the previous inequality we obtain $V_I \le |\gamma|$, and finally conclude that $E_I \le 3|\gamma|$.

Case where at least two c-transversals intersect twice. We now suppose that two c-transversals u and v intersect twice in $\mathscr{A}$. The curves γ, σ, u and v induce a subdivision of $\mathscr{A}$ where u and v are each cut into three pieces, say u_1, u_2 and u_3 for u and v_1, v_2 and v_3 for v and the two generators are each cut into two pieces, say γ_1, γ_2 for γ and σ_1, σ_2 for σ (see Figure 5.9.C). A c-transversal w crossing γ_1 would have to cross u_1 or v_1 to enter the face bounded by $u_1 \cdot v_2 \cdot u_2^{-1} \cdot v_1^{-1} \cdot \gamma_2^{-1}$ in the subdivision induced by γ, u and v. It is easily seen that no matter how this face is crossed, it will be cut into subfaces, one of which must have three or four sides bounded by u, v and w. This contradicts

Proposition 5.3.3. Likewise no c-transversal can cross σ_1. Hence, any c-transversal distinct from u and v must extend between γ_2 and σ_2 and cut either u_2 or v_2. It is easily seen that any two such c-transversals cannot cross without creating a triangle or a quadrilateral bounded by c-transversals, which is again impossible. It follows that apart from u and v all c-transversals are pairwise disjoint inside $\mathscr{A}$. We deduce that all faces have one side on γ or σ (but not both), whence $F = 2|\gamma|$. Since any c-transversal distinct from u and v is cut into two pieces we also get $V_I = |\gamma|$ and $E_I = 3|\gamma|$. We note that the left triangle $\gamma_1 \cdot u_1 \cdot v_1^{-1}$ is not cut by any c-transversal. $\square$

The short edges. An edge of H_c^* whose relative interior is crossed by a minimal generator is said **short**.

Lemma 5.4.12. *If $\mathscr{A}$ is crossed by c-transversal curves only, then every edge of H_c^* in $\mathscr{A}$ is short.*

Proof. If $\mathscr{A}$ contains no vertex of H_c^* then $\mathscr{A}$ crosses the same edges as γ and σ and the lemma is trivial. Otherwise, we show how to sweep the entire arrangement inside $\mathscr{A}$ with a minimal generator from the left boundary γ to the right boundary σ. We claim that the subdivision of $\mathscr{A}$ induced by γ, σ and H_c^* contains a triangle face t with one side along γ (a left triangle in the above terminology). This is clear in the case where two c-transversals cross twice as noted at the end of the above proof. Otherwise, we consider two c-transversals crossing in $\mathscr{A}$; they obviously form a left triangle with γ. Adding the other c-transversals one by one we see that this left triangle is either not cut or that a new left triangle is cut out of it. We conclude the claim by a simple induction on $\left|\gamma\right|$. The left triangle t has one vertex x interior to $\mathscr{A}$ and incident to four edges u_1, v_1, u_2, v_2 where u_1, v_1 bound t. We can now sweep x with γ by crossing v_2, u_2 instead of u_1, v_1 to obtain a new minimal generator. This new generator bounds with σ a new annulus $\mathscr{A}' \subset \mathscr{A}$. Note that $\mathscr{A}'$ crosses the same set of edges as $\mathscr{A}$ except for u_1, v_1 that were crossed by γ. Moreover, the number of interior vertices is one less in $\mathscr{A}'$ than in $\mathscr{A}$. We conclude the proof of the lemma with a simple recursion on this number. $\square$

Lemma 5.4.13. *The set of short edges of a c-transversal u is a finite segment of u.*

Proof. Let a and b be two short edges along u. Let γ and σ be minimal generators crossing a and b respectively. By Lemma 5.4.10, we can assume that γ and σ are disjoint and still cross a and b (a single generator cannot cross a c-transversal twice). If the annulus bounded by γ and σ is crossed by c-transversals only then we can apply Lemma 5.4.12 to conclude that the edges between a and b along u are short. Otherwise, we can cut this annulus into two parts as in the proof of Lemma 5.4.11 and reach the same conclusion. This latter lemma also implies that there are $O(|\gamma|)$ short edges between a and b. It follows that set of short edges along u is connected and bounded. $\square$

Consider two minimal generators μ and ν. We claim that there exists a minimal generator that crosses the rightmost of the short edges crossed by μ and ν along each c-transversal. Indeed, the two disjoint minimal generators γ and σ returned by Lemma 5.4.10 cannot invert their order of crossings along c-transversals. Hence one of them uses all the leftmost short edges, while the other uses all the rightmost short edges. By a simple induction on the number $|\gamma|$ of c-transversals, this implies in turn that there exists a minimal generator γ_R that crosses the rightmost short edge of each c-transversal. We define the **canonical generator** with respect to c as the cycle in H_c dual to the sequence of short edges crossed by γ_R.

The canonical belt As for γ_R, we can show the existence of a minimal generator γ_L crossing the leftmost short edges. We define the **canonical belt** $\mathscr{B}_c$ as the union of the vertices, edges and faces crossed by the annulus bounded by γ_L and γ_R. By Lemma 5.4.12, the edges in $\mathscr{B}_c$ are the short edges and the edges of the projection of the τ-invariant line, if any. In particular, any minimal generator is included in the canonical belt.

Lemma 5.4.14. *Any intersection of c-transversals is interior to the canonical belt.*

Proof. Consider two c-transversals u and v crossing at a vertex x. Let γ be a minimal generator; it bounds, together with u and v, a triangle t in $\mathscr{S}_c$. Let λ be the generator obtained from γ by substitution of the γ-side of t with the complementary part of the boundary of t, slightly pushed out of t. By the triangle-free Property, the projection of a line of $\tilde{\mathscr{S}}$ that crosses the u-side or v-side of t must exit t through its γ-side. It follows that λ has the same crossing weight as γ. Hence, λ is minimal and the vertex x, which is interior to the annulus bounded by γ and λ, is also interior to the canonical belt. □

We consider the subgraph $\tilde{K}^*$ of $\tilde{H}^*$ induced by the lines in $\tilde{H}^*$ that are neither τ-transversal nor τ-invariant. The projection $\varphi_c(\tilde{K}^*)$ of $\tilde{K}^*$ in $\mathscr{S}_c$ is denoted by K_c^*; it is the union of the projections of the lines that are neither τ-transversal nor τ-invariant. The following lemma gives a simple characterisation of the canonical belt.

Lemma 5.4.15. *$\mathscr{B}_c$ is the unique component of $\mathscr{S}_c \setminus K_c^*$ that contains a generator.*

Proof. It is easily seen from its definition and from the previous lemma that $\mathscr{B}_c$ is a component of $\mathscr{S}_c \setminus K_c^*$. Let σ be any generator contained in $\mathscr{S}_c \setminus K_c^*$ and let γ be a (simple) minimal generator. Since γ is included in $\mathscr{B}_c$, we just need to show that σ and γ are contained in the same component of $\mathscr{S}_c \setminus K_c^*$. If σ is not a simple curve, we first remove its contractible loops to make it simple. If σ and γ intersect then there is nothing to show. Otherwise, σ and γ bound a compact annulus $\mathscr{A}$ in $\mathscr{S}_c$. As already observed, any line whose projection stays in $\mathscr{A}$ must be τ-invariant. It follows that $\mathscr{A}$ is entirely contained in a single component of $\mathscr{S}_c \setminus K_c^*$. □

5.4.3 Computing the canonical generator

We now explain how to compute the canonical generator associated with the loop c in time proportional to $|c|$. Let Π be the relevant region of the loop c^6 obtained by six concatenations of c. According to Section 5.3.2, we can build the adjacency graph Γ of the faces of Π in $O(|c|)$ time. The edges dual to the edges of Γ are the edges of $\tilde{H}^*$

interior to Π. They induce a subgraph of $\tilde{H}^*$ which we denote by Γ^*. As opposed to Γ, the graph Γ^* may have multiple components (see Figure 5.3). A vertex of Γ^* may have degree one or four depending on whether it is a flat boundary vertex or an interior vertex of Π. As explained in the mirror sub-procedure of Section 5.3.2, if $\tilde{e}^* \in \Gamma^*$ is an edge dual to $\tilde{e}$ then the two siblings of $\tilde{e}$ correspond to the facing edges of $\tilde{e}^*$ and they can be computed in constant time (if they belong to Γ^*). Similarly, we can easily compute in constant time the circular list of (one or four) edges sharing a same vertex of Γ^*.

Identifying and classifying lines in Γ^*. From the preceding discussion we can traverse Γ^* to give a distinct tag to each maximal component of facing edges in constant time per edge. Lemma 5.3.5 ensures that each such component is supported by a distinct line. We denote by $\ell(\tilde{e})$ the identifying tag of the line supporting the edge dual to $\tilde{e}$. With a little abuse of notation we will identify a line with its tag.

Let $c_1,\dots,c_6$ be the successive lifts of c in the lift of c^6 and let $x_0,x_1,\dots,x_6$ be the successive lifts of $c(0)$ in the lift of c^6. Let $\tilde{e}_{i,j}$ be the j-th edge of c_i. Since by construction $\tau(c_i)=c_{i+1}$, we have $\tau(\ell(\tilde{e}_{i,j}))=\ell(\tilde{e}_{i+1,j})$. This allows us to compute the translate of any line crossing one of $c_1,\dots,c_5$ in constant time per line. Notice that the interior of Π is crossed by a τ-invariant line if and only if $\tau(\ell)=\ell$ for some line ℓ crossing one (thus any) of $c_1,\dots,c_5$. We can now fill a table $C[\ell]$ whose Boolean value is true if ℓ intersects $\tau(\ell)$ in Π and false otherwise. This clearly takes $O(|c|)$ time. In order to identify the τ-transversals separating x_i from x_{i+1}, we need the following property.

Lemma 5.4.16. *Two intersecting lines crossing the interior of Π intersect in the interior of Π.*

PROOF. Otherwise there would be a line λ separating Π from the intersection of the two intersecting lines. These two lines would form a triangle with λ, in contradiction with the triangle-free Property. $\square$

We first identify and orient the τ-transversals separating x_2 from x_3. We start by filling a table $P[\ell,i]$ counting the parity of the number of intersections of each line ℓ with c_i for $i\in\{2,3,4\}$. This can be done in $O(|c|)$ time: we initialize all the entries of the table P to 0 and, for each $i\in\{2,3,4\}$ and each edge $\tilde{e}$ of c_i, we invert the current parity of $P[\ell(\tilde{e}),i]$. By Proposition 5.4.1, the τ-transversals separating x_2 from x_3 are exactly those ℓ for which $C[\ell]$ is false, $P[\ell,3]$ is odd, and $P[\ell,2]$ and $P[\ell,4]$ are even. We then orient these transversals from left to right as follows. We traverse in order the oriented edges of the lift c_3. As we traverse an oriented edge $\tilde{e}$ of c_3, we check whether its dual belongs to an identified transversal that was not already oriented. In the affirmative, we use the left-to-right orientation of the dual edge of $\tilde{e}$ — given by the correspondence between the oriented cellular embedding and its dual — to orient this transversal.

We can now identify and orient all the τ-transversals separating x_i from x_{i+1}, for $i\in[0,5]$, by translation of those separating x_2 from x_3. Corollary 5.4.7 ensures the existence of at least one transversal separating x_2 from x_3; we choose one and denote it by ℓ in the sequel. We shall concentrate on the part of Π contained in the closure $\bar{B}_\ell$ of the fundamental domain of $\langle\tau\rangle$ comprised between ℓ and $\tau(\ell)$. We put $\mathscr{C}:=\Pi\cap\bar{B}_\ell$.

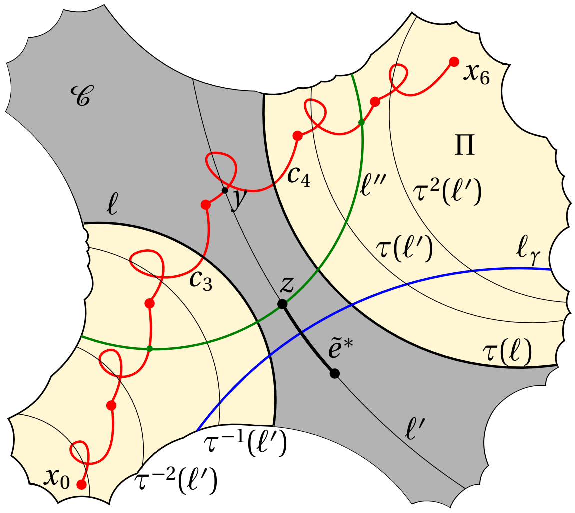


Figure 5.10: The grey region $\mathscr{C}$ is the intersection of the relevant region Π of the lift of c^6 with the fundamental domain bounded by ℓ and $\tau(\ell)$.

Finding a lift of the canonical generator. We want to show that $\mathscr{C}$ contains either a whole lift of the canonical belt or half of it. In this latter case, Π is bounded by a τ-invariant line.

Lemma 5.4.17. *Exactly one of the two following situations occurs*

1. *Π contains the intersection $\bar{B}_\ell \cap \varphi_c^{-1}(\mathscr{B}_c)$.*

2. *There exists a τ-invariant line λ, whose projection $\varphi_c(\lambda)$ cuts the canonical belt into two open parts $\mathscr{B}_L$ and $\mathscr{B}_R$, each one containing a generator and intersecting exactly one edge of each c-transversal. The relevant region Π contains either (i) the intersection $\bar{B}_\ell \cap \varphi_c^{-1}(\mathscr{B}_L)$ or (ii) the intersection $\bar{B}_\ell \cap \varphi_c^{-1}(\mathscr{B}_R)$ and exclude the other one.*

Proof. We first consider the case where there is no τ-invariant line. Let $\tilde{e}^*$ be an edge of $\tilde{H}^* \cap \bar{B}_\ell$ projecting to an edge e^* of the canonical belt. We want to prove that $\tilde{e}^*$ is interior to Π. Denote by ℓ' the supporting line of $\tilde{e}^*$; this is a τ-transversal by Proposition 5.4.6. On the one hand, if ℓ' does not cut any other transversal then it stays inside $\bar{B}_\ell$ and must separate x_3 from either x_2 or x_4. In particular, ℓ' intersects the concatenation $c_3 \cdot c_4$, hence Π. If one of the c_i's intersect $\tilde{e}^*$, then $\tilde{e}^*$ is inside Π by definition. Otherwise, we consider a minimal generator γ crossing e^* and its reciprocal image ℓ_γ. Let $y \in \ell' \cap (c_3 \cdot c_4)$ and let z be the endpoint of $\tilde{e}^*$ between y and the point $\ell_\gamma \cap \tilde{e}^*$. See Figure 5.10. The line ℓ'' crossing ℓ' at z does not cut ℓ_γ by Proposition 5.4.6. Lemma 5.4.9 implies that ℓ'' does not cut $\tau^{-2}(\ell')$ nor $\tau^3(\ell')$. It ensues that ℓ'' intersects the lift of c^6 between $\tau^{-2}(y)$ and $\tau^2(y)$. Since ℓ' and ℓ'' both cross the interior of Π, their intersection z is interior to Π by Lemma 5.4.16. It ensues that $\tilde{e}^*$ is also interior to Π. On the other hand, if ℓ' indeed crosses some other transversal, then by Lemma 5.4.13 one of the endpoints of $\tilde{e}^*$ is the intersection of ℓ' with some other τ-transversal μ. Lemma 5.4.9 implies that these two transversals cannot intersect $\tau^{-2}(\ell)$ nor $\tau^3(\ell)$. It ensues that ℓ' and μ cross the lift of c^6 (this is where we need six lifts of c) and we conclude as above that $\tilde{e}^*$ is interior to Π.

We now consider the case where there exists a τ-invariant line λ. As observed in the proof of Lemma 5.4.8, any τ-transversal must cross λ. The triangle-free Property then forbids these transversals to intersect. It follows that every τ-transversal contains exactly two short edges separated by a vertex of λ. Said differently, the edges of the canonical belt precisely lift to the edges incident to λ (including the edges of λ). On the one hand, if λ does not cross any of the c_i's, then the lift of c^6 lies on one side of λ, say its left. The line λ may play the role of ℓ_γ and we may argue as above that all short edges to the left of λ and contained in $\bar{B}_\ell$ are interior to Π. Equivalently, every lift in $\bar{B}_\ell$ of an edge of $\mathscr{B}_L$ is interior to Π. On the other hand, if c crosses $\varphi_c(\lambda)$, then λ crosses the interior of Π. Lemma 5.4.16 ensures that all its intersections with the τ-transversals in $\bar{B}_\ell$ are interior to Π. We conclude this time that Π contains $\bar{B}_\ell \cap \varphi_c^{-1}(\mathscr{B}_c)$. □

We now explain how to identify the lift of the canonical belt, or half of it, contained in $\mathscr{C}$.

Lemma 5.4.18. *Let Σ^* be the subgraph of $\Gamma^* \cap \mathscr{C}$ projecting to the canonical belt. We can identify the edges of Σ^* in $O(|c|)$ time.*

PROOF. From the preceding lemma, $\varphi_c(\mathscr{C}) \cap \mathscr{B}_c$ is connected and contains a generator. Recall that $\tilde{K}^*$ is the union of the lines that are neither τ-transversal nor τ-invariant, and that K_c^* is its projection into $\mathscr{S}_c$. Lemma 5.4.15 ensures that $\varphi_c(\mathscr{C}) \cap \mathscr{B}_c$ is the only component of $\varphi_c(\mathscr{C}) \setminus K_c^*$ that contains a generator. Equivalently, $\mathscr{C} \cap \varphi_c^{-1}(\mathscr{B}_c)$ is the only component of $\mathscr{C} \setminus \tilde{K}^*$ that contains an edge $\tilde{e}^* \in \ell$ together with its translate $\tau(\tilde{e}^*)$.

Thanks to Lemma 5.4.9 and following the paragraph on line identification and classification, we can correctly detect all the τ-transversal and τ-invariant lines crossing $\mathscr{C}$. By complementarity, we identify the edges of $\tilde{K}^*$ in $\mathscr{C}$. We also identify by a simple traversal the subgraph $\Gamma_{\mathscr{C}}$ of Γ whose dual edges are contained in $\mathscr{C}$. The graph $\Gamma_{\mathscr{C}}$ is connected ($\mathscr{C}$ is a "convex" region of the plane) and each component of $\mathscr{C} \setminus \tilde{K}^*$ is a union of faces corresponding to a connected component of $\Gamma_{\mathscr{C}} \setminus \tilde{K}$, where $\tilde{K}$ is the set of primal edges corresponding to $\tilde{K}^*$. We eventually select the component Σ of $\Gamma_{\mathscr{C}} \setminus K$ that includes an edge $\tilde{e}$ together with its translate $\tau(\tilde{e})$. It clearly takes time proportional to $|c|$ to select the edges of Σ. We finally remark from the initial discussion that the dual of the edges in Σ are the edges of Σ^*. □

Proposition 5.4.19. *We can compute the canonical generator in $O(|c|)$ time.*

PROOF. We first compute Σ^* as in Lemma 5.4.18. We then determine if we are in situation 2*(i)* of Lemma 5.4.17. To this end, we check that all the edges of Σ^* are supported by pairwise distinct τ-transversals. If this is the case, we further check if the right extremities of the edges of Σ^* are linked by edges of a (necessarily τ-invariant) line. This is easily seen in constant time per edge by projecting the edges of Σ^* back into H^* on $\mathscr{S}$. If we are indeed in situation 2*(i)*, the canonical generator is composed of the projection on $\mathscr{S}_c$ of the dual of the edges facing the edges of Σ^* to their right (keeping only one of the two edges supported by ℓ and $\tau(\ell)$). In the other situations 1 and 2*(ii)*, the edges crossed by the lift of the canonical generator in $\bar{B}_\ell$ are the edges of Σ^* that are supported by τ-transversal lines and whose right endpoint is not a crossing with any other τ-transversal or τ-invariant line. In other words, these are the rightmost edges in Σ^* of the pieces of τ-transversals crossing Σ^*, unless they abut on ℓ or $\tau(\ell)$. In either case, we can clearly determine the sequence of edges of the canonical generator in $O(|c|)$ time. $\square$

5.4.4 End of the proof of Theorem 5.1.2

Let c and d be two non-contractible cycles represented as closed walks in H. Assuming that $\mathscr{S}$ is orientable with genus at least two, we compute the canonical generators γ_R and δ_R corresponding to c and d respectively. This takes $O(|c|+|d|)$ time according to Proposition 5.4.19. Following the discussion in the Introduction, c and d are freely homotopic if and only if the projections $\pi_c(\gamma_R)$ and $\pi_c(\delta_R)$ in $\mathscr{S}$ are equal as cycles of H. This can be determined, under the obvious constraint that these two projections have the same length, in $O(|c|+|d|)$ time using the classical Knuth-Morris-Pratt algorithm [CLRS09] to check whether $\pi_c(\gamma_R)$ is a substring of the concatenation $\pi_c(\delta_R)\cdot\, \pi_c(\delta_R)$.

It remains to consider the cases of $\mathscr{S}$ being a torus. As noted for the contractibility test, its fundamental group is commutative. Being conjugate is thus equivalent to being equal as group elements and the free homotopy test reduces to the contractibility test.

We have thus solved the free homotopy test for closed orientable surfaces. We finally consider the free homotopy test when $\mathscr{S}$ is non-orientable. A possibly self-crossing cycle c on $\mathscr{S}$ is **two-sided** if a consistent orientation of $\mathscr{S}$ can be propagated all along c. The cycle is otherwise **one-sided**, which can easily be decided in $O(|c|)$ time with the edge signature of the combinatorial map encoding $\mathscr{S}$ (see Definition 2.2.15). Since the square c^2 of c is two-sided, we may assume that the two given cycles c and d are two-sided. Indeed, for non-orientable surfaces of genus ≥ 3, two one-sided cycles are (freely) homotopic if and only if their square are (freely) homotopic[3]. We can thus decide is c and d are homotopic by testing for the (closed) lifts of c^2 and d^2 in the orientation covering of $\mathscr{S}$ (see Exercise 2.2.38). This covering being orientable, we are brought back to the orientable case.

We finally note that if $\mathscr{S}$ is a projective plane, its fundamental group is again commutative and the test is trivial. If $\mathscr{S}$ is a Klein bottle, the test was already resolved by Max Dehn [Sti87, p.153] in linear time.

[3]For a quick argument, the automorphisms of the Poincaré disk $\tilde{\mathscr{S}}$ induced by a loop and its square have the same axis. If the square of two orientation reversing automorphisms are equal then their restriction to their common axis are equal, implying that the automorphisms are themselves equal.

Part II

Isometric Embedding of the Square Flat Torus

Contents

In this part, I give an informal overview of the main ideas that lead to the computation of an isometric embedding of the square flat torus. The result of this computation first appeared as a short announcement [BJLT12] in the Proceedings of the National Academy of Sciences (PNAS). The present text lies somewhere in-between the presentation for the general public of *Pour la Science* [BLT13] and the complete description including proofs and implementation details that appeared in Ensaios Matemáticos [BJLT13]. During more than six years this project gathered Vincent Borrelli, Boris Thibert, Saïd Jabrane and myself. Vincent and myself were supervising the PhD thesis of Saïd during most of this time. Because the project had such a success and because Mathematics is not my major, it is probably useful for the purpose of an *Habilitation à Diriger des Recherches* to recall the circumstances of the birth and realization of the HEVEA project. It was initiated by Vincent Borrelli, a mathematician from the *Institut Camille Jordan* in Lyon and a specialist of the h-principle and convex integration. Vincent had the intuition that the mathematical tools developed by John Nash and later by Mikhail Gromov could be materialized by computer programs dedicated to the resolution of differential relations. In particular, considering the differential relation for isometries, one could *compute* an isometric embedding of a flat torus in $\mathbb{R}^3$. As a benefit, one would visualize and *thus* understand the structure of the solutions built by convex integration. Those were indeed still mysterious and their degree of differentiability remains an active subject of research. The HEVEA project required the skills of colleagues able to understand the details of a hard theory while keeping in mind the constraints of computer calculations. Boris Thibert had the same thesis advisor as Vincent and is more concerned with numerical and algorithmic aspects of geometry. He was thus an ideal colleague for this project. Although I am personally more inclined towards algorithms and combinatorial structures, I have always been interested in various aspects of geometry. As a Master (at the time DEA) student, I indeed took a course on differential geometry by Marc Chaperon, a distinguished French geometer. Much later, during a seminar in Paris, Vincent noticed my interested questions and a couple years after we met again in Grenoble. We then decided to form a group with Boris who had just settled in Grenoble. Thanks to some funding of the region Rhône-Alpes, we have been able to finance the thesis of Saïd, a former teacher willing to study mathematics again. We were now a group of four and during a first period we learned, thanks to the extraordinary teaching ability of Vincent, the necessary chunk of theory to start the project. From then, the HEVEA team merged into a single and multidisciplinary body with four heads, each one contributing to every aspect of the project: theorems, proofs, implementation, analysis of the results. . . The exchanges and rearrangements were so intensive that it is impossible to attribute a specific part to anyone of us. The papers were written with four hands, each one rewriting the version of its predecessor. Eventually, the discovery of smooth fractals was the result of a progressive and multidisciplinary work, in the best sense. This way of working in a spirit of sharing for a non-profit Science is one the strong values and claims of the HEVEA project.

1 Isometric Embeddings

We recall that a **Riemannian manifold** $(\mathscr{M}, g)$ is a differentiable manifold $\mathscr{M}$ equipped with an inner product g_p over the tangent space of each point $p \in \mathscr{M}$. The inner product g_p is required to vary continuously from point to point. In other words, if X, Y are tangent vector fields on $\mathscr{M}$, the map $p \mapsto g_p(X(p), Y(p))$ must be continuous. We call g a **Riemannian metric** as it allows to measure the length of any differentiable path $\gamma : [a, b] \to \mathscr{M}$ by the formula

$$L(\gamma) = \int_a^b \sqrt{g_{\gamma(t)}(\gamma'(t), \gamma'(t))}\mathrm{d}t$$

If $(\mathscr{N}, h)$ is a Riemannian manifold and $f : \mathscr{M} \to \mathscr{N}$ is a differentiable map, we can define the inner product of two vectors u, v tangent at a point $p \in \mathscr{M}$ by

$$h_{f(p)}(df(p).u, df(p).v)$$

We obtain this way a Riemannian metric on $\mathscr{M}$ called the **pullback** of h by f and denoted by f^*h. See Figure 6.11. A differentiable map $f : (\mathscr{M}, g) \to (\mathscr{N}, h)$ between two Riemannian manifolds is an **isometry** (not necessarily one-to-one or onto) if the length of every path $\gamma : [a, b] \to \mathscr{M}$ is the same as the length of the image path $f \circ \gamma$. Equivalently, f is an isometry if

$$g = f^*h,$$

i.e., g is equal to the pullback of h. We shall consider isometric **immersions** or **embeddings** (also written *imbeddings*). In the first case the differential of the map f is required to be injective. In the second case, f is also required to be an homeomorphism between $\mathscr{M}$ and its image $f(\mathscr{M})$.

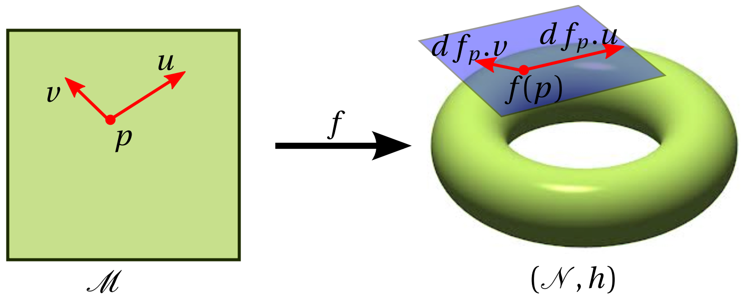


Figure 6.11: We can define the inner product of two tangent vectors u and v by transforming them via f into vectors tangent to $\mathcal{N}$.

The idea of such abstract metric spaces was introduced by Bernhard Riemann in 1854 in his famous lecture for his habilitation at the University of Göttingen [Rie67, Kap05]. This intrinsic point of view came right after Carl Friedrich Gauss had developped the extrinsic study of differential geometry. Gauss was considering surfaces in the Euclidean space $\mathbb{E}^3$ so that the geometry of a surface would inherit from the Euclidean geometry: the length of a path is obtained by tightening the path (made of an inelastic rope) to a line and measuring it with an Euclidean ruler. The *non-Euclidean* geometry of Riemann deliberately ignored the axiomatic geometry of Euclid and paved the way to the study of hyperbolic geometry, of four and higher dimensional spaces and of the general theory of relativity. It was supposed to be much more general that the classical Euclidean geometry. But was it really the case? In other words, given a Riemannian manifold $(\mathcal{M}, g)$ is it always possible or not to get an isometric image of $(\mathcal{M}, g)$ into some Euclidean space? And even so, what is the minimal dimension of the Euclidean space where $(\mathcal{M}, g)$ can be isometrically embedded? Those questions have haunted mathematicians for most of the twentieth century. Among the milestones of the history of isometric embeddings we may cite the Janet-Cartan theorem (1926-1927) stating that a real analytic Riemannian manifold of dimension n can be *locally* isometrically embedded (analytically) in the Euclidean space $\mathbb{E}^{s_n}$ with $s_n = n(n+1)/2$. This local result was followed by the global but purely topological embedding theorem of Whitney (1936): every C^∞ manifold of dimension n embeds into $\mathbb{R}^{2n}$. The existence of global isometric embeddings was eventually proved by Nash [Nas54] in 1954: every compact n-dimensional C^1 manifold $(\mathcal{M}, g)$ that topologically embeds into $\mathbb{R}^k$, with $k \geq n+2$, isometrically embeds into $\mathbb{E}^k$. Figure 6.12 reproduces the first half page of the celebrated Nash's paper on C^1 isometric embeddings. Kuiper improved Nash's theorem the year after to replace the condition $k \geq n+2$ by $k \geq n+1$. The now called Nash-Kuiper's theorem actually states that isometric embeddings are dense (although this was formally recognized only later by M. Gromov) in the following sense. For any *(strictly) short* embedding $f_0 : (\mathcal{M}, g) \to \mathbb{E}^k$, i.e. such that $f_0^*\langle\cdot,\cdot\rangle_{\mathbb{E}^k} < g$, and for any $\varepsilon > 0$ there exists an isometric embedding f with

$$\|f - f_0\|_{C^0} < \varepsilon$$

In particular, the Euclidean sphere of radius one can be isometrically embedded into a three dimensional Euclidean ball of any strictly positive radius! (See Figure 6.13.) This seems to contradict the analytic [CV36] or smooth [AP50, Nir53, Pog64] version of Cauchy's rigidity theorem, claiming that surfaces of strictly positive Gaussian curva-

ANNALS OF MATHEMATICS
Vol. 60, No. 3, Nnvember, 1954
Printed in U.S.A.

C^1 ISOMETRIC IMBEDDINGS

BY JOHN NASH

(Received February 26, 1954)
(Revised June 21, 1954)

Introduction

The question of whether or not in general a Riemannian manifold can be isometrically imbedded in Euclidean space has been open for some time. The local problem was discussed by Schlaefli [1] in 1873 and treated by Janet [2] and Cartan [3] in 1926 and 1927.

This question comes up in connection with the alternative extrinsic and intrinsic approaches to differential geometry. The historically older extrinsic attitude sees a manifold as imbedded in Euclidean space and its metric as derived from the metric of the surrounding space. The metric is considered to be given abstractly from the intrinsic viewpoint.

This intrinsic approach has seemed the more general, so long as there was no contravening evidence. Now it develops that the two attitudes are equally general, and any (positive) metric on a manifold can be realized by an appropriate imbedding in Euclidean space.

Figure 6.12: The hilighted sentences put the work of Nash in the context of comparison between Riemannian and Euclidean geometries. From its discovery of C^1 isometric embeddings Nash concludes that the abstract point of view of Riemannian geometry is no more general than Euclidean geometry.

ture have a unique isometric embedding in $\mathbb{E}^3$. Another baffling consequence is the existence of isometric embedding of flat tori. A **flat torus** is the quotient of the Euclidean plane by some discrete lattice in the plane. The **square flat torus** is denoted by $(\mathbb{T}^2, \langle\cdot,\cdot\rangle_{\mathbb{E}^2})$. It corresponds to the lattice $\mathbb{Z}^2$ and can be seen as the Euclidean unit square, a fundamental domain for the action of $\mathbb{Z}^2$, whose opposite sides have been pointwise identified. Being locally isometric to the Euclidean plane, a flat torus has zero Gaussian curvature everywhere. By Gauss's Theorema Egregium, any isometric embedding in $\mathbb{E}^3$ must have zero curvature as well. But a simple argument shows that any closed surface embedded in $\mathbb{E}^3$ must have strictly positive curvature at some point (see Figure 6.14). The existence of an isometric embedding of a flat torus would thus contradict Gauss's Theorema Egregium! Both paradoxes are resolved by noting that Nash's embeddings are only required to be C^1 while the intrinsic curvature is only meaningful for C^2 surfaces. Two years after its outstanding discovery, Nash announced in 1956 the existence of C^r, $3 \le r \le \infty$, global isometric embedding into $\mathbb{E}^k$ with $k = n(3n+11)/2$ (in the compact case). Nash will extend this result to the analytic case ten years later [Nas66]. Those results where further enhanced by M. Gromov who obtained smaller dimensions for the embedding space. A. Zeghib [Zeg05] provides more details on the long history of isometric immersions.

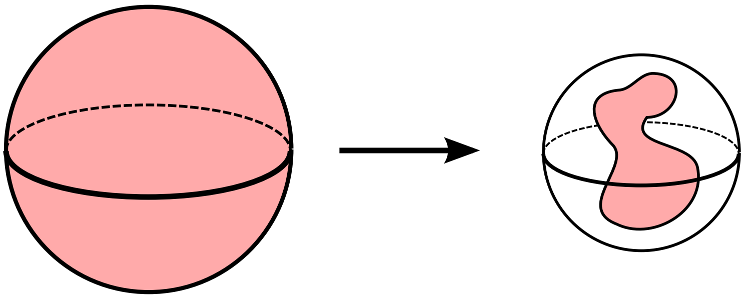


Figure 6.13: The left sphere of radius one can be isometrically embedded into a ball of smaller radius.

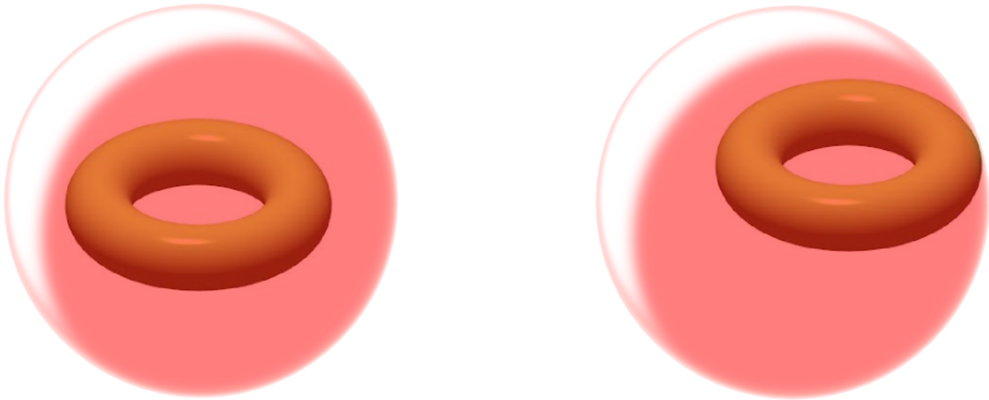


Figure 6.14: To show that a smooth embedded closed surface cannot have zero Gaussian curvature everywhere, one can surround the surface with a big sphere (Left) and move the surface until it touches the surface (Right). At the contact point it can be proved that the surface curvature is larger than the sphere curvature.

2 Differential Relations

The work of Nash concerning C^1 isometric embeddings was reformulated and generalized by Gromov in an extraordinary way. Gromov's point of view is the following. According to the above definition of an isometry, an embedding of the square flat torus $f:(\mathbb{T}^2,\langle\cdot,\cdot\rangle_{\mathbb{E}^2})\to\mathbb{E}^3$ is isometric if $\langle\cdot,\cdot\rangle_{\mathbb{E}^2}=f^*\langle\cdot,\cdot\rangle_{\mathbb{E}^3}$. Using the canonical Cartesian coordinates u,v in the plane, this condition can be written as

$$\langle\frac{\partial f}{\partial u},\frac{\partial f}{\partial u}\rangle_{\mathbb{E}^3}=1,\quad \langle\frac{\partial f}{\partial u},\frac{\partial f}{\partial v}\rangle_{\mathbb{E}^3}=0,\quad \langle\frac{\partial f}{\partial v},\frac{\partial f}{\partial v}\rangle_{\mathbb{E}^3}=1$$

This is an instance of a system of first-order partial differential equations. In general, such a system applies to the set of differentiable maps $f:X\to Y$ for given spaces X and Y. It is described by k equations of the form

$$R_i(x,f(x),f'(x))=0,\qquad 1\le i\le k,$$

involving the parameter $x\in X$, the value $f(x)$ and the derivative $f'(x)$ of the unknown map f (in some local coordinate charts, $f'(x)$ is the sequence of partial derivatives of f). Dealing with abstract manifolds or with partial differential systems of higher order requires to introduce the rather abstract jet-spaces[4]. To keep the presentation simple, we will stick to the first order case and suppose that $X=\mathbb{R}^p$ and $Y=\mathbb{R}^q$. The one-jet space then reduces to a product $J^1(X,Y)=J^1(\mathbb{R}^p,\mathbb{R}^q):=\mathbb{R}^p\times\mathbb{R}^q\times\mathbb{R}^{pq}$. A point $(x,y,z)\in J^1(X,Y)$ corresponds to a parameter $x\in X$, a possible value $y\in Y$ for a map $X\to Y$ and a possible value for its differential, which is a linear map $\mathbb{R}^p\to\mathbb{R}^q$ and can be considered as a point in $\mathbb{R}^{pq}$. The projection on the first component $\pi_X : J^1(X,Y)\to X$ provides the one-jet space with a bundle structure. In our simple framework, this just means that we have a product $J^1(X,Y)=X\times F$ with $F=\mathbb{R}^q\times\mathbb{R}^{pq}$. For $x\in X$, the set $\{x\}\times F$ is called the **fiber** above x. The elements of this fiber are precisely those elements of $J^1(X,Y)$ that projects onto x. A **section** of $(J^1(X,Y),\pi_X)$ is a map $s: X\to J^1(X,Y)$ such that $\pi_X\circ s = Id_X$. In other words, a section of $(J^1(X,Y),\pi_X)$ sends each $x\in X$ to an element of its fiber. In our simple framework, any section $s: X\to J^1(X,Y)$ has the form $s(x)=(x,f(x),g(x))$ and can be seen as the graph of a map f with some possible derivative $g(x)$ attached at the point $(x,f(x))$. *A priori* g is not related to the

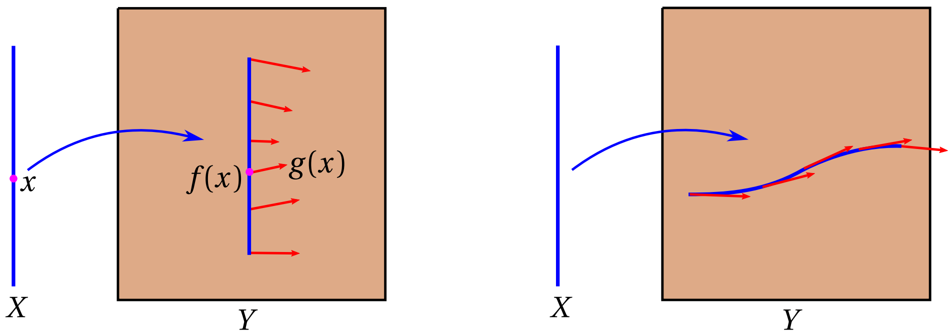


Figure 6.15: Left, A non-holonomic section of $J^1(X,Y)$ with $X=\mathbb{R}$ and $Y=\mathbb{R}^2$, represented as a map $x\mapsto f(x)$ with the vector $g(x)$ attached at $f(x)$. Right, An holonomic section.

derivative of f. The section is said **holonomic** precisely when g is the derivative of f as illustrated Figure 6.15. If we introduce the projection on the two first components $\pi_{XY} : J^1(X,Y)\to X\times Y$ (it also provides $J^1(X,Y)$ with another bundle structure), the composition $\pi_{XY}\circ s$ has the form $x\mapsto(x,f(x))$ for some map $f: X\to Y$. So that the map s is a holonomic section if $s(x)=(x,f(x),f'(x))$.

Turning back to our differential system, each equation $R_i=0$ can be seen as an equation over the one-jet space $J^1(X,Y)$. The system of equations $R_i=0$ thus defines a *subset* of $J^1(X,Y)$. More generally, Gromov considers **differential relations** as subsets of $J^1(X,Y)$, not just those defined by equations. See Figure 6.16. In particular an **open differential relation** is an open subset of $J^1(X,Y)$. Differential relations are the central object of the famous book of Gromov [Gro86]. In the usual spirit of partial differential

[4]Intuitively, the r**-th order jet space** is the set of order r Taylor polynomials of all C^r maps at every point $x\in X$. When X and Y are abstract differentiable manifolds, an invariant notion of Taylor polynomials seems hard to define. However, we can still decide that two functions defined in some neighborhood of a point $x\in X$ have a **contact of order** r at x if they have the same Taylor polynomial of order r in *some* coordinate charts. This contact relation is clearly independent of the choice of coordinate charts. The r-th order jet space is then the set of equivalence classes of contact for all $x\in X$. Using coordinate charts, it can be equipped with a smooth structure. Obviously, functions with a contact of order r have a contact of order $k\le r$. It follows that the r-th order jet space projects naturally onto the k-th order jet space.

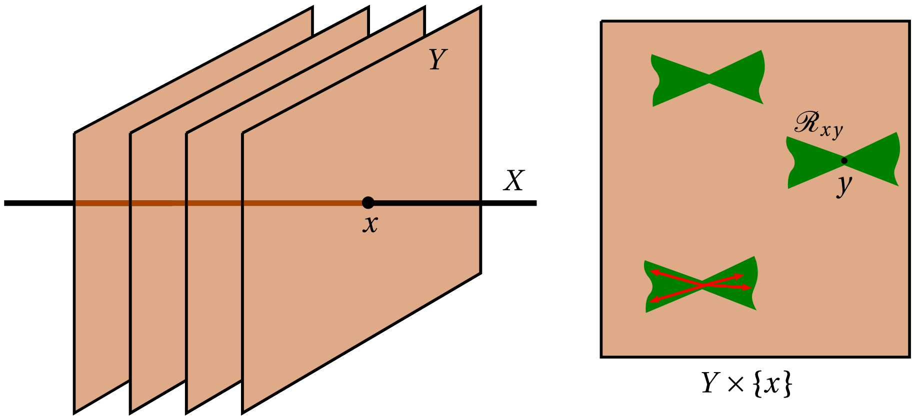


Figure 6.16: There is a natural projection $\pi_{XY} : J^1(X,Y) \to X \times Y$. So, for $p = 1$ and $q = 2$, i.e., $X = \mathbb{R}$ and $Y = \mathbb{R}^2$, $J(X,Y)$ can be viewed as a fiber bundle with base $X \times Y$ and fibers $\mathbb{R}^{pq} = \mathbb{R}^2$. The restriction of a differential relation $\mathscr{R}$ to each fiber gives us a subset $\mathscr{R}_{xy}$ colored in green on the right figure. Each $\mathscr{R}_{xy}$ is a subset of possible derivatives, which are velocity vectors of planar paths $\mathbb{R} \to \mathbb{R}^2$ in the present case. The red arrows emphasize that the green sets are sets of velocity vectors. In general $\mathscr{R}_{xy}$ may vary with x and y. For simplicity, we used the same set at every point in the drawing.

equations, a **solution** of a differential relation $\mathscr{R}$ would be a map $f : X \to Y$ such that $f'(x) \in \mathscr{R}_{xf(x)}$ for all $x \in X$. Here, $\mathscr{R}_{xf(x)}$ is the restriction of $\mathscr{R}$ to the fiber of $(x, f(x))$ in $J^1(X,Y)$. In the point of view of differential relations, we rather view a solution as a holonomic section $s : X \to \mathscr{R}$ with $s(x) = (x, f(x), f'(x))$ and $f'(x) \in \mathscr{R}_{xf(x)}$, where $f = \pi_Y \circ s$. This reformulation leads naturally to a weaker notion of solution: a **formal solution** of $\mathscr{R}$ is a section $s : X \to \mathscr{R}$. Here, it is important to emphasize that s is *not* required to be holonomic. We shall denote by (f, φ) the formal solution (or more generally the section of $J^1(X,Y)$) $x \mapsto (x, f(x), \varphi(x))$. Again, for a formal solution it is not required that $\varphi = f'$. Figure 6.17 illustrates the two notions of solutions.

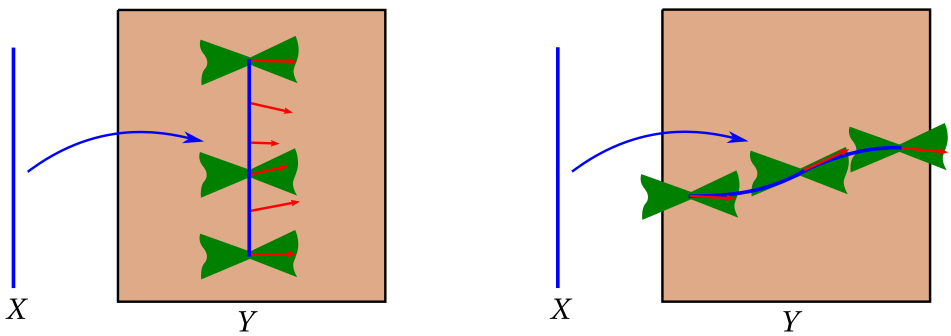


Figure 6.17: Left, A formal solution of some differential relation $\mathscr{R}$ (in green). Only a few $\mathscr{R}_{xy}$ are represented. Right, A solution of the same relation.

3 The h-Principle

Why is it interesting to introduce the formal solutions of a differential relation $\mathscr{R}$? On the one hand, what is really desired is a solution of $\mathscr{R}$, not just a formal solution. On the other hand, it seems much easier to decide if $\mathscr{R}$ has any formal solution than to decide if $\mathscr{R}$ has an actual solution. In fact, the existence of a formal solution relates to Topology rather than to Analysis. To illustrate this claim, let us describe a somehow artificial example. We consider real functions over the sphere $\mathbb{S}^2 \subset \mathbb{E}^3$. With our previous notations, this means $X = \mathbb{S}^2$ and $Y = \mathbb{R}$. The derivative of a function $\mathbb{S}^2 \to \mathbb{R}$ at some $x \in \mathbb{S}^2$ is a linear form on the plane $T_x\mathbb{S}^2$ tangent to $\mathbb{S}^2$ at x. We can identify this linear form with a vector in $T_x\mathbb{S}^2$ (the dual vector with respect to the inner product induced by $\langle \cdot, \cdot \rangle_{\mathbb{E}^3}$ in $T_x\mathbb{S}^2$). This way, a section of $(J^1(X,Y), \pi_X)$ has the form $x \mapsto (s, f(x), v(x))$ where f is a real function over $\mathbb{S}^2$ and v is a vector field tangent to $\mathbb{S}^2$. We now consider the differential relation $\mathscr{R}$ that forbids the differential to cancel. In other words $\mathscr{R}_{xy} = T_x\mathbb{S}^2 \setminus \{0\}$ for all $x \in \mathbb{S}^2$ and $y \in \mathbb{R}$. See Figure 6.18. Are there any formal solution of $\mathscr{R}$?

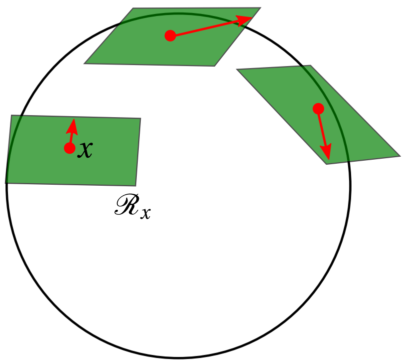


Figure 6.18: The differential relation $\mathscr{R}$ is the collection of pointed tangent planes. Since $\mathscr{R}_{xy}$ does not depend on y, only one copy $\mathscr{R}_x$ of all the $\mathscr{R}_{xy}$ is represented for each x.

The answer is negative and the reason comes from a purely topological argument: if $\mathscr{R}$ had a formal solution $x \mapsto (s, f(x), v(x))$, then v would be a continuous non-vanishing vector field tangent to the sphere. But the famous hairy ball theorem states that this is impossible. Since every solution is also a formal solution, we immediately infer that $\mathscr{R}$ has no solution. (This could also have been inferred directly from true solutions, but using formal solutions isolates the topological part.)

In fact, the relationship between solutions and formal solutions can be made much stronger. According to Gromov, a differential relation $\mathscr{R}$ satisfies the **homotopic principle**, or h-principle for short, when every formal solution of $\mathscr{R}$ can be continuously deformed to a solution. Obviously, one need to give reasonable conditions for $\mathscr{R}$ to satisfy the h-principle, as otherwise this principle would be vacuous. Gromov has developed several approaches to prove that the h-principle holds for certain relations. One of them uses the notion of *ampleness*. A subset of an affine space $\mathscr{A}$ is **ample** if the convex hull of each path connected component of the subset is equal to the whole space $\mathscr{A}$. See Figure 6.19. For maps $\mathbb{R} \to \mathbb{R}^q$, a differential relation $\mathscr{R}$ is said **ample** if $\mathscr{R}_{xy}$ is ample in $\mathbb{R}^q = T_x\mathbb{R}^q$. For maps $\mathbb{R}^p \to \mathbb{R}^q$, $p \geq 1$, we define a coordinate slice of $\mathscr{R}_{xy} \subset \mathbb{R}^{pq} = (\mathbb{R}^q)^p$ as the intersection of $\mathscr{R}_{xy}$ with any q-dimensional affine plane parallel to one of the p factors of $(\mathbb{R}^q)^p$. The relation $\mathscr{R}$ is said **ample in the coordi-**

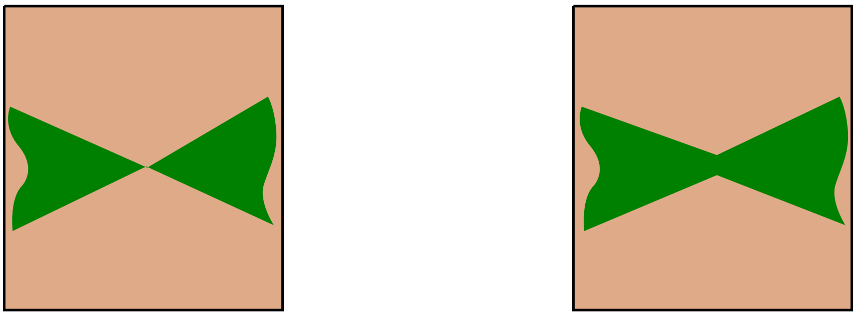


Figure 6.19: Left, the green open (infinite) double wedge is not ample in the plane. Its two components are already convex and do not cover the plane. Right, a thickened version of the left green subset with the shape of an infinite bow tie. This ample set has a single component whose convex hull fills in the whole plane.

nate directions if all its coordinate slices are ample. An **ample differential relation** is required to be ample in the coordinate directions for all change of coordinates [EM02]. Thanks to the method of convex integration, Gromov proves that the h-principle holds for open ample differential relations. In fact Gromov proves a much stronger version:

Theorem 3.1 (Gromov, 69-73)**.** *An open ample differential relation satisfies a C^0 dense and parametric h-principle.*

The C^0 dense property just means that any formal solution $x \mapsto (x, f_0(x), \varphi(x))$ can be deformed to solutions that are as close as desired to f_0: for any $\varepsilon > 0$, there is a solution f such that $\|f - f_0\|_{C^0} := \sup_{x\in X} \|f(x) - f_0(x)\| < \varepsilon$. The parametric property tells that we can actually deform a whole family (over a certain parameter space) of formal solutions toward a family of solutions in a continuous manner. In Figure 6.20, we consider the same open and ample differential relation as on Figure 6.17, starting with a formal solution (f_0, φ) where f_0 is a vertical path. In Figure 6.21, we consider

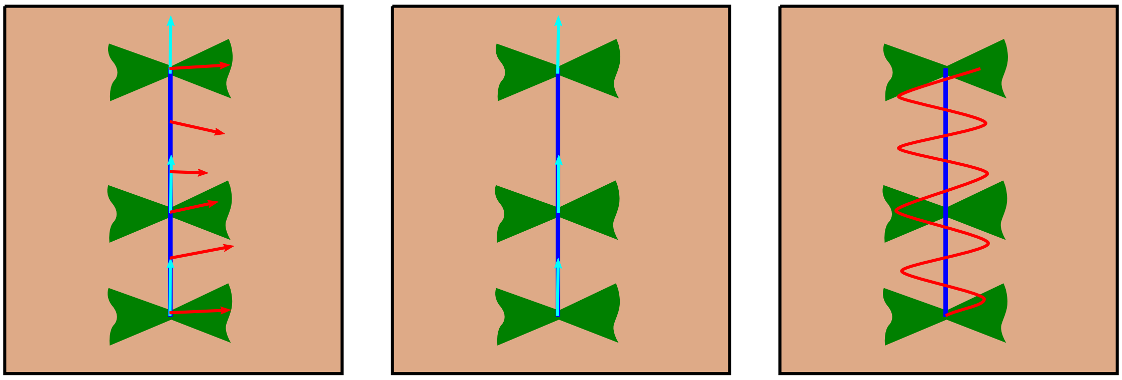


Figure 6.20: Left, A formal solution (f_0, φ) of the uniform differential relation $\mathscr{R}$ corresponding to the right figure 6.19. The (red) arrows represent the field of derivatives φ. Middle, the parametrization of f_0 is such that its derivative leaves the differential relation: f_0 is not a solution of $\mathscr{R}$. Right, By Gromov's theorem there exist solutions of $\mathscr{R}$ arbitrarily close to f_0. Indeed, using more but smaller oscillations, one can get solutions closer and closer to f_0.

the differential relation corresponding to the left figure 6.19. This relation tells us that the angle of the velocity vector of a path with the horizontal direction should belong to some small interval $(-\theta, \theta)$. It is not ample. This time, (f_0, φ) is still a formal solution but cannot be deformed to an arbitrarily close solution. It is interesting to note that

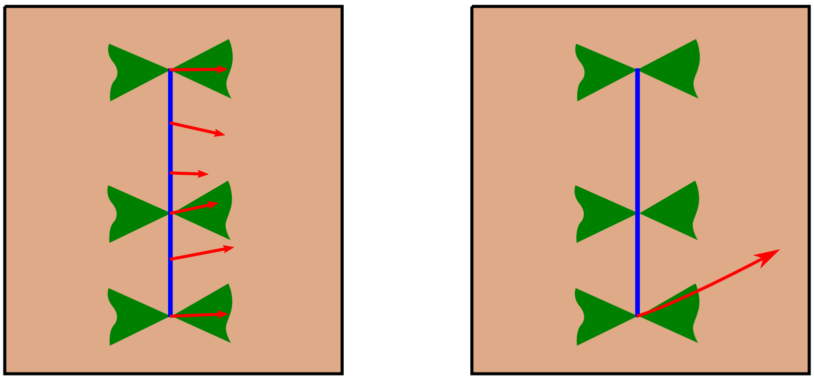


Figure 6.21: Left, A formal solution (f_0, φ) of the uniform differential relation $\mathscr{R}$ corresponding to the left figure 6.19. Right, Because $\mathscr{R}$ is not ample we cannot apply Gromov's theorem. Starting a solution from the lower endpoint and going to the right, the relation $\mathscr{R}$ forbids to ever come back to the right by imposing a small angle with the horizontal. The same would be true inverting right and left. The solution will thus get further away from f_0 as we reach its other endpoint.

in the three dimensional space, the analogous relation saying that the velocity vector of a path with the horizontal direction should be small becomes ample. By Gromov's theorem we have solutions arbitrarily close to the vertical path as on Figure 6.22. This says that we can travel vertically while staying almost horizontal!

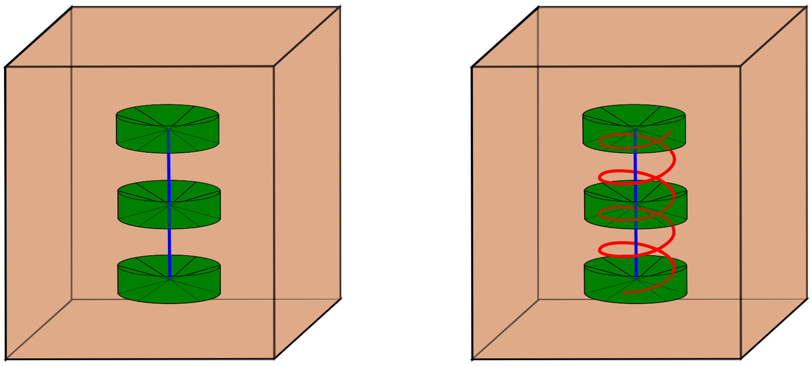


Figure 6.22: Left, the uniform differential relation in three dimension analogous to the one in Figure 6.21. This time, the fibers of the relation are connected and becomes ample. Right, A solution with a helix shape close to the vertical map.

4 One Dimensional Convex Integration

In practice, being ample is a quite strong requirement for differential relations. Hopefully, one can work with weaker conditions assuming extra properties on the initial formal solution (f_0, φ). Let I be the unit interval in $\mathbb{R}$ and consider a differential relation $\mathscr{R} \subset J^1(I, Y)$ for maps $I \to Y$. A formal solution (f_0, φ) is said **short** if for all $x \in I$, the derivative $f_0'(x)$ is a convex combination of some derivatives (depending on x and y) that all belong to the component of $\varphi(x)$ in $\mathscr{R}_{x f_0(x)}$. Formally, we ask that $f_0'(x) \in Conv_{(x, f_0(x), \varphi(x))} \mathscr{R}_{x f_0(x)}$ for all $x \in I$, where $Conv_{(x,y,z)} \mathscr{R}_{xy}$ is the convex hull in the fiber of $(J^1(I, Y), \pi_{XY})$ above (x, y) of the component of $\mathscr{R}_{xy}$ that contains (x, y, z). See Figure 6.23 for an illustration. Suppose that Y is endowed with a Riemannian metric g

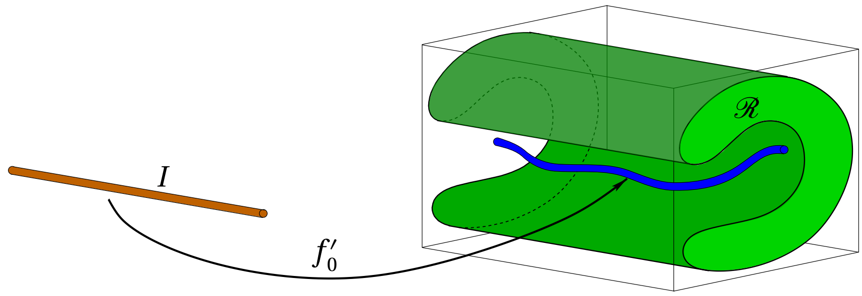


Figure 6.23: A short map $f_0 : I \to Y$ with respect to some differential relation.

and that $\mathscr{R}$ is the relation enforcing maps $(I, \langle \cdot, \cdot \rangle_{\mathbb{E}^1}) \to (Y, g)$ to be isometric. Then $\mathscr{R}_{xy}$ is the unit sphere in $T_y Y$ (for g_y) and (f_0, φ) is short if $f_0'(x)$ lies inside this unit sphere, i.e. $g_{f_0(x)}(f_0'(x), f_0'(x)) \leq 1$. It follows that $f_0(I)$ is shorter than I, whence the terminology. We can now state the one dimensional convex integration lemma.

Lemma 4.1 (Gromov, 1973)**.** *Let $\mathscr{R} \subset J^1(I, Y)$ be an open differential relation and let (f_0, φ) be a short solution. Then, for all $\varepsilon > 0$ there exist a solution f of $\mathscr{R}$ with*

$$\|f - f_0\|_{C^0} < \varepsilon$$

Sketch of Proof. We first construct a continuous family of loops

$$\begin{array}{cccc} h: & I & \longrightarrow & C^\infty([0,1], \mathscr{R}) \\ & x & \longmapsto & h_x \end{array}$$

such that $f_0'(x)$ is the average value of h_x as on Figure 6.24:

$$\forall x \in I, \quad f_0'(x) = \int_0^1 h_x(u) du$$

Intuitively, h_x exists because by assumption $f_0'(x) = \sum_{i=1}^k w_i z_i$ for some positive weights w_i with $\sum_i w_i = 1$ and some $z_1, \ldots, z_k \in \mathscr{R}_{x f_0(x)}$. Using that $\mathscr{R}$ is open, it is possible to turn this discrete sum into an integral along some C^∞ path in $\mathscr{R}_{x f_0(x)}$ through the z_i. It remains to make this path move continuously with x.

We then set

$$f(x) := f_0(0) + \int_0^x h_s(\{Ns\}) \mathrm{d}s \tag{6.1}$$

where $N \in \mathbb{N}^*$ and $\{Ns\}$ is the fractional part of Ns. It amounts to integrate h along a line in the cylinder $I \times \mathbb{R}/\mathbb{Z}$, which can also be viewed as a spiralling curve. See Figure 6.25. From (6.1) we obtain $f'(x) = h_x(\{Nx\}) \in \mathscr{R}$. Integrating over a period, we get

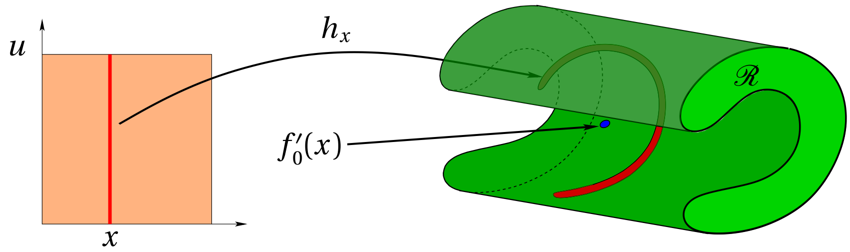


Figure 6.24: $f_0'(x)$ is the average value of h_x.

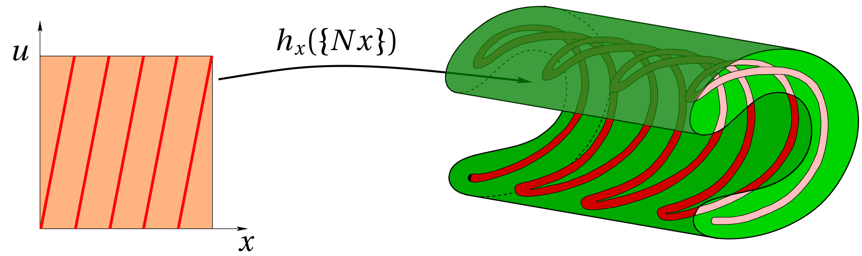


Figure 6.25: When identifying the top and bottom edges of the left parameter domain, the red segments join to form a single spiralling curve.

$$\int_t^{t+\frac{1}{N}} h_s(\{Ns\})\mathrm{d}s \approx \frac{1}{N}\int_0^1 h_t(u)\mathrm{d}u = \frac{1}{N}f_0'(t)$$

if N is large enough. Whence

$$f(x) \approx f_0(0) + \sum_{i=1}^{Nx} \frac{1}{N} f_0'(\frac{i}{N}) \approx f_0(x)$$

Choosing N large enough, we can make f as close as desired to f_0. $\square$

5 The Relation of Isometries

A one dimensional metric on I is given by a strictly positive function $r : I \to \mathbb{R}_+^*$. A map $f : (I, r) \to Y = \mathbb{E}^2$ is an isometry if $\|f'(x)\|_{\mathbb{E}^2} = r(x)$ for all $x \in I$. The differential relation $\mathscr{R}$ for isometries thus constrains the norm of the derivative. Precisely, R_{xy} is the circle of radius $r(x)$ centered at the origin in $T_y\mathbb{E}^2$. Let $f_0 : (I, r) \to Y = \mathbb{E}^2$ be a short map, i.e. $\|f_0'(x)\|_{\mathbb{E}^2} < r(x)$ for all $x \in I$. In order to apply the preceding formula (6.1) to f_0 we need to build explicitely the loops h_x satisfying

$$f_0'(x) = \int_0^1 h_x(u)du \tag{6.2}$$

The simplest is to perform a round trip along a circular arc as on Figure 6.26. This leads to set

$$h_x(s) = r(x)\mathbf{e}_x^{i\alpha_x \cos(2\pi s)} \tag{6.3}$$

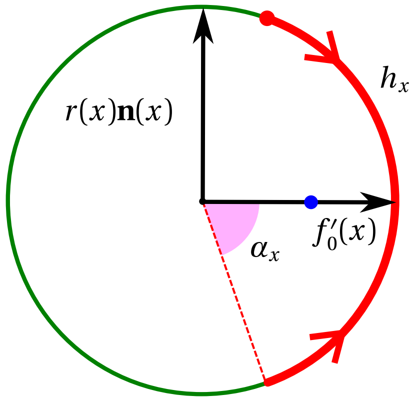


Figure 6.26: Because f_0 is short, its derivative $f_0'(x)$ lies inside the circle of radius $r(x)$. We can choose h_x to travel along a circular arc of angle $2\alpha_x$ and come back. The angle α_x must be chosen so that the average point of h_x is $f_0'(x)$. Here, $\mathbf{n}(x)$ is the unit normal of the path f_0 at x.

where $\mathbf{e}_x^{i\theta} = \cos\theta \frac{f_0'(x)}{\|f_0'(x)\|} + \sin\theta\, \mathbf{n}(x)$ is the unit vector forming the angle θ with $f_0'(x)$, and the angular excursion α_x is chosen to satisfy

$$\int_0^1 r(x)\mathbf{e}_x^{i\alpha_x \cos(2\pi s)} ds = f_0'(x)$$

Equivalently,

$$\alpha_x := J_0^{-1}\left(\frac{\|f_0'(x)\|}{r(x)}\right)$$

where $J_0(x) = \int_0^1 \cos(x\cos 2\pi u)\mathrm{d}u$ is the Bessel function of 0 order (restricted to the interval $[0,z]$ where $z \approx 2.4$ is the smallest positive root of J_0). As can be seen on Figure 6.27 the convex integration formula (6.1) leads naturally to a solution oscillating N times along the initial path. Those oscillations are sometimes called **corrugations**. The term seems to have been originally coined by Bill Thurston to denote the oscillations introduced in his famous sphere eversion demonstrated in the beautiful movie "Outside In" [ea94].

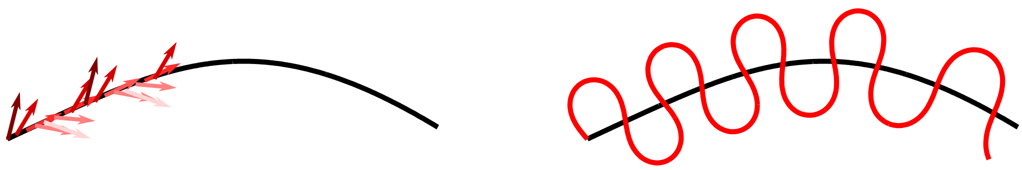


Figure 6.27: Left, The vector $h_s(\{Ns\})$ attached at each point of f_0 should be integrated to compute the isometric solution on the Right. The oscillations of $h_s(\{Ns\})$ induce oscillations of the solution.

6 Isometric Embedding of the Square Flat Torus

We are now ready to describe the construction of an isometric embedding of the square flat torus in $\mathbb{E}^3$. Most of the following description applies to the embedding in $\mathbb{E}^3$ of any closed orientable[5] Riemannian surface. But the details were specifically tuned for the square flat torus. We are thus looking for an embedding $f : (\mathbb{T}^2, g) \to \mathbb{E}^3$ such that $f^*\langle\cdot,\cdot\rangle_{\mathbb{E}^3} = g$. Here, g is the Euclidean metric in the plane but could be any given Riemannian metric. We suppose that we are given a strictly short embedding $f_0 : (\mathbb{T}^2, g) \to \mathbb{E}^3$, i.e. such that $f_0^*\langle\cdot,\cdot\rangle_{\mathbb{E}^3} < g$. The domain $\mathbb{T}^2$ is now two dimensional and the easiest way to apply the one dimensional convex integration process underlying lemma 4.1 is to cover $\mathbb{T}^2$ with a one parameter family of curves and apply the one dimensional process to each curve in the family as illustrated Figure 6.28. This will

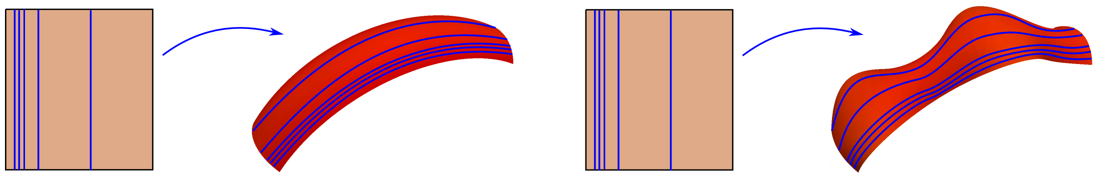


Figure 6.28: Left, each curve in $\mathbb{T}^2$ is sent by the initial embedding f_0 to a short curve in $\mathbb{E}^3$ (only a portion is represented). Right, we apply the one dimensional convex integration to each curve in the family to obtain a corrugated surface.

elongate each curve in the family. However, curves that are transverse to the one parameter family will not change much during this process and will still be shorter than desired. In order to get closer to an isometric embedding one should again corrugate the surface using a transverse family of curves. It is not clear though how to choose this new family of curves or whether new families of curves should be used to corrugate the surface. Following Nash [Nas54], we look to the isometric default

$$\Delta_0 = g - f_0^*\langle\cdot,\cdot\rangle_{\mathbb{E}^3}$$

Because f_0 is strictly short, Δ_0 is a metric, i.e., the quadratic form $(\Delta_0)_x$ is positive definite for all $x \in \mathbb{T}^2$. We observe that the set of inner products in $\mathbb{R}^2$ is a convex open cone

$$Q_+ = \{E\partial_x^* \otimes \partial_x^* + F(\partial_x^* \otimes \partial_y^* + \partial_y^* \otimes \partial_x^*) + G\partial_y^* \otimes \partial_y^*,\ EG - F^2 > 0,\ F > 0,\ G > 0\},$$

whose boundary is composed of squares of linear forms $\ell \otimes \ell$. Those positive non-definite forms are called **primitive forms**. In his original paper, Nash samples Q_+ with an infinite number of points in order to express any inner product q as a smooth convex combination of those points (only a finite number of coefficients are nonzero). This approach would practically prevents us from any effective computations. We rather choose f_0 so that Δ_0 is contained in the open cone $\mathscr{C}$ spanned by exactly three fixed[6] primitive forms $\ell_i \otimes \ell_i$, $i = 1, 2, 3$. See Figure 6.29. In practice we have chosen

$$\ell_1 = \partial_x^*,\ \ \ell_2 = \frac{1}{\sqrt{5}}(\partial_x^* + 2\partial_y^*),\ \ \ell_3 = \frac{1}{\sqrt{5}}(\partial_x^* - 2\partial_y^*)$$

we can now write

$$\Delta_0 = \rho_1\,\ell_1 \otimes \ell_1 + \rho_2\,\ell_2 \otimes \ell_2 + \rho_3\,\ell_3 \otimes \ell_3$$

[5]For non-orientable surfaces we can only get isometric immersions since closed non-orientable surfaces do not embed topologically in $\mathbb{E}^3$.

[6]Here, we use the fact that $\mathbb{T}^2$ is parallelizable, as a Lie group, to transport the same forms everywhere.

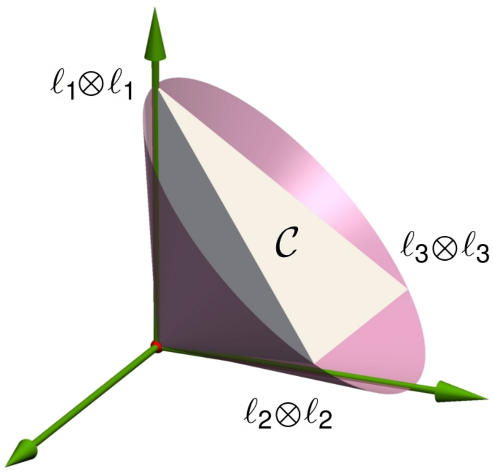


Figure 6.29: The cone of metrics and the subcone $\mathscr{C}$ spanned by $\ell_i \otimes \ell_i$, $i = 1,2,3$.

where the $\rho_i : \mathbb{T}^2 \to \mathbb{R}$ are strictly positive functions. We then apply for each $i = 1,2,3$ a one dimensional convex integration along the integral curves of the field of directions $(\ker \ell_i)^{\perp}$ in order to cancel ρ_i. However, there are basically two reasons why this simple approach cannot succeed. The first reason appears as we look more carefully to the effect of a convex integration in the transverse direction: the isometric default cannot be kept small even when the number of oscillations N grows at infinity; the field $(\ker \ell_i)^{\perp}$ should be replaced by some carefully chosen perturbed directions (see [BJLT13]). Even so, it is not possible to cancel the ρ_i exactly after a stage of three convex integrations has been applied. This second reason is unavoidable and can be inferred from the fact that the differential relation for isometries, call it $\mathscr{R}_{iso}$, is closed while Lemma 4.1 requires $\mathscr{R}$ to be *open*. The situation is schematically described in Figure 6.30. In

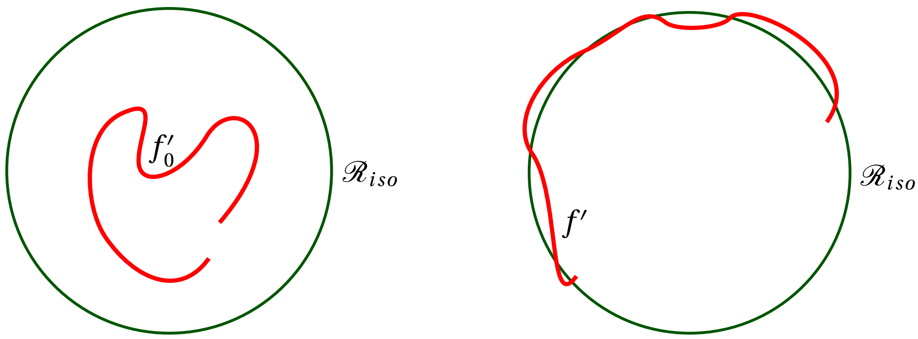


Figure 6.30: Left, the relation for isometries $\mathscr{R}_{iso}$ is depicted as a circle. Being strictly short, the initial embedding f_0 has its derivative inside this circle. Right, after a stage of three convex integrations the new map f is *approximately* isometric, but it may be non-short at some points. This prevents us to re-apply the convex integration process.

order to deal with the closed relation $\mathscr{R}_{iso}$ we replace it by an increasing sequence of open relations $\mathscr{R}_1, \mathscr{R}_2, \ldots$ converging towards $\mathscr{R}_{iso}$. A first stage of three convex integrations approximately cancels the three coefficients of the isometric default and leads to a C^∞ embedding f_1 that is a solution for $\mathscr{R}_1$ and that is short for $\mathscr{R}_2$. We can then apply a second stage of three convex integrations to obtain a C^∞ solution f_2 of

$\mathscr{R}_2$ that is short for $\mathscr{R}_3$. See Figure 6.31. Continuing *ad infinitum*, we build a sequence

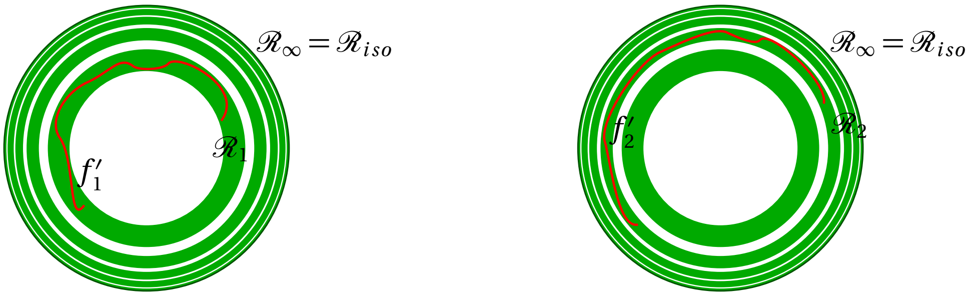


Figure 6.31: Left, starting with f_0, we apply three convex integrations to obtain an embedding f_1 satisfying $\mathscr{R}_1$. Right, we then apply three convex integrations to obtain an embedding f_2 satisfying $\mathscr{R}_2$. This process is repeated *ad infinitum*.

of C^∞ embeddings $f_0, f_1, f_2, \dots$ Choosing the number of oscillations large enough for each convex integration, we can finally show that this sequence C^1 converges to a C^1 isometric embedding. Figure 6.32 illustrates the result of four successive corrugations on a standard torus of revolution while Figure 6.33 shows the images of four "circles" of the flat torus by this embedding.

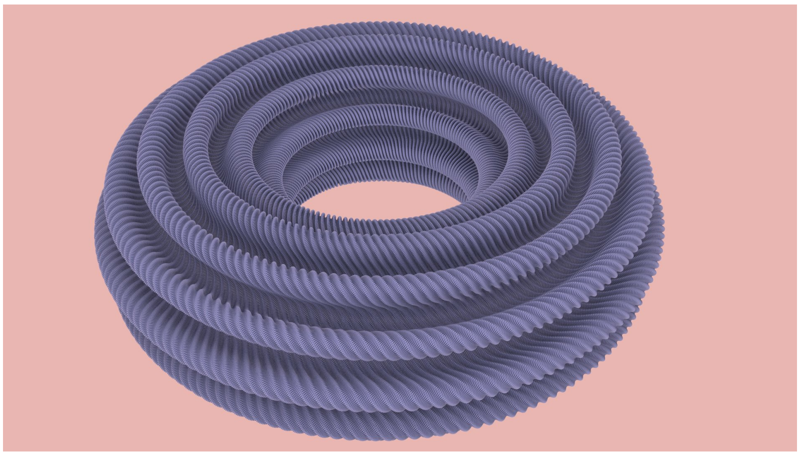


Figure 6.32: A rendered view of a computer model of a corrugated torus. The underlying three dimensional mesh used for the computations contains two milliards of vertices. Further corrugations would not be visible to the naked eye. The embedding is close to an isometry.

7 Fractal Structure

Looking closer and closer to a meridian, the image of a vertical circle on Figure 6.34, we observe a somehow repetitive pattern evoking a fractal. The popular von Koch snowflake fractal is compared to a meridian of the isometric embedding of the square flat torus in Figure 6.35. However, the meridian looks smoother than the snowflake. The meridian is indeed a C^1 curve so that its Hausdorff dimension is one as for any other smooth curve. The fractal structure is rather apparent on its derivative. This is best seen

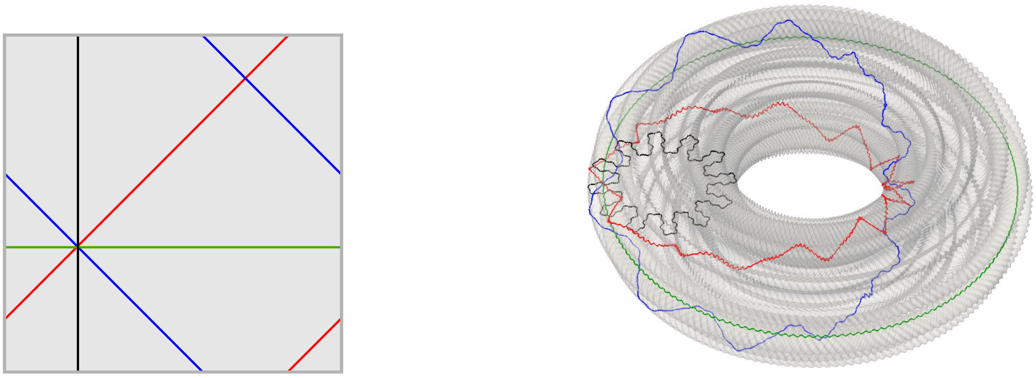


Figure 6.33: The horizontal and vertical segments in the left square represent closed curves of length one in the flat torus while the two pairs of parallel diagonal segments represent two closed curves of length $\sqrt{2}$. Their almost isometric images are shown on the right.

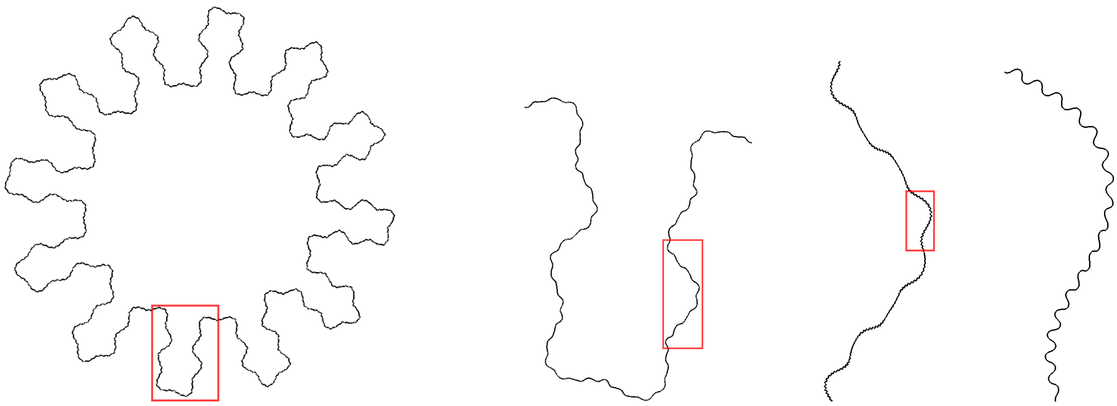


Figure 6.34: Successive zooms in the image curve of a meridian.

by applying the whole convex integration process to a circle (see Figure 6.36). In this case, using the integration loops (6.3), we can obtain an explicit formula for the normal to the limit isometric embedding:

$$\forall x \in \mathbb{S}^1, \quad \mathbf{n}_\infty(x) = \left(\prod_{k=1}^{\infty} e^{i\alpha_k(x)\cos 2\pi N_k x} \right) \mathbf{n}_0(x)$$

where $\mathbf{n}_0(x) = e^{ix}$ is the normal to the initial embedding, α_k is the excursion angle of the integration loop and N_k is the chosen number of oscillations for the kth convex integration. The limit normal is thus turned by an angle $A_\infty(x) = \sum_{k=1}^{\infty} \alpha_k(x)\cos(2\pi N_k x)$ that looks similar to a Weierstrass function $\sum_{k=0}^{\infty} a^k \cos(2\pi b^k x)$ whose graph (Figure 6.37) has a suspected Hausdorff dimension $\ln(a)/\ln(b) + 2$ [Fal03]. The expression for the limit normal can be written in a matrix form:

$$\begin{pmatrix} \mathbf{t}_\infty \\ \mathbf{n}_\infty \end{pmatrix} = \left(\prod_{k=0}^{\infty} \mathscr{C}_k \right) \begin{pmatrix} \mathbf{t}_0 \\ \mathbf{n}_0 \end{pmatrix} \tag{6.4}$$

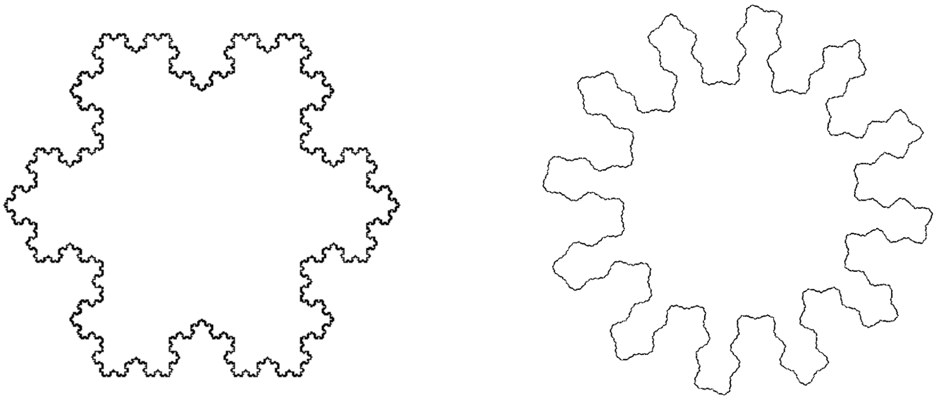


Figure 6.35: Left, The von Koch curve. Right, A meridian of the isometric embedding.

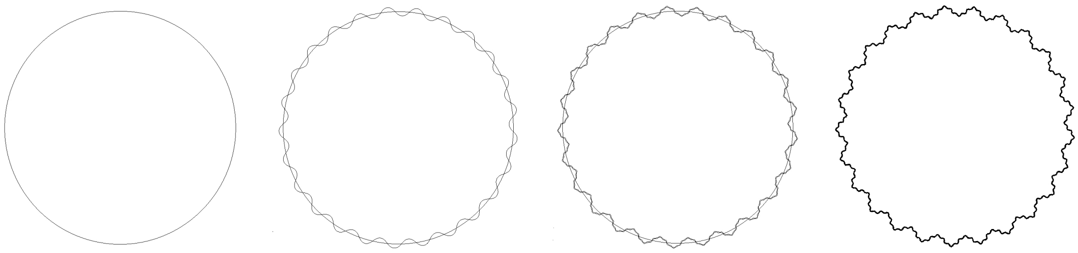


Figure 6.36: Starting with an embedding of the unit circle as a circle of radius < 1, we apply the whole convex integration process to obtain an isometric embedding, i.e., a closed curve of length 2π. Some embeddings in the resulting sequence are shown.

where $\mathscr{C}_k = \begin{pmatrix} \cos\theta_k & \sin\theta_k \\ -\sin\theta_k & \cos\theta_k \end{pmatrix}$, $\theta_k(x) = \alpha_k(x)\cos(2\pi N_k x)$ and $\mathbf{t}_\infty, \mathbf{t}_0$ are the unit tangent to the limit and initial embedding respectively. The frame $\begin{pmatrix} \mathbf{t}_\infty \\ \mathbf{n}_\infty \end{pmatrix}$ is thus obtained by applying an infinite number of rotations to the initial frame. It turns out that a similar structure can be proved for the limit isometric embedding of the square flat torus. More precisely, if $\begin{pmatrix} v_\infty^\perp \\ v_\infty \\ \mathbf{n}_\infty \end{pmatrix}$ are $\begin{pmatrix} v_0^\perp \\ v_0 \\ \mathbf{n}_0 \end{pmatrix}$ are frames adapted to the limit and initial embeddings of the torus, it can be shown [BJLT12] that

$$\begin{pmatrix} v_\infty^\perp \\ v_\infty \\ \mathbf{n}_\infty \end{pmatrix} = \prod_{k=0}^{\infty} \left(\prod_{j=1}^{3} \mathscr{C}_{k,j} \right) \begin{pmatrix} v_0^\perp \\ v_0 \\ \mathbf{n}_0 \end{pmatrix} \tag{6.5}$$

where each $\mathscr{C}_{k,j}$ is a rotation approximately equal to the following simple product:

$$\mathscr{C}_{k,j} \approx \begin{pmatrix} \cos\theta_{k,j+1} & 0 & \sin\theta_{k,j+1} \\ 0 & 1 & 0 \\ -\sin\theta_{k,j+1} & 0 & \cos\theta_{k,j+1} \end{pmatrix} \begin{pmatrix} \cos\beta_j & \sin\beta_j & 0 \\ -\sin\beta_j & \cos\beta_j & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

where $\theta_{k,j} = \alpha_{k,j}(x)\cos(2\pi N_{k,j} x)$ and β_j is the angle between the kernels of the primitive forms ℓ_j and ℓ_{j+1}. Comparing (6.4) and (6.5), we qualified the limit embedding as a

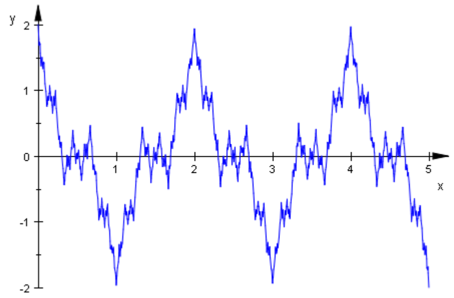


Figure 6.37: The graph of a Weierstrass function.

smooth fractal. This is the analogue of the primitive of a fractal curve but for two dimensional surfaces. In nature, some seashells present successive corrugations very similar to the corrugations of the flat torus embedding. Smooth fractals could be the answer of mother Nature to certain physical constraints...

8 Some pictures

We have implemented the convex integration process to produce computer rendering of the (unit) square flat torus. Calculations were performed on a 8-core CPU (3.16 GHz) with 32 GB of RAM and parallelised C++ code. The initial map f_0 is the standard parametrization of the torus of revolution with minor and major radii respectively $\frac{1}{10\pi}$ and $\frac{1}{4\pi}$. We applied four convex integrations with oscillation numbers $12, 80, 500, 9000$. Each map f of the sequence is encoded by a $n \times n$ grid whose node (i, j) contains the coordinates of $f(i/n, j/n)$. Flows and integrals are common numerical operations for which we have used Hairer's solver [HNW91] based on DOPRI5 for non-stiff differential equations. We used a $10,000^2$ grid mesh ($n = 10,000$) for the three first corrugations and refined the grid to 2 milliards nodes for the last corrugation. To illustrate the metric improvement we have compared the lengths of a collection of meridians, parallels, and diagonals on the flat torus (see Figure 6.33) with the lengths of their images by the last computed embedding. The length of any curve in the collection differs by at most 10.2% with the length of its image. By contrast, the deviation reaches 80% when the standard torus f_0 is taken in place of the last embedding. In fact, the situation is even better since we actually decreased the isometric default with respect to an intermediate metric (corresponding to the open relation $\mathscr{R}_1$) rather than the flat metric.

I do not resist to end this part with some of the beautiful pictures output by our program. See Figures 6.38,6.39 and 6.40. Damien Rohmer, a specialist of image rendering, joined the project and greatly help us for the rendering. More pictures and a short movie can be viewed on the HEVEA website http://hevea.imag.fr/.

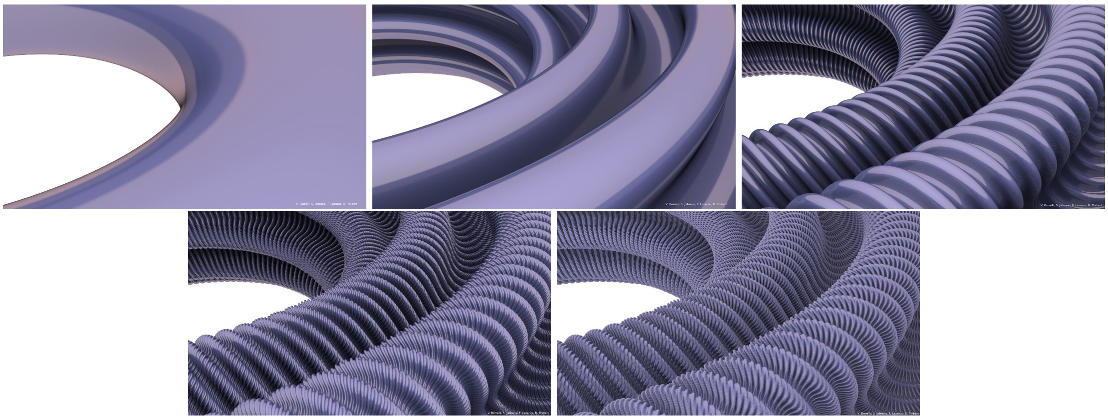


Figure 6.38: The first four corrugations, starting from the standard torus.

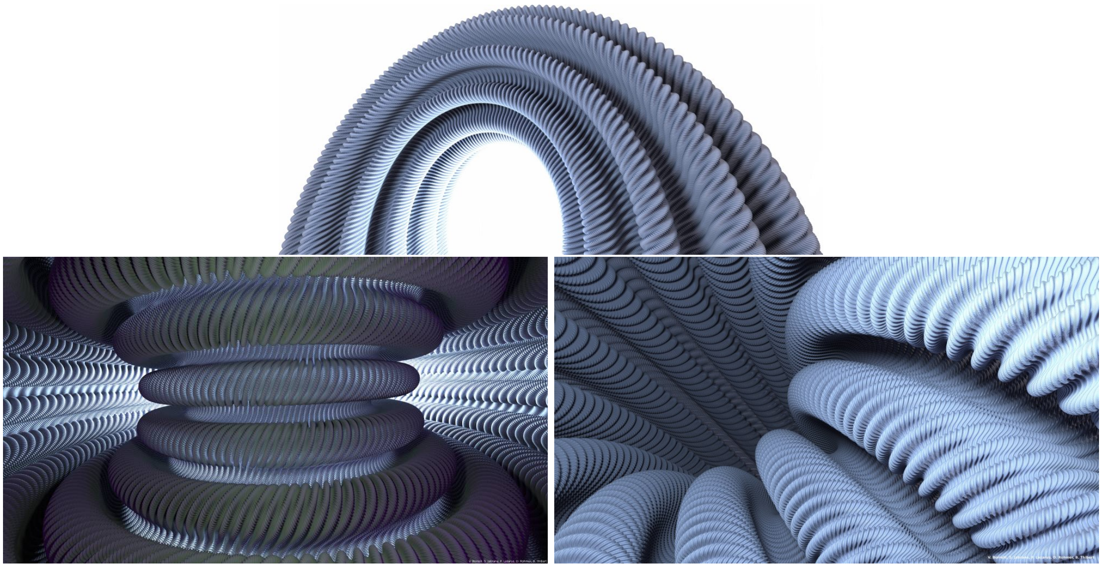


Figure 6.39: More views from the outside and inside of the last computed embedding.

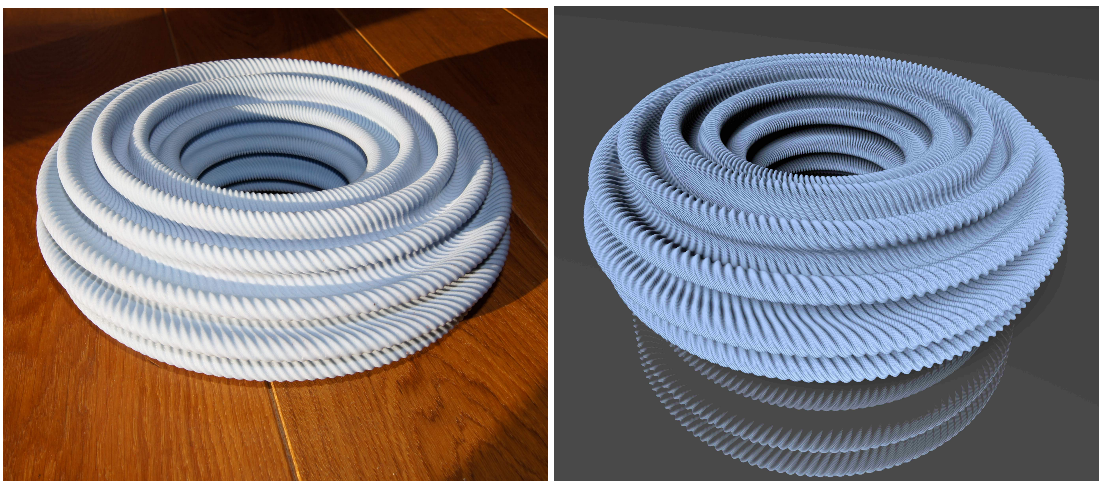


Figure 6.40: 3D printing (left) and computer rendering of the last computed embedding.

Appendix A

Main Research Activities

In this part I give a brief overview of my past research activities. The presentation is deliberately informal to emphasize a thematic evolution rather than a succession of independent works. I started working on deformation tools for the modeling and animation of shapes in computer graphics. Most of the deformation tools proposed at that time for three dimensional meshes were based on the computation of a field of deformation. This field was interpolated from sampled values specified interactively by the user. Such deformation fields were not so much adapted to articulated bodies. I thus developed with Sabine Coquillard and Pierre Jancène a more convenient tool based on axial deformations [19]. The deformation relies on a cylindrical coordinate system defined by an axis curve. Some results are pictured on Figure A.1. The tool could

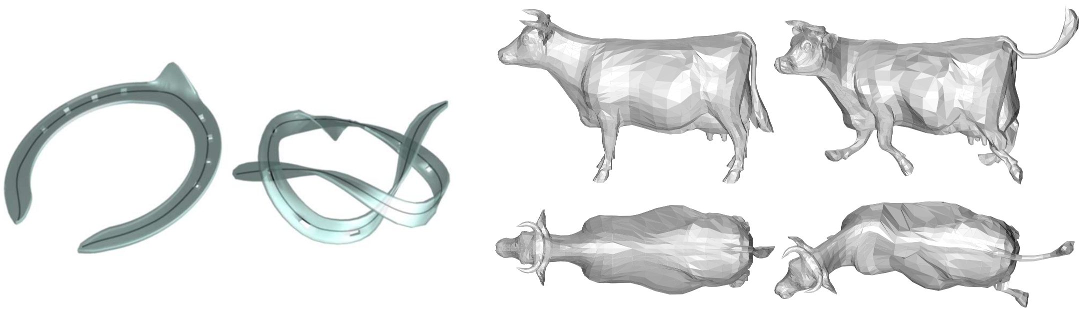


Figure A.1: An axis curve is interactively drawn to deform a 3D mesh. Each vertex of the mesh are given cylindrical coordinates with respect to a moving frame along the axis. The deformation of the axis is passed to the mesh by preserving the cylindrical coordinates of each vertex referring this time to the deformed axis. In the right part of the figure the axis is not shown.

be applied locally as demonstrated on Figure A.2. It was used by the artist Arghyro Paouri to create an animated sequence. See Figure A.3. In order to avoid the tedious task of positioning an axis adapted to a given shape, I implemented with Anne Verroust a simple automatic construction of a spin curve inside a mesh [26, 30]. In practice, we have considered the barycenters of the level sets of some distance function to provide a skeleton as illustrated on Figure A.4 and A.5.

I also studied an interpolation schema between two axis curves in order to animate a deformation [16, 17]. A simple linear interpolation may lead to the apparition of

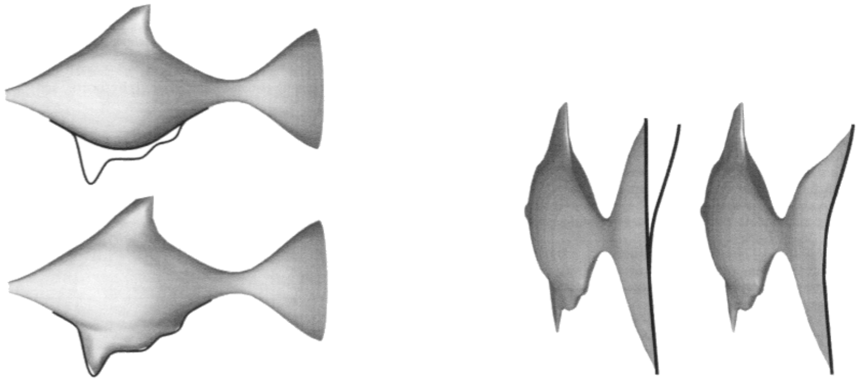


Figure A.2: The deformation is only applied in a neighborhood of the axis defined a zone of influence.

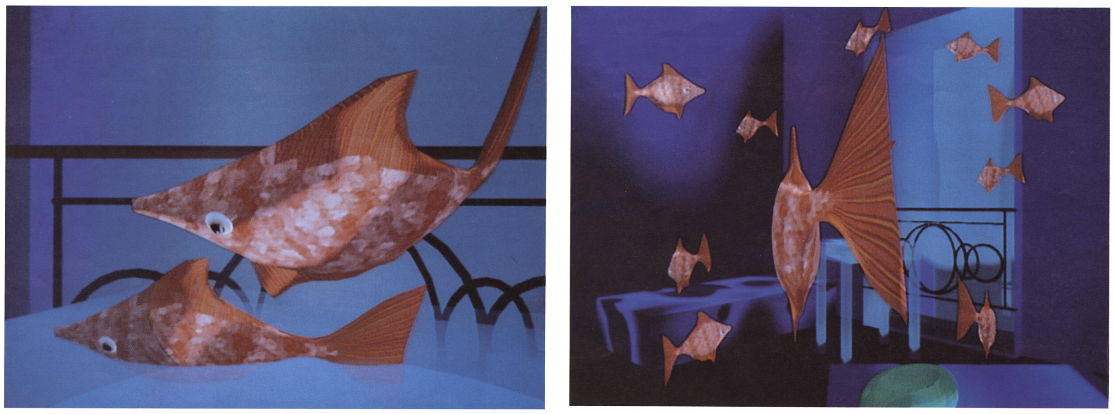


Figure A.3: Two pictures of *Fish Story* by A. Paouri, 1993.

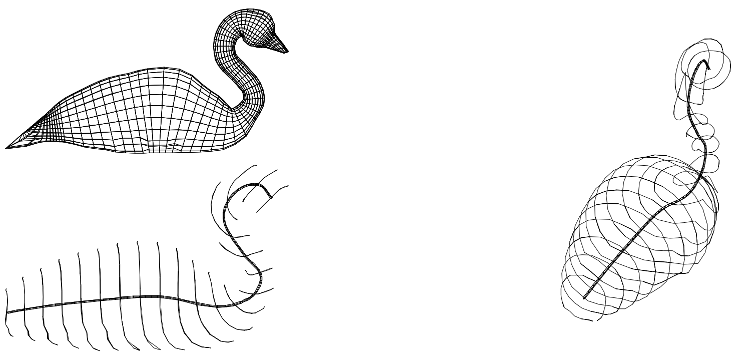


Figure A.4: The computation of a spin curve is based on the level sets of the geodesic distance from a given vertex of a mesh.

cusps or the cancellation of the tangent vector of the interpolated curves. In order to avoid such undesired effects the two axis are endowed with an adapted moving frame and the interpolation applies to the frames rather than the curves. It appears that this interpolation process is rather general and applies to regular differentiable curves. In other words, it yields a homotopy in the space of curve immersions. The interpolation consists in viewing the curves as trajectories of a space uniform time varying vector field

Figure A.5: The spin curve may ramify to produce a skeleton like structure.

and to interpolate the vector fields. Again, we cannot use a linear interpolation of the vector fields as the interpolated field may cancel if the given vector fields have opposite directions at some point. A natural solution is to interpolate the moving frames instead of the velocity vectors. A moving frame can be seen as a curve in the (Lie) group of rotations $SO(3,\mathbb{R})$. The derivative of such a curve can be transported in a canonical way into the associated Lie algebra of skew-symmetric 3×3 matrices. We can now use linear interpolation of those derivatives, then integrate the interpolated derivative to obtain an interpolated curve in $SO(3,\mathbb{R})$. This finally corresponds to the moving frame of some interpolated curve in $\mathbb{R}^3$, thus providing the desired homotopy. The process can be seen as a non trivial interpolation of the Cartan matrices of the moving frames, leading to the fact that the curvature and torsion varies in a monotonic manner (see lower left Figure A.6). This is an interesting property when one wants to avoid excessive bending during the interpolation. In practice, the process is applied to polylines sampled from splines or other smooth curves and the various integrations are obtained with Newton like schemes. The same spline curves sampled with 40 and 10 vertices respectively were used for the animation on the left part of Figure A.6 demonstrating the robustness of the computation.

I eventually combined the axial deformation tool with the computation and animation of axes to produce metamorphoses of 3D meshes [23, 24, 25]. Figure A.7 illustrates the transformation process applied to a sickle and a fish. I implemented the metamorphosis tool which was used by A. Paouri for a short movie sequence and the creation of sculptures. See Figures A.8 and A.9.

I then joined the team of Gabriel Taubin for a post-doc at the IBM T.J. Watson center. Gabriel and Jarek Rossignac had just designed a new compression method for efficient encoding of triangulations [TR98]. Their method is based on a decomposition of a triangulation into long strips of triangles obtained by an orange peeling algorithm. Together with Gabriel, William Horn and André Guéziec we extended this method to handle non-manifold meshes [15] and produced a complete specification for a standard compression format to transmit meshes through the Web [29, 13]. We also proposed progressive multi-level compression schemes [14, 28]. While these works were aimed at

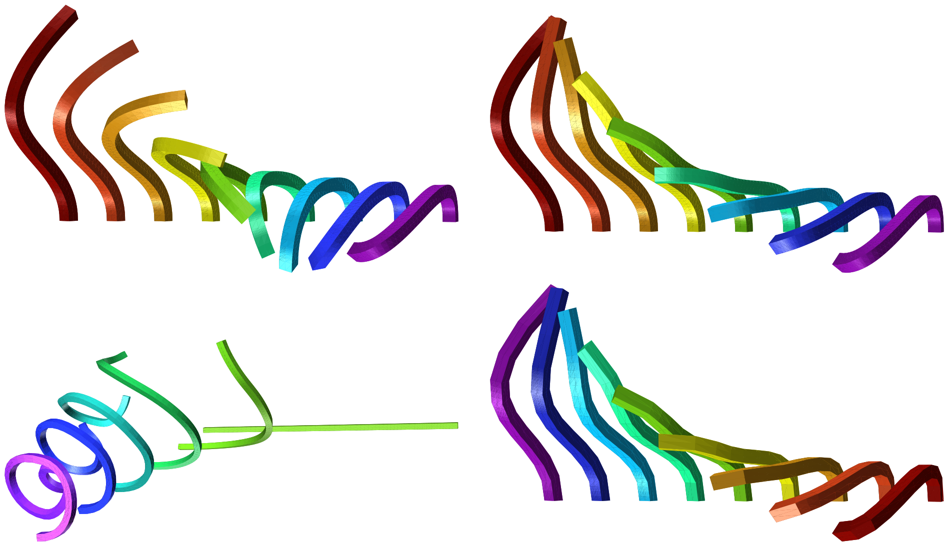


Figure A.6: Upper row, two interpolations corresponding to the Serret-Frenet frame (left) and to the rotation minimizing frame (right). Lower left, Interpolation between a bended curve and a line segment. Curvature and torsion decrease continuously during the transformation. Lower right, The computation is applied to polylines with only 10 vertices, showing the remarkable stability of the process.

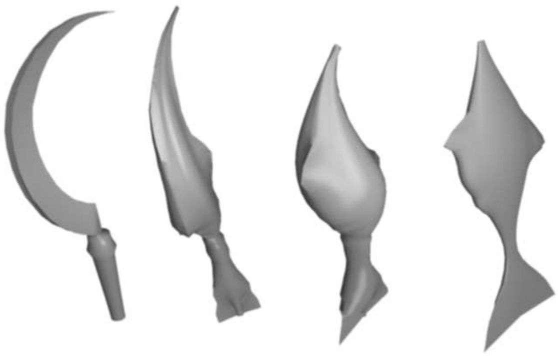


Figure A.7: Transformation of a sickle into a fish.

solving practical issues, they were appealing to concepts of combinatorial topology.

Back to France I decided to pursue in this direction. I first obtained, in a joint work with Michel Pocchiola, Gert Vegter and Anne Verroust, two efficient algorithms to compute a piecewise linear canonical system of loops on a triangulated orientable surface [LPVV01]. A system of loops is a cut graph made of simple loops sharing a common basepoint and otherwise disjoint; a canonical system of loops is a system of loops that meet in a particular cyclic order around the basepoint. It may be used to cut surfaces into discs in order to compute an actual homeomorphism between two surfaces of the same genus. The computation of a canonical system had already been tackled by Gert Vegter and Chee Yap [VY90]. Their paper claims that some inductively constructed graph can be maintained planar. For a proof the authors refer to a future full version of the paper that never appeared, and according to one of the authors the claim remains unproved. I implemented the two new algorithms. Some outputs are

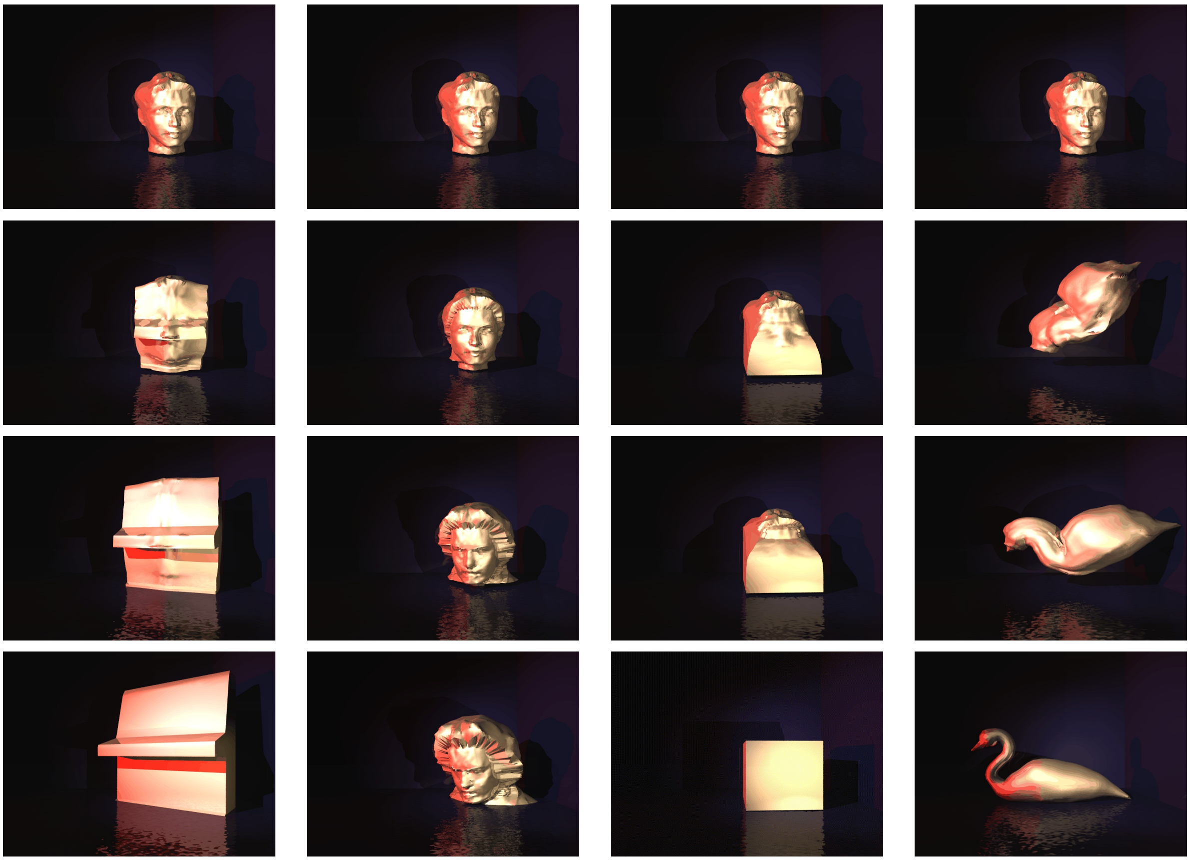


Figure A.8: Four transformations (one by column) extracted from animations by A. Paouri. Low resolution files can be visualized at http://www.gipsa-lab.fr/ francis.lazarus/Documents/Tete_Cyg.mpg and http://www.gipsa-lab.fr/ francis.lazarus/Documents/Piano_Bee.mpg.

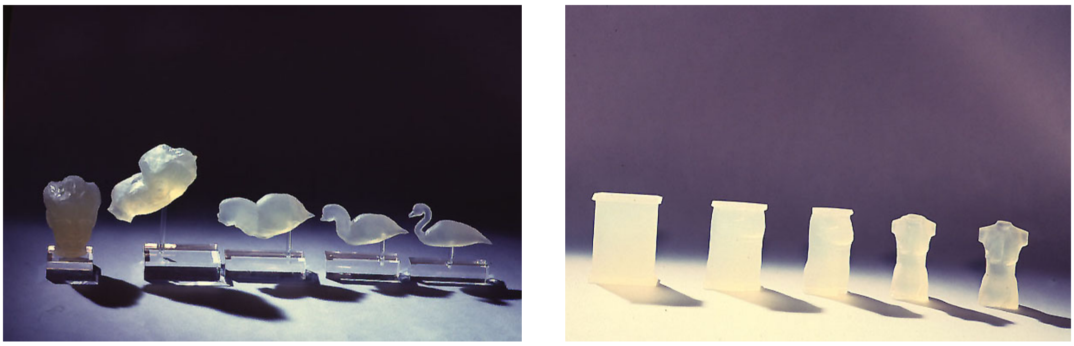


Figure A.9: METADATA: two series of sculptures by A. Paouri obtained from stereolithography.

shown in Figure A.10. Their complexity is proportional to the surface genus g times its number n of triangles. It is optimal in the sense that for every g and n there is a triangulation of size n and genus g all of whose canonical systems of loops have size $\Omega(gn)$.

I then started a long collaboration with Éric Colin de Verdière, at the time preparing

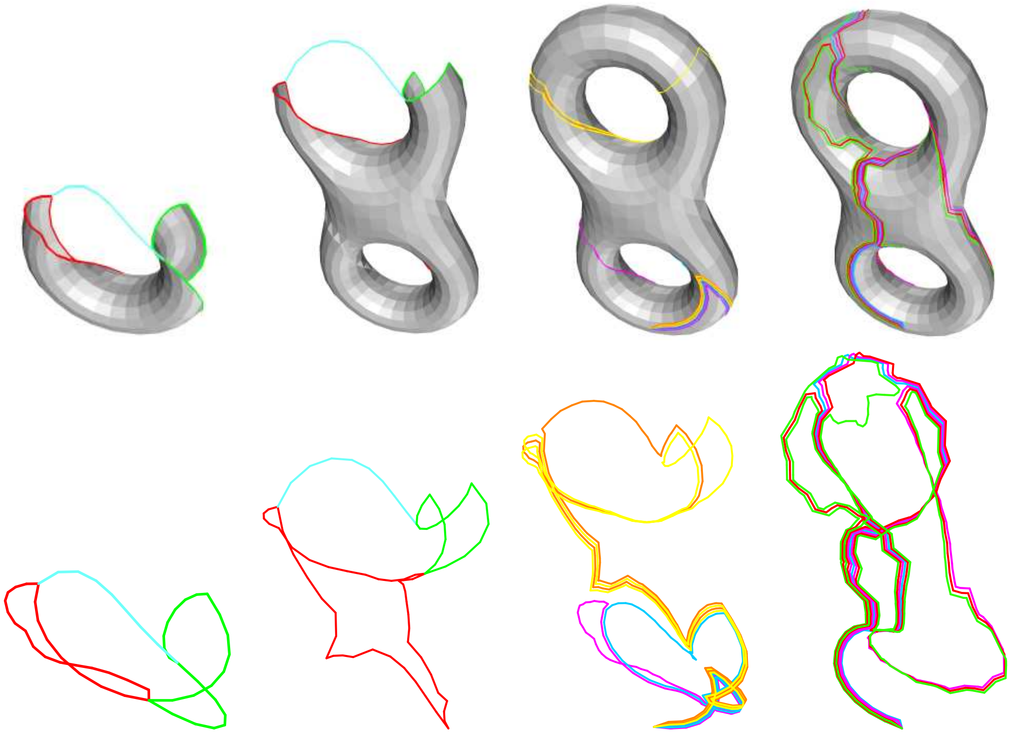


Figure A.10: One of the two algorithms is based on a traversal of the surface triangulation corresponding to the evolving level set of some discrete function. A, B, When two boundary components merge (highlighted in red and green), a linking (blue) path is searched in the unvisited part of the surface. Two more canonical generators are computed from these boundaries and path. C, The resulting canonical system. D, A canonical system computed with another algorithm based on the original Brahana's algorithm [Bra21].

a thesis under the supervision of Michel Pocchiola. We have proposed algorithms to compute a shortest simple loop homotopic to a given simple loop on a combinatorial surface. The edges of the surface are supposed to be positively weighted. The key point is to augment the loop into some family of loops and to shorten the whole family rather than a single loop. For homotopy with fixed basepoint we take as a family a fundamental system of loops [CdVL05]. For free homotopy, we consider a pants decomposition [12]. The family of loops cuts the surface into a disc in the former case and into a collection of pair of pants in the latter case. Each loop in the family is shorten as much as possible inside the cut surface to obtain a new family of curves. This is repeated for each loop in turn until no loop can be shorten. It can be proved that each loop in the resulting family is as short as possible among the homotopic simple loops. Figure A.11 illustrates the process on a flat torus (a locally Euclidean torus). Our analysis gave a polynomial time complexity up to a multiplicative factor proportional to the maximum ratio between any two edge lengths. This factor was shown to be unnecessary in a further analysis by Éric

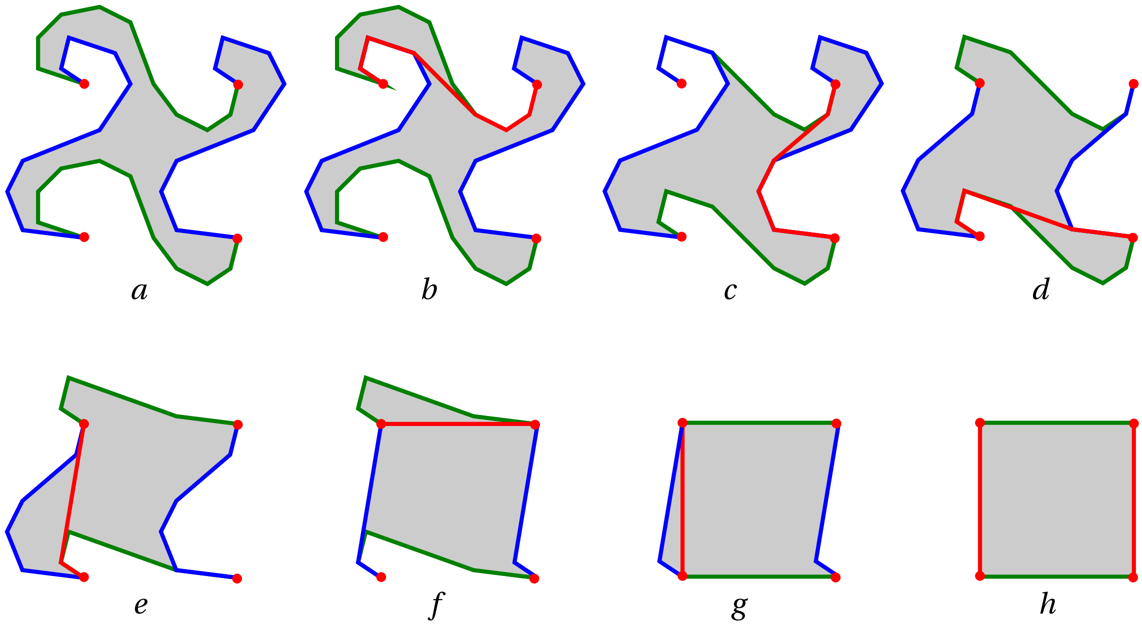


Figure A.11: a, The initial set of two loops cuts the torus into a disc. The two (blue) vertical sides identify to the first loop and the (green) horizontal sides identify to the second loop. b, A shortest (red) loop homotopic to the green loop is computed inside the disc. c, Replacing the green loop by the new one we get another disc in which we compute a shortest loop homotopic to the blue loop. d – h, the process is repeated until we reach a shortest system.

and Jeff Erickson [CE10]. I implemented the shortening algorithm in C++. Figure A.12 shows some shortening steps on an initial canonical system of loops.

During a visit with Éric to Jeff at the University of Urbana-Champain, we worked on another optimization problem: the computation of a minimal non-contractible separating cycle, also called a *splitting* cycle, on a combinatorial surface with weighted edges. Here, a cycle is a closed walk in the graph of the surface such that some infinitesimal perturbation of the cycle is a simple curve on the surface. On a surface without boundary, a splitting cycle cuts the surface into non-trivial components each of non-zero genus. Interestingly, the set of splitting cycles does not satisfy the 3-path condition introduced by Thomassen [Tho90]. This condition leads to polynomial time algorithms for computing a minimal cycle. It turns out that the computation of a minimal splitting cycle is NP-hard [9] as shown by a reduction of the Hamilton cycle problem in grid graphs [IPS82]. We nonetheless propose as a fixed parameter tractable algorithm with respect to the genus of the surface. It relies on the existence of a splitting cycle that cuts every shortest path $O(g)$ times. Figure A.13 and A.14 pictures some steps of the argument.

The collaboration with Urbana-Champain was continued, grouping six people around the specific problem of the Fréchet distance of curves in the perforated plane. The Fréchet distance between two curves in the plane is the minimum length of a leash that allows a dog and its owner to walk along their respective curves, from one end to the other, without backtracking. Alt and Godau [AG95] describe a polynomial time algorithm to compute the Fréchet distance between polygonal curves in the plane. For

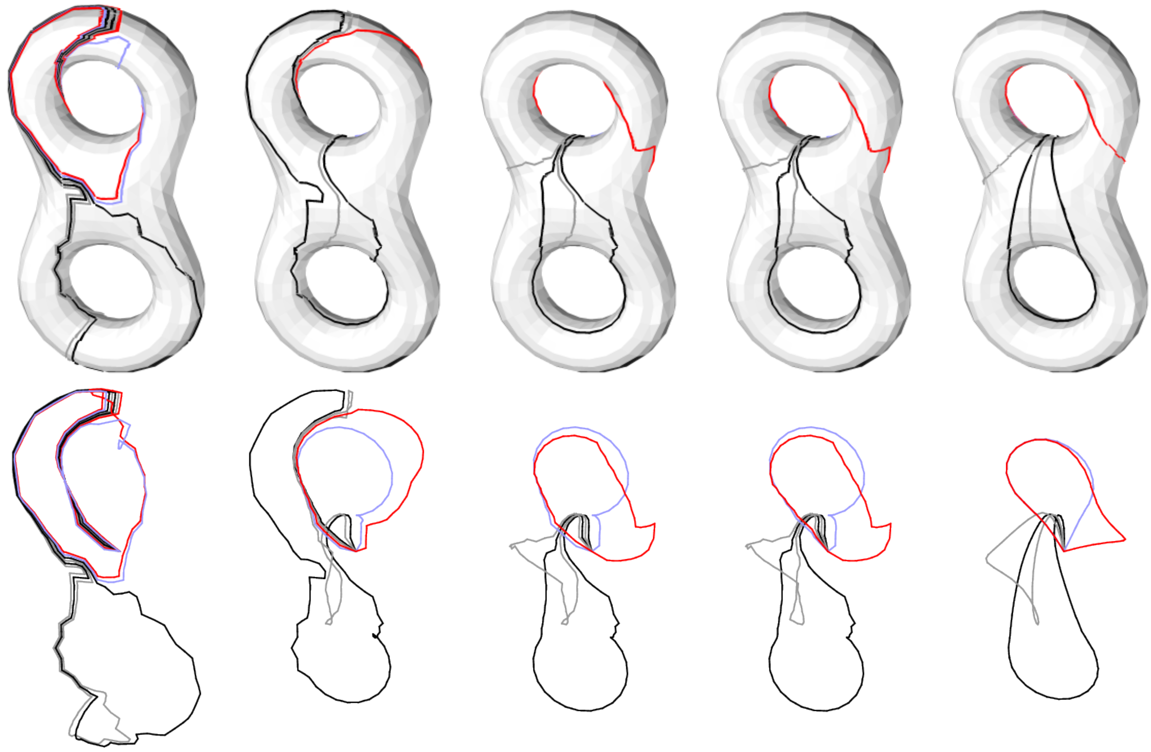


Figure A.12: A, A canonical system of four loops obtained after [LPVV01]. The system after shortening each loop once (B), twice (C) and thrice (D). D, The system is minimal; it cannot be shorten anymore. Using a continuous geometric model, the system was further optimized locally to get a geodesic system of loops on the piecewise linear surface.

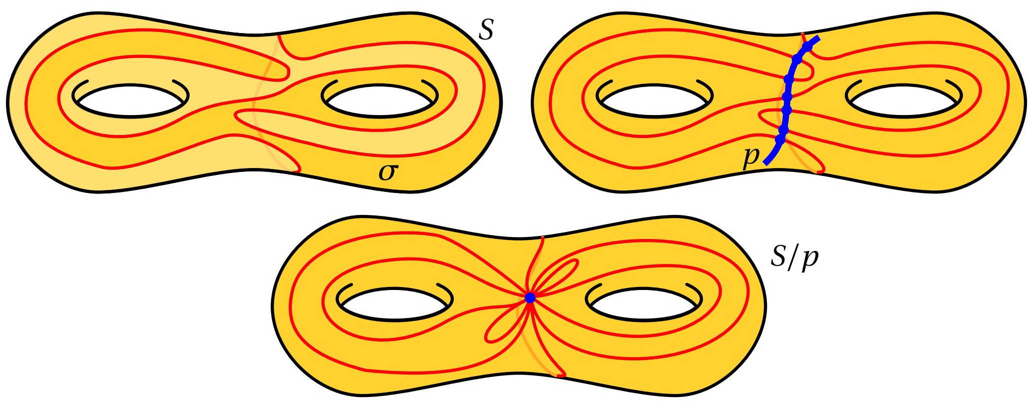


Figure A.13: Upper left, a splitting cycle σ on a genus two surface S. Upper right, A shortest path p cuts the splitting cycle into pieces. Middle, Contracting the path, we obtain a homeomorphic surface S/p where the cycle pieces become non-intersecting loops. From the Euler characteristic it is seen that those loops define $O(g)$ distinct homotopy classes.

a non-trivial topology and geometry, such as the plane minus a set of polygons, we also require that the leash moves continuously over time. This condition constrains

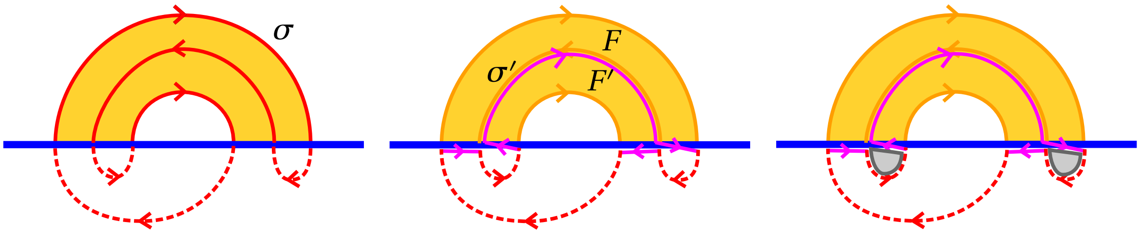


Figure A.14: Left, Three homotopic loops or inverse loops in S/p correspond to three parallel pieces of σ in S. Since the splitting cycle is separating, the orientations of the pieces must alternate as on the figure. Middle, a shorter and homologous cycle $\sigma' = \sigma + \partial_2(F + F')$. Right, Assuming the minimality of σ, the shaded regions cannot be discs, attesting the non-contractibility of σ'. It follows that σ' is splitting, which contradicts the minimality of σ. As a result, the number of pieces of σ is at most twice the number of homotopy classes in σ/p.

the homotopy class of the leash (with endpoints on the curves) to remain constant over time. We show that the homotopy class corresponding to a shortest leash must contain a line segment between the curves. This allows us to enumerate a small subset of potentially optimal homotopy classes and to determine the Fréchet distance between polygonal curves in polynomial time [10].

Turning back to curve optimization on surfaces, we studied with Éric and Sergio Cabello the existence of simple cycles without repeated vertices constrained to certain topological properties. Given a combinatorial surface with n edges we provide $O(n)$ algorithms for deciding the existence of a contractible, non-contractible, or non-separating cycle without repeated vertices. In accordance with [9], we prove that the problem is NP-complete for splitting or separating cycles and propose a fixed parameter tractable with respect to the genus [7]. We have also proposed an output sensitive algorithm for computing a non-contractible (or non-separating) cycle with the least number of edges. If the wanted cycle has k edges, it can be found in $O(gnk)$ time on a genus g surface [8]. This is to be compared with the probabilistic $O(g^2 n \log n)$ time algorithm of Cabello et al. [CCE13]. Our algorithm starts with a subset V of the surface vertices that intersects any non-contractible cycle. We may take for V the vertices of a cut graph. We then decide for each vertex of V if some non-contractible cycle with at most k edges uses that vertex. We can partition V into $O(gk)$ bins so as to perform the decision in parallel for all the vertices in a bin. This leads to our $O(gnk)$ time algorithm.

When the edges of the surface graph G are weighted, we may compute a shortest non-contractible or non-separating cycle. In [6], we describe two algorithms for the directed case where each edge has two weights (possibly infinite) depending on the direction in which the edge is traversed. For the first algorithm, we revisit the 3-path condition by Thomassen [Tho90]. A family $\mathscr{F}$ of loops with basepoint s satisfies the 3-path condition if (1) the inverse of a path in $\mathscr{F}$ is in $\mathscr{F}$ and (2) for any three paths between s and another common endpoint, if two of the three loops formed by these paths are in $\mathscr{F}$ so is the third one. We may extend Thomassen's generic method to compute the shortest loop in the complement $\overline{\mathscr{F}}$ of $\mathscr{F}$. To this end we consider a shortest path tree T with root s and a shortest path tree R with sink s. For each edge xy of G we let $\ell_{TR}(xy) := T[s,x] \cdot xy \cdot R[y,s]$ be the loop composed of the path from s to x in T, the edge xy and the path from y to s in R. Similarly, we put $\ell_{TT}(xy) := T[s,x]\cdot xy \cdot T[y,s]$. We next consider the set of loops

$$\mathscr{M} = \{\ell_{TR}(xy) \mid xy \in G \text{ and } \ell_{TR}(xy) \in \overline{\mathscr{F}}\}$$

and

$$\mathscr{N} = \{\ell_{TR}(xy) \mid xy \in G \text{ and } \ell_{TT}(xy) \in \overline{\mathscr{F}}\}$$

As in Thomassen's method, it is easily shown that a shortest loop in $\mathscr{M}$ is a shortest loop in $\overline{\mathscr{F}}$. More surprisingly, a shortest loop in $\mathscr{N}$ is a shortest loop in $\overline{\mathscr{F}}$. Note that $\mathscr{M} \subset \overline{\mathscr{F}}$ but we cannot generally say that $\mathscr{N} \subset \overline{\mathscr{F}}$. When $\mathscr{F}$ is the set of contractible loops or the set of separating loops, working with $\mathscr{N}$ leads to more efficient algorithms. We obtain a $O(n^2 \log n)$ time algorithm for the computation of a shortest non-contractible or non-separating cycle, where n is the number of edges of the genus g surface. The second algorithm relies on a divide-and-conquer approach based on the existence of small size ($O(\sqrt{gn})$) separators that can computed in $O(n)$ time [Epp03, Th. 5]. After computing such a separator X, the graph $G - X$ is the union of two disjoint subgraphs G_1, G_2. We compute a shortest non-contractible (or non-separating) cycle intersecting X. We also compute a shortest non-contractible cycle recursively in G_1 and G_2. Each G_i, $i = 1,2$, need not be cellularly embedded, but the classification of cycles in G_i requires to maintain an embedding in the original surface Σ. For this purpose, we contract the complement of G_i as much as possible while preserving a cellular embedding in Σ. This results in a relatively small cellular embedding containing G_i. Combining all the steps we get a $O(g^{1/2} n^{3/2} \log n)$ time algorithm. For the non-separating case, Jeff Erickson [Eri11] proposes a $O(g^2 n \log n)$ time algorithm. This is more efficient as soon as $g = O(n^{1/3})$.

I then studied the fundamental problem of computational topology for curves on surfaces: decide whether a loop is contractible or whether two loops are freely homotopic. A first study is due to Dey and Guha [DG99]. They identify the decision problems as the word and conjugacy problems in surface groups. However, their paper contains serious flaws [LR11, LR12] leaving the question open. With my student Julien Rivaud we have obtain the first linear time algorithm for the contractibility test or the free homotopy test [LR11, LR12]. After a preprocessing in time proportional to the surface complexity (its number of edges) the homotopy tests can be done in time proportional to the size of the loops. Our presentation is geometric and relies on the construction of a portion of some cyclic covering of the surface. Jeff Erickson and Kim Whittlesey greatly simplified our approach to obtain the same complexity [EW13]. Their method is based on classical results on small cancellation theory [GS90] and do not necessitate the construction of a covering. In [RL12], Julien and I provide a simpler homotopy test for surfaces with non-empty boundary.

Another active subject of research in computational topology is concerned with the notion of homological persistence. This notion appears in papers by Robins [Rob99] and Edelsbrunner et al. [ELZ00] in the context of approximation theory. It was followed by numerous developments [EH08]. The principle is to encode an evolution process of an object rather than the object itself. We start with a filtration of a simplicial complex K:

$$K_1 \subset K_2 \subset \ldots \subset K_n = K$$

This filtration induces a sequence of linear maps at the homology level (with coefficients in a field):

$$H_*(K_1) \to H_*(K_2) \to \ldots \to H_*(K_n)$$

where each arrow is induced by inclusion. Such a sequence has a complete invariant given by the *persistence intervals*. Those are pairs of indices in the filtration. When the filtration corresponds to the sub-level sets of a scalar function $f : K \to \mathbb{R}$, the intervals can be represented by a *persistence diagram* $D(f)$ in the plane: to an interval $[a, b[$ we associate the point (f_a, f_b) in $D(f)$, where f_a is the maximal f-value over the sub-level with index a. Intuitively, points that are close to the diagonal $\{x = y\}$ represent topological incidents that last a short amount of time in the filtration. One can prove [CSEH07] that the persistence diagram is stable. More precisely:

$$d(D(f), D(g)) \leq \|f - g\|_\infty$$

where $\|f - g\|_\infty = \sup_{\sigma \in K} |f(\sigma) - g(\sigma)|$ and $d(.,.)$ is a suitable distance over the set of diagrams. See my course notes [Laz12b] for an introduction to homological persistence and the stability theorem. The problem of simplification consists somehow in reverting this result. Starting from the diagram D obtained by deleting the points of $D(f)$ at a distance less than ε from the diagonal, the problem is to construct a function $g : K \to \mathbb{R}$ that is ε-close to f and such that $D(g) = D$. The function g is called a *topological simplification* of f. When K is a triangulated surface, Edelsbrunner et al. [EMP06] have described a rather intricate process to construct such a g. Due to is intricacy, the complexity of the process can hardly be analysed. I found a very simple linear time algorithm for this problem. This work was initiated during a workshop and originally implied five people [1]. The algorithm relies on a duality for persistence intervals [CSEH08] that allows to restrict f to the 1-skeleton of K. The simplification problem for graphs can be further reduced to the case of trees for which there is a simple linear time algorithm. In this work [1] we also show that the simplification problem may have no solution when K is not a surface. Counter-examples are easily built from a triangulated dunce cap or a Poincaré homology sphere. The first counter-example indicates that an obstruction already exists in dimension 2, while the second one shows that manifolds of dimension 3 may not have a simplification.

As described in the second part of this document, the HEVEA project was my second main activity during the last six or seven years. It was on the cover of the Proceedings of the National Academy of Sciences (PNAS) [BJLT12] and of the French edition of the Scientific American [BLT13]. A long version appeared in the Brazilian journal Ensaios Matemáticos [BJLT13]. Another article describes in details the convex integration process for curves [3]. See Part II for a description of the project.

Appendix B

A Quick Primer on Combinatorial Group Theory

For completeness, I give a short introduction to the basics of combinatorial group theory. A classical reference is [LS77]. See also [III04] for a nice introduction or [vCGKZ98] for a connection with Topology.

Let S be set of symbols. We associate to S a set S^{-1} made of the symbols of S with an additional -1 upperscript, as in s^{-1}. Intuitively, s^{-1} is the group inverse of s. We consider an equivalence relation over the finite words with alphabet $S \cup S^{-1}$: two words are equivalent if one can be transformed into the other by a sequence of insertions and/or removals of factors of the form ss^{-1} or $s^{-1}s$ with $s \in S$. The cosets of this equivalence form a group for the concatenation of words. It is called the **free group** over S and is often denoted by $F(S)$. The unit in the group is the coset of the empty word. Another useful point of view is to define $F(S)$ by the following *universal* property: for every function $f : S \to G$ into a group G, there exists a unique group morphism $\varphi : F(S) \to G$ that restricts to f on S (formally, one should assume a function $\iota : S \to F(S)$ and require that $\iota\varphi = f$. It appears that ι must be 1-1, so that S can be assumed a subset of $F(S)$.)

The **conjugate** of a group element v by another group element u is the element $u^{-1}vu$. Let R be a set of words over $S \cup S^{-1}$. The **normal closure** $C(R)$ of R is the smallest normal subgroup of $F(S)$ that contains (the cosets of) R. It is composed of the elements in R, their inverses, all their conjugates and the products of such conjugates. We can now define the group with **combinatorial presentation** $\langle S \mid R\rangle$ as the quotient group

$$\langle S \mid R\rangle = F(S)/C(R)$$

This is the group of words over $S \cup S^{-1}$ where two words are identified if one can be transformed into the other by

1. the insertion or removal of factors ss^{-1} or $s^{-1}s$,

2. or, the insertion or removal of words in R or their inverses.

The elements of R are called **relations**. They are often denoted by equalities as in $r = 1$ or $a = b$ to mean $r \in R$ or $ab^{-1} \in R$. Note that every group G has a combinatorial presentation given by $S = G$ and by taking for R the multiplication table of the group. A group may or may not have a finite presentation.

Given a combinatorial presentation $\langle S \mid R \rangle$, one obtain a presentation of an isomorphic group by either one of the following **Tietze transformations**

- (add consequences) add a subset of $C(R)$ to the relations,
- (remove redundancies) remove a subset of relations that are consequences of the others,
- (add generators) add a new symbol $s \notin S$ and a relation $s = w$ for some $w \in F(S)$. This can be done with an infinite set of new symbols,
- (remove generators) remove a symbol s of S involved in a unique relation $s = w$ where w does not use s or s^{-1}. This can be done with an infinite set of symbols.

Conversely, it is a theorem that any combinatorial presentations of isomorphic groups can be transformed one into the other by a sequence of Tietze transformations. If the group have finite presentations we can use a finite sequence of elementary Tietze transformations.

Given a combinatorial presentation $\langle S \mid R \rangle$ and a group G, Dick's theorem states that a function $f : S \to G$ extends to a morphism $\langle S \mid R \rangle \to G$ if and only if for every $r \in R$, we have $f(r) = 1_G$ where $f(r)$ is the natural extension of f to words.

There are two important constructions to obtain new groups from old ones. Given two groups A, B and two embeddings (i.e., monomorphisms) $\varphi : C \hookrightarrow A$ and $\psi : C \hookrightarrow B$ of a group C, there is a group $A \star_C B$ which is in some sense the gluing of A and B along C. This group is called the **free product with amalgamation** of A and B. It can be defined by a universal property. More simply, if $A = \langle S_A \mid R_A \rangle$ and $B = \langle S_B \mid R_B \rangle$, then

$$A \star_C B = \langle S_A \cup S_B \mid R_A \cup R_B \cup \{\varphi(c) = \psi(c) \mid c \in C\} \rangle$$

The other important construction involves two isomorphic subgroups $\varphi_1 : C \hookrightarrow A$, $\varphi_2 : C \hookrightarrow A$ of a group A. The **HNN extension** $A\star_C$ is a new group containing A where $\varphi_2(C)$ identifies with a conjugate $t^{-1}\varphi_1(C)t$ for some new element $t \notin A$. If $A = \langle S \mid R \rangle$ then

$$A\star_C = \langle S \cup \{t\} \mid R \cup \{\varphi_2(c) = t^{-1}\varphi_1(c)t \mid c \in C\} \rangle$$

Every coset of words in a free group has a canonical representative; this is the unique **reduced word** in the coset, i.e., the unique word without ss^{-1} or $s^{-1}s$ factor. Having canonical representative is a nice property when one needs to compare cosets. Although there is in general no canonical representatives for a combinatorial presentation, free product with amalgamations and HNN-extensions have some kind of normal forms that allow to deduce structural properties. In particular, the inclusions of generators induce inclusions of A, B and C in $A \star_C B$ and of A, C in $A\star_C$.

Appendix C

Counter-Examples to Dey and Guha's Approach

In a first stage, Dey and Guha [DG99] obtain a term product representation of c and d as in the present Lemma 5.2.3. Suppose $f_H|_A = a_1 a_2 \cdots a_{4g}$, then a term $a_i a_{i+1} \cdots a_j$ is denoted (i,j). This term is equivalent in $\langle A\,;\, f_H|_A\rangle$ to the complementary term $a_{i-1}^{-1} a_{i-2}^{-1} \cdots a_{j+1}^{-1}$ *going backward* along $f_H|_A$. This complementary term is denoted $\overline{(i-1, j+1)}$. The **length** $|(i,j)|$ of a term (i,j) is the length of the sequence $a_i a_{i+1} \cdots a_j$. The length of a complementary term is defined analogously, so that $|(i,j)| + |\overline{(i-1,j+1)}| = 4g$. The **length** of a product of (possibly complementary) terms is the sum of the lengths of its terms. Let us rename the above term and complementary term as respectively a **forward term** and a **backward term**. A term will now designate either a forward or backward term. Note that a term being equivalent to its complementary term, we may use a forward or backward term in place of each term. By convention, we will write a term in backward form only if it is strictly shorter than its complementary forward term. This convention will be implicitly assumed in this section and corresponds to the notion of *rectified* term in [DG99].

Let us say that a product $t_1 t_2$ of two terms

- 0-**reacts** if $t_1 t_2 = 1$, the unit element in the group $\langle A\,;\, f_H|_A\rangle$,
- 1-**reacts** if $t_1 t_2 = t$ in $\langle A\,;\, f_H|_A\rangle$, for a term t such that $|t| \le |t_1| + |t_2|$, and
- 2-**reacts** if $t_1 t_2 = t_1' t_2'$ in $\langle A\,;\, f_H|_A\rangle$, for two terms t_1', t_2' such that $|t_1'| + |t_2'| < |t_1| + |t_2|$.

The aim of Dey and Guha is to apply reactions to a given term product in order to reach a canonical form where no two consecutive terms react in that form. For this, they define a function `apply` that recursively applies reductions to a product of terms. This function is in turn called by another function `canonical`, supposed to produce a canonical form.

The following claim appears as points 2 and 3 in Lemma 4 of [DG99] and aims at showing that the function `apply` does terminate.

Let u, v, w be 3 terms such that uv does not react. If vw 1-reacts or 2-reacts with $vw = v'$ or $vw = v'w'$ (and $v'w'$ does not react), then uv' does not 1-react.

The non-existence of such 1-reactions is essential in the proof that the function `canonical` indeed returns a canonical form [DG99, Prop. 7]. However, this claim is false as demonstrated by the following examples. Consider a genus 2 surface with $f_H|_A = abcda^{-1}b^{-1}c^{-1}d^{-1}$. Put $u = (2,4)$, $v = \overline{(1,7)}$, and $w = (7,8)$. Then $uv = bcd \cdot a^{-1}dc$ does not react and $vw = a^{-1}dc \cdot c^{-1}d^{-1}$ 1-reacts, yielding $v' = a^{-1}$. But uv' 1-reacts, in contradiction with the claim, since $uv' = bcd \cdot a^{-1} = (2,5)$. Likewise, if we now set $u = (2,4)$, $v = \overline{(1,8)}$ and $w = \overline{(4,2)}$, we have: uv does not react, vw 2-reacts, yielding $v'w' = (5,5) \cdot \overline{(3,2)}$, and uv' 1-reacts, in contradiction with the claim, since $uv' = bcd \cdot a^{-1} = (2,5)$.

Define the **expanded word** of a term product as the word in the elements of A (and their inverses) obtained by replacing each term in the product with the corresponding factor of $f_H|_A$ or $(f_H|_A)^{-1}$. Again, $f_H|_A$ and $(f_H|_A)^{-1}$ should be considered cyclically. Call a product of terms **stable** if no two consecutive terms react. Another important claim [DG99, Lem. 8] states that

The expanded word of a stable product of terms does not contain a factor of length $2g+1$ that is also a factor of $f_H|_A$ or $(f_H|_A)^{-1}$.

This claim is used to prove that the (supposed) canonical form of a product is equivalent to 1 if and only if it is the empty product [DG99, Prop. 6]. However this claim is again false as demonstrated by the following example. Consider the same genus 2 surface as in the previous examples. Then the product $\overline{(1,7)} \cdot (2,4) \cdot \overline{(1,7)} = cba \cdot bcd \cdot a^{-1}dc$ is stable and contains the factor $a \cdot bcda^{-1}$ of length $2g+1 = 5$ that is also a factor of $(f_H|_A)$.

Finally, the canonical form defined by Dey and Guha is not canonical. By definition of the `canonical` function in [DG99, p. 314], a stable (rectified) product w is canonical, *i.e.*, `canonical`$(w) = w$. Using the same genus 2 surface as before, consider the products $w_1 = \overline{(8,6)} \cdot \overline{(8,6)} = dcb \cdot dcb$ and $w_2 = (1,4) \cdot (2,5) = abcd \cdot bcda^{-1}$. It is easily seen that none of these products react. It follows that `canonical`$(w_i) = w_i$, $i = 1,2$. However $w_1 = w_2$ in $\langle A; f_H|_A \rangle$. Indeed, since $\overline{(8,6)} = (1,5)$ in $\langle A; f_H|_A \rangle$, we have

$$w_1 = abcda^{-1} \cdot abcda^{-1} = abcd \cdot bcda^{-1} = w_2$$

This contradicts the fact that an element of $\langle A; f_H|_A \rangle$ can be expressed as a unique canonical product of terms. In particular, Proposition 7 in [DG99] is wrong.

The method proposed in [DG99] seems bound to fail: the above counterexamples[1] show that there are not enough rules to ensure that stable products are canonical forms, but adding more rules would make more difficult to keep under check chains of reactions which are already out of control. Another issue is that for the end of the comparison algorithm to be usable, appending a term to a stable product should not trigger chains of reaction deep down the stack. This adds even more constraints on the reaction rules, especially as lemma 13 in [DG99] does not seem to help. Indeed, the function `good_conjugate` introduced in the lemma is supposed to reinforce the function `reduced_conjugate` defined in [DG99, p. 318]. From its description, `good_conjugate` transforms a canonical product w in $\tilde{w}$ by inserting some term product $c \cdot c^{-1}$ into w before computing $\tilde{\tilde{w}} =$ `canonical`$(\tilde{w})$ and `reduced_conjugate`$(\tilde{\tilde{w}})$. If `canonical` were canonical then $\tilde{\tilde{w}}$ and w would be equal as products of terms; it

[1] The counterexamples easily generalize to genus $g > 2$ orientable surfaces with $f_H|_A = a_1a_2 \cdots a_{2g}a_1^{-1}a_2^{-1} \cdots a_{2g}^{-1}$. Similar counterexamples for non-orientable surfaces can also be found starting with the product of squares as a (canonical) relator.

follows that the functions `good_conjugate` and `reduced_conjugate` would have exactly the same effect.

Personal publications

[1] Dominique Attali, Marc Glisse, Samuel Hornus, Francis Lazarus, and Dmitriy Morozov. Persistence-sensitive simplification of functions on surfaces in linear time. In *TOPOINVIS'09 (Topological Methods In Data Analysis and Visualization)*, 23-24 Feb. 2009. TOPOINVIS'09 (Topological Methods In Data Analysis and Visualization), Snowbird.

[2] Vincent Borrelli, Saïd Jabrane, Francis Lazarus, and Boris Thibert. Flat tori in three dimensional space and convex integration. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 109(19):7218–7223, May 2012. 177, 194

[3] Vincent Borrelli, Saïd Jabrane, Francis Lazarus, and Boris Thibert. The Nash-Kuiper process for curves. In *Actes du séminaire de théorie spectrale et géométrie*, 2012.

[4] Vincent Borrelli, Saïd Jabrane, Francis Lazarus, and Boris Thibert. Isometric embeddings of the square flat torus in ambient space. *Ensaios Matemáticos*, 24:1–91, 2013. 177, 191

[5] Vincent Borrelli, Francis Lazarus, and Boris Thibert. Les fractales lisses. *Pour la Science*, (425):20–27, March 2013. 177

[6] Sergio Cabello, Éric Colin de Verdière, and Francis Lazarus. Finding shortest non-trivial cycles in directed graphs on surfaces. In *26th Annual ACM Symposium on Computational Geometry*, pages 156–165, 2010.

[7] Sergio Cabello, Éric Colin de Verdière, and Francis Lazarus. Finding cycles with topological properties in embedded graphs. *SIAM Journal of Discrete Mathematics*, 25(4):1600–1614, 2011.

[8] Sergio Cabello, Éric Colin de Verdière, and Francis Lazarus. Algorithms for the edge-width of an embedded graph. *Comput. Geom. Theory Appl.*, 45(5-6):215–224, 2012.

[9] Erin W. Chambers, Éric Colin de Verdière, Jeff Erickson, Francis Lazarus, and Kim Whittlesey. Splitting (complicated) surfaces is hard. *Computational Geometry, Theory and Applications*, 41:94–110, 2008.

[10] Erin W. Chambers, Éric Colin de Verdière, Jeff Erickson, Sylvain Lazard, Francis Lazarus, and Shripad Thite. Homotopic Fréchet distance between curves, or walking Your Dog in the Woods in Polynomial Time. *Computational Geometry, Theory and Applications*, 43:295–311, 2010.

[11] Éric Colin de Verdière and Francis Lazarus. Optimal System of Loops on an Orientable Surface. *Discrete & Computational Geometry*, 33(3):507 – 534, March 2005. 130

[12] Éric Colin de Verdière and Francis Lazarus. Optimal Pants Decomposition and Shortest Cycles on an Orientable Surface. *Journal of the ACM*, 54(4):art. 18, July 2007.

[13] André Guéziec, Gabriel Taubin, William Horn, and Francis Lazarus. A framework for streaming geometry in vrml. *IEEE Computer Graphics and Applications*, 19(2):68–78, 1999.

[14] André Guéziec, Gabriel Taubin, Francis Lazarus, and William Horn. Simplicial maps for progressive transmission of polygonal surfaces. In *Third symposium on Virtual reality modeling language (VRML'98)*, pages 25–31. ACM, 1998.

[15] André Guéziec, Gabriel Taubin, Francis Lazarus, and William Horn. Cutting and stitching: converting sets of polygons to manifold surfaces. *IEEE Transactions on Visualization and Computer Graphics*, 7(2):136–151, 2001.

[16] Francis Lazarus. *Courbes, cylindres et métamorphoses pour l'image de synthèse*. Thèse de doctorat en sciences, Université PARIS VII, December 1995.

[17] Francis Lazarus. Smooth interpolation between two polylines in space. *Computer Aided Design*, 29(3):189–196, November 1997.

[18] Francis Lazarus. Topologie combinatoire et algorithmique. Notes de cours. `http://www.gipsa-lab.fr/~francis.lazarus/Enseignement/geoAlgo.html`, 2012. 126

[19] Francis Lazarus, Sabine Coquillart, and Pierre Jancène. Axial deformations: an intuitive deformation technique. *Computer Aided Design*, 26(8):607–613, August 1994.

[20] Francis Lazarus, Michel Pocchiola, Gert Vegter, and Anne Verroust. Computing a Canonical Polygonal Schema of an Orientable Triangulated Surface. In *Proc. of the 17th Annual Symposium on Computational Geometry*, pages 80–89, 2001. 130, 141, 143, 204

[21] Francis Lazarus and Julien Rivaud. On the homotopy test on surfaces. preprint. arXiv : 1110.4573v2 [cs.CG], 2011. 147

[22] Francis Lazarus and Julien Rivaud. On the homotopy test on surfaces. In *Proceedings of the 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 440–449, 2012. 126, 147

[23] Francis Lazarus and Anne Verroust. Feature-based shape transformation for polyhedral objects. In *Fifth Eurographics Workshop on Animation and Simulation*, Oslo, September 1994. Eurographics association.

[24] Francis Lazarus and Anne Verroust. Metamorphosis of cylinder-like objects. *International Journal of Visualization and Computer Animation*, 8(3):131–146, July 1997.

[25] Francis Lazarus and Anne Verroust. 3D metamorphosis: a survey. *The Visual Computer*, 8(3):131–146, 1998.

[26] Francis Lazarus and Anne Verroust. Level set diagrams of polyhedral objects. In *fifth ACM symposium on Solid modeling and applications*, 1999.

[27] Julien Rivaud and Francis Lazarus. On the homotopy test on surfaces with boundaries. In *28th European Workshop on Computational Geometry (EUROCG)*, pages 189–192, Perugia, Italy, March 2012. 122

[28] Gabriel Taubin, André Guéziec, William Horn, and Francis Lazarus. Progressive forest split compression. In *Siggraph'98 Conference Proceedings*, pages 123–132. ACM, 1998.

[29] Gabriel Taubin, William Horn, Francis Lazarus, and Jarek Rossignac. Geometry coding and vrml. *Proceedings of the IEEE*, 86(6):1228–1243, 1998.

[30] Anne Verroust and Francis Lazarus. Extracting skeletal curves from 3d scattered data. *The Visual Computer*, 16:15–25, 2000.

Bibliography

[AG95] Helmut Alt and Michael Godau. Computing the Fréchet distance between two polygonal curves. *Int. J. Comput. Geometry Appl.*, 5(1 & 2):75–91, 1995. 203

[Ago76] Max K. Agoston. *Algebraic Topology: A First Course*. Monographs and Textbooks in Pure and Applied Mathematics. Marcel Dekker, Inc, New York, 1976. 3

[AHT02] Ian Agol, Joel Hass, and William Thurston. 3-manifold knot genus is NP-complete. In ACM, editor, *Proceedings of the Thiry-Fourth Annual ACM Symposium on Theory of Computing*, pages 761–766. ACM Press, 2002. 129

[AHU74] Alfred V. Aho, John E. Hopcroft, and Jeffrey D. Ullman. *The Design an Analysis of Computer Algorithm*. Addison Wesley, 1974. 5, 160

[AP50] Aleksandr D. Aleksandrov and Alekseĭ V. Pogorelov. Uniqueness of convex surfaces of revolution. *Mat. Sb.*, 26(2):183–204, 1950. 179

[Big94] Norman L. Biggs. *Algebraic graph theory*. Cambridge University Press, second edition edition, 1994. 7

[Bil13] Erwan Biland. Le langage des catégories ii. *Quadrature*, 88:16–21, 2013. 67

[BJLT12] Vincent Borrelli, Saïd Jabrane, Francis Lazarus, and Boris Thibert. Flat tori in three dimensional space and convex integration. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 109(19):7218–7223, May 2012. 177, 194

[BJLT13] Vincent Borrelli, Saïd Jabrane, Francis Lazarus, and Boris Thibert. Isometric embeddings of the square flat torus in ambient space. *Ensaios Matemáticos*, 24:1–91, 2013. 177, 191

[BL92] C. Paul Bonnington and Charles H.C. Little. The classification of combinatorial surfaces using 3-graphs. *The Australasian Journal Of Combinatorics*, 5:87–102, 1992. 96

[BL94] C. Paul Bonnington and Charles H.C. Little. A combinatorial generalisation of the Jordan curve theorem. *The Australasian Journal Of Combinatorics*, 9:179–199, 1994. 116

[BL95] C. Paul Bonnington and Charles H.C. Little. *The Foundations of Topological Graph Theory*. Springer-Verlag, 1995. 3

[BLT13] Vincent Borrelli, Francis Lazarus, and Boris Thibert. Les fractales lisses. *Pour la Science*, (425):20–27, March 2013. 177

[BLW98] Norman L. Biggs, E. Keith Lloyd, and Robin J. Wilson. *Graph theory 1736-1936*. Oxford University Press, 1998. 14

[Bra21] Thomas R. Brahana. Systems of circuits on 2-dimensional manifolds. *Ann. Math.*, 23(2):144–168, 1921. 95, 141, 202

[BS85] Robin P. Bryant and David Singerman. Foundations of the theory of maps on surfaces with boundary. *Quart. J. Math. Oxford*, 36(2):17–41, 1985. 39, 46, 63, 66, 67, 78, 79, 116

[BW09] Lowell W. Beineke and Robin J. Wilson, editors. *Topics in Topological Graph Theory*, volume 128 of *Encyclopedia of Mathematics and its Applications*. Cambrigde university Press, 2009. 5, 23, 33, 38, 46

[CCE13] Sergio Cabello, Erin W. Chambers, and Jeff Erickson. Multiple-source shortest paths in embedded graphs. *SIAM Journal on Computing*, 42(4):1542–1571, 2013. 21, 205

[CdV10] Éric Colin de Verdière. Shortest cut graph of a surface with prescribed vertex set. In *Proceedings of the 18th Annual European Symposium on Algorithms*, number 6347 in Lecture Notes in Computer Science, pages 100–111, 2010. 122, 125

[CdV12] Éric Colin de Verdière. *Topological algorithms for graphs on surfaces*. Habilitation à diriger des recherches, École normale supérieure, 2012. Available at `http://www.di.ens.fr/~colin/`. 3

[CdVL05] Éric Colin de Verdière and Francis Lazarus. Optimal System of Loops on an Orientable Surface. *Discrete & Computational Geometry*, 33(3):507 – 534, March 2005. 130

[CdVL10] Sergio Cabello, Éric Colin de Verdière, and Francis Lazarus. Output-sensitive algorithm for the edge-width of an embedded graph. In *26th Annual ACM Symposium on Computational Geometry*, pages 147–155, 2010. 152

[CE10] Éric Colin de Verdière and Jeff Erickson. Tightening non-simple paths and cycles on surfaces. *SIAM Journal on Computing*, 39(8):3784–3813, 2010. 3, 130, 147, 148, 152, 153, 154, 155, 156, 157, 158, 161, 203

[CLRS09] Thomas Cormen, Charles Leiserson, Ronald Rivest, and Clifford Stein. *Introduction to Algorithms*. MIT Press, third edition, 2009. 14, 133, 174

[CSEH07] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of Persistence Diagrams. *Discrete & Computational Ceometry*, 37(1):103–120, 2007. 207

[CSEH08] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Extending persistence using poincaré and lefschetz duality. *Foundations of Computational Mathematics*, 2008. 207

[CV36] Stephan E. Cohn-Vossen. Die verbiegung von flachen im grossen. *Fortschr. Math. Wiss*, 1:33–76, 1936. 179

[DA85] Bernard Domanski and Michael Anshel. The complexity of dehn's algorithm for word problems in groups. *Journal of Algorithms*, 6:543–549, 1985. 149

[DEG98] Tamal K. Dey, Herbert Edelsbrunner, and Sumanta Guha. Computational Topology. In B. Chazelle, J.E. Goodman, and R. Pollack, editors, *Advances in Discrete and Computational Geometry*. Contemporary Mathematics, AMS, 1998. 38

[DG99] Tamal K. Dey and Sumanta Guha. Transforming Curves on Surfaces. *Journal of Computer and System Sciences*, 58(2):297–325, 1999. 147, 153, 206, 211, 212

[DH07] Max Dehn and Poul Heegaard. *Encyklopädie der mathematischen Wissenschaften mit Einschluss ihrer Anwendungen*, volume vol. III.AB, chapter Analysis situs, chapter 3, pages 153–220. B.G. Teubner, Leipzig, 1907. 95

[DM68] P. H. Doyle and D. A. Moran. A short proof that compact 2-manifolds can be triangulated. *Inventiones Mathematicae*, 5(2):160–162, 1968. 95

[DPeK82] Narsingh Deo, Gurpur M. Prabhu, and Mukkai S. et Krishnamoorty. Algorithms for generating fundamental cycles in a graph. *ACM Trans. Math. Software*, 8:26–42, 1982. 19

[DS95] Tamal K. Dey and Haijo Schipper. A new technique to compute polygonal schema for 2-manifolds with application to null-homotopy detection. *Discrete and Computational Geometry*, 14:93–110, 1995. 38, 148

[ea94] Bill Thurston et al. Outside in, 1994. 189

[Edm60] Jack R. Edmonds. A combinatorial representation of polyhedral surfaces. *Notices of the American Society*, 7:646, 1960. 37

[EH08] Herbert Edelsbrunner and John Harer. Persistent homology — a survey. In J. E. Goodman, J. Pach, and R. Pollack, editors, *Twenty Years After*. AMS, 2008. 206

[EHP04] Jeff Erickson and Sariel Har-Peled. Optimally Cutting a Surface into a Disk. *Discrete & Computational Geometry*, 31(1):37 – 59, January 2004. 124

[ELZ00] Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological Persistence and Simplification. In *IEEE Symposium on Foundations of Computer Science*, pages 454–463, 2000. 206

[EM02] Yakov Eliashberg and Nikolai Mishachev. *Introduction to the h-principle*, volume 48 of *Graduate Studies in Mathematics*. A.M.S., Providence, 2002. 185

[EMP06] Herbert Edelsbrunner, Dmitriy Morozov, and Valerio Pascucci. Persistence-Sensitive Simplification of Functions on 2-Manifolds. In *22nd Annual ACM Symposium on Computational Geometry*, pages 127–134, 2006. 207

[EN13] Jeff Erickson and Amir Nayyeri. Tracing compressed curves in triangulated surfaces. *Discrete & Computational Geometry*, 49(4):823–863, 2013. 130

[Epp03] David Eppstein. Dynamic generators of topologically embedded graphs. In *Proceedings of the fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA-03)*, pages 599–608. ACM Press, jan 2003. 120, 152, 206

[Eps66] David B. A. Epstein. Curves on 2-manifolds and isotopies. *Acta Mathematica*, 115:83–107, 1966. 115

[Eri11] Jeff Erickson. Shortest non-trivial cycles in directed surface graphs. In *Proc. of the 27th Annual ACM Symposium on Computational Geometry*, pages 236–243, 2011. 206

[Eri12] Jeff Erickson. Combinatorial optimization of cycles and bases. In Afra Zomorodian, editor, *Advances in Applied and Computational Topology.*, Proceedings of Symposia in Applied Mathematics, 2012. 122

[EW05] Jeff Erickson and Kim Whittelsey. Greedy optimal homotopy and homology generators. In *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1038–1046, 2005. 122, 124, 128

[EW13] Jeff Erickson and Kim Whittelsey. Transforming curves on surfaces redux. In *Proc. of the 24rd Annual ACM-SIAM Symposium on Discrete Algorithms*, 2013. 126, 149, 206

[Fal03] Kenneth Falconer. *Fractal Geometry*. Wiley, second edition edition, 2003. 193

[FW94] Michael L. Fredman and Dan E. Willard. Trans-dichotomous algorithms for minimum spanning trees and shortest paths. *Journal of Computer System Sciences*, 48(3):533–551, 1994. 124

[GGD12] Ernesto Girondo and Gabino González-Diez. *Introduction to compact Riemann Surfaces and dessins d'enfants*. LMS, 2012. 39, 42

[GH02] Alexander Golynski and Joseph D. Horton. A Polynomial-Time Algorithm to Find the Minimal Cycle Basis of a Regular Matroid. In *8th Scandinavian Workshop on Algorithm Theory*, 2002. 22

[Gle01] Petra Manuela Gleiss. *Short Cycles*. PhD thesis, University of Vienna, 2001. `http://www.tbi.univie.ac.at/papers/Abstracts/pmg-diss.pdf`. 22

[GR01] Christopher Godsil and Gordon Royle. *Algebraic graph theory*. Springer-Verlag, 2001. 7

[Gre60] Martin D. Greendlinger. On Dehn's algorithm for the conjugacy and word problems with applications. *Comm Pure Appl. Math.*, 13:641–677, 1960. 147

[Gro86] Mikhail Gromov. *Partial Differential Relations*. Springer-Verlag, 1986. 182

[GS90] Steve M. Gersten and Hamish B. Short. Small cancellation theory and automatic groups. *Inventiones mathematicae*, 102:305–334, 1990. 149, 206

[GT87] Jonathan L. Gross and Thomas W. Tucker. *Topological graph theory*. Dover, reprint 2001 from wiley edition, 1987. 3, 5, 23, 96

[Hak61] Wolfgang Haken. Theorie der normalflächen: Ein isotopiekriterium für den kreisknoten. *Acta Mathematica*, 105:245–375, 1961. 129

[Hal07a] Thomas C. Hales. The Jordan Curve Theorem, Formally and Informally. *The American Mathematical Monthly*, 114:882–894, dec 2007. 116

[Hal07b] Thomas C. Hales. Jordan's Proof of the Jordan Curve Theorem. *STUDIES IN LOGIC, GRAMMAR AND RHETORIC*, 10(23):45–60, 2007. 116

[Hat02] Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 2002. `http://www.math.cornell.edu/~hatcher/AT/ATpage.html`. 17, 108, 115

[Hef91] Lothar Heffter. über das problem der nachbargebiete. *Math. Ann*, 38:477–508, 1891. 37

[Hef98] Lothar Heffter. über metacyklische gruppen und nachbarconfigurationen. *Math. Ann*, 50:261–268, 1898. 37

[HM93] David Hartvigsen and Russel Mardon. When Do Short Cycles Generate the Cycle Space. *Journal of Combinatorial Theory, Series B*, 57:88–99, 1993. 19

[HNW91] Ernst Hairer, Syvert Paul Nørsett, and Gerhard Wanner. *Solving ordinary differential equations*, volume 2. Springer, 1991. 195

[Hol00] Derek F. Holt. Word-hyperbolic groups have real-time word problem. *International Journal of Algebra and Computation*, 10(2):221–227, 2000. 149

[Hor87] Joseph D. Horton. A Polynomial-Time Algorithm to Find the Shortest Cycle Basis of a Graph. *SIAM Journal of Computing*, 16(2):358–366, 1987. 19, 21

[HR01] Derek F. Holt and Sarah Rees. Solving the word problem in real time. *Journal London Math Soc.*, 63:623–639, 2001. 149

[HS85] Joel Hass and Peter Scott. Intersections of curves on surfaces. *Israel Journal of Mathematics*, 51(1-2):90–120, 1985. 151

[III04] Charles F. Miller III. Combinatorial group theory. , 2004. 209

[IPS82] Alon Itai, Christos H Papadimitriou, and Jayme Luiz Szwarcfiter. Hamilton paths in grid graphs. *SIAM Journal on Computing*, 11(4):676–686, 1982. 203

[JS78] Gareth A. Jones and David Singerman. Theory of maps on orientable surfaces. *Proc. London Math. Soc.*, 37(2):273–307, 1978. 43, 67, 116

[Kap05] B. F. Kaplan. Riemann's geometric ideas. *American Mathematical Monthly*, 121(1):79–86, Jan 2005. 179

[KMMP04] Telikepalli Kavitha, Kurt Mehlhorn, Dimitrios Michail, and Katarzyna Paluch. A Faster Algorithm for Minimum Cycle Basis of Graphs. In *ICALP*, 2004. 22

[KMMP08] Telikepalli Kavitha, Kurt Mehlhorn, Dimitrios Michail, and Katarzyna E. Paluch. An Õ(m2 n) algorithm for minimum cycle basis of graphs. *Algorithmica*, 52:333–349, 2008. 22

[Kne29] Hellmuth Kneser. Geschlossene flächen in dreidimensionalen mannigfaltigkeiten. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 38:248–259, 1929. 129

[Laz12a] Francis Lazarus. Géométrie algorithmique. Notes de cours. `http://www.gipsa-lab.fr/~francis.lazarus/Enseignement/geoAlgo.html`, 2012. 115

[Laz12b] Francis Lazarus. Topologie combinatoire et algorithmique. Notes de cours. `http://www.gipsa-lab.fr/~francis.lazarus/Enseignement/geoAlgo.html`, 2012. 126

[Lin82] Sóstenes Lins. Graph-encoded maps. *Journal of Combinatorial Theory, Series B*, 32:171–181, 1982. 152

[Lit88] Charles H. C. Little. Cubic combinatorial maps. *Journal of Combinatorial Theory, Series B*, 44:44–63, 1988. 116

[LPVV01] Francis Lazarus, Michel Pocchiola, Gert Vegter, and Anne Verroust. Computing a Canonical Polygonal Schema of an Orientable Triangulated Surface. In *Proc. of the 17th Annual Symposium on Computational Geometry*, pages 80–89, 2001. 130, 141, 143, 204

[LR11] Francis Lazarus and Julien Rivaud. On the homotopy test on surfaces. preprint. arXiv : 1110.4573v2 [cs.CG], 2011. 147

[LR12] Francis Lazarus and Julien Rivaud. On the homotopy test on surfaces. In *Proceedings of the 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 440–449, 2012. 126, 147

[LS77] Roger C. Lyndon and Paul E. Shupp. *Combinatorial Group Theory*, volume 89 of *A Series of Modern Surveys in Mathematics*. Springer-Verlag, 1977. 103, 114, 115, 209

[LZ04] Sergei K. Lando and Alexander K. Zvonkin. *Graphs on surfaces and their applications*. Springer-Verlag, 2004. 44, 63

[Mar04] Martin Mareš. Two linear time algorithms for mst on minor closed graph classes. *Archivum Mathematicum (Brno)*, 40:315–320, 2004. 124

[Mas77] William S. Massey. *Algebraic Topology: An Introduction*, volume 56 of *Graduate Texts in Mathematics*. Springer Verlag, 1977. 23, 95

[Mas91] William S. Massey. *A Basic Course in Algebraic Topology*, volume 127 of *Graduate Texts in Mathematics*. Springer Verlag, 1991. 115, 148, 149, 152

[MM09] Kurt Mehlhorn and Dimitrios Michail. Minimum Cycle Bases: Faster and Simpler. *ACM Transactions on Algorithms*, 2009. 22

[Moh01] Bojan Mohar. Graph minors and graphs on surfaces. Invited talk at the 18th British Combinatorial Conference, july 2001. 53

[Moi77] Edwin E. Moise. *Geometric topology in dimensions 2 and 3*, volume 47 of *Graduate Texts in Math*. Springer-Verlag, 1977. 95

[MT01] Bojan Mohar and Carsten Thomassen. *Graphs on Surfaces*. John Hopkins University Press, 2001. 5, 38, 39, 53, 96

[MW02] Jonathan P. McCammond and Daniel T. Wise. Fans and ladders in small cancellation theory. *Proceedings of the London Mathematical Society*, 84:599–644, 2002. 103

[Nas54] John F. Nash. C^1-isometric imbeddings. *Ann. of Math.*, 60(3):383–396, 1954. 179, 190

[Nas66] John F. Nash. Analycity of the solutions of implicit functions problems with analysic data. *Ann. of Math.*, 84(2):345–355, 1966. 180

[Nir53] Louis Nirenberg. The weyl and minkowski problems in differential geometry in the large. *Comm. Pure Appl. Math.*, 6:337–394, 1953. 179

[Pog64] Alekseĭ V. Pogorelov. Some results on surface theory in the large. *Advances in Math.*, 1(2):191–264, 1964. 179

[Rad25] Tibor Radó. über den Begriff der Riemannschen Fläche. *Acta Litt. Sci. Szeged*, 2:101–121, 1925. 95

[Ran96] Andrew A. Ranicki. *The Hauptvermutung Book: A Collection of Papers on the Topology of Manifolds (K-Monographs in Mathematics)*, chapter On the Hauptvermutung, pages 3–31. Springer, 1996. 3

[Rie67] Bernhard Riemann. Über die hypothesen welche der geometrie zu grunde liegen. *Abhandlungen der Königlichen Gesellschaft der Wissenschaften zu Göttingen*, 13, 1967. 179

[RL12] Julien Rivaud and Francis Lazarus. On the homotopy test on surfaces with boundaries. In *28th European Workshop on Computational Geometry (EUROCG)*, pages 189–192, Perugia, Italy, March 2012. 122

[Rob99] Vanessa Robins. Towards computing homology from finite approximations. *Topology proceedings*, 24, 1999. 206

[Ser77] Jean-Pierre Serre. *Trees*. Springer-Verlag, 1977. translation from the French by J. Stillwell. 8, 26, 115

[SSS08] Marcus Schaefer, Eric Sedgwick, and Daniel Stefankovic. Computing Dehn twists and geometric intersection numbers in polynomial time. In *CCCG*, pages 111–114, 2008. 129

[SSv02] Marcus Schaefer, Eric Sedgwick, and Daniel Štefankovič. Algorithms for normal curves and surfaces. In *8th Annual International Conference on Computing and Combinatorics (COCOON 2002)*, August 2002. 129

[Sta83] Saul Stahl. A combinatorial analog of the jordan curve theorem. *Journal of Combinatorial Theory, Series B*, 35(1):28–38, 1983. 38, 116

[Sti87] John Stillwell. *Papers on group theory and topology*. Springer-Verlag, New York, 1987. 95, 149, 174

[Sti93] John Stillwell. *Classical topology and combinatorial group theory*. Springer-Verlag, New York, 1993. 95, 149, 155

[Str90] Ralph Strebel. *Sur les groupes hyperboliques d'après Mikhael Gromov*, chapter Small cancellation groups, pages 227–276. Birkhäuser, 1990. 103

[Tar83] Robert E. Tarjan. *Data Structures and Network Algorithms*. Number 44 in CBMS-NFS Regional conference series in applied mathematics. SIAM, 1983. 123

[Tho90] Carsten Thomassen. Embeddings of graphs with no short noncontractible cycles. *Journal of Combinatorial Theory, Series B*, 48:155–177, 1990. 203, 205

[Tho92] Carsten Thomassen. The Jordan-Schönflies Theorem and the Classification of Surfaces. *American Mathematical Monthly*, pages 116–130, feb 1992. 95, 116

[TR98] Gabriel Taubin and Jarek Rossignac. Geometric compression through topological surgery. *ACM Transactions on Graphics*, 17:84–115, 1998. 199

[Tut73] William T. Tutte. What is a map? In F. Harary, editor, *New Directions in the Theory of Graphs*, pages 309–325, New-York, 1973. Academic Press. 38, 63, 95

[Tut79] William T. Tutte. Combinatorial Oriented Maps. *Can. J. Math.*, XXXI(5):986–1004, 1979. 38, 116

[Tut01] William T. Tutte. *Graph Theory*, volume 21 of *Encyclopedia of mathematics and its applications*. Cambridge University Press, 2001. first edition 1984. 63, 95, 116

[Tve80] Helge Tverberg. A Proof of the Jordan Curve Theorem. *Bull. London Math. Soc.*, 12:34–38, 1980. 115

[vCGKZ98] D.J. von Collins, R.I. Grigorchuk, P.F. Kurchanov, and H. Zieschang. *Combinatorial Group Theory and Applications to Geometry*. Springer Verlag, 1998. 4, 115, 209

[Veg97] Gert Vegter. Computational Topology. In Jacob E. Goodman and Joseph O'Rourke, editors, *Handbook of Discrete and Computational Geometry*. CRC Press, 1997. 3, 38

[VL89] Andrew Vince and Charles H.C Little. Discrete jordan curve theorems. *Journal of Combinatorial Theory, Series B*, 47(3):251–261, 1989. 38, 116

[VY90] Gert Vegter and Chee K. Yap. Computational complexity of combinatorial surfaces. In *Proc. 6th Annu. ACM Sympos. Comput. Geom.*, pages 102–111, 1990. 38, 141, 200

[Zeg05] Abdelghani Zeghib. Histoire des immersions isométriques. In J. Kouneiher, D. Flament, P. Nabonnand, and J. Szczeciniarz, editors, *Géométrie au XXe siècle, Editeurs : J. Kouneiher , D. Flament , P. Nabonnand , J. Szczeciniarz*. Hermann, 2005. 180

[ZVC80] Heiner Zieschang, Elmar Vogt, and Hans-Dieter Coldewey. *Surfaces and Planar Discontinuous Groups*, volume 835 of *Lecture Notes in Mathematics*. Springer Verlag, 1980. 4, 52, 96

[Zvo] Alexander K. Zvonkin. Cartes et dessins d'enfants. `http://www.labri.fr/Perso/~zvonkin`. 38

Index

Abstract. In the first and main part of this document I propose some elements of a theory of graphs and surfaces from the topological and combinatorial point of view. L. Heffter introduced the representation of surfaces by means of permutations at the end of the nineteenth century. This representation is at the origin of combinatorial maps, a very rich concept in relation with the famous Grothendieck's dessins d'enfants. In this manuscript, I am essentially concentrating on topological and combinatorial aspects of maps in connection with the fundamental group or the first homology group. I thus describe efficient computations for their bases, or efficient homotopy and homology tests. Those notions are defined in a purely combinatorial framework without ever resorting to pointset Topology. It appears that a large amount of the published algorithms concerning topological properties of surfaces has never been implemented. My purpose is to provide a formal framework that will hopefully facilitate their implementation.

A second part is devoted to the isometric embedding of a square flat torus. This is a joint work with three other colleagues that had quite a success, with some beautiful images. It is based on an amazing result by John F. Nash; any n-dimensional Riemannian manifold that embeds into $\mathbb{R}^k$ with $k \geq n+1$ can be isometrically embedded into the Euclidean space of the same dimension k in a smooth (C^1) manner. In particular, a flat torus (a locally Euclidean torus whose Gaussian curvature thus vanishes everywhere) can be embedded isometrically into the three dimensional Euclidean space. An easy argument using Gaussian curvature shows that this is actually impossible if the embedding is required to be C^2. Nash's exploit was to prove that this argument could be bypassed with C^1 embeddings. Our contribution was to effectively compute such a C^1 embedding and, doing so, to discover a new geometric structure: the C^1 fractals.

Résumé. Dans la première et principale partie de ce mémoire, je propose une théorie élémentaire des surfaces d'un point de vue combinatoire et topologique. La représentation des surfaces à l'aide de permutations a été introduite par L. Heffter à la fin du XIX[e] siècle. Cette représentation a donné lieu au concept de carte combinatoire, objet d'une richesse sans pareil ayant abouti aux fameux dessins d'enfants de Grothendieck. Dans ce manuscrit je m'intéresse essentiellement aux aspects topologiques et algorithmiques des cartes combinatoires liées au groupe fondamental ou au premier groupe d'homologie. Je décris ainsi comment calculer efficacement des bases de ces groupes ou comment décider si deux courbes sont homotopes ou homologues. Ces notions sont décrites dans un cadre purement combinatoire sans jamais faire appel à des concepts de topologie générale. Il apparaît qu'un nombre conséquent d'algorithmes en lien avec la topologie des surfaces ont été publiés sans être implémentés. Mon objectif est de fournir un cadre formel qui facilite ses implémentations.

Une seconde partie est dédiée au plongement isométrique du tore carré plat. Il s'agit d'une collaboration avec trois collègues qui a obtenu un certain succès. Il en ressort quelques images surprenantes. Ce travail repose sur un incroyable résultat de John F. Nash montrant que toute variété riemannienne de dimension n plongée dans $\mathbb{R}^k$ avec $k \geq n+1$ peut être plongée isométriquement dans l'espace euclidien de même dimension k de manière lisse (C^1). En particulier, un tore plat (un tore localement euclidien dont la courbure gaussienne est donc identiquement nulle) peut être plongé isométriquement dans l'espace euclidien de dimension 3. Un argument simple reposant sur la courbure gaussienne montre cependant qu'il n'existe pas de plongement isométrique C^2 d'un tore plat. L'exploit de Nash a été de prouver que cet argument pouvait être contourné dans le cas C^1. Notre contribution a été de calculer de manière effective un tel plongement C^1 et, ce faisant, de découvrir une nouvelle structure géométrique : les fractales C^1.