Greedy Optimal Homotopy and Homology Generators *

Jeff Erickson† Kim Whittlesey‡

Abstract

We describe simple greedy algorithms to construct the shortest set of loops that generates either the fundamental group (with a given basepoint) or the first homology group (over any fixed coefficient field) of any oriented 2-manifold. In particular, we show that the shortest set of loops that generate the fundamental group of any oriented combinatorial 2-manifold, with any given basepoint, can be constructed in $O(n \log n)$ time using a straightforward application of Dijkstra's shortest path algorithm. This solves an open problem of Colin de Verdière and Lazarus.

In memory of John R B Whittlesey (1927–2003)

1 Introduction

Several geometric problems call for topologically complex surfaces to be cut into one or more topological disks. Examples in computer graphics include denoising, texture mapping, remeshing, compression, and morphing (see references in [8]); more theoretical examples include computing separators and tree decompositions of non-planar graphs [7]. In light of these applications, a natural algorithmic problem is to find *optimal* method for cutting surfaces to simplify their topology.

Erickson and Har-Peled [8] were the first to consider the problem of optimally cutting a surface into a single topological disk by removing a so-called *cut graph* of minimum total length. Erickson and Har-Peled showed that computing minimum-length cut graphs is NP-hard, by a reduction from the classical Steiner tree problem. They also developed a brute-force polynomial-time algorithm for manifolds with constant genus and a constant number of boundary components, as well as a greedy algorithm that computes a $O(\log^2 g)$-approximation of the shortest cut graph in $O(g^2 n \log n)$ time.

Colin de Verdière and Lazarus [5] considered the special case of one-vertex cut graphs, which they called *systems of loops*. Every system of loops for an orientable surface of genus g with no boundary contains $2g$ loops through a common basepoint. Systems of loops also provide a minimal presentation for the fundamental group of the surface. Given a triangulated manifold and a system of loops as input, their algorithm computes the shortest system of loops in the same homotopy class, in polynomial time under a mild assumption about the input geometry. A more recent extension of this algorithm by the same authors optimizes *pants decompositions* within a given *free* homotopy class [6].

At the end of their paper [5], Colin and Verdière and Lazarus ask, "How does one compute the shortest system of loops, among all systems, relaxing the homotopy condition? Comparing with the work of Erickson and Har-Peled, we expect this last problem to be much less tractable than those solved in the present paper." In this paper, we show that finding the shortest system of loops is considerably *more* tractable than either finding the shortest cut graph or finding the shortest system of loops in a given homotopy class. We describe a simple greedy algorithm, based on Dijkstra's shortest path algorithm, that computes the shortest system of loops with a given basepoint in $O(n \log n)$ time. Running this algorithm once for every basepoint gives us the overall shortest system of loops in $O(n^2 \log n)$ time.

The difference between finding the shortest cut graph and finding the shortest system of loops boils down to how the lengths of these objects are defined. The length of a cut graph is just the sum of the lengths of its edges. The length of a system of loops, however, is the sum of the lengths of the loops; if any path is traversed more than once by the loops, or even by the same loop, its length is counted more than once. The situation is similar to the difference between *Steiner* trees and minimum *spanning* trees for points in the plane. The minimum Steiner tree is a set of paths that touches every point whose total length is as small as possible; the Euclidean minimum spanning tree is a set of *point-to-point* paths that touch every point, the *sum of whose lengths* is as small as possible. Computing minimum Steiner trees is NP-hard; computing minimum spanning trees is easy. (We develop this analogy further in Section 3.3; see Lemma 3.8.)

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

†Department of Computer Science, University of Illinois at Urbana-Champaign; jeffe@cs.uiuc.edu; http://www.cs.uiuc.edu/~jeffe/. Partially supported by NSF CAREER award CCR-0093348 and NSF ITR grants DMR-0121695 and CCR-0219594

‡Department of Mathematics, University of Illinois at Urbana-Champaign; kwhittle@math.uiuc.edu; http://www.math.uiuc.edu/~kwhittle/.

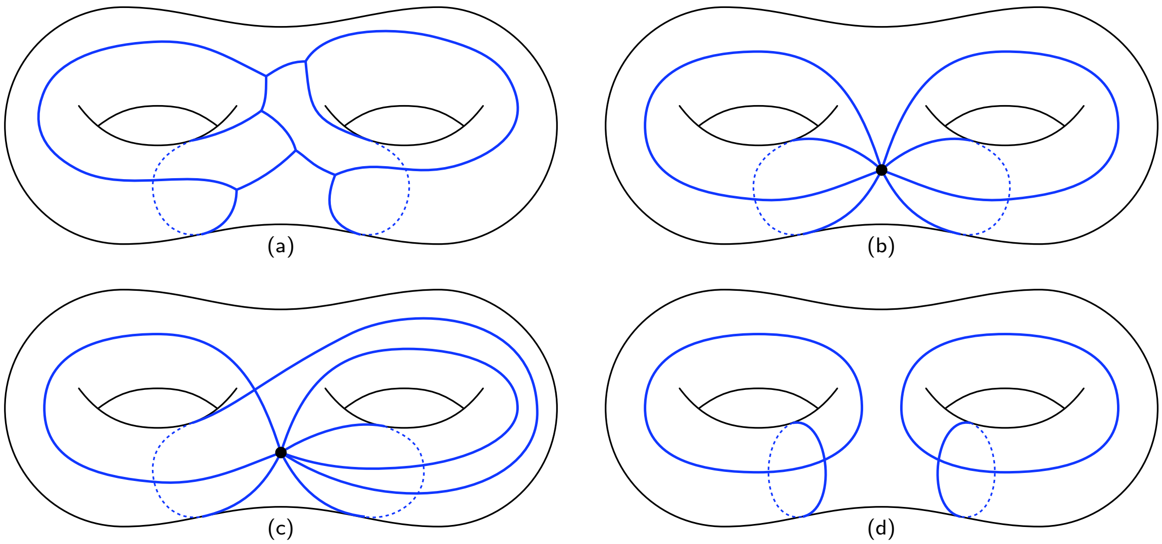

Figure 1. Four types of graphs on a two-holed torus. (a) A cut graph that is not a system of loops. (b) A system of loops. (c) A homotopy basis that is not a system of loops. (d) A homology basis that is neither a homotopy basis nor a cut graph.

The rest of the paper is organized as follows. In Section 2, we describe some relevant topological background. We develop our greedy algorithm for computing optimal systems of loops in Section 3. Finally, in Section 4, we generalize our techniques to find the shortest set of loops that generate the *homology* group of a given 2-manifold, over any coefficient field, in polynomial time.

2 Topological Background

We begin with some standard definitions from topology; for a more thorough introduction, we refer the interested reader to Hatcher [11] or Stillwell [28].

A *2-manifold* is a topological space in which every point has a neighborhood homeomorphic to $\mathbb{R}^2$. This paper considers only connected, compact, orientable 2-manifolds without boundary. The *genus* of a 2-manifold is the number of disjoint cycles that can be removed without disconnecting the manifold. Two connected, compact, orientable, 2-manifolds without boundary are homeomorphic if and only if they have the same genus.

2.1 Loops and Homotopy. Let x be a fixed *basepoint* in some 2-manifold M. A *loop based at* x is (the image of) a continuous function $\ell : [0,1] \to M$ such that $\ell(0 = \ell(1) = x$. Two loops ℓ and ℓ' with the same basepoint are *homotopic* (relative to the basepoint) if there is a continuous function $h : [0,1]\times[0,1] \to M$ such that $h(0,t) = \ell(t)$, $h(1,t) = \ell'(t)$, and $h(s,0) = h(s,1) = x$ for all $s,t \in [0,1]$. A loop is *contractible* if it is homotopic to the constant loop. The set of homotopy equivalence classes of loops based at x forms a group under concatenation, called the *fundamental group* and denoted $\pi_1(M,x)$. The identity element of the fundamental group is the homotopy class of contractible loops. Fundamental groups of the same connected space with different basepoints are isomorphic.

We define a **homotopy basis** to be any set of $2g$ loops whose homotopy classes generate the fundamental group $\pi_1(M,x)$; see Figure 1(b). Homotopy bases are a generalization of the *systems of loops* studied by Colin de Verdière and Lazarus [5]; a system of loops is a set of $2g$ simple loops (with a common basepoint) whose complement in the manifold is a topological disk. Every system of loops is a homotopy basis, but the converse is not true; homotopy bases can contain (self-) intersections that cannot be removed by homotopy. See Figure 1(c) for an example.

2.2 Cycles and Homology. Let R be an arbitrary ring. A *k-chain* is a formal linear combination of oriented k-simplices[1] with coefficients in the ring R. The set of k-chains forms a *chain group* $C_k(M;R)$ under addition. The *boundary operator* $\partial_k : C_k \to C_{k-1}$ is a linear map taking any oriented simplex to the chain consisting of its oriented boundary facets. A k-chain is called a *k-cycle* if its boundary is empty and a *k-boundary* if it is the boundary of a $(k+1)$-cycle; every k-boundary is a k-cycle. Let Z_k and B_k denote the subgroups of k-cycles and k-boundaries in C_k. The *kth homology group* $H_k(M;R)$ is the quotient group Z_k/B_k. If M is an oriented 2-manifold of genus g, then $H_1(M;R) \cong R^{2g}$.

More intuitively, a *homology cycle* is a formal linear combination of oriented cycles with coefficients in R.

[1] In *simplicial* homology, we assume that M is a simplicial complex and build chains from its component simplices. In *singular* homology, continuous maps from the canonical k-simplex to M play the role of 'k-simplices'. These two definitions yield isomorphic homology groups for manifolds.

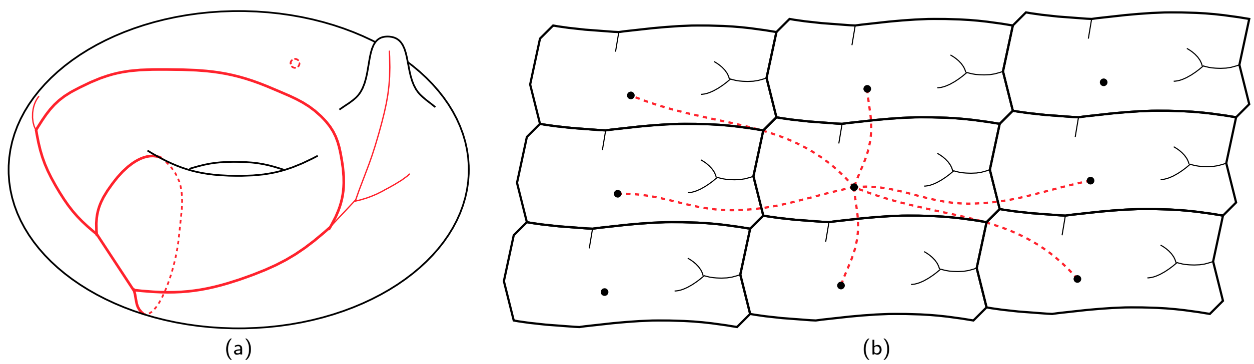

Figure 2. (a) The cut locus of a topological torus, after Sinclair and Tanaka [27]; the basepoint is hidden by the surface. The reduced cut locus has two branch points and three cut paths. (b) The reduced cut locus is the projection of a Voronoi diagram in the universal cover. Dashed paths are lifts of the shortest loops crossing the cut paths.

The identity element of the homology group is the equivalence class of *separating* cycles, that is, cycles whose removal disconnects the surface. Two homology cycles are in the same homology class 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 separating cycles. Throughout the paper, we let $[\ell]$ denote the homology class of a loop ℓ.

We define a **homology basis** for M to be any set of $2g$ cycles whose homology classes generate $H_1(M;R)$. Any homotopy basis is also a homology basis, but not vice versa, since the cycles in a homology basis generally do not have a common point; see Figure 1(d).

2.3 The Cut Locus and Its Friends. Most of our topological proofs assume that the manifold M is a smooth surface, or more precisely, a complete real-analytic Riemannian manifold [9]. This restriction allows us to exploit nice properties of geodesics without the additional technical machinery required for piecewise-linear [24, 23, 25] and combinatorial surfaces [5]. Our algorithmic results, on the other hand, are necessarily restricted to piecewise-linear and combinatorial surfaces.[2]

The *cut locus* $X = X(M,x)$ is the closure of the set of points in M with at least two shortest paths from a basepoint $x \in M$. For smooth surfaces (but *not* in general for piecewise-linear surfaces), the cut locus X is the embedding of a finite graph onto M, and X is a deformation retract of $M \setminus \{x\}$, where the retraction follows the shortest paths from x out to the cut locus. Symmetrically, $M \setminus X$ can be retracted to the basepoint x along the same shortest paths. Thus, the cut locus is a *cut graph* in the sense of Erickson and Har-Peled [8]; that is, $M \setminus X$ is a topological disk.

[2] We are deliberately avoiding the term "polyhedral surface", since this term has been used to describe both combinatorial surfaces, where paths are restricted to the 1-skeleton [8], and piecewise-linear surfaces, where paths can cross through the interior of faces [24].

We define the *reduced cut locus* $\Phi = \Phi(M,x)$ as the set of points in M with at least two *non-homotopic* shortest paths to x. The reduced cut locus is a subgraph of the cut locus, obtained by repeatedly removing vertices of degree 1. The reduced cut locus can also be defined as the projection to M of the Voronoi diagram (with respect to the shortest path metric) of the lifts of x in the universal cover $\tilde{M}$. The reduced cut locus is also a cut graph; in particular, any non-contractible loop in M must cross Φ at least once. We view Φ as an embedded graph in M with minimum degree 3, whose vertices we call *branch points* and whose edges we call *cut paths*. Euler's formula implies that if M is an oriented manifold of genus $g > 0$, then Φ has at most $4g-2$ branch points and $6g-3$ cut paths, with equality if every branch point has degree 3 [8, Lemma 4.2].

Let ϕ be a cut path in Φ, oriented arbitrarily. For any point c in the interior of ϕ, let $\sigma(c,\phi)$ denote the shortest non-contractible loop that contains c, oriented so that it crosses ϕ from left to right. This loop is the union of two non-homotopic shortest paths from x to c. For any two points $c, c' \in \phi$, the loops $\sigma(c,\phi)$ and $\sigma(c',\phi)$ are homotopy-equivalent. For each branch point b adjacent to ϕ, let $\sigma(b,\phi)$ denote the shortest non-contractible loop that contains b and is homotopic to $\sigma(c,\phi)$ for some $c \in \phi$. Finally, let $\sigma(\phi)$ denote the shortest loop of the form $\sigma(c,\phi)$ over all points c in the *closure* of ϕ; intuitively, this is the shortest non-contractible loop based at x that crosses ϕ. In general, $\sigma(\phi)$ may not be the shortest loop in its homotopy class.

Our proofs implicitly rely on the assumption that for any two distinct points c, c' on the reduced cut locus Φ, any two shortest crossing loops $\sigma(c,\phi)$ and $\sigma(c',\phi')$ intersect only at the common basepoint x. This assumption holds for any smooth Riemannian manifold [9, Corollary 2.111] but requires some additional machinery when the manifold is combinatorial [5] or piecewise linear [24]. We omit further details from this extended abstract.

3 The Greedy Homotopy Basis

3.1 Definition. Fix an oriented 2-manifold M and a basepoint $x \in M$. We inductively define the *greedy homotopy basis* $\gamma_1, \gamma_2, \ldots, \gamma_{2g}$ as follows:

For each i, γ_i is the shortest loop ℓ such that $M \setminus (\gamma_1 \cup \gamma_2 \cup \cdots \cup \gamma_{i-1} \cup \ell)$ is connected.

(In general, there may be several candidates for each greedy loop γ_i, possibly even in different homotopy classes; choose one arbitrarily.) Our main result is that the greedy homotopy basis is the shortest set of generators for the fundamental group $\pi_1(M, x)$. Since the greedy homotopy basis is a system of loops, and every system of loops is a homotopy basis, our result implies that the greedy loops also comprise the shortest system of loops.

3.2 No Free Lunch. Most problems for which greedy algorithms provide optimal solutions can be modeled as *matroids*. A matroid is a non-empty collection of subsets of a ground set X, called *independent sets*, that satisfies two axioms: (1) any subset of an independent set is an independent set, and (2) if A and B are independent sets with $|A| > |B|$, then there is an element $a \in A \setminus B$ such that $B \cup \{a\}$ is also an independent set. Maximal independent sets are called *bases* of a matroid; every basis of a matroid has exactly the same cardinality.

As we will see in Section 4, computing the shortest *homology* basis is a straightforward matroid optimization problem. Unfortunately, finding the optimal homotopy basis does not fit the matroid framework quite as readily. For example, we can define a *partial homotopy basis* to be any finite set of loops $L = \{\ell_1, \ell_2, \ldots, \ell_r\}$ that generates a (not necessarily proper) subgroup G of the fundamental group $\pi_1(M, x)$, such that no subset of L also generates G.

Theorem 3.1. *The set of partial homotopy bases of a 2-manifold is not necessarily a matroid.*

Proof: Let M be a two-holed torus, and let $\{a, b, c, d\}$ be a canonical generating set for its fundamental group $\pi_1(M) = \langle a, b, | \; ab\bar{a}\bar{b}cd\bar{c}\bar{d} \rangle$. Consider the sets $A = \{aba, a^2, b\}$ and $B = \{a, b\}$. A generates a free subgroup of rank 3 in $\pi_1(M)$, and B generates a free subgroup of rank 2. Thus, both A and B are partial homotopy bases. However, neither $\{a, b, aba\}$ nor $\{a, b, a^2\}$ is a partial homotopy basis, because in each case the newly added element is redundant.

By exploiting the fact that any free group has free subgroups of arbitrary large rank [28], we can also construct arbitrarily large partial homotopy bases for M. Specifically, for any non-negative integer n, the set $\{b, a^2, ab^2a, aba^2ba, \ldots, (ab)^n(ba)^n, (ab)^n a(ba)^n\}$ generates a free subgroup of rank $2n + 2$. □

Similarly, we can define a *partial system of loops* to be any finite set of loops $L = \{\ell_1, \ell_2, \ldots, \ell_r\}$ such that $M \setminus (\ell_1 \cup \cdots \cup \ell_r)$ is connected.

Theorem 3.2. *The set of partial systems of loops of a 2-manifold is not necessarily a matroid.*

Proof: Let M be a standard one-holed torus, and let (a, b) denote the standard torus knot that wraps 'around the hole' a times and 'through the hole' b times. This loop is simple if and only if $\gcd(a, b) = 1$, and two loops (a, b) and (c, d) form a system of loops if and only if $|ad - bc| = 1$. The sets $A = \{(0, 1), (1, 0)\}$ and $B = \{(3, 5)\}$ are both partial systems of loops, but neither $\{(3, 5), (0, 1)\}$ nor $\{(3, 5), (1, 0\})$ is a partial system of loops. □

3.3 Structure and Optimality.

Lemma 3.3. *Every loop in the greedy homotopy basis is has the form $\sigma(\phi)$ for some cut path ϕ.*

Proof: The greedy homotopy basis can be defined in a fashion similar to Dijkstra's single-source shortest path algorithm. Imagine a circular wavefront growing around the basepoint x. At any time t, the wavefront contains all points whose shortest path to x has length t. At various values of t, the wavefront meets itself; the meeting point c is (by definition) on the cut locus. If the two shortest paths to c form a non-contractible loop, then c lies on a cut path ϕ in the *reduced* cut locus, and the non-contractible loop is actually $\sigma(c, \phi) = \sigma(\phi)$. If $\sigma(\phi)$ can be added to the greedy basis without separating M, then $\sigma(\phi)$ is a greedy loop. Otherwise, no greedy loop is even in the same homotopy class as $\sigma(\phi)$.

On the other hand, the set of all loops $\sigma(\phi)$ redundantly generates the fundamental group, so the greedy construction must eventually halt with a (non-redundant) homotopy basis. □

This lemma allows us to redefine the greedy loop γ_i as the shortest loop **of the form $\boldsymbol{\sigma(\phi)}$** such that $M \setminus (\gamma_1 \cup \cdots \cup \gamma_{i-1} \cup \sigma(\phi))$ is connected.

Any set of loops that generates the fundamental group $\pi_1(M)$ also generates the homology group $H_1(M, \mathbb{Z})$. The homology class of any loop ℓ can be written as a formal linear combination (with integer coefficients) of the homology classes of the greedy loops: $[\ell] = \sum_{i=1}^{2g} \lambda_i [\gamma_i]$. We say that γ_i is a *greedy factor* of ℓ if the corresponding integer coefficient λ_i is non-zero. If we express the *homotopy* class of ℓ as a concatenation of greedy loops, then any greedy factor γ_i and its inverse $\overline{\gamma}_i$ appear a different number of times.

Lemma 3.4. *If γ is a greedy factor of $\sigma(\phi)$, then $|\gamma| \le |\sigma(\phi)|$.*

Proof: Let $\gamma_1, \dots, \gamma_i$ be all the greedy loops shorter than $\sigma(\phi)$. Define two embedded graphs $\Gamma_i = \gamma_1 \cup \dots \cup \gamma_i$ and $\Gamma'_i = \gamma_i \cup \sigma(\phi)$. By definition, $M \setminus \Gamma_i$ is connected. If $M \setminus \Gamma'_i$ is also connected, then $\gamma_{i+1} = \sigma(\phi)$. In this case, γ_{i+1} is the *only* greedy factor of $\sigma(\phi)$.

On the other hand, suppose $M \setminus \Gamma'_i$ is disconnected. Let C be a minimal separating subgraph of Γ'_i; this subgraph must contain $\sigma(\phi)$. If C contains no greedy loops, then $\sigma(\phi)$ is null-homologous and therefore has no greedy factors, and the lemma is satisfied vacuously. So suppose C contains the greedy loops $\gamma_{j_1}, \gamma_{j_2}, \dots, \gamma_{j_r}$. Then assuming these loops are oriented appropriately, the cycle $\sigma(\phi)(\gamma_{j_1}\gamma_{j_2}\cdots\gamma_{j_r})^{-1} = \sigma(\phi)\overline{\gamma}_{j_r}\cdots\overline{\gamma}_{j_2}\overline{\gamma}_{j_1}$ is separating, which implies that $\sigma(\phi)$ is in the same *homology* class as the cycle $\gamma_{j_1}\gamma_{j_2}\cdots\gamma_{j_r}$. Thus, the only greedy factors of $\sigma(\phi)$ are the greedy loops in C, none of which are longer than $\sigma(\phi)$. □

Lemma 3.5. *If γ is a greedy factor of an arbitrary loop ℓ, then $|\gamma| \le |\ell|$.*

Proof: If ℓ is a contractible cycle, then ℓ has no greedy factors, and the lemma is vacuously satisfied. Otherwise, ℓ must cross the reduced cut locus Φ at least once. Suppose ℓ crosses the cut paths $\phi_1, \phi_2, \dots, \phi_r$ in that order. Then ℓ is homotopy equivalent to the concatenation of loops $\sigma(\phi_1)\sigma(\phi_2)\cdots\sigma(\phi_r)$. See Figure 3. It follows that any greedy factor of ℓ is a greedy factor of at least one $\sigma(\phi_j)$ in this sequence. The definition of $\sigma(\phi_j)$ imply that $|\ell| \ge |\sigma(\phi_j)|$. The result now follows immediately from the previous lemma. □

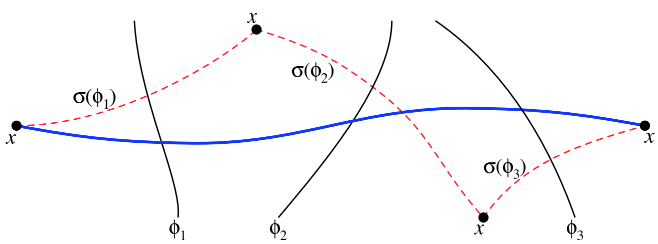

Figure 3. A loop crossing the reduced cut locus three times, as seen in the universal cover.

Lemma 3.6. *Let $\{\alpha_1, \alpha_2, \dots, \alpha_{2g}\}$ be any set of loops that generate the fundamental group $\pi_1(M, x)$. There is a permutation $\pi \in S_{2g}$ such for each i, $\gamma_{\pi(i)}$ is a greedy factor of α_i.*

Proof: Since the α_i's and γ_j's each generate the fundamental group, their homology classes also generate the homology group $H_1(M, \mathbb{Z}) \cong \mathbb{Z}^{2g}$. Any basis for the lattice $\mathbb{Z}^{2g}$ can be transformed into any other by a nonsingular linear transformation. Let $M \in \mathbb{Z}^{2g \times 2g}$ be the matrix representing the linear transformation that maps the greedy homology basis onto the α-basis. For all i and j, we have

$$[\alpha_i] = \sum_{j=1}^{2g} m_{ij}[\gamma_j].$$

Observe that γ_j is a greedy factor of α_i if and only if $m_{ij} \ne 0$. The matrix M has full rank; in particular its determinant is non-zero:

$$\det M = \sum_{\pi \in S_{2g}} (-1)^{\operatorname{sgn}(\pi)} \prod_{i=1}^{2g} m_{i,\pi(i)} \ne 0.$$

(In fact, $\det M = \pm 1$.) Thus, there is at least one permutation π such that $m_{i,\pi(i)} \ne 0$ for all i. □

Lemmas 3.5 and 3.6 immediately imply our main result.

Theorem 3.7. *For any 2-manifold M and any basepoint $x \in M$, the greedy homotopy basis is the shortest set of generators of $\pi_1(M, x)$.*

Finally, the following lemma provides a further analogy between greedy homotopy bases and Euclidean minimum spanning trees. Let P be a set of points in the plane. Two points $p, q \in P$ are joined by a *Gabriel edge* if the circle with diameter pq has no points of P in its interior, or equivalently, if pq is an edge in the Delaunay triangulation of P that intersects its dual Voronoi edge. It is well known that every edge in the Euclidean minimum spanning tree of P is a Gabriel edge [22]. Intuitively, the following lemma implies that every loop in the greedy homotopy basis is a 'Gabriel' loop.

Lemma 3.8. *Each loop γ_i in the greedy homotopy basis is the shortest loop in its homotopy class.*

Proof: Colin de Verdière and Lazarus [5] prove that for any system of loops, the shortest homotopically equivalent system of loops consists of simple loops that are shortest in their individual homotopy classes. The lemma follows from this result and the optimality of the greedy homotopy basis (see below), but we can give a simpler, self-contained proof.

Let ϕ^* denote the shortest loop in the homotopy class of $\sigma(\phi)$. (Intuitively, ϕ^* is a 'Delaunay' path.) Suppose $\phi^* \ne \sigma(\phi)$, or equivalently, that ϕ^* does not cross ϕ. Let $\phi_1, \phi_2, \dots, \phi_r$ be the sequence of cut paths crossed by ϕ^*. The proof of the Lemma 3.5 implies that $|\sigma(\phi_i)| \le |\phi^*| < |\sigma(\phi)|$ for each i. Thus, every greedy factor of $\sigma(\phi)$ is strictly shorter than $\sigma(\phi)$. It follows from Lemma 3.4 that $\sigma(\phi)$ is not a loop in the greedy homotopy basis. □

3.4 Computing the Greedy Loops from the Cut Locus. Recall that the reduced cut locus Φ is a deformation retract of the punctured manifold $M \setminus \{x\}$. Let $G = \{\phi_1, \ldots, \phi_r\}$ be any subgraph of Φ, and let $\Sigma = \{\sigma(\phi_1), \ldots, \sigma(\phi_r)\}$ be the corresponding set of shortest crossing loops. Because each loop $\sigma(\phi_i)$ intersects Φ in exactly one point, $\Phi \setminus G$ is a deformation retract of $M \setminus \Sigma$. In particular, $\Phi \setminus G$ is connected if and only if $M \setminus \Sigma$ is connected.

Thus, once we have computed the reduced cut locus Φ, the sequence of greedy loops can be constructed as follows. First, weight each cut path ϕ with the length of its shortest crossing loop $\sigma(\phi)$. Then consider the cut paths in order of increasing weight. For each cut path ϕ, if $\Phi \setminus \phi$ is connected, remove ϕ from Φ and declare $\sigma(\phi)$ to be the next greedy loop.

In addition to computing the greedy homotopy basis, this algorithm also reduces Φ to its *maximum* spanning tree. The algorithm follows a 'reverse greedy' strategy [1, Exercise 13.19] based on Tarjan's 'red rule': The shortest edge in any cycle is not in the maximum spanning tree [29]. Although this procedure could be implemented efficiently as stated with appropriate dynamic graph data structures [13], it is much faster to compute the maximum spanning tree directly. Φ has $O(g)$ edges and $O(g)$ vertices, so classical MST algorithms run in $O(g \log g)$ time [29]; Chazelle's deterministic algorithm [2] runs in $O(g\alpha(g))$ time; and the randomized algorithm of Klein, Karger, and Tarjan [17] runs in $O(g)$ expected time.

Finally, we observe that it is unnecessary to compute the *reduced* cut locus, since the maximum spanning tree of the unreduced cut locus must contain all the topologically redundant cut paths.

3.5 Combinatorial Surfaces. Following Colin de Verdière and Lazarus [5], a *combinatorial surface* $M = (\mathcal{M}, G)$ consists of an abstract 2-manifold $\mathcal{M}$ and a weighted graph G embedded on $\mathcal{M}$, such that every face of the embedding is a topological disk; the weight of each edge in G is the length of the corresponding path on $\mathcal{M}$; and we only allow paths that are subgraphs of G. In particular, the basepoint x must be a vertex of G. Combinatorial surfaces generalize surfaces of non-convex polyhedra where all paths are constrained to the 1-skeleton. The faces of a combinatorial surface are not necessarily flat—in fact, we don't care about the internal geometry of faces at all—and the lengths of edges need not correspond to Euclidean distances. Combinatorial surfaces are sometimes also called *maps* [7].

For any combinatorial surface $(\mathcal{M}, G)$, the *dual graph* G^* is the graph whose vertices are the faces of G, with an edge e^* between any pair of faces that share an edge e in G.

Erickson and Har-Peled [8] describe an algorithm to compute the shortest non-separating loop through a given basepoint in $O(n \log n)$ time. Naïvely applying this algorithm $O(g)$ times, we can compute the greedy homotopy basis of a combinatorial manifold in $O(gn \log n)$ time. This can be improved to $O(n \log n)$ by being more careful, but we can derive an even simpler algorithm with the same running time using a special case of Eppstein's *tree-cotree decomposition* [7].

Let T be the tree of shortest paths in G from the basepoint x to every other vertex of G. For each edge $e \in G \setminus T$, let $\sigma(e)$ be the shortest loop that contains e; this loop consists of the two shortest paths from x to the endpoints of e plus the edge e itself. The graph $(G \setminus T)^*$, consisting of edges of the dual graph G^* that do not correspond to edges of T, is essentially the cut locus of M with respect to the basepoint x. Let T^* be the maximum spanning tree of $(G \setminus T)^*$, where the weight of any dual edge e^* is $|\sigma(e)|$. The greedy homotopy basis consists of all loops $\sigma(e)$, where e is an edge of G that is neither in T nor crossed by T^*. Euler's formula implies that there are exactly $2g$ such loops, and it is easy to verify that they do actually form a homotopy basis [7].

Theorem 3.9. *Given any combinatorial surface M and any basepoint $x \in M$, we can compute the shortest system of loops with basepoint x in $O(n \log n)$ time, or in $O(n)$ time if M has genus $O(n^{1-\varepsilon})$ for some $\varepsilon > 0$.*

Proof: We can compute T in $O(n \log n)$ time using Dijkstra's algorithm [29], or in $O(n)$ time if $g = O(n^{1-\varepsilon})$ for some $\varepsilon > 0$ using an algorithm of Henzinger *et al.* [12].[3] We can easily compute the length of $\sigma(e)$ for each edge $e \in G \setminus T$ in $O(n)$ time from the shortest path tree. The maximum spanning tree T^* can be computed directly in $O(n \log n)$ time with any classical algorithm. Alternately, in $O(n)$ time, we can compute an abstract reduced cut locus Φ from X by repeatedly removing any degree-1 vertices and contracting any paths of degree-2 vertices; we can then compute the maximum spanning tree of Φ in $O(g \log g)$ time as described in the previous section. □

For any two basepoints $x, y \in M$, the groups $\pi_1(M, x)$ and $\pi_1(M, y)$ are isomorphic. Thus, it is natural to ask for the shortest system of loops over all possible base points. For combinatorial manifolds, we can simply try all n basepoints.

Corollary 3.10. *Given any combinatorial surface M, we can compute the shortest system of loops for M in $O(n^2 \log n)$ time, or in $O(n^2)$ time if M has genus $O(n^{1-\varepsilon})$ for some $\varepsilon > 0$.*

[3] The shortest-path algorithm of Henzinger *et al.* [12] can be modified to run in $O(n)$ time for any minor-closed family of graphs with separators of size $O(n^{1-\varepsilon})$. The family of graphs of genus g is closed under taking minors, and any graph of genus-g graph has a separator of size $O(\sqrt{ng})$ [7, 10, 18].

3.6 Piecewise-Linear Surfaces. Now suppose M is a piecewise-linear manifold—for example, a non-convex polyhedron in $\mathbb{R}^3$—and we are interested in arbitrary loops on the surface of M. In this case, for any basepoint x, we can compute the exact cut locus of x, often called the *geodesic Voronoi diagram* of x in this context, in $O(n^2)$ time using the 'continuous Dijkstra' algorithm of Chen and Han [3]; see also [16, 15, 24]. With the cut locus in hand, we can compute the greedy homotopy basis as described in Section 3.4.

Theorem 3.11. *Given any piecewise-linear manifold M in $\mathbb{R}^3$ and any basepoint $x \in M$, we can compute the shortest system of loops for M based at x in $O(n^2)$ time.*

Alternately, one can express the expanding distance wave as a differential equation and compute a numerically accurate solution using any number of efficient numerical methods [19, 20, 30, 25, 31]. For example, Kimmel and Sethian's popular fast marching method [20] numerically approximates geodesic distances on a piecewise-linear surface with fat[4] triangular facets in $O(n \log n)$ time. Using this method, we can compute a numerical approximation to the shortest system of loops in $O(n \log n)$ time.

Computing the shortest homotopy basis over all possible basepoints is considerably harder for piecewise-linear manifolds than for combinatorial manifolds, because the number of possible basepoints is no longer finite. In fact, even the simpler problem of finding the shortest non-separating cycle in a piecewise-linear manifold appears to be open.

4 The Greedy Homology Basis

Homology groups for compact oriented 2-manifolds without boundary have an extremely simple structure: If the manifold M has genus g, then $H_1(M, R) \cong R^{2g}$. If the coefficient ring R is a field, such as $\mathbb{Z}_2$ or $\mathbb{Q}$, then $H_1(M; R)$ is a vector space, and thus the sets of independent homology classes form a matroid. It immediately follows that the shortest homology basis over any field has the following greedy characterization: For all i, γ_i is the shortest simple cycle whose homology class is not a linear combination (over the coefficient field) of the homology classes $[\gamma_1], [\gamma_2], \ldots, [\gamma_{i-1}]$.

In this section, we describe an algorithm that efficiently constructs this greedy homology basis, over any fixed field of coefficients, for a given combinatorial surface. Our algorithm is essentially an adaptation of a greedy algorithm of Horton [14] for computing the shortest cycle basis—or in our terminology, the shortest $\mathbb{Z}_2$-homology basis [4]—for a weighted undirected graph. To simplify the exposition, we describe the algorithm for $\mathbb{Z}_2$ coefficients only; only minimal changes are required for other coefficient fields.[5]

A simple cycle ℓ is *tight* if it contains a shortest path between every pair of points in ℓ.

Lemma 4.1. *Every cycle in the shortest homology basis is tight.*

Proof: Let $\ell_1, \ell_2, \ldots, \ell_{2g}$ be an arbitrary homology basis. If ℓ_1 is not a simple cycle, then it can be decomposed into two smaller cycles ℓ_1' and ℓ_1''. At least one of these two cycles, say ℓ_1', is not spanned by the other basis cycle $\ell_2, \ldots, \ell_{2g}$; otherwise, ℓ_1 would not be in the basis. Thus, $\ell_1', \ell_2, \ldots, \ell_{2g}$ is a shorter homology basis.

Now suppose ℓ_1 is a simple cycle but is not tight. Let x and y be arbitrary points on ℓ_1 such that ℓ_1 does not contain a shortest path from x to y; these points break ℓ_1 into two paths α and β from y to x. Let σ be a shortest path from x to y. Finally, define two cycles $\ell_1' = \alpha\sigma$ and $\ell_1'' = \beta\sigma$. At least one of these cycles, say ℓ_1', is not spanned by the other basis loops $\ell_2, \ldots, \ell_{2g}$; otherwise, $\ell_1 \simeq \ell_1'\overline{\ell_1''}$ would not be in the basis. Thus, $\ell_1', \ell_2, \ldots, \ell_{2g}$ is a shorter homology basis. □

To compute the greedy homology basis, we will leverage our greedy homotopy algorithm, but we need to modify our notation since we no longer have a fixed basepoint. Let $\Phi(x)$ be the reduced cut locus of M with respect to any point $x \in M$; and let $\sigma(x, \phi)$ denote the shortest loop that contains x and crosses the cut path ϕ in $\Phi(x)$, possibly at one of the endpoints of ϕ.

Lemma 4.2. *Every cycle in the shortest homology basis has the form $\sigma(x, \phi)$ for some cut path ϕ in $\Phi(x)$.*

Proof: Let $\sigma(x, c)$ denote the shortest loop that contains the points $x \in M$ and $c \in \Phi(x)$. Every tight loop, and thus every loop in the greedy homology basis, has the form $\sigma(x, c)$ for some points x and c. Moreover, if c and c' lie on the same cut path in $\Phi(x)$, the greedy homology basis cannot contain the longer of the two loops $\sigma(x, c)$ and $\sigma(x, c')$. □

Now we describe an algorithm to compute the greedy homology basis for a given combinatorial surface. For some arbitrary vertex x, compute the reduced cut locus $\Phi(x)$ and the greedy *homotopy* basis $\gamma_1(x), \gamma_2(x) \ldots, \gamma_{2g}(x)$ with respect to x. Like any homotopy basis, these loops also form a homology basis. For each edge e, compute a vector of $2g$ bits

[4] Sethian and Vladimirsky [26] discuss the numerical instability of fast marching methods caused by extremely obtuse triangles.

[5] Actually, it is unclear whether *any* changes are required; perhaps the shortest homology basis is the same for every coefficient ring! We leave this as an open question.

representing the homology class $[\sigma(x,e)]$ in the basis $\{[\gamma_1(x)],\ldots,[\gamma_{2g}(x)]\}$. There are only $O(g)$ different homology vectors to store, one for each cut path in $\Phi(x)$ plus the zero vector for loops that do not cross $\Phi(x)$. With this information, we can compute the homology class of any cycle ℓ simply by adding (modulo 2) the bit vectors associated with the edges of ℓ.

Next, for each basepoint y, compute the reduced cut locus $\Phi(y)$ and every loop of the form $\sigma(y,\phi)$. Since $\Phi(y)$ has $O(g)$ cut paths, there are only $O(g)$ such loops. We can compute the homology vectors for all of these in $O(gn)$ time by accumulating the bit vectors along paths in the shortest path tree rooted at y. Altogether, over all basepoints y, we compute $O(gn)$ candidate loops and their homology classes in time $O(n^2\log n + gn^2)$, or $O(gn^2)$ if g is sufficiently small. Note that any tight loop ℓ will appear in this list of candidates once for every vertex of ℓ.

Finally, we consider the candidate loops $\sigma_1,\sigma_2,\ldots$ in order from shortest to longest, and for each σ_j, we determine whether the homology class of σ_j is a linear combination of shorter greedy homology classes. We can test linear independence of the homology vectors in $O(g^2)$ time by Gaussian elimination. Thus, the total time spent scanning the list of candidate loops is $O(ng^3)$.

Theorem 4.3. *For any combinatorial manifold M, we can compute the shortest set of cycles that generates $H_1(M,\mathbb{Z}_2)$ in time $O(n^2\log n + n^2g + ng^3)$, or in time $O(n^2g + ng^3)$ if $g = O(n^{1-\varepsilon})$ for some $\varepsilon > 0$.*

5 Open Problems

By definition, removing a cut graph G from an oriented 2-manifold M leaves a single topological disk, which is often called a *polygonal schema*. Each edge of G appears twice on the boundary of the polygonal schema, and we can obtain the original manifold M by identifying every corresponding pair of boundary edges in opposite orientations. The pattern of corresponding boundary edges defines a combinatorial signature for the cut graph, which we call its *gluing pattern*. For example, the so-called *canonical* systems of loops constructed by Lazarus *et al.* [21] have the gluing pattern $a_1b_1\bar{a}_1\bar{b}_1a_2b_2\bar{a}_2\bar{b}_2\cdots a_gb_g\bar{a}_g\bar{b}_g$. Given cut graphs G and G' on two combinatorial manifolds M and M', where the cut graphs have the same gluing pattern, it is relatively straightforward to construct a homeomorphism from M to M'. How quickly can we compute the shortest cut graph, with a given (or possibly fixed) gluing pattern?

How quickly can we compute the shortest set of generators for homology with *integer* coefficients? As far as we know, this problem is open even for weighted undirected graph. Our matroid argument fails because the homology group $H_1(M,\mathbb{Z}) \cong \mathbb{Z}^{2g}$ is not a vector space, but merely an integer lattice. Sets of linearly independent integer vectors still form a matroid, so we can use the same greedy algorithm to compute the shortest set of $2g$ homologically independent loops, but these loops may not form a basis. We pessimistically conjecture that this problem is NP-hard.

Finally, how difficult is it to compute other types of optimal graphs on surfaces, such as pants decompositions or one-vertex triangulations?

Acknowledgments. Thanks to Francis Lazarus for his detailed comments on an earlier version of our paper and for pointing us to work on cycle bases of graphs. Thanks also to Stephen Gortler, Rob Ghrist, and Sariel Har-Peled for helpful discussions.

References

[1] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, Englewood Cliffs, NJ, 1993.

[2] B. Chazelle. A minimum spanning tree algorithm with inverse-Ackermann type complexity. *J. ACM* 47:1028–1047, 2000.

[3] J. Chen and Y. Han. Shortest paths on a polyhedron. *Internat. J. Comput. Geom. Appl.* 6:127–144, 1996.

[4] W. Chen. On vector spaces associated with a graph. *SIAM J. Appl. Math.* 20(3):526–529, 1971.

[5] È. Colin de Verdière and F. Lazarus. Optimal systems of loops on an orientable surface. *Proc. 43rd Annu. IEEE Sympos. Found. Comput. Sci.*, 627–636, 2002. Full version to appear in *Discrete Comput. Geom.*

[6] È. 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.

[7] D. Eppstein. Dynamic generators of topologically embedded graphs. *Proc. 15th ACM-SIAM Sympos. Discrete Algorithms*, 599–608, 2004.

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

[9] S. Gallot, D. Hulin, and J. Lafontaine. *Riemannian Geometry*. Springer-Verlag, 1987.

[10] J. R. Gilbert, J. P. Hutchinson, and R. E. Tarjan. A separation theorem for graphs of bounded genus. *J. Algorithms* 5:391–407, 1984.

[11] A. Hatcher. *Algebraic Topology*. Cambridge University Press, 2001. ⟨http://www.math.cornell.edu/~hatcher/⟩.

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

[13] J. Holm, K. de Lichtenberg, and M. Thorup. Polylogarithmic deterministic fully-dynamic algorithms for connectivity, minimum spanning tree, 2-edge, and biconnectivity. *J. ACM* 48(4):723–760, 2001.

[14] J. D. Horton. A polynomial-time algorithm to find the shortest cycle basis of a graph. *SIAM J. Comput.* 16(2):358–366, 1987.

[15] B. Kaneva and J. O'Rourke. An implementation of Chen & Han's shortest paths algorithm. *Proc. 12th Canadian Conf. Comput. Geom.*, 139–146, 2000. ⟨http://www.cs.unb.ca/conf/cccg/eProceedings/⟩.

[16] S. Kapoor. Efficient computation of geodesic shortest paths. *Proc. 32nd Annu. ACM Sympos. Theory Comput.*, 770–779, 1999.

[17] D. Karger, P. N. Klein, and R. E. Tarjan. A randomized linear-time algorithm to find minimum spanning trees. *J. ACM* 42:321–328, 1995.

[18] J. R. Keller. Spectral partitioning, eigenvalue bounds, and circle packings or graphs of bounded genus. *Proc. 36th Annu. ACM Sympos. Theory Comput.*, 455–464, 2004.

[19] R. Kimmel, A. Amir, and A. M. Bruckstein. Finding shortest paths on surfaces using level sets propogation. *IEEE Trans. PAMI* 17:635–640, 1995.

[20] R. Kimmel and J. A. Sethian. Computing geodesic paths on manifolds. *Proc. Nat. Acad. Sci.* 95(15):8431–8435, 1998.

[21] 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.

[22] D. W. Matula and R. R. Sokal. Properties of Gabriel graphs relevant to geographic variation research and clustering of points in the plane. *Geogr. Anal.* 12(3):205–222, 1980.

[23] E. Miller and I. Pak. Metric combinatorics of convex polyhedra: Cut loci and nonoverlapping unfoldings. Preprint, December 2003.

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

[25] K. Polthier and M. Schmies. Straightest geodesics on polyhedra. *Mathematical Visualization*, 135–150, 1998. Springer Verlag.

[26] J. A. Sethian and A. Vladimirsky. Fast methods for the Eikonal and related Hamilton-Jacobi equations on unstructured meshes. *Proc. Nat. Acad. Sci.* 97(11):5699–5703, 2000.

[27] R. Sinclair and M. Tanaka. Loki: Software for computing cut loci. *Experiment. Math.* 11(1):1–25, 2002.

[28] J. Stillwell. *Classical Topology and Combinatorial Group Theory*, 2nd edition. Graduate Texts in Mathematics 72. Springer-Verlag, 1993.

[29] R. E. Tarjan. *Data Structures and Network Algorithms.* CBMS-NSF Regional Conference Series in Applied Mathematics 44. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1983.

[30] J. N. Tsitsiklis. Efficient algorithms for globally optimal trajectories. *IEEE Trans. Automatic Control* 40:1528–1538, 1995.

[31] G. Zigelman, R. Kimmel, and N. Kiryati. Texture mapping using surface flattening via multidimensional scaling. *IEEE Trans. Visual. Comput. Graph.* 8:198–207, 2001.