# Optimal System of Loops on an Orientable Surface*

Éric Colin de Verdière† Francis Lazarus‡

Submitted to *Discrete & Computational Geometry* August 18, 2003
Revised and resubmitted June 29, 2004

## Abstract

Every compact orientable boundaryless surface $\mathcal{M}$ can be cut along simple loops with a common point $v_0$, pairwise disjoint except at $v_0$, so that the resulting surface is a topological disk; such a set of loops is called a *system of loops* for $\mathcal{M}$. The resulting disk may be viewed as a polygon in which the sides are pairwise identified on the surface; it is called a *polygonal schema*. Assuming that $\mathcal{M}$ is a combinatorial surface, and that each edge has a given length, we are interested in a *shortest* (or *optimal*) system of loops homotopic to a given one, drawn on the vertex-edge graph of $\mathcal{M}$. We prove that each loop of such an optimal system is a shortest loop among all simple loops in its homotopy class. We give an algorithm to build such a system, which has polynomial running time if the lengths of the edges are uniform. As a byproduct, we get an algorithm with the same running time to compute a shortest simple loop homotopic to a given simple loop.

## 1 Introduction

From the classification of surfaces in topology, any compact orientable boundaryless surface $\mathcal{M}$ is, up to homeomorphism, a sphere, a torus, or, more generally, a $g$-torus – a gluing of $g$ tori – for some integer $g$, called the *genus* of the surface. It is a well-known fact that such a surface can be obtained from a polygon by pairwise identifications of its sides; such a polygon is called a *polygonal schema*. We focus on surfaces homeomorphic to a $g$-torus ($g > 0$). A *reduced* polygonal schema of such a surface is a special type of polygonal schema where all the vertices of the polygon get identified to a single point $v_0$, the *basepoint*, on the surface. In that case, the polygonal schema has $4g$ sides; after identification, these sides correspond to a set of $2g$ simple loops through $v_0$, pairwise disjoint except at $v_0$, whose removal transforms the surface into a disk. Such a set of loops is called a *fundamental system* of loops, or *system* of loops for short. See Figure 1 and, for a general reference on this subject, Stillwell [18, Chapter 1.4].

*A preliminary version of this paper was presented at the 43rd Annu. IEEE Symp. Found. Comp. Sci. [5].

†Dépt. d'informatique, École normale supérieure (Paris), France. Eric.Colin.de.Verdiere@ens.fr
‡CNRS, University of Poitiers, France. Francis.Lazarus@sic.univ-poitiers.fr

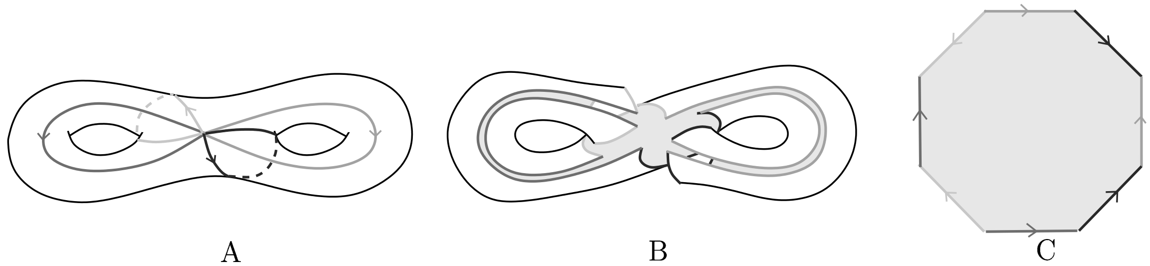


Figure 1: A: a system of loops of a double-torus ($g = 2$). B: the same surface after cutting along the system of loops is a topological disk. C: the same surface as in B, represented in the plane, is a polygonal schema of the double-torus, *i.e.*, a polygon where the sides can be pairwise identified to re-obtain the surface.

We shall consider systems of loops on a combinatorial surface (a topological 2-manifold obtained by assembling finitely many simple polygons) whose vertex-edge graph is weighted. In our combinatorial setting, $v_0$ is a vertex of $\mathcal{M}$, and the loops are closed walks on the vertex-edge graph $G$ of $\mathcal{M}$, based at $v_0$. To mimic the continuous framework, the loops may share edges and vertices of $G$, provided that they can be spread apart on the surface with a thin space so that they become simple and disjoint except at $v_0$.

We describe a conceptually simple, iterative procedure that takes a given system of loops and outputs a shorter homotopically equivalent system. Figure 2 illustrates the principle of the method. We prove that, at the end of the process, each loop is a shortest simple loop in its homotopy class. In particular, the resulting system of loops is *optimal* in the sense that it is as short as possible among all homotopic systems, and any optimal system of loops is made of shortest simple homotopic loops. Furthermore, we implemented our algorithm (see Appendix A) with a running time that is polynomial in the complexity of the surface, the complexity of the input system, and the longest-to-shortest edge ratio of $G$.

We can apply this result to the computation of a shortest simple loop in the homotopy class of a given simple loop $\gamma$: if $\gamma$ is non-separating, extend $\gamma$ to a system of loops on $\mathcal{M}$; after optimization, this system of loops contains the desired loop. Section 6 also describes the case when $\gamma$ is non-contractible and separating.

Let us stress out that these results are *a priori* non obvious. Firstly, the shortest loop homotopic to a simple loop may itself not be simple, as can be seen on Figure 3 (this situation contrasts with cycles, when considering *free homotopy* [4]). As a consequence, computing a shortest system homotopic to a given system of loops cannot be obtained by just searching for the shortest loop homotopic to each loop in the system. Also, even if we are able to compute shortest *simple* homotopic loops, it could still happen that these loops intersect. However, as we show in this paper, it is possible to form a system composed of shortest simple homotopic loops only.

Secondly, there is a natural strategy for computing a shortest loop within a given ho-

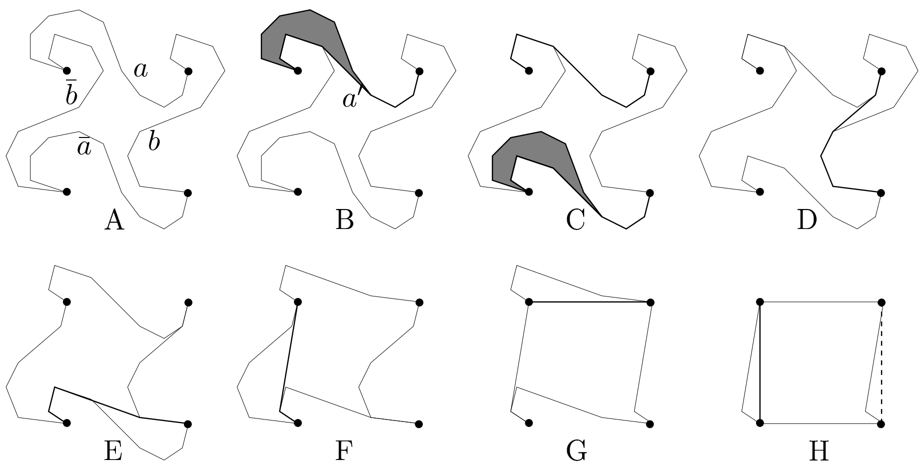


Figure 2: A: A polygonal schema of a torus endowed with the Euclidian metric. Sides $a$ and $\bar{a}$ (resp. $b$ and $\bar{b}$) map to a single loop on the torus. B: The shortest path $a'$ between the endpoints of $a$ is computed *inside* the polygonal schema. C: The grey shaded region between $a$ and $a'$ is cut and pasted on the opposite side $\bar{a}$. D: The same process is repeated with $b$ instead of $a$. E, F, G, H: The process is repeated for each side of the polygonal schema until a minimal configuration (a square) is reached.

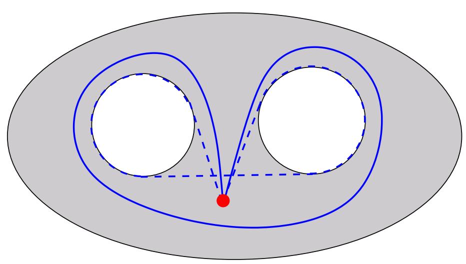

Figure 3: A simple loop (plain thick line) whose shortest homotopic loop (dashed line) self-intersects twice, on a surface homeomorphic to a sphere with three boundaries. A similar example can be built on a double-torus, for instance.

motopy class in $\mathcal{M}$: this problem reduces to computing a shortest path in the universal cover $\tilde{\mathcal{M}}$ of $\mathcal{M}$, between the endpoints of a lift of the input loop. But this algorithm has exponential running time, even if all edges have unit lengths: if the shortest path we are looking for is composed of $k$ edges in $G$, then we should *a priori* visit all vertices at a distance at most $k$ from a lift of the basepoint in the lift of $G$ in $\tilde{\mathcal{M}}$. However, as shown in Appendix B, the number of such vertices can be exponential in $k$ (for surfaces of genus at least 2).

The fact we obtain an algorithm with polynomial running time (in the case of uniform weights) contrasts with the NP-hardness of computing a shortest polygonal schema of a combinatorial surface [9]. This might indicate that the computation of a minimal system of loops, independently of any homotopy class, is a much harder problem.

On a practical level, computing a system of loops on a surface is known to be useful in several problems where a correspondence between the surface and a topological disk needs to be established. Important applications are surface parameterization [10] and texture mapping [15, 17]. Canonical systems (*i.e.*, systems with a canonical ordering of the loops around the basepoint) also allow to construct homeomorphisms between surfaces of the same genus: given a canonical system for each of two such surfaces, it is sufficient to establish a correspondence between their two complementary disks that preserves the order of the loops on their boundary. Brahana's classical proof of the classification theorem for compact 2-manifolds [2] gives an abstract algorithm to transform an arbitrary polygonal schema into a canonical schema using a sequence of elementary cutting and pasting operations. Vegter and Yap [19] and, later, Lazarus *et al.* [13], studied more closely the complexity of the computation of a canonical system on a triangulated surface. In particular, Lazarus *et al.* [13] gave two algorithms with worst-case optimal asymptotic running time complexity. However, their algorithms usually produce jaggy and irregular loops as they do not take into account the geometry of the surface. The previously cited work by Erickson and Har-Peled [9] partly overcomes the geometric aspect, but their method does not allow to specify the combinatorial structure of the polygonal schema (*i.e.*, the way of identifying the sides of the schema); this precludes the simple construction of homeomorphisms between surfaces that would extend a homeomorphism between the complementary disks. In the present paper, the homotopy class is maintained, and thus also the combinatorial structure of the schema.

The structure of this paper is as follows. In Section 2, we review elementary topological notions, and present the framework and our main theorem. Its proof is given in the next three sections. Finally, we discuss the computational issues and give the complexity of our algorithm.

## 2 Framework and Result

### 2.1 Homotopy and system of loops

We begin with some useful standard definitions [18].

Let $M$ be a surface with or without boundary. A *path* is a continuous mapping $p : [0,1] \to M$; its *endpoints* are $p(0)$ and $p(1)$. A path is *simple* if it is one-to-one. If $M$ is pointed with basepoint $v$, a *loop* is a path with both endpoints equal to $v$; a loop is *simple* if its restriction to $[0,1)$ is one-to-one. We say that two loops are *disjoint* if their intersection reduces to $v$.

The *concatenation* of two paths $p_1$ and $p_2$ satisfying $p_1(1) = p_2(0)$ is the path $p_1.p_2$ defined by:

$$p_1.p_2(t) = \begin{cases} p_1(2t) & \text{if } t \leq 1/2 \\ p_2(2t-1) & \text{if } t \geq 1/2 \end{cases}$$

The *inverse* of a path $p$ is the path $\bar{p}$ defined by $\bar{p}(t) = p(1-t)$.

Two paths $p$ and $q$, both with endpoints $a$ and $b$, are *homotopic* if there is a continuous family of paths with endpoints $a$ and $b$ that joins $p$ and $q$. More formally, a *homotopy* between $p$ and $q$ is a continuous mapping $h : [0,1] \times [0,1] \to M$ such that $h(0,.) = p$, $h(1,.) = q$, $h(.,0) = a$, and $h(.,1) = b$. A loop is *contractible* if it is homotopic to the constant loop.

A *system of loops* of a pointed surface $(M, v)$ is an ordered set of disjoint simple loops that cut $M$ into a single topological disk. Every pointed surface that is compact, orientable, and without boundary, admits a system of loops. From the theory of the classification of surfaces [18], it is known that any system on $(M, v)$ is made of $2g$ loops, where $g$ is the *genus* of the surface $M$. See Figure 1.

### 2.2 Length of paths

A *combinatorial surface* is a topological 2-manifold obtained by gluing together finitely many polygons without holes along their edges. (In particular, it does not necessarily embed as a piecewise linear surface in $\mathbb{R}^3$.) An alternating way of describing a combinatorial surface is to provide a compact surface with an embedding of a graph such that the faces are topological disks. A piecewise linear (PL) path on a combinatorial surface is a path whose intersection with each face of the surface is a (finite set of) piecewise linear path(s) with respect to some previously fixed homeomorphisms that send the faces onto plane polygons.

In this paper, $M$ denotes an arbitrary surface, while $\mathcal{M}$ denotes our input surface. $\mathcal{M}$ is assumed to be a pointed surface, with basepoint $v_0$. Moreover, $\mathcal{M}$ is assumed to be an orientable combinatorial surface without boundary whose edges have positive weights, such that $v_0$ is a vertex of this surface. Let $G$ be the (weighted) vertex-edge graph of $\mathcal{M}$, and $G^*$ be its dual graph embedded into $\mathcal{M}$: there is a vertex of $G^*$ in each face of $G$ and an edge of $G^*$ crossing each edge of $G$. In all this paper, we are interested in sets of piecewise linear paths (or loops) drawn on $\mathcal{M}$ that are *regular* with respect to $G^*$. More precisely,

- no path contains a vertex of $G^*$;
- the set of intersection points of each path with the edges of $G^*$ is finite, and each such intersection is a crossing;
- the set of (self-)intersection points between the paths is finite, and disjoint from the union of the edges of $G^*$;
- at each point of (self-)intersection between paths, exactly two curve parts meet and actually cross;
- for loops, the same definition holds except at $v_0$, where more than two loops can intersect and may or may not cross.

*Throughout this paper, we assume all paths on $\mathcal{M}$ to be regular and piecewise linear, although we omit this assumption in most statements.* If a path $p$ crosses the edges $e_1^*, \ldots, e_k^*$ of $G^*$, its *length* $|p|$ is defined to be the sum of the weights of $e_1, \ldots, e_k$, counting multiplicities. Note that if two regular paths are connected by a point not in $G^*$, then the length of the concatenated path is the sum of the lengths of these paths.

Any regular system of PL loops on $\mathcal{M}$ can be mapped to a set of closed walks on $G$ without changing their homotopy classes, see Figure 4; the length of a loop is the length of the corresponding closed walk. The resulting walks can fail to be simple and can overlap, as they can travel several times through a given vertex or edge of $G$; however, it is always possible to perturb them to get a system of loops on the surface.

In this paper, we are interested in loops on $G$ that can be perturbed to become simple and disjoint on the surface. But, for our purposes, it is is conceptually simpler to deal with PL loops that are really simple and disjoint: this enables to use topological arguments. We will come back to a setting where the loops are on $G$ when describing the computational framework: in Section 5.1, we will define a *combinatorial set of loops*, which corresponds to the data of closed walks on $G$, along with the left-to-right ordering of the walks along each edge of $G$. All algorithms described in this paper can be implemented using this data structure, without need to store actual PL loops on the surface.

### 2.3 Our results

**Definition 1** *Let $s = (s_1, s_2, \ldots, s_{2g})$ be a system of loops on $(\mathcal{M}, v_0)$. An* elementary step *$f_j(s)$ (resp. $\bar{f}_j(s)$) consists in replacing the $j$th loop $s_j$ by a* simple *loop $r_j$ such that:*

- *$r_j$ does not cross any loop in $s$ except at $v_0$;*
- *in the circular ordering of these $2g+1$ loops around $v_0$, $r_j$ starts and ends just on the left (resp. right) of $s_j$;*

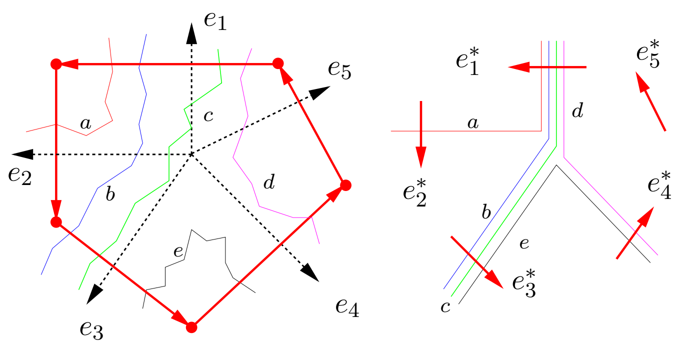


Figure 4: The set of simple, pairwise disjoint paths $(a, b, c, d, e)$ is mapped to a set of walks in $G$, in the neighborhood of a vertex whose incident edges are $e_1, \ldots, e_5$.

- *$r_j$ is as short as possible.*

*A* main phase *$f(s)$ is the sequential application of all $4g$ elementary steps $f_j$ and $\bar{f}_j$ in any order (*e.g.*, $f = \bar{f}_{2g} \circ \ldots \circ \bar{f}_1 \circ f_{2g} \circ \ldots \circ f_1$) to $s$. These operations transform a system of loops into another one in the same homotopy class.*

Pictorially, for an elementary step $f_j$, we consider the two sides $a$ and $\bar{a}$ of the polygonal schema that correspond to $s_j$, and we find a shortest path $a'$ inside this schema between the endpoints of $a$. Now we cut the "disk" delimited by $a$ and $a'$, and we glue it back to the other side $\bar{a}$. (See Figure 2.)

Here is our main theorem:

**Theorem 2** *Let $s^0$ be a system of loops on $(\mathcal{M}, v_0)$, and let $s^{n+1} = f(s^n)$. For some $m \in \mathbb{N}$, $s^m$ and $s^{m+1}$ have the same length, and, in this situation, $s^m$ is a system homotopic to $s^0$ made of loops individually as short as possible among all simple loops in their homotopy class. In particular, $s^m$ is an optimal system on $(\mathcal{M}, v_0)$.*

# 3 Crossing words

In this section, we introduce the main ingredient of this paper: the crossing word of a given path with a set of disjoint, simple paths.

## 3.1 Universal cover and lifts

We briefly recall the notion of universal covers. See Stillwell [18] or Massey [16] for more details.

Let $M$ and $M'$ be two manifolds. A map $\pi : M' \to M$ is called a *covering map* if each point $x \in M$ lies in an open neighborhood $U$ such that (1) $\pi^{-1}(U)$ is the union of disjoint open sets $(U_i)_{i\in I}$ and (2) for every $i \in I$, the restriction $\pi|_{U_i} : U_i \to U$ is a homeomorphism. If there is a covering map from $M'$ to $M$, we call $M'$ a *covering space* of $M$.

Every connected manifold $M$ has a unique simply-connected covering space (*i.e.*, a covering space in which each loop is homotopic to a constant loop), called the *universal cover* of $M$ and denoted $\tilde{M}$. Let $\pi : \tilde{M} \to M$ be the associated covering map. A *lift* of a path $p : [0, 1] \to M$ to the universal cover $\tilde{M}$ is a path $\tilde{p} : [0, 1] \to \tilde{M}$ such that $\pi \circ \tilde{p} = p$. We will use the following properties of paths and universal covers.

- The *lift property*: Let $p$ be an arbitrary path in $M$, and let $x = p(0)$. For any point $\tilde{x} \in \tilde{M}$ such that $\pi(\tilde{x}) = x$, there is a unique path $\tilde{p} : [0, 1] \to \tilde{M}$ such that $\tilde{p}(0) = \tilde{x}$ and $\pi \circ \tilde{p} = p$.
- The *homotopy property*: Two paths $p_1$ and $p_2$ with the same endpoints are homotopic in $M$ if and only if they have lifts $\tilde{p}_1$ and $\tilde{p}_2$ with the same endpoints in $\tilde{M}$.
- The *intersection property*: A path $p$ in $M$ self-intersects if and only if either some lift of $p$ self-intersects or two lifts of $p$ intersect.

Let $\pi_1(M, v)$ be the set of homotopy classes of loops on a pointed surface $(M, v)$. If $\ell$ is a loop on $(M, v)$, let $[\ell]$ denote its homotopy class. The set $\pi_1(M, v)$ equipped with the law $[\ell_1].[\ell_2] = [\ell_1.\ell_2]$ is a group, called the *fundamental group* of $(M, v)$; its unit element (the class of null-homotopic loops) will be denoted by $\epsilon$.

We conclude this section with the following lemma.

**Lemma 3** *Let $M$ be a connected compact surface with boundary. Let $c$ be a simple path on $M$ whose intersections with the boundary of $M$ are precisely its endpoints. Then any lift of $c$ separates the universal cover $\tilde{M}$ of $M$ in two connected components.*

PROOF. We show that the universal cover $\tilde{M}$ of $M$ is homeomorphic to the unit closed disk $D$ with some points on $\partial D$ removed. Since any lift of a boundary-to-boundary path in $M$ is a boundary-to-boundary path in $\tilde{M}$, the lemma follows by a direct application of the Jordan curve theorem.

Consider a polygonal schema $S$ of $M$: the universal cover $\tilde{M}$ of $M$ consists of copies of $S$ glued together along the sides of $S$ that do not correspond to boundaries of $M$. Let $\tilde{M}_0$ be such a copy appearing in $\tilde{M}$. For each integer $i$, let $\tilde{M}_i$ be the part of $\tilde{M}$ that can be reached from a point in the interior of $\tilde{M}_0$ by crossing at most $i$ sides of the boundary of copies of $S$; each $\tilde{M}_i$ is a topological disk. Furthermore, $\tilde{M}_i \subseteq \tilde{M}_{i+1}$ and $\tilde{M} = \bigcup_{i=0}^{\infty} \tilde{M}_i$.

Write $\partial\tilde{M}_i$ as the disjoint union of the points $B_i$ that are on the boundary of $\tilde{M}$ and of the points $F_i$ that are not. Let $h_0 : \tilde{M}_0 \to D$ map $\tilde{M}_0$ homeorphically onto some polygon containing the center of $D$ and such that $h_0(\tilde{M}_0) \cap \partial D = h_0(B_0)$. By induction on $i \geq 0$, we define a homeomorphism $h_i$ from $\tilde{M}_i$ into the unit disk $D$ such that:

- $h_i(B_i) \subseteq \partial D$;

- $h_i(F_i) \cap \partial D = \emptyset$;

- $h_i(\tilde{M}_i)$ contains the disk of radius $1 - \frac{1}{i+1}$.

- $h_{i+1}|_{\tilde{M}_i} = h_i|_{\tilde{M}_i}$;

Now, if $x \in \tilde{M}_i$, put $h(x) = h_i(x)$. It is routine to check that $h$ is a homeomorphism from $\tilde{M}$ onto its image. This image contains the interior of $D$ and is included in $D$. □

## 3.2 Crossing words and reductions

Let $M$ be an oriented surface with or without boundary. Let $(s_i)_{i \in I}$ be an indexed set of paths on $M$, and let $p$ be a path on $M$; assume that $p$ crosses the paths $s_i$ only transversely and at a finite number of points. Let $A$ be the set of *symbols* $\{i, i \in I\} \cup \{\bar{\imath}, i \in I\}$. The set $A^*$ of *words* on $A$ is the set of finite sequences of elements in $A$. The *crossing word* of $p$ with the paths in $s$ is the word defined as follows: walk along $p$ and, at each crossing encountered with a path $s_i$, write the symbol $i$ if $p$ pierces $s_i$ from right to left at this point and $\bar{\imath}$ otherwise. A word on such a set of symbols $A$ is called *parenthesized* if it can be transformed to the empty word by successive removal of subwords of the form $i\bar{\imath}$ or $\bar{\imath}i$, where $i \in I$.

The proof of Theorem 2 relies on the study of crossing words of paths in the universal cover $\tilde{\mathcal{M}}$ of $\mathcal{M}$. Fix a lift $v_0^\epsilon$ of $v_0$ in $\tilde{\mathcal{M}}$. For $\alpha \in \pi_1(\mathcal{M}, v_0)$, all lifts of the loops in $\alpha$ starting at $v_0^\epsilon$ end at the same lift of $v_0$, which we call $v_0^\alpha$; this gives a one-to-one correspondence between $\pi_1(\mathcal{M}, v_0)$ and the lifts of $v_0$. If $\ell$ is a loop on $(\mathcal{M}, v_0)$ and $\alpha \in \pi_1(\mathcal{M}, v_0)$, we denote by $\ell^\alpha$ the lift of $\ell$ starting at $v_0^\alpha$. If $a$ is a loop in the homotopy class $\alpha$, we have in particular $a^\epsilon.\ell^\alpha = (a.\ell)^\epsilon$.

Fix an index $i$ between 1 and $2g$; on $(\mathcal{M}, v_0)$, consider a system of loops $s$ and a simple loop $t_i$ homotopic to $s_i$. Assume that the loop $t_i^\epsilon$ crosses the loops $s_k^\alpha$ (for $k \in [1, 2g]$ and $\alpha \in \pi_1(\mathcal{M}, v_0)$) transversely and at a finite number of points[1]. We define $[s/t_i]$ to be the crossing word, in $\tilde{\mathcal{M}}$, of the path $t_i^\epsilon$ with the paths $s_k^\alpha$, $k \in [1, 2g]$ and $\alpha \in \pi_1(\mathcal{M}, v_0)$: it is thus a word on the set of symbols of the form $k^\alpha$ or $\bar{k}^\alpha$.

A special interesting way of uncrossing the loop $t_i$ with those in $s$ is the following. Geometrically, $t_i$ and the loops in $s$ cut $\mathcal{M}$ into pieces. Each piece is a topological disk bounded by subpaths of $t_i$ and $s$. If a piece is bounded by exactly two subpaths, one of $t_i$ and one of some $s_k$, then by deforming $t_i$ continuously we can remove crossings between $t_i$ and $s_k$ (see Figure 5). More precisely, let $a$ and $b$ be the endpoints of the subpaths of $t_i$ and $s_k$. The number of crossings actually removed depends on whether $a$ and/or $b$ are the basepoint $v_0$. If neither $a$ nor $b$ equals $v_0$, then two crossings can be removed and we have a *parenthesized* uncrossing (Figure 5, left). If exactly one of $a$ or $b$ equals $v_0$, one crossing

[1] Of course, the endpoints of $t_i^\epsilon$ intersect the endpoints of some loops $s_k^\alpha$; in the following, these intersections should not be considered as a crossing.

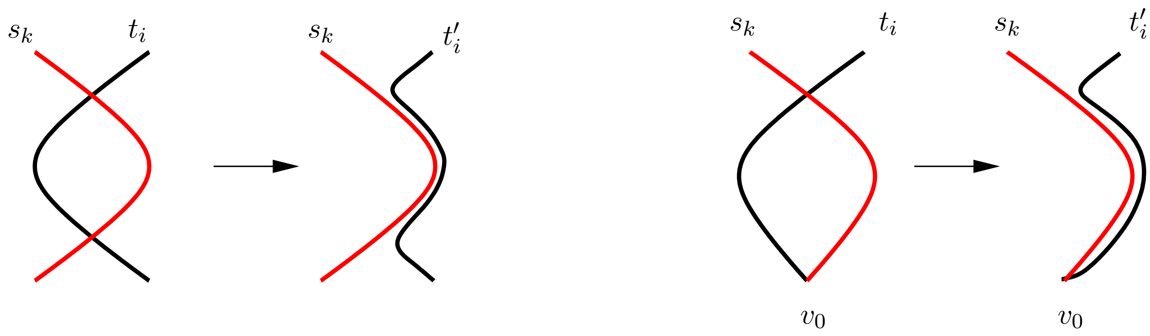


Figure 5: Uncrossing curves $s_k$ and $t_i$ by deforming continuously $t_i$. Left: a parenthesized uncrossing. Right: an extremal uncrossing.

is removed and we have an *extremal* uncrossing (Figure 5, right). If $a = b = v_0$, no crossing is removed.

We now want to describe more precisely the effect of uncrossings on crossing words in the universal cover. We say that a symbol $k^\alpha$ or $\bar{k}^\alpha$ is *initial* if $v_0^\epsilon$ is one of the endpoints of $s_k^\alpha$, or, equivalently, if $\alpha \in \{\epsilon, [\bar{s}_k]\}$ (where $\bar{s}_k$ means the inverse of $s_k$). Similarly, such a symbol is *final* if the target of $t_i^\epsilon$ coincides with one of the endpoints of $s_k^\alpha$, or, equivalently, if $\alpha \in \{[t_i], [t_i.\bar{s}_k]\}$. We define two types of *reductions* on a word in $A^*$:

- a *parenthesized reduction* consists in removing a subword of the form $k^\alpha \bar{k}^\alpha$ or $\bar{k}^\alpha k^\alpha$;
- an *extremal reduction* consists in removing the first (resp. last) symbol of the word, if it is an initial (resp. final) symbol.

Note that a parenthesized or extremal uncrossing gives rise to a reduction of the same type on the crossing word.

Let $w, w' \in A^*$. We say that $w$ *reduces* to $w'$ if $w'$ can be obtained from $w$ by some reductions; $w$ is an *irreducible word* if no reduction is possible on $w$. Given $w \in A^*$, it is easy to prove that there is a unique irreducible word to which $w$ reduces.

# 4 Proof of Theorem 2

Let $s$ be a system of loops on $(\mathcal{M}, v_0)$; fix an index $i$ between 1 and $2g$, and let $t_i$ be a loop homotopic to $s_i$. The proof of our theorem essentially relies on the two following facts:

- the crossing word $[s/t_i]$ reduces to the empty word;
- if $t_i$ is, in addition, as short as possible among all simple loops in the homotopy class of $s_i$, then each main phase corresponds to at least one reduction.

The next two subsections are devoted to the proof of these facts.

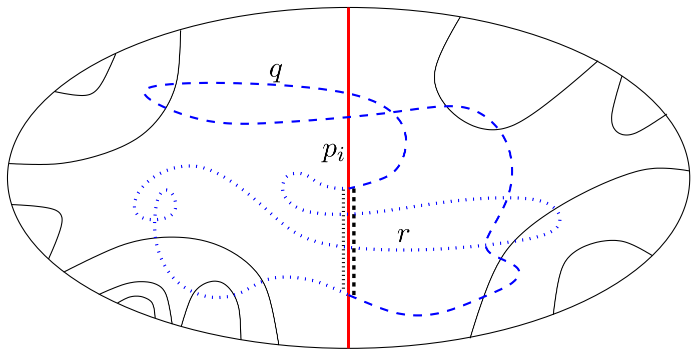


Figure 6: The induction step of Lemma 5.

### 4.1 Reducibility of $[s/t_i]$

**Proposition 4** *$[s/t_i]$ reduces to the empty word.*

Let us first introduce two lemmas.

**Lemma 5** *Let $M$ be an (oriented, possibly non-compact) simply connected surface with boundary. Let $p_1, \ldots, p_n$ be simple, pairwise disjoint paths in $M$, each of which separates $M$ into two connected components. Let $\ell$ be a loop in the interior of $M$ that intersects each path $p_k$ only transversely and at a finite number of points. Then the crossing word of the loop $\ell$ with the paths $p_k$ is parenthesized.*

PROOF. We prove the result by induction on the number of crossings between $\ell$ and the paths $p_k$. The lemma is trivial if $\ell$ crosses no $p_k$. Assume on the contrary there is at least one crossing between $\ell$ and some path $p_i$.

Since $p_i$ is separating, $\ell$ must cross $p_i$ once more with the opposite orientation. We temporarily view $\ell$ as a cycle (*i.e.*, we forget the basepoint of $\ell$ and consider $\ell$ as a map $S^1 \to M$). The two crossings split $\ell$ into two paths $q$ and $r$ (Figure 6). It is possible to extend $q$ and $r$ into *loops* $q'$ and $r'$ without changing their crossings with any $p_k$. By the induction hypothesis, the crossing word of $q$ or $r$ with the paths $p_k$ is parenthesized. But the crossing word of $\ell$ with the paths $p_k$ results from the concatenation of these two crossing words, with two additional symbols $i$ and $\bar{\imath}$, followed by a cyclic permutation: it is thus parenthesized. □

Recall that an *isotopy* between two simple loops is a homotopy with the additional property that the loop remains simple at each stage of the homotopy. We will need the following theorem [8, Theorem 4.1].

**Theorem 6** *Let $\ell_0$ and $\ell_1$ be piecewise linear, homotopic, simple loops in $(\mathcal{M}, v_0)$, such that they are not null-homotopic. Then, there is a piecewise linear isotopy between $\ell_0$ and $\ell_1$ (keeping the basepoint fixed).*

Let $D$ be an open disk on $\mathcal{M}$ containing $v_0$. We say that a loop $u$ on $(\mathcal{M}, v_0)$ is *$D$-clean* if $u$ "enters $D$ only once", *i.e.*, considering $u$ to be a mapping from the circle into $\mathcal{M}$, $u^{-1}(D)$ is connected. If $u$ is $D$-clean, let $\dot{u}$ denote the path consisting of the part of $u$ outside $D$.

**Lemma 7** *Let $D$ be an open disk containing $v_0$, and let $u$ and $u'$ be piecewise linear homotopic loops on $(\mathcal{M}, v_0)$ that are $D$-clean and simple. Then, there are two paths $p$ and $p'$ on the boundary of $D$ so that $\dot{u}^{-1}.p.\dot{u}'.p'$ is null-homotopic in $\mathcal{M} \setminus D$.*

PROOF. $u$ and $u'$ are piecewise linearly isotopic on $\mathcal{M}$, with the basepoint $v_0$ fixed, by Theorem 6. Let $h : [0,1] \times [0,1] \to \mathcal{M}$ be the isotopy: for each $t$, $h(t, .) : [0,1) \to \mathcal{M}$ is one-to-one, and $h(t,0) = h(t,1) = v_0$; $h(0,.) = u$, and $h(1,.) = u'$.

$h^{-1}(D)$ is a neighborhood of the compact set $[0,1] \times \{0,1\}$; hence there exists an $\varepsilon > 0$ such that $h([0,1] \times ([0,\varepsilon] \cup [1-\varepsilon, 1])) \subset D$. Let $h'$ be the restriction of $h$ to $[0,1] \times [\varepsilon, 1-\varepsilon]$.

Let $r : \mathcal{M} \setminus \{v_0\} \to \mathcal{M} \setminus D$ be a continuous map that is the identity on $\mathcal{M} \setminus D$ and that maps $D \setminus \{v_0\}$ onto the boundary of $D$. Since $h$ is an isotopy and $h(.,0) = v_0$, $h'' = r \circ h'$ is a well-defined homotopy. $h''(., \varepsilon)$ and $h''(., 1-\varepsilon)$ are on the boundary of $D$; $h''(0,.)$ (resp. $h''(1,.)$) is made of a path on the boundary of $D$, $\dot{u}$ (resp. $\dot{u}'$), and another path on the boundary of $D$; from these facts, it is easy to derive the paths $p$ and $p'$, and the desired homotopy. □

PROOF OF PROPOSITION 4. Let $s'_i$ be a simple loop homotopic to $s_i$ such that it does not cross any of the loops $s_k$, $k \in [1, 2g]$ (for example, let $s'_i$ "go along" $s_i$, sufficiently near $s_i$). Let $D$ and $D'$ be two open disks such that $v_0 \in D'$ and the closure of $D'$ is included in $D$. By choosing disks small enough, we can ensure that $t_i$, $s'_i$, and all loops $s_k$ are $D$- and $D'$-clean. By further reducing $D$ and $D'$, we can also assume that $t_i$ does not cross $s$ in $D$ (recall that the basepoint is not considered as a crossing).

Let $\mathcal{M}' = \mathcal{M} \setminus D'$. By Lemma 7, there are two paths $p$ and $p'$ on the boundary of $D$ such that $\ell := \dot{s}'^{-1}_i.p.\dot{t}_i.p'$ is a contractible loop in $\mathcal{M} \setminus D$, hence also in $\mathcal{M}'$. Hence any lift of $\ell$ in the universal cover $\tilde{\mathcal{M}}'$ of $\mathcal{M}'$ is a loop.

For any $D'$-clean loop $u$, denote by $\ddot{u}$ the part of $u$ outside $D'$. Applying Lemma 3 to the $\ddot{s}_k$ and Lemma 5 in $\tilde{\mathcal{M}}'$, we obtain that the crossing word of any lift of $\ell$ with the lifts of $\ddot{s} = (\ddot{s}_1, \ldots, \ddot{s}_{2g})$ is parenthesized. Let $q = p.\dot{t}_i.p'$, which implies that $\ell = \dot{s}'^{-1}_i.q$. Since $\dot{s}'_i$ does not cross $\ddot{s}$, the crossing word of any lift of $q$ with the lifts of $\ddot{s}$ in $\tilde{\mathcal{M}}'$ is also parenthesized.

Let $\pi$ be the covering map from $\tilde{\mathcal{M}}$ onto $\mathcal{M}$. Note that $\tilde{\mathcal{M}}'$ is a covering space — in fact, the universal cover — of $\tilde{\mathcal{M}} \setminus \pi^{-1}(D')$. In particular, crossings between paths in $\tilde{\mathcal{M}}'$ project to crossings in $\tilde{\mathcal{M}}$. Hence the crossing word of any lift of $q$ in $\tilde{\mathcal{M}}$ with the lifts of $\ddot{s}$ is also parenthesized.

Let $\tilde{q}$ be, in $\tilde{\mathcal{M}}$, the lift of $q = p.\dot{t}_i.p'$ that contains, as a subpath, the lift of $\dot{t}_i$ contained in $t^{\epsilon}_i$. The crossing word of $\tilde{q}$ with lifts of $s$ is parenthesized; it is the concatenation of the crossing word $w_1$ of a lift of $p$ with the lifts of $s$, the crossing word $[s/t_i]$ (because $t_i$ does not cross $s$ inside $D$), and the crossing word $w_2$ of a lift of $p'$ with the lifts of $s$. By the choice of $\tilde{q}$, $p$, and $p'$, the word $w_1$ is made of initial symbols and, similarly, the word $w_2$ is made of final symbols. This implies that the crossing word of $\tilde{q}$ with lifts of $s$ not only reduces to the empty word but also to $[s/t_i]$; hence $[s/t_i]$ reduces to the empty word. □

### 4.2 Uncrossing the loops

We introduce operations on crossing words that should reflect the effect of an elementary step on systems of loops as defined in Section 2.3. A *$j$-symbol* is a symbol of the form $j^\alpha$ or $\bar{\jmath}^\alpha$. We call a *left $j$-reduction* on a word $w$ the removal of either

1. the first $j$-symbol, if it is initial and has a bar, or
2. all subwords of the form $j^\alpha \bar{\jmath}^\alpha$, or
3. the last $j$-symbol, if it is final and has no bar.

An *elementary step* on a word $w$ is the application of all possible *left $j$-reductions* performed *simultaneously*: the process is non-recursive; for example, $g_j(\bar{\jmath}^\epsilon \bar{\jmath}^\epsilon j^\alpha j^\beta \bar{\jmath}^\beta \bar{\jmath}^\alpha) = \bar{\jmath}^\epsilon j^\alpha \bar{\jmath}^\alpha$. The resulting word is denoted $g_j(w)$. Clearly, $w$ reduces to $g_j(w)$. *Right $j$-reductions* and $\bar{g}_j(w)$ are defined analogously, inverting the role of a bar and no bar.

Let $s$ be a system of loops. Consider $i, j \in [1, 2g]$ and let $t_i$ be a shortest simple loop homotopic to $s_i$. Let $r = f_j(s)$. Intuitively, the goal of this section is to prove that applying any elementary step $f_j$ or $\bar{f}_j$ to $s$ induces (roughly) an elementary step on the crossing word $[s/t_i]$:

**Proposition 8** *Assume that $s$, $r_j$, and $t_i$ constitute a regular set of loops. There exists a simple loop $t'_i$ of the same length as and homotopic to $t_i$, such that $s$, $r_j$, and $t'_i$ constitute a regular set of loops and $[r/t'_i] = g_j([s/t_i])$.*

Of course, a similar proposition holds if we let $r = \bar{f}_j(s)$. The immediate but important corollary is:

**Corollary 9** *Assume that $[s/t_i]$ is not the empty word. Then there exists a simple loop $t'_i$ of the same length as and homotopic to $t_i$ such that $[f(s)/t'_i]$ is shorter than $[s/t_i]$.*

Proof of Corollary 9. By Proposition 8, there exists $t'_i$ such that

$$[f(s)/t'_i] = g([s/t_i]),$$

where $g$ is the composition of all the $g_k$'s and $\bar{g}_k$'s in some order. $[s/t_i]$ reduces to the empty word by Proposition 4, hence, for some $j$, a $j$-reduction is possible. This $j$-reduction is performed with the application of either $g_j$ or $\bar{g}_j$, depending on the $j$-reduction. □

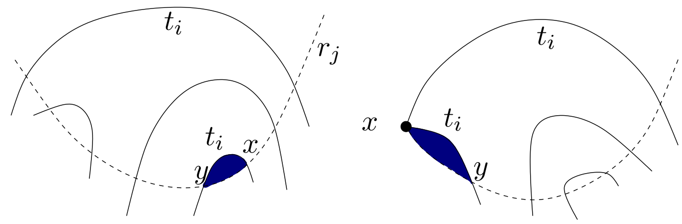


Figure 7: The two situations considered in the proof of Lemma 10: on the left, none of the corners are the basepoint; on the right, $x$ is the basepoint.

An $r_k$-symbol (resp. $s_k$-symbol) is of the form $r_k^\alpha$ or $\bar{r}_k^\alpha$ (resp. $s_k^\alpha$ or $\bar{s}_k^\alpha$). Recall that $r_k = s_k$ for each $k \neq j$. In view of proving Proposition 8, we define $[r+s/t_i]$ as the crossing word of $t_i^\epsilon$ with all lifts of $s$ and of $r_j$ (*e.g.* $s_k^\alpha r_j^\beta \bar{s}_j^\gamma \ldots$); this word is well-defined because the set of loops is regular. The knowledge of $[r+s/t_i]$ determines $[s/t_i]$ and $[r/t_i]$. We define an $r_j$-reduction as being a parenthesized or an extremal reduction involving $r_j$-symbols.

**Lemma 10** *Suppose that an $r_j$-reduction can be processed on the word $[r+s/t_i]$. Then one can replace $t_i$ by a simple loop $t'_i$, of the same length as and homotopic to $t_i$, such that $[r+s/t'_i]$ is deduced from $[r+s/t_i]$ by an $r_j$-reduction (and the resulting set of loops is regular).*

Proof. Define a *lens* of two loops $\ell_1$ and $\ell_2$ on a given surface to be two homotopic subpaths of $\ell_1$ and $\ell_2$ with the same endpoints $x$ and $y$, called the *corners* of the lens, so that these two subpaths concatenated together make a simple loop bounding a topological disk on that surface. Because either a parenthesized or an extremal $r_j$-reduction is possible on $[r+s/t_i]$, there must exist, in $\tilde{\mathcal{M}}$, a lens of $t_i^\epsilon$ and a lift of $r_j$, such that at most one of its corners projects to $v_0$ and no lift of $r$ and $s$ crosses this lens.

The projection of this lens on $\mathcal{M}$ crosses no loop in $r$; it can thus be viewed as being drawn in the polygonal schema defined by $r$, hence bounds a topological disk $D$ and is also a lens. No loop of $r$ and $s$ enters $D$; the intersection of $t_i$ with $D$ is a set of simple paths whose endpoints lie on $r_j$. Considering an innermost such curve (Figure 7), we get a lens of $r_j$ and $t_i$ that is crossed by none of the paths in $r$ and $s$, nor by $t_i$. Let $x$ and $y$ be the corners of the lens, and $t_i^{xy}$ and $r_j^{xy}$ be the parts of $t_i$ and $r_j$ constituting this lens; assume that $y$ is not the basepoint.

Consider the loop $r'_j$ obtained from $r_j$ after the replacement of $r_j^{xy}$ by a path going along $t_i^{xy}$, just outside the lens; by the definition of $r = f_j(s)$, $r'_j$ cannot be shorter than $r_j$; hence, $t_i^{xy}$ cannot be shorter than $r_j^{xy}$. Change $t_i$ into $t'_i$ as follows: replace $t_i^{xy}$ by a path going along $r_j^{xy}$, just outside the lens. $[r+s/t'_i]$ is deduced from $[r+s/t_i]$ by an $r_j$-reduction, and $t'_i$ is no longer than $t_i$, hence has the same length as, $t_i$; furthermore, $t'_i$ is simple. □

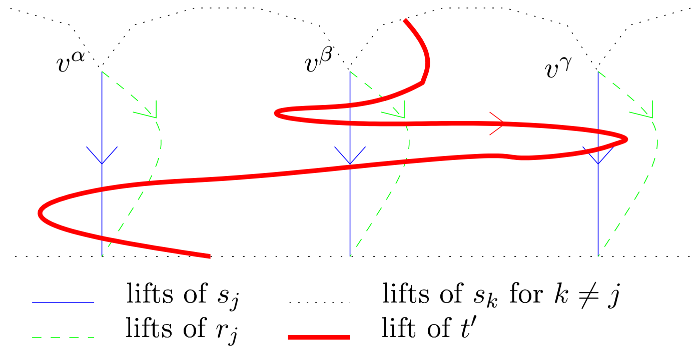


Figure 8: A view of $\tilde{\mathcal{M}}(s, j, \tilde{t}')$, the lifts of $r_j$, $s_j$, and the lift $\tilde{t}'$. Here, $[r + s/t'] = \bar{r}_j^\beta \bar{s}_j^\beta s_j^\beta r_j^\beta s_j^\gamma \bar{s}_j^\gamma \bar{r}_j^\beta \bar{s}_j^\beta \bar{r}_j^\alpha \bar{s}_j^\alpha s_j^\alpha r_j^\alpha$.

Proof of Proposition 8. Let $t_i'$ be the loop obtained by repeated applications of Lemma 10 until no $r_j$-reduction is possible on $[r + s/t_i]$. Note that $[s/t_i] = [s/t_i']$; we shall prove that $[r/t_i'] = g_j([s/t_i'])$.

Define the $j$-*intervals* of $t_i'$ to be the maximal subpaths of $t_i'$ that do not cross any $s_k$, for $k \neq j$. If $t'$ is a $j$-interval of $t_i'$, let $\tilde{t}'$ be the lift of $t'$ that is included in $t_i'^{\epsilon}$. Let $[s/t']$ denote the crossing word of $\tilde{t}'$ with the lifts of $s$ (thus, $[s/t']$ is a subword of $[s/t_i']$). Actually, $[s/t']$ contains only $j$-symbols. Define an occurrence of a symbol of $[s/t']$ to be initial or final if and only if that occurrence is initial or final in $[s/t_i']$. To prove our proposition, it is sufficient to prove: $[r/t'] = g_j([s/t'])$.

Assume first that $t'$ is not the first $j$-interval, nor the last one. Similarly as above, let $[r+s/t']$ be the subword of $[r+s/t_i']$ that is the crossing word of $\tilde{t}'$ with the lifts of $r$ and $s$.

Consider $\tilde{\mathcal{M}}$ where all lifts of $s_k$, for $k \neq j$, have been removed. Let $\tilde{\mathcal{M}}(s, j, \tilde{t}')$ be the connected component of this surface that contains $\tilde{t}'$ (Figure 8). On $\tilde{\mathcal{M}}(s, j, \tilde{t}')$, we see that two distinct lifts of $r_j$ are separated by a lift of $s_j$, and vice-versa. Hence, there are no consecutive $r_j$-symbols in $[r + s/t']$, because they cannot involve the same lift of $r_j$ (an $r_j$-reduction would be possible on $[r+s/t']$), nor different lifts of $r_j$ in $\tilde{\mathcal{M}}(s, j, \tilde{t}')$ (since two such lifts are separated by a lift of $s_j$).

$[r + s/t']$ is deduced from $[s/t']$ by insertions of $r_j$-symbols. We know that at most one $r_j$-symbol can be inserted between two $s_j$-symbols (and also at the beginning and the end) of $[s/t']$. Using this property and obvious separation properties of the lifts of $r_j$ and $s_j$ in $\tilde{\mathcal{M}}(s, j, \tilde{t}')$, one can deduce $[r + s/t']$ directly from $[s/t']$. The following table gives the rules: the first two columns indicate two successive $s_j$-symbols in $[s/t']$, and the right column indicates the $r_j$-symbol to be inserted between these $s_j$-symbols to get $[r+s/t']$, or "$-$" if no symbol should be inserted. The beginning or the end of the word $[s/t']$ is denoted by a star.

| between | and | insert | between | and | insert |
|---|---|---|---|---|---|
| $\star$ | $\bar{s}_j^\alpha$ | $\bar{r}_j^\alpha$ | $s_j^\alpha$ | $\bar{s}_j^\alpha$ | $-$ |
| $\star$ | $s_j^\alpha$ | $-$ | $\bar{s}_j^\alpha$ | $s_j^\alpha$ | $-$ |
| $\bar{s}_j^\alpha$ | $\star$ | $-$ | $\bar{s}_j^\alpha$ | $\bar{s}_j^\beta$ | $\bar{r}_j^\beta$ |
| $s_j^\alpha$ | $\star$ | $r_j^\alpha$ | $s_j^\alpha$ | $s_j^\beta$ | $r_j^\alpha$ |
| $\star$ | $\star$ | $-$ | | | |

The word $[r/t']$ is exactly the list of inserted symbols, and it is easy to see, using this table and by induction on the length of $[s/t']$, that this list is precisely $g_j([s/t'])$.

Assume now that $t'$ is the first $j$-interval of $t_i$. The reasoning is analogous, except that the first rule (top left) is not valid any more; instead, the following rule has to be used: between $\star$ and $\bar{s}_j^\alpha$, we insert nothing if $\bar{\jmath}^\alpha$ is an initial symbol, and $\bar{r}_j^\alpha$ otherwise. By induction, $[r/t'] = g_j([s/t'])$. The case where $t'$ is the last $j$-interval of $t_i$ is analogous. □

## 4.3 Conclusion of the proof

**Lemma 11** *Let $u$ be a system of loops on $(\mathcal{M}, v_0)$, and let $v_i$ be a simple loop homotopic to $u_i$, with minimal length, so that $u$ and $v_i$ constitute a regular set of loops. If $v_i$ does not cross any loop in $u$, then the ith loop of $f(u)$ has the same length as $v_i$.*

PROOF. Without loss of generality, suppose we can write $f = a \circ \bar{f}_i \circ b \circ f_i \circ c$, where $a, b$, and $c$ are compositions of elementary steps other than $f_i$ and $\bar{f}_i$; the argument is similar if $f_i$ and $\bar{f}_i$ appear in the opposite order in $f$. There are two cases to consider. First suppose $v_i$ is to the left of $u_i$. By Proposition 8, there exists a simple loop $v_i'$ of the same length as and homotopic to $v_i$ such that $v_i'$ does not cross any loop in $c(u)$ and $v_i'$ is to the left of $c(u)_i$. The definition of $f_i$ implies that $f_i(c(u))_i$ cannot be longer than $v_i'$. The result follows for an elementary step can only shorten a loop and $v_i'$ has minimal length among all the simple loops homotopic to $u_i$.

Now suppose $v_i$ is to the right of $u_i$. Again, there exists a simple loop $v_i'$ of the same length as and homotopic to $v_i$ such that $v_i'$ does not cross any loop in $c(u)$, this time with $v_i'$ to the right of $c(u)_i$. Then, there exists a simple loop $v_i''$ of the same length as and homotopic to $v_i'$, where $v_i''$ is to the right of $u_i' = f_i(c(u))_i$ and does not cross any loop in $f_i(c(u))$. We may complete the proof as for the first case by exchanging left and right. □

PROOF OF THEOREM 2. Let $s^0$ be a system of loops; for all $n \geq 0$, define $s^{n+1} = f(s^n)$. Fix an index $i$ between 1 and $2g$. Let $t_i$ be a simple loop homotopic to $s_i^0$, such that $t_i$ has minimal length and such that $t_i$ and $s^0$ constitute a regular set of loops. Let $p$ be the length of $[s^0/t_i]$. By at most $p$ applications of Corollary 9, one can construct a loop $t_i'$ homotopic to and of the same length as $t_i$ such that $t_i'$ does not cross any loop in $s^p$. By Lemma 11, $s_i^{p+1}$ has the same length as $t_i^0$. This proves that the length of $(s^n)_{n\in\mathbb{N}}$ becomes stationary at some stage of the algorithm.

Let $m$ be the smallest integer such that the length of $s^m$ equals the length of $s^{m+1}$. To finish the proof of Theorem 2, it remains to prove that all systems $s^k$, for $k \geq m$, have the same length. This is obviously the case if the following property is true (and it is possible to assume it with a trivial modification of the algorithm): if, in some elementary step, the $j$th loop of $f_j(s)$ has the same length as $s_j$, then $f_j(s) = s$. This restriction is not necessary, however. The interested reader may consult the lengthy proof in the appendix. □

# 5 A practical algorithm

This section explains how Theorem 2 can be turned into a practical algorithm as claimed in the introduction. Specifically, we describe how to perform an elementary step in a combinatorial framework, which we describe below.

## 5.1 Combinatorial set of loops

We (temporarily) view $G$, the vertex-edge graph of $\mathcal{M}$, as a directed graph: each edge of $G$ is replaced by two opposite directed edges. A *combinatorial set of loops* $S$ on the graph $G$ is a set of closed walks (with basepoint $v_0$) in $G$, with, for each directed edge $e$, a left-to-right order $\preceq_e$ of the traversals of $e$ or its reversal $-e$ by walks in $S$. These orders should be consistent in the following sense: $a \preceq_e b$ if and only if $b \preceq_{-e} a$. Intuitively, $a \preceq_e b$ if and only if $a$ is on the left of $b$ on edge $e$. A *combinatorial loop* is a combinatorial set of loops composed of a single closed walk.

Let $v$ be a vertex of $G$, and $e_1, \ldots, e_n$ be the clockwise-ordered list of oriented edges of $G$ whose source is $v$. We define a cyclic order $\preceq_v$ over the edges of the walks in $S$ meeting at $v$, by enumerating its elements in this order: first, the edges of the walks in $S$ on $e_1$ or $-e_1$, in $\preceq_{e_1}$-order; then the edges of the walks in $S$ on $e_2$ or $-e_2$, in $\preceq_{e_2}$-order; and so on. We say that two subpaths of length two, $a_1, a_2$ and $b_1, b_2$, of walks in $S$ *cross at* $v$ if $v$ is the target of both $a_1$ and $b_1$ and if, in the cyclic order $\preceq_v$, $a_1$ and $a_2$ separate $b_1$ and $b_2$. A combinatorial set of loops is *simple* if no crossing occurs in it (except possibly at $v_0$ if the two subpaths involved in a crossing are composed of the last and first edges of each of their corresponding closed walks).

It is easy to derive a data structure to store a combinatorial set of loops, in which accessing the predecessor and successor edges of a loop, and the predecessor and successor edges of the walks with respect to some $\preceq_v$-order, takes constant time.

Clearly, any regular set $s$ of disjoint simple loops on $\mathcal{M}$ can be mapped to a combinatorial set of loops $S$ in $G$ with the same lengths and homotopy classes. The order $\preceq_e$ of the traversals of $e$ or its reversal $-e$ by walks in $S$ is directly deduced from the order of the crossings of $s$ along the dual edge $e^*$; see Figure 4. We denote by $\rho$ this mapping. The converse is also true: the closed walks of a simple combinatorial set of loops can be expanded along the edges to get a set of disjoint, simple loops $s \in \rho^{-1}(S)$. A combinatorial set of loops $S$ is called a *combinatorial system* of loops if it is the image $\rho(s)$ of a (regular) system of loops $s$.

### 5.2 Processing an elementary step

We briefly indicate how an elementary step (as illustrated in Figure 2) is processed algorithmically. Our goal is to show that this problem reduces to finding a shortest path in a graph $G(\mathcal{M} \setminus s)$, which we now describe.

Let $S = \rho(s)$ be a combinatorial system of loops. Define $G(\mathcal{M} \setminus s)$ to be the weighted graph whose vertices are the components of $\mathcal{M} \setminus \{s \cup G^*\}$ and whose edges join two vertices separated by (a piece of) an edge $e^*$ of $G^*$. Let such an edge have the same weight as $e$. Clearly, $G(\mathcal{M} \setminus s)$ depends on $S$ and $G$ only; it can be computed as follows.

The set of edges of $G(\mathcal{M} \setminus s)$ is the union, over the non-oriented edges $e$ of $G$, of the intervals between edges of the walks in $S$ going along $e$. Each edge $e$ of $G$ traversed by $k$ walk edges of $S$ thus gives rise to $k+1$ edges in $G(\mathcal{M} \setminus s)$ with the same weight as $e$. Next, identify the extremities of edges in $G(\mathcal{M} \setminus s)$ whenever they correspond to the same vertex $v$ in $G$ and the corresponding intervals are not separated by $S$ (inserting a 2-path in place of these intervals in the combinatorial set of loops $S$ would introduce no crossing at $v$). $G(\mathcal{M} \setminus s)$, and its topological embedding in $\mathcal{M} \setminus s$ (*i.e.*, the circular order of its edges around its vertices) can be computed in time proportional to its size.

Note that the basepoint $v_0$ in $G$ gives rise to at least $4g$ vertices in $G(\mathcal{M} \setminus s)$ corresponding to the $4g$ sectors cut by $s$ around $v_0$.

Let $\overline{\mathcal{M} \setminus s}$ denote the topological completion of $\mathcal{M} \setminus s$. It is a topological disk whose boundary can be subdivided into $4g$ paths, or "sides", exactly two of which project to any loop in $s$. Clearly, $G(\mathcal{M} \setminus s)$ can be embedded into $\overline{\mathcal{M} \setminus s}$ such that the $4g$ vertices in $G(\mathcal{M} \setminus s)$ corresponding to $v_0$ coincide with the replicates of $v_0$ in $\overline{\mathcal{M} \setminus s}$. Moreover, each path in $\overline{\mathcal{M} \setminus s}$ between such replicates can be mapped to $G(\mathcal{M} \setminus s)$, while preserving its extremities and its length (with the appropriate definitions). A converse statement is also true: a shortest path $p_j$ between the endpoints of a side of $\overline{\mathcal{M} \setminus s}$ — corresponding, for example, to one of the duplicates of $s_j$ — can be computed as a shortest path in $G(\mathcal{M} \setminus s)$ between the corresponding vertices.

We can eventually re-identify the extremities of this shortest path to obtain a combinatorial loop $R_j$ that can be inserted in place of $S_j$ in the combinatorial set of loops $S$. According to our discussion, this new combinatorial set of loops is of the form $\rho(f_j(s))$.

### 5.3 Complexity analysis

Let $n$ be the complexity of $\mathcal{M}$ (total number of vertices, edges, and faces) and $g$ be its genus. Define $\alpha$ as the longest-to-shortest edge ratio of $\mathcal{M}$. Consider a combinatorial system of loops $S = \rho(s)$ on $\mathcal{M}$, and let $m$ be the maximal number of edges of any combinatorial loop in $S$.

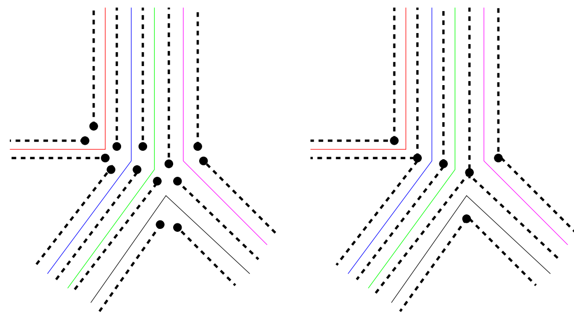

Figure 9: The construction of $G(\mathcal{M} \setminus s)$ (dashed), in the neighborhood of the vertex shown in Figure 4. First stage (left): create one edge for $G(\mathcal{M} \setminus s)$ for each interval between walk-edges on the same edge of $G$. Second stage (right): connect edges $e$ and $e'$ if a path of length two could be inserted at the intervals $e$ and $e'$ without crossing.

Since a combinatorial loop can only get shorter in length, its maximal number of edges in the course of the algorithm is $O(\alpha m)$. Let $t_i$ be a shortest loop homotopic to $s_i$, and let $T_i = \rho(t_i)$. Similarly, the complexity of $T_i$ is $O(\alpha m)$. A trivial bound on the lengths of the crossing words is thus $O(g\alpha m^2)$; it is also a bound on the number of main phases required. Any system of loops considered can have complexity $O(g\alpha m)$, whence the complexity of an elementary step, which is dominated by the computation of a shortest paths using Dijkstra's algorithm, is $O((g\alpha m + n)\log(g\alpha m + n))$. Finally:

**Theorem 12** *Let $\mathcal{M}$ be an orientable combinatorial closed surface of complexity $n$, genus $g$, and longest-to-shortest edge ratio $\alpha$. Given a system of loops on $\mathcal{M}$, with each loop composed of at most $m$ edges, there is an algorithm that computes a homotopic system with minimal length in $O(g^2\alpha m^2(g\alpha m + n)\log(g\alpha m + n))$ time.*

By applying the data structures of Henzinger *et al.* [11], the logarithmic factor in the running time can be removed.

If $m = O(n)$, the time complexity reduces to $O(g^3\alpha^2 n^3)$. If $n = O(m)$, we obtain a $O(g^3\alpha^2 m^3)$ time complexity. In both cases, we get a cubic dependency on the input data, disregarding the genus or geometric parameter $\alpha$. This rather coarse analysis can be refined by introducing a multiplicity parameter.

We define the *multiplicity* of a simple combinatorial loop $R$ as follows. For each vertex $v$ of $\mathcal{M}$, consider the parenthesized word formed by the pairs of consecutive edges of $R$ meeting at $v$. The *multiplicity* of vertex $v$ (w.r.t. $R$) is the maximal number of nested parentheses in (any cyclic permutation of) this word. Thus any 2-path $a_1, a_2$ crosses $R$ a number of times that is at most the multiplicity of $R$ at the target of $a_1$. Define the multiplicity of $R$ to be the maximal multiplicity of any vertex in $\mathcal{M}$; thus, if a closed walk $C$ with $k$ edges is introduced in the combinatorial set of loops made of the single loop $R$, the number of crossings between $C$ and $R$ is no more than $k$ times the multiplicity of $R$.

Also, $R$ has a number of edges bounded by $n$ times its multiplicity, assuming it has no spur (*i.e.*, no consecutive edges on the same edge of $G$).

Let $\mu$ be the maximum, over $j$, of the multiplicity of the combinatorial loop $\rho(s_j)$. Hence (assuming no spur in $\rho(s)$), the number of edges of a combinatorial loop at the beginning of the algorithm is $O(\mu n)$, and its maximal number of edges in the course of the algorithm is $O(\alpha\mu n)$. Any system of loops considered can have complexity $O(g\alpha\mu n)$, whence the complexity of an elementary step is $O(g\alpha\mu n \log(\alpha\mu n))$.

Let $T_i$ be defined as in the previous analysis. Its complexity is $O(\alpha\mu n)$. According to the above observations, the lengths of the crossing words is $O(g\alpha\mu^2 n)$. We also note that any spurs in the input can be removed beforehand. Putting all together we obtain:

**Theorem 13** *Given a system of loops on an orientable combinatorial closed surface $\mathcal{M}$, such that the maximal multiplicity of each loop is $\mu$, there is an algorithm that computes a homotopic system with minimal length in $O(\alpha^2\mu^3 g^3 n^2 \log \alpha\mu n)$ time. ($m$, $n$, $g$, and $\alpha$ are defined as in the previous theorem.)*

Again, the logarithmic factor in the running time can be removed. Also remark that each loop of any system of loops constructed from a spanning tree of a maximal non-deconnecting graph (see, for example, Dey and Schipper [6]) has multiplicity 2. Following the algorithm of Lazarus *et al.* [13], we can even construct in time $O(gn)$ a *canonical* system where each loop has multiplicity 2.

## 6 Shortest simple loop

In this section, we apply the computation of optimal systems to obtain an algorithm for the computation of a shortest simple loop homotopic (relative to a fixed basepoint) to a given simple loop.

We say that a combinatorial loop $T = \rho(t)$ is *non-separating* if it is the image by $\rho$ of a regular simple non-separating loop $t$. Any non-separating simple loop can be augmented to form a system of loops with roughly the same multiplicity:

**Lemma 14** *Let $T$ be a non-separating simple combinatorial loop on $\mathcal{M}$ with multiplicity $\mu$. Then, $T$ can be completed in $O(g\mu n)$ time to a combinatorial system of loops for $\mathcal{M}$, each loop having multiplicity $O(\mu)$.*

PROOF. While the result may seem intuitively clear, the details are not so obvious. Because these details are also helpful for an implementation, we give a precise construction of a system of loops. To this end we will subdivide $\mathcal{M}$ with the help of some $t \in \rho^{-1}(T)$, in order to simulate a combinatorial loop of multiplicity one. The completion to a system of loops with constant multiplicity will follow easily, and we eventually contract the subdivided surface with its system to its original subdivision. The tricky part is to show that this contraction provides a system with the required multiplicity.

Our first task is to "clean up" $T$ so as to simplify the construction of a $t \in \rho^{-1}(T)$. We may assume that $\mathcal{M}$ is triangulated. If not, we can triangulate each polygonal face of $\mathcal{M}$ by starring around one of its vertices. This only adds edges to $\mathcal{M}$, does not change the multiplicity of $T$, and does not modify the order of complexity of $\mathcal{M}$. We may also assume that two consecutive edges of $T$ are not incident to the same triangle of $\mathcal{M}$, except for the first and last edge of $T$. If not, we can replace two such edges by the third edge of the incident triangle. Doing so for all pairs of edges in $T$ will not introduce crossing and can only reduce the multiplicity of a vertex w.r.t. $T$. Note that this simplification can be performed (recursively, together with the removal of spurs) in time proportional to the complexity of $T$.

We now construct explicitly a simple $t \in \rho^{-1}(T)$: first insert, at an arbitrary place in the ordered list of each edge of $G$, an element representing the edge itself. Consider the circular list, formed by such augmented ordered lists, of a vertex $v$ of $G$. Break this circular list to obtain a linear list. Each element representing an edge $e$ in this list is surrounded by a number $n_e \leq \mu$ of nested parentheses corresponding to consecutive pairs of edges of $T$. Subdivide $e$ near $v$ by inserting $n_e$ vertices. Each such new vertex can thus be associated a consecutive pair of edges of $T$. The order on these vertices along $e$ should reflect the order of the nested parentheses. Let $V(v)$ be the set of vertices introduced this way near $v$. It is now easy to link such new vertices with line segments, while traversing $T$, to form a simple loop $t \in \rho^{-1}(T)$.

We next subdivide $\mathcal{M}$ with $t$. Call $\mathcal{M}'$ the subdivided surface. Note that $\mathcal{M}$ is obtained from $\mathcal{M}'$ by identifying each set of the form $V(v) \cup v$ to a single vertex. In this process $t$ contracts to $T$. Following, for example, Dey and Schipper [6], it is easy to form a system $S'$ of loops for $\mathcal{M}'$, each of the loops having multiplicity 2, that contains $t$. For completeness, we briefly recall the technique: augment $t$ to form a maximal non-disconnecting subgraph $G'$ of the vertex-edge graph of $\mathcal{M}'$. Select a spanning tree $\mathcal{T}'$ in $G'$ containing $t$ minus one of its edges. For each non-tree edge $e$ in $G'$ construct a loop by concatenating $e$ with the two simple paths in $\mathcal{T}'$ joining its extremities to the basepoint. It is not difficult to show that we obtain this way the required combinatorial system of loops [13].

Contracting back $\mathcal{M}'$ to $\mathcal{M}$, the system $S'$ contracts to a combinatorial system $S$ on $\mathcal{M}$ that contains $T$. It remains to show that the multiplicity of the loops in $S$ is $O(\mu)$.

Let $R$ be a combinatorial loop in $S$ deduced by contraction of a combinatorial loop $R'$ in $S'$. For a vertex $v$ in $\mathcal{M}$ (resp. $\mathcal{M}'$) we consider the cyclic parenthesized word formed by the pairs of consecutive edges of $R$ (resp. $R'$) meeting at $v$. This cyclic word can be encoded by a plane tree whose oriented edges are labeled with edges of $R$ (resp. $R'$) incident with $v$. (See Figure 10.) We call this tree the *word-tree* of $v$. The multiplicity of $v$ is easily seen to be the diameter (the length of the longest simple path) of its word-tree. If an edge $(y, z)$ of $\mathcal{M}$ is contracted to a vertex $x$, then the word-tree of $x$ results from a pasting of the word-trees of $y$ and $z$. It follows that $x$ has a multiplicity bounded by the sum of the multiplicities of $y$ and $z$.

Now, consider a vertex $v$ in $\mathcal{M}$. Its word-tree is obtained by merging in some way

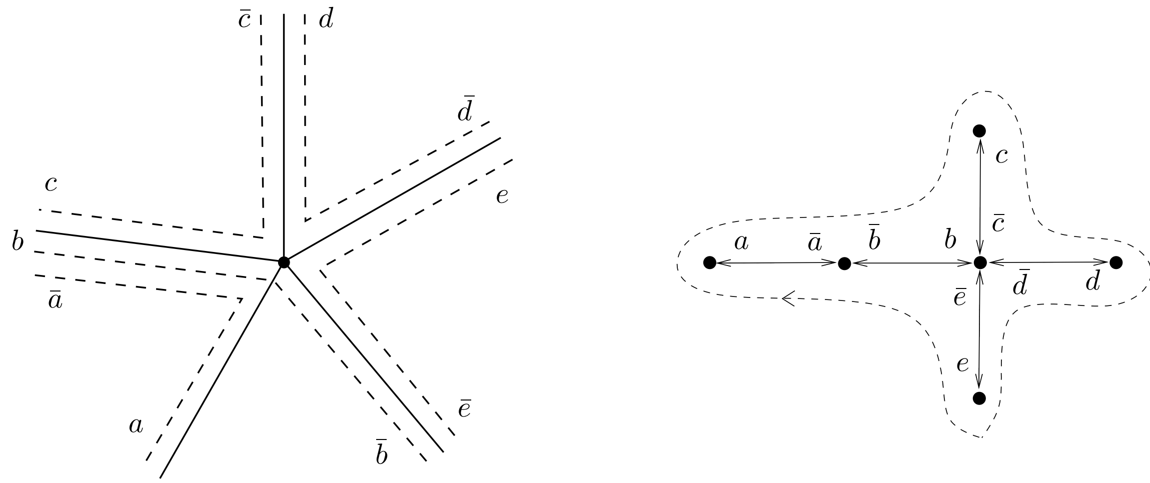


Figure 10: The cyclic parenthesized word formed by the pairs of consecutive edges of a combinatorial loop (dashed lines on the left) can be represented by a plane labeled tree (on the right). Walking around the tree, we get the cyclic word $a\bar{a}bc\bar{c}d\bar{d}e\bar{e}\bar{b}$ by concatenation of the labels of the oriented edges. The multiplicity of the vertex equals the diameter (3) of the tree.

the word-trees of the vertices in $V(v) \cup v$ on $\mathcal{M}'$. More precisely, we first merge, by edge contractions, the $n_e$ vertices inserted in each edge $e$ incident to $v$ in $\mathcal{M}$ to get a vertex $v_e$. Since $R'$ has multiplicity 2, the word-tree of $v_e$ has diameter at most $2n_e \leq 2\mu$. Then, the vertex $v$ in $\mathcal{M}$ results from merging the vertices $v_e$ with the vertex $v$ in $\mathcal{M}'$. This merging amounts to paste the $v_e$ word-trees to the word-tree of $v$ in $\mathcal{M}'$. It follows that the word-tree of $v$ in $\mathcal{M}$ has a diameter bounded by twice the diameter of any $v_e$ word-tree plus 2. This in turn implies that the multiplicity of $v$ in $\mathcal{M}$ is bounded by $4\mu + 2$. □

**Theorem 15** *Let $T$ be a simple combinatorial loop on $\mathcal{M}$ with multiplicity $\mu$. Among all the simple combinatorial loops homotopic to $T$, a shortest one can be computed in $O(\alpha^2\mu^3g^3n^2 \log \alpha\mu n)$ time. (The parameters $\alpha$, $g$, and $n$ are defined as in Theorem 13.)*

PROOF. First suppose that $T$ does not separate $\mathcal{M}$. By the preceding lemma, we can construct in $O(g\mu n)$ time a combinatorial system with loops of multiplicity $O(\mu)$ that contains $T$. It remains to apply Theorem 13 to compute an optimal system containing a simple loop homotopic to $T$; by Theorem 2, this is the desired loop.

Suppose on the contrary that $T$ separates $\mathcal{M}$. Consider $t \in \rho^{-1}(T)$ as constructed in Lemma 14. The loop $t$ cuts $\mathcal{M}$ into two surfaces $\mathcal{M}_1$ and $\mathcal{M}_2$, each with one boundary cycle. Close each component by gluing a single face to the boundary cycle, and construct a system of loops for each component whose basepoint corresponds to the original basepoint on $t$. The union of these two systems is a system $s$ for $\mathcal{M}$ that does not cross $t$. By contraction, as in Lemma 14, $s$ provides a combinatorial system $S$ with loops of multiplicity $O(\mu)$.

Applying Theorem 13, we compute an optimal system $S'$ homotopic to $S$. Then, we compute a shortest loop $T'$ homotopic to $T$ that does not cross $S'$: this reduces to find a shortest path in a topological disk (the polygonal schema associated to $S'$), hence $T'$ is simple. We claim that $T'$ is a shortest simple loop homotopic to $T$. Indeed, let $T''$ be a shortest simple loop homotopic to $T'$; as in the proof of Proposition 4, $[S'/T'']$ (or more precisely, $[s'/t'']$, with $s' \in \rho^{-1}(S')$ and $t'' \in \rho^{-1}(T'')$) reduces to the empty word; as in the proof of Lemma 10, there exists a simple loop homotopic to $T''$ that does not cross $S'$; hence $T'$ and $T''$ have the same length. □

# 7 Discussion

Our work suggests further research in several directions. Considering a loop or a system of loops as embedded graphs, a natural extension of this work concerns the optimization of the embedding of a general graph on a surface. In the plane, the problem has been studied by Bespamyatnikh [1] and Efrat *et al.* [7]). The algorithm described in the present paper extends to the optimization of graph embeddings, as proved recently [3]: one can compute the shortest graph embedding isotopic, with fixed vertices, to a given graph embedding.

Relieving the vertices from being fixed leads to the study of optimal cycles in a free homotopy class. In a recent paper [4], we proved that the use of similar techniques can help in computing the shortest cycle (*i.e.*, loop without basepoint) freely homotopic to a given simple cycle. The idea is to use a *pants decomposition* of the surface, which plays the same role as a system of loops in the present paper.

Another interesting continuation of this work is to replace the combinatorial systems by piecewise linear systems using the induced metric on some polyhedral surface immersed into $\mathbb{R}^3$. This would somehow extend the works of Hershberger and Snoeyink [12], Bespamyatnikh [1], and Efrat *et al.* [7] to more general surfaces.

Our work also suggests some open questions. What is the influence of the basepoint position? How does one compute the shortest system of loops, among all systems, relaxing the homotopy condition? Comparing with the work of Erickson and Har-Peled [9], we expect this last problem to be much less tractable than those solved in the present paper. Finally, it would be nice to get rid of the longest-to-shortest edge ratio in the complexity analysis.

## Acknowledgement

We thank Michel Pocchiola for interesting discussions and comments. We are also especially grateful to one of the anonymous referees for his numerous valuable suggestions which allowed us to greatly improve the presentation of the paper.

## References


[1] S. Bespamyatnikh. Computing homotopic shortest paths in the plane. In *Proc. 14th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 609–617. ACM Press, 2003.

[2] H. R. Brahana. Systems of circuits on 2-dimensional manifolds. *Annals of Mathematics*, 23:144–168, 1921.

[3] É. Colin de Verdière. *Shortening of curves and decomposition of surfaces*. PhD thesis, University of Paris 7, Dec. 2003. Available at `http://www.di.ens.fr/~colin/`.

[4] É. Colin de Verdière and F. Lazarus. Optimal pants decompositions and shortest homotopic cycles on an orientable surface. In *Proc. 11th Symposium on Graph Drawing*, volume 2912 of *LNCS*, pages 478–490. Springer Verlag, 2003.

[5] É. Colin de Verdière and F. Lazarus. Optimal system of loops on an orientable surface. In *Proc. IEEE Symposium on Foundations of Computer Science*, pages 627–636, 2002.

[6] T. Dey and H. Schipper. A new technique to compute polygonal schema for 2-manifolds with application to null-homotopy detection. *Discrete & Computational Geometry*, 14:93–110, 1995.

[7] A. Efrat, S. Kobourov, and A. Lubiw. Computing homotopic shortest paths efficiently. In *Proc. 10th European Symposium on Algorithms*, pages 411–423, 2002.

[8] D. Epstein. Curves on 2-manifolds and isotopies. *Acta Mathematica*, 115:83–107, 1966.

[9] J. Erickson and S. Har-Peled. Optimally cutting a surface into a disk. *Discrete & Computational Geometry*, 31(1):37–59, 2004.

[10] M. Floater. Parametrization and smooth approximation of surface triangulations. *Computer Aided Geometric Design*, 14(3):231–250, 1997.

[11] M. Henzinger, P. Klein, S. Rao, and S. Subramanian. Faster shortest-path algorithms for planar graphs. *Journal of Computer and System Sciences*, 55(1, part 1):3–23, 1997.

[12] J. Hershberger and J. Snoeyink. Computing minimum length paths of a given homotopy class. *Computational Geometry: Theory and Applications*, 4:63–98, 1994.

[13] F. Lazarus, M. Pocchiola, G. Vegter, and A. Verroust. Computing a canonical polygonal schema of an orientable triangulated surface. In *Proc. 17th Annual Symposium on Computational Geometry*, pages 80–89, 2001.

[14] R. Lyndon and P. Shupp. *Combinatorial Group Theory*, volume 89 of *A Series of Modern Surveys in Mathematics*. Springer-Verlag, 1977.

[15] J. Maillot, H. Yahia, and A. Verroust. Interactive texture mapping. In *SIGGRAPH 93*, pages 27–34, 1993.

[16] W. S. Massey. *Algebraic Topology: An Introduction*, volume 56 of *Graduate Texts in Mathematics*. Springer-Verlag, 1977.

[17] D. Piponi and G. Borshukov. Seamless texture mapping of subdivision surfaces by model pelting and texture blending. In *SIGGRAPH 2000*, pages 471–478, 2000.

[18] J. Stillwell. *Classical topology and combinatorial group theory*. Springer-Verlag, New York, 1993.

[19] G. Vegter and C. K. Yap. Computational complexity of combinatorial surfaces. In *Proc. 6th Annual ACM Symposium on Computational Geometry*, pages 102–111, 1990.

# APPENDICES

# A Implementation

We have implemented the algorithm described in Section 5 using the C++ based CGAL library.

In order to make comparisons between combinatorial and PL "geodesics", we also implemented a simple local optimization that produces geodesic loops *on the surface* of $\mathcal{M}$: we visit each vertex star of $\mathcal{M}$ and replace pieces of loops in the star by shortest paths in the star. We repeat this operation until the shortening gain is below a given threshold. The resulting loops are geodesics (not necessarily the shortest ones) and keep their homotopy class.

Figure 11 shows a simple example run on a genus 2 torus with 1536 facets. Euclidean distances were used for the edges. More pictures can be seen on http://www-sic.univ-poitiers.fr/lazarus/opt-sys.html.

# B The size of the universal cover

Let $\mathcal{M}$ be a combinatorial surface of genus $g \geq 2$, with $n$ edges, and let $v_0$ be a fixed vertex on $\mathcal{M}$. We wish to show the following lemma:

**Lemma 16** *The number of vertices in the (vertex-edge graph of the) universal cover of $\mathcal{M}$ at a distance at most $k$ from a given lift, $v_0^\epsilon$, of $v_0$ can be (at least) exponential in $k$.*

PROOF. To this end we reduce the problem to finding the number of lifts of $v_0$ at a distance at most $k$ from $v_0^\epsilon$. We further simplify the computation by using topological distances, i.e. assuming that all edges have unit length.

Consider a presentation of the fundamental group of $\mathcal{M}$ made of $2g$ generators and of a single cyclically reduced relation. Suppose in addition that some proper subset of $p > 1$ generators have representatives made of $O(1)$ edges. (Suitable surfaces can easily be constructed by connecting small handles through a common base-vertex $v_0$.) Then, any word $w$ on these generators and their inverses corresponds to a homotopy class that can be realized by a loop with $O(|w|)$ edges. Equivalently, $w$ corresponds to a lift of $v_0$ at a distance $O(|w|)$ from $v_0^\epsilon$.

Now, it is known [14, p. 104] that any proper subset of the generators constitutes a basis for a free subgroup in the fundamental group of $\mathcal{M}$. Whence the number of reduced words of length $m$ on the $p$ generators and their inverses is $\Omega((2p-1)^m)$. But these correspond to as many lifts of the basepoint at a distance $O(m)$ from $v_0^\epsilon$. □

Furthermore, we can assume that some shortest loop in a schema has size $\Omega(n)$, either considering a "large" handle (in addition to the previously evoked small handles) or a simple

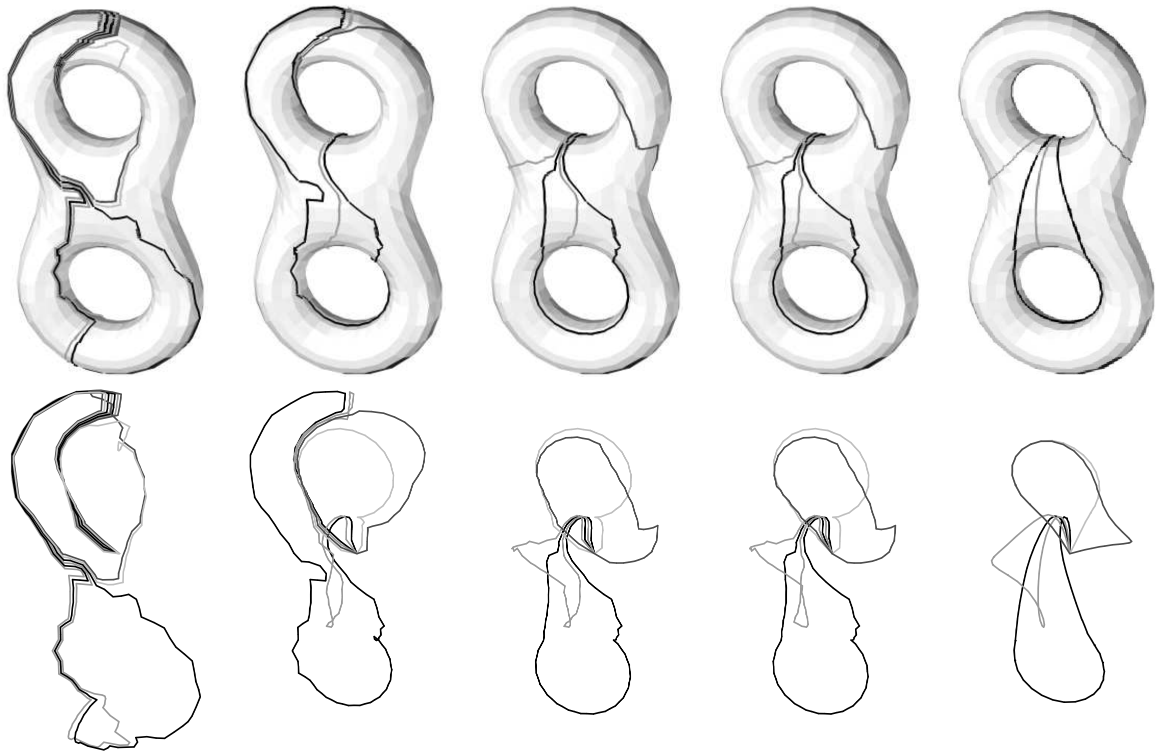


Figure 11: A: A canonical system $s$ computed by the algorithm of Lazarus *et al.* [13]. The basepoint is on the back side of the double torus. B: $f(s)$. C: $f^2(s)$. D: $f^3(s) = f^4(s)$. E: The local optimization was applied to this optimal system to get a geodesic system on the surface (4,000 star optimizations were performed).

loop expressed as a large product of small loops. Hence shortening this loop would *a priori* involve considering $(2p-1)^{\Omega(n)}$ vertices.

## C Test for termination

We show here that Theorem 2 holds without the additional assumption on the computation of the elementary step $f_j(s)$ described in the proof of this theorem: here we may have $|s_j| = |f(s)_j|$ while $s_j \neq f(s)_j$. To prove this result, it suffices to prove that whenever a system of loops $s$ has the same length as $f(s)$, then $s$ is optimal. Hereafter, we assume $|s| = |f(s)|$. The loop $t_i$ is simple, homotopic to $s_i$, and as short as possible.

Without loss of generality, we write $f = a \circ \bar{f}_j \circ b \circ f_j \circ c$, where $a$, $b$ and $c$ are compositions of elementary steps other than $f_j$ and $\bar{f}_j$. Changing the orientation of $s_j$ would indeed allow to invert the occurrences of $f_j$ and $\bar{f}_j$ in $f$. By re-indexing the loops of $s$, if necessary, we can also assume that $f_k$ occurs before $f_l$ if $k < l$. Let $q \in [0, 4g]$. If the $q$th elementary step in $f$ consists of replacing a loop $\ell$ by a shorter loop $\ell'$ in the current system, we say that $\ell'$ *appeared at time* $q$ and that $\ell$ *disappeared at time* $q$. In particular, the loops in $s$ appeared at time zero and $s_1$ disappeared at time 1.

A crossing word $w$ is said *left $j$-reduced* (resp. *right $j$-reduced*) if $w = g_j(w)$ (resp. $w = \bar{g}_j(w)$).

**Lemma 17** *There exists a simple loop $t'_i$, homotopic to and no longer than $t_i$, such that $[s/t_i]$ reduces to $[s/t'_i]$, and for any $j \in [1, 2g]$, $[s/t'_i]$ is left $j$-reduced.*

PROOF. Suppose that a left $j$-reduction is possible on $[s/t_i]$ for some $j \in [1, 2g]$. Then, as in the proof of Lemma 10, we can exhibit an empty lens bounded by $s_j$ and $t_i$, with $t_i$ to the left of $s_j$. Call the segments of $s_j$ and $t_i$ bounding this lens $\bar{s}_j$ and $\bar{t}_i$, respectively. We shall prove that $\bar{s}_j$ and $\bar{t}_i$ have the same length. It will follow that we can modify $t_i$ without changing its length nor its homotopy class to proceed to the left $j$-reduction in $[s/t_i]$.

Suppose the factor $c$ before $f_j$ in $f$ is composed of $p$ elementary steps. For any $q \in [1, p]$, define $c_q$ as the composition of the $q$ first elementary steps. We show by induction on $q$ that $\bar{t}_i$ can be replaced by a path $\tau_q$ with the same length as and homotopic to $\bar{t}_i$, which crosses no loop of $c_q(s)$, for any $q \in [1, p]$. (If $j = 1$, then $c$ must be an empty composition and there is nothing to prove.)

Consider $\tau_q$ as in the induction hypothesis, with $q < p$. Let $\ell$ be the loop that appeared at time $q + 1$. Only $\ell$ can cross the lens bounded by $\tau_q$ and $\bar{s}_j$; applying the Jordan curve theorem to (a lift of) this lens, we see that $[c_{q+1}(s)/\tau_q]$ reduces to the empty word (possibly using extremal reductions if one corner of the lens is a lift of the basepoint). By the minimality of $\ell$, all lenses between $\tau_q$ and $\ell$ have their two subpaths of equal length, and we can iteratively remove all the crossings between them. This process replaces $\tau_q$ by

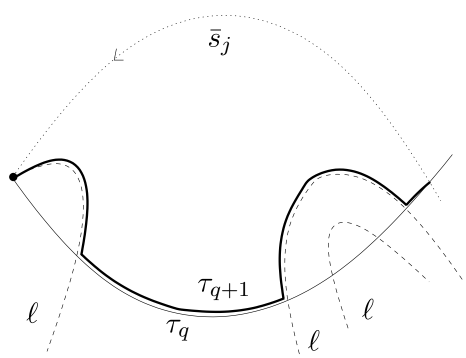


Figure 12: The induction step in Lemma 17.

a homotopic path $\tau_{q+1}$ of the same length that does not cross $c_{q+1}(s)$; see Figure 12.

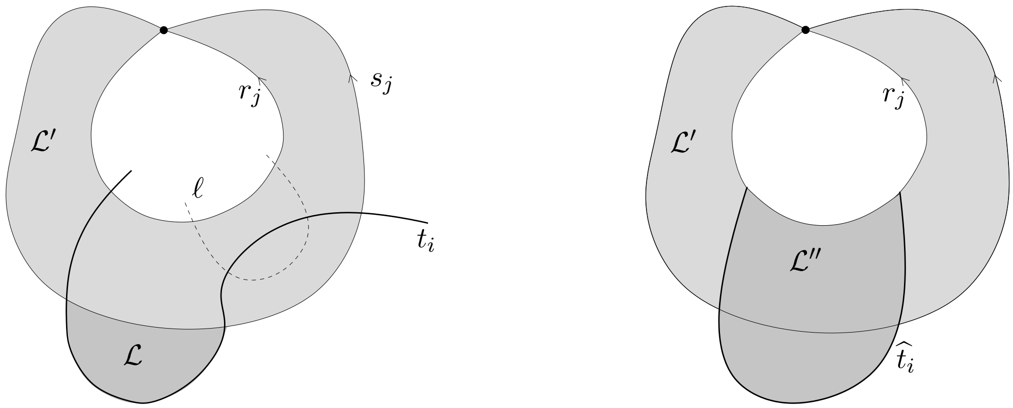


Figure 13: The situation on the left cannot occur: the loop $\ell$ must have disappeared before $r_j$ appeared; hence $\ell$ has an index (as a loop of $s$) smaller than $j$ and cannot form an empty lens with $t_i$ as this would either contradict that $[s/t_i]$ is left-reduced or that $j$ is minimal, depending on the orientation of $\ell$.

By induction, we can replace $\bar{t}_i$ by a path $\tau_p$ with the same length as and homotopic to $\bar{t}_i$, which crosses no loop of $c_p(s) = c(s)$. Furthermore, $\bar{s}_j$ and $\tau_p$ form a lens with $\tau_p$ to the left of $\bar{s}_j$. Because $|s_j| = |f(s)_j|$ by assumption, $s_j = c(s)_j$ must have the same length as $f_j(c(s))_j$. It follows that $\bar{s}_j$ cannot be longer than $\tau_p$. Hence $|\bar{s}_j| = |\tau_p| = |\bar{t}_i|$. □

**Lemma 18** *There exists a simple loop $t'_i$, homotopic to and no longer than $t_i$, such that $[s/t_i]$ reduces to $[s/t'_i]$, and for any $j \in [1, 2g]$, $[s/t'_i]$ is right $j$-reduced.*

PROOF. From the previous lemma we can assume that $[s/t_i]$ is left-reduced. Suppose that a right $j$-reduction is possible. Choose $j$ minimal among the possible right $j$-reductions. Then, we can exhibit an empty lens $\mathcal{L}$ bounded by $s_j$ and $t_i$, with $t_i$ to the right of $s_j$. Call again $\bar{s}_j$ and $\bar{t}_i$ the two respective subpaths of $s_j$ and $t_i$ bounding this lens.

As for Lemma 17, we shall prove that some part of $t_i$, including $\bar{t}_i$, can be modified without changing its length nor its homotopy class to proceed to – at least – a right $j$-reduction in $[s/t_i]$.

Following the proof of Lemma 17, we can assume that $\bar{t}_i$ crosses no loop of $c(s)$. Put $r_j = f_j(c(s))_j$. We have the following situation; $\bar{s}_j$ and $\bar{t}_i$ form the lens $\mathcal{L}$ with $\bar{t}_i$ to the right of $\bar{s}_j$, and $r_j$ is on the left of $s_j$; let us call $\mathcal{L}'$ the space bounded by $s_j$ and $r_j$. If a corner of $\mathcal{L}$ is interior to $s_j$, then the loop $t_i$ enters $\mathcal{L}'$ through this corner and must exit $\mathcal{L}'$ through a point interior to $r_j$: it cannot exit across $s_j$ as this would either contradict that $[s/t_i]$ is left-reduced or that $j$ is minimal. (See Figure 13, left, for an illustration.)

Hence, there are a subpath $\widehat{t}_i$ of $t_i$ containing $\bar{t}_i$ and a subpath $\widehat{r}_j$ that form a lens $\mathcal{L}'' \subset \mathcal{L} \cup \mathcal{L}'$; see Figure 13, right. We first claim that $\widehat{t}_i$ and $\widehat{r}_j$ have the same length.

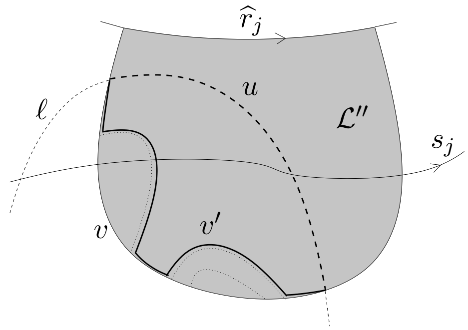


Figure 14: The induction step in Lemma 18.

Proof of this claim. Put $s' = b(f_j(c(s)))$. Suppose $b \circ f_j \circ c$ is composed of $p$ elementary steps. Let $\ell$ be a loop that appeared at time $q$, for some $q \in [0, p]$. We consider a subpath $u$ of $\ell \cap \mathcal{L}''$ with the following property (P): $u$ is a chord of $\widehat{t_i}$ and, if $u$ crosses $\mathcal{L}'$, the extension of $u$ in $\ell$ leaves $\mathcal{L}'$ through $s_j$. Let $v$ be the subarc of $\widehat{t_i}$ sustained by $u$. We show by induction on $q$ that $u$ and $v$ have equal length.

If $q = 0$, there is no subpath $u$ of $\ell$ with the property (P) for $\ell$ is in $s$ and cannot cross $s_j$ (nor $\mathcal{L}$ by assumption). So, suppose the induction hypothesis is true for some $q \leq 0$ and consider the loop $\ell$ that appeared at time $q+1$. Let $u$ and $v$ be as above. By replacing, if necessary, $u$ by an innermost subpath of $\ell$ in the lens formed by $u$ and $v$, we can assume that $\ell$ does not cross this lens. If the lens is empty at time $q+1$, then we must have $|u| = |v|$ by the minimality of $\ell$. Otherwise, this lens contains subpaths of loops that appeared before $\ell$. These subpaths have the required property (P) with respect to these loops and, by the induction hypothesis, they must have the same lengths as the subarcs of $v$ they sustain. We can thus replace $v$ by a homotopic arc $v'$ with equal length so that the lens between $u$ and $v'$ is empty at time $q+1$; Figure 14 depicts the situation. It follows from the minimality of $\ell$ that $|u| = |v'| = |v|$.

Now, let $u$ be an arc of the intersection of a loop of $s' = b(f_j(c(s)))$ with $\mathcal{L}''$. This arc does not intersect $r_j$ since $s'$ contains $r_j$. Hence, $u$ has the above property (P) and must have the same length as the subarc it sustains on $\widehat{t_i}$. Again, this means that we can replace the subpath $\widehat{t_i}$ of $t_i$ by a homotopic path $w$ with equal length so that no loop of $s'$ crosses the lens between $\widehat{r_j}$ and $w$. Now, $w$ is to the right of $\widehat{r_j}$ so they must have the same length; otherwise, the elementary step $\bar{f_j}$ would have next decreased the length of $r_j$, a contradiction. □

Next, we consider a component $u$ in $s \cap \mathcal{L}'$ that crosses $\mathcal{L}''$. The arc $u$ is supported by a loop $h$ that disappeared before $r_j$ appeared. Again, this implies that $u$ cannot form a lens with $t_i$ (by the minimal choice of $j$ or the left-reduction of $[s/t_i]$). Hence $u$ must be of one of the three following types (see Figure 15, left):

1. it either crosses the two components of $\widehat{t_i}$ in $\mathcal{L}'$,

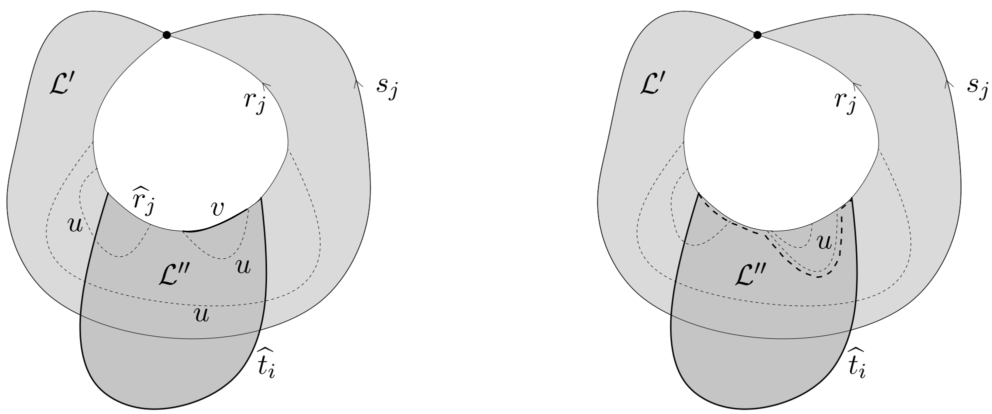


Figure 15: Left: The three types of component in $s \cap \mathcal{L}'$ that crosses $\mathcal{L}''$. Right: The subpath $\widehat{t}_i$ of $t_i$ can be replaced by the thick dashed curve.

2. or has one extremity on $\widehat{r}_j$,
3. or has both of its extremities on $\widehat{r}_j$.

In the last case we claim that $u$ must have the same length as its chord $v$ on $\widehat{r}_j$.

Proof of this claim. We can restrict to the case where $u$ is an innermost arc; the general case follows from a recursion on the level of nesting of arcs of type 3. If $v$ is to the left of $u$, then considering the elementary step when $h$ disappeared, it is easily seen that $u$ cannot be longer than $v$; for the length of $h$ would have decreased otherwise. Now, if $v$ is to the right of $u$, call $k$ the loop that replaced $h$ when $h$ disappeared and call $\ell$ the loop that replaced $k$ when $k$ disappeared. We know that $h$ disappeared before $r_j$ appeared, $k$ is to the left of $h$, and $k$ does not intersect $s_j$. As a consequence, $k$ intersects $r_j$ and must have disappeared before $r_j$ appeared. It follows that $\ell$ does not intersect $\mathcal{L}'$. Hence, the lens $\mathcal{L}_{k\ell}$ formed by $k$ and $\ell$ contains the lens formed by $u$ and $v$. We eventually exhibit a simple loop $h'$ in $\mathcal{L}_{k\ell}$ with the following property: $h'$ is homotopic to $h$, $|h'| = |h|$, and $h'$ contains $u$. This will in turn imply that $u$ cannot be longer than $v$, since otherwise $k$ could have been shortened. To construct $h'$, we consider the lenses formed by $h$ and $\ell$ and for each such lens outside $\mathcal{L}_{k\ell}$, we replace in $h$ the side on $h$ by the side on $\ell$; see Figure 16. This clearly produces a loop with the desired property. □

We can eventually replace $\widehat{t}_i$ in $t_i$ by an arc inside $\mathcal{L}''$ that follows the outermost arcs of type 3 as considered in the last claim. Figure 15, right, illustrates this replacement. We obtain a simple loop $t'_i$ no longer than and homotopic to $t_i$. $[s/t'_i]$ is deduced from $[s/t_i]$ by a right $j$-reduction and possibly other reductions if some arcs of $s \cap \mathcal{L}'$ are of type 1. Note that the intersections of $\widehat{t}_i$ with arcs of type 2 have been moved, but their number and order on $t'_i$ are left unchanged. □

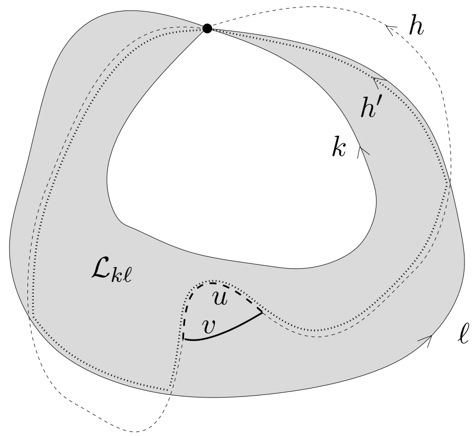


Figure 16: $\mathcal{L}_{k\ell}$ contains the lens with sides $u$ and $v$ as well as a simple loop $h'$ containing $u$, homotopic to $h$, and with the same length as $h$.

**Corollary 19** *If* $|s| = |f(s)|$, *then* $s$ *is optimal.*

PROOF. This is a direct consequence of Proposition 4, Lemma 11 and the last two lemmas. □