Splitting (Complicated) Surfaces Is Hard*

Erin W. Chambers†
Dept. of Computer Science
University of Illinois
erinwolf@uiuc.edu

Éric Colin de Verdière
CNRS, Laboratoire d'informatique
École normale supérieure, Paris, France
Eric.Colin.de.Verdiere@ens.fr

Jeff Erickson‡
Dept. of Computer Science
University of Illinois
jeffe@cs.uiuc.edu

Francis Lazarus
CNRS, Laboratoire des Images et des Signaux
Grenoble, France
Francis.Lazarus@lis.inpg.fr

Kim Whittlesey
Department of Mathematics
University of Illinois
kwhittle@math.uiuc.edu

ABSTRACT

Let $\mathcal{M}$ be an orientable combinatorial surface without boundary. A cycle on $\mathcal{M}$ is *splitting* if it has no self-intersections and it partitions $\mathcal{M}$ into two components, neither of which is homeomorphic to a disk. In other words, splitting cycles are simple, separating, and non-contractible. We prove that finding the shortest splitting cycle on a combinatorial surface is NP-hard but fixed-parameter tractable with respect to the surface genus. Specifically, we describe an algorithm to compute the shortest splitting cycle in $g^{O(g)} n \log n$ time.

Categories and Subject Descriptors: F.2.2 [Analysis of Algorithms and Problem Complexity]: Nonnumerical algorithms and problems—*Geometric problems and computations*; I.3.5 [Computer Graphics]: Computational geometry and object modeling—*Geometric algorithms, languages, and systems*; G.2.2 [Discrete Mathematics]: Graph theory—*Graph algorithms; path and circuit problems*

Keywords: computational topology, topological graph theory, combinatorial surface, splitting cycle

General Terms: Algorithms, Performance

*See http://www.cs.uiuc.edu/~jeffe/pubs/splitting.html for the most recent version of this paper.

†Research partially supported by an NSF Graduate Research Fellowship and by NSF grant DMS-0528086.

‡Research partially supported by NSF grants CCR-0093348, CCR-0219594, and DMS-0528086.

SCG'06, June 5–7, 2006, Sedona, Arizona, USA.

1. INTRODUCTION

Optimization problems on surfaces in the fields of computational topology and topological graph theory have received much attention in the past few years. Erickson and Har-Peled [12] prove that computing the shortest graph whose removal cuts a surface into a topological disk is NP-hard. Colin de Verdière and Lazarus consider the problem of finding the shortest *simple* loop [8] or cycle [7] within a given homotopy class. A polynomial time algorithm for the generalization of this problem to non-simple curves was recently obtained by Colin de Verdière and Erickson [6]. Erickson and Whittlesey [13] provide simple polynomial-time algorithms to compute the shortest homology basis and the shortest fundamental system of loops on a given surface.

Several authors have considered the problem of a computing shortest cycle with some prescribed topological property, such as non-contractibility. When the set of cycles with the desired property satisfies the so-called *3-path condition*, a generic algorithm of Mohar and Thomassen finds a shortest such cycle in $O(n^3)$ time [21, Sect. 4.3]. For example, the sets of non-separating and non-contractible cycles on any surface satisfy the 3-path condition, but the set of separating cycles on a surface does not. Erickson and Har-Peled describe a faster algorithm to compute non-separating and non-contractible cycles in $O(n^2 \log n)$ time [12]. Cabello and Mohar [4], Cabello [3], and Kutz [18] develop faster algorithms when the surface has small genus; the fastest of these runs in $g^{O(g)} n \log n$ time [18].

In this paper, we consider the following natural optimization problem: Given a orientable 2-manifold $\mathcal{M}$ with genus $g \geq 2$ and without boundary, compute a shortest simple cycle that separates $\mathcal{M}$ into two topologically non-trivial components. For simplicity, we will call a simple non-contractible separating cycle a *splitting* cycle. The set of splitting cycles does not satisfy the 3-path condition, so a different approach is required to compute the shortest one.

After reviewing a few necessary concepts from topology and proving some preliminary results, we prove in Section 3 that computing the shortest splitting cycle on a given surface is NP-hard. In Section 4, we then prove that a splitting cycle on a surface of genus g cuts any shortest path on the surface $O(g)$ times, and that this bound is tight in the worst case. This property leads to an algorithm to compute a shortest splitting cycle in $g^{O(g)} n \log n$ time, which we describe in

Section 5. Thus, we show that the shortest splitting problem is fixed-parameter tractable with respect to the genus of the surface. This is the first result of this kind among the previously cited works. In particular, although Erickson and Har-Peled provide an algorithm to compute the minimum cut graph on any surface of constant genus in polynomial time, the order of the polynomial depends on the genus [12].

2. PRELIMINARIES

2.1 Topological Background

We recall several notions from combinatorial and computational topology. See also Hatcher [16], Stillwell [22], or Zomorodian [24] for more details.

Curves on surfaces.

A *surface* (or 2-manifold with boundary) $\mathcal{M}$ is a topological Hausdorff space where each point has a neighborhood homeomorphic either to the plane or to the closed half-plane. The points with neighborhood homeomorphic to the closed half-plane comprise the *boundary* of $\mathcal{M}$. All the surfaces considered in this paper are *compact*, *connected*, and *orientable*. Such a surface is homeomorphic to a sphere with g handles attached and b open disks removed, for some unique non-negative integers g and b called respectively the *genus* and the *number of boundaries* of $\mathcal{M}$. In this paper, all surfaces are without boundary, unless specifically stated otherwise.

Let $\mathcal{M}$ be a surface. A *path* on $\mathcal{M}$ is a continuous map $p\colon [0,1] \to \mathcal{M}$. The *endpoints* of p are $p(0)$ and $p(1)$. A *loop* with *basepoint* x is a path with both endpoints equal to x. A *cycle* is a continuous map $\gamma\colon S^1 \to \mathcal{M}$, where S^1 denotes the unit circle. A path, loop, or cycle is *simple* if it is one-to-one, except, of course, for the endpoints of a loop. Two cycles are *disjoint* if they do not intersect. Two paths are *disjoint* if they do not intersect, except possibly at their endpoints; in particular, two disjoint loops can intersect only at their common basepoint.

If p is a path, its *reversal* is the path $\bar{p}(t) = p(1-t)$. The *concatenation* $p \cdot q$ of two paths p and q with $p(1) = q(0)$ is defined by setting $(p \cdot q)(t) = p(2t)$ for all $t \leq 1/2$, and $(p \cdot q)(t) = q(2t-1)$ for all $t \geq 1/2$.

Systems of loops and homotopy.

If L is a set of pairwise disjoint simple loops, $\mathcal{M} \setminus L$ denotes the surface with boundary obtained by *cutting* $\mathcal{M}$ along the loops in L. A *system of loops* on $\mathcal{M}$ is a set of pairwise disjoint simple loops L with a common basepoint such that $\mathcal{M} \setminus L$ is a topological disk. Any system of loops contains exactly $2g$ loops. $\mathcal{M} \setminus L$ is a $4g$-gon where each loop appears as two boundary edges; this $4g$-gon is called the *polygonal schema* associated with L.

A *homotopy* between two paths p and q is a continuous map $h\colon [0,1] \times [0,1] \to \mathcal{M}$ such that $h(0,\cdot) = p$, $h(1,\cdot) = q$, $h(\cdot,0) = p(0) = q(0)$, and $h(\cdot,1) = p(1) = q(1)$. A (free) *homotopy* betweeen two cycles γ and δ is a continuous map $h\colon [0,1] \times S^1$ such that $h(0,\cdot) = \gamma$ and $h(1,\cdot) = \delta$. Two paths or cycles are *homotopic* if there is a homotopy between them. For any two loops ℓ and ℓ' with the same basepoint, their concatenations $\ell \cdot \ell'$ and $\ell' \cdot \ell$ are freely homotopic as cycles but not necessarily homotopic as loops. A loop or cycle is *contractible* if it is homotopic to a constant loop or cycle.

Two cycles are *homologous* (with $\mathbb{Z}_2$ coefficients) if one can be continuously deformed into the other via a deformation that may include splitting cycles at self-intersection points, merging intersecting pairs of cycles, or adding or deleting contractible cycles. Thus, if two cycles are homotopic, they are also homologous, but the converse is not necessarily true. A cycle or loop is *null-homologous* if it is homologous to a constant loop. A simple cycle γ is null-homologous if and only if it is *separating*, that is, if $\mathcal{M} \setminus \gamma$ has two components. Every contractible simple cycle is separating (in fact, it bounds a disk), but not all simple cycles are contractible.

We say that a cycle *splits* a surface $\mathcal{M}$ if it is simple, non-contractible, and separating.

Combinatorial and cross-metric surfaces.

Like most earlier works [4, 7, 8, 12, 13, 19], we state and prove our algorithmic results in the *combinatorial surface* model. A combinatorial surface is an abstract surface $\mathcal{M}$ together with a weighted undirected graph $G(\mathcal{M})$, embedded on $\mathcal{M}$ so that each open face is a disk. In this model, the only allowed paths are walks in G; the length of a path is the sum of the weights of the edges traversed by the path, counted with multiplicity. Paths on a combinatorial surface may visit a vertex or run along an edge several times without crossing. The *complexity* of a combinatorial surface is the total number of vertices, edges, and faces of G.

It is often more convenient to work in an equivalent dual formulation of this model introduced by Colin de Verdière and Erickson [6]. A *cross-metric surface* is also an abstract surface $\mathcal{M}$ together with an undirected weighted graph $G^* = G^*(\mathcal{M})$, embedded so that every open face is a disk. However, now we consider only *regular* paths and cycles on $\mathcal{M}$, which intersect the edges of G^* only transversely and away from the vertices. The *length* of a regular curve p is defined to be the sum of the weights of the dual edges that p *crosses*, counted with multiplicity. See [6] for further discussion of these two models.

We emphasize that most of our structural results (Lemmas 2.1 and 2.2, Proposition 4.2, and Theorem 4.3) are *not* limited to the combinatorial surface model, but rather apply to surfaces with a wide range of metrics, including piecewise-linear, piecewise-algebraic, and abstract Riemann surfaces. Even the NP-hardness reduction in Section 3 can be easily modified to apply in these more general surface models. The restriction to the combinatorial surface model is only necessary for our algorithmic results in Section 5, largely because this is the only known surface model in which exact shortest paths can be computed efficiently without additional assumptions.[1]

[1] Efficient shortest-path algorithms for piecewise-linear surfaces [5, 20] require exact real arithmetic. Even if the input coordinates are integers, shortest path lengths are sums of square roots of integers; it is an open question whether two such sums can be compared in polynomial time on an integer RAM [2]. The analysis of these algorithms also assumes that the shortest path between any two points in the same face of the surface is contained in that face. Although this condition is satisfied by non-convex polyhedra in any Euclidean space, it is not true for arbitrary abstract PL surfaces.

2.2 Preliminary Lemmas

A set of loops, all with the same basepoint, are *independent* if they are simple, non-contractible, pairwise non-homotopic, and pairwise disjoint. An independent set L of loops is *strongly independent* if no loop in L is homotopic to the reverse of another loop in L.

Lemma 2.1. *Let $\mathcal{M}$ be an orientable surface of genus $g \geq 1$. Any independent set of loops in $\mathcal{M}$ contains at most $12g-6$ loops.*

Proof: Let L be a strongly independent set of loops with basepoint x; it suffices to show that $|L| \leq 6g-3$. Each component of $\mathcal{M} \setminus L$ has at least one boundary cycle, and each boundary cycle contains at least one copy of the basepoint x. If any component $\mathcal{M}'$ is a disk, that disk must have at least three copies of the basepoint x on its boundary, because no loop in L is contractible or homotopic to another loop in L or its inverse. We will add loops to the set L, maintaining its strong independence, until we reach a maximal strongly independent set.

First, if some component $\mathcal{M}'$ has non-zero genus, let ℓ be a simple non-separating loop in $\mathcal{M}'$ based at (some copy of) x. $\mathcal{M}' \setminus \ell$ has smaller genus than $\mathcal{M}'$, and $L \cup \{\ell\}$ is still a strongly independent set of loops in $\mathcal{M}$.

Second, if some component $\mathcal{M}'$ has more than one boundary cycle, let ℓ' be a path in $\mathcal{M}'$ between any two copies of x on different boundary cycles. $\mathcal{M}' \setminus \ell'$ has the same genus but one fewer boundary than $\mathcal{M}'$. This path corresponds to a loop ℓ in $\mathcal{M}$, and $L \cup \{\ell\}$ is still strongly independent.

Finally, suppose every component $\mathcal{M}'$ of $\mathcal{M} \setminus L$ is homeomorphic to a disk. As we argued earlier, each component has at least three copies of x on its boundary. If four or more copies of x appear on the boundary of some component $\mathcal{M}'$, we can triangulate $\mathcal{M}'$ by adding more loops to L without violating strong independence.

Thus, if L is a *maximal* strongly independent set of loops, each component of $\mathcal{M} \setminus L$ is a triangle, that is, a disk with exactly three copies of x on its boundary. Let t be the number of triangles. Each triangle is incident to three loops, and each loop to two triangles, so $3t = 2|L|$, and Euler's formula implies that $1 - |L| + t = 2 - 2g$. We conclude that $|L| = 6g-3$. □

The intersection of any loop ℓ with a small neighborhood of the basepoint consists of an *initial subpath* ℓ^+ and a *terminal subpath* ℓ^-. Let L be a set of loops with common basepoint. The *incidence pattern* of L records the cyclic order of its initial and terminal subpaths around the basepoint.

Lemma 2.2. *Let a and b be two simple, non-contractible, disjoint, homotopic loops on an orientable surface $\mathcal{M}$. The incidence pattern of these loops is $a^+a^-b^-b^+$. Moreover, the loop $a \cdot \bar{b}$ bounds an open disk whose intersection with a neighborhood of the basepoint consist of a wedge between a^+ and b^+ and a wedge between b^- and a^-.*

Proof: If the incidence pattern were either $a^+b^+a^-b^-$ or $a^+a^-b^+b^-$, there would be a simple cycle homotopic to the loop $a \cdot b \simeq a \cdot a$. (See Figure 1.) But this is impossible, because no loop homotopic to the square of a non-contractible simple loop is simple [11, Theorem 4.2]. So the incidence pattern must be $a^+a^-b^-b^+$.

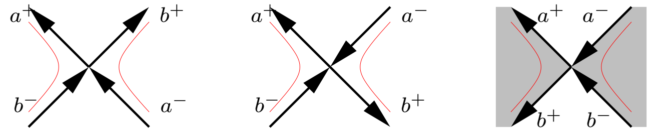

Figure 1. Lemma 2.2. Only the third incidence pattern is possible; a and b bound a disk containing the grey wedges.

Now, consider a contractible simple cycle γ that is a slightly translated copy of $a \cdot b^{-1}$ and intersects neither a nor b. In the neighborhood of the basepoint, γ is inside the wedge between a^+ and b^+ on one side, and inside the wedge between b^- and a^- on the other side. The disk bounded by γ cannot contain the basepoint, because it would then contain the non-contractible loops a and b. Thus, in the neighborhood of the basepoint, this disk also lies inside the same two wedges. □

2.3 Finding a Splitting Cycle

If length is not a consideration, we can compute a splitting cycle on any surface $\mathcal{M}$ in $O(n)$ time as follows. First we construct a simple non-contractible cycle, using a variant of an algorithm of Erickson and Har-Peled [12], which finds the shortest non-contractible loop with a given basepoint in $O(n \log n)$ time. The running time is dominated by the computation of a shortest-path tree rooted at x using Dijkstra's algorithm. If we ignore the edge lengths and use breadth-first search instead, the running time drops to $O(n)$; the modified algorithm still returns a simple non-contractible cycle α. If α is separating, then we are done. Otherwise, we compute another non-contractible cycle that crosses α exactly once, in $O(n)$ time. To do this, we choose an arbitrary vertex $x \in \alpha$ and compute a simple path β from one copy of x to the other copy in $\mathcal{M} \setminus \alpha$, using (for example) depth-first search. The cycle $\gamma = \alpha \cdot \beta \cdot \bar{\alpha} \cdot \bar{\beta}$ is simple and null-homologous, but not contractible (because α and β are not homotopic). Thus, γ is a splitting cycle; specifically, one of the components of $\mathcal{M} \setminus \gamma$ is a punctured torus.

3. NP-HARDNESS

Theorem 3.1. *Finding the shortest splitting cycle on a combinatorial surface is NP-hard.*

Proof: We describe a reduction from the following special case of the Euclidean traveling salesman problem [17]: *Given a set P of n points on the two-dimensional integer grid, are they connected by a tour whose length is exactly n?* Any such tour must lie entirely on the grid, and must consist of n axis-parallel unit-length segments, each joining a pair of points in P. Our reduction is similar to (and was inspired by) a proof by Eades and Rappaport [9] that computing the minimum-perimeter polygon separating a set of red points from a set of blue points in the plane is NP-hard; see also Arkin *et al.* [1].

We describe a two-step reduction. Let P be a set of n points on the $n \times n$ integer grid in the plane. To begin the first reduction, we overlay n 4×4 square grids of width $1/4n$, one centered on each point in P, on top of the $n \times n$

integer grid. In each small grid, we color the square in the second row and second column *red* and the square in the third row and third column *blue*. We now easily observe that the following question is NP-complete: *Does the modified grid contain a cycle of length at most $n + 1/2$ that separates the red squares from the blue squares?* Any TSP tour of P of length n can be modified to produce a separating cycle of length at most $n + 1/2$ by locally modifying the tour within each small grid, as shown in Figure 2(a) and (b). Conversely, any separating cycle must pass through the center points of all n small grids, which implies that any separating cycle of length at most $n + 1/2$ must contain n grid edges that comprise a TSP tour of P.

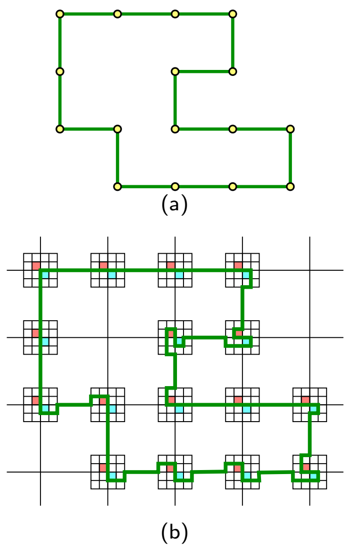

Figure 2. (a) A TSP tour of length n. (b) The corresponding red/blue separating cycle (not to scale).

Next we reduce the square-separation problem to finding a minimum-length splitting cycle. We isometrically embed the modified grid on a sphere, which we call *Earth*. We remove the red and blue squares to create $2n$ punctures, which we attach to two new punctured spheres, called *heaven* and *hell*. We attach the n punctures in heaven to the n blue punctures on Earth; similarly, we attach the n punctures in hell to the n red punctures on Earth. We append edges of length n^3 to the resulting surface so that each face of the final embedded graph is a disk. The resulting combinatorial surface $\mathcal{M}(P)$ has genus $2n - 2$ and complexity $O(n^2)$, and it can clearly be constructed in polynomial time. See Figure 3.

The shortest cycle γ that splits $\mathcal{M}(P)$ must lie entirely on Earth, since the edges in heaven and hell are far too long. Moreover, γ must separate the blue punctures from the red punctures; otherwise, $\mathcal{M}(P) \setminus \gamma$ would be connected by a path through heaven or through hell. Thus, γ is precisely the shortest cycle that separates the red and blue squares in our intermediate problem. This cycle has length at most $n + 1/2$ if and only if the original points P have a tour of length n. □

With a few trivial modifications, our reduction also implies that computing the shortest splitting cycle on a polyhedral or Riemann surface is NP-hard, although it is an open question whether the corresponding decision problems are still in NP.

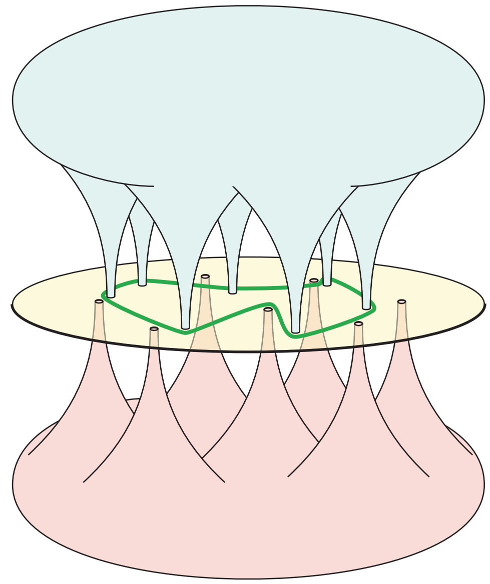

Figure 3. Separating heaven from hell (not to scale); the central disk is a small portion of Earth.

4. STRUCTURAL PROPERTIES

For any two points x and y on a cycle α, we let $\alpha[x, y]$ denote the path from x to y along α, taking into account the orientation of α. For a path or a dual edge α, the same notation is used for the unique simple path between x and y on α.

4.1 Multiplicity Bound

Lemma 4.1. *Any shortest splitting cycle on an orientable cross-metric surface $\mathcal{M}$ crosses each edge e^* of $G^*(\mathcal{M})$ at most once in each direction.*

Proof: Assume for the purpose of contradiction that some shortest splitting cycle γ crosses some dual edge e^* twice in the same direction, say left to right. Let x and z be consecutive left-to-right intersection points along e^*; that is, γ does not cross $e^*[x, z]$ from left to right. The path $e^*[x, z]$ is on one side of γ near x, but on the other side of γ near z. Because γ separates the surface, γ must cross e^* from right to left at some point y between x and z.

The cycles $\gamma[x, y] \cdot e^*[y, x]$ and $\gamma[y, z] \cdot e^*[z, y]$ are non-contractible; otherwise, we could shorten γ by removing two crossings with e^* without changing its homotopy class.

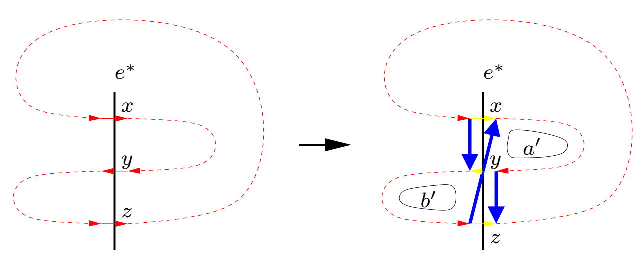

Figure 4. Lemma 4.1. If γ crosses e^* twice in the same direction, we can remove two crossings.

Define a new cycle $\gamma' = \gamma[x, y] \cdot e^*[y, z] \cdot \gamma[z, x] \cdot e^*[x, y] \cdot \gamma[y, z] \cdot e^*[z, x]$, as shown in Figure 4. The new cycle γ'

is simple, because x, y, z are consecutive along e^*. The cycle γ' is in the same homology class as γ and is therefore null-homologous. By translating the non-contractible cycles $\gamma[x,y] \cdot e^*[y,x]$ and $\gamma[y,z] \cdot e^*[z,y]$ away from γ', we obtain two non-contractible cycles a' and b', one in each component of $\mathcal{M} \setminus \gamma'$; so γ' is non-contractible. Finally, γ' crosses e^* two times fewer than γ and crosses every other edge in $G^*(\mathcal{M})$ the same number of times as γ. We conclude that γ' is a splitting cycle that is shorter than γ, which is impossible. □

4.2 Shortest-Path Crossing Bound

Proposition 4.2. *Let P be a set of pairwise interior-disjoint shortest paths on an orientable cross-metric surface $\mathcal{M}$. There is a shortest splitting cycle that crosses each path in P at most $48g - 24$ times.*

Proof: Let γ be a shortest splitting cycle with the minimum number of crossings with paths in P. We can assume that γ does not pass through the endpoints of any path in P, since we are on a cross-metric surface and could simply perturb γ slightly. Consider any path p in P that intersects γ.

The intersection points $\gamma \cap p$ partition γ into *arcs*. These arcs may intersect other paths in P. Let $\mathcal{M}/p$ be the quotient surface obtained by contracting p to a point p/p. Each arc corresponds to a loop in $\mathcal{M}/p$ with basepoint p/p. We say that two such arcs are *homotopic rel p* or *relatively homotopic* if the corresponding loops in $\mathcal{M}/p$ are homotopic.

For any consecutive intersection points x and y along γ, the arc $\gamma[x,y]$ cannot be homotopic to $p[x,y]$, since otherwise we can obtain a no longer splitting cycle $\gamma[y,x] \cdot p[x,y]$ that has fewer crossings with the paths in P. It follows that none of the arcs of γ are contractible rel p. Since the arcs are disjoint except at their common endpoints, Lemma 2.1 implies that there are at most $12g - 6$ relative homotopy classes of arcs.

We can partition the arcs into four types—LL, RR, LR, and RL—according to whether the arcs start on the left or right side of p, and whether they end on the left or right side of p. To complete the proof, we argue that there is at most one arc of each type in each relative homotopy class.[2] We explicitly consider only types LL and RL; the other two cases follow from symmetric arguments.

Case 1 [LL]: Suppose for purposes of contradiction that there are two LL arcs $u = \gamma[a,z]$ and $w = \gamma[c,x]$ that are homotopic rel p. By Lemma 2.2, without loss of generality, the intersection points appear along e^* in the order a, c, x, z, and the cycle $u \cdot p[z,x] \cdot \bar{w} \cdot p[c,a]$ bounds a disk. Without loss of generality, we assume that no other arc homotopic rel p intersects this disk. Since γ is separating, there must be exactly one arc $v = \gamma[y,b]$ between u and w that is relatively homotopic to $\bar{u}$ and $\bar{w}$.

Without loss of generality, suppose the path $\gamma[x,a]$ does not contain any of the arcs u, v, or w. Consider the cycle

$$\gamma' = p[a,b] \cdot \gamma[b,c] \cdot p[c,b] \cdot \bar{\gamma}[b,y] \cdot p[y,z] \cdot \gamma[z,y] \cdot p[y,x] \cdot \gamma[x,a]$$

obtained by removing u and w from γ, reversing v, and connecting the remaining pieces of γ with subpaths of p; see

[2]More careful analysis implies that there are at most three arcs in any relative homotopy class.

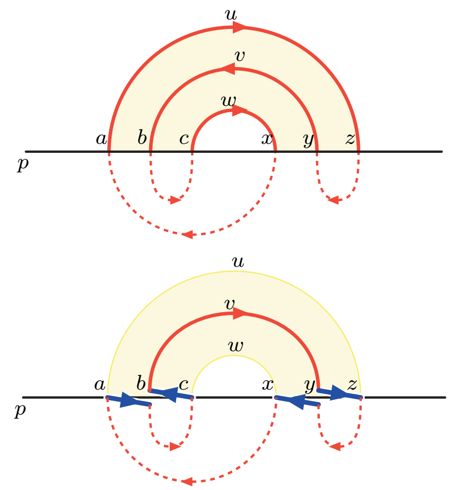

Figure 5. The exchange argument for LL arcs.

Figure 5. This cycle crosses p fewer times than γ, and crosses any other path in p no more than γ. An argument identical to the proof of Lemma 4.1 implies that γ' is simple, null-homologous, and non-contractible. Because p is a shortest path, u cannot be shorter than $p[a,z]$, which implies that γ' is no longer than γ, contradicting our assumption that γ is a shortest splitting cycle with the minimal number of crossings with paths in P.

We conclude that any two LL-arcs must be in different relative homotopy classes.

Case 2 [RL]: Now suppose for purposes of contradiction that there are two RL arcs $u = \gamma[a,x]$ and $w = \gamma[c,z]$ that are relatively homotopic. By Lemma 2.2, the cycle $u \cdot p[x,z] \cdot \bar{w} \cdot p[c,a]$ bounds a disk D. Without loss of generality no arc relatively homotopic to u and w lies inside this disk. Up to symmetry, we can assume that a precedes c and x precedes z along edge e^*. Point x cannot lie between a and c, because then arc u would end on the boundary of D, and γ could not exit D without creating a self-intersection, an arc that is contractible rel p, or an arc in D that is relatively homotopic to u and w, none of which are possible. Thus, the four intersection points must appear in the order a, c, x, z, possibly with $c = x$. See Figure 6.

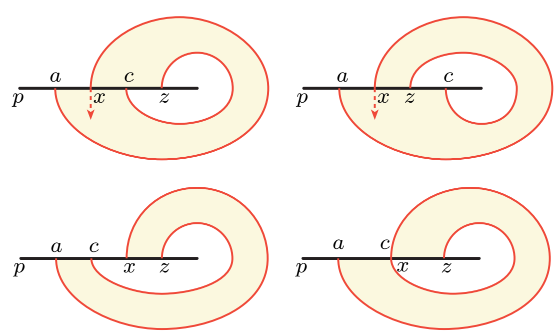

Figure 6. Two impossible incidence orders (top) and two possible incidence orders (bottom) for two RL arcs; disk D is shaded.

As in the previous case, because γ is a splitting cycle, there must be exactly one arc $v = \gamma[y,b]$ between u and w that is homotopic to $\bar{u}$ and $\bar{w}$. If $c = x$, there is only one way to connect these three arcs to form the cycle γ; if $c \neq x$, there are two possibilities. See Figure 7. In all three cases, by deleting arcs u and w, reversing arc v, and connecting the remaining pieces of γ with subpaths of p, we create a splitting cycle γ' that is no longer than γ and crosses fewer times the paths in P than γ, which is impossible. We omit the tedious details.

We conclude that any two RL-arcs must be in different relative homotopy classes. □

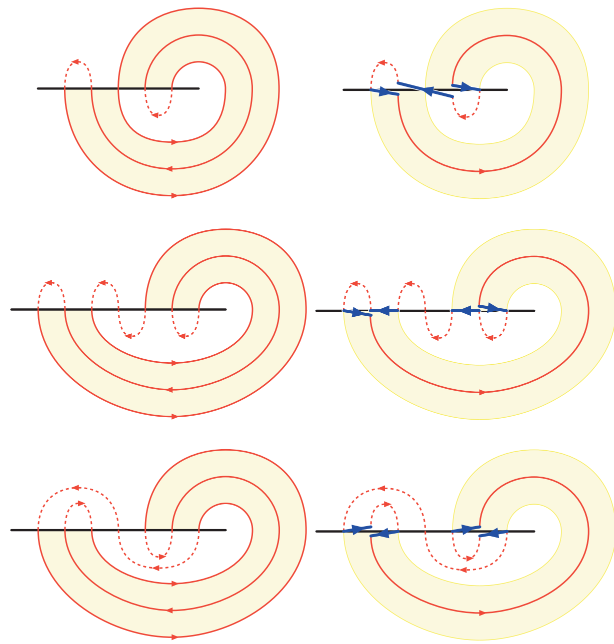

Figure 7. Three exchange arguments for RL arcs.

4.3 Crossing Lower Bound

Earlier algorithms for computing shortest non-contractible, non-separating, or essential cycles rely on the fact that the desired cycle is the concatenation of two equal-length shortest paths [21, 12].[3] Cabello and Mohar [4], Cabello [3], and Kutz [18] exploit a slightly different property to compute shortest nontrivial cycles more quickly on surfaces of constant genus: The desired shortest cycle crosses any shortest path at most twice. As we prove next, neither of these properties holds for the shortest splitting cycle; in particular, the upper bound of Proposition 4.2 is tight up to constant factors.

Theorem 4.3. *For any $g \geq 2$, there is an orientable combinatorial surface $\mathcal{M}_g$ of genus g whose unique shortest splitting cycle (up to orientation) crosses a shortest path g times and (therefore) cannot be decomposed into fewer than g shortest paths and edges.*

Proof: We consider only the case where g is even. We construct a combinatorial surface $\mathcal{M}_g$ of genus g such that the shortest non-contractible cycle γ and the shortest splitting cycle μ cross g times; see Figure 8. The shortest non-contractible cycle γ can be partitioned into two shortest paths, one of which will be disjoint from μ; it follows that μ crosses some shortest path g times.

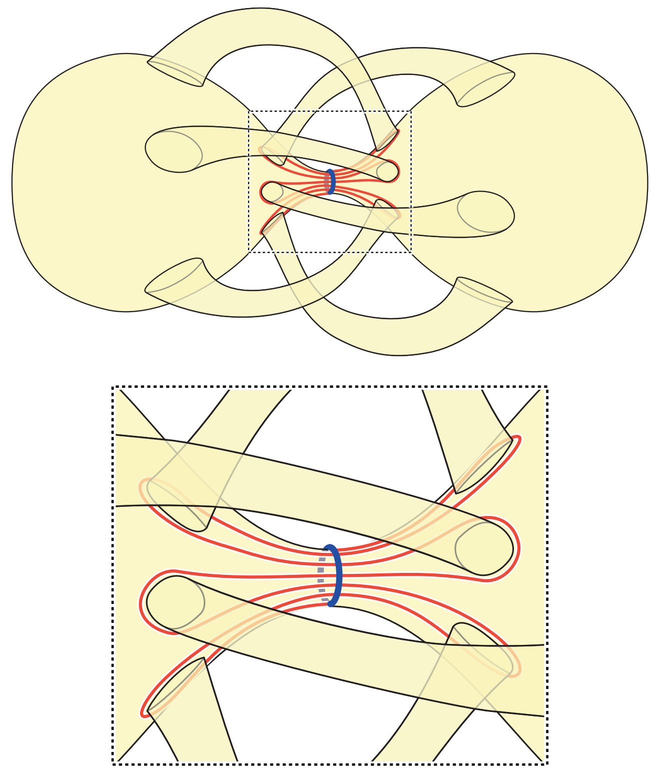

Figure 8. A surface whose shortest splitting cycle cuts a shortest path g times, and a closeup of the undulating shortest splitting cycle.

The base surface $\mathcal{M}_0$ is a sphere whose geometry approximates an hourglass. Let γ be the central cycle that partitions the two lobes of the hourglass. To construct $\mathcal{M}_g$, we attach g handles to this hourglass, each joining a small circle c_i on the neck of the hourglass, just to one side of γ, to a large circle C_i far away on the opposite lobe. The small circles c_i are arranged symmetrically around the neck of the hourglass, alternating between the two sides of γ; the large circles C_i are also partitioned evenly between the two lobes of the hourglass.

Let μ be a cycle that undulates around the small circles in order, crossing γ a total of g times, as shown in Figure 8. We easily verify that μ splits $\mathcal{M}_g$ into two surfaces of genus $g/2$. Let $x_1, x_2, \ldots x_g = x_0$ denote the g intersection points of μ and γ. For each i, let μ_i and γ_1 respectively denote the subpaths of μ and γ between x_{i-1} and x_i. Finally, we partition path γ_1 into three subpaths $\gamma_1^\flat$, $\gamma_1^\natural$, and $\gamma_1^\sharp$ at arbitrary points y and z. Let F denote the union of these $2g+2$ paths.

To obtain a combinatorial structure on $\mathcal{M}_g$, we embed a weighted graph G that contains F as a subgraph, such that every face of the embedding is a topological disk. We assign length 1 to edge γ_i for each $i \neq 1$, length 1 to edges $\gamma_1^\flat$ and $\gamma_1^\sharp$, length $g+1$ to edge $\gamma_1^\natural$, length $4g^2$ to each edge μ_i, and length at least $10g^5$ to every other edge in G. The cycle μ does not contain a single (even approximate!)

[3] In general, this characterization requires shortest paths that terminate in the interior of edges, but if we refine edges appropriately, the shortest cycle will indeed be the concatenation of two shortest vertex-to-vertex paths.

vertex-to-vertex shortest path. Even if we allow shortest paths between points in the interior of edges, each such path contains at most one vertex x_i. The cycle γ can clearly be partitioned into two shortest paths of length $g+1$ at points y and z, and μ crosses one of these paths g times. Thus, to complete the proof, it suffices to show that μ is in fact the shortest splitting cycle in $\mathcal{M}_g$.

Let α be a shortest splitting cycle; the assigned edge weights guarantee that α is a circuit in F. By Lemma 4.1, α traverses each path γ_i and μ_i at most once in each direction. For any i, there is a path in $\mathcal{M}_g \setminus F$ from one side of γ_i to the other, so α must traverse each edge γ_i either twice (in opposite directions) or not at all.

Consider any simple cycle β in a tubular neighborhood of F that traverses every edge in F either once in each direction or not at all. This cycle must be null-homologous, and therefore separating. Since β lies inside a small tubular neighborhood of F, which has genus zero, some component of $\mathcal{M} \setminus \beta$ has genus zero. Moreover, β is the only boundary of this component. We conclude that β is contractible.

The splitting cycle α is not contractible, so it must traverse some edge μ_i exactly once. But α must traverse the edges adjacent to any vertex x_i an even number of times, which implies that α traverses *every* edge μ_i exactly once. Thus, every splitting cycle in F is at least as long as μ. We conclude that μ is indeed the unique shortest splitting cycle. □

5. ALGORITHM

In this section, we prove that computing the shortest splitting cycle is fixed-parameter tractable with respect to the genus of the surface.

Theorem 5.1. *Let $\mathcal{M}$ be an orientable cross-metric surface without boundary; let g be its genus and n be its complexity. We can compute a shortest splitting cycle in $\mathcal{M}$ in $g^{O(g)} n \log n$ time.*

The algorithm proceeds in several stages, described in detail in the following subsections. First, we compute the shortest system of loops from some arbitrary basepoint, using the greedy algorithm of Erickson and Whittlesey [13]. Next, we enumerate all possible sequences of crossings of this system of loops by a simple cycle that crosses each loop $O(g)$ times. Proposition 4.2 implies that the shortest splitting cycle must have one of these crossing sequences. We discard any crossing sequence that does not correspond to a splitting cycle. For each valid crossing sequence, we compute a shortest cycle γ with that crossing sequence using the recent algorithm of Kutz [18]. Finally, we post-process γ to remove any self-intersections, without changing its length or its free homotopy type. Out of all cycles constructed this way, we return the shortest one.

5.1 Greedy Loops

Let v be any point of $\mathcal{M}$ in the interior of a face of $G^*(\mathcal{M})$. Let $\alpha_1, \alpha_2, \ldots, \alpha_{2g}$ be the shortest system of loops of $\mathcal{M}$ with basepoint v; this system of loops can be computed in $O(n \log n + gn)$ time using a greedy algorithm of Erickson and Whittlesey [13].

The key property of this system is that each loop α_i is composed of two shortest paths in the primal graph $G(\mathcal{M})$; however, in general these two paths meet at a point m_i in the interior of some edge e_i. To simplify our algorithm, we split e_i into two edges at m_i—or equivalently, in the dual graph, we replace the dual edge e_i^* with two parallel edges—partitioning the length appropriately, so that α_i consists of two vertex-to-vertex shortest paths β_i and β_i' in $G(\mathcal{M})$.

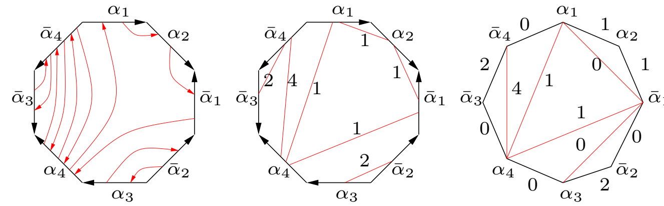

Figure 9. Left: A splitting cycle. Middle: The corresponding subsets of segments; each label indicates the number of segments contained in a subset. Right: The corresponding weighted triangulation.

5.2 Simple Crossing Sequences

The *crossing sequence* of a cycle γ records the intersections of γ with the greedy loops α_i, in cyclic order along γ. Any two cycles with the same crossing sequence are homotopic, although two homotopic cycles can have different crossing sequences. A crossing sequence is *simple* if can be generated by a simple cycle; non-simple cycles can have simple crossing sequences.

Proposition 4.2 implies that some shortest splitting cycle γ crosses each path β_i or β_i' $O(g)$ times, and thus crosses each loop α_i $O(g)$ times. Our algorithm therefore enumerates a superset of all simple crossing sequences that contain each loop α_i $O(g)$ times. Note that there are $g^{\Theta(g^2)}$ crossing sequences with $O(g)$ occurrences of each loop, but the vast majority of these are not simple. Thus, we cannot naively enumerate crossing sequences that satisfy Proposition 4.2 and then check each sequence for simplicity. Our enumeration algorithm enforces simplicity from the beginning.

We begin by cutting $\mathcal{M}$ along the loops α_i to obtain a polygonal schema $\mathcal{D}$; this is a cross-metric disk with complexity $O(gn)$. This cutting operation also cuts the unknown splitting cycle γ into *segments* that cut across $\mathcal{D}$. Because γ is simple, no two of these segments cross. Because v does not lie on $G^*(\mathcal{M})$, we can slightly perturb γ without changing its length (in the crossing metric) or its homotopy class. Thus, we assume that γ does not pass through the basepoint v.

The segments of γ can be grouped into subsets according to which pair of greedy loops they meet on the boundary of $\mathcal{D}$ (Figure 9). We abstract and dualize the polygonal schema, replacing each boundary path with a vertex and connecting vertices that correspond to consecutive edges. Now each subset of segments corresponds to an diagonal between two vertices of the dual $4g$-gon. Since no two segments cross, these diagonals cannot cross. In particular, all the diagonals belong to some triangulation of the dual polygon.

Thus the candidate crossing sequences of a shortest splitting cycle are described by *weighted triangulations*, which consist of a triangulation of the dual polygon, each of whose edges is weighted with an integer between 0 and $O(g)$. The label of an edge in the triangulation represents the number of times that the cycle runs along that edge. There are $C_{4g-2} = O(4^{4g})$ possible triangulations, where C_n is the nth

Calatan number, which we can enumerate in $O(g)$ time each. (This is essentially identical to the enumeration of binary trees.) There are $g^{O(g)}$ ways to label each triangulation, which we can enumerate in constant amortized time per labeling.

We thus obtain a total of $g^{O(g)}$ weighted triangulations. Most of these do not correspond to a splitting cycle, or indeed to *any* cycle. We now explain how to discard these possibilities.

5.3 Testing Weighted Triangulations

Given a candidate weighted triangulation T, we want to test whether T corresponds to a splitting cycle. We must check that (1) T corresponds to a set of cycles, (2) this set contains exactly one cycle, and (3) this cycle is separating but non-contractible. (We already know that the cycle is simple.) We describe how to perform these tests in $O(g^2)$ time. If any of these properties is not satisfied, we simply discard T.

To check that T corresponds to a set of cycles, it suffices to check that the two boundary edges of $\mathcal{D}$ that correspond to the same α_i are crossed the same number of times.

For the remaining tests, we build a combinatorial surface $\mathcal{M}'$ homeomorphic to $\mathcal{M}$, whose graph G' is the arrangement of the greedy system of loops and the candidate cycle(s) defined by T. We cut the abstract polygonal schema along the subsets of edges given by the triangulation and then identify corresponding subpaths on the boundary of the polygonal schema. If we have multiple edges running along the same edge of the triangulation, these define thin rectangular strips. The complexity of the resulting surface is $O(g^2)$. The $O(n)$ internal complexity of the original surface $\mathcal{M}$ is ignored.

Once we have constructed $\mathcal{M}'$, we can check whether T defines a single cycle γ by a simple depth-first search.

To test that γ is separating but non-contractible, we use a simplification of an algorithm of Erickson and Har-Peled [12]. To test separation, we perform a depth- or breadth-first search on the faces of $\mathcal{M}'$, starting from any initial face, but forbidding crossings of any segment of γ. The cycle γ is separating if and only if the search halts before visiting every face of the surface. (Alternately, γ is separating if and only if γ crosses each α_i the same number of times in both directions.) We also compute the Euler characteristic of the reachable portion of the surface during the search by counting vertices, edges, and faces. The cycle is non-contractible if and only if this Euler characteristic is neither 1 (a disk) nor $1 - 2g$ (the complement of a disk).

5.4 From Crossing Sequence to Cycle

For each valid weighted triangulation, we compute the shortest cycle with the corresponding crossing sequence in time $O(g^2 n \log n)$ using the recent algorithm of Kutz [18]. For the sake of completeness, we sketch the algorithm here.

First we glue together a cycle of $O(g^2)$ distinct copies of the polygonal schema $\mathcal{D}$, one per crossing in the sequence. Each successive pair of copies is glued along the edge specified by the corresponding entry in the crossing sequence. Because we consider only crossing sequences without *spurs*—the cycle does not cross a loop α_i and then immediately recross the same loop α_i in the opposite direction—the resulting combinatorial surface is an annulus, which we denote $\mathcal{D}^\circ$. (In Kutz's terms, we are considering only *curl-free* splitting cycles.) The polygonal schema $\mathcal{D}$ has complexity $O(gn)$, so the complexity of $\mathcal{D}^\circ$ is $O(g^3 n)$.

We then compute the shortest cycle γ° in $\mathcal{D}^\circ$ that is freely homotopic to the boundaries of $\mathcal{D}^\circ$, using an algorithm of Frederickson [14]; see also [6, Lemma 3.5(d)]. Given a combinatorial annulus of complexity N, Frederickson's algorithm finds a shortest generating cycle in $O(N \log N)$ time. Thus, we compute γ° in $O(g^3 n \log(g^3 n)) = O(g^3 n \log n)$ time.

Finally, projecting γ° back to the original surface $\mathcal{M}$ gives us a shortest cycle γ with the given crossing sequence.

Since there are $g^{O(g)}$ valid crossing sequences, the total time spent in this phase of our algorithm is $g^{O(g)} \times O(g^3 n \log n) = g^{O(g)} n \log n$.

5.5 Removing Self-intersections

Let γ be the shortest cycle computed in the previous phase, over all possible valid crossing patterns. This cycle is null-homologous, non-contractible, as short as possible in its homotopy class, and freely homotopic to a simple cycle, but *not* necessarily simple. Although the cycle γ° is simple, projecting it back to the original surface may introduce self-intersections.

Because γ is homotopic to a simple cycle, a theorem of Hass and Scott [15] (see also [7]) implies that if γ is self-intersecting, some pair of subpaths of γ bounds a disk, which we call a *bigon*. Any bigon can be removed by simple surgery, as shown in Figure 10, without changing its length or its homotopy type. Within small neighborhoods of the two self-intersection points, we replace the two intersecting subpaths with two parallel subpaths. A straightforward inductive argument now implies that there is a simple cycle γ' homotopic to γ and with the same length as γ. This is our shortest splitting cycle.

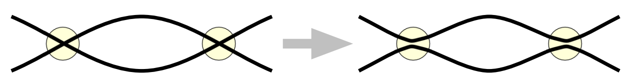

Figure 10. Surgery to remove a bigon.

Moreover, γ' crosses the greedy loops α_i at precisely the same points as the original cycle γ, although possibly in a different order. By Lemma 4.1, the cycle γ has complexity $O(n)$, and thus we can easily determine these $O(g^2)$ intersection points in $O(gn)$ time. These points partition γ' into $O(g^2)$ *segments*. The fact that γ' is as short as possible in its homotopy class implies that each segment is a shortest path in $\mathcal{D}$ between its two endpoints.

There is exactly one way to connect these intersection points by paths into a simple cycle with the same crossing sequence as γ. This pairing can be determined easily by referring back to the weighted triangulation used to determine the crossing sequence. We can compute non-crossing shortest paths in the planar graph $\mathcal{D}$ between the $O(g^2)$ pairs of boundary points in $O(gn \log g) = O(gn \log n)$ time, using an algorithm of Takahashi *et al.* [23]. Concatenating these paths at their common endpoints gives us the shortest splitting cycle γ'.

This concludes the proof of Theorem 5.1.

6. CONCLUSIONS

The results of this paper suggest several open problems. Most notably, can we approximate the shortest splitting cycle, or is that also NP-hard? The following high-level approach seems promising. Compute shortest simple cycles in each non-trivial homotopy class in order of increasing length, stopping either when we find a separating cycle, or when we find two cycles α and β that intersect an odd number of times. If we find a separating cycle, it is of course the shortest splitting cycle. If we find two cycles with odd intersection number, an exchange argument implies that the intersection number is 1. For some orientation of α and β, the cycle $\alpha \cdot \beta \cdot \bar{\alpha} \cdot \bar{\beta}$ is a splitting cycle whose length is at most four times the length of the shortest splitting cycle. Can this algorithm be implemented efficiently? How quickly can we enumerate the k shortest homotopy classes of (simple) cycles? Techniques of Eppstein [10] for enumerating k shortest paths may be useful here.

If we iterately cut along splitting cycles, we optain a decomposition of the surface into punctured tori. However, repeatedly cutting the surface along its shortest splitting cycle does not necessarily yield the shortest such decomposition. Is computing the shortest torus decomposition NP-hard? Similarly, a *pants decomposition* is a set of disjoint simple cycles decomposing a surface into pairs of pants, or spheres with three boundary components [7]. Is it NP-hard to compute the shortest pants decomposition? Are either of these problems fixed-parameter tractable?

Finally, Erickson and Har-Peled [12] prove that finding the minimum cut graph is NP-hard, by a reduction from the fixed-parameter tractable rectilinear Steiner tree problem. Is computing the shortest cut graph fixed-parameter tractable? The most serious bottleneck here seems to be computing the shortest cut graph in a given homotopy class, where the homotopy is allowed to move the vertices.

Acknowledgments. Portions of this work were done while all five authors were at the University of Illinois and later at École normale supérieure, thanks to the support of a UIUC/CNRS/INRIA travel grant. We thank the other participants of the 1st UIUC/CNRS/INRIA Workshop on Computational Topology for interesting discussions and meals. We also thank Therese Biedl, Sergio Cabello, Sariel Har-Peled, Martin Kutz, Kevin Milans, and Günter Rote for useful discussions while this work was in progress. We are especially grateful to Martin Kutz for sending us an advance copy of his paper [18].

References

[1] E. M. Arkin, S. Khuller, and J. S. B. Mitchell. Geometric knapsack problems. *Algorithmica* 10:399–427, 1993.

[2] J. Blömer. Computing sums of radicals in polynomial time. *Proc. 32nd Annu. IEEE Sympos. Found. Comput. Sci.*, 670–677, 1991.

[3] S. Cabello. Many distances in planar graphs. *Proc. 17th Annu. ACM-SIAM Sympos. Discrete Algorithms*, 1213–1220, 2006.

[4] S. Cabello and B. Mohar. Finding shortest non-separating and non-contractible cycles for topologically embedded graphs. *Proc. 13th Annu. European Sympos. Algorithms*, 131–142, 2005. Lecture Notes in Computer Science 3669, Springer-Verlag.

[5] J. Chen and Y. Han. Shortest paths on a polyhedron, part I: Computing shortest paths. *Int. J. Comput. Geom. Appl.* 6(2):127–144, 1996.

[6] É. Colin de Verdière and J. Erickson. Tightening non-simple paths and cycles on surfaces. *Proc. 17th Annu. ACM-SIAM Sympos. Discrete Algorithms*, 192 – 201, 2006.

[7] É. Colin de Verdière and F. Lazarus. Optimal pants decompositions and shortest homotopic cycles on an orientable surface. *Proc. 11th Sympos. Graph Drawing*, 478–490, 2003. Lecture Notes Comput. Sci. 2912, Springer-Verlag.

[8] É. Colin de Verdière and F. Lazarus. Optimal systems of loops on an orientable surface. *Discrete Comput. Geom.* 33(3):507–534, 2005.

[9] P. Eades and D. Rappaport. The complexity of computing minimum separating polygons. *Pattern Recogn. Lett.* 14(9):715–718, 1993.

[10] D. Eppstein. Finding the k shortest paths. *SIAM J. Computing* 28(2):652–673, 1998.

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

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

[13] J. Erickson and K. Whittlesey. Greedy optimal homotopy and homology generators. *Proc. 16th Annu. ACM-SIAM Sympos. Discrete Algorithms*, 1038–1046, 2005.

[14] G. N. Frederickson. Fast algorithms for shortest paths in planar graphs, with applications. *SIAM J. Comput.* 16(6):1004–1022, 1987.

[15] J. Hass and P. Scott. Intersections of curves on surfaces. *Israel J. Math.* 51:90–120, 1985.

[16] A. Hatcher. *Algebraic Topology*. Cambridge University Press, 2002. ⟨http://www.math.cornell.edu/~hatcher/⟩.

[17] D. S. Johnson and C. H. Papadimitriou. Computational complexity and the traveling salesman problem. *The Traveling Salesman Problem*, 68–74, 1985. John Wiley & Sons.

[18] M. Kutz. Computing shortest non-trivial cycles on orientable surfaces of bounded genus in almost linear time. These proceedings, 2006.

[19] F. Lazarus, M. Pocchiola, G. Vegter, and A. Verroust. Computing a canonical polygonal schema of an orientable triangulated surface. *Proc. 17th Annu. ACM Sympos. Comput. Geom.*, 80–89, 2001.

[20] J. S. B. Mitchell, D. M. Mount, and C. H. Papadimitriou. The discrete geodesic problem. *SIAM J. Comput.* 16(4):647–668, 1987.

[21] B. Mohar and C. Thomassen. *Graphs on Surfaces*. Johns Hopkins University Press, 2001.

[22] J. Stillwell. *Classical Topology and Combinatorial Group Theory*. Springer-Verlag, New York, 1993.

[23] J. Takahashi, H. Suzuki, and T. Nishizeki. Shortest noncrossing paths in plane graphs. *Algorithmica* 16:339–357, 1996.

[24] A. Zomorodian. *Topology for Computing*. Cambridge University Press, 2005.