Measuring Similarity Between Curves on 2-Manifolds via Homotopy Area

Erin Wolf Chambers [*] Yusu Wang [†]

Abstract

Measuring the similarity of curves is a fundamental problem arising in many application fields. There has been considerable interest in several such measures, both in Euclidean space and in more general setting such as curves on Riemannian surfaces or curves in the plane minus a set of obstacles. However, so far, efficiently computable similarity measures for curves on general surfaces remain elusive. This paper aims at developing a natural curve similarity measure that can be easily extended and computed for curves on general orientable 2-manifolds. Specifically, we measure similarity between homotopic curves based on how hard it is to deform one curve into the other one continuously, and define this "hardness" as the minimum possible surface area swept by a homotopy between the curves. We consider cases where curves are embedded in the plane or on a triangulated orientable surface with genus g, and we present efficient algorithms (which are either quadratic or near linear time, depending on the setting) for both cases.

[*] Dept. of Math and Computer Science, Saint Louis University, Saint Louis, MO. Research partially supported by NSF grant CCF 1054779.

[†] Dept. of Computer Science and Engineering, The Ohio State University, Columbus, OH 43210.

1 Introduction

Measuring curve similarity is a fundamental problem arising in many application fields, including graphics, computer vision, and geographic information systems. Traditionally, much research has been done on comparing curves embedded in the Euclidean space. However, in many cases it is natural to study curves embedded in a more general space, such as a terrain or a surface.

In this paper, we study the problem of measuring curve similarity on surfaces. Specifically, given two simple homotopic curves embedded on an orientable 2-manifold (including the plane), we measure their similarity by the minimum total area swept when deforming one curve to the other (the "area" of the homotopy between them), and present efficient algorithms to compute this new measure.

Related work. From the perspective of computational geometry, the most widely studied similarity measures for curves is the Fréchet distance. Intuitively, imagine that a man and his dog are walking along two paths with a leash between them. The Fréchet distance between these two paths is the minimum leash length necessary for them to move from one end of the paths to the other end without back-tracking. Since the Fréchet distance takes the "flow" of the curves into account, in many settings it is a better similarity measure for curves than alternatives such as the Hausdorff distance [5, 6].

Given two polygonal curves P and Q with n total edges in $\mathbb{R}^d$, the Fréchet distance can be computed in $O(n^2 \log n)$ time [4]. An $\Omega(n \log n)$ lower bound for the decision problem in the algebraic computation tree model is known [11], and Alt has conjectured that the decision problem is 3SUM-Hard [2]. Recently, Buchin et al. [12] show that there is a real algebraic decision tree to solve the Fréchet problem with sub-quadratic depth, suggesting that perhaps this is not the case. They also give an improved algorithm which runs in $O(n^2\sqrt{\log n}(\log\log n)^2)$ time. Very recently, Agarwal et al. present a novel approach to compute the discrete version of the Fréchet distance between two polygonal curves in sub-quadratic time [1]. This is the first algorithm for any variant of the Fréchet distance to have a sub-quadratic running time for general curves. No previous algorithm, *exact or approximate*, with running time $o(n^2)$ is known for general curves, although sub-quadratic approximation algorithms for special families of curves are known [6, 7, 22].

While the Fréchet distance is a natural curve similarity measure, it is sensitive to outliers. Variants of it, such as the summed-Fréchet distance, and the partial Fréchet similarity, have been proposed [13, 14, 23], usually at the cost of further increasing the time complexity.

The problem of extending and computing the Fréchet distance to more general metric space has also received much attention. Geodesic distance between points is usually considered when the underlying domain is not $\mathbb{R}^d$. For example, Maheshwari and Yi [28] computed the geodesic Fréchet distance between two polygonal paths on a convex polytope in roughly $O(n^3K^4\log(Kn))$ time, where n and K are the complexity of the input paths and of the convex polytope, respectively. Raichel and Har-Peled consider approximating the weak Fréchet distance between simplicial complexes in $\mathbb{R}^d$ [26]. Geodesic Fréchet distance between polygonal curves in the plane within a simple polygon has been studied in [8, 18, 24].

One issue with generalizing Fréchet distance directly to surfaces is that the underlying topology is not taken into account; for example, in geodesic Fréchet distance, while the length of the leash varies continuously, the actual leash itself does not. As a result, several measures of similarity have been proposed which take the underlying topology into account. Chambers et al. [15] proposed the so-called homotopic Fréchet distance and gave a polynomial (although not efficient) algorithm for when the curves reside in a planar domain with a set of polygonal obstacles. The extra requirement for this homotopic Fréchet distance is that the leash itself and not just its length has to vary in a continuous manner, essentially restricting the homotopy class which the leash is in. A stronger variant called isotopic Fréchet distance has also been proposed and investigated, although no algorithms at all are known to even approximate this parameter [16].

Orthogonal to homotopic Fréchet distance is the concept of the *height* of a homotopy; instead of minimizing the maximum leash length, this measure views the homotopy as tracing a way for the first curve to deform to the second curve, where the goal is to minimize the longest intermediate curve length. Introduced independently in two very different contexts [10, 17], it is not even known if the problem is in NP.

Recent work on approximating the homotopy height and the homotopic Fréchet distance has yielded efficient $O(\log n)$ approximation algorithms for both of these problems [25]. However, exact algorithms on surfaces for either problem are still unknown.

New work. In this paper, we develop a natural similarity measure for curves on general surfaces that can be computed both quickly and exactly. Intuitively, we measure distances between homotopic curves based on how hard it is to deform one curve into the other one, and define this "hardness" as the minimal total surface area swept by a homotopy between them. Our similarity measure is natural, and robust against noise (as the area captures average, instead of maximum, deviation from one curve to the other). To the best of our knowledge, this is the first similarity measure for curves on general surfaces with efficient polynomial-time algorithms to compute it exactly.

It is worth noting that this definition in essence combines homotopic Fréchet distance with homotopy height; those measures compute the "width" and "height" of the homotopy, while our measure calculates the total area. It is thus surprising that while no exact algorithms are known for either of those measures on surfaces, we are able to provide a polynomial running time for computing the area of a homotopy.

We consider both cases where curves are embedded in the plane, or on a closed, triangulated orientable surface with genus g. For the former case, our algorithm runs in $O(n + I^2 \log n)$ time, where n is the total complexity of input curves and I is the number of intersections between them. The running time is thus *linear* in n if there are a constant number of intersection points. On a surface, if the input is a triangulation of complexity N, then our algorithm runs in time $O(I^2 \log I + n \log n + n \log g + N)$. While our similarity measure is more expensive to compute than the Fréchet distance when $I = \omega(\sqrt{n})$, one major advantage is that this measure can be computed on general orientable surfaces efficiently. In fact, the ideas and algorithms behind the planar case form the foundation for the handling of the case on general surfaces.

The main ideas behind our approach are developed by examining some properties of one natural class of homotopies, including a relation with the winding number of a closed curve. Specifically, the use of the winding number enables us to compute the optimal homotopy area efficiently in the plane. This forms the basis of our dynamic programming framework to compute similarity between curves in the plane. We also show how to build efficient data structures to keep the total cost of the dynamic program low.

For the case where the underlying surface is a topological sphere, we extend the winding number in a natural way and show how to adapt our planar algorithm without additional blow-up in the time complexity. For the case when the surface has non-zero genus, we must extend our algorithm to run efficiently in the universal cover (which is homeomorphic to the plane) by using only a small portion of it.

We remark that the idea of measuring deformation areas has been used before in practice [19, 29]. For example, similarity between two convex polygons can be measured by their symmetric difference [3, 35]. In [9], the area sandwiched between a *x-monotone curve* and another curve is used to measure their similarity. However, "area" between general curves and its computation have not been investigated.

2 Definitions and Background

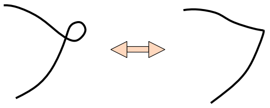
Paths and cycles. We will assume that we are working on an orientable 2-manifold M (which could be the plane). A *curve* (or a *path*) on a surface M is a map $P : [0, 1] \to M$; a *cycle* (or a *loop*) is a continuous map $\gamma : S^1 \to M$ where S^1 is the unit circle. A curve P or a cycle γ is *simple* if $P(t_1) \neq P(t_2)$ (resp. $\gamma(t_1) \neq \gamma(t_2)$) for any $t_1 \neq t_2$.

For a loop γ in the plane, we will also be concerned with the *winding number* of γ at a point $x \in \mathbb{R}^2$. Intuitively, imagine starting from a point y on Γ, and connecting x and y by a string. $\mathrm{wn}(x; \Gamma)$ is an integer measuring how many times this string winds, in a clockwise manner, around x as y traverses Γ. More formally, consider an infinite ray f based at x which is generic (meaning it has a finite set of intersections with γ). An intersection point $\gamma(t)$ between r and γ is a *crossing* if for all sufficiently small ϵ, the points $\gamma(t-\epsilon)$ and $\gamma(t+\epsilon)$ lie on opposite sides of r. A crossing is positive if the triangle through points 0, $\gamma(t)$, and $\gamma(t+\epsilon)$ is oriented counterclockwise, and is negative if oriented clockwise. Then $\mathrm{wn}(x;\gamma)$ is the number of positive crossings minus the number of negative crossings with respect to any generic ray.

Homotopy A *homotopy* between two paths P and Q (with the same endpoints) is a continuous map $H : [0,1]\times[0,1] \to M$ where $H(0,\cdot) = P$, $H(1,\cdot) = Q$, $H(\cdot,0) = P(0) = Q(0)$ and $H(\cdot,1) = P(1) = Q(1)$. A homotopy describes a continuous deformation between the two paths or curves: for any value $t \in [0,1]$, we let $H_t = H(t,\cdot)$ to be the intermediate curve at time t, so that $H_0 = P$ and $H_1 = Q$.

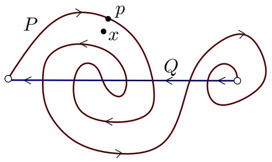
We define the *area* of a homotopy H to be the total area covered by the image of the homotopy on the surface, where an area that is covered multiple times will be counted with multiplicity. More precisely, given a homotopy H whose image is piecewise differentiable, $\text{Area}(H) = \int_{s\in[0,1]}\int_{t\in[0,1]} \left|\frac{dH}{ds} \times \frac{dH}{dt}\right| dsdt$. The *minimum homotopy area* between P and Q is the infimum of the areas of all homotopies between P and Q, denoted by $\sigma(P,Q)$. If such an infimum does exist, we call the homotopy which produces the minimum area the *optimal homotopy*.

We note that it is not immediately clear that this value exists, depending on the curves and underlying homotopy. Minimum area homotopies were considered by Douglas [21] and Rado [31] in the context of Plateau's problem; they noted that not only is the integral improper in general, but the infimum itself may not be continuous. The eventual proof that these exist in $\mathbb{R}^n$ relies on a definition using Dirichlet integrals which ensure (almost) conformal parameterizations of the homotopy. See the book by Lawson [27] for an overview of this result as well as several extensions to minimal area submanifolds in more general settings.



However, beyond a proof of existence, we are interested in computing such homotopies, or at least measuring their actual area, in much simpler settings such as $\mathbb{R}^2$ or a surface. To this end, we restrict the input curves to be *simple* curves which consist of a finite number of piecewise analytic components. We also need H to be continuous and piecewise differentiable, so that the integral can be defined. Finally, we will also require that at any time t, the intermediate curve H_t is *regular* (see [36] for smooth curves and [30] for piecewise-linear curves). Intuitively, this means that the deformation is "kink"-free [30], and cannot create or destroy a local loop as shown in the right figure (the singular point in the right curve is a kink).

Decomposing arrangements. Consider two simple piecewise analytic curves P and Q with the same endpoints. Their concatenation forms a (not necessarily simple) closed curve denoted by $C = P \circ Q$. Let $\mathrm{Arr}(C)$ denote the arrangement formed by C, where vertices in $\mathrm{Arr}(C)$ are the intersection points between P and Q. An edge / arc in $\mathrm{Arr}(C)$ is a subcurve of either P or Q.



We give C (and thus P and Q) an arbitrary orientation. Hence we can talk about the *sidedness* with respect to C at a point $p \in P$. Specifically, a point $x \in \mathbb{R}^2$ is to the *right* of C at p if it is a counter-clockwise turn from the orientation of the vector px the orientation of (tangent of) C at p (see the right figure for an example). Given two oriented curves γ_1 and γ_2, an intersection point p of them is *positive* if it is a counter-clockwise turn from the orientation of γ_1 to that of γ_2 at p. (Note that this also agrees with our crossing definition in the context of winding numbers.) For a curve γ and a point $x \in \gamma$, the *index of* x is the parameter of x under the arc-length parameterization of γ. We sometimes use x to represent its index along γ when its meaning is clear from the context. Given two points $x, y \in \gamma$, we will use $\gamma[x,y]$ to denote the unique sub-curve of γ between points x and y.

We say that a homotopy H from P to Q is *right sense-preserving* if for any $t, s \in [0,1]$, we have that either $H_{t+dt}(s) = H_t(s)$ or $H_{t+dt}(s)$ is to the right of the oriented curve H_t at $H_t(s)$. If it is the former case, then we say that $p = H_t(s)$ is a *fixed point* at time t. Similarly, we say that H is *left sense-preserving* if for any $t, s \in [0,1]$, $H_t(s)$ is either a fixed point or deforms to the *left* of the curve H_t. Our homotopy H is *sense-preserving* if it is either right or left sense-preserving. The sense-preserving property means that we can continuously deform the curve P always in the same direction, without causing local folds in the regions swept. Intuitively, any optimal homotopy should have this property to some extent, which we will make more precise and prove later.

3 Structure of Optimal Homotopies

Given two simple curves P and Q (with the same end points) embedded on an orientable 2-manifold M, let $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_I\}$ denote the set of I intersection points between them, sorted by their order along P. Given a homotopy H from P to Q, a point $p \in M$ is called an *anchor point with respect to* H if it remains on $H(t, \cdot) = H_t$ at all times $t \in [0,1]$. If p is an anchor point, then it is necessarily an intersection point between P and Q, as $p \in H_0 = P$ and $p \in H_1 = Q$. We exclude the beginning and ending end points of P and Q from the list of anchor points, as they remain fixed for all homotopys. In what follows, we show that any optimal homotopy can be decomposed by anchor points such that each of the resulting smaller homotopies has a simple structure.

Specifically, consider an arbitrary optimal homotopy H^*. Let $\mathbf{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$ be the set of anchor points with respect to H^*, the minimum area homotopy. We order the $\mathbf{b}_i$'s by their indices along P. It turns out that the order of their indices along Q is the same, and the proof of this simple observation is in Appendix A.

Observation 3.1 *The order of* $\mathbf{b}_i$*'s along* P *and along* Q *are the same.*

This observation implies that we can decompose H^* into a list of sub-homotopys, where H_i^* morphs $P[\mathbf{b}_i, \mathbf{b}_{i+1}]$ to $Q[\mathbf{b}_i, \mathbf{b}_{i+1}]$. Obviously, each H_i^* is necessarily optimal, and it induces no anchor points. The following result states that an optimal homotopy without anchor points has a simple structure, which is sense-preserving. Intuitively, if any point changes its deformation direction at any moment, the deformation will sweep certain redudant area and thus cannot be optimal. The detailed proof is in Appendix B.

Lemma 3.2 *If an optimal homotopy* H *from* P *to* Q *has no anchor points, then it is sense-preserving.*

4 Minimum Area Homotopies In The Plane

In this section, we consider the case where the input consists of two simple polygonal curves in the plane. We develop an algorithm to compute the similarity between P and Q in $O(I^2 \log I + n \log n)$ time, where n is the total complexity of input curves and I is the number of intersections. Note that $I = \Theta(n^2)$ in worst case. Although efficient algorithms for comparing curves in the plane exist (such as the Fréchet distance), our planar algorithm will be the fundamental component for comparing curves on general surfaces in the next section. It turns out that our approach can easily be extended to measure similarity between simple cycles in the plane; see Appendix F.

4.1 Relations to Winding Numbers

We are given two simple (open) curves in the plane which share common endpoints. Previously, we have shown that if an optimal homotopy does not induce anchor points, then it is sense-preserving. The implication of this result is manifested by using the *winding number*.

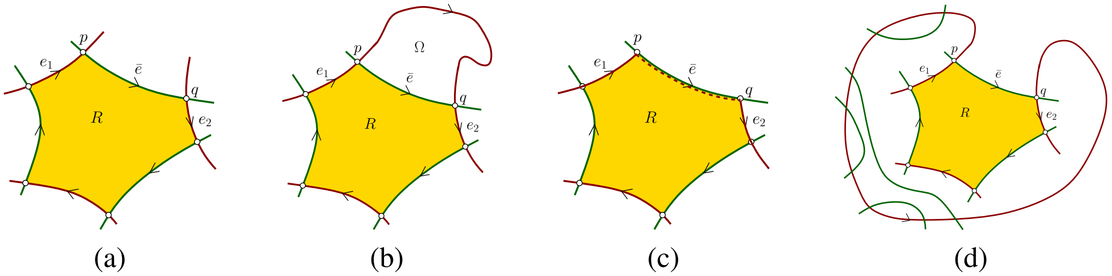
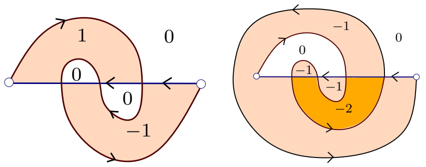


Figure 1: (a) The cell R with highest positive winding number. It boundary consists of alternating P-arcs (red) and Q-arcs (green). The two cases of relations between $P[p, q]$ and R are shown in (b) and (d), respectively. For case (b), we can deform P to sweep through Ω as shown in (c), and reduce the number of intersections by 2.



We say an oriented curve Γ has *consistent winding numbers* if $\mathrm{wn}(x, \Gamma)$ is either all non-negative, or all non-positive, for all $x \in \mathbb{R}^2$. Note that for a curve with consistent winding numbers, we can always orient the curve appropriately so that $\mathrm{wn}(x, \Gamma)$ is all non-negative. Two examples are shown in the figure on the right, where the second example has consistent winding numbers. Let $\mathrm{Arr}(\Gamma)$ denote the arrangement formed by the curve Γ. All points in the same cell of the arrangement $\mathrm{Arr}(\Gamma)$ of Γ have the same winding number, and the winding numbers of two neighboring cells differ by 1. The relation of consistent winding numbers and sense-preserving homotopys is given below, and the proof can be found in Appendix C.

Lemma 4.1 *If there is a sense-preserving homotopy H from P to Q, then the closed curve $P \circ Q$ has consistent winding numbers.*

Next, we describe two results to connect the above lemma to the computation of optimal homotopy. First, we define the *total winding number* $\mathrm{Tw}(\gamma)$ of a curve γ as $\mathrm{Tw}(\gamma) = \int_{\mathbb{R}^2} \mathrm{wn}(x; \gamma) d\nu(x)$, where $d\nu(x)$ is the area form[1]. The following observation is straightforward.

Observation 4.2 *For any P and Q in the plane, $\sigma(P, Q) \geq |\mathrm{Tw}(P \circ Q)|$.*

Proof: Take any regular homotopy H from P to Q. The area of a regular homotopy H in our setting can be reformulated as an integral on the image domain as $Area(H) = \int_{\mathbb{R}^2} deg_H(x) d\nu(x)$, where $deg_H(x)$ is defined as the number of connected components in the pre-image of x under H. In other words, $deg_H(x)$ is the number of times that any intermediate curve H_t sweeps through x. Now consider the function $F : [0, 1] \to \mathbb{R}$ defined as $F(t) = \mathrm{Tw}(H_t^* \circ Q)$. Obviously, $F(0) = \mathrm{Tw}(P \circ Q)$, $F(1) = 0$, and F is a continuous function. Furthermore, each time the winding number at a point x changes by 1 for some $t \in [0, 1]$, it means that some intermediate curve $H(t)$ sweeps through it. Hence $|\mathrm{wn}(x)|$ is a lower bound for $deg_H(x)$. We thus have that $|\mathrm{Tw}(P \circ Q)| \leq \int_{\mathbb{R}^2} |\mathrm{wn}(x)| d\nu(x) \leq \int_{\mathbb{R}^2} deg_H(x) d\nu(x)$ for any regular homotopy H, implying that $|\mathrm{Tw}(P \circ Q)| \leq \inf_H Area(H) = \sigma(P, Q)$. ■

Lemma 4.3 *Given P and Q, if $\Gamma = P \circ Q$ has consistent winding numbers, then $\sigma(P, Q) = |\mathrm{Tw}(\Gamma)|$.*

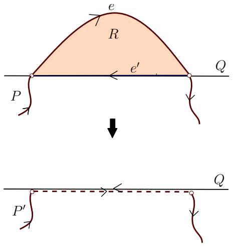
Proof: We provide a sketch of the proof here to illustrate the main idea; see Appendix D for the full proof. We prove the claim by induction on the number of intersections between P and Q. The base case is when there is no intersection between P and Q. In this case, Γ is a Jordan curve which decomposes $\mathbb{R}^2$ into two regions, one inside Γ and one ouside. The optimal homotopy area $\sigma(P, Q)$ in this case is the area of the

[1] Note that this allows us to use any Riemannian metric on the plane (including the standard Euclidean metric). This will be necessary later when we use the same algorithm to curves in a universal covering space.

bounded cell. By orienting Γ appropriately, every point in the bounded cell has winding number 1 and the claim follows.

Now assume that the claim holds for cases with at most $k-1$ intersections. We now prove it for the case with k intersections. Let an *X-arc* denote a subcurve of curve X. Consider the arrangement $\mathrm{Arr}(\Gamma)$ formed by $\Gamma = P \circ Q$. Since P and Q are simple, every cell in this arrangement has boundary edges alternating between P-arcs and Q-arcs. Assume without loss of generality that Γ has all non-negative winding numbers. Consider a cell $R \in \mathrm{Arr}(\Gamma)$ with largest (and thus positive) winding number. Since its winding number is greater than all its neighbors, it is necessary that all boundary arcs are oriented consistently as shown in Figure 1 (a), where the cell R (shaded region) lies to the right of its boundary arcs.

If R has only two boundary arcs, e from P and e' from Q, respectively, then we can morph P to another simple curve P' by deforming e through R to $-e'$ (where '$-$' means reversing the orientation) as illustrated on the right. The area swept by this deformation is exactly the area of cell R. Furthermore, after the deformation, every point $x \in R$ decreases their winding number by 1, and no other point changes its winding number. Since points in this cell initially has strictly positive winding number, the resulting curve $\Gamma' = P' \circ Q$ still has all non-negative winding number. The number of intersections between P' and Q is $k-2$. By induction hypothesis, $\sigma(P', Q) = \mathrm{Tw}(\Gamma')$. Since $\mathrm{Tw}(\Gamma) - \mathrm{Tw}(\Gamma') = \mathrm{Area(R)}$, we have that $\mathrm{Tw}(\Gamma) = \sigma(P', Q) + \mathrm{Area(R)}$. It then follows from Observation 4.2 and the fact $\sigma(P, Q) \le \sigma(P', Q) + \mathrm{Area(R)}$ that $\sigma(P, Q) = \mathrm{Tw}(\Gamma)$.



Otherwise, the cell R has more than one P-arc. Take the P-arc e_1 with the smallest index along P, and let p be its ending point. Let e_2 be the next P-arc along the boundary of R, and q its starting point, and $Q[p, q]$ the Q-arc between e_1 and e_2, denoted by $\bar{e}$ in Figure 1. $P[p, q]$ and $-Q[p, q]$ bound a simple polygon, which we denote by Ω. Since Ω does not intersect R, either Ω is on the opposite side of the Q-arc $\bar{e}$ from the interior of R (Figure 1 (b)), or they are on the same side (Figure 1 (d)). It turns out that in both cases, we can deform P to another simple curve P' such that (i) the number of intersections is reduced by 2, and (ii) $P' \circ Q$ still has consistent winding numbers. For example, in the case of Figure 1 (b), P is then deformed to sweep the region Ω as shown in Figure 1 (c). By applying the induction hypothesis to $P' \circ Q$, we are able to obtain that $\sigma(P, Q) = \mathrm{Tw}(\Gamma)$. The details are in Appendix D. ■

4.2 The Algorithm

Lemma 3.2 and 4.1 imply that if the closed curve $P \circ Q$ produces both positive and negative winding numbers, then any optimal homotopy from P to Q must have at least one anchor point. On the other hand, if it has consistent winding numbers, then by Lemma 4.3 we can compute the optimal cost to deform them by simply computing the total winding number. This leads to a simple dynamic-programming (DP) approach to compute $\sigma(P, Q)$.

Specifically, let $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_I$ denote the intersection points between P and Q, ordered by their indices along P, where $\mathbf{x}_0$ and $\mathbf{x}_I$ are the beginning and ending points of P and Q, respectively. Let $T(i)$ be the cost of the optimal homotopy between $P[\mathbf{0}, \mathbf{x}_i]$ and $Q[\mathbf{0}, \mathbf{x}_i]$, and $C[i, j]$ the closed curve formed by $P[\mathbf{x}_i, \mathbf{x}_j] \circ Q[\mathbf{x}_i, \mathbf{x}_j]$. We say that a pair of indices (i, j) is *valid* if (1) $\mathbf{x}_i$ and $\mathbf{x}_j$ have the same order along P and along Q; and (2) the closed curve $C[i, j]$ has consistent winding numbers. We have the following recursion:

$$T(i) = \begin{cases} \mathrm{Tw}(C[0, i]), & \text{if } C[0, i] \text{ has consistent winding numbers} \\ \min_{j<i \text{ and } (j,i) \text{ is valid}} \{\ \mathrm{Tw}(C[j, i]) + T(j)\ \}, & \text{otherwise} \end{cases}$$

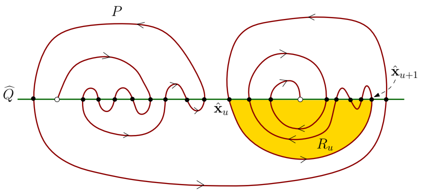
Time complexity. The main components of the above DP framework are (i) to compute $\mathrm{Tw}(C[i, j])$ for all pairs of i, js, and (ii) to check whether each pair (i, j) is valid or not. These can be done in $O(I^2 n)$

total time in a straightforward manner. We now show how to compute them in $O(I^2 \log I)$ time after $O(I \log I + n \log n)$ pre-processing time. Specifically, we describe how to compute such information in $O(I \log I)$ time for all $C[r, i]$s for a fixed $r \in [1, I]$ and all indices $i > r$.

To simplify the description of the algorithm, we extend Q on both sides until infinity, and obtain $\hat{Q}$. Now collect all intersection points between P and $\hat{Q}$, $\{\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_I\}$, which is a super-set of previous intersection points, and sort them by their order along the curve P. The algorithm can be made to work with Q directly, but using $\hat{Q}$ makes the intuition behind our algorithm, as well as its description, much more clear.

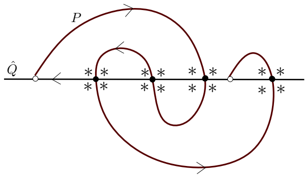
Note that $\hat{Q}$ divides the plane to two half-planes. For illustration purpose, we will draw $\hat{Q}$ as a horizontal line, and use the upper and lower half-planes to refer to these two sides of $\hat{Q}$. Another way to see that regarding $\hat{Q}$ as a horizontal line does not cause any loss of generality is that one can always find a homeomorphism from $\mathbb{R}^2 \to \mathbb{R}^2$ such that the image of $\hat{Q}$ is a horizontal line under this homeomorphism.

Now for a fixed integer r, we traverse P starting from $\hat{\mathbf{x}}_r$. We aim to maintain appropriate data structures so that each time we pass through an intersection point $\hat{\mathbf{x}}_i$ with $\hat{Q}$, we can, in $O(\log I)$ time, (1) check whether (r, i) is valid, and (2) obtain total winding number for $C[r, i]$.



Total winding numbers. We first explain how to maintain the total winding number for the closed curve $C[r, i] = P[\hat{\mathbf{x}}_r, \hat{\mathbf{x}}_i] \circ Q[\hat{\mathbf{x}}_i, \hat{\mathbf{x}}_r]$ as i increases. Assume i changes from u to $u + 1$. Since $\hat{\mathbf{x}}_u$ and $\hat{\mathbf{x}}_{u+1}$ are two consecutive intersection points along P, the arcs $P[\hat{\mathbf{x}}_u, \hat{\mathbf{x}}_{u+1}]$ and $\hat{Q}[\hat{\mathbf{x}}_u, \hat{\mathbf{x}}_{u+1}]$ form a simple closed polygon which we denote by R_u (shaded region in the right figure). Comparing the arrangement $\mathrm{Arr}(C[r, u{+}1])$ with $\mathrm{Arr}(C[r, u])$, regardless of where r is, only points within R_u will change their winding number, either all by $+1$ or all by -1, depending on whether R_u is to the right side or the left side of the P-arc $P[\hat{\mathbf{x}}_u, \hat{\mathbf{x}}_{u+1}]$, respectively. The winding numbers for points outside R_u are not affected. Hence the change in the total winding number is simply $\alpha_u \mathrm{Area(R_u)}$, where α_u is either $+1$ or -1. See the previous figure, where all points in R_u will decrease their winding number by 1 as we move from $C[r, u]$ to $C[r, u{+}1]$. We can pre-compute the area of R_u's for all u in $O(n \log n + I \log I)$ time, by observing that the set of R_us satisfy the parentheses property: Namely, either R_u and R_v are disjoint in their interior, or one contains the other. The details can be found in Appendix E. With the area of R_us known, updating the total winding number from $C[r, u]$ to $C[r, u + 1]$ takes only constant time.

Checking the validity of (r, i)s. To check whether (r, i) is valid or not, we need to check whether all cells in the arrangement $\mathrm{Arr}(C[r, i])$ have consistent winding numbers. First observe that for any r and i, $\mathrm{Arr}(P + \hat{Q})$ is a refinement of the arrangement $\mathrm{Arr}(C[r, i])$. That is, a cell in $\mathrm{Arr}(P + \hat{Q})$ is always contained within some cell in $\mathrm{Arr}(C[r, i])$. Hence all points within the same cell of $\mathrm{Arr}(P + \hat{Q})$ always have the same winding number, and we simply need one point from each cell in $\mathrm{Arr}(P + \hat{Q})$ to maintain the winding number for all cells in $\mathrm{Arr}(C[r, i])$, for any r and i. We now describe how to maintain the winding number for cells of $\mathrm{Arr}(P + \hat{Q})$ (thus for $\mathrm{Arr}(C[r, u])$s) as we pass each $u > r$, so that we can check whether $C[r, u]$ has consistent winding numbers or not efficiently.



To this end, take four points around each intersection point $\hat{\mathbf{x}}_i$ of P and $\hat{Q}$ (shown as stars in the right figure). The collection of such *representative points* hit all cells in $\mathrm{Arr}(P + \hat{Q})$. (It does not matter whether there may be more than one point taken from a cell of $\mathrm{Arr}(P + \hat{Q})$.) Hence $\mathrm{Arr}(C[r, i])$ has consistent winding number if and only if all these representative points have consistent winding numbers. Next, we build a data structure to maintain the winding numbers for these points as i increases. Specifically, let U be the set of representatives that are to the right of $\hat{Q}$, which are the stars above $\hat{Q}$ in the right figure. (Those to the left of $\hat{Q}$ will be handled in a symmetric manner). Each point has a key associated with it which is its index along $\hat{Q}$. We build a standard balanced 1-D range tree on U based on such keys, where each leaf f stores a point from U. Every internal node v is associated with an interval

$[l_v, r_v]$, where l_v and r_v are the smallest and largest keys stored in the subtree rooted at v. In other words, all representatives with an index along $\hat{Q}$ within $[l_v, r_v]$ are stored in the subtree rooted at v. At every node v, interior or not, we also store a value $addW_v$. To compute the winding number for the representative point p_f stored at a leaf node f, we identify the path $\{v_0, v_1, \ldots, v_a = f\}$ from the root v_0 to f. The winding number for p_f is simply $\sum_{i=0}^{a} addW_{v_i}$. Finally, each internal node v also stores the maximum and minimum winding numbers associated with all leaves in its subtree. At the beginning, all winding numbers are zero. The size of this tree is $O(I)$ with height $O(\log I)$, and can be built in $O(I \log I)$ time once the arrangement $\mathrm{Arr}(P + \hat{Q})$ is known.

Let $\mathbf{q}_i$ denote the index of point $\hat{\mathbf{x}}_i$ along $\hat{Q}$ (or can be considered as the x-coordinate of $\hat{\mathbf{x}}_i$). As we move from $C[r, u]$ to $C[r, u+1]$, cells of $\mathrm{Arr}(P + \hat{Q})$ contained in R_u should either all increase or all decrease their winding number by 1. Note that representatives of these cells are simply those contained in the interval $[\mathbf{q}_u, \mathbf{q}_{u+1}]$ (or $[\mathbf{q}_{u+1}, \mathbf{q}_u]$ if $\mathbf{q}_{u+1} < \mathbf{q}_u$). Hence updating the winding number is similar to an interval query of $[\mathbf{q}_u, \mathbf{q}_{u+1}]$, and the $O(\log I)$ number of nodes in the canonical decomposition of $[\mathbf{q}_u, \mathbf{q}_{u+1}]$ update their $addW_v$ values by either $+1$ or -1 depending on the sideness of R_u with respect to the arc $P[\hat{\mathbf{x}}_u, \hat{\mathbf{x}}_{u+1}]$. The minimum and maximum winding numbers can also be updated $O(1)$ time per visited node. The entire process visits $O(\log I)$ nodes, and thus takes $O(\log I)$ time as i increases from u to $u+1$. To see whether $C[r, u+1]$ has consistent winding numbers or not, we only need to check the minimum and maximum winding numbers stored at the root of the tree, denoted by w_{min} and w_{max}, respectively. If $w_{min} \times w_{max}$ equals to zero, then all winding numbers w.r.t. $C[r, u+1]$ are either all non-negative or all non-positive. Otherwise, $(r, u+1)$ is not valid.

Repeat the above process for every $r \in [1, I]$. Overall, after $O((n+I)\log n)$ pre-processing, we can check whether (r, i) is valid or not and compute $\mathrm{Tw}(C[r, i])$ for all $r \in [1, I]$ and all $i > r$ in $O(I^2 \log I)$ time. Putting everything together, we have the following result.

Theorem 4.4 *Given two simple polygonal chains P and Q (with the same endpoints) in the plane of n total complexity, and with I intersection points between them, we can compute the optimal homotopy and its area in $O(I^2 \log I + n \log n)$ time and $O(I^2 + n)$ space.*

The case where we have two simple cycles P and Q in the plane is discussed in Appendix F, and we obtain the following extension:

Corollary 4.5 *Given two polygonal cycles P and Q in the plane of n total complexity and with I intersection points, we can compute the optimal homotopy and its area in $O(I(I^2 \log I + n \log n))$ time if $I > 0$; and compute the optimal homotopy area in $O(n \log n)$ time if $I = 0$.*

5 Minimum Area Homotopies on 2-Manifolds

In this section, we consider optmial homotopy between curves P and Q on an orientable and triangulated 2-manifold M without boundary. Our input is a triangulation K of M with complexity N, and two simple homotopic polygonal curves P and Q sharing endpoints. Edges in P and Q are necessarily edges from the triangulation K. The total complexity of P and Q is n, and there are I number of intersections between them. Note that in this setting, $I = O(n)$. Below we discuss separately the cases when M has non-zero genus and when M is a topological sphere.

5.1 Curves on a Surface with Non-zero Genus

Given an orientable 2-manifold M, let $\mathcal{U}(M)$ be a universal covering space of M with $\phi : \mathcal{U}(M) \to M$ the associated covering map. Note that ϕ is continuous, surjective, and a local homeomorphism. For any path γ in M, if we fix the lift (pre-image) of its starting point, then it lifts to a unique path $\tilde{\gamma}$ in $\mathcal{U}(M)$, such that $\phi(\tilde{\gamma}) = \gamma$ [20, 33]. Since P and Q are homotopic with common endpoints, the closed curve formed by $C = P \circ Q$ is contractible on M, and the lift of C, denoted by $\tilde{C}$, is a closed curve in $\mathcal{U}(M)$. More generally, by the Homotopy Lifting Property of the universal cover [32], we have the following observation:

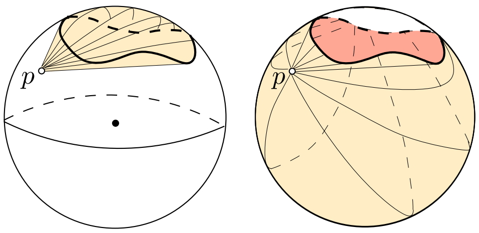
Observation 5.1 *Once we fix the lift of the starting point of P and Q in $\mathcal{U}(M)$, there is a one-to-one correspondence between homotopies between P and Q in M and those between $\tilde{P}$ and $\tilde{Q}$ in $\mathcal{U}(M)$.*

We now impose an area measure in $\mathcal{U}(M)$ by lifting the area measure in M; this can be done via the map ϕ, which is a local homeomorphism. Now the area of a homotopy in M is the same as the area of its lift in $\mathcal{U}(M)$. As such, we can convert the problem of finding an optimal homotopy in M to finding one in $\mathcal{U}(M)$. Furthermore, for any orientable compact 2-manifold with genus $g > 0$, its universal cover is topologically equivalent to $\mathbb{R}^2$. We can thus apply algorithms and results from previous section to the universal covering space. Specifically, given two homotopic paths with n edges from a triangulation K of a surface M with N simplices, the entire algorithm takes $O(I^2 \log I + n \log n + n \log g + N)$ time. Roughly speaking, we construct the portion of the universal covering space that the lift of C will traverse (and enclose in some sense), which consists of $O(n)$ copies of some polygonal schema of M [20]. The main observation is that only the combinatorial structure of $\tilde{C}$ is needed, so we can avoid filling in each copy of the polygonal schema with all triangles from K. Details are in Appendix G.

5.2 The Case of the Sphere

We now consider the remaining case where the input M is a (topological) sphere $\mathbf{S}$ (that is, genus is zero). All paths on $\mathbf{S}$ are homotopic. The universal cover of a sphere is itself, hence compact. However, the previous algorithm in Section 4.2 works for a domain homeomorphic to $\mathbb{R}^2$ and cannot be directly applied. We sketch how we handle the sphere case here; see Appendix H for the full discussion.

We observe that the results in Section 3 still hold. However, as the sphere is compact, the winding number is not well-defined. For example, see the right figure, where there are two ways that the curve γ winds around the point p. In the first case, the winding number is 0, while in the second case, the winding number is -1. In order to use a dynamic programming framework as before to compute the optimal homotopy between P and Q, we need to develop analogs of Lemma 4.1 and 4.3 for curves on the sphere.



To this end, observe that if we remove one point, say $\mathbf{z} \in \mathbf{S}$ from the sphere $\mathbf{S}$, then the resulting space $\mathbf{S_z} = \mathbf{S} - \mathbf{z}$ is homeomorphic to $\mathbb{R}^2$, and the concept of the winding number is well defined for $\mathbf{S_z}$. Specifically, $\mathbf{z}$ can be considered as the point of infinity in $\mathbb{R}^2$. The *winding number of $x \in \mathbf{S_z}$ w.r.t. C and z*, denoted by $\mathrm{wn}(x; \mathbf{z}, C)$ (C omitted when its choice is clear), is simply the summation of signed crossing number for any path connecting x to $\mathbf{z}$. As in the planar case, we say that C is *consistent w.r.t.* $\mathbf{z}$ if $\mathrm{wn}(x; \mathbf{z}, C)$ is either non-negative, or non-positive for all $x \in \mathbf{S_z}$. Let $\sigma(P, Q; \Omega)$ denote the best cost to morph P to Q within domain Ω.

Observation 5.2 *If there is an optimal homotopy between P and Q that does not sweep through some point* $\mathbf{z}$*, then $\sigma(P, Q; \mathbf{S}) = \sigma(P, Q; \mathbf{S_z})$.*

Observation 5.3 *Suppose H^* is an optimal homotopy between P and Q with no anchor points. For any cell R in $\mathrm{Arr}(P + Q)$, if H^* sweeps through one point in its interior, then it sweeps through all points in R.*

The simple proof for the above observation is in Appendix H.1. The key result is the following lemma, the proof of which can be found in Appendix H.2.

Lemma 5.4 *If there is an optimal homotopy H^* of P and Q with no anchor point, then the image of this optimal homotopy cannot cover all points in* $\mathbf{S}$.

Given two homotopic paths P and Q from $\mathbf{S}$ sharing common endpoints, Lemma 5.4 and Observation 5.2 imply that if P can be morphed to Q optimally without anchor points, then there exists some point $\mathbf{z} \in \mathbf{S}$ such that $\sigma(P, Q; \mathbf{S}) = \sigma(P, Q; \mathbf{S}_\mathbf{z})$. For this choice of $\mathbf{z}$, it is necessary that the closed curve $P \circ Q$ has consistent winding numbres. Once this $\mathbf{z}$ is identified, $\sigma(P, Q; \mathbf{S}_\mathbf{z})$ is simply the total winding number of $P \circ Q$ w.r.t. $\mathbf{z}$, as suggested by Lemma 4.3, because $\mathbf{S}_\mathbf{z}$ is homeomorphic to the plane. Furthermore, by Observation 5.3, we only need to pick one point from each cell of $\mathrm{Arr}(P+Q)$ to check for the potential $\mathbf{z}$. Specifically, let $\{\mathbf{z}_1, \ldots, \mathbf{z}_l\}$ be a set of such *representatives*, where $l = O(I)$. The optimal homotopy area $\sigma(P, Q)$ is simply the smallest of all $\mathrm{Tw}(P \circ Q; \mathbf{z}_i)$ for those $\mathbf{z}_i$s with respect to whom the curve $P \circ Q$ has consistent winding numbers. Hence if we assume that if there is an *optimal* homotopy between P and Q with *no* anchor points, then the results in this section provide an algorithm to compute $\sigma(P, Q)$.

The algorithm for sphere. To compute the optimal homotopy between P and Q, we follow the same dynamic programming framework as before. If there is no anchor point in an optimal homotopy, then we use the discussion above to compute the optimal homotopy area. Otherwise, we identify the intersection point that serves as next anchor point, and recurse. The main difference lies in the component of computing $\sigma(i, j) := \sigma(P[\mathbf{x}_i, \mathbf{x}_j], Q[\mathbf{x}_i, \mathbf{x}_j])$, assuming that there is an optimal homotopy from $P' = P[\mathbf{x}_i, \mathbf{x}_j]$ to $Q' = Q[\mathbf{x}_i, \mathbf{x}_j]$ with no anchor points. Previously, this is done by checking whether $P' \circ Q'$ has consistent winding numbers. Now, we need to check the same condition but against $l = O(I)$ number of potential representatives $\{\mathbf{z}_1, \ldots, \mathbf{z}_l\}$ as the potential point of infinity. This gives a linear-factor blow-up in the time complexity compared to the algorithm for the planar case. However, we show that this linear blow-up can be tamed down and we can again compute *all* $\sigma(r, j)$s for all rs and all $j > r$ in $O(I^2 \log I)$ time, after $O(n \log n + N)$ pre-processing time. See Appendix H.3 for details. Overall, the total time complexity remains the same as before.

Putting both cases ($g > 0$ and $g = 0$) together, we conclude with the following main result.

Theorem 5.5 *Given a triangulation K of an orientable compact 2-manifold M with genus g, let N be the complexity of K. Given two homotopic paths P and Q of total complexity n with I intersection points, we can compute an optimal homotopy and its area $\sigma(P, Q; M)$ in $O(I^2 \log I + n \log n + n \log g + N)$ time.*

6 Conclusion

In this paper, we propose a new curve similarity measure which captures how hard it is to deform from one curve to the other based on the amount of total area swept. It is robust to noise (as it is area-based), and can be computed efficiently; to our knowledge, there is no other efficiently computable similarity measure for curves on surfaces. Our algorithm can be extended for cycles in the plane (see Appendix F). It appears that our algorithm can also be extended to cycles on the surfaces. Indeed, if the optimal free homotopy has an anchor point, then we can break cycles into curves that share a common start and end point, which then reduces to the problem of comparing curves on surfaces. However, on a surface the analog of Lemma F.1 no longer holds, so that two curves may intersect in M but not have an anchor point in the optimal homotopy; in this case, it is not clear in this case how to bound the size of the universal cover necessary for our algorithm.

Measuring similarity of curves on surfaces is an interesting problem, and many open areas remain. One interesting new idea would be to develop an area-based curve similarity measure that allows topological changes, such as allowing a region to be swept as long as it has trivial homology.

Acknowledgment. The authors would like to thank Joseph O'Rourke, Rephael Wenger, Michael Davis, and Tadeusz Januszkiewicz for useful discussions at the early stage of this work, and David Letscher, Brody Johnson, and Bryan Clair for helpful discussions at the later stage of this paper. We would also like to thank the anonymous reviewers for their comments.

References

[1] P. K. Agarwal, R. B. Avraham, H. Kaplan, and M. Sharir. Computing the discrete Fréchet distance in subquadratic time. In *Proc. 24th Ann. ACM-SIAM Sympos. Discrete Alg. (SODA)*, 2013.

[2] H. Alt. The computational geometry of comparing shapes. In *Efficient Algorithms*, volume 5760 of *Lecture Notes in Computer Science*, pages 235–248. Springer Berlin / Heidelberg, 2009.

[3] H. Alt, U. Fuchs, G. Rote, and G. Weber. Matching convex shapes with respect to the symmetric difference. *Algorithmica*, 21:89–103, 1998.

[4] H. Alt and M. Godau. Computing the Fréchet distance between two polygonal curves. *International Journal of Computational Geometry and its Applications*, 5:75–91, 1995.

[5] H. Alt and L. J. Guibas. Discrete geometric shapes: Matching, interpolation, and approximation. In J.-R. Sack and J. Urrutia, editors, *Handbook of Computational Geometry*, pages 121–153. Elsevier Science Publishers B. V. North-Holland, Amsterdam, 2000.

[6] H. Alt, C. Knauer, and C. Wenk. Comparison of distance measures for planar curves. *Algorithmica*, 38(1):45–58, 2004.

[7] B. Aronov, S. Har-Peled, C. Knauer, Y. Wang, and C. Wenk. Fréchet distance for curves, Revisited. In *Proceedings of the European Symposium on Algorithms*, pages 52–63, 2006.

[8] S. Bespamyatnikh. An optimal morphing between polylines. 12(3):217–228, 2002.

[9] P. Bose, S. Cabello, O. Cheong, J. Gudmundsson, M. var Kreveld, and B. Speckmann. Area-preserving approximations of polygonal paths. *Journal of Discrete Algorithms*, 4:554–566, 2006.

[10] G. R. Brightwell and P. Winkler. Submodular percolation. *SIAM J. Discret. Math.*, 23(3):1149–1178, July 2009.

[11] K. Buchin, M. Buchin, C. Knauer, G. Rote, and C. Wenk. How difficult is it to walk the dog? In *Proc. 23rd Europ. Workshop Comput. Geom.*, pages 170–173, 2007.

[12] K. Buchin, M. Buchin, W. Meulemans, and W. Mulzer. Four Soviets walk the dog—with an application to Alt's conjecture. *arXiv/1209.4403*, 2012.

[13] K. Buchin, M. Buchin, and Y. Wang. Partial curve matching under the fréchet distance. In *Proc. ACM-SIAM Sympos. Discrete Alg. (SODA)*, 2009.

[14] M. Buchin. *On the Computability of the Fréchet Distance Between Triangulated Surfaces*. PhD thesis, Dept. of Comput. Sci., Freie Universität Berlin, 2007.

[15] E. W. Chambers, É. Colin de Verdière, J. Erickson, S. Lazard, F. Lazarus, and S. Thite. Homotopic frchet distance between curves or, walking your dog in the woods in polynomial time. *Computational Geometry*, 43(3):295 – 311, 2010. Special Issue on 24th Annual Symposium on Computational Geometry (SoCG'08).

[16] E. W. Chambers, T. Ju, D. Letscher, and L. Liu. Isotopic fréchet distance. In *CCCG11*, 2011.

[17] E. W. Chambers and D. Letscher. On the height of a homotopy. In *CCCG'09*, pages 103–106, 2009.

[18] A. F. Cook and C. Wenk. Geodesic Fréchet distance inside a simple polygon. In *Proc. 25th Internat. Sympos. Theoret. Asp. Comp. Sci.*, pages 193–204, 2008.

[19] R. G. Cromley. *Digital Cartography*. Prentice Hall, Englewood Cliffs, NJ, 1992.

[20] T. K. Dey and H. Schipper. A new technique to compute polygonal schema for 2-manifolds with application to null-homotopy detection. *Discrete and Computational Geometry*, 14(1):93–110, 1995.

[21] J. Douglas. Solution of the problem of plateau. *Transactions of the American Mathematical Society*, 33:263–321, 1931.

[22] A. Driemel and S. Har-Peled. Jaywalking your dog: computing the fréchet distance with shortcuts. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '12, pages 318–337. SIAM, 2012.

[23] A. Efrat, Q. Fan, and S. Venkatasubramanian. Curve matching, time warping, and light fields, new algorithms for computing similarity between curves. *J. Math. Imaging Vis.*, 27(3):203–216, 2007.

[24] A. Efrat, S. Har-Peled, L. J. Guibas, J. S. Mitchell, and T. Murali. New similarity measures between polylines with applications to morphing and polygon sweeping. *Discrete and Computational Geometry*, 28:535–569, 2002.

[25] S. Har-Peled, A. Nayyeri, M. Salavatipour, and A. Sidiropoulos. How to walk your dog in the mountains with no magic leash. In *Proceedings of the 2012 symposuim on Computational Geometry*, SoCG '12, pages 121–130, New York, NY, USA, 2012. ACM.

[26] S. Har-Peled and B. Raichel. The fréchet distance revisited and extended. In *Proceedings of the 27th annual ACM symposium on Computational geometry*, SoCG '11, pages 448–457, New York, NY, USA, 2011. ACM.

[27] H. Lawson. *Lectures on minimal submanifolds*, volume 1 of *Mathematics lecture series*. Publish or Perish, 1980.

[28] A. Maheshwari and J. Yi. On computing Fréchet distance of two paths on a convex polyhedron. In *European Workshop on Computational Geometry*, pages 41–44, 2005.

[29] R. B. McMaster and K. S. Shea. *Generalization in Digital Cartography*. Association of American Cartographers, Washington DC, 1992.

[30] K. Mehlhorn and C.-K. Yap. Constructive Whitney-Graustein Theorem: Or how to untangle closed planar curves. *SIAM J. Comput.*, 20(4):603–621, 1991.

[31] T. Rado. On plateau's problem. *Annals of Mathematics*, 31:457–469, 1930.

[32] J. J. Rotman. *An Introduction to Algebraic Topology*. Graduate Texts in Mathematics; 119. Springer-Verlag New York Inc., 1988.

[33] H. Schipper. Determining contractibility of curves. In *Proc. Sympos. on Computational Geometry*, pages 358–367, 1992.

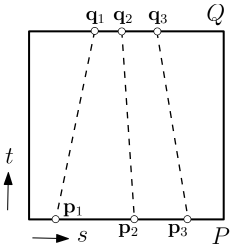
[34] G. Vegter and C. K. Yap. Computational complexity of combinatorial surfaces. In *Proc. Sympos. on Computational Geometry*, pages 102–111, 1990.

[35] R. C. Veltkamp. Shape matching: similarity measures and algorithms. In *Proc. Shape Modeling International*, pages 188–199, 2001.

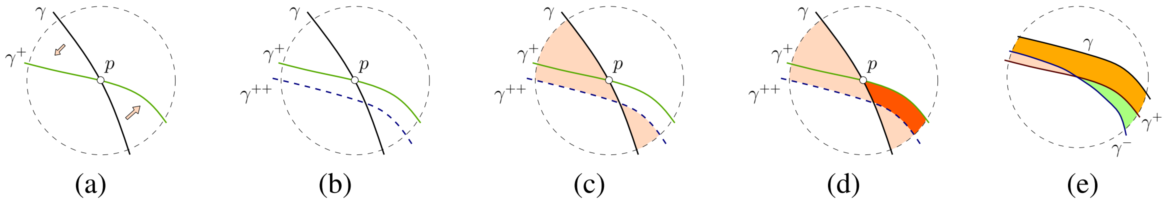
[36] H. Whitney. On regular closed curves in the plane. *Compositio Mathematica*, 4:276–284, 1937.

A Proof for Observation 3.1

Note that H^* is a map from $\square \to M$, where $\square = [0,1] \times [0,1]$ is the unit square and a point $(s,t) \in \square$ will be mapped to $H^*_t(s)$. See the right figure for an illustration. (Since P and Q share starting and ending endpoint, the left and right sides of $\square$ should be contracted to a point. We use the square view for simpler illustration.) The top and bottom boundary edges of this square are mapped to Q and P, respectively. Given an anchor point $\mathbf{b}_i$, let $\mathbf{p}_i$ and $\mathbf{q}_i$ be the parameters of $\mathbf{b}_i$ in H^*_0 and H^*_1, respectively; that is, $H^*_0(\mathbf{p}_i) = H^*_1(\mathbf{q}_i) = \mathbf{b}_i$. By definition of anchor points, the pre-image of $\mathbf{b}_i$ under the map H^* necessarily includes a curve in $\square$ connecting $\mathbf{p}_i$ on the bottom edge to $\mathbf{q}_i$ on the top boundary edge of $\square$. Since $\mathbf{b}_i \neq \mathbf{b}_j$, the pre-images of $\mathbf{b}_i$ cannot intersect with that of $\mathbf{b}_j$. Hence no two such curves can intersect each other, which means that $\mathbf{p}_i$s must be ordered in the same way as $\mathbf{q}_i$s.



B Proof for Lemma 3.2



(a) (b) (c) (d) (e)

Figure 2: (a) and (b) p is fixed from γ to γ^+, but not so in γ^{++}. Sweeping γ to γ^{++} directly through the shaded region in (c) has a smaller area than first to γ^+ then to γ^{++} (see shaded region in (d)). The darker shaded region in (d) is swept twice. (e) If the deformation changes orientation at γ, then there is a local fold in the regions swept.

Consider a time t in the homotopy, and let $\gamma = H_t$. We first show that H deforms γ consistently, so that every point on γ is either fixed or deforms to the same side of γ.

First note that if some portion of γ is left sense-preserving at time t and then reverses its direction and becomes right sense preserving at time t^+ a small amount later, some portion of the domain has been swept twice. Hence this homotopy cannot have minimal area, since we can create a smaller one which stops at time t and moves directly to some intermediate curve at a time greater than t^+ without sweeping any portion twice. See Figure 2 (e).

Now suppose that some portion of γ is deforming to one direction and another is morphing in the opposite direction. Since H is a homotopy and is therefore continuous, this means that there is at least one interval of fixed points between these two regions (which may possibly consist of a single point). Let p be an extremal point on this interval; see Figure 2 (a) for a picture when p is the only fixed point. In addition, since p is a fixed point but not an anchor point, where know there is some $t^+ = t + dt$ where p is still on $h_{t^+} = \gamma^+$ and another $t^{++} = t^+ + dt$ where p is not on $h_{t^{++}} = \gamma^{++}$.

Now we have several cases to consider. First, consider if H has directly reversed the direction of either portion of the curve (before p or after p), we are in a similar situation to the one previously discussed, since the curve goes from locally forward to locally backward (or vice versa). In this case, we again know that some area of the domain has been swept twice, which means γ^{++} has been swept over once and then was returned to, so we can reduce the area swept by H by reparameterizing H to move directly to γ^{++} without passing it and then reversing. (See Figure 2 (b),(c), and (d) for an illustration.)

Now if neither portion directly reverses, then γ^{++} must also be deforming to two different directions. Also, we know that γ^{++} must intersect γ at some point $q \neq p$, since we are essentially rotating around a central set of fixed points on these curves. In this case, we can again alter H to attain a smaller area swept

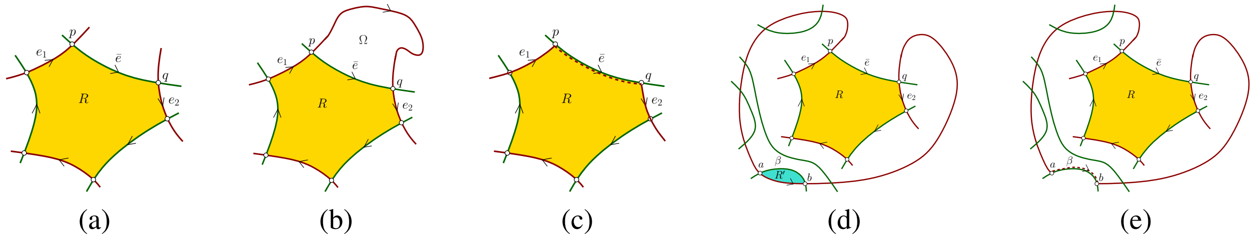


Figure 3: (a) The cell R with highest positive winding number. It boundary consists of alternating P-arcs (red) and Q-arcs (green). The two cases of relations between $P[p, q]$ and R are shown in (b) and (d), respectively. For case (b), we can deform P to sweep through Ω as shown in (c), and reduce the number of intersections by 2. Similarly, for case (d), we can identify any bigon R' and deform P to reduce the number of intersections by 2 as well.

by simply sweeping directly from γ to γ^{++}; this will reduce the area since the triangular region in the center bounded by γ, γ^{+}, and γ^{++} will be swept only one time instead of twice.

The claim thus follows, since any homotopy with no anchor points that is not sense preserving cannot be minimal.

C Proof for Lemma 4.1

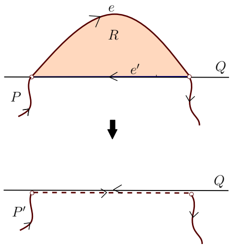
Without loss of generality, assume that the map H is right sense-preserving, always deforming an intermediate curve to its right. Consider the time-varying function $F : [0, 1] \times \mathbb{R}^2 \to \mathbb{Z}$, where $F(t, x) = \mathrm{wn}(x; H_t)$ is the winding number at $x \in \mathbb{R}^2$ with respect to the curve parameterized by H_t. Obviously, $F(0, x) = \mathrm{wn}(x; P \circ Q)$, and $F(1, x) = 0$. During the deformation, $F(t, x)$ changes by either 1 or -1 whenever the intermediate curve sweep over it. Since the homotopy is right sense-preserving, when an intermediate curve sweeps x, x always moves from the left side of the intermediate curve to its right side. Hence the winding number x decreases monotonically. Since in the end, the winding number at each point is zero, $\mathrm{wn}(x; P \circ Q) = F(0, x) \geq 0$.

If the map H is left sense-preserving, then a symmetric argument shows that $\mathrm{wn}(x; P \circ Q) \leq 0$ for all $x \in \mathbb{R}^2$.

D Proof for Lemma 4.3

We prove the claim by induction on the number of intersections between P and Q. The base case is when there is no intersection between P and Q. In this case, Γ is a Jordan curve which decomposes $\mathbb{R}^2$ into two regions, one inside Γ and one unbound. By orienting Γ appropriately, every point in the bounded cell has winding number 1 and the claim follows.

Now assume that the claim holds for cases with at most $k - 1$ intersections. We now prove it for the case with k intersections. Let an *X-arc* denote a subcurve of curve X. Consider the arrangement $\mathrm{Arr}(\Gamma)$ formed by $\Gamma = P \circ Q$. Since P and Q are simple, every cell in this arrangement has boundary edges alternating between P-arcs and Q-arcs. Assume without loss of generality that Γ has all non-negative winding numbers. Consider a cell $R \in \mathrm{Arr}(\Gamma)$ with largest (and thus positive) winding number. Since its winding number is greater than all its neighbors, it is necessary that all boundary arcs are oriented consistently as shown in Figure 3 (a), where the cell R (shaded region) lies to the right of its boundary arcs.



If R has only two boundary arcs, e from P and e' from Q, respectively, then we can morph P to another simple curve P' by deforming e through R to $-e'$ (where '$-$' means reversing the orientation). See the right figure for an illustration. The area swept by this deformation is exactly the area of cell R. Furthermore, after the deformation, every point $x \in R$ decreases their winding number by 1, and no other point changes its winding number. Since points in this cell initially has strictly positive winding number, the resulting curve $\Gamma' = P' \circ Q$ still has all non-negative winding number. The number of intersections between P' and Q is $k-2$. By induction hypothesis, $\sigma(P', Q) = \mathrm{Tw}(\Gamma')$. Since $\mathrm{Tw}(\Gamma) - \mathrm{Tw}(\Gamma') = \mathrm{Area(R)}$, we have that $\mathrm{Tw}(\Gamma) = \sigma(P', Q) + \mathrm{Area(R)}$. It then follows from Observation 4.2 and the fact $\sigma(P, Q) \le \sigma(P', Q) + \mathrm{Area(R)}$ that $\sigma(P, Q) = \mathrm{Tw}(\Gamma)$.

Otherwise, the cell R has more than one P-arc. Take the P-arc e_1 with the smallest index along P, and let p be the ending endpoint of it. Let e_2 be the next P-arc along the boundary of R, and q its starting endpoint, and $Q[p, q]$ the Q-arc between e_1 and e_2, denoted by $\bar{e}$ in Figure 3. Obviously, the subcurve $P[p, q]$ cannot intersect R, and $P[p, q]$ and $-Q[p, q]$ bound a simple polygon, which we denote by Ω. Either Ω is on the opposite side of the Q-arc $\bar{e}$ from the interior of R (Figure 3 (b)), or they are on the same side (Figure 3 (d)).

Case (1): R and Ω are on the opposite side of $\bar{e}$. In this case, the region Ω is to the right of the oriented arc $P[p, q]$. Note that P does not intersect the interior of Ω; as otherwise, P will either intersect itself or intersect $\bar{e}$, neither of which is possible. Hence only Q can intersect Ω. Since Q is also a simple curve, there is no vertices of $\mathrm{Arr}(\Gamma)$ contained in the interior of Ω. As a result, every cell of $\mathrm{Arr}(\Gamma)$ contained in Ω must have at least one boundary edge coming from $P[p, q]$. This implies that each cell contained in Ω has strictly positive winding number; that is, $\mathrm{wn}(x; \Gamma) > 0$ for any $x \in \Omega$. This is because if a cell $\xi \subseteq \Omega$ has winding number 0, then its neighbor across its boundary on the other side of $P[p, q]$ will have winding number -1, as Ω is to the right of $P[p, q]$. This violates the condition that Γ has all non-negative winding numbers and thus cannot happen.

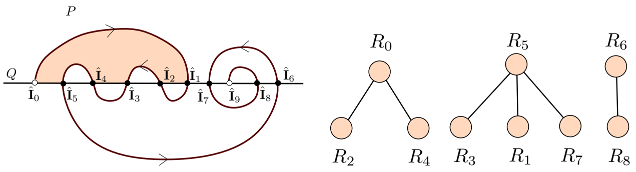
We now deform P to P' by sweeping $P[p, q]$ through Ω to $Q[p, q]$. See Figure 3 (c). The cost of this sweeping is $\mathrm{Area}(\Omega)$ and $\mathrm{Tw}(\Gamma) - \mathrm{Tw}(P' \circ Q) = \mathrm{Area}(\Omega)$. P' is still simple, and the number of intersection points between P' and Q is now $k-2$. Since $\mathrm{wn}(x; \Gamma) > 0$ for any $x \in \Omega$, we have $\mathrm{wn}(x; P' \circ Q) \ge 0$ for $x \in \Omega$. No other point will change their winding number after this deformation. Thus the curve $P' \circ Q$ has all non-negative winding numbers as well. Hence by induction hypothesis, we have that $\sigma(P', Q) = \mathrm{Tw}(P' \circ Q)$. Since $\sigma(P, Q) - \sigma(P', Q) \le \mathrm{Area}(\Omega)$ and $\mathrm{Tw}(\Gamma) - \mathrm{Tw}(P' \circ Q) = \mathrm{Area}(\Omega)$, it then follows from Observation 4.2 that $\sigma(P, Q) = \mathrm{Tw}(\Gamma)$.

Case (2): R and Ω are both from the same side of $\bar{e}$. We now consider the remaining case as shown in Figure 3 (d). Take the unbounded region $\overline{\Omega} := \mathbb{R}^2 \setminus \Omega$ which is the complement of Ω. This unbounded region lies to the right of the oriented curve $P[p, q]$. Since both P and Q are simple, only Q can intersect Ω and there is no vertices of $\mathrm{Arr}(\Gamma)$ contained in the interior of $\overline{\Omega}$. First, observe that it is not possible that $\overline{\Omega} \cap Q = \emptyset$. This is because otherwise, $\overline{\Omega}$ is the unbounded face of $\mathrm{Arr}(\Gamma)$ and thus the winding number for all points in $\overline{\Omega}$ is 0. This however is not possible as this will imply that any point y to the immediate left of $P[p, q]$ has winding number -1, violating our assumption that all cells in $\mathrm{Arr}(\Gamma)$ have consistent (non-negative) winding numbers.

Hence $\overline{\Omega} \cap Q \neq \emptyset$, and there are set of arcs from Q intersecting $P[p, q]$. Then there must exist a bigon cell R' bounded by only two arcs, one P-arc $P[a, b] \subseteq P[p, q]$ and a Q-arc β. See Figure 3 (d). Similar to the argument from the previous paragraph, we can show that points in R' must have strictly positive winding number. Now deform P to P' by sweeping $P[a, b]$ through R' to β as shown in Figure 3 (e). P' is still simple, the number of intersection points between P' and Q is now $k-2$. Only points in R' reduce their winding number by 1, and the resulting arrangement still has consistent winding numbers. As such, by induction hypothesis, we have that $\sigma(P', Q) = \mathrm{Tw}(P' \circ Q)$. Since $\sigma(P, Q) - \sigma(P', Q) \le \mathrm{Area(R')}$ and

$\mathrm{Tw}(\Gamma) - \mathrm{Tw}(P' \circ Q) = \mathrm{Area}(\mathrm{R}')$, it then follows from Observation 4.2 that $\sigma(P, Q) = \mathrm{Tw}(\Gamma)$.

E Computing the Area of R_u for All $u \in [1, I]$



We can pre-compute the area of R_u's for all $u \in [1, I]$ in $O(n \log n + I)$ time as follows. First, we compute the arrangement of $\mathrm{Arr}(P + \hat{Q})$ and the area of all cells in it in $O(n \log n + I)$ time. Each R_u is the region bounded between a P-arc $P[\hat{\mathbf{x}}_i, \hat{\mathbf{x}}_{i+1}]$ and a corresponding $\hat{Q}$-segment $\hat{Q}[\hat{\mathbf{x}}_i, \hat{\mathbf{x}}_{i+1}]$. Since no two P-arcs intersect, the containment relationship between such P-arcs satisfies parentheses property. In particular, we can use a collection of trees to represent the containment relation among all regions R_us. See the right figure for an illustration. The difference between the region represented at a parent node and the union of regions represented by all its children is a cell in $\mathrm{Arr}(P + Q)$. For example, the shaded cell in the right figure is the difference between R_0 and its children R_2 and R_4. We can thus compute the area of all R_us by a bottom-up traversal of these trees. Computing these trees take $O(I \log I)$ time by first sorting all intersection points with respect to their order along $\hat{Q}$. Traversing these trees to compute all R_us takes $O(I)$ time. Putting everything together, we need $O(n \log n + I \log I)$ time.

F Cycles in the Plane

We now consider the case where we have two simple cycles P and Q in the plane. We have the following characterization:

Lemma F.1 *If the two simple cycles P and Q intersect, then there is an anchor point in the optimal homotopy between them.*

Proof: Suppose that P and Q intersect but there is no anchor point in the optimal homotopy H^*. By Lemma 3.2, we know that H^* must be sense preserving. However, this means that H^* continually moves one curve to the other in one local direction, which means that one curve must be entirely contained within the other, contradicting the assumption that they intersect. ■

At this point, the algorithm for cycles which intersect each other reduces to the one for curves: if we know which intersection point between P and Q is the anchor point, we can simply "break" the cycles into two curves at this point; this will become the start and end point for each of the two curves. Hence our algorithm for cycles will take a multiplicative factor of $O(I)$ extra time than the algorithm for curves, since we need to try each possible intersection point as the required anchor point.

It then follows from Theorem 4.4 that:

Corollary F.2 *Given two polygonal cycles P and Q in the plane of n total complexity and with $I > 0$ intersection points, we can compute the optimal homotopy and its area in $O(I(I^2 \log I + n \log n))$ time.*

The remaining case is that when the two polygonal cycles P and Q are disjoint. If one of the cycle contains the other, then the area of the optimal homotopy is simply the area sandwiched between these two simple cycles. This can be computed in $O(n \log n)$ time easily.

However, if the cycles are disjoint but neither contains the other, then in a sense the "optimal" free homotopy between them will simply be the area bounded by each curve, since the homotopy can collapse each curve separately to a point and then deform the points to each other. Indeed, that the sum of area bounded by the two Jordan cycles P and Q is the minimum possible homotopy area follows from a similar argument as the proof of Observation 4.2. However, in this case, the free homotopy described above is not regular since it collapses a curve to a single point. Nevertheless, one can argue that there exists a sequence of regular homotopies whose areas converge to this sum. In other words, the optimal area homotopy is still well-defined (as the *infinum* of the area of regular free homotopies between P and Q), although there does not exist a regular homotopy to achieve this optimal area. (This is analogous to similar issues that arise in the general case, and is the reason for introducing more restricted integrals in the mathematical literature [27].)

G Algorithm for the Cases with Non-zero Genus

We now explain our algorithm to compute the optimal homotopy for two polygonal chains P and Q from a triangulation K of a compact orientable surface M with genus $g > 0$. There are n total edges in P and Q, each being an edge of K. The two homotopic paths P and Q share the same starting and ending points. Let I be the number of intersection points between P and Q; obviously, $I = O(n)$ in this scenario. Denote by N the complexity of K.

The high level idea is to compute a lift of the curve $C = P \circ Q$ in the universal covering space, which we denote as $\tilde{C} = \tilde{P} \circ \tilde{Q}$, and then apply the algorithm from Section 4 to the lifted curves. We now detail the steps involved and their complexity.

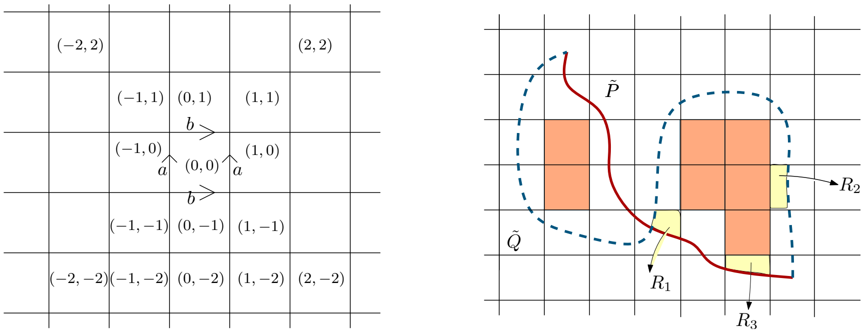
Step 1: Compute relevant portion of a universal covering space. Eventually, we will construct a portion of a univeral covering space $\mathcal{U}(K)$ by tiling the so-called polygonal schema of M [34, 20]. We use the algorithm from [20] to construct a polygonal schema T (and a reduced version of it) in $O(N)$ time. The universal covering space will consist of infinite copies of this polygonal schema glued together appropriately. In what follows, we call each copy of the polygonal schema in the constructed univesal covering space *a tile*.

For $g \geq 1$, the universal covering space $\mathcal{U}(K)$ is homeomorphic to $\mathbb{R}^2$. By fixing the lift of the starting endpoint of P and Q in $\mathcal{U}(K)$, we obtain a specific lift $\tilde{P}$ and $\tilde{Q}$ for P and Q respectively. Since P and Q are homotopic, $\tilde{P}$ and $\tilde{Q}$ form a closed curve, denoted by $\tilde{C} = \tilde{P} \circ \tilde{Q}$. Note that the number of intersection points between $\tilde{P}$ and $\tilde{Q}$ is at most I, as every intersection point in the lift necessarily maps to an intersection point of P and Q under the covering map, but not vice versa. Consider the arrangement $\mathrm{Arr}(\tilde{C})$ in $\mathcal{U}(K)$, and we want to construct the portion of the universal covering space $U \subseteq \mathcal{U}(K)$ such that all *bounded* cells in $\mathrm{Arr}(\tilde{C})$ will intersect (either along the boundary of these cells or enclosed in their interior).

As shown in [20], the lifted curve $\tilde{C}$ passes through $O(n)$ number of tiles in $\mathcal{U}(K)$. However, while the total number of tiles intersecting all bounded cells in $\mathrm{Arr}(\tilde{C})$ is $O(n)$ for the case where $g > 1$, it can be $\Theta(n^2)$ for the case when $g = 1$. Hence to this end, we will separate the case for $g = 1$ and $g > 1$ and handle them differently.

For the case $g > 1$, we again use the algorithm by Dey and Schipper [20] to compute the relevant portion U of the universal covering space in $O(n \log g + N)$ time. The output contains all $O(n)$ copies of the polygonal schema intersecting U, where each tile is represented by a reduced $4g$-gon without being explicitly filled with triangles from K. Once these are known, we can compute the combinatorial structure of the arrangement of $\tilde{C}$ in U, as well as the description of the set of tiles each bounded cell in $\mathrm{Arr}(\tilde{C})$ intersects, in $O(n + I \log I)$ time.

For the case $g = 1$, the input manifold is a torus, and the canonical polygon schema for it is a rectangle with oriented boundary arcs $aba^{-1}b^{-1}$. Imagine now that we give the *base polygonal schema* T_0 (which is the tile that contains the lift of the starting point of P and Q) a *coordinate* $(0, 0)$, and the coordinate for every other copy of the polygon is shown in Figure 4 (a). Specifically, a copy of polygonal schema T has coordinate (i, j) if those closed loops whose lift starting from T_0 and ending in T have the same homotopy



(a) (b)

Figure 4: (a) A combinatorial view of the universal covering space $\mathcal{U}(M)$. a and b are the generators and we can give each cell a coordinate. (b) The lift of P (solid curve) and the lift of Q (dashed curve). The heavily shaded region are copies of polygonal schema contained inside cells of $\mathrm{Arr}(\tilde{C})$, and their total number can be easily computed by a scanning algorithm. R_1 is an essential cell; R_2 and R_3 are two non-essential cells.

type as $a^i b^j$. We can easily obtain the sequence of the rectangles (and their coordinates) that the curve $\tilde{C}$ will pass through in $O(n+N)$ time [20]. Once these coordinates are known, the combinatorial structure of the arrangement of $\tilde{C}$ in U can also be computed in $O(n + I \log I)$ time. Note that in this case, we do not explicitly enumerate the set of tiles fully enclosed within a bounded cell of $\mathrm{Arr}(\tilde{C})$ (the shaded tiles in Figure 4 (b)), whose number can be $\Theta(n^2)$ instead of $O(n)$ as in the case when $g > 1$.

Step 2: Area of cells in $\mathrm{Arr}(\tilde{C})$. In order to perform our algorithm introduced in Section 4 to the lifted curves $\tilde{P}$ and $\tilde{Q}$, in addition to the combinatorial structure of $\mathrm{Arr}(\tilde{C})$, we also need the area of each cell in $\mathrm{Arr}(\tilde{C})$. We first describe how to compute it for the case $g = 1$.

Take any cell X in $\mathrm{Arr}(\tilde{C})$ and assume the boundary of X intersects m copies of polygonal schema. Even though that X may contain $\Theta(m^2)$ copies of polygonal schema in its interior, we do not need to enumerate these interior tiles explicitly to compute their total area.

Indeed, by a scanning algorithm from left to right, we can compute in $O(m)$ time how many tiles are completely contained inside X (heavily–shaded regions in Figure 4 (b)) (note that the coordinates of each rectangle traversed by the boundary of X are known). Since the area of every polygonal schema is simply the total area of the input triangulation, we can compute the total area of rectangles contained inside X in $O(m)$ time.

Now let $\mathbf{R}$ be the collection of tiles that intersect the boundary of X. It remains to compute the total area of $\mathbf{R} \cap X$. Call $T \cap X$ a *sub-cell*, for each tile $T \in \mathbf{R}$. Let G denote the boundary curves of the polygonal schema T. There are two types of sub-cells: the *essential* ones which contain at least one intersection point between $\tilde{P}$ and $\tilde{Q}$ as their vertices, and the *non-essential* ones which are bounded either by arcs from G and arcs from P, or bounded by arcs from G and arcs from Q. See Figure 4 (b) for examples.

First let us consider all non-essential sub-cells (for all cells X of $\mathrm{Arr}(\tilde{C})$) formed by boundary arcs of a tile and arcs from P. If we plot all these P-arcs within one tile, no two P-arcs in this tile intersect, as P is a simple curve. Imagine that we pick an arbitrary but fixed point on the boundary G of the polygonal schema T as the origin $\mathbf{o}$. Each P-arc α subdivides T into two regions, and let T_α denotes the canonical one excluding $\mathbf{o}$. Note that since P is a simple curve, the set of canonical regions T_αs for all P-arcs again satisfy the parenthesis property, and these regions and their areas can be computed in $O(n \log n + N)$ time using a data structure similar to one used in Section E to compute the area of R_us. See Figure 5 for an illustration. Once these areas are known, the area of each non-essential sub-cell can be computed in $O(s)$

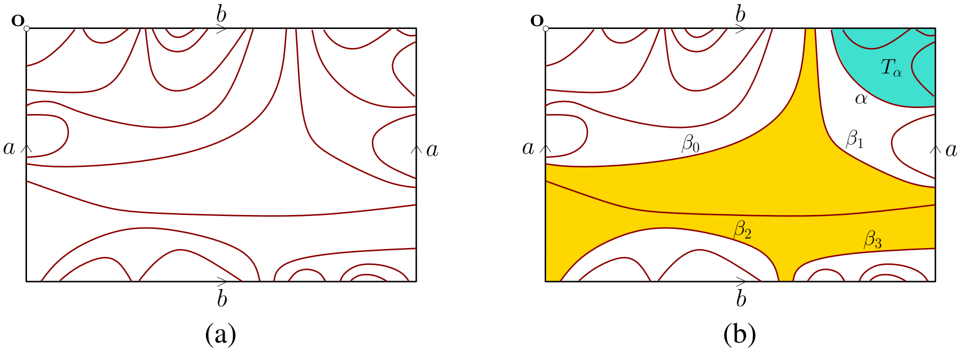


Figure 5: (a) We overlay all non-essential sub-cells involving P-arcs into one copy of the polygonal schema. (b) An example of the canonical region T_α is shown for arc α (shaded region in the top-right corner). The shaded region in the middle is a sub-cell X which is computed as $T_{\beta_0} - T_{\beta_1} - T_{\beta_2} - T_{\beta_3}$, where β_is are the boundary P-arc for X. Among these P-arcs, β_0 is the top-most arc in the data structure as described in Section E.

time where s is the number of P-arcs on the boundary of this sub-cell. See Figure 5 (b). Hence the areas of all non-essential sub-cells can be computed in $O(n)$ time once all T_αs are known. The total time complexity required here is thus $O(n \log n + N)$. The areas for all non-essential sub-cells (for all cells X of $\mathrm{Arr}(\tilde{C})$) formed by G and Q-arcs can be computed in a symmetric manner.

What remains is to compute the area of all essential sub-cells. Again, let G denote the boundary of the polygonal schema T. We use a PQ-arc to refer to an arc that starts and ends with points on G and consists of alternating P- and Q-arcs. Note that an essential sub-cell is either completely contained within a polygonal schema, or its boundary consists of PQ-arcs, G-arcs, P-arcs and Q-arcs. In particular, no two PQ-, P- or Q-arcs can be consecutive: they are separated by G-arcs. Now collect all P-arcs and Q-arcs that are involved in PQ-arcs. Plot them within the same tile T and compute their arrangement A as well as the area for each cell in A. This can be done in $O(n \log n + N)$ time and A has only $O(I)$ (not $O(n)$) number of cells. If an essential sub-cell X is completely contained within a polygonal schema, then it is a union of a set of cells from A. We can simply spend $O(I)$ time to go through cells in A, identify those contained in X and return their total area. If an essential sub-cell X has G-arcs on its boundary, then we need a slightly more complicated way to handle it.

Specifically, for all the remaining essential sub-cells, there can be $O(I)$ number of PQ-arcs along their boundaries. (Note that two PQ-arcs can share common P-arcs or Q-arcs among them). Each P-arc or Q-arc from the arrangement A can appear in at most $O(I)$ PQ-arcs. Hence it takes $O(I^2)$ times to compute and enumerate all PQ-arcs. Similar to before, each PQ-arc α divides the tile T into two regions, and we define T_α to be the canonical one that excluding a specific origin $\mathbf{o}$ on G. Take an essential sub-cell X that has s number of P-, Q-, or PQ-arcs along its boundary, denoted by $\alpha_1, \alpha_2, \ldots, \alpha_s$. Let α_1 be the arc (which can be P-, Q- or PQ-arc) whose endpoints along G spans the largest interval. Then, X can be represented as $X = T_{\alpha_1} - \bigcup_{i \in [2,s]} T_{\alpha_i}$, where T_{α_i} is the canonical region defined by an arc α_i. Since the area of all canonical regions are known (for P-arcs or Q-arcs, we have computed them before), the area of X ts area can be computed in $O(s)$ time. Computing the area of all remaining essential sub-cells thus takes $O(I^2)$ time.

Putting everything together, the total time needed to compute the area of all cells in $\mathrm{Arr}(\tilde{C})$ takes $O(n \log n + N + I^2)$ time for the case of $g = 1$.

The case when $g > 1$ is similar and simpler. Indeed, we now can afford to compute all the tiles contained within any cell of $\mathrm{Arr}(\tilde{C})$ explicitly, as their total number is bounded by $O(n)$ [20, 33]. The areas of essential and non-essential sub-cells are computed using the same algorithm as above. The total time complexity is $O(n \log n + n \log g + N + I^2)$.

Step 3: Putting everything together. With the combinatorial structure of $\mathrm{Arr}(\tilde{C})$ and the area of each cell computed, we now apply the algorithm from Section 4.2 to compute the optimal homotopy in $O(I^2 \log I + n \log n)$ time in $\mathcal{U}(M)$, which, by 5.1, gives the optimal homotopy between P and Q in M in the same time bound. The total time complexity for the entire algorithm is $O(n \log n + n \log g + I^2 \log I + N)$.

H Missing Details for the Sphere Case

Observation H.1 *Given a closed curve* Γ *and any two points* $\mathbf{z}, \mathbf{w} \in \mathbf{S}$*, we have that:* $\mathrm{wn}(x; \mathbf{w}) = \mathrm{wn}(x; \mathbf{z}) + \mathrm{wn}(\mathbf{z}; \mathbf{w})$ *(all winding numbers are w.r.t the curve* Γ*). In particular, for any two points* $\mathbf{z}_1, \mathbf{z}_2$ *from the same cell of* $\mathrm{Arr}(\Gamma)$*, we have that* $\mathrm{wn}(x; \mathbf{z}_1) = \mathrm{wn}(x; \mathbf{z}_2)$ *for all* $x \neq \mathbf{z}_1, \mathbf{z}_2$.

Proof: Let $\gamma(x, y)$ be a path connecting point x to y. Note that the concatenation between $\gamma(x, \mathbf{z})$ and $\gamma(\mathbf{z}, \mathbf{w})$ is a path from x to $\mathbf{w}$. Since $\mathrm{wn}(x; \mathbf{w})$ is simply the summed signed crossing number of any path from x to $\mathbf{w}$ with respect to Γ, the claim follows immediately. ■

H.1 Proof for Observation 5.3

Suppose x and y are two points from the interior of R such that H^* sweeps through x, but not y. Connect x with y by any path γ in the interior of R. This path has to intersect the boundary of the region swept by H^*, and let z be one such intersection point on γ. Obviously, there is a local fold in the optimal homotopy as it sweeps through z; namely, some intermediate curve will touch z and immediately trace back. Thus the input homotopy H^* cannot be sense-preserving. Contradiction. Hence H^* sweeps y as well.

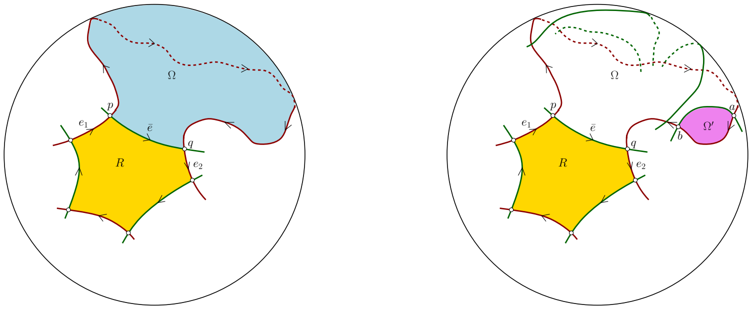
H.2 Proof for Lemma 5.4

We prove the lemma by induction on the number of intersections between P and Q. When there is no intersection between P and Q (other than the common endpoints), the Jordan curve $P \circ Q$ divides the sphere into two connected components, and the optimal homotopy is the smaller area of the two. The lemma holds for this base case.

Now assume that the lemma holds for P and Q with at most k intersection points. We wish to show the result for the case where P and Q have $k+1$ intersection points. Since H^* has no anchor points, this optimal homotopy is sense-preserving by Lemma 3.2. Assign an orientation to the closed curve $C = P \circ Q$ so that locally, every point on the curve P will continuously deform to its right during the optimal homotopy. Now pick an *arbitrary* point $\mathbf{z}$ not on P and Q, and compute the winding number for each cell of $\mathrm{Arr}(P + Q)$ w.r.t. $\mathbf{z}$. Take the cell R with the largest winding number. We assume that $\mathbf{z} \notin R$. Suppose this is not the case and that $\mathbf{z} \in R$. Then we show that we can change the choice of $\mathbf{z}$ to make this hold.

Specifically, if $\mathbf{z} \in R$, then the cell R must have winding number 0. Now take the cell R' of $\mathrm{Arr}(P+Q)$ with the smallest winding number, and let $\mathbf{w}$ be a point from R'. Obviously, $\mathrm{wn}(\mathbf{w}; \mathbf{z}) \leq \mathrm{wn}(x; \mathbf{z}) \leq 0$ for any $x \in \mathbf{S}$. Now we consider the winding numbers w.r.t. to $\mathbf{w}$ instead of $\mathbf{z}$. By Observation H.1 we have that $\mathrm{wn}(x; \mathbf{w}) = \mathrm{wn}(x; \mathbf{z}) + \mathrm{wn}(\mathbf{z}; \mathbf{w})$. On the other hand, we have that $\mathrm{wn}(\mathbf{z}; \mathbf{w}) = -\mathrm{wn}(\mathbf{w}; \mathbf{z})$. Hence $\mathrm{wn}(\mathbf{z}; \mathbf{w}) \geq \mathrm{wn}(x; \mathbf{z}) \geq 0$ for any $x \in \mathbf{S}$. In other words, for this new choice of point $\mathbf{w}$, we have that R still has the largest winding number and in this case, $\mathbf{w} \notin R$.

Similar to the proof of Lemma 4.3, the boundary of this cell consists of alternating arcs from P and from Q, and they necessarily have the orientation as shown in Figure 1 (a) (otherwise, one of the neighboring cell if R will have a larger winding number). Choose the P-arc e_1 that appears earliest along P, with p being its



(a) (b)

Figure 6: (a) Take Ω as the region to the left of $Q[p, q]$. (b) There always exists a bigon Ω' (possibly Ω is no Q-arc intersects $P[p, q]$.

ending endpoint. Let $P[q]$ be the next intersection between P and R. We have that $P[p, q]$ and $\bar{e} = Q[p, q]$ do not intersect each other. The Jordan curve $P[p, q] \circ (-Q[p, q])$ bounds two regions on the sphere (instead of a bounded one and an unbounded one in the planar case as shown in Figure 1 (b) and (d)). We consider the region that lies to the right of $P[p, q]$ (thus left of $\bar{e}$), and denote it by Ω. See Figure 6 (a). Since R is to the right of $\bar{e}$, $\Omega \cap R = \emptyset$. As P is simple, only Q can potentially intersect the cell Ω. Hence there always exists a bigon Ω' in Ω which is a cell in $\mathrm{Arr}(P + Q)$. See Figure 6 (b). Note that it is possible that $\Omega' = \Omega$. Let $P[a, b]$ denote the P-arc that bounds the bigon Ω'. Let P' be a new curve obtained by replacing $P[a, b]$ with (slightly above) $Q[a, b]$. Since P deforms always to the right in the optimal homotopy, and in the end, $P[a, b]$ needs to deform to some portion of Q (which is not necessarily $Q[a, b]$ though), one can show that there is an optimal homotopy between P and Q that consists of first sweeping $P[a, b]$ to $Q[a, b]$ through Ω', and then optimally morph P' to Q. On the other hand, by the induction hypothesis, there is an optimal homotopy H' from P' to Q that does not sweep some point, say $\mathbf{z}_1$ in $\mathbf{S}$. There are now two cases:

(i) If $\mathbf{z}_1 \in \mathbf{S} - \Omega'$, then there is an optimal homotopy from P to Q that does not sweep $\mathbf{z}_1$ as well. The induction step then holds and the claim follows.

(ii) Otherwise, $\mathbf{z}_1 \in \Omega'$. Consider the cell $R' \in \mathrm{Arr}(P'+Q)$ that contains $\mathbf{z}_1$. Note that $R' \cap (\mathbf{S} - \Omega') \neq \emptyset$, as there is no vertices of $\mathrm{Arr}(P' + Q)$ contained neither on nor inside Ω'. Hence R' must also contain some point, say $\mathbf{z}_2$, that is outside of Ω'. It then follows from Observation 5.3 that $\mathbf{z}_2$ is not swept either. This leads us back to case (i), and the induction step again holds.

The claim then follows by induction.

H.3 Details of Algorithm for Sphere Case

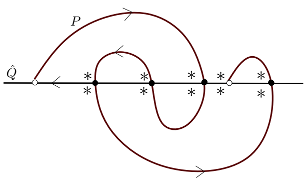
Dynamic programming framework. Similar to the planar case, let $\mathbf{x}_0, \ldots, \mathbf{x}_I$ denote the intersection points between P and Q, ordered by their indices along P, with $\mathbf{x}_0$ and $\mathbf{x}_I$ being the beginning and ending points of P and Q. Let $T(i)$ denote the optimal homotopy area between $P[\mathbf{0}, \mathbf{x}_i]$ and $Q[\mathbf{0}, \mathbf{x}_i]$, and $C[i, j]$ the closed curve formed by $P[\mathbf{x}_i, \mathbf{x}_j] \circ Q[\mathbf{x}_j, \mathbf{x}_i]$. However, now we say that a pair of indices (i, j) is *valid* as long as $\mathbf{x}_i$ and $\mathbf{x}_j$ have the same order along P and along Q. This is different from the definition of valid pairs of indices as in the planar case.

Specifically, for any closed curve γ, it turns out that γ can always have consistent winding number for *some* choices of the point of infinity $\mathbf{z}$: that is, there always exists $\mathbf{z} \in \mathbf{S}$ such that $\mathrm{wn}(x, \mathbf{z}; \gamma)$ is consistent for all $x \in \mathbf{S}_{\mathbf{z}}$. We call such choices of $\mathbf{z}$ *consistent representatives w.r.t.* γ. Let $\mathrm{Tw}^*(\gamma) := \min_{\mathbf{z}} |\mathrm{Tw}(\gamma; \mathbf{S}_{\mathbf{z}})|$ where $\mathbf{z}$ ranges over all possible choices of consistent representatives w.r.t. γ. Then, Lemma 4.3 and 5.4 imply that if there is an optimal homotopy between $P[\mathbf{x}_i, \mathbf{x}_j]$ and $Q[\mathbf{x}_i, \mathbf{x}_j]$ with no anchor points, then the optimal homotopy area is $\mathrm{Tw}^*(\gamma)$. However, different from the planar case, $\mathrm{Tw}^*(C[i,j])$ is defined for all valid pairs of i, js, and it may not in general be the optimal homotopy area for $P[\mathbf{x}_i, \mathbf{x}_j]$ and $Q[\mathbf{x}_i, \mathbf{x}_j]$. We now have the following recurrence:

$$T(i) = \begin{cases} 0, & \text{if } i == 0 \\ \min_{j<i \text{ and } (j,i) \text{ is valid}} \{ \ \mathrm{Tw}^*(C[j,i]) + T(j) \ \}, & \text{otherwise} \end{cases}$$

As before, the final goal is to compute $T(I) = \sigma(P, Q)$.

Computing Tw^***s.** Here we describe how to compute $\mathrm{Tw}^*(C[i,j])$ efficiently. Specifically, we show how to compute all $\mathrm{Tw}^*(C[r,j])$s for all $j > r$ in $O(I)$ time, for any fixed r, after $O(n \log n + I \log I + N) = O(n \log n + N)$ preprocessing.



First, let us choose the representatives by taking two points around each intersection points $\mathbf{x}_i$ between P and Q as shown in the right figure. Consider only those representatives to the right of Q (which are those above Q in the right figure), and denote them by $Z = \{\mathbf{z}_1, \ldots, \mathbf{z}_I\}$. Z is sorted by their indices along Q. Those to the left of it can be handled similarly. The first observation is that for any two consecutive representatives, $\mathrm{wn}(x; \mathbf{z}_i) - \mathrm{wn}(x; \mathbf{z}_{i+1})$ is 1 or -1, depending on the orientation of P-arc that separating them.

Now to compute which $\mathbf{z}_i$ will give consistent winding numbers, we first compute the winding number of each cell in $\mathrm{Arr}(P+Q)$ for $\mathbf{z}_1$. Next, take the cells R_1 and R_2 with minimum and maximum winding numbers, and assume that $\mathbf{z}_{i_1}$ and $\mathbf{z}_{i_2}$ are their representatives. If there are more than one cells with largest (or smallest) winding numbers, just pick an arbitrary one. By Observation H.1, the closed curve $C[r,j] := P[\mathbf{x}_r, \mathbf{x}_j] \circ Q[\mathbf{x}_r, \mathbf{x}_j]$ has all non-negative winding number w.r.t $\mathbf{z}_{i_1}$, and all non-positive winding number w.r.t. $\mathbf{z}_{i_2}$. Hence we simply compute the total winding number $\mathrm{Tw}(C[r,j]; \mathbf{z}_{i_1})$ and $\mathrm{Tw}(C[r,j]; \mathbf{z}_{i_2})$, and return the one with the smaller absolute value as $\mathrm{Tw}^*(C[r,j])$. We refer to the indices i_1 and i_2 as the *wn-min and wn-max indices*, respectively, and these two total winding numbers *valid total winding numbers*. Basically, the smaller of the absolute values of the two valid total winding numbers is simply the best cost to deform $P[\mathbf{x}_r, \mathbf{x}_j]$ to $Q[\mathbf{x}_r, \mathbf{x}_j]$ without using anchor points. This improves the time complexity of computing each $\mathrm{Tw}^*(C[r,j])$ to $O(I)$ time, instead of the naive $O(In)$ time by computing all $\mathrm{Tw}(C[r,j]; \mathbf{z}_i)$s, for $i \in [1, I]$, from scratch.

To further improve the time complexity, we will start with $C[r, r+1]$, and update the winding number information in each cell as well as the valid total winding numbers, as we traverse P and pass through each intersection point $\mathbf{x}_i$. To this end, we use the same range tree data structure as in Section 4.2. Specifically, we use this data structure to maintain the winding number information w.r.t $\mathbf{z}_1$. The wn-min and wn-max indices can be easily maintained by storing at each internal node the minimum and maximum winding number within its subtree. The time complexity for updates remains the same as before (i.e, $O(\log I)$ time per update).

The remaining question is to maintain the valid total winding numbers. Consider the valid total winding number corresponding to wn-min index (that for wn-max index can be maintained similarly). Let A denote the total area of topological sphere **S**. First, observe that for any C, $\mathrm{Tw}(C; \mathbf{z}_{i+1})$ is simply $\mathrm{Tw}(C; \mathbf{z}_i) + \alpha A$, where $\alpha = 1$ or -1, depending whether P and Q forms a positive or negative crossing at $\mathbf{x}_i$ (a positive crossing means that P rotates in a clockwise manner to Q at the intersection point). Let κ_u denote the wn-min index for $C[r,u]$, and $V[u]$ denote $\mathrm{Tw}(C[r,u]; \kappa_u)$. Note that as we pass $\mathbf{x}_{u+1}$, the wn-min index κ_{u+1} either stays the same, or change from κ_u by 1 or -1. Let V' be the total winding number w.r.t. κ_u after we pass through intersection $\mathbf{x}_{u+1}$. If $\kappa_u = \kappa_{u+1}$, then $V[u+1] = V'$. Otherwise, $V[u+1] =$

$V' + (\kappa_{u+1} - \kappa_u)A$. Hence the update of valid total winding numbers takes only $O(1)$ time per intersection point for each $u > r$.

Hence with $O(n \log n + N)$ pre-processing time, we can compute all $\sigma(r, j)$s for all $j > r$ for any fixed r, in $O(I \log I)$ time, and thus computing all $\mathrm{Tw}^*(C[r, u])$s for all $r \in [1, I]$ and all $u < r$, in $O(I^2 \log I)$ total time. Putting everything together, the dynamic programming problem can be solved in $O(n \log n + I^2 \log I + N)$ total time.