

Original citation:

Chang, Hsien-Chih, Erickson, Jeff, Letscher, David, de Mesmay, Arnaud, Schleimer, Saul, Sedgwick, Eric, Thurston, Dylan and Tillmann, Stephan (2017) Tightening curves on surfaces via local moves. In: ACM-SIAM Symposium on Discrete Algorithms, New Orleans, USA, 7-10 Jan 2018. Published in: SODA 18 : Proceedings of ACM-SIAM Symposium on Discrete Algorithms

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/94044>

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:

Accepted for publication for SODA 18 : Proceedings of ACM-SIAM Symposium on Discrete Algorithms published by the Society for Industrial and Applied Mathematics (SIAM). Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

A note on versions:

The version presented here may differ from the published version or, version of record, if you wish to cite this item you are advised to consult the publisher's version. Please see the 'permanent WRAP URL' above for details on accessing the published version and note that access may require a subscription.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

Tightening Curves on Surfaces via Local Moves

Hsien-Chih Chang*

Jeff Erickson[†]

David Letscher[‡]

Arnaud de Mesmay[‡]

Saul Schleimer[§]

Eric Sedgwick[¶]

Dylan Thurston^{||}

Stephan Tillmann^{**}

Abstract

We prove new upper and lower bounds on the number of homotopy moves required to tighten a closed curve on a compact orientable surface (with or without boundary) as much as possible. First, we prove that $\Omega(n^2)$ moves are required in the worst case to tighten a contractible closed curve on a surface with non-positive Euler characteristic, where n is the number of self-intersection points. Results of Hass and Scott imply a matching $O(n^2)$ upper bound for contractible curves on orientable surfaces. Second, we prove that any closed curve on any orientable surface can be tightened as much as possible using at most $O(n^4)$ homotopy moves. Except for a few special cases, only naïve exponential upper bounds were previously known for this problem.

1 Introduction

A closed curve γ on a surface Σ can be *tightened* via homotopy to another closed curve γ' with the minimum number of self-intersections within its homotopy class. Classical results in combinatorial topology [1, 2, 43] imply that any generic curve γ can be tightened using a finite sequence of the following local transformations, which we call *homotopy moves*:

- $1 \leftrightarrow 0$: Remove/add an *empty monogon*.
- $2 \leftrightarrow 0$: Remove/add an *empty bigon*.
- $3 \rightarrow 3$: Flip a *triangle*; equivalently, move a strand across a self-intersection.

Each homotopy move is performed inside an open disk embedded in Σ , meeting γ as shown in Figure 1.1. These moves are analogues of the *Reidemeister moves* performed on planar diagrams of knots and links [2, 43].

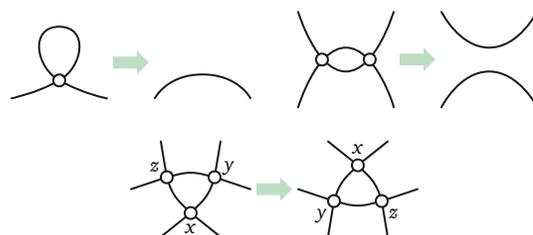


Figure 1.1. Homotopy moves $1 \rightarrow 0$, $2 \rightarrow 0$, and $3 \rightarrow 3$.

Previous work. A proof that $O(n^2)$ moves suffice to tighten a generic closed curve in the plane, with n self-intersections, is implicit in Steinitz’s proof that every three-connected planar graph is the one-skeleton of a convex polyhedron [45, 46]. Specifically, Steinitz proved that any non-simple closed curve with no empty *monogons* contains a minimal *bigon* which can be emptied with $3 \rightarrow 3$ moves and then removed by a $2 \rightarrow 0$ move. This quadratic bound has been reproved and generalized by several other authors [23, 24, 30, 40, 48, 49]. Chang and Erickson recently improved the upper bound to $O(n^{3/2})$ and proved that such bound is optimal in the worst case [11].

Hass and Scott [32] proved that on any orientable surface, every non-simple closed curve that is homotopic to a simple closed curve has either an empty monogon or a minimal bigon. Thus, any such curve can be simplified using $O(n^2)$ moves following Steinitz’s technique. Chang and Erickson proved a matching $\Omega(n^2)$ lower bound for *non-contractible* curves on surfaces with positive genus and no boundary [11]. de Graaf and Schrijver [28] proved that arbitrary curves on the annulus can be tightened using at most $O(n^2)$ moves.

However, no polynomial upper bound is known in the most general case when the surface and the curve are both unrestricted. Hass and Scott [33] proved that any closed curve on any surface can be simplified using a *finite* number of homotopy moves that never increase the number of self-intersections. The same result was later extended to multicurves by de Graaf and Schrijver [28], and proven again using a combinatorial algorithm by Paterson [41]. This monotonicity result, together with asymptotic bounds by Bender and Canfield [7] on the number of distinct (rooted) 4-regular maps with n vertices and genus g , immediately implies an upper bound of the form $n^{O(g)} 2^{O(n)}$; this is the best upper bound previously known. Chang and Erickson conjectured that an arbitrary

*Department of Computer Science, University of Illinois at Urbana-Champaign, USA. Work by these authors was partially supported by NSF grant CCF-1408763.

[†]Department of Computer Science, Saint Louis University, USA. Work by this author was partially supported by NSF grant IIS-1319944.

[‡]Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France. Work by this author was partially supported by the French project ANR-16-CE40-0009-01 (GATO).

[§]Warwick Mathematics Institute, University of Warwick, Coventry, UK.

[¶]School of Computing, DePaul University, Chicago, USA.

^{||}Department of Mathematics, Indiana University, Bloomington, USA.

^{**}School of Mathematics and Statistics, The University of Sydney, Australia. Work by this author was partially supported under the Australian Research Council’s Discovery funding scheme (project number DP160104502).

(multi-)curve on any surface can be tightened using at most a quadratic number of moves [11].

For a more extended survey of the problem and its close relation to *electrical transformations*, see Chang and Erickson [11].

New results. First, we construct an infinite family of contractible curves on the annulus that require at least $\Omega(n^2)$ moves to tighten.

Theorem 1.1. *In the annulus, there are contractible curves with n self-intersections that require $\Omega(n^2)$ homotopy moves to tighten to a trivial curve.*

Our new lower bound generalizes to any surface that has the annulus as a covering space—that is, any surface except for the sphere, the disk, or the projective plane; see Theorem 3.1. Thus, this result extends and generalizes Chang and Erickson’s $\Omega(n^2)$ lower bound for non-contractible curves on the torus [11]. This result is also applied in an upcoming companion paper [10] to derive a quadratic lower bound on the number of facial electrical transformations to reduce a 2-terminal plane graph in the worst case.

Next, we turn our attention to upper bounds. We separately consider orientable surfaces with and without boundary.

Theorem 1.2. *On an oriented surface of genus g with $b > 0$ boundary components, a closed curve with n self-intersections can be tightened using at most $O((g + b)n^3)$ homotopy moves.*

Theorem 1.3. *On an oriented surface without boundary, a closed curve with n self-intersections can be tightened using at most $O(n^4)$ homotopy moves.*

The latter result is surprisingly more complex and subtle, with multiple components and tools drawn from discrete and computational topology. We consider surfaces with boundary first, not only because we obtain a stronger bound (at least when the genus and number of boundary components are small), but also because the proof is simpler and provides important intuition for the more difficult proof of Theorem 1.3. We emphasize that the bound in Theorem 1.3 is independent of the genus of the surface.

Our main technical contribution is to extend Steinitz’s bigon reduction technique to *singular bigons*—bigons that wrap around the surface and overlap themselves but nevertheless have well-defined disjoint bounding paths—whose existence is guaranteed by a theorem of Hass and Scott [32, Theorem 2.7]. (A formal definition of the singular bigon can be found in Section 4.1.) To work with singular bigons, it is conceptually easier to look at a lift of the bigon in the universal cover. Unlike the case when the

bigon is embedded, moving the two bounding paths of the bigon now also moves all their *translates* in the universal cover, which potentially changes the structure inside the lifted bigon. We overcome this difficulty by carefully subdividing the homotopy into phases, each performed inside a subset of the universal cover that maps injectively onto the original surface.

Our proof of Theorem 1.3 also uses a discrete analog of the classical *isoperimetric inequality* in the hyperbolic plane to bound the number of vertices inside the lifted bigon (area) in terms of the number of vertices on its boundary (perimeter). To make the presentation self-contained, we provide an elementary proof of this inequality using the combinatorial Gauss-Bonnet theorem [6, 22, 37, 42].

Related work. The results of Hass and Scott [33] and de Graaf and Schrijver [28] both use discrete variants of curve-shortening flow. Grayson [29] and Angenent [3] provide similar results using differential curvature flow when the curves and surfaces are well-behaved. None of these algorithms provide any bound on the number of homotopy moves performed as a function of the number of self-intersections.

The *geometric intersection number* of a closed curve γ on a surface is the number of self-intersections of a tightening of γ . Several methods for characterizing and computing geometric intersection numbers are known [12, 13, 14, 27, 36]; however, none of these earlier results offers a full complexity analysis. Arettines [4] described a polynomial-time algorithm to compute geometric intersection number of a curve on an orientable surface with boundary, starting from the reduced crossing sequence of the curve with a system of arcs (defined in Section 4.2). Despré and Lazarus [17] described the first fully-analyzed polynomial-time algorithm to compute the geometric intersection number of arbitrary closed curves on arbitrary orientable surfaces. Both of these algorithms follow a high-level strategy similar to ours, based on Hass and Scott’s results about singular bigons, but neither algorithm computes an explicit sequence of homotopy moves. Instead, Arettines removes singular bigons by permuting their intersections along each arc, and Despré and Lazarus remove singular bigons by directly *smoothing* their endpoints. Further references can be found in Despré and Lazarus [17].

2 Background

Throughout the paper we assume the reader is familiar with some of the fundamentals of the combinatorial topology of surfaces. We refer the readers to Massey [39], Stillwell [47], and Giblin [26] for comprehensive introductions to the topic.

Surfaces and curves. A *surface* Σ is a 2-dimensional manifold. In this paper, all surfaces are assumed to be compact, connected, and oriented unless stated otherwise.

The authors have placed this paper in the public domain.

Formally, a **closed curve** in a surface Σ is a continuous map $\gamma: S^1 \rightarrow \Sigma$, and a **path** in Σ is a continuous map $\eta: [0, 1] \rightarrow \Sigma$. Depending on the context, we sometimes abuse the terminology and refer to a continuous map $\eta: (0, 1) \rightarrow \Sigma$ as a **path** as well. A **curve** is either a closed curve or a path; its parametrization equips the curve with an orientation. We consider only **generic** curves, which are injective except at a finite number of self-intersections, each of which is a transverse double point; thus the double points avoid the boundary of Σ . A curve is **simple** if it is injective. A **subpath** of a curve γ is the restriction of γ to an interval; again, a subpath is simple if the restriction is injective. Unless specified otherwise, we do not distinguish between γ and its image.

The image of any non-simple closed curve γ has a natural structure as a 4-regular map, whose **vertices** are the self-intersections of γ , **edges** are maximal subpaths between vertices, and **faces** are components of $\Sigma \setminus \gamma$. Every vertex x of γ has four **corners** adjacent to it; these are the four components of $D_x \setminus \gamma$ where D_x is a small disk neighborhood of x . Two curves γ and γ' are **isomorphic** if their images define combinatorially equivalent maps; we will not distinguish between isomorphic curves.

Monogons and Bigons. A **monogon** for a closed curve γ is a subpath of γ that begins and ends at some vertex x , intersects itself only at x , and bounds a disk in Σ containing exactly one of the four corners at x . A **bigon** for γ consists of two simple interior-disjoint subpaths of γ , sharing endpoints x and y , that together bound a disk in Σ containing exactly one corner at x and one at y . Since each subpath is simple, the vertices x and y are distinct. A monogon or bigon is **empty** if it does not intersect the rest of γ . Notice that a $1 \rightarrow 0$ move is applied to an empty monogon, and a $2 \rightarrow 0$ move is applied to an empty bigon. A bigon is **minimal** if the disk it bounds does not contain a smaller bigon. Note that a minimal bigon may share boundary paths with a larger bigon containing it.

Winding numbers. Let γ be a generic closed curve in the plane, and let p be any point not in the image of γ . Let ρ be any ray from p to infinity that intersects γ transversely. The **winding number** $wind(\gamma, p)$ is the number of times γ crosses ρ from right to left, minus the number of times γ crosses ρ from left to right. The winding number does not depend on the particular choice of ray ρ . All points in the same face of γ have the same winding number; the winding numbers of two adjacent faces differ by 1, with the higher winding number on the left side of the edge. If p lies on the curve γ , we define $wind(\gamma, p)$ to be the average of the winding numbers of the faces incident to p with appropriate multiplicity—two faces if p lies on an edge, four if p is a vertex. The winding number of a vertex is always an integer.

Smoothing. Suppose that γ is a generic closed curve and x is a vertex of γ . Let D_x be a small disk neighborhood of x . Then we may **smooth** the curve γ at x by: removing

$\gamma \cap D_x$ from γ , adding in two components of $\partial D_x \setminus \gamma$, and obtaining another 4-regular map. There are two possible smoothings. One results in another closed curve, with the orientation of one subpath of γ reversed; the other breaks γ into a pair of closed curves, each retaining its original orientation. In the latter smoothing, let γ_x^+ and γ_x^- respectively denote the closed curve locally to the left and to the right of x , as shown in Figure 2.1. For any vertex x and any other point o , we have $wind(\gamma, o) = wind(\gamma_x^+, o) + wind(\gamma_x^-, o)$.

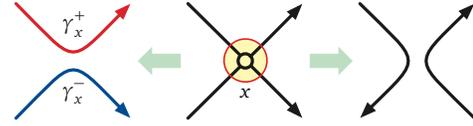


Figure 2.1. Smoothing a vertex. The left smoothing preserves orientation; the right smoothing preserves connectivity.

Homotopy. A **homotopy** between two closed curves γ and γ' on the same surface Σ is a continuous deformation from one curve to the other. Formally this is a continuous map $H: S^1 \times [0, 1] \rightarrow \Sigma$ such that $H(\cdot, 0) = \gamma$ and $H(\cdot, 1) = \gamma'$. Similarly, a **homotopy** between two paths η and η' is a continuous deformation that keeps the endpoints fixed. Formally this is a continuous map $H: [0, 1] \times [0, 1] \rightarrow \Sigma$ such that $H(\cdot, 0) = \eta$, and $H(\cdot, 1) = \eta'$, and both $H(0, \cdot)$ and $H(1, \cdot)$ are constant functions. Two curves are **homotopic**, or in the same **homotopy class**, if there is a homotopy from one to the other. A closed curve γ is **contractible** if it is homotopic to a constant curve; intuitively, this says that γ can be continuously contracted to a single point. Classical topological arguments [1, 2, 43] imply that two curves are homotopic if and only if one can be transformed into the other by a finite sequence of homotopy moves. As any contractible curve γ can be made simple through homotopy moves, we sometimes refer to the tightening process of γ as **simplifying** γ .

Covering spaces and lifts. A surface $\hat{\Sigma}$ is a **covering space** of another surface Σ if there is a **covering map** from $\hat{\Sigma}$ to Σ ; that is, a continuous map $\pi: \hat{\Sigma} \rightarrow \Sigma$ so that each point x on Σ has a neighborhood $U \subset \Sigma$ so that $\pi^{-1}(U)$ is a union of disjoint open sets $U_1 \cup U_2 \cup \dots$, and, for any i , the restriction $\pi|_{U_i}: U_i \rightarrow U$ is a homeomorphism. A **lift** of a curve α in Σ to the covering space $\hat{\Sigma}$ is a curve $\hat{\alpha}$ in $\hat{\Sigma}$ such that $\alpha = \pi \circ \hat{\alpha}$. When γ is a closed curve, we sometimes abuse notation and define a **lift** of γ to $\hat{\Sigma}$ to be an infinite path $\hat{\gamma}$ in $\hat{\Sigma}$ such that $\gamma(t \bmod 1) = \pi \circ \hat{\gamma}(t)$. A **translate** of a lift $\hat{\alpha}$ is any other lift of α to the same covering space; equivalently, two paths $\hat{\alpha}, \hat{\beta}: [0, 1] \rightarrow \hat{\Sigma}$ are translates of each other if and only if $\pi \circ \hat{\alpha} = \pi \circ \hat{\beta}$.

3 Quadratic Lower Bound

We now prove a quadratic lower bound on the worst-case number of homotopy moves required to tighten closed

curves in the annulus; we extend this lower bound to more complex surfaces in Section 3.3.

Rather than considering the standard annulus $S^1 \times [0, 1]$, it will be more convenient to work in the punctured plane $\mathbb{R}^2 \setminus \{o\}$, which is homeomorphic to the open annulus $S^1 \times (0, 1)$; here o is an arbitrary point, which we call the **origin**. Winding numbers for closed curves in the punctured plane are defined by considering them as curves in the plane. Two closed curves in the punctured plane are homotopic if and only if they have the same winding number around the origin: $wind(\gamma, o) = wind(\gamma', o)$ [34]. We refer to any closed curve in the punctured plane as an *annular curve*.

For any homotopy in the punctured plane, homotopy moves across the face containing o are forbidden. This makes the quadratic lower bound possible; without this restriction, any planar curve can be simplified using at most $O(n^{3/2})$ moves [11].

3.1 Traces and Types

To simplify the presentation, we identify the vertices before and after a $3 \rightarrow 3$ move as indicated in Figure 1.1. Each $3 \rightarrow 3$ move involves three subpaths of γ , which intersect in three vertices; intuitively, each of these vertices moves continuously across the opposite subpath. Thus, in any homotopy from one curve γ to another curve γ' , each vertex of the evolving curve either starts as a vertex of γ or is created by a $0 \rightarrow 1$ or $0 \rightarrow 2$ move, moves continuously through a finite sequence of $3 \rightarrow 3$ moves, and either ends as a vertex of γ' or is destroyed by a $1 \rightarrow 0$ or $2 \rightarrow 0$ move.

Let H be a homotopy that transforms γ into γ' , represented as a finite sequence of homotopy moves. We define a graph $T(H)$, called the **trace** of H , whose nodes are the vertices of γ , the vertices of γ' , and the $1 \leftrightarrow 0$ and $2 \leftrightarrow 0$ moves in H ; each edge of $T(H)$ corresponds to the lifetime of a single vertex of the evolving curve. Every node of $T(H)$ has degree 1 or 2; thus, $T(H)$ is the disjoint union of paths and cycles.

We define the **type** of any vertex x of any annular curve γ as the winding number of the simpler curve γ_x^+ around the origin o (not around the vertex x); that is, we define $type(\gamma, x) := wind(\gamma_x^+, o)$. Vertex x is **irrelevant** if either $type(\gamma, x) = 0$ or $type(\gamma, x) = wind(\gamma, o)$ and **relevant** otherwise. Two vertices x and y have **complementary** types if $type(\gamma, x) + type(\gamma, y) = wind(\gamma, o)$, or equivalently, if $wind(\gamma_x^+, o) = wind(\gamma_y^-, o)$. If two vertices have complementary types, then either both are relevant or both are irrelevant.

Lemma 3.1. *The following hold for any annular curve:*

(a) *Each $1 \leftrightarrow 0$ move creates or destroys an irrelevant vertex.*

- (b) *Each $2 \leftrightarrow 0$ move creates or destroys two vertices with complementary types and identical winding numbers.*
- (c) *Each $3 \leftrightarrow 3$ move changes the winding numbers of three vertices, each by exactly 1.*
- (d) *Except as stated in (a), (b), and (c), homotopy moves do not change the type or winding number of any vertex.*

Proof: Claim (a) is immediate. Up to symmetry, there are only two cases to consider to prove claim (b): The two sides of the empty bigon are oriented in the same direction or in opposite directions. In both cases, γ_x^+ and γ_y^- are homotopic and $wind(\gamma, x) = wind(\gamma, y)$, where x and y are the vertices of the bigon. See Figure 3.1. Claim (c) follows immediately from the observation that each vertex involved in a $3 \rightarrow 3$ move passes over the curve exactly once. Finally, claim (d) follows from the fact that winding number is a homotopy invariant; specifically, if there is a homotopy between two planar curves γ and γ' whose image does not include a point p , then $wind(\gamma, p) = wind(\gamma', p)$ [34]. \square

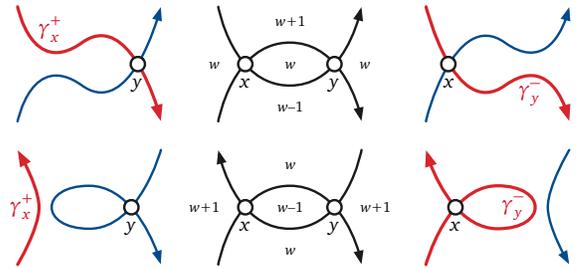


Figure 3.1. The vertices of empty bigons have complementary types and identical winding numbers.

Lemma 3.1 implies that no homotopy move transforms a relevant vertex into an irrelevant vertex or vice versa, and that relevant vertices are neither created by $0 \rightarrow 1$ moves nor destroyed by $1 \rightarrow 0$ moves. Let $R(H)$ denote the subgraph of edges in the trace $T(H)$ that correspond to relevant vertices of the evolving curve. Again, $R(H)$ is the disjoint union of paths and cycles. Each path in $R(H)$ connects either two vertices of γ with complementary types, two vertices of γ' with complementary types, or a vertex of γ and a vertex of γ' with identical types. Intuitively, each path in $R(H)$ is the record of a single relevant vertex alternately moving forward and backward in time, reversing directions and types at every $0 \leftrightarrow 2$ move.¹ We say that the nodes at the end of each path in $R(H)$ are **paired** by the homotopy H . We emphasize that different homotopies may lead to different pairings.

Between $2 \leftrightarrow 0$ moves, a relevant vertex can participate in any finite number of $3 \rightarrow 3$ moves. By Lemma 3.1(c),

¹Readers familiar with particle physics may recognize $R(H)$ as an elementary type of Feynmann diagram, where complementary relevant vertices play the role of particle-antiparticle pairs.

each 3→3 move changes the winding numbers of each of the three moving vertices by 1, and Lemma 3.1(d) implies that the winding number of a vertex changes only when it participates in a 3→3 move. Thus, the homotopy H must contain at least

$$\frac{1}{3} \sum_{x \sim y} |\text{wind}(x) - \text{wind}(y)|$$

3→3 moves, where the sum is over all pairs of paired vertices of $R(H)$, and the winding number of each vertex is defined with respect to the curve (γ or γ') that contains it.

3.2 A Bad Contractible Curve

Proof (of Theorem 1.1): For any pair of relatively prime integers p and q , the **flat torus knot** $T(p, q)$ is (any curve isomorphic to) the parametrized curve

$$((\cos(q\theta) + 2)\cos(p\theta), (\cos(q\theta) + 2)\sin(p\theta)),$$

which has exactly $(|p| - 1) \cdot |q|$ vertices and winding number p around the origin.

For any odd integer p , let Π_p denote the closed curve obtained by placing a scaled copy of $T(-p, 1)$ inside the innermost face of $T(p, 2)$ and attaching the two curves as shown in Figure 3.2. For purposes of illustration, we homotope all crossings into a narrow horizontal rectangle to the right of the origin, which is also where we join the two curves. The resulting curve Π_p has winding number zero around the origin and thus is contractible, and it has $3(p - 1)$ vertices. Within the rectangle, the curve consists of $2p$ simple paths, which we call **strands**; the endpoints of the strands are connected by disjoint parallel paths outside the rectangle. In the left half of the rectangle, strands are directed downward; in the right half, strands are directed upward. All but two strands connect the top and bottom of the rectangle; the only exceptions are the strands that connect the two flat torus knots.

We catalog the vertices of Π_p as follows. In the left half of the rectangle, Π_p has one vertex a_i with type i and winding number i , for each integer i from 1 to p . In the right half, Π_p has four vertices for each index i between 1 and $(p - 1)/2$ (see Figure 3.3):

- two vertices x_i and x'_i with type $-i$ and winding number $2i$;
- one vertex y_i with type i and winding number $p - 2i$; and
- one vertex z_i with type $i - p$ and winding number $p - 2i$.

Every homotopy from Π_p to a simple closed curve defines an essentially unique pairing of the vertices of Π_p ; without loss of generality, a_i is paired with x'_i , a_{p-i} is paired with z_i ,

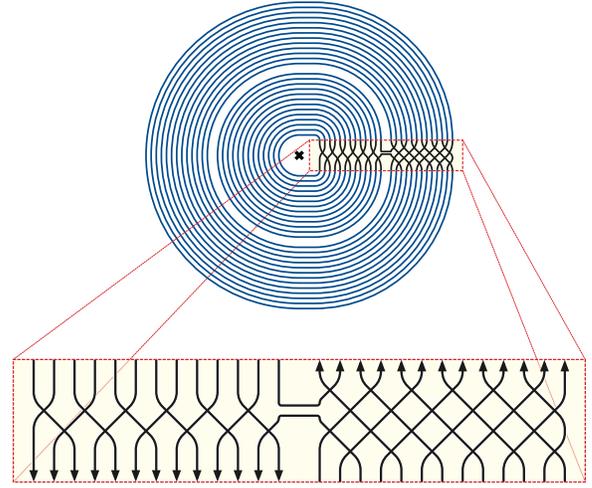


Figure 3.2. Our bad example curve Π_{13} in the punctured plane.

and x_i is paired with y_i , for each integer i between 1 and $(p - 1)/2$. Thus, the number of 3→3 moves in any homotopy that contracts Π_p is at least

$$\begin{aligned} & \frac{1}{3} \sum_{i=1}^{(p-1)/2} (|i - 2i| + |(p - i) - (p - 2i)| + |2i - (p - 2i)|) \\ &= \frac{1}{3} \sum_{i=1}^{(p-1)/2} (2i + |4i - p|) \\ &= \frac{1}{3} \left(\sum_{i=1}^{(p-1)/2} 2i + \sum_{j=1}^{(p-1)/2} (2j + 1) \right) \\ &= \frac{p(p - 1)}{6}. \end{aligned}$$

This completes the proof. \square

3.3 More complicated surfaces

We extend Theorem 1.1 to surfaces with more complex topology as follows. Recall that a closed curve in any surface Σ is **non-contractible** if it is not homotopic to a constant curve and **two-sided** if it has a neighborhood homeomorphic to the annulus. Let Σ be a compact surface, possibly with boundary or non-orientable, that contains a simple two-sided non-contractible cycle α ; the only compact surfaces that do *not* contain such a cycle are the sphere, the disk, and the projective plane. To create a bad example curve for Σ , we simply embed our previous annular curve Π_p in an annular neighborhood A of α . The resulting curves are still contractible in Σ and, as we will shortly prove, still require $\Omega(n^2)$ homotopy moves to simplify.

However, winding numbers are not well-defined in surfaces of higher genus, so we need a more careful argument to prove the quadratic lower bound. Instead

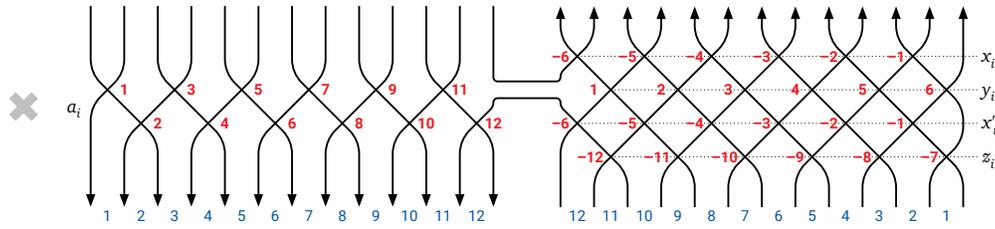


Figure 3.3. Vertices of Π_{13} annotated by type (bold red numbers next to each vertex) and winding number (thin blue numbers directly below each vertex).

of reasoning directly about homotopy moves on Σ , we lift everything to a certain covering space of Σ previously considered by several authors [15, 18, 32, 35, 44].

Theorem 3.1. *Let Σ be a compact connected surface, possibly with boundary or non-orientable (but not the sphere, the disk, or the projective plane). For any positive integer n , there is a contractible curve with n vertices in Σ that requires $\Omega(n^2)$ homotopy moves to simplify.*

Proof: Let α be a simple two-sided non-contractible closed curve in Σ , that is, a non-contractible curve that lies in an open neighborhood A homeomorphic to the open annulus $S^1 \times (0, 1)$. Every compact connected surface (other than the sphere, the disk, or the projective plane) contains such a curve.

The **cyclic covering space** $\hat{\Sigma}_\alpha$ of Σ with respect to α is the quotient of the universal covering space of Σ by the infinite-cyclic subgroup of the fundamental group $\pi_1(\Sigma)$ generated by α . Let $\pi: \hat{\Sigma}_\alpha \rightarrow \Sigma$ be the corresponding covering map. Standard covering space results imply that α has a unique lift $\hat{\alpha}$ in $\hat{\Sigma}_\alpha$ which is a simple closed curve. Also, $\hat{\alpha}$ has an open annular neighborhood \hat{A} with non-contractible boundary components in $\hat{\Sigma}$. (See, for example, Schrijver [44, Proposition 2].) Moreover, we may assume that the restriction of the covering map π to \hat{A} is a homeomorphism to A .

Let $\hat{\gamma}$ be an arbitrary contractible curve in \hat{A} , and let γ be the projection of $\hat{\gamma}$ to A . The two curves γ and $\hat{\gamma}$ have the same number of vertices and edges. Any homotopy $H: S^1 \times [0, 1] \rightarrow \Sigma$ from γ to a point lifts to a homotopy $\hat{H}: S^1 \times [0, 1] \rightarrow \hat{\Sigma}_\alpha$ from $\hat{\gamma}$ to a point. Each homotopy move in \hat{H} projects to a homotopy move in H , but H may include additional homotopy moves, where the strands involved are projected from different parts of the covering space. It follows that simplifying γ in Σ requires at least as many homotopy moves as simplifying $\hat{\gamma}$ in $\hat{\Sigma}$.

The lower bound now follows directly from Theorem 1.1, by setting $\hat{\gamma} = \Pi_p$ for some $p = \Theta(n)$, as defined in Section 3.2. \square

Theorem 3.1 strengthens an $\Omega(n^2)$ lower bound of Chang and Erickson for tightening non-contractible curves in orientable surfaces [11]. Results of Hass and Scott [32,

Theorem 2.7] imply that our lower bound is tight for the Möbius band, the Klein bottle, and any orientable surface except the sphere or the disk; any contractible curve on these surfaces can be simplified using at most $O(n^2)$ homotopy moves.

4 Upper Bound: Surfaces with Boundary

In this section, we consider the case of surfaces with boundary. The benefit of working with such surfaces is that the fundamental group of surface with boundary is *free*; intuitively on such surface one can always find a way to decrease the complexity of the bigon wrapping around the surface. Later in Section 5 we will describe a similar algorithm for closed curves on an arbitrary orientable surface without boundary. The reader is encouraged to follow the order of the presentation and get an intuitive sense of how the bigon removal algorithm operates in this simpler setting.

4.1 Singular Bigons and Singular Monogons

Let γ be a closed curve on a surface Σ . Recall that an (*embedded*) **bigon** in γ consists of two simple interior-disjoint subpaths of γ in Σ with the same endpoints that bound a disk and enclose one corner at each of those endpoints. Following Hass and Scott [32], a **singular bigon** in γ consists of two subpaths of γ that are disjoint in the domain, and the two subpaths are homotopic to each other in Σ . Similarly, a **singular monogon** is a subpath of γ whose two endpoints are identical in Σ , and that forms a null-homotopic closed curve in Σ .

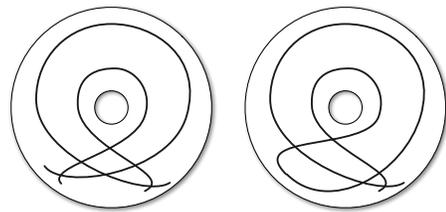


Figure 4.1. A basic singular bigon and a basic singular monogon in the annulus.

The authors have placed this paper in the public domain.

Our algorithm relies on the following simple property of singular monogons and bigons, which follows immediately from their definition.

Lemma 4.1. *The bounding paths of any singular monogon or bigon in γ cross γ at most $2n$ times.*

An important subtlety of Hass and Scott’s definition is that a lift of a singular bigon to the universal cover is not necessarily an *embedded* bigon. First, the lifted boundary paths of the bigon need not be simple or disjoint. More subtly, the endpoints of the lifted bigon might not enclose single corners: an embedded bigon looks like a lens \oslash , but a lift of a singular bigon might resemble a heart \heartsuit or a butt \ominus . Similarly, a lift of a singular monogon is not necessarily an *embedded* monogon; the lifted subpath might self-intersect way from its endpoint, and it may not enclose a single corner at its endpoint.

We define a singular monogon or singular bigon to be **basic** if any of its lifts on the universal cover is an *embedded* monogon or bigon, respectively. Hass and Scott proved that any closed curve with excess intersections on an arbitrary orientable surface, with or without boundary, must contain a singular monogon or a singular bigon [32, Theorem 4.2]. However, a close reading of their proof reveals that the singular monogon or singular bigon they find is in fact basic. We thus restate their result without repeating the proof.

Lemma 4.2 (Hass and Scott [32]). *Let γ be a closed curve on an arbitrary orientable surface. If γ has excess intersections, then there is a **basic** singular monogon or a **basic** singular bigon in γ .*

This lemma is the foundation of our upper bound proofs, both in this section and the next. Given a curve γ with n vertices that is not already tightened, we decrease the number of vertices of γ as follows. If γ contains an *embedded* monogon or bigon, we delete it, following Steinitz’s algorithm [45, 46], using $O(n)$ homotopy moves. Otherwise, if γ contains a basic singular bigon, we attempt to remove it, essentially by swapping the two bounding curves; however, if at any point γ has only $n-2$ vertices, we immediately abort the bigon removal. Finally, if γ contains no basic singular bigons, Lemma 4.2 implies that γ must contain a basic singular monogon; we perform a single $0 \rightarrow 1$ move to transform it into a basic singular bigon (as shown in Figure 4.2) and then defer to the previous case.

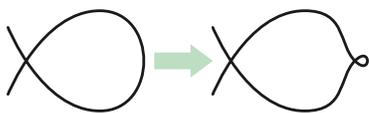


Figure 4.2. A single $0 \rightarrow 1$ move transforms a basic singular monogon into a basic singular bigon.

The curve γ is tightened after repeating the previous reduction process at most n times. Thus, Theorem 1.2 follows immediately from the following lemma, which we prove in the remainder of this section.

Lemma 4.3. *Let Σ be an orientable surface of genus g with $b > 0$ boundary components, and let γ be a closed curve in Σ with n vertices that contains a basic singular bigon, but no embedded monogons or bigons. The number of vertices of γ can be decreased by 2 using $O((g + b)n^2)$ homotopy moves.*

4.2 Removing a Basic Singular Bigon

Fix a surface Σ and a closed curve γ with n vertices, satisfying the conditions of Lemma 4.3. A **system of arcs** Δ on the surface Σ is a collection of simple disjoint boundary-to-boundary paths that cuts the surface Σ open into one single polygon. Euler’s formula implies that every system of arcs contains exactly $2g + b - 1$ arcs. Cutting the surface Σ along these arcs leaves a topological disk P whose boundary alternates between arcs (each arc in Δ appearing twice) and subpaths of the boundary. We refer to P as the **fundamental polygon** of Σ with respect to Δ .

For any closed curve γ on any orientable surface Σ with boundary, there is a system of arcs Δ satisfying the following **crossing property**: Each arc in Δ intersects each edge of γ at most twice, and only transversely. (For examples of such a construction, see Colin de Verdière and Erickson [15, Section 6.1] or Erickson and Nayyeri [21, Section 3].) The fundamental polygon induces a tiling of the universal cover of Σ ; we call each lift of the fundamental polygon a **tile**.

Any basic singular bigon β of γ in Σ lifts to a bigon $\hat{\beta}$ in the universal cover of Σ , with two bounding subpaths λ and ρ that are disjoint in the domain of γ except possibly at their endpoints. Since $\hat{\beta}$ bounds a disk in the universal cover, any lift of any arc of Δ intersects $\hat{\beta}$ an even number of times. The intersection of a tile with $\hat{\beta}$ may have several components; we call each component a **block**. A block is **transverse** if it is adjacent to both λ and ρ , and **extremal** otherwise. The **transverse** blocks have a natural linear ordering B_1, \dots, B_k along either λ or ρ .

Our process for removing the bigon $\hat{\beta}$ has three stages: (1) Sweep inward over the extremal blocks, (2) sweep across the sequence of transverse blocks, and finally (3) remove one small empty bigon at a corner of $\hat{\beta}$. The first two stages are illustrated in Figure 4.3. This homotopy projects to a homotopy on Σ . We will prove that at the end of this bigon removal process, γ has exactly $n - 2$ vertices.

To simplify our algorithm, we actually abort the bigon-removal process immediately as soon as γ has $n - 2$ vertices; however, for purposes of analysis, we conservatively assume that the removal process runs to

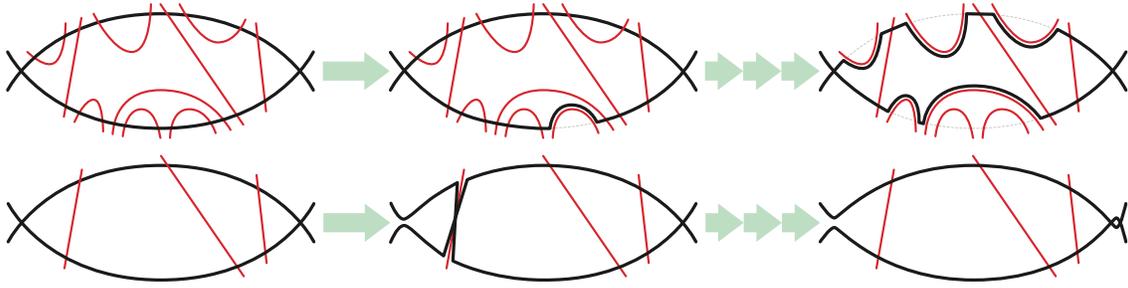


Figure 4.3. Removing a basic singular bigon on a surface with boundary. Top: Sweeping extremal blocks. Bottom: Sweeping transverse blocks.

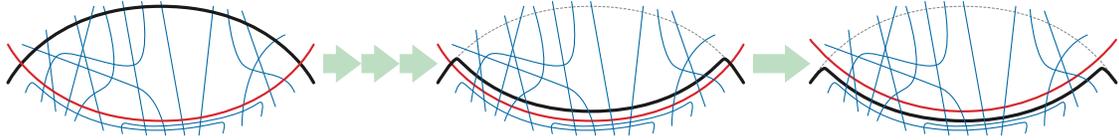


Figure 4.4. Sweeping a minimal embedded bigon bounded by a subpath of γ (black) and a subpath of Δ (red). Thin (blue) lines are other subpaths of γ .

completion. We separately analyze stage (1) and stage (2) next.

Lemma 4.4. *All extremal blocks can be removed from $\hat{\beta}$ using $O((g + b)n^2)$ moves, without changing the number of vertices of γ .*

Proof: We actually describe how to remove every embedded bigon formed by a subpath of γ and a subpath of any arc in Δ using at most $O((g + b)n^2)$ homotopy moves, each of which is a $3 \rightarrow 3$ move. Every extremal block in $\hat{\beta}$ projects to such an embedded bigon, because tiles (and a fortiori blocks) project injectively into the surface Σ .

We proceed inductively as follows. Suppose γ and Δ bound an embedded bigon, since otherwise there is nothing to prove. Let B be a *minimal* embedded bigon with respect to containment, bounded by a subpath δ of an arc in Δ , and a subpath α of the curve γ . Because γ has no embedded monogons or bigons, every subpath of γ inside B is simple, and every pair of such subpaths intersects at most once. Moreover, every such subpath has one endpoint on α and the other endpoint on δ . Thus, the number of intersections between δ and γ is equal to number of intersections between α and $\gamma \setminus \alpha$.

To remove B , we first sweep the subpath α across B until the bigon defined by α and δ has no vertices in its interior, and then sweep α across δ without performing any additional homotopy moves, as shown in Figure 4.4. Because the number of intersections between δ and γ is equal to number of intersections between α and $\gamma \setminus \alpha$, this sweep does not change the number of vertices of γ .

The pattern of subpaths inside B implies that if the interior of B contains any vertices of γ , then some triangular face of γ lies inside B and adjacent to an edge of α . Thus, we can reduce the number of interior vertices of B with a single $3 \rightarrow 3$ move. It follows inductively that the number

of moves required to sweep over B is equal to the number of vertices in the interior of B , which is trivially at most n .

Removing a minimal embedded bigon between γ and Δ takes at most n moves and decreases the number of intersections between γ and Δ by 2. Each of the $O(g + b)$ arcs in Δ intersects each of the $O(n)$ edges of γ at most twice by the crossing property of Δ , so the total number of such intersections is at most $O((g + b)n)$. Finally, because every move is a $3 \rightarrow 3$ move, we *never* change the number of vertices of γ . The lemma follows immediately. \square

Now let B_1, B_2, \dots, B_k denote the sequence of transverse blocks of $\hat{\beta}$, and for each index $1 \leq i < k$, let δ_i denote the common boundary B_i and B_{i+1} . Each path δ_i is a subpath of a lift of some arc in Δ . For notational convenience, let $x = \delta_0$ and $y = \delta_k$ denote the endpoints of $\hat{\beta}$, so that each block B_i has paths δ_{i-1} and δ_i on its boundary.

Recall that λ and ρ denote the bounding subpaths of $\hat{\beta}$. To sweep over the transverse blocks, we *intuitively* maintain a path ϕ from a point on λ to a point on ρ , which we call the **frontier**. The frontier starts as a trivial path at the endpoint δ_0 . Then we repeatedly sweep the frontier over B_i from δ_{i-1} to δ_i , as i goes from 1 to k . After these k iterations, the frontier lies at the endpoint δ_k .

Our actual homotopy modifies the bounding curves λ and ρ as shown in the bottom of Figure 4.3. Intuitively, the prefixes of λ and ρ “behind” ϕ are swapped; the frontier itself is actually an arbitrarily close pair of crossing subpaths connecting the swapped prefixes of λ and ρ with the unswapped suffixes “ahead” of the frontier. Replacing the single path ϕ with a close pair of crossing paths increases the number of homotopy moves to perform the sweep by only a constant factor.

Lemma 4.5. *Sweeping ϕ over one transverse block requires at most $O(n)$ homotopy moves.*

Proof: Consider a sweep over B_i , from δ_{i-1} to δ_i . We start by moving the frontier just inside B_i , without performing any homotopy moves. The main sweep passes ϕ over every vertex in B_i , including the vertices on the bounding paths λ and ρ , stopping ϕ just before it reaches δ_i . Finally, we move the frontier onto δ_i without performing any homotopy moves. Because the interior of each block projects injectively onto the surface, no other translate of ϕ intersects B_i during the sweep.

Up to constant factors, the number of homotopy moves required to sweep B_i is bounded by the number of vertices inside B_i , plus the number of intersections between γ and the bounding subpaths δ_{i-1} or δ_i . There are trivially at most n vertices in B_i , and the crossing property of the system of arcs Δ implies that each arc in Δ intersects γ at most $O(n)$ times. \square

With the two previous lemmas in hand, we are finally ready to prove Lemma 4.3. Let γ be a closed curve in Σ with a basic singular bigon β , let $\hat{\beta}$ be a lift of β to the universal cover of Σ , and let λ be one of the bounding paths of $\hat{\beta}$.

The definition of singular bigon immediately implies that λ contains at most $2n$ edges of γ . Each of these edges crosses each arc of Δ at most twice, and there are $O(g+b)$ arcs in Δ , so λ crosses Δ at most $O((g+b)n)$ times. Each transverse block B_i except the last can be charged to the unique intersection point $\delta_i \cap \lambda$. We conclude that $\hat{\beta}$ contains $O((g+b)n)$ transverse blocks.

Sweeping inward over all extremal blocks in $\hat{\beta}$ requires $O((g+b)n^2)$ homotopy moves and does not change the number of vertices of γ . Sweeping over all $O((g+b)n)$ transverse blocks requires a total of $O((g+b)n^2)$ homotopy moves. Sweeping the transverse blocks has the same effect as smoothing one endpoint of the bigon and doubling the other endpoint, as shown on the bottom right of Figure 4.3, which implies that γ still has n vertices. Removing the final empty bigon with a single $2 \rightarrow 0$ move reduces the number of vertices to $n-2$.

This completes the proof of Lemma 4.3, and therefore the proof of Theorem 1.2.

5 Upper Bound: Surfaces Without Boundary

Finally, we prove our upper bound for closed curves on surfaces without boundary. We follow the same high-level strategy described in Section 4; consequently, it suffices to prove that a basic singular bigon can be removed using $O(n^3)$ homotopy moves.

Instead of a system of arcs, we decompose the surface using a *reduced cut graph*; this cut graph induces a regular hyperbolic tiling in the universal cover of the surface. In Section 5.1 we describe how to compute a cut graph whose induced tiling intersects the bounding paths of any basic

singular bigon at most $O(n)$ times. In Section 5.2, we apply Dehn’s isoperimetric inequality for regular hyperbolic tilings [16] to bound the number of tiles lying in the interior of the bigon. Then we describe our process for removing a singular bigon at two levels of detail. First, in Section 5.3, we provide a coarse description of the homotopy as a sequence of moves in the *bigon graph*, which is the decomposition of the lifted bigon by the tiling. We process the regions in this decomposition in a particular order to keep the number of *chords* created by translates of the moving path under control. Finally in Section 5.4 we obtain the actual sequence of homotopy moves by carefully perturbing the curves in the previous homotopy into general position; bounding the intersections between perturbed chords is the most delicate portion of our analysis.

5.1 Dual Reduced Cut Graphs

A **combinatorial surface** G decomposes a closed surface Σ without boundary into vertices, edges, and faces, such that each vertex is a point, each edge is a simple path, and each face is an open disk on the surface Σ . The **dual** of a combinatorial surface G is another combinatorial surface G^* on the same surface Σ as G . The vertices, edges, and faces of G^* correspond to faces, edges, and vertices of G , respectively. The dual combinatorial surface G^* inherits a natural embedding from the embedding of G in the surface Σ . A **tree-cotree decomposition** of a combinatorial surface G is a partition (T, L, C) of the edges of G into three pairwise disjoint subsets, where T induces a spanning tree in G , C induces a spanning tree of G^* , and L is the set of leftover edges $E(G) \setminus (T \cup C)$. The number of edges in L is twice the genus of the underlying surface Σ , when Σ has no boundary.

Let γ be a closed curve on Σ ; we temporarily view γ as a 4-regular graph embedded on the surface. Let G be a refinement of this graph obtained by triangulating every face. A **dual reduced cut graph** X (hereafter, just *cut graph*) is a combinatorial surface obtained from a tree-cotree decomposition (T, L, C) of G as follows: Start with the subgraph of G^* induced by the dual spanning tree C^* and the leftover edges L^* , repeatedly delete vertices with degree one, and finally perform series reductions on all vertices with degree two [20]. To be consistent with the terminology in Section 4.2, we call the edges of X **arcs**.

Notice that X inherits a natural embedding into Σ from G^* , which in turn inherits an embedding into Σ from the embedding of G . Euler’s formula implies that the cut graph X has $4g-2$ vertices, $6g-3$ arcs, and 1 face. If one cut opens the surface Σ along X , one obtains a polygon with $12g-6$ sides, known as a **fundamental polygon** of Σ with respect to X . The fundamental polygon induces a tiling of the universal cover $\hat{\Sigma}$ of Σ ; we refer to each copy of the fundamental polygon as a **tile**. By construction, the

cut graph X is a combinatorial surface of Σ that satisfies the following **crossing property**: *Each edge of γ crosses X at most once.* We emphasize that this crossing property might no longer hold when we start moving the curve γ . Compared with the system of arcs we used in Section 4, the cut graph gives us an improved upper bound on the number of tiles intersecting the bounding paths of an embedded bigon in the universal cover of Σ .

5.2 Isoperimetric Inequality

Consider an embedded bigon $\hat{\beta}$ in the universal cover of surface Σ , which is a lift of a basic singular bigon in the curve γ on Σ . Unlike the case of surface with boundaries in Section 4, there will be tiles lying completely in the *interior* of the bigon $\hat{\beta}$, without intersecting the two bounding paths, and thus require additional work. We bound the number of such interior tiles using a discrete isoperimetric inequality, which is a consequence of Dehn’s seminal observation that the graph metric defined by a regular tiling of the hyperbolic plane is a good approximation of the continuous hyperbolic metric [16]. We provide a self-contained proof of this inequality, using a combinatorial version of the Gauss-Bonnet theorem described at varying levels of generality by Banchoff [6], Lyndon and Schupp [38], and Gersten and Short [25].

Let G be a combinatorial surface on surface Σ , and let $\chi(\Sigma)$ be the Euler characteristic of Σ , defined as the number of vertices and faces in G minus the number of edges in G , which is equal to $\chi(\Sigma) = 2 - 2g - b$, where g is the genus of Σ and b is the number of boundary components of Σ . One can view the definition of the Euler characteristic of Σ through a different lens. Assign an arbitrary real “interior angle” $\angle c$ (measured in circles²) to each corner c of Σ . Define the **curvature** $\kappa(v)$ of a vertex v in Σ as $1 - \frac{1}{2} \deg v - \sum_{c \in v} (\frac{1}{2} - \angle c)$, and the **curvature** $\kappa(f)$ of a face f in Σ as $1 - \sum_{c \in f} (\frac{1}{2} - \angle c)$. The following equality is known as the **combinatorial Gauss-Bonnet theorem**:

$$\sum_v \kappa(v) + \sum_f \kappa(f) = \chi(\Sigma).$$

Now we are ready to bound the number of tiles that lies completely within a bigon using the combinatorial Gauss-Bonnet theorem. Let Σ be a surface of genus g without boundaries, let γ be a closed curve on Σ , and let X be the cut graph of γ on Σ . Let $\hat{\beta}$ be an embedded bigon in the universal cover of Σ that is a lift of a basic singular bigon of γ in Σ , intersecting the lift of X transversely. Let the **perimeter** $L_X(\hat{\beta})$ of $\hat{\beta}$ be the number of intersections between $\hat{\beta}$ and edges of the tiling induced by X . Let the **area** $A_X(\hat{\beta})$ be the number of components in the

²As opposed to radians or degrees. Why would anyone refer to a right angle other than one-fourth of a circle?

intersection between all tiles of X and the disk bounded by $\hat{\beta}$.

Lemma 5.1. *Let Σ be a closed surface of genus $g > 1$, let γ be a closed curve on Σ , let X be the cut graph of γ on Σ , and let $\hat{\beta}$ be any embedded bigon that intersects the lift of X transversely. We have $A_X(\hat{\beta}) = O(L_X(\hat{\beta}))$.*

Proof: First we connect the number of tiles that lie completely in the interior of $\hat{\beta}$ with the number vertices on the boundary of the union of these tiles. Consider a region R in the universal cover of Σ that is the union of a subset of tiles. The **perimeter** L of R is the number of vertices on the boundary of R , and the **area** A of R is the number of tiles that lies in R . Every boundary vertex is either incident to either one interior face and has degree 2 (convex) or incident to two interior faces and has degree 3 (concave). If we assign angle $1/3$ (circles) to each corner, then every interior vertex has curvature 0, every face has curvature $2 - 2g$, every convex vertex has curvature $1/6$, and every concave vertex has curvature $-1/6$. The combinatorial Gauss-Bonnet theorem implies that the number of convex vertices minus the number of concave vertices is $(12g - 12)A + 6$. Thus, some face f must be incident to more than $12g - 12$ convex vertices, which implies that f is incident to at most 4 interior edges. (Otherwise, because each convex vertex must be incident no interior edges, a face f incident to at least 5 interior edges must be incident to at most $12g - 6 - (5 + 1) = 12g - 12$ convex vertices, a contradiction.) Removing f decreases A by 1 but decreases L by at least $12g - 14$. The isoperimetric inequality $A \leq L/(12g - 14)$ now follows immediately by induction.

Now consider the embedded bigon $\hat{\beta}$. Since each vertex in X has degree exactly 3, the perimeter $L_X(\hat{\beta})$ is an upper bound on the number of vertices on the boundary of the union of tiles that lies completely within the disk bounded by $\hat{\beta}$. The number of components in the intersection of the interior of $\hat{\beta}$ and the tiles of X that do not completely lie within $\hat{\beta}$ is also at most $L_X(\hat{\beta})$. Therefore the area $A_X(\hat{\beta})$ is at most $O(L_X(\hat{\beta}))$, as claimed. \square

5.3 Coarse Homotopy

Let β be a basic singular bigon in γ , let $\hat{\beta}$ be its lift to the universal cover, and let λ and ρ be the bounding paths of $\hat{\beta}$. Our goal is to remove this bigon by swapping the bounding paths λ and ρ , which has the same effect as smoothing the two endpoints of β , reducing the number of vertices of γ by 2. See Figure 5.1. In this section, we construct a homotopy from λ to ρ , not as a sequence of individual homotopy moves, but as a coarser sequence of moves in a certain planar graph. Applying the reversal of this sequence of moves to ρ moves it to the original position of λ , completing the exchange of the two bounding paths.

The authors have placed this paper in the public domain.

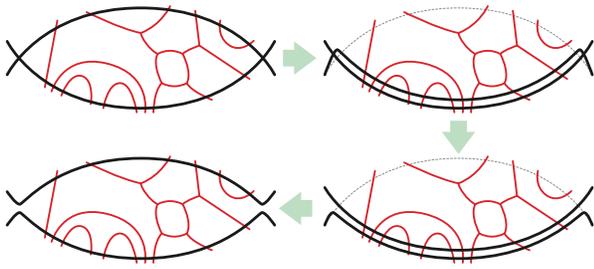


Figure 5.1. Swapping the two bounding paths of a bigon.

Bigon graph. The tiling of the universal cover of Σ induced by the cut graph X decomposes the disk bounded by $\hat{\beta}$ into pieces; we call this decomposition the **bigon graph**. More formally, we define the bigon graph G as follows. The vertices of G are the two endpoints of $\hat{\beta}$, the intersections of λ and ρ with edges of the tiling, and the vertices of the tiling in the interior of $\hat{\beta}$. The edges of G are subpaths of $\lambda \cup \rho$ and subpaths of tiling edges bounded by these vertices. Finally, the bounded faces of G are the components of the intersection of each tile with the interior of $\hat{\beta}$. We emphasize that the intersection of a single tile with the interior of $\hat{\beta}$ may have several components.

Lemma 4.1 and the crossing property of the cut graph X imply that at most $O(n)$ tiles of Σ intersect the two bounding paths λ and ρ of $\hat{\beta}$. Thus Lemma 5.1 implies that the bigon graph G has at most $O(n)$ faces, and therefore $O(n)$ vertices and edges by Euler’s formula.

Discrete homotopy. We now construct a *discrete homotopy* [8, 9, 31] through the bigon graph G that transforms one bounding path λ of the bigon into the other bounding path ρ . This discrete homotopy is a sequence of walks in G —which may traverse the same edge in G more than once—rather than a sequence of generic curves. In the next section, we will carefully perturb these walks into generic curves, and implement each step of the discrete homotopy as a finite sequence of homotopy moves.

Let W be a walk on the bigon graph G from one endpoint of the bigon to the other. A **spike** in W is an edge of G followed immediately by the same edge in the opposite direction. We define two local operations for modifying W ; see Figure 5.2.

- **Face move:** Replace a single edge e in W with the complementary boundary walk around some face f of G that is incident to e .
- **Spike move:** Delete a spike from W and decrease the length of W by two.

We emphasize that after a face move across face f , the frontier walk W may traverse some edges of f more than once; moreover, these multiple traversals may or may not be spikes. Because every face f is a disk and the edge e and its complementary boundary walk share endpoints, any face move can be implemented by a homotopy across f .

Similarly, a spike move can be implemented by a homotopy in the edge containing the spike. A discrete homotopy in G is a finite sequence of face moves and spike moves. We refer to the current walk W at any stage of this homotopy as the **frontier walk**.

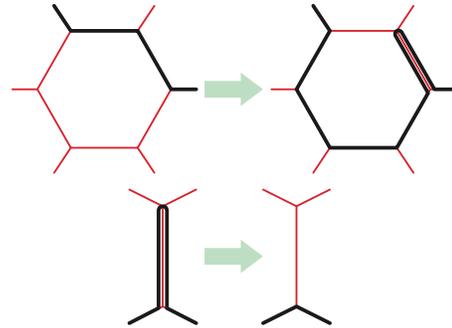


Figure 5.2. A face move and a spike move.

We construct our discrete homotopy from λ to ρ as follows. The initial frontier walk W is equal to λ (viewed as a walk in G). Let T be an arbitrary spanning tree of the bigon graph G that includes every edge in ρ , and let C be the complementary spanning tree of the dual graph G^* . We first perform a face move within in each face of G , following a preorder traversal of C ; that is, we move W across a face f only after moving W across the parent of f in C . (See Figure 5.3.) After all faces have been processed, we repeatedly remove spikes from W , until no more spikes remain. It is not hard to see that we require one spike move for each edge of T that is not an edge of ρ . At the end of this procedure, the frontier walk W coincides with ρ .

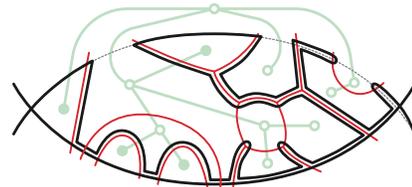


Figure 5.3. The frontier walk after eight face moves, following a preorder traversal of the dual spanning tree.

This sequence of graph moves guarantees a crucial **multiplicity property**: At every stage of the discrete homotopy, the frontier walk W traverses each edge of G at most twice. Specifically, W traverses each edge of λ at most once, and W traverses an edge e that is not in λ only if a face move has been performed on a face incident to e . Each edge is incident to at most two faces. It follows immediately that the length of W (that is, the number of edges) is always at most twice the number of edges of G .

5.4 Fine Homotopy

Finally, we refine the discrete homotopy in the previous section, first by perturbing the moving frontier walk so

that after every graph move, γ is a generic curve, and then by decomposing the perturbed graph moves into a finite sequence of homotopy moves. We will denote the perturbed **frontier** path by ω , reserving W for the original frontier walk in G .

Bubble-wrapping. We define a convenient family \mathcal{O} of open sets, which we call **bubbles**, that covers the bigon graph G and its bounded faces, following a construction of Babson and Chan [5]. (See also Erickson [19].) Each bubble in \mathcal{O} is either a vertex bubble, an edge bubble, or a face bubble. The vertex bubbles are disjoint open balls around the vertices of G . The edge bubbles are disjoint open neighborhoods of the portions of the edges of G outside the vertex bubbles. Finally, the face bubbles are disjoint open neighborhoods of the portions of the faces of G outside the vertex and edge bubbles. The intersection of all pairs of two bubbles of different types is the disjoint union of open disks, one for each incidence between the corresponding vertex and edge, vertex and face, or edge and face of G . Every bubble projects injectively onto the original surface Σ . We also require the intersection of all pairs of *translates* of edge bubbles to be the disjoint union of open disks, one for each intersection between the corresponding translates of edges of G . See Figure 5.4.

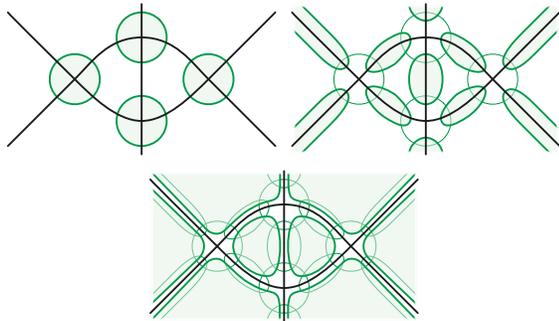


Figure 5.4. Vertex bubbles, edges bubbles, and face bubbles.

Perturbing the frontier. Interactions between translates of the moving frontier present a significant subtlety in our algorithm. To control the complexity of all the curves inside the bubbles of G , we carefully perturb the frontier (and therefore its translates) within each bubble.

The **augmented bigon graph** \tilde{G} is a refinement of the bigon graph G , defined as the arrangement of the bounding paths λ and ρ , the portion of the tiling induced by the cut graph X inside the disk of the bigon, and the portion of all translates of λ and ρ inside the disk of the bigon. Each edge of G is a walk in \tilde{G} , and each face of G is decomposed into one or more faces of \tilde{G} . The complexity of \tilde{G} within each face of G is at most the complexity of (the lift of) γ in the same face, as translates of λ and ρ are still subpaths of the lift of γ . As the interior of each tile embeds into the surface Σ , the complexity of \tilde{G} inside each face of G is at most $O(n)$, where n is the number of vertices in the original curve γ .

To see why this refinement is necessary, observe that during the discrete homotopy described in Section 5.3, *translates* of the frontier walk W do not necessarily follow the bigon graph G ; however, the intersection of every translate of W with the bigon does lie in the augmented bigon graph \tilde{G} . In summary, between face or spike moves, the frontier walk W and all its translates are walks in \tilde{G} .

Within any vertex bubble, G and \tilde{G} are identical, because translates of λ and ρ do not intersect the vertices of G . Within any face bubble of G , each translate of W follows only translates of λ or ρ , because face bubbles do not intersect the tiling. Finally, the intersection of \tilde{G} with any edge bubble of G is homeomorphic a straight line segment (the edge of G) crossed by multiple disjoint line segments (translates of λ and ρ).

We model each edge bubble as a *Euclidean* rectangle containing several straight segments parallel to the edge, which we call **tracks**, arranged so that if an edge e of G intersects an edge e' of a translate of G , each track in the edge bubble of G intersects each track in the translated edge bubble of e' transversely. (The metric is merely a convenience, so that we can write “straight” and “parallel”; the tracks can be defined purely combinatorially.)

Now consider a walk W in the augmented bigon graph \tilde{G} . This walk traverses a sequence of subpaths of edges of *translates* of the original bigon graph G . To define the perturbed walk ω , we perturb each maximal subpath along edge e onto a unique track in the edge bubble of e . Moreover, when W switches from edge e to another edge e' (including at the tip of a spike, which we view as a zero-length walk), the perturbed walk ω has a corner at the intersection of those two tracks. Thus, every subpath of ω inside the edge bubble of some edge e alternates between tracks parallel to other translated edges and tracks parallel to e . Intuitively, we say that the resulting frontier ω **sticks to** the original walk W , and in particular, subpaths of ω within an edge bubble **stick to** the corresponding edge of G . See Figure 5.5.

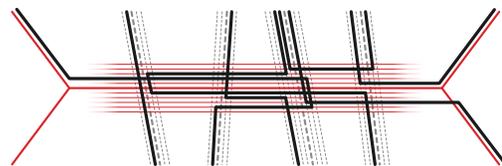


Figure 5.5. Closeup of the augmented bigon graph \tilde{G} near edge e of the bigon graph G , showing subpaths of γ sticking to subpaths in e , including the perturbations of two spikes.

Graph moves revisited. In our perturbed homotopy, we require every face move to be performed entirely within the corresponding face bubble, and every spike move to be performed entirely within the corresponding edge bubble, while maintaining the track structure of the perturbed frontier ω . To this end, we introduce two additional graph moves.

The authors have placed this paper in the public domain.

- **Edge move:** Move a subpath sticking to an edge e of G into an incident face bubble, within the edge bubble of e .
- **Vertex move:** Move the curve across a vertex v of G within the corresponding vertex bubble.

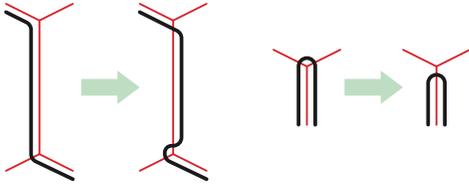


Figure 5.6. An edge move and a vertex move.

These moves can be seen as preprocessing steps to ensure that a subpath of ω lies in the proper face or edge bubble before performing a face or a spike move. Thus, our perturbed coarse homotopy still follows the outline given in Section 5.3, but now each face move is prefaced by a single edge move, and each spike move is prefaced by a single vertex move, as shown in Figure 5.7.

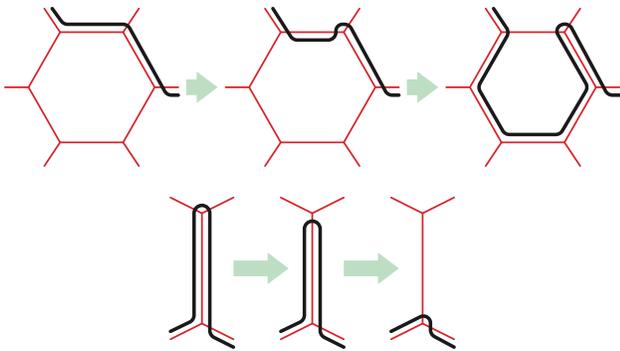


Figure 5.7. Top: An edge move followed by a face move. Bottom: A vertex move followed by a spike move.

We emphasize that every face move is performed entirely within a face bubble, every edge move and spike move is performed entirely within an edge bubble, and every vertex move is performed entirely within a vertex bubble.

The final homotopy. Finally, we construct a sequence of homotopy moves from one bounding path λ of the lifted bigon $\hat{\beta}$ to (a small open neighborhood of) the other bounding path ρ , by decomposing the perturbed graph moves.

Lemma 5.2. *Let Σ be an orientable surface without boundary, and let γ be a closed curve with n vertices on Σ that contains a basic singular bigon, but no embedded monogons or bigons. A basic singular bigon can be removed from γ using $O(n^3)$ homotopy moves, without changing the rest of γ .*

Proof: Let β be a basic singular bigon in γ ; let $\hat{\beta}$ be the lift of β to the universal cover; let λ and ρ be the bounding curves of $\hat{\beta}$; let G be the corresponding bigon graph; and let \tilde{G} be the augmented bigon graph. Our earlier analysis implies that G has at most $O(n)$ vertices, edges, and faces. Thus, moving λ to ρ requires at most $O(n)$ graph moves.

Each of these graph moves is performed within a bubble in \mathcal{O} that embeds in Σ , and therefore can be realized using $O(m)$ homotopy moves, where m is the number of vertices of γ within that bubble before the graph move begins. It remains only to prove the following claim:

At every stage of the algorithm, the number of vertices of γ inside any bubble is at most $O(n^2)$.

The proof of this claim is surprisingly delicate. All the properties we mentioned in each of the previous subsections contribute to avoid the danger of increasing the number of vertices in γ uncontrollably during the process: (a) Dividing the homotopy into graph moves based on faces in G (Section 5.1 and Section 5.2), (b) the order of face moves (Section 5.3), and (c) the way each graph move is implemented (Section 5.4). For the rest of the proof we refer to subpaths of translates of the frontier ω within a bubble simply as *chords*.

As there are at most $O(n)$ faces in G , the multiplicity property of the frontier walk W implies that W always has length at most $O(n)$, and therefore is incident to at most $O(n)$ faces. This in turn implies that there are at most $O(n)$ chords inside any bubble.

The maximum number of vertices of curve γ inside a bubble at any stage of the homotopy is at most the sum of the number of vertices of γ before the homotopy, the number of intersections between the original γ and the chords, and the number of intersections between pairs of chords. The first term is n by definition; Lemma 4.1 implies that the second term is also at most $O(n)$. To bound the last term, we separately consider each type of bubble:

- **Face bubbles:** Every translate of ω sticks to either edges of the tiling or translates of λ and ρ . Because vertices and edges of the tiling do not lie inside the bubble, every chord within a face bubble sticks to some edge in the translates of λ or ρ between any two graph moves. The multiplicity property of the frontier implies that at most two translates of ω stick to each edge e of translates λ or ρ ; these two translates intersect at most a constant number of times $O(1)$ times in the neighborhood of e . Thus, at most $O(n)$ vertices are created by intersecting chords within any face bubble.
- **Edge bubbles:** Fix an arbitrary edge e of G . Our construction ensures that that the chords within each the edge bubble of e are polygonal curves, and the number of intersections between two such chords does not exceed the sum of the number of segments in each

The authors have placed this paper in the public domain.

of them. The multiplicity property of the frontier walk W implies that each edge of G is traversed by W at most twice, and therefore each chord sticking to some walk on e consists of at most $O(1)$ segments. It follows that any pair of chords that stick to e intersect $O(1)$ times. We conclude that at most $O(n^2)$ vertices are created by intersecting chords within any edge bubble.

- *Vertex bubbles:* Fix an arbitrary vertex v of G . No translates of λ and ρ intersect the vertex bubble of v . Thus, each chord in the vertex bubble sticks to a walk on the edges of the tiling incident to v , which must have length 2. Our construction ensures that each pair of these chords intersects at most $O(1)$ times. We conclude that at most $O(n^2)$ vertices are created by intersecting chords within any vertex bubble.

This concludes the proof. \square

Summary. We conclude by summarizing our proof of Theorem 1.3. Let γ be a closed curve on an orientable surface without boundary. If γ is not yet tightened, Lemma 4.2 implies that after at most one $0 \rightarrow 1$ move (see Figure 4.2), γ contains at least one basic singular bigon. By Lemma 5.2, we can decrease the number of vertices of γ by two by removing one basic singular bigon. After $O(n)$ such bigon removals, all the excess intersections of γ must have been removed. We conclude that γ can be tightened using at most $O(n^4)$ homotopy moves.

Acknowledgments. The authors would like to express gratitude to Schloss Dagstuhl Leibniz-Zentrum für Informatik and the organizers of Seminar 17072, where they were brought together and work on the upper bounds was initiated. We also thank Erin W. Chambers, Gregory R. Chambers, Anne Driemel, Brittany Terese Fasy, Jessica S. Purcell, and Birgit Vogtenhuber for helpful discussions during the initial stage of this work. We also thank the anonymous referees for helpful feedback on the paper.

References

- [1] James W. Alexander. Combinatorial analysis situs. *Trans. Amer. Math. Soc.* 28(2):301–326, 1926.
- [2] James W. Alexander and G. B. Briggs. On types of knotted curves. *Ann. Math.* 28(1/4):562–586, 1926–1927.
- [3] Sigurd Angenent. Parabolic equations for curves on surfaces: Part II. Intersections, blow-up and generalized solutions. *Ann. Math.* 133(1):171–215, 1991.
- [4] Chris Arettines. A combinatorial algorithm for visualizing representatives with minimal self-intersection. *J. Knot Theory Ramif.* 24(11):1550058–1–1550058–17, 2015. arXiv:1101.5658.
- [5] Eric K. Babson and Clara S. Chan. Counting faces of cubical spheres modulo two. *Discrete Math.* 212(3):169–183, 2000. arXiv:9811085v1.
- [6] Thomas Banchoff. Critical points and curvature for embedded polyhedra. *J. Diff. Geom.* 1:245–256, 1967.
- [7] Edward A. Bender and E. Rodney Canfield. The asymptotic number of rooted maps on a surface. *J. Comb. Theory Ser. A* 43:244–257, 1986.
- [8] Erin W. Chambers and David Letscher. On the height of a homotopy. *CCCG*, vol. 9, 103–106, 2009.
- [9] Erin W. Chambers and David Letscher. Erratum for on the height of a homotopy, 2010. (<http://mathcs.slu.edu/~chambers/papers/hherratum.pdf>).
- [10] Hsien-Chih Chang and Jeff Erickson. Lower bounds for planar electrical reduction. Submitted, 2017.
- [11] Hsien-Chih Chang and Jeff Erickson. Untangling planar curves. *Discrete Comput. Geom.* 58(4):889–920, 2017.
- [12] Moira Chas and Fabiana Krongold. An algebraic characterization of simple closed curves on surfaces with boundary. *J. Topol. Anal.* 2(3):395–417, 2010.
- [13] Moira Chas and Fabiana Krongold. Algebraic characterization of simple closed curves via Turaev’s cobracket. *J. Topology* 9(1):91–104, 2015.
- [14] Marshall Cohen and Martin Lustig. Paths of geodesics and geometric intersection numbers: I. *Combinatorial Group Theory and Topology*, 479–500, 1984. *Annals of Math. Studies* 111, Princeton Univ. Press.
- [15] Éric Colin de Verdière and Jeff Erickson. Tightening non-simple paths and cycles on surfaces. *SIAM J. Comput.* 39(8):3784–3813, 2010.
- [16] Max Dehn. Über unendliche diskontinuierliche Gruppen. *Math. Ann.* 71(1):116–144, 1911.
- [17] Vincent Despré and Francis Lazarus. Computing the geometric intersection number of curves. *Proc. 33rd Int. Symp. Comput. Geom.*, 35:1–35:15, 2017. *Leibniz Int. Proc. Informatics* 77. arXiv:1511.09327.
- [18] David B. A. Epstein. Curves on 2-manifolds and isotopies. *Acta Mathematica* 115:83–107, 1966.
- [19] Jeff Erickson. Efficiently hex-meshing things with topology. *Discrete Comput. Geom.* 52(3):427–449. Springer, 2014.
- [20] Jeff Erickson and Sarel Har-Peled. Optimally cutting a surface into a disk. *Discrete Comput. Geom.* 31(1):37–59, 2004.
- [21] Jeff Erickson and Amir Nayyeri. Minimum cuts and shortest non-separating cycles via homology covers. *Proc. 22nd Ann. ACM-SIAM Symp. Discrete Algorithms*, 1166–1176, 2011.
- [22] Jeff Erickson and Kim Whittlesey. Transforming curves on surfaces redux. *Proc. 24th Ann. ACM-SIAM Symp. Discrete Algorithms*, 1646–1655, 2013.
- [23] Thomas A. Feo and J. Scott Provan. Delta-wye transformations and the efficient reduction of two-

The authors have placed this paper in the public domain.

- terminal planar graphs. *Oper. Res.* 41(3):572–582, 1993.
- [24] George K. Francis. The folded ribbon theorem: A contribution to the study of immersed circles. *Trans. Amer. Math. Soc.* 141:271–303, 1969.
- [25] Steve M. Gersten and Hamish B. Short. Small cancellation theory and automatic groups. *Invent. Math.* 102:305–334, 1990.
- [26] Peter Giblin. *Graphs, Surfaces and Homology*, 3rd edition. Cambridge Univ. Press, 2010.
- [27] Daciberg L. Gonçalves, Elena Kudryavtseva, and Heiner Zieschang. An algorithm for minimal number of intersection points of curves on surfaces. *Proc. Seminar on Vector and Tensor Analysis* 26(139–167), 2005.
- [28] Maurits de Graaf and Alexander Schrijver. Making curves minimally crossing by Reidemeister moves. *J. Comb. Theory Ser. B* 70(1):134–156, 1997.
- [29] Matthew A. Grayson. Shortening embedded curves. *Ann. Math.* 129(1):71–111, 1989.
- [30] Branko Grünbaum. *Convex Polytopes*. Monographs in Pure and Applied Mathematics XVI. John Wiley & Sons, 1967.
- [31] Sariel Har-Peled, Amir Nayyeri, Mohammad Salavatipour, and Anastasios Sidiropoulos. How to walk your dog in the mountains with no magic leash. *Discrete Comput. Geom.* 55(1):39–73, 2016.
- [32] Joel Hass and Peter Scott. Intersections of curves on surfaces. *Israel J. Math.* 51:90–120, 1985.
- [33] Joel Hass and Peter Scott. Shortening curves on surfaces. *Topology* 33(1):25–43, 1994.
- [34] Heinz Hopf. Über die Drehung der Tangenten und Sehnen ebener Kurven. *Compositio Math.* 2:50–62, 1935.
- [35] Francis Lazarus and Julien Rivaud. On the homotopy test on surfaces. *Proc. 53rd Ann. IEEE Symp. Foundations Comput. Sci.*, 440–449, 2012. arXiv:1110.4573.
- [36] Martin Lustig. Paths of geodesics and geometric intersection numbers: II. *Combinatorial Group Theory and Topology*, 501–544, 1987. *Annals of Math. Studies* 111, Princeton Univ. Press.
- [37] Roger C. Lyndon. On Dehn’s algorithm. *Math. Ann.* 166:208–228, 1966.
- [38] Roger C. Lyndon and Paul E. Schupp. *Combinatorial Group Theory*. Springer-Verlag, 1977.
- [39] William S. Massey. *A basic course in algebraic topology*. Springer-Verlag, 1991.
- [40] Tahl Nowik. Complexity of planar and spherical curves. *Duke J. Math.* 148(1):107–118, 2009.
- [41] Jane M. Paterson. A combinatorial algorithm for immersed loops in surfaces. *Topology Appl.* 123:205–234, 2002.
- [42] George Pólya. An elementary analogue to the Gauss-Bonnet theorem. *Amer. Math. Monthly* 61:601–603, 1954.
- [43] Kurt Reidemeister. Elementare Begründung der Knotentheorie. *Abh. Math. Sem. Hamburg* 5:24–32, 1927.
- [44] Alexander Schrijver. Decomposition of graphs on surfaces and a homotopic circulation theorem. *J. Comb. Theory Ser. B* 51(2):161–210, 1991.
- [45] Ernst Steinitz. Polyeder und Raumeinteilungen. *Enzyklopädie der mathematischen Wissenschaften mit Einschluss ihrer Anwendungen* III.AB(12):1–139, 1916.
- [46] Ernst Steinitz and Hans Rademacher. *Vorlesungen über die Theorie der Polyeder: unter Einschluß der Elemente der Topologie*. Grundlehren der mathematischen Wissenschaften 41. Springer-Verlag, 1934. Reprinted 1976.
- [47] John Stillwell. *Classical Topology and Combinatorial Group Theory*, 2nd edition. Graduate Texts in Mathematics 72. Springer-Verlag, 1993.
- [48] Klaus Truemper. On the delta-wye reduction for planar graphs. *J. Graph Theory* 13(2):141–148, 1989.
- [49] Gert Vegter. Kink-free deformation of polygons. *Proceedings of the 5th Annual Symposium on Computational Geometry*, 61–68, 1989.