

Optimal System of Loops on an Orientable Surface*

Éric Colin de Verdière[†]

Francis Lazarus[‡]

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 fundamental system of loops for \mathcal{M} . The resulting disk is a polygon in which the edges are pairwise identified on the surface; it is called a polygonal schema. Assuming that \mathcal{M} is triangulated, and that each edge has a given length, we are interested in a shortest (or optimal) system 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 a polynomial (under some reasonable assumptions) algorithm to build such a system. As a byproduct, we get a polynomial algorithm to compute a shortest simple loop homotopic to a given simple loop.

1. Introduction

1.1. Background and previous work

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 . We focus on surfaces homeomorphic to a g -torus ($g > 0$). It is a well-known fact that such a surface can be obtained from a $4g$ -gon by pairwise identifications of its edges; such a polygon is called a *polygonal schema*. For a general reference on this subject, see for example [11, Chapter 1.4].

*Partially supported by the IST Programme of the EU as a Shared-cost RTD (FET Open) Project under Contract No IST-2000-26473 (ECG - Effective Computational Geometry for Curves and Surfaces). A preliminary short version of this paper appeared in the Abstracts of the 18th European Workshop on Computational Geometry, April 2002.

[†]Laboratoire d'informatique de l'École normale supérieure, UMR 8548 (CNRS), Paris, France. Eric.Colin.de.Verdiere@ens.fr

[‡]Laboratoire IRCOM-SIC, UMR 6615 (CNRS), Poitiers, France. Francis.Lazarus@sic.univ-poitiers.fr.

In this paper, we consider the case of *reduced* polygonal schemata: all the vertices of the polygon get identified to a single point v_0 , the *basepoint*, on the surface. This corresponds to a cutting of the surface along simple loops having v_0 as common vertex, pairwise disjoint except at v_0 , such that the complement of the union of these loops is a topological disk. Such a set of loops is called a *fundamental system of loops*, or *system* for short. A system is called *canonical* if, in addition, the loops around the corresponding reduced polygonal schema appear in the order of the form $S_1, S_2, \bar{S}_1, \bar{S}_2, S_3, S_4, \bar{S}_3, \bar{S}_4, \dots$, where the bar indicates opposite orientation.

Computing a (canonical) 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 [4] and texture mapping [9, 10]. Canonical systems also allow to construct homeomorphisms between surfaces of 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.

The surface \mathcal{M} is assumed to be triangulated (or, more generally, to be a polyhedral 2-manifold – the faces are arbitrary simple polygons, and the intersection of two faces is either empty, a vertex, or a common edge). In a combinatorial setting, the loops are closed paths on the vertex-edge graph \mathcal{G} of \mathcal{M} ; to mimic the continuous setting, two loops of a system are allowed to go along a same edge of \mathcal{G} , provided that they “do not cross” if we conceptually spread them apart with a thin space. This informal idea yields a rigorous combinatorial structure, the *edge-ordered set of loops*. Furthermore, we assume that each edge of \mathcal{G} has a positive *length*, or *weight*; the length of a system is the sum of the lengths of its loops.

Lazarus, Pocchiola, Vegter, and Verroust [8] gave two methods to compute a canonical system of loops on a triangulated surface. While both their algorithms have worst-case optimal asymptotic complexity, they usually produce jaggy and irregular loops as they do not take into account the geometry of the surface. The work by Erickson and Har-Peled [3] partly overcomes the geometric aspect: they study the problem of cutting a combinatorial surface into a topo-

logical disk whose boundary has minimal length. However, their method leads to schemata that are neither reduced nor canonical, hence not suited to the construction of homeomorphisms between surfaces.

1.2. Novelty of this paper

We say that two systems with the same basepoint are *homotopic* if the sets of homotopy classes of the loops in both systems are the same. A system with minimal length in its homotopy class is called *optimal*.

This paper is based on a conceptually very simple Elementary Step which transforms a system into another homotopic system by shortening one loop as much as possible while keeping the other ones fixed. A natural question is to ask what we obtain when this process is iterated forever. Quite surprisingly, this simple iterative scheme converges (in a polynomial number of steps) to an optimal system. We prove that:

- the iterative scheme reaches stability in length and yields a system in which each loop is a shortest loop among all simple loops in its homotopy class (hence this system is optimal). This directly implies a theoretical, non-trivial fact: any optimal system is made of shortest simple homotopic loops;
- this scheme can be implemented efficiently, leading to an algorithm which, given a system S , computes a homotopic optimal system in time $O(\mu^5 \alpha^3 g^3 n^3 \log(\mu \alpha n))$, where n is the combinatorial complexity of the surface, α is the longest-to-shortest edge ratio, and μ is the maximal number of times a given loop in S passes through a given vertex in M ;
- these results can be used to compute a shortest simple loop homotopic to a given simple loop.

A slightly different version of the algorithm (yielding the same result) has been implemented.

Let us stress out that these results are *a priori* non-obvious. First of all, a shortest loop homotopic to a simple loop may itself not be simple; hence, computing a shortest system homotopic to a given system cannot be obtained by just searching for a shortest loop homotopic to each loop in the system. Even if we find shortest *simple* homotopic loops, it could still happen that these loops intersect. Furthermore, consider the related problem of finding a shortest loop within a given homotopy class. A natural tool for this is the universal covering of M , since the problem reduces to find a shortest path in this space. However, if the shortest path we are looking for is composed of k edges, then we should *a priori* visit all vertices at a distance at most k from a lift of the basepoint, and this number of vertices can be exponential in k .

This paper is organized as follows. We first adapt the definition of a system of loops to combinatorial surfaces in Section 2. Section 3 is devoted to the proof that we end with loops individually as short as possible among simple homotopic loops; Section 4 provides the details to obtain a practical algorithm. In Section 5, we analyze the complexity of the algorithm. Section 6 applies these results to the computation of a shortest simple loop homotopic to a given simple loop. We end with experimental results.

2. Framework of the paper

Consider a triangulated oriented surface M (or, more generally, a polyhedral 2-manifold), possibly with boundary; let $G = (V, E)$ be its vertex-edge graph. In this paper, we consider three types of paths: piecewise linear paths on M , denoted by lowercase letters (e.g., p), paths in G , written in typewriter fonts (e.g., \mathbf{P}), and loops in a combinatorial structure, the edge-ordered set of loops (EOSL for short), written in uppercase letters (e.g., P). This section aims at describing more precisely these settings. PL paths shall be used as a tool for the proof of the algorithm correctness, as it relies on topological theorems, while the EOSL is the relevant framework for systems on the 1-skeleton of a combinatorial manifold.

2.1. The PL setting

Let $G^* = (V^*, E^*)$ be the dual graph of G , naturally embedded in M (each vertex f^* of G^* is in the face f of G ; each edge e^* of G^* crosses its dual edge e and only this edge).

In this paper, a *loop* ℓ is a piecewise linear path (i.e., a continuous mapping $[0, 1] \rightarrow M$) such that $\ell(0) = \ell(1)$; this point is called the *basepoint* of ℓ . A path p (resp. a loop ℓ) is *simple* if p (resp. $\ell|_{[0,1]}$) is one-to-one. A *bouquet of circles* is an ordered set of simple loops meeting at their common basepoint, which are pairwise disjoint except at this basepoint.

An *admissible set of paths* on M is a set of piecewise linear paths on M which is in general position in the following sense:

- 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^* , and each such intersection is a crossing.

If M is boundaryless, and v_0 is a vertex in M , a (fundamental) *system of loops* $s = (s_1, \dots, s_n)$ on M , or *system*

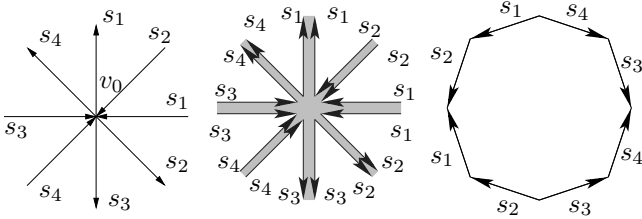


Figure 1. A system of loops (example for $g = 2$). From left to right: the loops meeting at the basepoint; the same situation after cutting along the loops; and a view of the polygonal schema after this cutting. In this example, the system is canonical.

for short, is a bouquet of circles with basepoint v_0 , such that the complement of this bouquet of circles is a topological disk. From the theory of the classification of surfaces (see [11]), it is known that any system on M is made of $n = 2g$ loops, where g is the genus of the surface. See Figure 1. In this paper, all systems of loops considered are on the manifold M , and the basepoint v_0 is fixed.

2.2. Paths on the vertex-edge graph

We choose an arbitrary orientation on the edges of G , and we denote by E^+ the set of edges of E with this orientation and by E^- the set of edges with opposite orientation. Let $E^{\text{or}} = E^+ \cup E^-$. If $e \in E^+$, let e^* be the corresponding edge of E^* , oriented such that e crosses e^* “from left to right”. If $e \in E^-$, $-e$ means the edge e with reverse orientation.

Let $p = (p_1, \dots, p_n)$ be an admissible set of paths on M . For $i = 1, \dots, n$, consider the list of edges of G^* crossed by p_i ; by duality, this yields a list of edges $e_i^0, \dots, e_i^{m_i}$ in E^{or} which is a path \mathbf{P}_i on G . The walk-edges of this set of paths are the pairs (i, j) , where $0 \leq j \leq m_i$; (i, j) corresponds to edge $e_i^j \in E^{\text{or}}$.

The same process can be done for an admissible set of loops ℓ on M ; in this case, the basepoint of path \mathbf{L}_i (corresponding to ℓ_i) is the source of e_i^0 ; the predecessor of a walk-edge (i, j) is $(i, j - 1)$ if $j \neq 0$ and (i, m_i) otherwise.

We assume that each (undirected) edge of G has a positive length, or *weight*; the length of a path (or loop) p is the length of \mathbf{P} in the weighted graph G . In this paper, we do never consider the length of a path on the manifold itself.

2.3. Edge-Ordered Set of Loops

An *edge-ordered set of loops* (EOSL for short) is a set L of loops in G , with the data, for each edge $e \in E^+$, of a linear order \preceq_e on the set \mathcal{W}_e of the walk-edges of L that

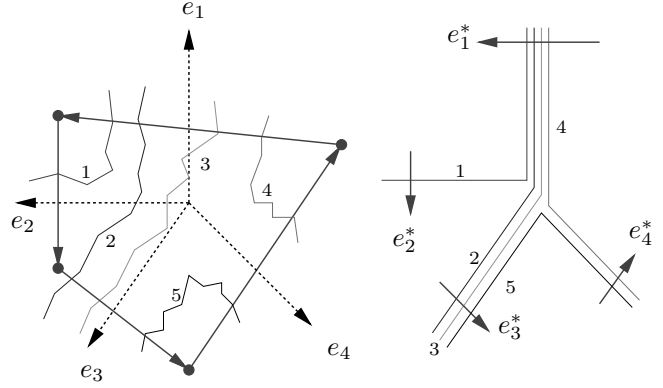


Figure 2. A representation of an edge-ordered set of loops in the star of a vertex of G , whose incident edges are e_1, \dots, e_4 . On the left, the intersections between the loops and the dual edges e_1^*, \dots, e_4^* are represented. This yields the EOSL represented on the right: each walk-edge corresponds to a crossing of a loop with an edge of G^* . In this example, no crossing occurs between the loops.

correspond to e or $-e$. If L is an EOSL, and $e \in E^-$, we define the order \preceq_e on \mathcal{W}_e by the rule: $a \preceq_e b$ if and only if $b \preceq_{-e} a$. Let ℓ be an admissible set of loops on M ; we define an EOSL $L = \rho(\ell)$ on G as follows (Figure 2): we consider the corresponding set of loops \mathbf{L} in G ; if $e \in E^+$, the elements of \mathcal{W}_e correspond to the intersection points of ℓ with e^* , and this set is linearly ordered by the orientation of e^* .

Let $v \in V$, and e_1, \dots, e_n be the CCW-ordered list of edges in E^{or} whose source is v . We define a cyclic order \preceq_v on the walk-edges meeting at v by enumerating its elements in this order: first the walk-edges in \mathcal{W}_{e_1} in \preceq_{e_1} -order; then the walk-edges in \mathcal{W}_{e_2} in \preceq_{e_2} -order; and so on. Consider two walk-edges w_1 and w_2 of L with common source $v \in V$; we say that w_1 and w_2 *cross* if, in the cyclic order \preceq_v , w_1 and its predecessor separate w_2 from its predecessor.

A *combinatorial bouquet of circles* is an edge-ordered set of loops with common basepoint, such that no crossing occurs except between walk-edges of the form $(i_1, 0)$ and $(i_2, 0)$. If ℓ is a bouquet of circles, then $\rho(\ell)$ is a combinatorial bouquet of circles.

If L is a combinatorial bouquet of circles, there exists a bouquet of circles ℓ such that $\rho(\ell) = L$: pictorially, this means that we consider all walk-edges along a given edge of E^+ and spread them by a thin space. A *combinatorial system of loops* is a combinatorial bouquet of circles S such that a bouquet of circles in $\rho^{-1}(S)$ is a system (clearly, this property does not depend on the particular bouquet considered in $\rho^{-1}(S)$).

3. Our theoretical result

We consider our polyhedral 2-manifold \mathcal{M} , whose vertex-edge graph is denoted by \mathcal{G} .

Let s be a system on \mathcal{M} with basepoint v_0 , and $k \in [1, 2g]$; we define $f_k(s)$ to be the system s where s_k has been removed and replaced by a loop s'_k , such that the resulting set of loops is a system homotopic to s and s'_k has minimal length with this property. Similarly, if S is a combinatorial system, we define $F_k(S)$ to be a shortest homotopic system resulting from S by the replacement of S_k by another loop S'_k . Throughout this paper, we refer to this shortening process as an Elementary Step.

Let $f = f_{2g} \circ \dots \circ f_1$ and $F = F_{2g} \circ \dots \circ F_1$. We call an application of f (or F) a Main Step.

Theorem 1 *Let s^0 be a system with basepoint v_0 , and $s^{n+1} = f(s^n)$. 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 which are individually as short as possible among all simple loops with basepoint v_0 in their homotopy class. In particular, s^m is an optimal system¹.*

Of course, this theorem can be rephrased in purely combinatorial terms, replacing s by S and f by F , and both versions are trivially equivalent. The goal of this section is devoted to the proof of Theorem 1; the remaining sections make use of this theorem.

3.1. Topological preliminaries

In this subsection, we briefly recall the vocabulary of homotopies and universal coverings. See [5] or [11] for more details.

A *homotopy* between two loops ℓ_0 and ℓ_1 with basepoint v_0 on a surface M is a continuous map $h : [0, 1] \times [0, 1] \rightarrow M$ such that $h(0, \cdot) = \ell_0$, $h(1, \cdot) = \ell_1$, and $h(\cdot, 0) = h(\cdot, 1) = v_0$. If such a map exists, we say that ℓ_0 and ℓ_1 are *homotopic*: in an intuitive language, there is a continuous deformation from one to the other. A loop is *null-homotopic* if it is homotopic to a constant loop.

The universal covering of M is a simply connected surface, \tilde{M} (i.e., each loop is null-homotopic) together with a continuous *projection* π from \tilde{M} onto M satisfying: each point x of M has an open, arcwise connected neighborhood U so that $\pi^{-1}(U)$ is a union of disjoint open sets $(U_i)_{i \in I}$ and $\pi|_{U_i} : U_i \rightarrow U$ is a homeomorphism. A *translation* τ in \tilde{M} is a projection preserving homeomorphism: $\pi \circ \tau = \pi$. The main properties of \tilde{M} used in this paper are:

- the *lift property*: let p be a path in M with source point y ; let $x \in \tilde{M}$ be such that $\pi(x) = y$. Then there is a unique path \tilde{p} in \tilde{M} , starting at x , such that $\pi(\tilde{p}) = p$; \tilde{p} is a *lift* of 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 two 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 a lift of p self-intersects, or two lifts of p intersect.

Let $\pi_1(M, v)$ be the set of homotopy classes of loops with basepoint v on M . If ℓ is a loop with basepoint v , let $[\ell]$ denote its homotopy class; $\ell_1.\ell_2$ denotes the concatenation of ℓ_1 and ℓ_2 . 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 ϵ .

3.2. Crossing words and reductions

We consider the universal covering $\tilde{\mathcal{M}}$ of \mathcal{M} . Fix a lift v_0^ϵ of v_0 in $\tilde{\mathcal{M}}$. For $\alpha \in \pi_1(\mathcal{M}, v_0)$, the lifts of all paths in α starting at v_0^ϵ end at the same lift of v_0 , which we call v_0^α ; this gives a one-to-one correspondence between $\pi_1(\mathcal{M}, v_0)$ and the lifts of v_0 . If ℓ is a loop with basepoint v_0 , $\alpha \in \pi_1(\mathcal{M}, v_0)$, we denote by ℓ^α the lift of ℓ starting at v_0^α .

Let A be the set of *symbols* of the form k^α or \bar{k}^α , where $k \in [1, 2g]$ and $\alpha \in \pi_1(\mathcal{M}, v_0)$. The set A^* of *words* on A is the set of finite sequences of elements in A . Fix $i \in [1, 2g]$; consider a system of loops s , and a simple loop t_i with basepoint v_0 , homotopic to s_i . We assume that this set of $2g + 1$ loops is admissible. t_i^ϵ crosses the loops s_k^α (for $k \in [1, 2g]$ and $\alpha \in \pi_1(\mathcal{M}, v_0)$) at a finite number of points. We walk along t_i^ϵ and, at each crossing encountered with a lift s_k^α of s , we write the symbol k^α or \bar{k}^α , according to the orientation of the crossing (with respect to a fixed orientation of $\tilde{\mathcal{M}}$). The resulting element of A^* is called the *crossing word* of t_i with s , and denoted by $[s/t_i]$. Note that, for this definition, the fact that two loops meet at the basepoint is not considered as a crossing.

We say that a symbol k^α or \bar{k}^α is *initial* if $\alpha \in \{\epsilon, [\bar{s}_k]\}$ (here, bar is used for reverse orientation), or, equivalently, if v_0^ϵ is one of the endpoints of s_k^α . Similarly, such a symbol is *final* if $\alpha \in \{[t_i], [t_i.\bar{s}_k]\}$; i.e., if the target of t_i^ϵ coincides with one of the endpoints of s_k^α . We define two types of *reductions* on a word in A^* :

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

¹**Remark.** The proof of Theorem 1 extends to the case where we consider the real *length* of PL systems drawn on \mathcal{M} (and not on its vertex-edge graph), provided that the suitable definition of a crossing is used: we have to take into account that two loops can partly overlap without crossing.

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 . A reduction consisting in the removal of k^α or \bar{k}^α (or both) is called a k -reduction. We define in the straightforward way the k -irreducible words and the fact that a word k -reduces to another word. We define $g(w)$ (resp. $g_k(w)$) to be the *unique* irreducible (resp. k -irreducible) word derived from w . Let ε be the empty word of A^* . The next subsection is devoted to the proof of the following key proposition:

Proposition 2 $g([s/t_i]) = \varepsilon$.

3.3. Reducibility of $[s/t_i]$

Lemma 3 *Let M be an oriented surface with boundary. We consider an admissible set of paths in M made of simple pairwise non-intersecting paths c_1, \dots, c_n , each of which separates M into two connected components, and of a loop ℓ in the interior of M . Let L be the list of crossings between ℓ and the c_k , in this order on ℓ , taking into account the orientation of the crossing (with or without a bar). Then L is a parenthesized expression.*

PROOF. Omitted. \square

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 by Epstein [2, Theorem 4.1].

Theorem 4 *Let ℓ_0 and ℓ_1 be piecewise linear, homotopic, simple loops with common basepoint in M , such that they are not null-homotopic. Then, there is a piecewise linear isotopy between ℓ_0 and ℓ_1 , keeping the basepoint fixed.*

Let D be an open disk around v_0 . We say that a loop u with basepoint v_0 is D -clean if u “enters D only once”, i.e., considering u to be a mapping from the circle into 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 5 *Let D be an open disk containing v_0 , and let u and u' be piecewise linear homotopic loops on M with basepoint v_0 , which 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 $M \setminus D$.*

PROOF. u and u' are piecewise linearly isotopic on M , with the basepoint v_0 fixed, by Theorem 4. Let $h : [0, 1] \times [0, 1] \rightarrow M$ be the isotopy: for each t , $h(t, \cdot) : [0, 1] \rightarrow M$ is one-to-one, and $h(t, 0) = h(t, 1) = v_0$; $h(0, \cdot) = u$, and $h(1, \cdot) = u'$.

$h^{-1}(D)$ 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 $r : M \setminus \{v_0\} \rightarrow M \setminus D$ be a continuous map which is the identity on $M \setminus D$ and which maps $D \setminus \{v_0\}$ onto the boundary of D . Since h is an isotopy and $h(\cdot, 0) = h(\cdot, 1) = v_0$, $h'' = r \circ h'$ is a well-defined continuous map. $h''(\cdot, \varepsilon)$ and $h''(\cdot, 1 - \varepsilon)$ are on the boundary of D ; $h''(0, \cdot)$ (resp. $h''(1, \cdot)$) 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. \square

PROOF OF PROPOSITION 2. Let s'_i be a simple PL 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 all loops s_k , t_i and s'_i are D - and D' -clean and do not cross inside D (except at v_0).

Let $\mathcal{M}' = M \setminus D'$, and let $\tilde{\mathcal{M}}'$ be its universal covering. By Lemma 5, there are two paths p and p' on the boundary of D such that any lift ℓ of $s'_i{}^{-1}.p.t_i.p'$ is a loop in $\tilde{\mathcal{M}}'$ (D' is included in D).

For any D' -clean loop u , denote by \dot{u} the part of u outside D' . It is easy to see that each lift of \dot{s}_k is a separating curve in $\tilde{\mathcal{M}}'$; applying Lemma 3, we obtain that the list of crossings of ℓ with the lifts of $\dot{s} = (\dot{s}_1, \dots, \dot{s}_{2g})$ is parenthesized.

Recall that π is the projection from $\tilde{\mathcal{M}}$ onto M . Note that $\tilde{\mathcal{M}}'$ is a covering space (in fact, the universal covering) 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 list of crossings of any lift of $\ell' = p.t_i.p'$ in $\tilde{\mathcal{M}}$ with the lifts of \dot{s} is also parenthesized. Considering the lift of ℓ' containing the part of t_i^ε which is outside $\pi^{-1}(D)$, we see that $[s/t_i]$ can be deduced from this expression by extremal reductions, which concludes the proof. \square

3.4. Uncrossing the loops

In this subsection, we fix i and j in $[1, 2g]$, and we consider a continuous system s and a simple loop t_i homotopic to s_i . Let $r = f_j(s)$. We always assume that s and t_i (resp. r and t_i) constitute an admissible set of loops. Intuitively, the strategy is to show that, if t_i is as short as possible, then applying f_j to s unties the intersections between t_i and s_j . More precisely, we prove in this subsection:

Proposition 6 *There exists a simple loop t'_i homotopic to and not longer than t_i , such that $[r/t'_i] = g_j([s/t_i])$ (and, of course, r and t'_i constitute an admissible set of paths).*

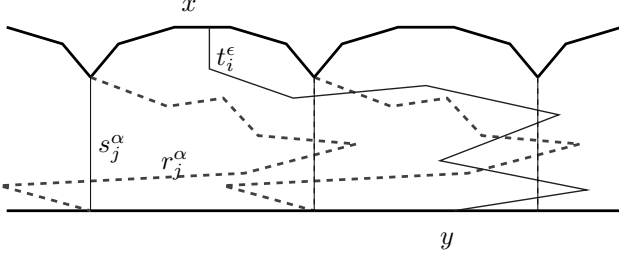


Figure 3. The crossings of the part of t_i^ϵ between x and y with lifts of s_j and r_j . The same properties hold if x and y are both on the same component of the boundary of the strip.

Lemma 7 $g_j([r/t_i]) = g_j([s/t_i])$.

PROOF. A k -symbol is a symbol of the form k^α or \bar{k}^α . Consider $[s/t_i]$ where all j -symbols are removed; this list is the same as $[r/t_i]$ where all j -symbols are removed. To prove our lemma, it is sufficient to consider, in $[s/t_i]$ and $[r/t_i]$, the lists of j -symbols between two consecutive non- j -symbols (and also before the first and after the last non- j -symbol), and to prove that they reduce to the same expression by applications of parenthesized j -reductions (and possibly initial or final j -reductions for the initial and final parts of $[s/t_i]$).

Two successive non- j -symbols in $[s/t_i]$ correspond, in $\tilde{\mathcal{M}}$, to two intersections x and y of t_i^ϵ with the lifts of s_k ($k \neq j$). Let p denote the part of t_i^ϵ between x and y , and $[s/p]$ be the list of crossings of p with the lifts of s (hence only lifts of s_j); p is on an infinite strip bounded by lifts of $s_k = r_k$ for $k \neq j$ (see Figure 3). It is easy to see that $[s/p]$, to which we apply as much as possible parenthesized j -reductions, is the list of the lifts of s_j that separate x from y . But s_j^α separates x from y if and only if r_j^α does, because they are paths with the same endpoints on the boundary of the strip; hence $[s/p]$ and $[r/p]$ are equal up to parenthesized j -reductions.

The reasoning for the initial and last parts of $[s/t_i]$ and $[r/t_i]$ is quite similar. \square

Define a *lens* of two paths (or loops) p_1 and p_2 to be two strict subpaths of p_1 and p_2 , with the same endpoints, so that these two subpaths concatenated together make a simple loop bounding a topological disk. The *corners* of this lens are the two endpoints of the subpaths.

Lemma 8 *Suppose a j -reduction is possible on $[r/t_i]$. Then, there is a lens of t_i and r_j which is crossed neither by t_i nor by any of the r_k .*

PROOF. We first claim that there is a lens $\tilde{\mathcal{L}}$, in $\tilde{\mathcal{M}}$, of t_i^ϵ and r_j^α for some α , which is crossed by no lift of r . Indeed,

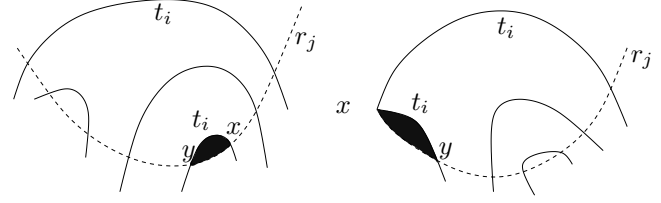


Figure 4. The two situations considered in the proof of Lemma 8: on the left, none of the corners is the basepoint; on the right, x is the basepoint.

this is true for parenthesized and extremal j -reductions (in the latter case, one corner of the lens is a lift of the basepoint).

The projection \mathcal{L} of $\tilde{\mathcal{L}}$ on \mathcal{M} crosses no path r_k . It is thus contained in the polygonal schema defined by r . By the Schoenflies Theorem, \mathcal{L} bounds a disk D and is also a lens of t_i and r_j . 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 4), we get the result. \square

PROOF OF PROPOSITION 6. We will use the following property: if p is a path with endpoints a and b , and c is a point on p not lying on any edge of \mathcal{G}^* , then the length of p is the sum of the lengths of the parts of p between a and c and between c and b .

If $[r/t_i]$ is j -irreducible, then we are over by Lemma 7. Otherwise, we apply Lemma 8; let x and y denote the corners of the lens, and t_i^{xy} and r_j^{xy} be the parts of t_i and r_j constituting the lens. By the admissibility property, x and y are on no edge of \mathcal{G}^* .

Suppose that t_i^{xy} is shorter than r_j^{xy} . Then, by replacing the part r_j^{xy} of r_j by a path going along t_i^{xy} , we get a path r'_j , homotopic to r_j , such that the replacement of r_j by r'_j in r yields a homotopic system of loops, and r'_j is shorter than r_j . This contradicts the definition of f_j .

We replace the part t_i^{xy} of t_i by a path going along r_j^{xy} , just outside the lens; call t'_i the resulting path. $[r/t'_i]$ is deduced from $[r/t_i]$ by a j -reduction. Note that r and t'_i form an admissible set of loops, and t'_i is not longer than t_i . We end by induction. \square

3.5. Conclusion of the proof

Lemma 9 *Let u be an admissible system, and v_i be a simple loop homotopic to u_i , with minimal length, so that u and v_i constitute an admissible set of loops. If $[u/v_i] = \varepsilon$, then the i th loop of $f(u)$ has the same length as v_i .*

PROOF. Omitted (use Proposition 6). \square

PROOF OF THEOREM 1. Fix $i \in [1, 2g]$. Let t_i^0 be a shortest simple loop homotopic to s_i^0 , such that t_i^0 and s^0 constitute an admissible set of loops.

By Propositions 6 and 2, one can construct a sequence $(t_i^n)_{n \in \mathbb{N}}$ of loops, having the same length as t_i^0 , homotopic to s_i^0 , such that the length of the expression $[s^n/t_i^n]$ strictly decreases with n until it is the empty word. Then for some n , $[s^n/t_i^n] = \varepsilon$, and, by Lemma 9, s_i^{n+1} has the same length as t_i^0 . This proves that the length of s^n becomes stationary at some stage of the algorithm.

Let m be the first integer such that at s^m and s^{m+1} have the same length. One can prove: for any $j \in [1, 2g]$, there exists t'_j , of minimal length, such that $[s^m/t'_j] = g_j([s^m/t_i^m])$ (using arguments as in Subsection 3.4, and induction). Hence, iterating this process, we get that there exists t'_j , of minimal length, such that $[s^m/t'_j] = \varepsilon$. Then, by Lemma 9, s_i^{m+1} has the same length as t_i^0 . This concludes the proof. \square

4. Shortest paths on cylinders

We will develop the tools used to process an Elementary Step. Note that $F_i(S)$ consists in the replacement of S_i by S'_i such that this loop:

1. crosses no S_j for $j \neq i$;
2. is homotopic to S_i ;
3. is as short as possible;
4. is simple.

The key idea for this section is that, if we are able to compute (in a particular way) a loop satisfying the first three hypotheses, then the last one will be automatically established. Finally, we will reduce the problem to that of finding a shortest path in some graph. Recall from Section 2 the notations p , \mathbf{P} , and S_i ; in this section, we will use the applications $p \mapsto \mathbf{P}$ and $S_i \mapsto \mathbf{S}_i$ implicitly.

4.1. The cylinder $\mathcal{M}(s, i)$

Let $s = (s_1, \dots, s_{2g})$ be an admissible system of loops on \mathcal{M} . We define $\mathcal{M}(s, i)$ to be the bounded cylinder obtained after the cutting of \mathcal{M} along the loops s_1, \dots, s_{2g} , except s_i . The extremities of s_i are no more identified on $\mathcal{M}(s, i)$. We denote by ϕ the quotient map from $\mathcal{M}(s, i)$ onto \mathcal{M} . Note that the i th loop of $f_i(s)$ satisfies Condition 1, hence is in $\mathcal{M}(s, i)$. Analogously, the i th loop of $F_i(S)$ is to be searched for in some graph embedded on $\mathcal{M}(s, i)$, which we now describe.

We first explain how \mathcal{G}^* is transformed after the cutting of \mathcal{M} into $\mathcal{M}(s, i)$. Intuitively, we refine \mathcal{G}^* on \mathcal{M} by adding vertices at the intersection points between its edges

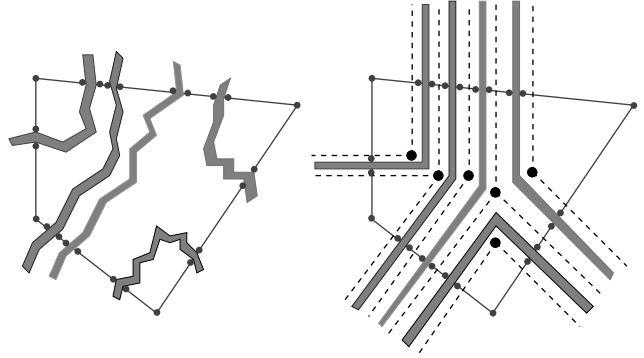


Figure 5. The situation on the part of the surface shown in Figure 2. The grey area represents the cutting of the manifold along all loops. We see that \mathcal{G}^* is cut into pieces, but the result is still a graph $\mathcal{G}^*(s, i)$. The right side illustrates the algorithmic construction of $\mathcal{G}(s, i)$, represented in dashed lines.

and the paths s_j (for $j \neq i$); after that, by cutting along the loops s_j , we are able to build a new graph $\mathcal{G}^*(s, i)$, embedded on $\mathcal{M}(s, i)$, such that $\phi(\mathcal{G}^*(s, i)) = \mathcal{G}^*$ (see Figure 5). More formally, the set of vertices of $\mathcal{G}^*(s, i)$ is made of the preimage, by ϕ , of the union of two sets: the set of vertices of \mathcal{G}^* , and the intersection points between the edges of \mathcal{G}^* and the paths s_j ($j \neq i$). Two vertices of $\mathcal{G}^*(s, i)$ are adjacent if and only if they can be linked, in $\mathcal{M}(s, i)$, by a path whose image by ϕ is included in an edge of \mathcal{G}^* .

Each edge of $\mathcal{G}^*(s, i)$ maps naturally to an edge of \mathcal{G}^* . Consider now the dual graph $\mathcal{G}(s, i)$ of $\mathcal{G}^*(s, i)$ in $\mathcal{M}(s, i)$. By duality, each edge of $\mathcal{G}(s, i)$ also maps to an edge of \mathcal{G} ; by abuse of notation, we still denote this map by ϕ . $\mathcal{G}(s, i)$ can be embedded in $\mathcal{M}(s, i)$ (Figure 5, right), and the edges of $\mathcal{G}(s, i)$ in a given set $\phi^{-1}(e)$ are “parallel”.

By definition, the weight of an edge e of $\mathcal{G}(s, i)$ is the weight of $\phi(e)$. The fundamental property of $\mathcal{G}(s, i)$ is the following: any admissible path in $\mathcal{M}(s, i)$ “retracts” to a combinatorial path in $\mathcal{G}(s, i)$ of the same length, and the converse is true. Hence, computing $F_i(S)$ reduces to compute a path in $\mathcal{G}(s, i)$ satisfying only Conditions 2, 3, and 4.

Let $\tilde{\mathcal{M}}(s, i)$ be the universal covering of $\mathcal{M}(s, i)$, and let π be its projection. $\tilde{\mathcal{M}}(s, i)$ can be viewed as the surface with boundary resulting from stitching along s_i infinitely many copies of the polygonal schema associated to s (Figure 6). We now define a graph $\tilde{\mathcal{G}}(s, i)$ embedded on $\tilde{\mathcal{M}}(s, i)$ by: $\tilde{\mathcal{G}}(s, i) = \pi^{-1}(\mathcal{G}(s, i))$. In other words, $\tilde{\mathcal{G}}(s, i)$ is the natural covering of $\mathcal{G}(s, i)$ in $\tilde{\mathcal{M}}(s, i)$.

The algorithmic construction of all these graphs is easy and skipped in this abstract.

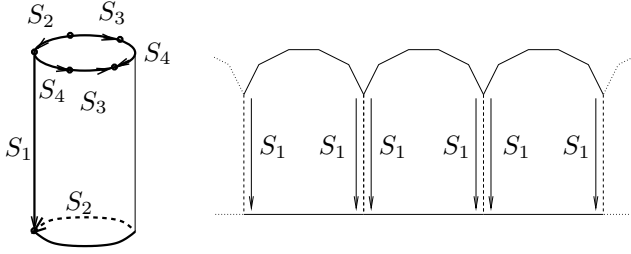


Figure 6. On the left, $\mathcal{M}(s, 1)$ (case $g = 2$), which is a topological cylinder. On the right, the manifold $\tilde{\mathcal{M}}(s, 1)$.

4.2. Simplicity of a shortest path

To achieve Condition 4, a shortest homotopic path in $\mathcal{G}(s, i)$ must be chosen in a particular way to “break the ties”. Specifically, we need an algorithm which, given two vertices a and b of $\tilde{\mathcal{G}}(s, i)$, computes a shortest path $\text{SP}(a, b)$ such that:

- if $c \in \text{SP}(a, b)$, then $\text{SP}(a, b)$ is the concatenation of $\text{SP}(a, c)$ and $\text{SP}(c, b)$;
- SP is invariant by any translation τ in the universal covering $\tilde{\mathcal{G}}(s, i)$: $\text{SP}(\tau(a), \tau(b)) = \tau(\text{SP}(a, b))$.

Note that $\text{SP}(a, b)$ and $\text{SP}(b, a)$ may differ. It can be checked that a slight variant of Dijkstra’s algorithm can be used for that problem: before everything, we choose a linear ordering on the oriented edges of $\mathcal{G}(s, i)$. During the relaxation step of Dijkstra’s algorithm (see [1]), if equality occurs between the stored distance and the new computed distance, we have two shortest paths arriving to a vertex, whose last edges differ: we simply select the path whose last edge (projected in $\mathcal{G}(s, i)$) is minimal.

Consider the path s_i viewed in $\mathcal{M}(s, i)$, and let a and b be two vertices of $\tilde{\mathcal{G}}(s, i)$ which are the endpoints of a lift of s_i in $\tilde{\mathcal{M}}(s, i)$. Let $\tilde{\mathbf{P}} = \text{SP}(a, b)$, and $\mathbf{P} = \pi(\tilde{\mathbf{P}})$. The following proposition says that \mathbf{P} is nearly the i th loop of $F_i(S)$:

Proposition 10 *Let \hat{S} be the EOSL resulting from S by the removal of S_i . It is possible to insert a loop S'_i in \hat{S} such that $\mathbf{S}'_i = \phi(\mathbf{P})$, and the resulting EOSL S' is a system of loops homotopic to S .*

Note that p is a path which crosses no s_j for $j \neq i$, is homotopic to s_i , and as short as possible. Hence the i th loop of $F_i(S)$ cannot be shorter than \mathbf{P} . By the above proposition, we can let $F_i(S) = S'$.

SKETCH OF PROOF. We prove the existence a simple admissible path p in $\mathcal{M}(s, i)$ whose corresponding path in

$\mathcal{G}(s, i)$ is \mathbf{P} . This implies that there exists an admissible system of loops s' , resulting from s by the replacement of s_i by a loop $s'_i = \phi(p)$, such that the length of s'_i equals the length of \mathbf{P} . Considering $\rho(s')$ yields the result.

The idea of the proof is to start with some simple admissible path \tilde{p}_0 corresponding to a lift $\tilde{\mathbf{P}}_0$ of \mathbf{P} . If $p = \pi(\tilde{p}_0)$ self-intersects then, by the intersection property, \tilde{p}_0 must intersect some other lift \tilde{p}_1 of p . By the Jordan Curve Theorem in $\tilde{\mathcal{M}}(s, i)$, these two lifts must intersect at least twice. We then consider two subpaths of \tilde{p}_0 and \tilde{p}_1 sharing their extremities and oriented the same way. By the properties of our shortest-path algorithm, these two subpaths are shortest paths, and the parts of $\tilde{\mathbf{P}}_0$ and $\tilde{\mathbf{P}}_1$ corresponding to these subpaths are identical. We can thus swap their projections in p without changing its corresponding path \mathbf{P} . This strictly reduces the number of self-intersections of p . The proof is ended by induction. \square

Actually, it can be checked that there is only one way to insert S'_i in \hat{S} . Moreover, this insertion can be done in time linear in the complexity of S'_i .

5. Complexity analysis

We now give a (crude?) upper bound on the complexity of our algorithm. Let n be the complexity of \mathcal{M} and g be its genus. Let S be a combinatorial system homotopic to some given S^0 on \mathcal{M} ; let s and s^0 be respective associated continuous systems.

Lemma 11 *Suppose there exists a shortest simple loop t_i homotopic to s_i such that both s_i^0 and $f_i(s)_i$ cross t_i at most C_i times. Then, computing $F_i(S)$ is possible in time $O(p \log p)$, where $p = (n + |S| + |S^0|)C_i$. ($|S|$ is the number of edges of system S .)*

PROOF. Consider a part \bar{t}_i of t_i (resp. a part \bar{s}_i^0 of s_i^0) which goes from one boundary of $\mathcal{M}(s, i)$ to the other one. A lift of $f_i(s)_i$ in $\tilde{\mathcal{M}}(s, i)$ crosses at most $2C_i + 1$ lifts of \bar{s}_i^0 , hence $f_i(s)_i$ is confined to the space made of $2C_i + 2$ patches delimited by two consecutive lifts of \bar{s}_i^0 in $\tilde{\mathcal{M}}(s, i)$. This corresponds to a search in a graph of complexity $O(p)$, since the graph $\mathcal{G}(s, i)$ has size $O(n + |S|)$. In practice, it is not required to know C_i in advance: searching through an exponentially increasing number of patches leads to the announced complexity. \square

Let μ be the multiplicity of any loop in S^0 , i.e., the maximal multiplicity of any vertex of \mathcal{M} in a loop of S^0 . In particular the complexity of a loop is bounded by μn . The next lemma follows easily:

Lemma 12 *Consider an EOSL made of S^0 and of a loop T ; let $k \in [1, 2g]$. The number of crossings between S_k^0 and T is bounded by $\mu|T|$.*

Let L be the maximum over i of the maximal number of edges of any shortest simple loop T_i homotopic to S_i^0 . By the preceding lemma, there exists a shortest simple loop t_i homotopic to s_i^0 such that s^0 and t_i cross $O(g\mu L)$ times; this is also an upper bound on the number of Main Steps required, by Propositions 2 and 6. Whence the total number of Elementary Steps is $O(g^2\mu L)$.

Note that, during the algorithm, the values of C_i (Lemma 11) are bounded from above by the maximal number of crossings between S_i^0 and any shortest simple loop T_i homotopic to S_i^0 , hence by μL . Let m be the maximum of $(n + |S^j| + |S^0|)$ over the shortening iterations. From Lemma 11, an Elementary Step can be processed in time $O(\mu L m \log(\mu L m))$. The total time spent by the algorithm is thus $O(g^2\mu^2 L^2 m \log(\mu L m))$. We further introduce the longest-to-shortest edge ratio, α , in \mathcal{M} . Since an Elementary Step cannot increase the length of a loop, we have $L = O(\alpha\mu n)$ and $m = O(g\alpha\mu n)$. Putting all this together, we get:

Theorem 13 *Given a system of loops with multiplicity μ on an orientable triangulated surface \mathcal{M} , with longest-to-shortest edge ratio α , there is an algorithm that computes an optimal homotopic system in $O(\mu^5\alpha^3g^3n^3 \log(\mu\alpha n))$ time.*

Remark. According to [6], the logarithmic term in the theorem can be removed.

The definition of μ can be slightly modified to imply a weaker condition so that the above complexity still applies (details are omitted). Moreover, with this new definition, it can be proved that $\mu = 2$ holds true for any system computed as in [8].

6. Shortest simple loop

In this section, we study the problem of computing a shortest simple loop homotopic to a given simple loop. While an exhaustive search in the universal cover of \mathcal{M} again leads to an exponential algorithm, a simple application of the previous results gives us a polynomial algorithm. Let T be an EOSL on \mathcal{G} made of a single simple loop with multiplicity μ .

Theorem 14 *Among all EOSLs made of a simple loop homotopic to T , a shortest one can be computed in $O(\mu^5\alpha^3g^3n^3 \log(\mu\alpha n))$ time (the parameters α , g , and n are defined as in Theorem 13).*

PROOF. First suppose that T does not separate \mathcal{M} . Using [8, Part 5, Step 1], it is possible to extend T to a combinatorial system of loops containing T , so that the multiplicity of each loop is at most 2μ . This is easy if $\mu = 1$: indeed, in this case, \mathbf{T} is a simple loop on \mathcal{M} , and it is

possible to ensure that \mathbf{T} is a part of the graph G described in that paper. In the general case, we have to refine \mathcal{M} to simulate that $\mu = 1$, compute the system, and then reconstruct the surface. In all cases, it remains to apply Theorem 13 to compute an optimal system containing a simple loop homotopic to T ; by Theorem 1, this is the desired loop.

Suppose on the contrary that T separates \mathcal{M} . Our intermediate goal is to compute a *simple* EOSL of \mathcal{M} made of a system for \mathcal{M} and of T , so that the multiplicity of each loop is at most 2μ . Again, this is easy to do if $\mu = 1$: in this case, we consider the manifolds \mathcal{M}_1 and \mathcal{M}_2 which result from the cutting of \mathcal{M} along T ; we close \mathcal{M}_1 and \mathcal{M}_2 with a single face, and we compute systems for each of these two surfaces, with the basepoint corresponding to the basepoint of T . The “union” of T with both systems in \mathcal{M} yields the result. If $\mu > 1$, like above, it is possible to conclude using a refinement of the manifold.

Applying Theorem 13, we compute an optimal system S' homotopic to S . Then, we insert in S' a shortest loop T' among all loops homotopic to T and which do 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 simple loop associated to T' , and let t'' be a shortest simple loop homotopic to t' ; as in the proof of Proposition 2, $[s'/t'']$ reduces to the empty word; as in the proof of Lemma 8, there exists a simple loop homotopic to t'' and which does not cross s' ; hence t' and t'' have the same length. \square

7. Implementation

We have implemented a slightly different version of the algorithm using the C++ based CGAL library (in this version, we do not need to compute partial covering spaces of $\mathcal{M}(s, i)$ as suggested in Section 4.1).

In order to make comparisons, 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 7 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>.

8. Discussion

Natural extensions of our work concern the case of non-orientable and/or bordered surfaces. Another interesting

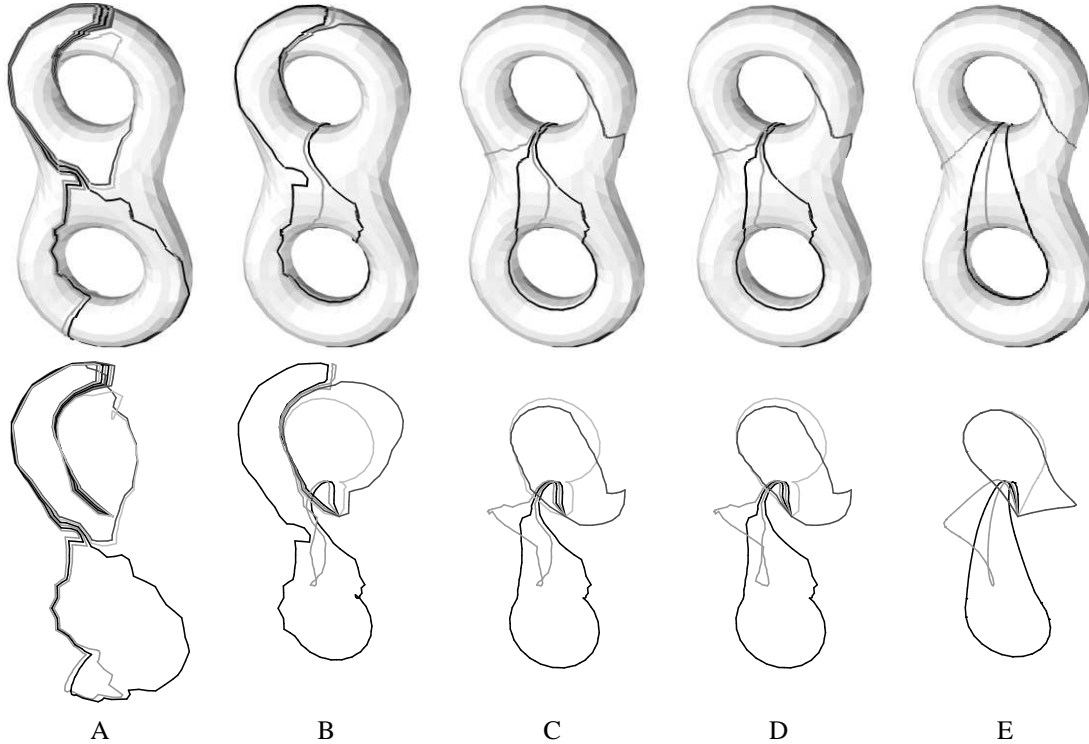


Figure 7. A: A canonical system, S , obtained after [8]. The basepoint is on the back side of the double torus. B: $F(S)$. C: $F^2(S)$. D: $F^3(S) = F^4(S)$. E: The local optimization was applied to this optimal system to get a geodesic system on the surface (4,000 star optimizations were performed).

continuation 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 work of Hershberger and Snoeyink [7] to much more general surfaces.

Our work also suggests some open questions. What is the influence of the basepoint position? How would it be possible to get the shortest system, among all systems, relaxing the homotopy condition? Comparing with the work of Erickson and Har-Peled [3], we expect this last problem to be much less tractable than those solved in the present paper.

Acknowledgements

We wish to thank Michel Pocchiola for interesting discussions and comments.

References

- [1] T. Cormen, C. Leiserson, R. Rivest, and C. Stein. *Introduction to algorithms*. MIT Press, Cambridge, MA, 2001.
- [2] D. Epstein. Curves on 2-manifolds and isotopies. *Acta Mathematica*, 115:83–107, 1966.
- [3] J. Erickson and S. Har-Peled. Optimally cutting a surface into a disk. In *Proc. of the 18th Annual Symposium on Computational Geometry*, pages 244–253, 2002.
- [4] M. Floater. Parametrization and smooth approximation of surface triangulations. *Computer Aided Geometric Design*, 14(3):231–250, 1997.
- [5] M. Greenberg. *Lectures on Algebraic Topology*. Benjamin, Reading, MA, 1967.
- [6] M. Henzinger, P. Klein, S. Rao, and S. Subramanian. Faster shortest-path algorithms for planar graphs. *J. of Computer and System Sciences*, 55(1, part 1):3–23, 1997.
- [7] J. Hershberger and J. Snoeyink. Computing minimum length paths of a given homotopy class. *Comput. Geom. Theory Appl.*, 4:63–98, 1994.
- [8] F. Lazarus, M. Pocchiola, G. Vegter, and A. 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.
- [9] J. Maillot, H. Yahia, and A. Verroust. Interactive texture mapping. In *SIGGRAPH 93*, pages 27–34, 1993.
- [10] D. Piponi and G. Borshukov. Seamless texture mapping of subdivision surfaces by model pelting and texture blending. In *SIGGRAPH 2000*, pages 471–478, 2000.
- [11] J. Stillwell. *Classical topology and combinatorial group theory*. Springer-Verlag, New York, 1993.