Continuous similarity measures for curves and surfaces

Citation for published version (APA):
Ophelders, T. A. E. (2018). *Continuous similarity measures for curves and surfaces*. [Phd Thesis 1 (Research TU/e / Graduation TU/e), Mathematics and Computer Science]. Technische Universiteit Eindhoven.

Document status and date:
Published: 29/08/2018

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Download date: 16. Nov. 2023

Continuous Similarity Measures for Curves and Surfaces

Tim Anton Elisa Ophelders

Continuous Similarity Measures for Curves and Surfaces

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Technische Universiteit Eindhoven, op gezag van de
rector magnificus prof.dr.ir. F.P.T. Baaijens, voor een
commissie aangewezen door het College voor
Promoties, in het openbaar te verdedigen
op woensdag 29 augustus 2018 om 16:00 uur

door

Tim Anton Elisa Ophelders

geboren te Geleen

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter: prof.dr. M.G.J. van den Brand

promotor: prof.dr. B. Speckmann

copromotor: dr. K.A. Buchin

leden: prof.dr. G. Vegter (Rijksuniversiteit Groningen)
prof.dr. J. Erickson (University of Illinois at Urbana-Champaign)
dr. S.Y. Oudot (École Polytechnique)
prof.dr. L.M.J. Florack
prof.dr. M.T. de Berg

Het onderzoek of ontwerp dat in dit proefschrift wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.

Netherlands Organisation for Scientific Research

The work in this thesis is supported by the Netherlands' Organization for Scientific Research (NWO) under project no. 639.023.208.

The work in the thesis has been carried out under the auspices of the research school IPA (Institute for Programming research and Algorithmics).

Cover design: Tim Ophelders
Printing: Ipskamp Printing
ISBN: 978-90-386-4546-9

© 2018 by Tim Anton Elisa Ophelders. All rights are reserved. Reproduction in whole or in part is prohibited without the written consent of the copyright owner.

A catalogue record is available from the
Eindhoven University of Technology Library

Technische Universiteit
Eindhoven
University of Technology

Contents

1 Introduction **1**
- 1.1 On Measuring Similarity · · · 3
- 1.2 Contributions · · · 5
  - 1.2.1 Part I—Fréchet Distance · · · 5
  - 1.2.2 Part II—Homotopy Height · · · 6

I Fréchet Distance

2 The Fréchet Distance between Real-Valued Surfaces **11**
- 2.1 Preliminaries · · · 12
- 2.2 Contour tree distance · · · 13
  - 2.2.1 Free space diagrams · · · 16
  - 2.2.2 Contour tree distance in NP · · · 17
  - 2.2.3 NP-hardness · · · 19
- 2.3 Surfaces · · · 21
  - 2.3.1 Euler diagrams · · · 21
  - 2.3.2 Euler diagrams from matchings · · · 22
  - 2.3.3 Matchings from Euler diagrams · · · 23
  - 2.3.4 NP-hardness · · · 26
- 2.4 An unrepresentable matching · · · 27
- 2.5 Related measures · · · 27
- 2.6 Discussion · · · 29

3 Horizontal Fréchet Isotopies and Monotonicity **31**
- 3.1 Preliminaries · · · 33
- 3.2 Disproving a conjecture · · · 34
- 3.3 Simple curves · · · 36
- 3.4 Isotopies between monotone curves · · · 38
- 3.5 Isotopies to monotone curves · · · 38
  - 3.5.1 Non-monotone isotopies · · · 39
  - 3.5.2 Horizontal length · · · 40
  - 3.5.3 Monotonizing isotopies · · · 40
  - 3.5.4 Monotonicity and relation to persistence · · · 42
- 3.6 Monotone homotopies · · · 44
- 3.7 Discussion · · · 50

II Homotopy Height

4 Computational Complexity of Optimal Homotopies 55
- 4.1 Preliminaries · · · 58
- 4.2 Isotopies and monotonicity of optimal homotopies · · · 60
- 4.3 Retractions and pausing at short cycles · · · 62
- 4.4 Computing homotopy height in NP · · · 64
- 4.5 Variants and approximation algorithms · · · 71
  - 4.5.1 Homotopic Fréchet distance · · · 71
  - 4.5.2 Minimum height linear layouts · · · 73
- 4.6 Discussion · · · 74

5 Optimal Homotopies over a Spiked Polygonal Plane 77
- 5.1 Variable-cost spikes · · · 80
- 5.2 Unit-cost spikes · · · 82
  - 5.2.1 Regular homotopies · · · 85
  - 5.2.2 Computation · · · 87
- 5.3 Discussion · · · 90

6 Concluding Remarks 93

References 97

Summary 105

Curriculum Vitæ 107

1

Introduction

Shapes are all around us. Some can be recognized by their silhouette alone, while others demand closer inspection. Shapes appear at all scales: in polymers, plants, veins, organs, sculptures, mountains, coast lines, asteroids, and even galaxies. Not all shapes are visible: consider for example the orbit of a planet, or the trajectory of a moving person or particle. A collection of database entries can be thought of as an even more abstract shape.

The ability to compare shapes is crucial to our understanding of the world. For instance, when classifying organisms, one must decide whether two plants are of the same species, or otherwise how and how much they differ. Many of the aforementioned objects change their shape over time. Tracking such a change in shape often aids in the prediction of disease outbreaks, floods, and other hazards. For objects that change their shape over time, one can ask whether two pictures are of the same object, perhaps at different points in time, or whether and why the pictured objects are different.

To answer the question of why two objects differ, it is helpful to know which parts of those objects correspond. In particular, if one can accurately assign to each feature of one shape, a distinct corresponding feature on the other shape, then one can easily pinpoint the differences between the objects by comparing those corresponding features. We illustrate this in Figure 1.1, where the two objects to be compared are sets of locations: one set is indicated by dots and the other by crosses. Comparing these objects is significantly easier if the locations that correspond are indicated.

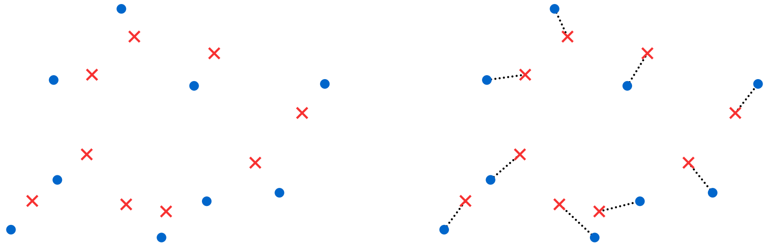


Figure 1.1: Two collections of objects. Without the corresponding features indicated (left), and with the corresponding features indicated (right).

Generally, finding a suitable correspondence is the major difficulty in comparing objects. Consider a second example: CT scans of human lungs. Two such scans are depicted in the top row of Figure 1.2, where brightness corresponds to density. In the figure, a person's lungs are visibly more compressed when lying on their back than when lying on their stomach. Since it is easy for humans to locate the lungs on both pictures, this comparison can readily be made. In contrast, comparing the two pictures becomes considerably harder if one does not know the context of what the pictures represent, or where the lungs are located. Since the top left picture is of a human lying on their back, whereas the top right picture is of a human lying on their stomach, it makes sense to rotate the second picture by 180 degrees. The result of this rotation can be seen on the bottom left. This rotation is helpful for comparing the two pictures, but still does not account for misalignments due to breathing and gravity, as can be seen in the bottom right, where the pictures are overlaid. To account for this, the image would need to be deformed more, such that corresponding features overlap perfectly.

This thesis presents methods for finding suitable correspondences between shapes in a variety of settings. In the absence of context, there tend to be so many possible correspondences that it becomes unclear which one to use. Naturally, we are looking for a correspondence that aligns features as well as possible. For this reason, we usually aim to find a correspondence for which the difference between corresponding features is as small as possible. However, without further restrictions, such a correspondence is often unrealistic, as two neighboring points on one shape might correspond to two points on the other shape that are far apart.

If the context is already known partially, we can discard correspondences that are not supported by that context. This usually reduces the number of correspondences that need to be considered significantly. Sometimes, context can be given in terms of what the compared objects represent, but sometimes the context is more abstract. For example, in contrast to the locations of Figure 1.1, the pixels of Figure 1.2 have an inherent adjacency structure, such that adjacent pixels of one

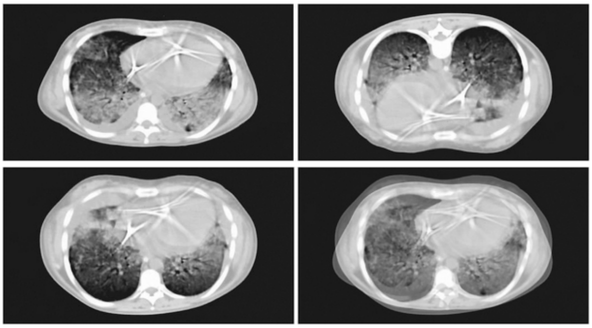


Figure 1.2: Computed tomography scans of lungs in supine (top left) and prone (top right) positions. The prone picture rotated 180 degrees (bottom right) and overlaid with the supine picture (bottom right). Original figure by L. Gattinoni and A. Protti [62] and reused with permission.

picture would naturally correspond to adjacent pixels of the other picture.

As they take into account the essential connectivity of the compared shapes, adjacency restrictions play an important role in the comparison of shapes with a structure of connectivity. Adjacency restrictions are taken into account for all shapes compared in this thesis. Such adjacency restrictions lead to fewer candidate correspondences. However, perhaps counter-intuitively, this does generally not simplify the task of finding the best correspondence, as techniques for finding an unrestricted optimum may no longer apply. This is because a correspondence that is optimal in the absence of context may not adhere to the restrictions imposed by a given context.

1.1 On Measuring Similarity

Before describing the technical contributions in this thesis in detail, we first discuss what it means to compare two shapes, starting with the simple case of two single points. For instance, an intuitive way to compare two points on a terrain is to measure the distance between them. However, there are many different ways in which the distance between two points can be defined. In fact, depending on the application, different ways are more suitable than others. The simplest definition of the distance between two points would be the length of the line segment connecting them. In the context of points on a terrain, this segment may hover above, or even pass through the terrain. As such, when the application is to hike from one side of a mountain to another, such a line segment might not be the desired measure of distance.

A much more general approach is to define the distance in terms of an optimal path between the two points. This path can be constrained to suit a specific application. The length of a shortest such path is an obvious candidate for the distance between the two points. There are various constraints one can impose on a path. In the context of a terrain, one can for example require paths to lie exactly on the terrain, or allow them to hover above the terrain, or even allow them to pass through the terrain.

We wish to extend this notion of distance measured via paths to shapes more complex than points. A next step in this direction is to move from comparing pairs of points to comparing two sets of multiple points. Such a comparison is based on a correspondence between the points of the two sets. Generally, a correspondence matches each point in one set to one or more points of the other set and vice-versa. Again, one can connect matched points with paths. We can then measure some property of the resulting set of paths. Commonly used properties are the total and the maximum length of those paths.

Depending on the application, it is not always clear which correspondence between two point sets should be used in the comparison. If there are no further constraints on this correspondence, a common approach is to use a correspondence that minimizes the measured property. An example is the Hausdorff distance, which minimizes the maximum length of a path between corresponding points. Most distance measures however, impose additional constraints on the correspondence. For instance, if the two sets have an equal number of points, the correspondence could require each point to be matched to exactly one point of the other set. When minimizing the total length of paths, this is called a minimum cost bipartite matching. When minimizing the maximum length, we speak of a bottleneck matching.

This thesis consists of two parts, the measures studied in both parts are based on bottleneck matchings. As with the Hausdorff distance, the distance measures considered in Part I of this thesis are based on the concept of finding a matching that minimizes the maximum distance between matched points.

In some applications, when considering multiple paths, it matters how these paths interact. For instance, when the application is to visualize a correspondence, paths may be required not to cross, allowing the correspondence to be seen at a glance. See also Figure 1.3. When the application is to plan a walk for a group of friends, paths may be required to follow similar routes. See also the Figure 1.3 (right), where three friends walk from the blue locations to the red locations along similar paths.

Since the paths themselves give no information on where a person is located at a given time, we cannot yet conclude that the friends will be walking close together. A more desirable way to model a walk for a group of friends is to equip each path with a temporal parameter, tracking the location of a person over time. We refer to a path equipped with such a temporal parameter as a trajectory. Any given time corresponds to a single point on the path and as time progresses, that point moves continuously from the matched point on one shape, along the path, to the matched point on the other shape. If each pair of matched points has a trajectory between them, we speak of a deformation. As such, a deformation is an abstract way of moving points matched by a correspondence in a synchronized manner. Such a deformation constitutes a way to interpolate between the two compared shapes: at any given time, the deformation yields an intermediate shape consisting of the points on each of the trajectories between matched points at that given time. Indeed, constructing morphs between given shapes is a major application of such deformations.

In addition to constraints on trajectories themselves, one can constrain the interactions between trajectories. For instance, requiring that the points of no two trajectories coincide at any given time. More generally, several aspects of a deformation can be constrained or optimized in order to make the comparison between the shapes more meaningful. An example of such a constraint is that all intermediate shapes are in some sense similar to the two compared shapes. In particular, for

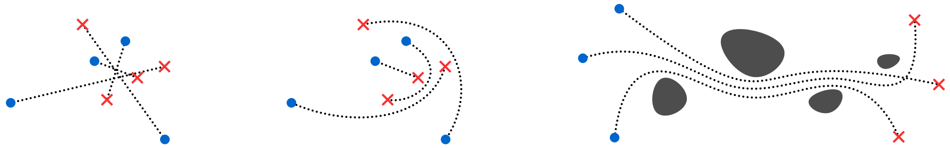


Figure 1.3: Crossing (left), non-crossing (center), and similarly routed paths between corresponding locations (right).

the distance measures considered in this thesis, the intermediate shapes preserve the connectivity of the compared shapes throughout the deformation. A deformation that preserves connectivity is called a homotopy. In Part II, we consider curves as the compared shapes, and minimize the length of the longest intermediate curve in the homotopy.

1.2 Contributions

The results in this thesis concern the computation of similarity measures between curves and surfaces. The similarity measures studied are all variations on the Fréchet distance, which is well-studied in the fields of computational geometry and topology. Abstractly, the Fréchet distance takes two shapes such as curves or surfaces as input, and outputs how similar they are based on the cost of a cheapest deformation from one input shape to the other.

One important difference between the original formulation of the Fréchet distance and the variants considered in this thesis is that the deformations studied are required to be continuous (homotopies). We sometimes even require the intermediate shapes throughout the deformation to remain simple (isotopies). The continuity requirement plays an important role if the shapes move through a non-Euclidean space.

Throughout a deformation, points can be traced as they move from their position on the first input shape to a position on the second input shape. Two options are considered for the cost of a deformation, namely the *width* and the *height*. The width of a deformation is the maximum total distance that a point moves throughout the deformation. The height of a deformation is the maximum length (if the input shapes are curves) attained by an intermediate curve of the deformation.

The following four combinations are covered in this thesis.

1. Minimum width homotopies between real-valued surfaces.
2. Minimum width isotopies between simple curves in the plane.
3. Minimum height homotopies between curves on surfaces.
4. Minimum height homotopies between curves in polygonal domains with weighted point-obstacles.

In addition to the contributions contained in this thesis, the author obtained results related to the Fréchet distance, moving objects, and combinatorial optimization [22, 70, 41, 68, 87, 42, 27, 43].

1.2.1 Part I—Fréchet Distance

Real-valued surfaces. In Chapter 2 we study the Fréchet distance between surfaces. Minimum width homotopies between surfaces have until recently hardly been studied. It is a long-standing open problem whether the cost of such a homotopy is even computable. We give an affirmative answer for the special case of real-valued surfaces of genus zero. In fact, in this special case, we show that the problem is NP-complete, even to approximation within a factor two. Furthermore, we give a lower bound for this measure in terms of a natural distance measure between the contour trees of the respective surfaces. Although this measure is intuitively easier to compute, we show that it is also NP-complete to approximate within a factor two.

This chapter is based on joint work with Kevin Buchin and Bettina Speckmann. This work appeared in the Proceedings of the 28th Symposium on Discrete Algorithms [23].

Isotopies. In Chapter 3 we consider minimum width isotopies between simple curves in the plane. In this setting, we disprove a conjecture about the width of an isotopy between two particular curves by providing an isotopy of almost half the conjectured width. Moreover, we show that if there is a direction in which the input curves are monotone, then the width of an optimal isotopy is the same as that of an optimal homotopy, which can be efficiently computed using known algorithms. We also give first steps towards computing optimal isotopies for curves that do not satisfy this monotonicity property. In particular, given an input curve and a direction, we compute an isotopy of minimal movement in that direction, in order to obtain a curve that is monotone in the same direction. Moreover, given two curves and a parameterization for each of them, we present an algorithm that decides whether there exists an isotopy that moves each point monotonely from one parameterized curve to the other.

This chapter is based on joint work with Kevin Buchin, Erin Chambers, and Bettina Speckmann. Part of this work appeared as an abstract in the 33rd European Workshop on Computational Geometry [21].

1.2.2 Part II—Homotopy Height

General computation. In Chapter 4 we study minimum height homotopies between curves on surfaces. We apply several recent results on the structure of optimal homotopies to achieve polynomial bounds on the combinatorial complexity of such a homotopy. This has as a consequence that this problem lies in the complexity class NP. Moreover, we relate this problem to an equivalent simple graph drawing problem.

This chapter is based on joint work with Erin Chambers and Arnaud de Mesmay. This work appeared in the Proceedings of the 29th Symposium on Discrete Algorithms [28].

Polygonal domains. In Chapter 5 we consider a more restricted setting of a polygonal domain with weighted point-obstacles. We show that for some instances, any optimal homotopy passes through a curve with a linear number of inflection points, even when the polygonal domain is convex. In contrast, we show that curves require at most one inflection point to sweep a convex polygonal domain if the point-obstacles have unit weight. More generally, we give a polynomial time algorithm to compute a minimum height homotopy in arbitrary polygonal domains with unit weight point-obstacles.

This chapter is based on joint work with Benjamin Burton, Erin Chambers, Marc van Kreveld, Wouter Meulemans, and Bettina Speckmann. This work appeared in the Proceedings of the 25th European Symposium on Algorithms [25].

Part I

Fréchet Distance

2

The Fréchet Distance between Real-Valued Surfaces

In the first part of this thesis, we aim to find deformations that minimize the distance that points travel as they move from one shape to the other. In this chapter, we consider the problem of comparing real-valued functions on surfaces, focusing in particular on spheres and disks of constant boundary. That is, functions $f\colon M \to \mathbb{R}$ where $f(x) = f(x')$ for all $x, x' \in \partial M$. The kind of similarity we investigate is that under continuous deformations of surfaces, such as in Figure 2.1. Here shapes that can be deformed into each other have distance 0, and shapes have some meaningful positive distance otherwise. There are two natural computational problems that arise for each measure, namely deciding whether two images have distance 0, and the more general problem of computing the distance between two images.

A problem with similar use cases is elastic image matching, which is known to be NP-complete [67]. In practice, this problem is currently approached using various heuristics [84].

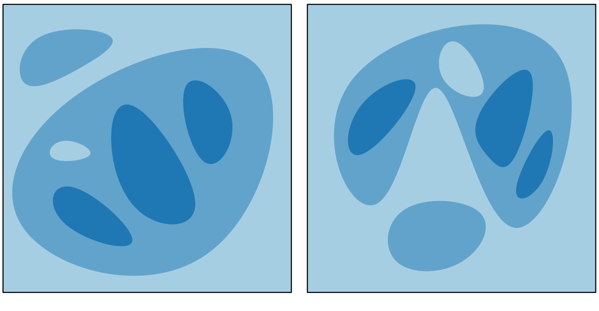


Figure 2.1: Two pictures that can continuously be deformed into each other.

2.1 Preliminaries

Given two functions f and $g\colon M \to \mathbb{R}^k$ with common parameter space M, their *Fréchet distance* [59, 60] is defined as

$$d_F(f,g) = \inf_{\mu\colon M\to M} \sup_{x\colon M} d(f(x), g\circ\mu(x)),$$

where $\mu\colon M \to M$ ranges over all orientation preserving homeomorphisms and $d(\cdot,\cdot)$ is the underlying norm of $\mathbb{R}^k$. Essentially, the Fréchet distance captures the similarity between two functions by realigning their parameter spaces to minimize the maximum difference in the function value of aligned points. It is often assumed that f and g are piecewise-linear functions.

Efficient algorithms for computing $d_F(f,g)$ exist for L^p norms if f and g are polylines [5], so if $M = [0,1]$ or $M = S^1$ for closed polylines. The computational complexity of this case is well understood with a significant amount of recent work addressing the fine-grained complexity of this case and its variants [1, 14, 17, 19]. Considerable progress has also been made recently on the problem of approximating the Fréchet distance between polylines [15, 50]. With many efficient algorithms at hand the Fréchet distance between polylines has found numerous applications, with recent work particularly focusing on geographic applications such as map-matching tracking data [12] and moving objects analysis [16].

The computational complexity in the case that f and $g\colon M \to \mathbb{R}^k$ are (triangulated) surfaces is much less understood. The problem is known to be NP-hard [63] also when $k = 2$ [18, 22]. But it is not known whether it is actually in NP, in fact until recently it was only known to be upper semi-computable for surfaces in $\mathbb{R}^k$ [4, 24]. Recently, major results towards computability [77, 78] for surfaces of genus 0 were found, by introducing an exponential time approximation algorithm.

We show that for $k = 1$, the problem is in NP[1] if M is a topological sphere or disk with constant boundary. Additionally, we show that already for $k = 1$ computing the Fréchet distance is NP-hard. Even for this case, no efficient constant-factor approximation algorithms have been found so far, and we show that computing a factor $2 - \varepsilon$ approximation is NP-hard.

In previous work, a few variants on the comparison of surfaces under the Fréchet distance have been investigated. For instance, there are efficient algorithms for computing the Fréchet distance with certain constraints on the homeomorphisms μ [22] and for computing the *weak Fréchet distance* [4] between triangulated surfaces homeomorphic to the disk. The Fréchet distance can be computed in polynomial time for simple polygons (including interior) in the plane [20], even if they have one [18] or more holes [76]. An efficient constant-factor approximation exists for so-called folded polygons, which need not lie in the plane [39].

In addition to our results on surfaces (Section 2.3), we define a suitable similarity measure between contour trees, which we show to be NP-complete to compute as well (Section 2.2). Recently, other measures [8, 9, 10, 48, 75] such as the interleaving distance or functional distortion distance have been investigated for contour trees. In particular, it is expected that the interleaving distance and the functional distortion distance are equal, but so far it has been proven only that they differ by at most a constant factor. For computing the interleaving distance, no efficient exact algorithms are known. We also show a relation to the Gromov-Hausdorff distance between trees, for which an NP-hardness proof was recently found [2]. The same paper proposes polynomial-time approximation algorithms with an approximation ratio of $O(\min(n, \sqrt{rn}))$ (where r is the length ratio between the longest and shortest edge) for the Gromov-Hausdorff and interleaving distances. We highlight how the Fréchet distance and weak Fréchet distance differ from these measures in Section 2.5.

---

[1] Here we assume that we can compare numbers in the input (and sums and differences between pairs of input numbers) in polynomial time.

Outline. Our main result is the NP-completeness of the Fréchet distance between $\mathbb{R}$-valued surfaces. Because of its combinatorial complexity, we start by defining a combinatorially simpler problem, the contour tree distance. The contour tree distance is a similarity measure related to the Fréchet distance, that abstracts from the geometric representation of the matching μ. As such, the contour tree distance is a lower bound on the Fréchet distance.

We show that computing the contour tree distance is in NP using properties of the free space diagram, which is a commonly used tool in algorithms for computing the Fréchet distance between curves. Using extensions of the free space diagram, valid matchings between contour trees can be analyzed easily. Nonetheless, it turns out that approximating the contour tree distance within a factor 2 is NP-hard by a reduction from the NP-hard problem Exact Cover by 3-sets.

Using specific properties of the matchings used in this reduction, these matchings can be extended to matchings μ between surfaces. A reduction from the contour tree distance to the Fréchet distance between $\mathbb{R}$-valued surfaces then shows that approximating the Fréchet distance within a factor 2 is also NP-hard. However, not all matchings between contour trees have an extension to a matching between surfaces, as discussed in Section 2.4.

To show that computing the Fréchet distance is in NP, we make use of a polynomial amount of information of a matching μ. This information is used to represent a matching as an Euler diagram, for which the recognition problem is in NP.

2.2 Contour tree distance

The Reeb graph [80] of a function $f\colon M \to \mathbb{R}$ is the quotient space $M/{\sim_f}$ where $a \sim_f b$ if and only if a and b are in the same connected component of the level set $f^{-1}(f(a))$. Denote by $\mathcal{R}_f$ the corresponding quotient map. Because f associates a single real number to each equivalence class of $\sim_f$, the resulting Reeb graph has a natural $\mathbb{R}$-valued function associated with it, namely the (unique) function $f'\colon M/{\sim_f} \to \mathbb{R}$ satisfying $f' \circ \mathcal{R}_f = f$. If M is the disk or the 2-sphere, the Reeb graph forms a tree called a contour tree.

For the sake of representation, in this chapter we assume each surface to be triangulated, to form a simplicial 2-complex. Moreover, we assume without loss of generality that function values along edges of Reeb graphs are interpolated linearly between the values of the vertices at their endpoints. In this representation, the contour tree of a surface with n faces has complexity $O(n)$ and can be computed in $O(n \log n)$ time [88]. Intuitively, the contour tree has a vertex for each minimum, maximum and saddle point, and edges between 'adjacent' vertices.

Based on the Fréchet distance between f and g, we derive a computationally simpler measure that abstracts from the realizability of the matching μ between spheres or disks. Throughout this chapter, we use the notation $\mathbb{X} = M/{\sim_f}$ and $\mathbb{Y} = M/{\sim_g}$ for the contour trees of f and g, respectively. We shall denote the vertex set of $\mathbb{X}$ by $V(\mathbb{X})$ and its edge set by $E(\mathbb{X})$. With slight abuse of notation, we reuse function names f and g for the natural $\mathbb{R}$-valued functions associated with the contour trees $\mathbb{X}$ and $\mathbb{Y}$. Our distance measure d_C compares the contour trees $\mathbb{X}$ and $\mathbb{Y}$ of f and g. We define the contour tree distance d_C as

$$d_C(f\colon \mathbb{X} \to \mathbb{R}, g\colon \mathbb{Y} \to \mathbb{R}) = \inf_{\tau \in \mathcal{M}(\mathbb{X},\mathbb{Y})} \sup_{(x,y)\in\tau} |f(x) - g(y)|,$$

where $\tau \subseteq \mathbb{X} \times \mathbb{Y}$ is drawn from some class of matchings $\mathcal{M}(\mathbb{X}, \mathbb{Y})$, defined below. So τ defines a correspondence between contour trees, such that $(x, y) \in \tau$ if some points on contours x and y were matched by a corresponding matching μ on M. Denote $\tau(x) = \{y \mid (x,y) \in \tau\}$ and $\tau^{-1}(y) = \{x \mid (x,y) \in \tau\}$. The class $\mathcal{M}(\mathbb{X}, \mathbb{Y})$ captures the essential (but not all) properties of an orientation preserving matching μ. We define $\mathcal{M}(\mathbb{X}, \mathbb{Y})$ as the set of matchings τ for which

1. τ is a connected subset of $\mathbb{X} \times \mathbb{Y}$; and
2. $\tau(x)$ is a nonempty subtree of $\mathbb{Y}$ for each $x \in \mathbb{X}$; and
3. $\tau^{-1}(y)$ is a nonempty subtree of $\mathbb{X}$ for each $y \in \mathbb{Y}$.

Here, the term *subtree* is used to denote a connected subset of a tree, not necessarily containing leaves or complete edges of that tree. By Conditions 2. and 3., each connected set matches with a connected set, and Condition 1. ensures continuity. Let $x \rightsquigarrow_{\mathbb{X}} x'$ denote all (possibly backtracking) paths from x to x' in the underlying space $\mathbb{X}$. The following properties can be derived immediately:

1. $\tau(\mathbb{X}) = \mathbb{Y}$ and $\tau^{-1}(\mathbb{Y}) = \mathbb{X}$;
2. for each $p_f\colon x \rightsquigarrow_{\mathbb{X}} x'$, some path $p\colon (x,y) \rightsquigarrow_{\mathbb{X}\times\mathbb{Y}} (x', y')$, whose projection onto $\mathbb{X}$ equals p_f, satisfies $p \subseteq \tau$;
3. for each $p_g\colon y \rightsquigarrow_{\mathbb{Y}} y'$, some path $p\colon (x,y) \rightsquigarrow_{\mathbb{X}\times\mathbb{Y}} (x', y')$, whose projection onto $\mathbb{Y}$ equals p_g, satisfies $p \subseteq \tau$;
4. if $\{(x,y), (x',y)\} \subseteq \tau$ then $p \times \{y\} \subseteq \tau$ for some $p\colon x \rightsquigarrow_{\mathbb{X}} x'$;
5. if $\{(x,y), (x,y')\} \subseteq \tau$ then $\{x\} \times p \subseteq \tau$ for some $p\colon y \rightsquigarrow_{\mathbb{Y}} y'$.

An example of a matching between two trees is shown in Figure 2.2. The two-dimensional patch illustrates a many-to-many correspondence. For a matching

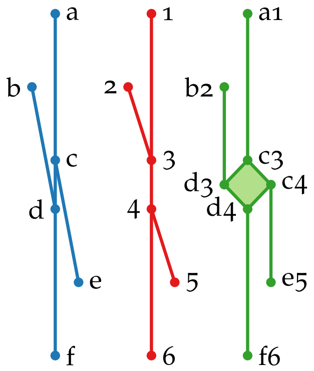


Figure 2.2: Two trees (left) and a matching (right).

$\mu \colon M \to M$ between surfaces f and g, define τ_μ to be the corresponding matching between the Reeb graphs of f and g:

$$\tau_\mu = \{(\mathcal{R}_f(x), \mathcal{R}_g \circ \mu(x)) \mid x \in M\}.$$

By Lemma 2.1 we have for each orientation preserving homeomorphism μ, that the matching $\tau_\mu \in \mathcal{M}(\mathbb{X}, \mathbb{Y})$, and hence that $d_C(f,g) \leq d_F(f,g)$. On the other hand, a matching $\tau \in \mathcal{M}(\mathbb{X}, \mathbb{Y})$ does not need to correspond to an orientation preserving homeomorphism on M. We give an example of such a matching in Section 2.4.

Lemma 2.1. *$\tau_\mu \in \mathcal{M}(\mathbb{X}, \mathbb{Y})$ for any orientation preserving homeomorphism μ on M.*

Proof. Consider such a matching $\mu \colon M \to M$. We prove all three conditions of $\mathcal{M}(\mathbb{X}, \mathbb{Y})$ for τ_μ. The set $\{(x,y) \mid \mu(x) = y\}$ is a connected subset of $S^2 \times S^2$. Hence its image under the quotient map $(x,y) \mapsto (\mathcal{R}_f(x), \mathcal{R}_g(y))$ is connected, so Condition **1**. holds. Because μ is a homeomorphism on a connected set, $\tau_\mu(x)$ is connected, and by surjectivity of τ_μ^{-1} nonempty. So, $\tau_\mu(x) \subseteq \mathbb{Y}$ and symmetrically $\tau_\mu^{-1}(y) \subseteq \mathbb{X}$ is a nonempty subtree. ☐

Corollary 2.1. *$d_C(f,g) \leq d_F(f,g)$.*

To test whether the contour tree distance between two trees is zero, one only needs to test whether the trees are equal. We represent trees canonically by exhaustively removing degree 2 vertices that lie on the line segment connecting the two adjacent vertices, and replacing them by a single edge between those vertices. This reduces the problem of testing a contour tree distance of zero to labeled unordered unrooted tree isomorphism, which can be solved in linear time [3].

Note that the contour tree distance and Fréchet distance between trees are different problems. In fact, one major limitation of the Fréchet distance for trees is that non-homeomorphic trees have infinite Fréchet distance. The Fréchet distance between trees can be computed in $O(n^{5/2})$ time [18], whereas we will see that computing the contour tree distance is considerably harder.

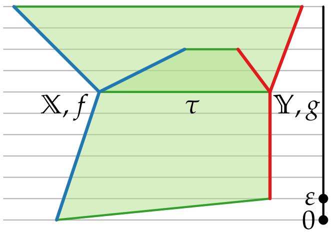


Figure 2.3: Trees $\mathbb{X}$ and $\mathbb{Y}$ with an ε-matching τ.

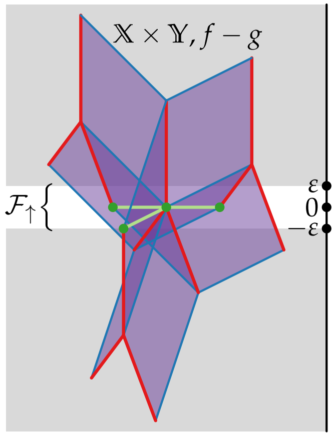


Figure 2.4: Matching τ in the free space $\mathcal{F}_\varepsilon$.

2.2.1 Free space diagrams

We consider the decision problem that asks whether the contour tree distance between two $\mathbb{R}$-valued trees is at most ε. The *free space diagram* is a commonly used data structure in algorithms for computing the Fréchet distance. We can also formulate a *free space* in relation to the contour tree distance. Define the *ε-free space* $\mathcal{F}_\varepsilon$ of f and g to be the set of pairs in the product of the parameter spaces whose respective images are at most ε apart:

$$\mathcal{F}_\varepsilon = \{(x, y) \in \mathbb{X} \times \mathbb{Y} \mid |f(x) - g(y)| \leq \varepsilon\}.$$

Define an *ε-matching* to be a matching $\tau \in \mathcal{M}(\mathbb{X}, \mathbb{Y})$ for which all $(x, y) \in \tau$ satisfy $|f(x) - g(y)| \leq \varepsilon$. Then an ε-matching $\tau \subseteq \mathcal{F}_\varepsilon$ exists if (and only if) the answer to the decision problem is yes. The contour tree distance is the minimum value of ε for which an ε-matching exists. In Figure 2.4, a product parameter space containing the ε-matching of Figure 2.3 is drawn in green as a projection from four-dimensional space.

Without loss of generality, assume function values along edges of $\mathbb{X}$ and $\mathbb{Y}$ to be interpolated linearly. A commonly used property of the free space is that for each edge $e_x \in E(\mathbb{X})$ and edge $e_y \in E(\mathbb{Y})$, the face $e_x \times e_y \in E(\mathbb{X}) \times E(\mathbb{Y})$ has a convex intersection with the free space $\mathcal{F}_\varepsilon$ (see Lemma 2.2, originally proven in [5]). An important consequence of this lemma which we will see in Section 2.2.2 is that given only information about the intersection of the ε-free space with the four edges bounding such a face, sufficient information about the interior of the face can be derived to decide whether an ε-matching exists.

Lemma 2.2. *For edges $e_x \in E(\mathbb{X})$ and $e_y \in E(\mathbb{Y})$ on which the respective functions f and g are linear, the face $F = e_x \times e_y$ has a convex intersection with $\mathcal{F}_\varepsilon$ for any ε.*

Proof. The map $(x, y) \mapsto f(x) - g(y)$ is affine when restricted to domain F. The preimage of a convex set (the ε-ball) under an affine map is convex. Intersecting this convex preimage with the convex set F yields again a convex set. Hence $F \cap \mathcal{F}_\varepsilon$ is convex. $\square$

Let ε be the value of the contour tree distance, then ε is the minimum value for which all constraints of $\mathcal{M}(\mathbb{X}, \mathbb{Y})$ can be satisfied by some matching τ. Hence ε is the minimum value for which either some vertex $(x, y) \in V(\mathbb{X}) \times V(\mathbb{Y})$ lies in $\mathcal{F}_\varepsilon$, or for two (possibly identical) edges e_1 and $e_2 \in V(\mathbb{X}) \times E(\mathbb{Y})$ or e_1 and $e_2 \in E(\mathbb{X}) \times V(\mathbb{Y})$ of $\mathbb{X} \times \mathbb{Y}$, the sets $e_1 \cap \mathcal{F}_\varepsilon$ and $e_2 \cap \mathcal{F}_\varepsilon$ contain points with the same y- or x-coordinates, respectively. If ε is such a minimum value, we call ε a *critical value*, and there are $O(n^3)$ of them. For two functions f and g the $O(n^3)$ critical values of ε can each be computed in constant time. However, as shown in Section 2.2.3, determining which critical value corresponds to the contour tree distance is NP-hard. First we show that computing the contour tree distance is in NP.

2.2.2 Contour tree distance in NP

Given only a polynomial amount of information about an ε-matching τ, we show that it can be verified in polynomial time that an ε-matching $\tau' \in \mathcal{M}(\mathbb{X}, \mathbb{Y})$ exists. The information we use is for each vertex x of $\mathbb{X}$ the endpoints and vertices of the corresponding subtree $\tau(x)$ of $\mathbb{Y}$, and for every vertex y of $\mathbb{Y}$ the endpoints and vertices of the corresponding subtree $\tau^{-1}(y)$ of $\mathbb{X}$. This set of points P_τ is defined more formally in the equation below. To illustrate, if vertex x maps to the subtree $\tau(x)$ of $\mathbb{Y}$ shown in Figure 2.5, then exactly the four points of $\mathbb{Y}$ highlighted in green will appear paired with x in P_τ. For any matching $\tau \in \mathcal{M}(\mathbb{X}, \mathbb{Y})$, there are only $O(n^2)$ points in the set P_τ. However, it should be noted that such points give no information about the behavior of τ in the interior of faces of $\mathbb{X} \times \mathbb{Y}$.

$$\begin{aligned} P_\tau = {} & \{(x, y) \mid x \in V(\mathbb{X}) \wedge y \in \partial(\tau(x))\} \cup \\ & \{(x, y) \mid y \in V(\mathbb{Y}) \wedge x \in \partial(\tau^{-1}(y))\} \cup \\ & (\tau \cap (V(\mathbb{X}) \times V(\mathbb{Y}))), \\ & \text{where } \partial \text{ is the boundary operator.} \end{aligned}$$

Let $CH(P)$ be the convex hull (including its interior) of a point set P. For a given set P_τ, derive τ' using Equation 2.2.2; such that for each face F, τ' consists of the

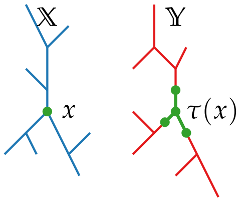


Figure 2.5: A subtree $\tau(x)$ of $\mathbb{Y}$ (green) matched with a vertex x of $\mathbb{X}$.

convex hull of $F \cap P_\tau$. In Lemma 2.3, we show that τ' is a valid ε-matching if P_τ is derived from an ε-matching τ.

$$\tau' = \bigcup_{F \in E(\mathbb{X}) \times E(\mathbb{Y})} CH(F \cap P_\tau).$$

Lemma 2.3. *If $\tau \in \mathcal{M}(\mathbb{X}, \mathbb{Y})$ is an ε-matching, then τ' derived from P_τ is an ε-matching with $\tau' \in \mathcal{M}(\mathbb{X}, \mathbb{Y})$.*

Proof. If τ was an ε-matching, then by Lemma 2.2, τ' is a subset of $\mathcal{F}_\varepsilon$. It remains to verify the three conditions of $\mathcal{M}(\mathbb{X}, \mathbb{Y})$ for τ'. Any face of $\mathbb{X} \times \mathbb{Y}$ will contain at most one component of τ by convexity. We show that all nonempty faces are connected, indeed the intersection of the boundary of a face with τ' is the same as that with τ. Hence τ' cannot have more components than τ. Then because τ is connected, τ' must also be connected, satisfying Condition **1.**.

For each vertex $x \in V(\mathbb{X})$, we know that $\tau'(x)$ is a nonempty tree. We show that this holds even if x is internal to some edge $e_x \in E(\mathbb{X})$. Because τ' is connected and has nonempty trees at the endpoints of e_x, the set $\tau'(x)$ cannot be empty. So suppose for contradiction that $\tau'(x)$ has multiple components, then because the interior of each face is convex, we must have $y \notin \tau'(x)$ for some vertex $y \in \tau(x)$ which is not possible. So Condition **2.** and by a symmetric argument Condition **3.** holds. ☐

Therefore, if ε is the contour tree distance, a set P_τ of polynomial size exists, whose derived τ' is an ε-matching. Given P_τ, it is easily tested whether for its derived τ', the properties required by $\mathcal{M}(\mathbb{X}, \mathbb{Y})$ are satisfied. Moreover, as a consequence of Lemma 2.2, it can be verified that τ' is an ε-matching by checking that $|f(x) - g(y)| \leq \varepsilon$ for all $(x, y) \in P_\tau$. Hence, Theorem 2.1 follows.

Theorem 2.1. *Deciding whether the contour tree distance between $\mathbb{X}$ and $\mathbb{Y}$ is at most ε is in NP.*

Remark. The proof that the contour tree distance is in NP extends to trees $\mathbb{X}$ and $\mathbb{Y}$ in higher dimensions. Specifically, if $f\colon \mathbb{X} \to \mathbb{R}^k$ and $g\colon \mathbb{Y} \to \mathbb{R}^k$, then computing $d_C(f, g) = \inf_{\tau \in \mathcal{M}(\mathbb{X}, \mathbb{Y})} \sup_{(x,y) \in \tau} \|f(x) - g(y)\|_p$ is also in NP for L^p norms $\|\cdot\|_p$.

2.2.3 NP-hardness

We show that approximating the contour tree distance between $\mathbb{R}$-valued trees within factor 2 is NP-hard by a reduction from the NP-hard problem Exact Cover by 3-sets [61].

Definition 2.1. *Exact Cover by 3-sets (X3C) Input: A set $\mathcal{S}$ of m subsets of $\{1,\dots,k\}$, each of size 3.*
Output: Does a subset (consisting of $k/3$ triples) of $\mathcal{S}$ partition $\{1,\dots,k\}$?

We introduce the gadgets used in our reduction in Figure 2.6. For this, define a zig-zag of radius r centered at position p along a segment $[a,b]$ to be a path visiting vertices at positions a, $p+r$, $p-r$ and b, in that order. Gadget Y^* is a long segment from position 0 to position $6k+6$. Gadget Y_l (with $l \in \{1,\dots,k\}$) is a path from position 1 to $6k+6$ with a zig-zag of radius 2 around position $6l$. Similarly, gadget $X_{i,j}$ ($i \in \{1,\dots,m\}$, $j \in \{1,2,3\}$) has a zig-zag around position $6 \cdot s(i,j)$, but with radius 1. The function s aligns the center of the zig-zag of $X_{i,j}$ with that of $Y_{s(i,j)}$, such that gadget $X_{i,j}$ has a contour tree distance of 1 to Y^* and $Y_{s(i,j)}$, but a contour tree distance of 2 to any gadget Y_l with $l \neq s(i,j)$.
The function s can be configured such that each triple of gadgets $(X_{i,1},\ X_{i,2},\ X_{i,3})$ corresponds to one of the m subsets of $\mathcal{S}$. We connect the three elements of each triple at a common vertex at position 1, and finally connect all triples at a common vertex at position 2 (blue in Figure 2.7) to form tree $f\colon \mathbb{X} \to \mathbb{R}$.

Similarly, each gadget Y_l corresponds to an element of $\{1,\dots,k\}$, and all Y_l are connected to a common vertex at position 1. The idea is that to obtain a low contour tree distance, $k/3$ triples of f must match the Y_l gadgets exactly; then what remains in f are $m-k/3$ triples that must be matched elsewhere. Each such unmatched triple of f will then be forced to match with three copies of Y^*, connected at a vertex at position 0 to form a so called Y^*-triple. We use $m-k/3$ such Y^*-triples, each connected to the Y_l gadgets at position 1 to form tree $g\colon \mathbb{Y} \to \mathbb{R}$. In Lemma 2.4, we use a solution to X3C to derive a matching using only many-to-one correspondences between f and g, even though many-to-many correspondences are permitted by $\mathcal{M}(\mathbb{X},\mathbb{Y})$.

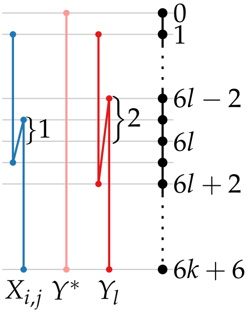


Figure 2.6: Gadgets.

Lemma 2.4. *$d_C(f,g) \leq 1$ if $\mathcal{S}$ admits a solution to X3C, even if only many-to-one correspondences are allowed.*

Proof. Consider the correspondence (as given by a solution to X3C) between gadgets $X_{i,j}$ of f and gadgets Y_l of g, and let each Y^*-triple correspond to an unused $(X_{i,1},\ X_{i,2},\ X_{i,3})$-triple of f. Then each individual gadget has distance at most 1 to the corresponding gadget of the other tree. The connections between gadgets (see Figure 2.7) also have a correspondence of distance at most 1 by matching connections of f that are part of the solution with the vertex of g at position 1 and matching remaining connections with connections to Y^*-triples. □

Lemma 2.5. *$d_C(f,g) \geq 2$ unless $\mathcal{S}$ admits a solution to X3C.*

Proof. Assume instead that $d_F(f,g) < 2$, then for some matching τ between f and g, the distance between all matched points is less than 2.

The endpoint $\textsc{End}(x_{i,j})$ at position $6k+6$ for each gadget $X_{i,j}$ of f must be matched to a connected component of the preimage of $(6k+4, 6k+8)$ under g. Since gadgets are connected only in the interval $[0,2]$ which is disjoint from $(6k+4, 6k+8)$, each such component corresponds to a single gadget of g. Therefore each $X_{i,j}$ is matched with a single gadget of g. By a symmetric argument, the endpoint of each gadget of g is matched with a single gadget of f. Hence, a one-to-one correspondence between gadgets of f and g exists.

Now suppose two gadgets $X = X_{i,j}$ and $X' = X_{i',j'}$ of different triples (that is, $i \neq i'$) correspond to two gadgets of the same Y^*-triple. Then each path connecting X and X' contains a point at position 2 which must be matched with the vertex at 0 that connects the two Y^* gadgets, yielding a contradiction. Thus, each triple $(X_{i,1},\ X_{i,2},\ X_{i,3})$ corresponds either to a single Y^*-triple, or three Y_l gadgets. Suppose $X_{i,j}$ corresponds to Y_l with $l \neq s(i,j)$, then the zig-zag of Y_l corresponds to a monotone path (without a zig-zag) of $X_{i,j}$. Let the vertices of Y_l at positions $6l+2$ and $6l-2$ be denoted v^+ and v^- respectively. By continuity of τ, we have $6l < f(\tau^{-1}(v^+)) \leq f(\tau^{-1}(v^-)) < 6l$, which is impossible. Hence, the contour tree distance between f and g is at least 2 if $\mathcal{S}$ admits no solution to X3C. □

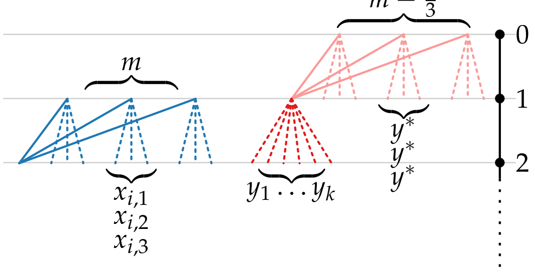


Figure 2.7: Gadgets are connected to form trees f (blue) and g (red).

Combining these lemmas with Theorem 2.1 we obtain Theorem 2.2.

Theorem 2.2. *Computing a $(2-\varepsilon)$-approximation of the contour tree distance is NP-complete, even if only many-to-one correspondences are allowed.*

2.3 Surfaces

We wish to use the fact that computing the contour tree distance is in NP to prove that the Fréchet distance between $\mathbb{R}$-valued surfaces is also in NP. We will consider two surfaces: the disk $[0,1]^2$ and the sphere S^2. It turns out that not all matchings between the contour trees $\mathbb{X}$ and $\mathbb{Y}$ can be realized as orientation preserving homeomorphisms on the sphere. An example of such a matching is described in Section 2.4. In the case of the disk, the boundaries must also be matched, which imposes additional constraints on the matching of the interiors.

In Section 2.3, we show that given a polynomial amount of information about an ε-matching τ between contour trees, we can verify in NP that an ε-matching μ on the disk exists. We use this to prove that the Fréchet distance between $\mathbb{R}$-valued spheres or disks with constant boundary is in NP. For this we use properties of Euler diagrams, which are described in Section 2.3.1. The relation between matchings and Euler diagrams is discussed in Section 2.3.2, and how Euler diagrams can be used to derive matchings is discussed in Section 2.3.3.

2.3.1 Euler diagrams

An Euler diagram is a set of topological disks, drawn in the plane to capture relations such as overlap or containment for pairs of disks. Eight such relations exist [54], namely $Rel = \{disjoint, equal, inside, contains, covered, cover, meet, overlap\}$, whose meanings are as follows.

$$\begin{aligned}
disjoint(a,b) &\equiv a \cap b = \emptyset;\\
equal(a,b) &\equiv a = b;\\
inside(a,b) &\equiv a \subseteq b \text{ and } \partial a \cap \partial b = \emptyset;\\
contains(a,b) &\equiv inside(b,a);\\
covered(a,b) &\equiv a \subseteq b \text{ and } \partial a \cap \partial b \neq \emptyset;\\
cover(a,b) &\equiv covered(b,a);\\
meet(a,b) &\equiv a \cap b \subseteq \partial a \cap \partial b \neq \emptyset;\\
overlap(a,b) &\equiv a \not\subseteq b \text{ and } b \not\subseteq a.
\end{aligned}$$

For a set S of n elements, a *topological expression* [82, 89] on S is a function $\mathcal{P} : S \times S \to Rel$. We say a topological expression $\mathcal{P}$ on S is satisfiable if and only if there is an Euler diagram of disks $\{D(i) \mid i \in S\}$, such that for any $i, j \in S$, the relation between $D(i)$ and $D(j)$ is as given by $\mathcal{P}(i,j)$. It was shown by Schaefer, Sedgwick and Štefankovič [81] that deciding whether a topological expression $\mathcal{P}$ on n elements is satisfiable (by an Euler diagram in the plane) is in NP.

2.3.2 Euler diagrams from matchings

We show that deciding whether the Fréchet distance between $\mathbb{R}$-valued surfaces f and $g\colon M \to \mathbb{R}$ is at most ε is in NP if M is a sphere. First consider the case where $M = [0,1]^2$ is a topological disk and that the boundary of f as well as g has a constant function value, so $f(p) = f(q)$ and $g(p) = g(q)$ for all $p, q \in \partial M$.

Define a *refinement* of a contour tree $\mathbb{X}$ to be a tree homeomorphic to $\mathbb{X}$, whose edges are subdivided using extra vertices. Assume that $\mathbb{X}$ and $\mathbb{Y}$ both have a vertex (or add it otherwise) for the contour containing the boundary, which we treat as the root of the trees. We assume all edges are directed away from the root.

For a vertex v of $\mathbb{X}$ with d outgoing edges, $\mathcal{R}_{\mathbb{X}}^{-1}(x)$ is a region homotopic to a d-holed disk on M. We index the boundary components of such a vertex by v and each of its outgoing edges, so that the set $W(\mathbb{X}) = V(\mathbb{X}) \cup E(\mathbb{X})$ of vertices and edges indexes all boundary components $\{c_w \mid w \in W(\mathbb{X})\}$ of vertices of $\mathbb{X}$. For a matching $\mu\colon M \to M$, for $w \in W(\mathbb{X})$ we are interested in the corresponding subtrees $\rho(w) = \mathcal{R}_{\mathbb{Y}} \circ \mathcal{R}_{\mathbb{X}}^{-1}(c_w)$ on $\mathbb{Y}$, and for $w \in W(\mathbb{Y})$ we are interested in the subtrees $\rho^{-1}(w) = \mathcal{R}_{\mathbb{X}} \circ \mu^{-1}(c_w)$ on $\mathbb{X}$. We can derive ρ given $\mu\colon M \to M$, and refine $\mathbb{X}$ into $\mathbb{X}'$ by adding vertices along edges at $\bigcup_{y \in W(\mathbb{Y})} \partial\rho^{-1}(y)$ and refine $\mathbb{Y}$ into $\mathbb{Y}'$ by adding vertices at $\bigcup_{x \in W(\mathbb{X})} \partial\rho(x)$. Since $\partial\rho^{-1}(y)$ and $\partial\rho(x)$ have $O(n)$ points per $y \in W(\mathbb{Y})$ or $x \in W(\mathbb{X})$, trees $\mathbb{X}'$ and $\mathbb{Y}'$ have $O(n^2)$ vertices.

For a vertex $x \in V(\mathbb{X}')$ of a rooted tree, let $sub(x) \subseteq \mathbb{X}$ be the subtree rooted at x, and for an edge $e = (x, x') \in E(\mathbb{X})$ from x to x', let $sub(e) = (e \cup sub(x')) \setminus \{x\}$. For each $i \in W(\mathbb{X}')$, there is a disk $D_\mu(i) = \mathcal{R}_f^{-1}(sub(i)) \subseteq M$, and for each $i \in W(\mathbb{Y}')$, there is a disk $D_\mu(i) = \mu^{-1}(\mathcal{R}_g^{-1}(sub(i))) \subseteq M$. We can now derive an Euler diagram from μ with $O(n^2)$ disks $\{D_\mu(i) \mid i \in W(\mathbb{X}') \cup W(\mathbb{Y}')\}$. The corresponding topological expression, call it $\mathcal{P}_\mu$, consists of $O(n^4)$ relations between these disks.

We capture the general structure of such expressions later, when we define topological expressions respecting τ_μ. For that, we first capture the structure of topological expressions for the disks of a single tree $\mathbb{X}'$, by saying a topological expression $\mathcal{P}$ on $S \supseteq W(\mathbb{X}')$ *respects* $\mathbb{X}$ if

1. for all $e = (i, j) \in E(\mathbb{X}')$, $\mathcal{P}(i, e) = contains$ and $\mathcal{P}(e, j) = contains$;
2. for all $e = (i, j) \in E(\mathbb{X}')$ and $e' = (i, j') \in E(\mathbb{X}')$ with $e \neq e'$, $\mathcal{P}(e, e') = disjoint$.

For a satisfiable topological expression respecting $\mathbb{X}$, the drawing of the disks of $W(\mathbb{X}')$ is the same (up to orientation preserving homeomorphism) for any satisfying Euler diagram Φ. Let $\{D(i) \mid i \in S\}$ be the disks of such an Euler diagram Φ. Then, for $x \in V(\mathbb{X}')$, define ϕ_x as the closed region bounded by $\partial D(x)$ and $\bigcup_{e=(x,x') \in E(\mathbb{X}')} \partial D(e)$, which is a disk with a hole for each outgoing edge of x. For $e = (x, x') \in E(\mathbb{X}')$, define ϕ_e as the closed annulus bounded by $\partial D(e)$ and $\partial D(x')$. Denote $\phi_{V(\mathbb{X}')} = \bigcup_{x \in V(\mathbb{X}')} \phi_x$ and $\phi_{E(\mathbb{X}')} = \bigcup_{e \in E(\mathbb{X}')} \phi_e$.

We say a topological expression $\mathcal{P}$ on $W(\mathbb{X}') \cup W(\mathbb{Y}')$ respects a matching $\tau \in \mathcal{M}(\mathbb{X}, \mathbb{Y})$ if the following conditions hold.

1. $\mathcal{P}$ respects both $\mathbb{X}$ and $\mathbb{Y}$;
2. $\mathcal{P}(root(\mathbb{X}'), root(\mathbb{Y}')) = equal$;
3. for $x \in V(\mathbb{X}')$ and $y \in V(\mathbb{Y}')$, ϕ_x intersects ϕ_y only if $y \in \tau(x)$ (and $x \in \tau^{-1}(y)$);
4. for each component of $\partial\phi_x$ with $x \in V(\mathbb{X}')$, it intersects ϕ_y for some $y \in V(\mathbb{Y}')$;
5. for each component of $\partial\phi_y$ with $y \in V(\mathbb{Y}')$, it intersects ϕ_x for some $x \in V(\mathbb{X}')$;

Any matching μ has an Euler diagram satisfying a topological expression respecting τ_μ.

Because we use only a polynomial amount of information about τ, and intersection relations between ϕ_x and ϕ_y can be formulated as topological expressions, it can be tested in NP that a topological expression respects τ, and by [81] that it is satisfiable.

2.3.3 Matchings from Euler diagrams

Call a matching μ an *ε-matching* if $|f(p) - g(\mu(p))| \leq \varepsilon$ for all $p \in M$. We show that there is some ε-matching μ for any satisfiable topological expression respecting some ε-matching τ. For this, let Φ be an Euler diagram satisfying a topological expression respecting ε-matching τ.

Let $M' = D(root(\mathbb{X}')) = D(root(\mathbb{Y}'))$. For two functions f' and $g' : M' \to \mathbb{R}$, let $\Delta(M') = \sup_{p \in M'} |f'(p) - g'(p)|$. We construct f' and g' with $d_F(f', f) = 0$, $d_F(g', g) = 0$, and $\Delta(M') \leq \varepsilon$

Start by assigning $f'(p) = f(x)$ for $p \in \phi_x$ for $x \in V(\mathbb{X}')$. Also assign $g'(p) = g(y)$ for $p \in \phi_y$ for $y \in V(\mathbb{Y}')$. Then $\Delta(\phi_{V(\mathbb{X}')} \cap \phi_{V(\mathbb{Y}')}) \leq \varepsilon$ because ϕ_x and ϕ_y intersect only if $f(x) - g(y) \leq \varepsilon$.

Next, we define g' on $\partial D(i)$ for each $i \in W(\mathbb{X}')$. By construction, $\partial D(i)$ intersects $\phi_{V(\mathbb{Y}')}$ for any $i \in W(\mathbb{X}')$, so g' is already defined for some point on $\partial D(i)$. We can hence assign g' to all points on $\partial D(i)$ by linear (arc length parameterized) interpolation between the points of $\phi_{V(\mathbb{Y}')} \cap \partial D(i)$. Because f' is constant on $\partial D(i)$, we have by interpolation that $\Delta(\partial D(i)) \leq \Delta(\phi_{V(\mathbb{Y}')} \cap \partial D(i)) \leq \varepsilon$. Assign f' analogously to $\partial D(i)$ for each $i \in W(\mathbb{Y}')$.

Let $\mathcal{F}$ be the faces of $M' \setminus \bigcup_{i \in W(\mathbb{X}') \cup W(\mathbb{Y}')} \partial D(i)$. Now, f' remains to be assigned for faces $F \in \mathcal{F}$ with $F \subseteq \phi_e$ for $e \in E(\mathbb{X}')$ (and g' remains to be assigned for faces with $e \in E(\mathbb{Y}')$). Consider such face $F \subseteq \phi_e = \phi_{(x,x')}$, then f' is already assigned on the boundary ∂F. If $\partial D(j) \subseteq \phi_e$ for $j \in W(\mathbb{Y}')$, then by construction, $\partial D(j)$ touches at least one component ($\partial D(e)$ or $\partial D(x')$) of $\partial\phi_e$. Therefore,

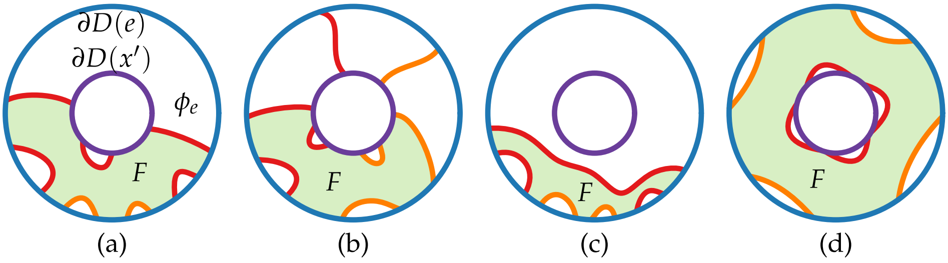


Figure 2.8: Possibilities for face $F \subseteq \phi_{e=(x,x')}$.

each component of ∂F touches at least one component of $\partial\phi_e$. Moreover, because both $\partial D(e)$ and $\partial D(x')$ are connected, at most one component of ∂F touches $\partial D(e)$, and at most one component of ∂F touches $\partial D(x')$. Hence, ∂F has at most two components.

Consider the arcs of $\partial F \setminus \partial\phi_e$ and observe that, because one can draw at most one path between the boundaries of an annulus without disconnecting it, at most two arcs connect $\partial D(e)$ and $\partial D(x')$. Therefore, ∂F has either zero or two arcs on which f' is not constant. Similarly, g' is not constant for zero or two arcs of $\partial F \setminus \partial\phi_{E(\Upsilon')}$, and $F \subseteq \phi_{E(\Upsilon')}$ if such arcs exist.

Based on this, we distinguish the following four cases, illustrated in Figure 2.8.

(a) ∂F touches at most one component of $\partial\phi_e$ (and hence ∂F has one component);

(b) ∂F has one component and touches both components of $\partial\phi_e$, and g is interpolated on both components of $\partial\phi_e$;

(c) ∂F has one component and touches both components of $\partial\phi_e$, and g is interpolated on ∂F on at most one component of $\partial\phi_e$;

(d) ∂F has two components (each touching a distinct component of $\partial\phi_e$).

In case (a), the value of f' is constant along ∂F, and we assign the same value for f' in the interior of F. We will see that the values of g' (which are assigned symmetrically) in the interior of F are values that are also assigned to g' on the boundary of F, so we have $\Delta(F) \leq \Delta(\partial F)$.

Consider the cases (b) and (c), where ∂F has two (interpolated) arcs, and call them α_0 and α_1, both starting on $\partial D(e)$ and ending on $\partial D(x')$. We find two simple non-intersecting paths β_x and $\beta_{x'}$ in F, where β_x connects the start points of α_0 and α_1, and $\beta_{x'}$ connects the end points of α_0 and α_1. We assign $f(x)$ to f' on β_x, and $f(x')$ to f' on $\beta_{x'}$. The paths β_x and $\beta_{x'}$ divide F into three components: a component C_x with $f'(p) = f(x)$ for $p \in \partial C_x$, a component $C_{x'}$ with $f'(p) = f(x')$ for $p \in \partial C_{x'}$, and finally a component C_e bounded by α_0, α_1, β_x and $\beta_{x'}$.

We assign x for f' on the interior of C_x, x' for f' on the interior of $C_{x'}$, and interpolate C_e using the second argument of an arbitrary parameterization $\pi\colon [0,1] \times [0,1] \to C_e$, with $\pi(0,t) = \alpha_0(t)$, $\pi(1,t) = \alpha_1(t)$, $\pi(\cdot,0) = \beta_x$ and $\pi(\cdot,1) = \beta_{x'}$. See Figure 2.9.

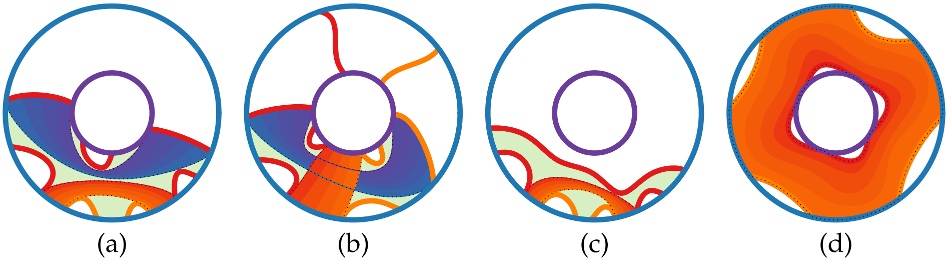


Figure 2.9: Disks C_e (and $C_{e'}$) interpolating f' (and g') in F for the four cases of Figure 2.8.

If g' has a constant value for F, then clearly $\Delta(F) \leq \Delta(\partial F)$, so assume $F \subseteq \phi_{e'}$ for $e' = (y, y') \in E(\mathbb{Y}')$. In case (b), we have $\Delta(F) \leq \Delta(\partial F)$ because each of $(f(x), g(y))$, $(f(x'), g(y))$, $(f(x), g(y'))$, and $(f(x'), g(y'))$ appear as values for $(f'(p), g'(p))$ for points p on ∂F, so interpolation will not exceed the maximum difference. To ensure $\Delta(F) \leq \Delta(\partial F)$ in case (b), we must be careful in choosing C_e which interpolates f' and at the same time $C_{e'}$ which interpolates g' in F. Since in case (b), interpolation of f' occurs on a single component of $\partial\phi_{e'}$, and interpolation of g' occurs on a single component of g', we can choose β_0 and β_1 in such a way that C_e does not intersect $C_{e'}$, such that the interpolation of f' occurs in a region where g' is constant and vice-versa. Doing so, differences between f' and g' never exceed their difference at the boundary, so $\Delta(F) \leq \Delta(\partial F)$.

Lastly consider case (d), where F is an annulus, and on each boundary component, f' and g' are constant. By interpolating f' and g' between their value on each boundary component using the second argument of a parameterization $\pi \colon S^1 \times [0, 1] \to F$, we ensure $\Delta(F) \leq \Delta(\partial F)$.

Theorems 2.3 and 2.4 follow.

Theorem 2.3. *Deciding whether the Fréchet distance between two $\mathbb{R}$-valued disks with constant values at their boundaries is at most ε is in NP.*

Proof. Observe that for $x \in V(\mathbb{X}')$, ϕ_x has the constant value $f(x)$ for f'. Moreover, for each $e = (x, x') \in E(\mathbb{Y}')$, ϕ_e interpolates f' between $f(x)$ and $f(x')$, so $d_F(f, f') = 0$. Similarly, $d_F(g', g) = 0$. Hence, if a topological expression on $W(\mathbb{X}') \cup W(\mathbb{Y}')$ respecting ε-matching τ is satisfiable, then $d_F(f, g) \leq \varepsilon$. Moreover, because such a topological expression exists for any ε-matching μ, deciding whether $d_F(f, g) < \varepsilon$ is in NP. □

Theorem 2.4. *Deciding whether the Fréchet distance between two $\mathbb{R}$-valued spheres is at most ε is in NP.*

Proof. Consider an ε-matching between the two spheres, and puncture one sphere at an arbitrary point, and puncture the other sphere at the matched point. The resulting surfaces are disks with constant values at their boundaries, and by Theorem 2.3 we can test in NP whether an ε-matching for the disks exists, in which case there is also an ε-matching between the spheres (because the boundaries of the disks must be aligned). □

2.3.4 NP-hardness

The contour tree distance is closely related to the Fréchet distance between surfaces because every $\mathbb{R}$-valued tree is the contour tree of some $\mathbb{R}$-valued surface. To construct such a surface from a contour tree, replace each edge of the tree by a cylinder and each degree d vertex by a sphere with d holes (one for each edge). The resulting surface is a topological sphere, representable as a simplicial 2-complex with linearly many triangles. Consider f and $g\colon S^2 \to \mathbb{R}$ to be $\mathbb{R}$-valued spheres with contour trees $\mathbb{X}$ and $\mathbb{Y}$.

Lemma 2.6. *Any many-to-one matching $\tau\colon \mathbb{X} \to \mathbb{Y}$ between contour trees of $\mathbb{R}$-valued spheres f and g can be lifted into an orientation preserving matching $\mu\colon S^2 \to S^2$, satisfying $\mathcal{R}_g \circ \mu = \tau \circ \mathcal{R}_f$.*

Proof. Assume τ is a many-to-one matching and denote by $C_f(x) = \mathcal{R}_f^{-1}(x)$ the contours in the preimage of $x \in \mathbb{X}$ under the quotient map $\mathcal{R}_f$ of the contour tree. For any degree d point $y \in \mathbb{Y}$, consider its preimage $X = \tau^{-1}(y)$. There are d components in $\mathbb{X} \setminus X$. We define μ to match $C_f(T)$ and $C_g(y)$ (both homeomorphic to d-holed spheres) by matching the boundary components as given by τ, so the interior of $C_f(T)$ is automatically matched to the interior of $C_g(y)$. Because τ is continuous, and $\tau^{-1}(y)$ and $\tau^{-1}(y')$ have an empty intersection for distinct y and $y' \in \mathbb{Y}$, this defines an orientation preserving homeomorphism μ between surfaces f and g. □

Corollary 2.2. *If τ is a many-to-one ε-matching between the contour trees $\mathbb{X}$ and $\mathbb{Y}$ of $\mathbb{R}$-valued spheres, then $d_F(f,g) \leq \varepsilon$.*

To prove NP-hardness for surfaces, we reuse the construction of the contour tree distance. Take $\mathbb{X}$ and $\mathbb{Y}$ to be the trees with functions f and g as constructed in Section 2.2.3, and construct $\mathbb{R}$-valued spheres with these trees as contour tree. To illustrate, f and g might be the contour trees of the surfaces depicted in Figure 2.10.

Theorem 2.5. *Computing a factor $(2-\varepsilon)$-approximation of the Fréchet distance between $\mathbb{R}$-valued spheres is NP-hard.*

Proof. By Lemma 2.5 and Corollary 2.1, $d_F(f,g) \geq d_C(f,g) \geq 2$ if $\mathcal{S}$ admits no solution to X3C. By Lemma 2.4 and Corollary 2.2, $d_F(f,g) \leq d_C(f,g) \leq 1$ if $\mathcal{S}$ admits a solution to X3C. So approximating the $d_F(f,g)$ for $\mathbb{R}$-valued spheres within factor 2 is NP-hard. □

Because the matching constructed in Lemma 2.4 always matched the two connection vertices, we can puncture the constructed surfaces at these vertices to obtain Corollary 2.3.

Corollary 2.3. *Computing a factor $(2-\varepsilon)$-approximation of the Fréchet distance between $\mathbb{R}$-valued disks is NP-hard.*

2.4 An unrepresentable matching

Consider the two rooted trees $\mathbb{X}$ and $\mathbb{Y}$ in Figure 2.11. The leaves of $\mathbb{X}$ are labeled $x_{i,j}$ ($i \in \{1,\dots,6\}$, $j \in \{1,2,3\}$) and the leaves of $\mathbb{Y}$ are labeled $y_{k,l}$ ($k \in \{1,\dots,9\}$, $l \in \{1,2\}$). For both trees, leaves with the same i or k are grouped in subtrees. Based on the complete bipartite graph $K_{3,3}$ with vertices $v_1,\dots,v_6$ and edges $e_1,\dots,e_9$, we construct a matching τ between those subtrees as follows. For an edge $e_k = (v_i, v_{i'})$ of $K_{3,3}$, match the path from $y_{k,1}$ to $y_{k,2}$ with the path between unused vertices $x_{i,j}$ and $x_{i',j'}$. Match the edge from the root to group i of $\mathbb{X}$ with the edges of $\mathbb{Y}$ from the root to the three groups that match with $x_{i,1}$, $x_{i,2}$ and $x_{i,3}$. Edges in the top layer of the tree are matched in a many-to-many fashion, whereas edges of the bottom layer are matched using linear interpolation. Then $\tau \in \mathcal{M}(\mathbb{X},\mathbb{Y})$ does not match any path from $y_{k,1}$ to $y_{k,2}$ (of edge $e_k = (v_i, v_{i'})$) with any group of $\mathbb{X}$ not containing any $x_{i,j}$ or $x_{i',j'}$. However, because $K_{3,3}$ is not a planar graph, this matching cannot be realized on the sphere, as illustrated in Figure 2.11.

The Fréchet distance between $\mathbb{R}$-valued spheres hence seems more discriminative than the contour tree distance. But this example merely shows that we cannot lift any ε-matching between contour trees to one between spheres; a different matching on spheres (see Figure 2.11) can still be an ε-matching. We have not found any instances (on spheres) for which the Fréchet distance is strictly greater than the contour tree distance.

2.5 Related measures

While the Fréchet distance is applicable to a broad class of functions, the functional distortion distance and interleaving distance specialize on Reeb graphs,

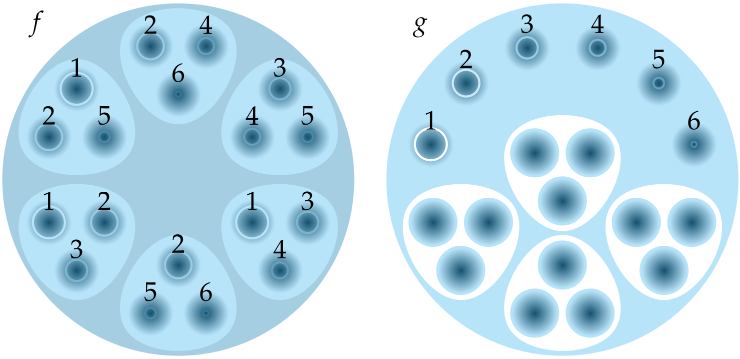


Figure 2.10: For input $\mathcal{S} = \{\{2,4,6\},\{3,4,5\},\{1,3,4\},\{2,5,6\},\{1,2,3\},\{1,2,5\}\}$, the disks resulting from the reduction from X3C with color-coded function values.

and can be used to compare $\mathbb{R}$-valued functions. They are closely related to the Gromov-Hausdorff distance. The functional distortion distance $d_D(f,g)$ between Reeb graphs $f\colon \mathbb{X} \to \mathbb{R}$ and $g\colon \mathbb{Y} \to \mathbb{R}$ is defined as

$$d_D(f,g) = \inf_{\Phi,\Psi} \max\{D(\Phi,\Psi), \|f - g \circ \Phi\|_\infty, \|g - f \circ \Psi\|_\infty\},$$

$$D(\Phi,\Psi) = \sup_{\substack{(x,y),(x',y') \\ \in C(\Phi,\Psi)}} \frac{1}{2} |\mathrm{height}_f(x,x') - \mathrm{height}_g(y,y')|,$$

$$C(\Phi,\Psi) = \{(x,y) \in \mathbb{X} \times \mathbb{Y} \mid \Phi(x) = y \text{ or } x = \Psi(y)\},$$

where $\Phi\colon \mathbb{X} \to \mathbb{Y}$ and $\Psi\colon \mathbb{Y} \to \mathbb{X}$ are continuous maps;

$$\mathrm{height}_f(x,x') = \inf_{\pi\colon x \rightsquigarrow_{\mathbb{X}} x'} \max_{x'' \in \pi} f(x'') - \min_{x'' \in \pi} f(x'')$$

is the length of the image (which is an interval) under f of the path in $\mathbb{X}$ from x to x' that minimizes this length. In the case of trees, this will be that of the unique simple path connecting x and x'.

Another measure is the weak Fréchet distance d_{wF}, which requires matchings to be surjective but—in contrast to the Fréchet distance—does not require the matching to be a homeomorphism. As such, it can compare trees that are not homeomorphic.

We highlight how the functional distortion distance, the weak Fréchet distance and the contour tree distance differ using small instances in Figure 2.12. Perhaps surprising is that the functional distortion distance between $\mathbb{X}'$ and $\mathbb{Y}'$ is 1 instead of 2. This is because the term $D(\Phi,\Psi)$ is only half the difference in height between pairs of matched points, so by ensuring the difference in function value of matched points is at most half the difference in height between any two points, we achieve a distance of 1. It therefore seems that the contour tree distance is more discriminative than the functional distortion distance, and both are much more discriminative than the weak Fréchet distance.

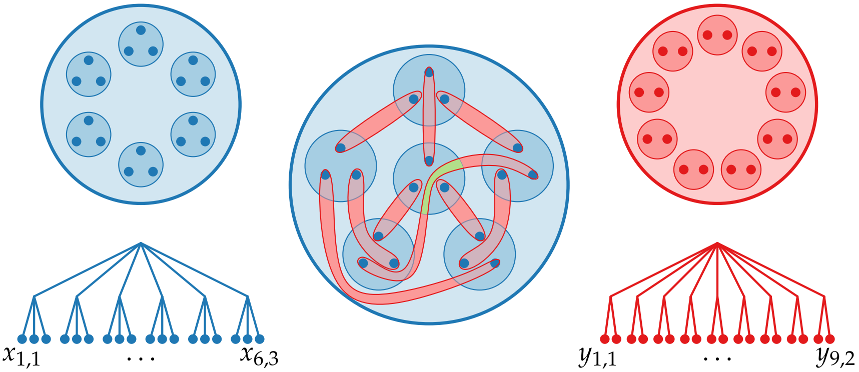


Figure 2.11: Trees $\mathbb{X}$ (left) and $\mathbb{Y}$ (right) with their corresponding surfaces. Center: a matching in which a subtree of $\mathbb{Y}$ intersects an additional subtree of $\mathbb{X}$ (green).

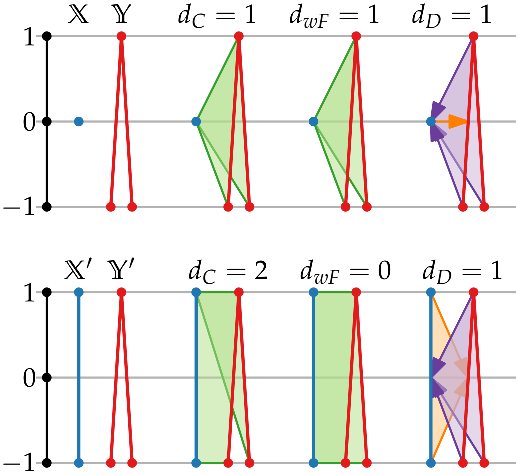


Figure 2.12: The contour tree distance d_C, weak Fréchet distance d_{wF} and functional distortion distance d_D for small trees. Correspondences green, Φ orange, Ψ purple.

2.6 Discussion

We have shown that computing the Fréchet distance between $\mathbb{R}$-valued surfaces is NP-complete if the surfaces are spheres or disks with constant boundary. The question whether the Fréchet distance is in NP for higher genus surfaces remains open. Using our techniques, the main bottleneck for this is that the string graph recognition problem is only known to lie in NEXP on higher genus surfaces [81], and it is therefore unknown whether Euler diagrams on arbitrary surfaces can be recognized in NP. So extensions of the techniques used in this chapter might only show that the Fréchet distance is in NEXP for $\mathbb{R}$-valued surfaces of higher genus. Finally, we are interested in an efficient constant-factor approximation algorithm for the Fréchet distance or the contour tree distance, and we would like to know whether these measures are equivalent on $\mathbb{R}$-valued spheres.

3

Horizontal Fréchet Isotopies and Monotonicity

In this chapter, we study the problem of comparing curves through deformations in which the intermediate curves remain simple. As in the previous chapter, we minimize the distance that a point travels as the deformation moves it between the two input curves.

Specifically, we consider the *isotopic Fréchet distance*, a distance measure between two curves f and g that captures one intuitive notion of an optimal morph between these two curves. The classic Fréchet distance between f and g, also called the 'dog leash distance', measures the length of the shortest possible straight leash needed to connect a man and a dog which are walking forward along f and g. Any two feasible walks using such a shortest leash induce a *Fréchet matching* between f and g. One can now imagine to build a morph between f and g by sliding each point of f along the leash that connects it to its matched point on g. Such an approach will work well in unrestricted Euclidean space, however, it is not suitable for more general spaces that might contain obstacles. In the presence of obstacles the leashes of the classic Fréchet distance can jump discontinuously and hence the resulting morph would be discontinuous as well.

The *homotopic Fréchet distance* [29, 66] forces leashes to move in a continuous way. More formally, for two curves f and $g\colon [0,1] \to \mathbb{R}^2$ in the plane, a homotopy $h\colon [0,1]^2 \to \mathbb{R}^2$ is a continuous map between f and g. Such a homotopy essentially morphs one curve into the other: each point of f traces a path $h(p,\cdot)$ to a point on g. The length of a homotopy is the length of the longest such path, and a *Fréchet homotopy* is one that minimizes this length. The homotopic Fréchet distance between f and g is then the length of a Fréchet homotopy between f and g. The homotopic Fréchet distance and the classic Fréchet distance are equivalent in $\mathbb{R}^2$. The morph that results from a Fréchet homotopy is continuous, but it may change the structure of the input curves during the morph: intermediate curves can self-intersect or collapse to a point, even if f and g are simple curves.

A homotopy is an isotopy if all its intermediate curves $h(\cdot,t)$ are simple. That is, a morph based on an isotopy does not allow the morphing curves to pass through themselves, a feature which is crucial for applications which involve curve-like objects in the real world. The *isotopic Fréchet distance* measures the length of an optimal isotopy between f and g; we call an optimal isotopy a *Fréchet isotopy*. The study of Fréchet isotopies was initiated in [32]. The authors gave some simple observations and examples and showed that the isotopic Fréchet distance in the plane can be arbitrarily larger than the homotopic Fréchet distance.

In this chapter we design the first algorithms to compute the isotopic Fréchet distance in various monotone settings, an important first setting to consider where structural observations and algorithms are nonetheless quite non-trivial.

Related work. Closely related are morphs based on *geodesic width* [53]: the intermediate curves are not allowed to cross the input curves f and g, and they are restricted to the area between the leashes connecting the endpoints of f and g. This restriction naturally enforces intermediate curves without self-intersections since 'geodesic leashes' do not cross each other. Morphs based on geodesic width minimize the maximum leash length. However, they are restricted to input curves that do not intersect each other; in contrast, Fréchet isotopies are also well-defined for input curves that intersect each other.

A variety of morphs have been considered in the graph drawing and computational geometry literature. Here, the intermediate curves are homeomorphic to the input and vary continuously. It is well known that any two drawings of the same planar graph (with the same faces and outer face) can be morphed into one another. More recent work has focused on bounding the number of steps in the optimal morph between any two input graphs [6, 7]. However, in contrast to Fréchet isotopies, the morphs do not minimize length.

Results. In this chapter, we present the first algorithmic results for the isotopic Fréchet distance in any setting. Our main focus is on two simple input curves, where one of the curves is monotone. Even in this restricted setting, it is unclear how to compute such isotopies, since the best homotopy between simple monotone curves is not always an isotopy.

We begin by refuting a conjecture by Chambers et al. [32] in the initial paper introducing the isotopic Fréchet distance (Section 3.2). We present some basic bounds on the lengths of isotopies in Section 3.3. In the same section we also show that in the context of the computation of short isotopies, the standard notion of a simple curve can be relaxed slightly (to allow the preimage of a point to be at most an interval) without affecting the length of an optimal isotopy significantly.

We then discuss algorithms to compute optimal isotopies between pairs of curves in monotone settings. First of all, in Section 3.4, we describe how to compute optimal isotopies if there is a direction in which both input curves are monotone. In this case, we show that the length of such an optimal isotopy is the same as that of an optimal homotopy.

In Section 3.5, we construct an isotopy that morphs a given curve into an x-monotone one using minimal horizontal movement. Surprisingly, this horizontal movement does not exceed the horizontal movement required by an optimal homotopy. We tie the cost of this isotopy to the notion of persistent homology, and moreover relate the stages of horizontal movement of our isotopy to this notion. For the related problem of the homotopic Fréchet distance, lower bounds based on persistent homology were recently developed [83].

Finally, in Section 3.6, we consider a pair of curves f and g parameterized by a variable p, and we capture the conditions under which an isotopy exists that (for all p) sends $f(p)$ to $g(p)$ along a curve that is monotone in either the positive or negative x-direction. Surprisingly the conditions under which such an isotopy exists are easy to test: since the source and destination of each point are known, one can check in advance whether a point moves in the positive or negative x-direction. From a global perspective, the isotopy we construct starts by moving certain extrema. For each extremum of f, there is a direction in which it 'shrinks', and if the points on this extremum move in that direction, we can shrink and possibly remove the extremum. Applying this procedure exhaustively yields a curve f'. The procedure forms an isotopy from f to f' if an isotopy from f to g exists. We apply a symmetric procedure to g to obtain a curve g'. Finally we show that if an isotopy between f and g exists, then it is easy to find an isotopy from f' and g' such that the result is an isotopy from f to f' to g' to g.

3.1 Preliminaries

A *curve* in the plane is a continuous map $f\colon [0,1] \to \mathbb{R}^2$. We denote the x and y-coordinates of $f(p)$ by $f(p).x$ and $f(p).y$, respectively. A continuous nondecreasing surjection $\alpha\colon [0,1] \to [0,1]$ is called a *reparameterization* of a curve. A *homotopy* is a continuous map $h\colon [0,1] \times [0,1] \to \mathbb{R}^2$. We denote its level curves by $h_t\colon p \mapsto h(p,t)$, and say h goes from curve f to g if $h_0 = f$ and $h_1 = g$. We say a curve is *simple* if the preimage of each point in $\mathbb{R}^2$ is at most one interval.[1] A homotopy is an *isotopy* if every curve h_t is simple.

A homotopy from f to g traces paths $\lambda_{h,p}\colon t \mapsto h(p,t)$ between the points $f(p)$ and $g(p)$, and such a path is traditionally referred to as a *leash*. Let the *length* of a homotopy h be the length $\mathit{length}(h) = \sup_p \mathit{length}(\lambda_{h,p})$ of its longest leash. We are interested in homotopies h minimizing this length and define the *homotopic Fréchet distance* between f and g as

$$d_{hom}(f,g) = \inf_{\substack{\alpha,\beta,h \\ h_0 = f\circ\alpha,\ h_1 = g\circ\beta}} \mathit{length}(h),$$

where h ranges over homotopies and α and β range over reparameterizations. The *isotopic Fréchet distance* d_{iso} is defined similarly, except h ranges over isotopies.

A map $m'\colon X \to \mathbb{R}^2$ is an *ε-perturbation* of $m\colon X \to \mathbb{R}^2$ if $\|m'(x) - m(x)\| \leq \varepsilon$ for all $x \in X$. We call a curve f *weakly simple* if for all $\varepsilon > 0$, there exists an ε-perturbation that is simple. Similarly, we call a homotopy h a *weak isotopy* if some ε-perturbation is an isotopy.

The *Fréchet distance* $d_F(f,g) = \inf_{\alpha,\beta} \sup_p \|f \circ \alpha(p) - g \circ \beta(p)\|$ is a related measure that does not require leashes to trace out a homotopy, so each leash can be assumed to be a shortest path. The pair (α,β) is called a *matching*. We define the *cost* of a matching (α,β) between f and g as $\mathit{cost}_{f,g}(\alpha,\beta) = \sup_p \|f \circ \alpha(p) -$

---

[1] The usual definition requires simple curves to be injective. However, simple curves are closed under reparameterizations by our definition, and we show that this distinction has little impact on our results. Intuitively, simple curves allow pausing at a point, but not returning to a previously visited point.

$g \circ \beta(p)\|$. A *Fréchet matching* between curves f and g in the plane is a matching with cost $d_F(f,g)$.

The map $Aff^{f,g}(p,t) = (1-t) \cdot f(p) + t \cdot g(p)$ using line segments (shortest paths) as leashes is a homotopy since it is an affine interpolation between continuous maps. We call $Aff^{f,g}$ the *affine homotopy* from f to g, and its length is $length(Aff^{f,g}) = \sup_p \|f(p) - g(p)\|$. It follows that the homotopic Fréchet distance and the Fréchet distance are equivalent in $\mathbb{R}^2$. On the other hand, the isotopic Fréchet distance in the plane can be arbitrarily larger than the homotopic Fréchet distance [32].

We call a homotopy h from $f \circ \alpha$ to $g \circ \beta$ a *Fréchet homotopy* if $length(h) = d_{hom}(f,g)$, and call h a *Fréchet isotopy* if h is an isotopy with $length(h) = d_{iso}(f,g)$. Since every isotopy is a homotopy, we have $d_{hom}(f,g) \leq d_{iso}(f,g)$ and any isotopy that is a Fréchet homotopy is also a Fréchet isotopy. However, Fréchet isotopies do not need to be Fréchet homotopies since there might exist a homotopy shorter than any isotopy.

In this chapter, we assume that all curves f and g are piecewise differentiable. For a curve f and a unit vector $(x,y) \in S^1$, we define the *directional length* of f in the direction (x,y) to be the total length that f moves *forward* in the direction of the vector, given by $length_{(x,y)}(f) = \int_0^1 \max(0, \langle \frac{\mathrm{d}f(p)}{\mathrm{d}p}, (x,y) \rangle)\mathrm{d}p$, where $\langle \cdot, \cdot \rangle$ is the inner product. We define the horizontal length of a curve as $length_{\mathrm{hor}}(f) = length_{(-1,0)}(f) + length_{(1,0)}(f)$ and define the *horizontal* homotopic and isotopic Fréchet distances using the horizontal length function. As usual, a horizontal Fréchet homotopy (respectively isotopy) is one minimizing the horizontal homotopic (respectively isotopic) Fréchet distance.

3.2 Disproving a conjecture

In Figure 3.1 we show an example of two zig-zag curves, originally presented in [32]. The Fréchet distance between these curves is at most ε, as the walks along each curve can have the same x-coordinate, keeping the leash between them at most that length. However, this Fréchet mapping yields a homotopy that collapses the zig-zag to a flat line before re-expanding to the other zig-zag, which does not result in an isotopy, as (for example) the two leftmost blue endpoints would map to a common point in the middle of the isotopy and then break into two red leftmost points again.

In [32] the authors conjectured that the isotopic Fréchet distance between the zig-zags is $\sqrt{L^2 + \varepsilon^2}$. However, the isotopy depicted by the green leashes on the right side of Figure 3.1 has length at most $\sqrt{L^2 + \varepsilon^2}/2 + \varepsilon/2$. Intermediately, the isotopy flattens the zig-zag to the center of a line (shown in the middle in purple) during the morph to the red zig-zag, but the parameterization is chosen in such a way that points move at most a distance of $\varepsilon/2$ either before or after reaching this line and at most $\sqrt{L^2 + \varepsilon^2}/2 + \varepsilon/2$ in total.

A lower bound. We will show that the isotopy of Figure 3.1 is arbitrarily close to

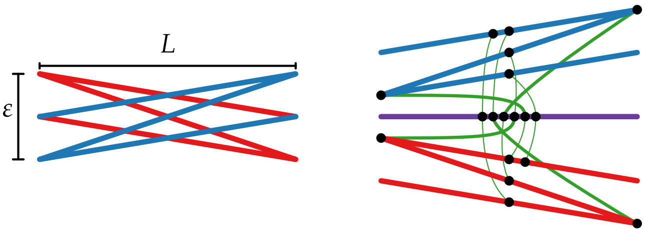


Figure 3.1: An isotopy between two 'opposite' zig-zags of length roughly $\sqrt{L^2+\varepsilon^2}/2+\varepsilon/2$. The fat arcs have horizontal length roughly $L/2$, whereas the others have negligible horizontal length.

optimal. Consider a convex region D and an isotopy h between curves f and g in the plane, where the endpoints of all intermediate curves lie in D; so $\operatorname{Im}(\lambda_{h,0}) \subseteq D$ and $\operatorname{Im}(\lambda_{h,1}) \subseteq D$. Fix some $p \in (0,1)$ and denote by $poly_t$ the polyline with an edge from $h(0,t)$ to $h(p,t)$, and an edge from $h(p,t)$ to $h(1,t)$. Let θ_t be the (counterclockwise) angle at $h_t(p)$ between the two edges of $poly_t$ (plus a multiple of 360 degrees), such that θ_t varies continuously with t. We show in Lemma 3.1 that (in any isotopy from $poly_0$ to $poly_1$) the leash $\lambda_{h,p}$ must intersect D if θ_0 and θ_1 differ by at least 180 degrees, see Figure 3.2.

Lemma 3.1. *If f is isotopic to $poly_0$ relative to its vertices*[2] *and g is isotopic to $poly_1$ relative to its vertices, and $|\theta_1-\theta_0| \geq 180$, then $h(p,t) \in D$ for some t.*

Proof. Because f and g are isotopic to $poly_0$ and $poly_1$ respectively (relative to their vertices), we may assume without loss of generality that $0 \leq \theta_0 < 360$ and $0 \leq \theta_1 < 360$. Because θ_t varies continuously, we have by the intermediate value theorem that $\theta_t = 180$ for some $t \in [0,1]$. In that case, $h(p,t)$ lies on the line segment between $h(0,t) \in D$ and $h(1,t) \in D$. By convexity, this line segment lies completely in D, so $h(p,t) \in D$. □

---

[2] That is, there exists an isotopy from f to $poly_0$ that does not move $f(0)$, $f(p)$ or $f(1)$.

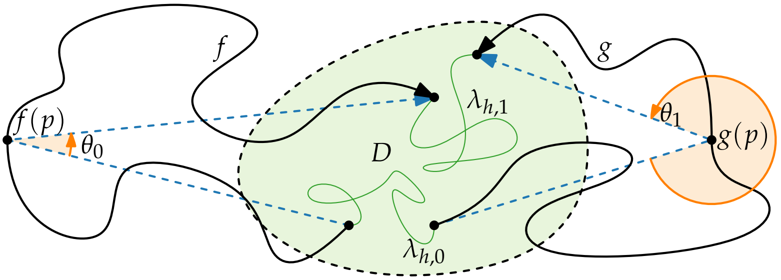


Figure 3.2: Curves $f = h_0$, $g = h_1$ and polylines $poly_0$ and $poly_1$ with endpoints in convex region D.

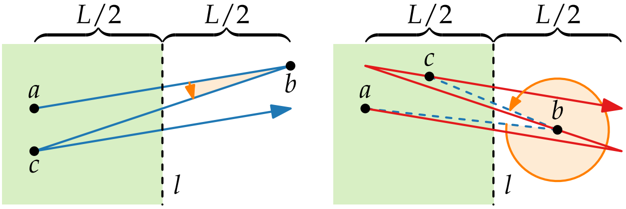


Figure 3.3: The vertices and region D used to obtain a lower bound for the curves of Figure 3.1.

Using Lemma 3.1, we can show that our isotopy of length $\sqrt{L^2+\varepsilon^2}/2+\varepsilon/2$ for the zig-zags of Figure 3.1 is optimal as ε approaches 0. For this, we show that any Fréchet isotopy has length at least $L/2$. Assume that the zig-zags f and g are parameterized such that there is an isotopy h of length less than $L/2$ between them. Let l be the vertical line in the middle between the vertices such that each vertex has distance $L/2$ to l, and let D be the half-plane to the left of l, see Figure 3.3. Let $f(a)$, $f(b)$ and $f(c)$ be the first three vertices of f. Because h has length less than $L/2$, the leashes $\lambda_{h,a}$ and $\lambda_{h,c}$ lie completely in D, and $\lambda_{h,b}$ lies completely outside D. Since $a < b < c$ and $g(b) \notin D$, we have that a and c lie in different components of $g^{-1}(D)$. The isotopy h induces a restricted isotopy between the subcurves of f and g from a to c, and these subcurves satisfy the conditions required by Lemma 3.1, meaning that $\operatorname{Im}(\lambda_{h,b})$ intersects D, contradicting that h has length less than $L/2$.

3.3 Simple curves

In this section, we consider our notion of simple curves, and connect it to the standard notion of simple. We prove that in the context of Fréchet isotopies, these concepts are roughly equivalent. While a curve f is generally called simple if it is an injective map, we will call such injective maps *strictly* simple instead. We call an isotopy through strictly simple curves a strict isotopy. Recall that in our modified definition of simple, a curve f is simple if the preimage of each point consists of at most one interval. A curve f is *constant* at $q \in \mathbb{R}^2$ if $f(p) = q$ for all $p \in [0,1]$. Any constant curve is simple but not strictly simple. For a non-constant simple curve f, let $\hat{f}$ be the unique arc-length parameterized curve, and $\hat{\alpha}_f$ be the unique reparameterization, such that $f = \hat{f} \circ \hat{\alpha}_f$. We prove Theorem 3.1 to justify our alternative definition of simple, since strict isotopies are roughly equivalent to isotopies in some sense.

Observation 3.1. *If f is a strictly simple curve and g is the constant curve at q, then $\mathit{Aff}_t^{f,g}$ is strictly simple for $t < 1$. To see this, observe that the curve $\mathit{Aff}_t^{f,g}$ is a copy of f scaled down towards q by a scaling factor of $(1-t)$. So $\mathit{Aff}_t^{f,g}$ is injective unless $t = 1$.*

Observation 3.2. *If f is simple and g is constant, then $\mathit{Aff}_t^{f,g}$ is simple for all t.*

Lemma 3.2. $d_{iso}(f,g) \leq \inf_{q\in\mathbb{R}^2} \sup_p \|f(p) - q\| + \sup_p \|g(p) - q\|$.

Proof. Let γ be the constant curve at q, and observe that $d_{iso}(f,\gamma) \leq \sup_p \|f(p) - q\|$ by the isotopy $\mathit{Aff}^{f,\gamma}$; similarly, $d_{iso}(\gamma,g) \leq \sup_p \|g(p) - q\|$. By the triangle inequality, we have $d_{iso}(f,g) \leq \sup_p \|f(p) - q\| + \sup_p \|g(p) - q\|$. $\square$

Lemma 3.3. *If there is an isotopy of finite length between curves f and g, then there is also one of length at most* $\inf_{q\in\mathbb{R}^2} \sup_p \|f(p) - q\| + \sup_p \|g(p) - q\| + \varepsilon$ *for any* $\varepsilon > 0$.

Proof. Take an isotopy of finite length δ between f and g. We are done if $\varepsilon \leq \delta$, so assume $\varepsilon > \delta$ and let γ be the constant curve at q. The idea is to shrink f and g to a sufficiently small neighborhood of q and apply a scaled version of the isotopy of length δ there. By Observation 3.1, $\mathit{Aff}_t^{f,\gamma}$ and $\mathit{Aff}_t^{g,\gamma}$ are strictly simple for $t < 1$. Consider the strictly simple curves $f' = \mathit{Aff}_t^{f,\gamma}$ and $g' = \mathit{Aff}_t^{\gamma,g}$ for $t = 1 - \varepsilon/\delta < 1$. The scaled version s' of s between f' and g' has length ε, so the composition of $\mathit{Aff}^{f,f'}$, s', and $\mathit{Aff}^{g',g}$ is a strict isotopy of length at most $\inf_{q\in\mathbb{R}^2} \sup_p \|f(p) - q\| + \sup_p \|g(p) - q\| + \varepsilon$. $\square$

Theorem 3.1. *For any isotopy h between strictly simple curves f and g, and* $\varepsilon > 0$*, there is a strict isotopy s of length at most* $\mathit{length}(h) + \varepsilon$ *from f to g.*

Proof. If $\mathit{length}(h) \geq \sup_p \|f(p) - q\| + \sup_p \|g(p) - q\|$ for some $q \in \mathbb{R}^2$, then we are done by Lemma 3.3. So we may assume that h_t is never a constant curve. We show that we can continuously perturb all h_t into strictly simple curves without significantly lengthening the isotopy. For a curve f and a value $\varphi \in [0,1]$, let $f^{\varphi}(p) = \hat{f}((1-\varphi)\hat{\alpha}(p) + \varphi p)$. Then f^{φ} is essentially an arc-length based interpolation between f (for $\varphi = 0$) and $\hat{f}$ (for $\varphi = 1$). Moreover, $\mathrm{Im}(f) = \mathrm{Im}(f^{\varphi})$, and f^{φ} is strictly simple whenever f is simple and $\varphi > 0$.

Let $\delta = \int_0^1 |\frac{d\mathit{length}(h_t)}{dt}|dt$ be the cumulative change in length of h_t throughout the homotopy. Define $\varphi_t = (\frac{1}{2} - |t - \frac{1}{2}|)\frac{\varepsilon}{2\delta}$, such that $\varphi_t = 0$ if and only if $t = 0$ or $t = 1$, and $\int_0^1 |\frac{d\varphi_t}{dt}|dt \leq \frac{\varepsilon}{2\delta}$. The map s defined by $s_t = h_t^{\varphi_t}$ is a homotopy because φ_t (and therefore $h_t^{\varphi_t}$) changes continuously. Moreover, s is a strict isotopy because $h_t^{\varphi_t}$ is strictly simple if $\varphi_t > 0$, which is the case for $0 < t < 1$, and the curves $f = s_0$ and $g = s_1$ are also strictly simple. It remains to show that s has length at most $\mathit{length}(h) + \varepsilon$. For this, we compare the length of $\lambda_{h,t}$ with that of $\lambda_{s,t}$. By construction of φ_t, we have that $\int_0^1 \|s(p,t) - h(p,t)\|dt \leq \delta\frac{\varepsilon}{2\delta} = \frac{\varepsilon}{2}$. Moreover, $\|s(p,t) - s(p,t')\| \leq \|h(p,t) - h(p,t')\| + \|s(p,t) - h(p,t)\| + \|s(p,t') - h(p,t')\|$ by the triangle inequality. Thus, $\mathit{length}(\lambda_{s,p}) \leq \mathit{length}(\lambda_{h,p}) + 2\int_0^1 \|s(p,t) - h(p,t)\|dt \leq \mathit{length}(\lambda_{h,p}) + \varepsilon$. Hence, $\mathit{length}(s) \leq \mathit{length}(h) + \varepsilon$. $\square$

3.4 Isotopies between monotone curves

A curve f is *x-monotone* if $f(p).x \leq f(p').x$ and strictly x-monotone if $f(p).x < f(p').x$ for all $p < p'$. We claim that for strictly x-monotone curves f and g, $h = Aff^{f,g}$ is an isotopy. For this, each intermediate curve h_t must be simple. In fact, we show in Lemma 3.4 that each h_t is strictly x-monotone. Because any strictly x-monotone curve is simple, Corollary 3.1 follows. Observe also that any x-monotone curve f is weakly simple, since it can be ε-perturbed into a simple x-monotone curve $f'(p) = (f(p).x, f(p).y + p\varepsilon)$, and we obtain Theorem 3.2.

Lemma 3.4. *For strictly x-monotone curves f and g, each curve h_t of $h = Aff^{f,g}$ is strictly x-monotone.*

Proof. Recall that $h_t(p) = (1-t)\cdot f(p) + t \cdot g(p)$. Consider the x-coordinates $x_t(p)$ of $h_t(p)$ and $x_t(p')$ of $h_t(p')$ for $p < p'$. Let $s_t = x_t(p') - x_t(p)$. Because f and g are strictly x-monotone, we have $x_0(p) < x_0(p')$ and $x_1(p) < x_1(p')$, so $s_0 > 0$ and $s_1 > 0$. Since s is affine, we have $s(t) > 0$ for $t \in [0,1]$, so $x_t(p) < x_t(p')$. Hence, h_t is strictly x-monotone. □

Corollary 3.1. *If f and g are strictly x-monotone curves, $Aff^{f,g}$ is an isotopy from f to g.*

Theorem 3.2. *In the limit, the isotopic Fréchet distance between simple x-monotone curves f and g in the plane is the same as their Fréchet distance.*

Proof. Pick any $\varepsilon > 0$, and perturb f and g into strictly x-monotone curves f' and g' with $d_{iso}(f,f') \leq \varepsilon$ and $d_{iso}(g,g') \leq \varepsilon$. We have $d_{iso}(f,g) \leq d_{iso}(f',g') + 2\varepsilon$ by the triangle inequality. Consider a matching (α,β) between f' and g' of cost δ and perturb it to obtain a matching (α',β') with $cost_{f',g'}(\alpha',\beta') \leq \delta + \varepsilon$ and the property that the curves $f' \circ \alpha'$ and $g' \circ \beta'$ are strictly x-monotone. By Corollary 3.1, $Aff^{f'\circ\alpha', g'\circ\beta'}$ is an isotopy of length at most $\delta + \varepsilon$, so $d_{iso}(f',g') \leq \delta + \varepsilon$. Taking a Fréchet matching (α,β) between f' and g', we get $\delta = d_F(f',g')$. So $d_{iso}(f,g) \leq cost_{f',g'}(\alpha',\beta') + 2\varepsilon \leq d_F(f',g') + 3\varepsilon \leq d_F(f,g) + 5\varepsilon$. Moreover, $d_F(f,g) \leq d_{iso}(f,g)$, so $d_{iso}(f,g)$ converges to $d_F(f,g)$ as ε approaches 0. □

The homotopy constructed in Theorem 3.2 may degrade into a weak isotopy if we were to take $\varepsilon = 0$. By definition, the length of Fréchet isotopies converges to that of their weak counterparts in the limit. Therefore, we choose in this chapter to work with weak isotopies instead, as they are often more conveniently constructed.

Corollary 3.2. *For any matching (α,β) between x-monotone curves f and g in the plane, the homotopy $Aff^{f\circ\alpha, g\circ\beta}$ is a weak isotopy of length $cost_{f,g}(\alpha,\beta)$.*

3.5 Isotopies to monotone curves

We say that f is *η-narrow* if it is contained in a rectangle of height η; so the interval $\mathrm{Im}(f.y)$ has length at most η. In this section, we give an algorithm to compute an isotopy to an x-monotone curve, and we can show that the length of this isotopy is within $\eta/2+\varepsilon$ of an optimal isotopy (for any $\varepsilon > 0$). In Section 3.5.1 we argue that this problem is interesting already even when η is negligibly small.

3.5.1 Non-monotone isotopies

Consider the curve $f = h_0$ of Figure 3.4 with $f(0) = p_0$ and $f(1) = p_5$, morphing into an x-monotone curve $g = h_1$ as depicted. No matter how small we pick $\eta > 0$, the depicted isotopy has length at most $r+\eta$ if $w < r$. Conversely by Lemma 3.6, there exists no homotopy to an x-monotone curve of length less than r, even if we pick a different x-monotone curve g. In this context, the lemma states that because $f(p_1).x \geq f(p_2).x + 2r$ and $g(p_1).x \leq g(p_2).x$, one of the leashes λ_{h,p_1} or λ_{h,p_2} has length at least r in any homotopy h.

What makes this instance interesting is that any optimal isotopy moves some points both forward (in the positive x-direction) and backward (in the negative x-direction) for a considerable distance. Formally, for any isotopy h from f to g, we have both $\mathit{length}_{(1,0)}(\lambda_{h,p}) \geq w/2-\eta$ and $\mathit{length}_{(-1,0)}(\lambda_{h,p}) \geq w/2-\eta$ for some p. In particular, consider the two endpoints $f(p_0)$ and $f(p_5)$ in Figure 3.4. Informally, these points must 'untangle' with respect to each other somewhere in any isotopy h. For this, the x-coordinates of $h_t(p_0)$ and $h_t(p_5)$ must be equal for some value of t, say for $t = t^*$. Because g is x-monotone, and an optimal isotopy has length at most $r+\eta$, we have $g(p_0).x \leq g(p_2).x \leq f(p_2).x + r + \eta = f(p_0).x + \eta$ and symmetrically $f(p_5).x - \eta = f(p_3).x - r - \eta \leq g(p_3).x \leq g(p_5).x$.

Let $\gamma = h_{t^*}$ and $x^* = \gamma(p_0).x = \gamma(p_5).x$. Since $f(p_5).x - f(p_0).x = w$, we have $x^* - f(p_0).x \geq w/2$ or $f(p_5).x - x^* \geq w/2$. If $x^* - f(p_0).x \geq w/2$, we also have $g(p_0).x - x^* \geq w/2 - \eta$, so p_0 moves forward for a distance of at least $w/2$ and backwards for a distance of at least $w/2-\eta$. Otherwise, $f(p_5).x - x^* \geq w/2$, and p_5 moves backwards for $w/2$ and forwards for $w/2-\eta$. The total distance such points move approaches $d_{iso}(f,g)$ as η approaches 0.

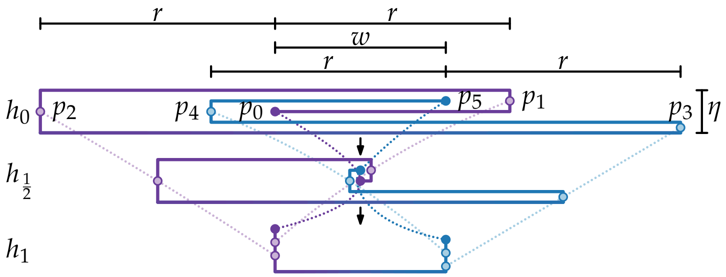


Figure 3.4: A curve for which any Fréchet isotopy to an x-monotone curve requires some point to move for a distance of at least $w/2-\eta$ in both the positive and the negative x-direction.

3.5.2 Horizontal length

Let f be an η-narrow curve. We are interested in how much an isotopy must move a point back and forth in the x-direction to obtain an x-monotone curve. Using Lemma 3.5, one can see that the isotopic Fréchet distance is at most $\eta/2+\varepsilon$ larger than the horizontal isotopic Fréchet distance. So as η approaches 0, the horizontal length approaches the isotopic Fréchet distance.

Lemma 3.5. *If f is an η-narrow simple curve, and $\eta > \varepsilon > 0$, then there is an isotopy of length $(\eta-\varepsilon)/2$ from f to an ε-narrow curve.*

Proof. Suppose $\mathrm{Im}(f_y) \subseteq [y-\eta/2, y+\eta/2]$, and let $f'(p) = (f(p).x, y)$. In the affine homotopy $h = \mathit{Aff}^{f,f'}$, the curves h_t are copies of f, scaled down vertically towards the horizontal line $\mathbb{R} \times \{y\}$ with scaling factor $1-t$. So for $0 \leq t < 1$, curve h_t is simple, and each h_t is $((1-t)\eta)$-narrow. In particular, h_t is ε-narrow if $t = 1-\frac{\varepsilon}{\eta} < 1$, so the subhomotopy $h|_{[0,1-\frac{\varepsilon}{\eta}]}$ is an isotopy to an ε-narrow curve, and this isotopy has length at most $(\eta-\varepsilon)/2$ because each point of f moves monotonically along a vertical line of length at most $(\eta-\varepsilon)/2$. □

For the remainder of the section, we will only optimize the horizontal length of isotopies. We morph an arbitrary simple curve f into an x-monotone one through an isotopy of minimum horizontal length. A lower bound on the horizontal length of such an isotopy is given in 3.6, and this lower bound not only holds for isotopies but also for homotopies. It is therefore perhaps surprising that we can construct a weak isotopy from any curve f to a monotone curve that has exactly this horizontal length, so no isotopy of less horizontal length exists.

Lemma 3.6. *If h is a homotopy from f to any x-monotone curve g, then $\mathit{length}_{hor}(h) \geq \frac{1}{2}\sup_{p \leq p'} f(p).x - f(p').x$.*

Proof. We have $\mathit{length}_{\mathrm{hor}}(\lambda_{h,p}) \geq |f(p).x - g(p).x|$ and because g is x-monotone, we have $g(p).x \leq g(p').x$. As $\mathit{length}_{\mathrm{hor}}(h) \geq \mathit{length}_{\mathrm{hor}}(\lambda_{h,p}) \geq f(p).x - g(p).x$ and $\mathit{length}_{\mathrm{hor}}(h) \geq \mathit{length}_{\mathrm{hor}}(\lambda_{h,p'}) \geq g(p').x - f(p').x$, we have $2 \cdot \mathit{length}_{\mathrm{hor}}(h) \geq f(p).x - g(p).x + g(p').x - f(p').x \geq f(p).x - f(p').x$. □

We call a curve *nice* if it is simple and contains no vertical subpaths. The projection of a nice curve onto the horizontal axis has no saddle points, and alternates between local minima and maxima using strictly increasing or decreasing paths. Consider two curves to be equivalent if there is a (degenerate) isotopy of zero cost between them. Then any simple curve has an equivalent nice curve, as such a curve can be obtained using only vertical movement by collapsing the neighborhood of vertical subpaths, such that each vertical subpath becomes a single point.

3.5.3 Monotonizing isotopies

For a simple curve f, we define a *monotonizing* isotopy Shr^f that yields a monotone curve using minimal horizontal movement. Our monotonizing isotopy will move all local minima (respectively maxima) in the positive (respectively negative) x-direction at a constant speed until two extrema cancel against each other, and repeat the procedure on the resulting curve. Formally, we will recursively define the monotonizing isotopy applied to a simple curve f. For this, call a maximal x-monotone subpath of f a *horizontal subpath*, and label those paths $\bar{f}_1, \ldots, \bar{f}_n$, in the order they appear along f. Let $I_i = [\ell_i, r_i] = \{x \mid (x, y) \in \mathrm{Im}(\bar{f}_i)\}$ be the interval of x-coordinates of points on $\bar{f}_i$. Let L be the length of the shortest such interval. For $t \leq L/2$, define $I_i^t = [\ell_i + t, r_i - t]$ and the arc $\bar{f}_i^t$ of $\bar{f}_i$ whose x-coordinates lie in I_i^t. We define $Shr_t^f = f^t$ as the curve that alternates between the horizontal paths $\bar{f}_i^t$ and the vertical segments connecting the end of $\bar{f}_i^t$ to the start of $\bar{f}_{i+1}^t$ (for $1 \leq i < n$).

Lemma 3.7. *For $t < L/2$, the curve f^t is simple.*

Proof. Suppose not, and consider the minimum t for which two points of f^t coincide. Then $t > 0$ since $f^0 = f$ is simple. Because f is simple, and different horizontal subpaths $\bar{f}_i^t$ and $\bar{f}_j^t$ of f^t are disjoint subpaths of f, subpaths $\bar{f}_i^t$ and $\bar{f}_j^t$ cannot intersect. So if f^t is non-simple, then at least one vertical segment is involved. Hence either (a) two vertical segments intersect, or (b) a vertical segment and a horizontal subpath $\bar{f}_i^t$ intersect.

We first consider case (a). The vertical segments move continuously as t varies. For $0 < t < L/2$, the $2n$ end points of the horizontal subpaths are all distinct, and at most one vertical segment starts or ends at each of them. Hence, if two vertical segments overlap, they overlap in each others interior. If both vertical segments move in the same direction, then their x-coordinates will be the same for any t, and their first overlap will occur when end points of two vertical segments coincide, which cannot happen for $0 < t < L/2$. So the overlapping vertical segments must move in opposite directions, and an end point of one lies in the interior of the other vertical segment. Since $t < L/2$, there is a horizontal subpath connected to this end point that would have intersected the other vertical segment for smaller t, which is a contradiction.

For case (b), no vertical segment has an end point interior to a horizontal subpath $\bar{f}_i^t$, as this would mean that $\bar{f}_i^t$ contains an end point of some $\bar{f}_j^t$. So either a vertical segment already intersected $\bar{f}_i^t$ for smaller t (a contradiction), or the vertical segment intersects an end point of $\bar{f}_i^t$. But since that end point of $\bar{f}_i^t$ moves in the same direction (at the same speed) as the vertical segment that intersects it, it must have already done so for $t = 0$, contradicting that f is simple. □

As t approaches $L/2$, at least one horizontal subpath shrinks to length 0. So if $f^{L/2}$ would be a nice curve, we could recursively apply the same procedure on $f^{L/2}$ until no more horizontal subpaths exist and obtain a (degenerate) isotopy that shrinks f to a point. Using the procedure of Section 3.5.2, one can indeed replace each of the curves f^t by a nice curve such that they form a degenerate isotopy through nice curves.

Lemma 3.8. *The curve $f^{L/2}$ is simple.*

Proof. The proof of Lemma 3.7 requires that $t < L/2$ only to prevent that two vertical segments share the same end point. At $t = L/2$, two vertical segments share the same end point only if some $\bar{f}_i^t$ contracts to a single point at $t = L/2$, which means that I_i has length L. Suppose for a contradiction that the interiors of the vertical segments overlap. Then the vertical segments move in different directions and are adjacent to the same horizontal subpath $\bar{f}_i^t$. Since the vertical segments have only one end point in common, an end point of one lies interior to the other vertical segment. Because I_{i-1}^t and I_{i+1}^t are at least as long as I_i^t, we have for slightly smaller t, that the horizontal subpath ($\bar{f}_{i-1}^t$ or $\bar{f}_{i+1}^t$) attached to the shorter vertical segment intersects the longer vertical segment, contradicting Lemma 3.7. ☐

An example. The major critical events are those where a minimum and maximum merge. As an example, consider a spiral as in Figure 3.5, which is similar to an example from [32]. In this spiral, a critical event first happens when the innermost maximum and minimum meet to form a new saddle. The final critical event causes the two end points to collapse. Two (minor) critical events where end points pass a saddle are not drawn.

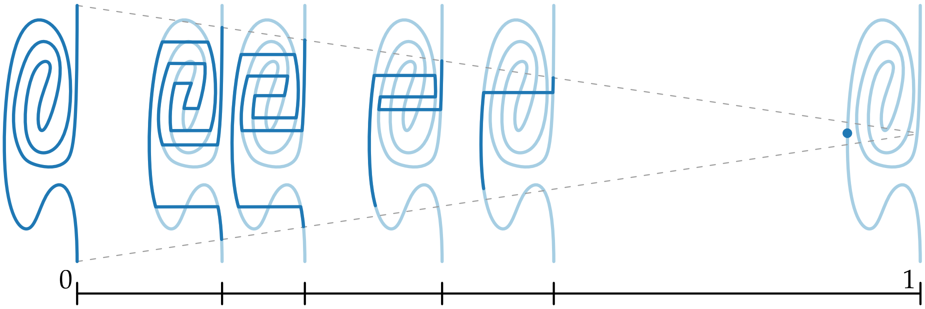


Figure 3.5: The major critical events of our shrinking homotopy (rotated 90 degrees) on a spiral.

3.5.4 Monotonicity and relation to persistence

Recall that Lemma 3.6 gave us the following lower bound on $length_{\text{hor}}(h)$ for a homotopy h turning f into an x-monotone curve.

$$length_{\text{hor}}(h) \geq \frac{1}{2} \sup_{p \leq p'} f(p).x - f(p').x.$$

We prove Lemma 3.9 to show that the isotopy Shr^f is optimal in the sense that it yields an x-monotone curve using minimal horizontal movement. It follows from Lemma 3.5 that the vertical movement required to achieve this is only $\eta + \varepsilon$ if f is η-narrow.

Lemma 3.9. *Shr^f_L is x-monotone if $L = \frac{1}{2}\sup_{p \leq p'} f(p).x - f(p').x$.*

Proof. Suppose $f' = Shr^f_L$ is not x-monotone, then there is some $q < q'$ for which $f'(q').x < f'(q).x$. Without loss of generality assume that q lies on a maximum of f' and q' lies on a minimum. Moreover, we may assume that $f'(q).x \leq f(q).x - L$, because some point on that maximum has moved distance L in the negative x-direction. Symmetrically, we may assume that $f'(q').x \geq f(q').x + L$. Therefore, $f(q').x + L \leq f'(q').x < f'(q).x \leq f(q).x - L$. This means that $f(q).x - f(q').x > 2L$ for some $q < q'$, contradicting that $L = \frac{1}{2}\sup_{p \leq p'} f(p).x - f(p').x$. □

Theorem 3.3. *Shr^f_L yields an x-monotone curve at $L = |\operatorname{Im}(f.x)|^{-1}\sup_{p \leq p'} f(p).x - f(p').x$, and there is no homotopy that achieves this with less horizontal movement.*

Persistence. The critical values of Shr^f are closely related to the notion of persistence, see [51] for an overview. In summary, we consider connected components of the sublevelsets $f.x^{-1}(-\infty, z]$ for some parameter $z \in \mathbb{R}$.

The minimum of each component is called its *representative* (we break ties consistently if there are multiple minima, so that we can speak of a minimum). When increasing z from $-\infty$ to ∞, new components are created in the sublevelset whenever z attains the x-coordinate of a minimum (which is the representative of that component), and we say these components are born at time z. On the other hand, when z attains the x-coordinate of a maximum, components may merge, causing the representative of components with larger representatives to be overtaken by smaller representatives. When such a representative is overtaken, we say the representative dies. The *persistence* of a minimum is the time between the birth and death of that component; that is, the length of the interval for which the minimum is a representative. Note that one minimum has infinite persistence.

To relate this to the critical events of Shr^f, we show that Shr^f_L has a critical event whenever $2L$ attains the value of the persistence of some minimum, see Lemma 3.10. Symmetrically, we can apply persistence in terms of maxima, and obtain critical event at Shr^f_L whenever some maximum has persistence value $2L$.

Lemma 3.10. *If the persistence of a minimum is $2L$, then a minimum and maximum merge at Shr^f_L.*

Proof. Assume that minimum p has persistence value $2L$. The connected component of $f.x^{-1}(-\infty, z]$ containing p merges with some other component at $z = f(p).x + 2L$, say q is the representative of the merged component. Then $f(q).x \leq f(p).x$. Let m be the maximum where the two components merge, so $f(m).x = f(p).x + 2L$. Let f' be the subcurve of f between p and m (directed such that $f'(0) = f(p)$ and $f'(1) = f(m)$), then $\operatorname{Im}(f'.x) \subseteq [f(p).x, f(m).x]$. Let A_i be the minimum of $\mathcal{A}^{|}(f)$ containing p, and A_j be the maximum of $\mathcal{A}^{|}(f)$ containing m. Let L' be the minimum of the critical value where A_i merges with a maximum and the critical value where A_j merges with a minimum.

Assume that $L' > L$. Then by Lemma 3.9, $Shr^{f'}_L$ is x-monotone. Because for $L'' < L'$, the arc of $Shr^{f}_{L''}$ between p and m behaves the same as $Shr^{f'}_{L''}$, we have that p and m are on consecutive extrema of $Shr^{f'}_{L''}$. But this means that A_i and A_j merge at $L'' = L$, which contradicts that $L' > L$, so we have $L' \leq L$.

To show that $L' \geq L$, assume for a contradiction that $L' < L$. Then A_i or A_j becomes a saddle at L'. First consider the case where the minimum A_i (containing p) becomes a saddle at L'. Then it merges with a maximum m' with $f(m').x < f(m).x$ and there is a minimum p' with $f(p').x < f(p).x$, where m does not lie between p and p'. But then p is not the representative of the connected component of $f.x^{-1}(-\infty, f(m').x]$ containing p, contradicting that p stops being a representative at $f(m).x > f(m').x$.

Now consider the case where the maximum A_j (containing m) becomes a saddle at L'. Then it merges with a minimum p' with $f(p').x > f(p).x$ and there is a maximum m' with $f(m').x > f(m).x$, where p and q do not lie between m and m'. But then, p and q lie in different components of the sublevelset at $z = f(p).x + 2L$ because m lies between p and q, contradicting that the components of p and q merge at $z = f(p).x + 2L$. Hence $L' = L$. □

3.6 Monotone homotopies

We say a homotopy h is *monotone* if the *leash* $h(p, \cdot)$ is a (forwards or backwards) monotone curve for all p. In this section, we define tools in order to decide whether a monotone isotopy between two curves f and g exists. We decide this by finding a monotone homotopy from f to a curve f' that lies somewhere 'between' f and g, and from a curve g' to g, where g' lies somewhere 'between' f' and g. We will see that if a monotone isotopy exists, these two homotopies are isotopies, and there is an obvious isotopy between f' and g', and these isotopies compose into a monotone isotopy from f to g.

We construct a monotone homotopy from a given curve f to some curve f' based on a continuous map τ (based on g) indicating the target x-coordinate $\tau(p)$ for each point $f(p)$. We will assume τ has finitely many local extrema. Our homotopy h from f to f' will have the property that if $f(p).x \leq \tau(p)$ and $t \leq t'$, then $f(p).x \leq h(p,t).x \leq h(p,t').x \leq \tau(p)$. Symmetric properties apply for $f(p).x \geq \tau(p)$. However, points need not always reach the target x-coordinate; that is, we will not necessarily have $f'(p).x = \tau(p)$ for all p. Instead, $f'(p).x$ will be the x-coordinate closest to $\tau(p)$, for which there is a point $f(q)$ with the same x-coordinate and $\tau(r)$ does not cross this x-coordinate for r between p and q, as illustrated in Figure 3.6. Equation 3.1 defines this coordinate formally as $\mathcal{E}_{f.x,g.x}(p)$, and we prove in Lemma 3.13 that the homotopy we construct indeed reaches these coordinates.

$$\mathcal{E}_{\sigma,\tau}(p) = \begin{cases} \min_q\{\, \sigma(q) \mid \forall_{p\leq r\leq q \text{ or } q\leq r\leq p}\ \tau(r) \leq \sigma(q) \leq \sigma(r)\,\} & \text{if } \tau(p) \leq \sigma(p) \\ \max_q\{\, \sigma(q) \mid \forall_{p\leq r\leq q \text{ or } q\leq r\leq p}\ \tau(r) \geq \sigma(q) \geq \sigma(r)\,\} & \text{if } \tau(p) \geq \sigma(p) \end{cases} \tag{3.1}$$

We then use the curve f' resulting from this homotopy to symmetrically construct a homotopy from g to g', using the x-coordinates of f' as target. Lemma 3.12 gives a key property of Equation $\mathcal{E}$, that tells us that if the projections of f and g onto the x-axis overlap, then the x-coordinates of the constructed curves are aligned. That is, $f'(p).x = g'(p).x$ for all p.

Lemma 3.11. $\mathcal{E}_{\sigma,\tau}(p) \in [\min\{\sigma(p), \tau(p)\}, \max\{\sigma(p), \tau(p)\}]$.

Proof. Assume without loss of generality that $\tau(p) \leq \sigma(p)$. Since the range r is quantified over contains p, we have $\tau(p) \leq \mathcal{E}_{\sigma,\tau}(p) \leq \sigma(p)$. □

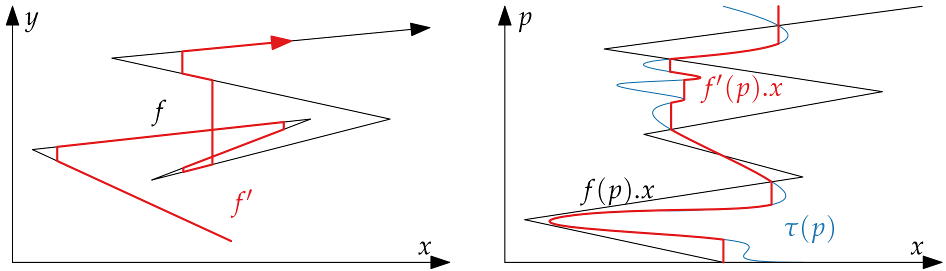
Lemma 3.12. *Let* $f'.x = \mathcal{E}_{f.x,g.x}$ *and* $g'.x = \mathcal{E}_{g.x,f'.x}$*. If the images of* $f.x$ *and* $g.x$ *overlap, then* $f'.x(p) = g'.x(p)$*.*

Proof. Consider any p, and without loss of generality assume that $f.x(p) \leq g.x(p)$. By the equation, we have $g.x(r) \geq f'.x(p) \geq f.x(r)$ for r between p and q, where $f.x(q) = f'.x(p)$. For $x > f'.x(p)$, there either is no q with $f.x(q) = x$, or for all such q there is some $g.x(r) = x$ with r between p and q. In the former case, $f.x(q)$ is a global maximum of f (and also of f'), and because $f'.x(p) \leq g.x(p)$ and the images of $f.x$ and $g.x$ overlap, there is some q' with $g.x(q') = f'.x(p)$. Using this value of q' as argument for the minimum in the definition of $g'.x(p)$ yields that $g'.x(p) \leq f'.x(p)$, and hence $g'.x(p) \leq f'.x(p)$. In the latter case where $f.x(q)$ is not a global maximum of f, such q' also exists (otherwise $f'.x(p)$ would have been larger), and the same argument shows that $g'.x(p) \leq f'.x(p)$. As Lemma 3.11 implies that $f'.x(p) \leq g'.x(p)$, we conclude that $f'.x(p) = g'.x(p)$. □

Let $P^{<}_{f,\tau} = \{p \mid f(p).x < \tau(p)\}$, $P^{=}_{f,\tau} = \{p \mid f(p).x = \tau(p)\}$, and $P^{>}_{f,\tau} = \{p \mid f(p).x > \tau(p)\}$. Let $f_<$, $f_=$, and $f_>$ be the respective restrictions of f to $P^{<}_{f,\tau}$, $P^{=}_{f,\tau}$, and $P^{>}_{f,\tau}$. For a local minimum[3] m of $f_<.x$, let a_m be the arc of $f_<$ consisting of the (one or two) horizontal subpaths of $f_<$, one of whose end points is $f_<(m)$. Let a^t_m be the subpath of a_m consisting of points $a_m(p)$ with $a_m(p).x \leq f_<(m).x + t$. For a

---

[3] Because the domains of $f_<$ and $f_>$ generally are not closed sets, we allow m to be drawn from their closures, so that $f(m)$ might actually be an endpoint of an arc of $f_=$.



(a) (b)

Figure 3.6: (a) The curve f, and the curve f' that results from the function τ shown in (b).

local maximum m of $f_>.x$, define a_m^t symmetrically as the subpath of a_m consisting of points $a_m(p)$ with $a_m(p).x \geq f_>(m).x - t$.

Let M be the set of values m for which $f.x(m)$, $f_=.x(m)$, or $\tau(m)$ is a local minimum or maximum. Let $X = f.x(M) \cup \tau(M)$ be the values of f and τ at such m, and let L be the minimum positive difference between distinct values of X.

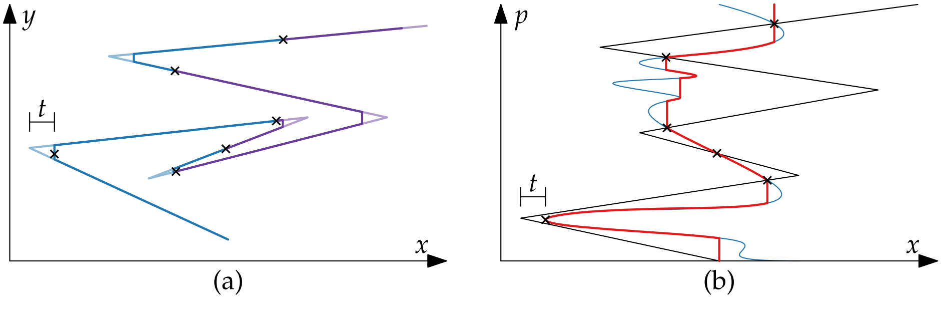
We define the curves f^t of our homotopy recursively. Consider the set M' of values of m for which $f(m)$ is a local minimum of $f_<$, or a local maximum of $f_>$. Let $f^0 = f$. For $0 < t \leq L$, we obtain f^t from f by replacing each a_m^t for each $m \in M'$, as follows.

(a) If $f(m)$ is an end point of f (and hence of a_m^t), we replace the arc a_m^t by its other end point.

(b) Otherwise, if a_m^t has only one horizontal path, then we may assume that $f(m).x = \tau(m)$. In this case, we replace a_m^t by a path b_m^t, with the same image, but possibly a different parameterization. If m is a local minimum of $f_<.x$, parameterize b_m^t such that $b_m^t(p).x = \tau(p)$ if $b_m^t(p).x < f(m).x + t$, and $b_m^t(p).x \leq \tau(p)$ if $b_m^t(p).x = f(m).x + t$. One can always find such a parameterization because $f(m).x = \tau(m)$, $a_m^t(p) \leq \tau(p)$, and as p goes from m to the other end of a_m^t, the value of $\tau(p)$ varies continuously and, by our choice of L, does not decrease. Symmetrically, if m is a local maximum of $f_>.x$, parameterize b_m^t such that $b_m^t(p).x = \tau(p)$ if $b_m^t(p).x > f(m).x - t$, and $b_m^t(p).x \geq \tau(p)$ if $b_m^t(p).x = f(m).x - t$.

(c) Otherwise, a_m^t must consist of two horizontal paths connected at $f(m)$, and we replace a_m^t by a (suitably parameterized) vertical segment between the two end points of a_m^t.

Note that f^t varies continuously with t and that $P^<_{f^t,\tau}$, $P^=_{f^t,\tau}$, and $P^>_{f^t,\tau}$ do not change combinatorially for $0 \leq t < L$. An example of a combinatorial change is illustrated in Figure 3.7. For $t = L$, the first combinatorial change may occur, so let $\hat{f} = f^L$, and recursively for $t > L$, define $f^t = \hat{f}^{t-L}$ to be the curve obtained by applying



(a) (b)

Figure 3.7: (a) $P^=_{f^t,\tau}$ indicated by crosses, $P^<_{f^t,\tau}$ in blue, $P^>_{f^t,\tau}$ in purple. The illustrated value of t corresponds to the smallest t for which $P^=_{f^t,\tau}$ changes combinatorially: the leftmost cross is newly added to $P^=_{f^t,\tau}$. (b) Functions $f.x$, $f'.x$ and τ.

the same procedure (with τ) to $\hat{f}$. Because τ has finitely many extrema, this procedure will eventually result in a curve f' for which the arcs of $f^t_<$ and $f^t_>$ all lie on vertical segments.

Lemma 3.13. *Let f' be the curve resulting from the above homotopy based on f and g, then $f'.x = \mathcal{E}_{f,g}$.*

Proof. Consider a value of p where $g(p).x \leq f(p).x$; the other case is symmetric. Then we have $f'(p) \leq f(p)$. Let $x = \min\{ f(q).x \mid \forall_{r \in [\min(p,q), \max(p,q)]}\ g(r) \leq f(q).x \leq f(r).x \}$ be the value that Equation 3.1 would assign to $f'(p).x$. It suffices to show that $f'(p).x = x$. Let $R \subseteq [0,1]$ be the maximal interval for which $\{p \in R \wedge g(r) \leq f'(p).x \leq f(r).x\}$ and $f(r) \neq f'(p).x$ except possibly if r is an end point of R. Let q_1 and q_2 be the endpoints of R. For q_1 we have $f(q_1).x = f'(p).x$, or $g(q_1).x = f'(p).x$, or $q_1 = 0$. Similarly, for q_2 we have $f(q_2).x = f'(p).x$, or $g(q_2).x = f'(p).x$, or $q_2 = 1$. If neither $f(q_1).x = f'(p).x$ nor $f(q_2).x = f'(p).x$, then p must lie on a vertical segment because if it were on a reparameterized horizontal subpath, we would have $q_1 = q_2 = p$. However, there cannot be a vertical segment containing p, as f' consists of subpaths of f with vertical segments between them. Hence $x \leq f'(p)$, and we have $f(q_1).x = f'(p).x$ or $f(q_2).x = f'(p).x$. If $x < f'(p)$, then since $x < f'(q_1).x$ and $x < f'(q_2).x$, $f'(p)$ would have been reparameterized or moved by a vertical segment. So $f'(p).x = x$. $\square$

Isotopies. We give conditions for which the above homotopy is an isotopy. To prevent vertical segments that are connected by a horizontal segment from creating intersections, we collapse vertical segments that form saddle-points (of $f^t.x$) to points using only vertical movement. Lemma 3.14 captures when f^t stops being simple already in the non-recursive part of the homotopy.

Definition 3.1 (Pocketed point)**.** *We say p is a pocketed point between $f.x$ and τ if there exist $u_1 < u_2$ with $p \notin [u_1, u_2]$ and $f(p).y$ strictly between $f(u_1).y$ and $f(u_2).y$, and for all $u \in [u_1, u_2]$, we have either $\tau(p) \leq f(p).x$ and $f(u).x \leq f(p).x \leq \tau(u)$, or $\tau(p) \geq f(p).x$ and $f(u).x \geq f(p).x \geq \tau(u)$.*

Lemma 3.14. *If f is weakly simple, but f^t is not for $t \leq L$, then $t = L$ and there exists a pocketed point between $f.x$ and τ.*

Proof. Let t be the minimal value for which f^t is not weakly simple. Then there exists a vertical segment on f^t whose causes an intersection, since the remainder of f^t is only reparameterized with respect to f. The interior of the vertical segment must touch an endpoint $f^t(p)$ of a horizontal subpath of f^t. Such an intersection can be created only when $t \geq L$, so let $t = L$.

Assume that the vertical segment causing the intersection moves in the positive x-direction (the other case is symmetric). If $p \in P^<_{f,\tau}$, there would either have been an intersection for smaller t, or $f^t(p)$ has not moved because there is a point $p' \in P^=_{f,\tau}$ for which $f^t(p')$ is an end point of a horizontal subpath, and the vertical segment also intersects $f^t(p')$. Thus, we may assume $p \in P^>_{f,\tau} \cup P^=_{f,\tau}$, so $\tau(p) \leq f(p).x$. Moreover, we may assume that $f(p).x = f^t(p).x$, because if $f^t(p)$ had moved, there would also be a point of $p' \in P^{>}_{f,\tau} \cup P^{=}_{f,\tau}$ with $f(p') = f^t(p)$ on the same horizontal subpath. The vertical segment causing the intersection connects two points $f(u_1)$ and $f(u_2)$ on horizontal subpaths of f. Because the vertical segment is moving in the positive x-direction, we have $f(u).x \leq f^t(p).x = f(p).x \leq \tau(u)$ for $u \in [u_1, u_2]$. □

Lemma 3.14 can be extended to Lemma 3.15, which shows that f^t stops being weakly simple in a (perhaps) recursive part of the homotopy only if there is a pocketed point.

Lemma 3.15. *If f is weakly simple, but f^t is not, then there exists a pocketed point between $f.x$ and τ.*

Proof. Assume f^t stops being weakly simple in the recursive step starting out from curve $\hat{f}$. Then $\hat{f}$ satisfies the properties of Lemma 3.14, say for values $\hat{u}_1$, $\hat{u}_2$, and $\hat{p}$. The points $\hat{f}(\hat{u}_1)$ and $\hat{f}(\hat{u}_2)$, and the point $\hat{f}(\hat{p})$ lie on horizontal subpaths of $\hat{f}$, and hence on f. We will find u_1, u_2 and p with $f(u_1) = \hat{f}(\hat{u}_1)$, $f(u_2) = \hat{f}(\hat{u}_2)$, and $f(p) = \hat{f}(\hat{p})$ on f. Suppose that $\tau(\hat{p}) \leq \hat{f}(\hat{p}).x$ and $\hat{f}(\hat{u}).x \leq \hat{f}(\hat{p}).x \leq \tau(\hat{u})$ for $\hat{u} \in [\hat{u}_1, \hat{u}_2]$ (the other case is symmetric). Since each point moves monotonely throughout the homotopy, u_1 and u_2 can be chosen such that $f(u).x \leq \hat{f}(\hat{p}).x \leq \tau(u)$ for $u \in [u_1, u_2]$. Moreover, p can be chosen such that $\tau(p) \leq f(p).x$, since otherwise $\hat{f}(\hat{p})$ would not have reached the same coordinate as $f(p)$. □

Conversely, it follows from Lemma 3.1 that if there is a pocketed point between $f.x$ and $g.x$ or between $g.x$ and $f.x$, then no x-monotone isotopy from f to g exists. So assume there are no such pocketed points, then there are also no pocketed points between g and the curve f', where f' is the curve resulting from applying the above homotopy to f with $\tau = g.x$. Hence the above homotopy (applied to $f.x$ with $\tau = g.x$) is an isotopy, and so is the homotopy applied to $g.x$ with $\tau = f'.x$.

To construct an x-monotone isotopy from f to g, it remains to construct an x-monotone isotopy between f' and g'. For this we consider two cases, namely when the images of $f.x$ and $g.x$ are disjoint, in which case we can construct a trivial isotopy, and the case where the images of $f.x$ and $g.x$ overlap, in which case Lemma 3.12 tells us that $f'.x(p) = g'.x(p)$ for all p. In this case, there is an isotopy if f' and g' are vertically equivalent, and we show that if f' and g' are not vertically equivalent, there is no x-monotone isotopy between f and g.

Lemma 3.16. *If for weakly simple curves f and g, the images of $f.x$ and $g.x$ are disjoint, there is an x-monotone isotopy between f and g.*

Proof. Assume that $\max_p f.x(p) < \min_p g.x(p)$ (the other case is symmetric). Pick x between $\max_p f.x(p)$ and $\min_p g.x(p)$. Let f' be a horizontal segment for which $\max_p f.x(p) < f'.x(p) < \min_p g.x(p)$. We will construct an x-monotone isotopy from f to f'. Such an isotopy can be obtained by translating the curves of the monotonizing isotopy of Section 3.5.3 in such a way that all points move in the positive x-direction. Symmetrically, construct an x-monotone isotopy from g to g'. These isotopies compose to form an x-monotone isotopy from f to g. □

Lemma 3.17. *If for weakly simple curves f and g, the images of $f.x$ and $g.x$ overlap, and there is no pocketed point between f and g or vice-versa, and f' and g' are not vertically equivalent, then no x-monotone isotopy between f and g exists.*

Proof. By Lemma 3.12, $f'.x(p) = g'.x(p)$ for all p. So if f' and g' are not vertically equivalent, then there exist p' and q' and x^* such that $f'(p').x = g'(p').x = f'(q').x = g'(q').x = x^*$, and $f'(p').y > f'(q').y$ and $g'(p').y < g'(q').y$. We may assume these four points all lie on horizontal subpaths (otherwise increase the difference between p' and q' to move the points vertically until they are all on horizontal subpaths). Since the points lie on horizontal subpaths, there exist p_f, p_g, q_f, and q_g such that $f(p_f) = f'(p')$, $g(p_g) = g'(p')$, $f(q_f) = f'(q')$, and $g(q_g) = g'(q')$. Moreover, we may assume that p' lies between p_f and p_g, and q' lies between q_f and q_g, and the interval between p_f and p_g is disjoint from that between q_f and q_g.

We claim that the points on the arc of f between p_f and p_g can be assumed to all move in the same x-direction. Assume that $f(p').x \leq g(p').x$ (a symmetric argument applies if $f(p').x \geq g(p').x$). Then because $f(p')$ reaches $f'(p')$, we may assume that $f(p).x \leq g(p).x$ for all p between p_f and p'. Similarly, we may assume that $f(p).x \leq g(p).x$ for all p between p' and p_g. By the same argument, the points on the arc of f between q_f and q_g can be assumed to all move in the same x-direction.

Suppose for a contradiction that there exists an x-monotone isotopy h from f to g. Let P_t and Q_t be the arcs of h_t between p_f and p_g, and between q_f and q_g, respectively. Let p_t be the value closest to p_g, for which the x-coordinate of $h(p_t, t)$ is x^*. Assume that the points of Q_t are moving in the positive x-direction (the other case is symmetric). Then all points on Q_0 have x-coordinate at most x^*, and all points of Q_1 have x-coordinate at least x^*.

Consider the subpaths of Q_t whose x-coordinates are at most x^*. We call such a subpath persistent if it contains a point whose target x-coordinate is at most x^*. We say that such a subpath lies below p_t if none of its points with x^* as x-coordinate lie above $h(p_t, t)$. We claim that for any t, all persistent subpaths of Q_t lie below p_t. Note that $f(q_f)$ lies on such a persistent subpath of Q_0 below p_0. We will derive a contradiction since $g(q_g)$ lies on a persistent subpath of Q_1 that does not lie below p_1. It remains to show that all persistent subpaths of Q_t lie below p_t. Initially all persistent subpaths of Q_0 lie below p_0, for the entirety of Q_0 is a persistent subpath, meaning that there would be a pocketed point on P_0 otherwise. We show that this property is maintained throughout the isotopy. As t increases, there are two cases where the isotopy might gain a persistent subpath of Q_t that does not lie below p_t: either a persistent subpath of Q_t gains a new point with x^* as x-coordinate above $h(p_t, t)$, or p_t jumps discontinuously below a point of a persistent subpath of Q_t with x^* as x-coordinate.

First consider the case where the points of P_t move in the negative x-direction. Then all points on P_0 have x-coordinate at least x^*, and all points of P_1 have x-coordinate at most x^*. If p_t jumps below a point of a persistent subpath of Q_t with x^* as x-coordinate, then that point of Q_t is pocketed by the arc of P_t between p_t and the location it jumped from, contradicting that an isotopy exists. If instead, a persistent path of Q_t gains a new point with x^* as x-coordinate above $h(p_t, t)$, then p_t is pocketed.

So consider the case where the points of P_t move in the positive x-direction. Then all points on P_0 have x-coordinate at most x^*, and all points of P_1 have x-coordinate at least x^*. If p_t jumps below a point of a persistent subpath of Q_t with x^* as x-coordinate, then that persistent subpath contains a point (that causes it to be persistent) that is pocketed by the arc of P_t between p_t and the location it jumped from. If instead, a persistent path of Q_t gains a new point with x^* as x-coordinate above $h(p_t, t)$, then p_g is pocketed. Hence, there does not exist an x-monotone isotopy. □

Theorem 3.4. *For weakly simple curves f and g, there is an x-monotone isotopy between them if and only if the images of $f.x$ and $g.x$ are disjoint, or f' and g' are vertically equivalent and there is no pocketed point between $f.x$ and $f'.x$ and there is no pocketed point between $g.x$ and $f'.x$.*

3.7 Discussion

In this chapter we have refuted a conjecture on the length of an optimal isotopy between two particular input curves by providing an optimal isotopy. Moreover, we have presented the first algorithms to compute the isotopic Fréchet distance in various monotone settings. Specifically, when f and g are both x-monotone curves in the plane we can compute optimal isotopies using traditional algorithms for computing the Fréchet distance between curves. Furthermore, we can provide a way to construct an isotopy of minimal horizontal movement from an arbitrary curve to an x-monotone one, the constructed isotopy is closely related to the notion of persistence.

Finally, given a correspondence between the points of f and g, we construct a corresponding isotopy in which each point moves monotonely in only the positive, or only the negative x-direction, or we conclude that no such isotopy exists. The more general question of computing the length of an optimal isotopy for a given correspondence between f and g remains open, even if only horizontal movement is measured. Moreover, if the correspondence between the points of f and g is not provided, we are interested in finding a correspondence between f and g that minimizes the length of the shortest isotopy.

Part II

Homotopy Height

4

Computational Complexity of Optimal Homotopies

In the second part of this thesis, we aim to find homotopies between curves on surfaces, in such a way that the length of the longest intermediate curve in an optimal homotopy is minimized.

Computational questions pertaining to homotopies can be formalized either in a continuous setting, where homotopies constitute one of the fundamental constructs of algebraic topology, but also in a more discrete one, where the input space is a simplicial, or more generally, a cellular description of a topological space. This latter setting will be the focus of this chapter. While considerably more restrictive than the more traditional (continuous) mathematical settings, this setting is nonetheless of key importance in applications areas such as graphics or medical imaging, where inputs are generally represented by triangular meshes built upon scanned point sets from an underlying 3D object.

Investigating homotopies from a computational perspective is a well-studied problem, dating back to the work of Dehn [44] on contractibility of curves, which has strong ties to geometric group theory. Whereas deciding whether two curves in a 2-dimensional complex are homotopic is well-known to be undecidable in general (see for example Stillwell [85]), efficient, linear-time algorithms have been designed to test homotopy [46, 73, 55] if the underlying space is a surface. In this chapter, we add a quantitative twist to this problem. The height of a homotopy is the length of the longest intermediate curve in the homotopy, as formalized in Section 4.1. The HOMOTOPY HEIGHT problem asks, given two homotopic curves on a combinatorial surface, to find the height of a homotopy of minimal height. The notion of homotopy height has obvious appeal from a practical perspective, as it quantifies how long a curve has to be to overcome a hurdle. For example, deciding whether a bracelet is long enough to slide off over a hand without breaking is closely related to the question of homotopy height. Deformations of minimal height minimize the necessary stretch and can in this way be used to quantify how similar closed curves are.

Results. We begin by considering two curves that form the boundary of a discrete annulus, and study the structure of minimum height homotopies between these curves. This chapter builds on several recent results in Riemannian geometry [36, 37], stating that an optimal homotopy exists that is an isotopy, so that all the intermediate curves remain simple. Moreover, we build on our own recent result [27] where we prove that in the Riemannian setting, such an optimal isotopy can be assumed to move in only one direction. That is, it never sweeps any portion of the annulus twice. We refer to this property as *monotonicity* and will define it more precisely in Section 4.2.

Once we translate those results from the Riemannian to the discretized setting, these isotopy and monotonicity properties turn out to be a key ingredient for computational purposes. In the discrete setting, we can easily modify an optimal monotone isotopy via operations that perform the same moves in a different order, and operations that avoid unnecessary moves. We call such operations surgeries, and they allow us to prove that HOMOTOPY HEIGHT is in NP (Theorem 4.5). We note that our setting is very general, as it also implies NP-membership for variants of HOMOTOPY HEIGHT in more restricted settings that were considered in earlier papers [31, 66], as well as for the HOMOTOPIC FRÉCHET DISTANCE, where this question was still open despite the recent articles investigating this distance [29, 66]. Further surgery arguments allow us to provide an efficient factor $O(\log n)$-approximation algorithm for HOMOTOPY HEIGHT (Corollary 4.7), by relying on an earlier $O(\log n)$-approximation algorithm of Har-Peled et al. [66] for homotopy height in a more restricted setting. Finally, we show that monotonicity directly implies an equivalence between the HOMOTOPY HEIGHT problem and a seemingly unrelated graph drawing problem which we call the MINIMAL HEIGHT LINEAR LAYOUT problem. Therefore, this problem is also in NP and we provide an efficient factor $O(\log n)$-approximation for it.

Related work. Optimal homotopies (for several notions of optimality) have been studied extensively in the mathematical community. Broadly speaking, such homotopies are considered in the field of quantitative homotopy theory, pioneered by Gromov [64]. This field aims at introducing a quantitative lens in the study of topological invariants on manifolds. The most extensively considered notion of optimality is probably that of minimizing the area swept by a homotopy; several variants of this problem are discussed in [72], and connections between minimum area homotopies and homologies in higher dimensions are discussed in [90]. The notion of controlling the width of a homotopy has also been studied [26, 69], see also Figure 4.1. Recent work on minimum height homotopies in the Riemannian setting [36, 37] laid the foundation for the results in this chapter.

The problem of comparing curves has heavily been motivated by the Fréchet distance. Generalizing the Fréchet distance to surfaces has led to the homotopic Fréchet distance, which is essentially the same as finding a minimum width homotopy given two input curves on a surface; algorithms can compute this quantity in polynomial time in the case where the surface is the plane minus a set of obstacles [29]. Moreover, in the discrete setting, approximation algorithms exist for the case where the two input curves bound a disk [66].

More directly, minimum height homotopies have been studied from the computational perspective in various discretized settings [31, 66], although this has mainly led to discussions of the local structure of optimal homotopies, and upper bounds on their height. Indeed, as it was not known if the optimal height homotopy was even monotone, the complexity of the problem was completely open.

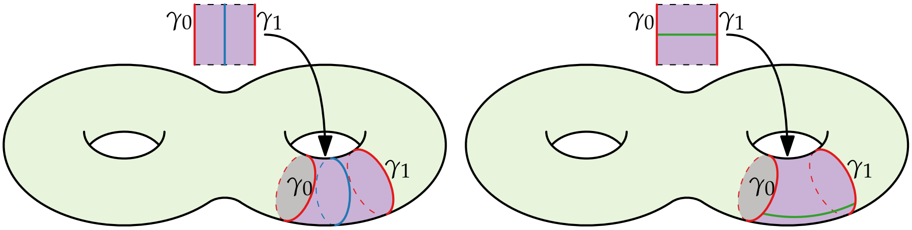


Figure 4.1: Left: the height of a homotopy between homotopic curves γ_0 and γ_1 measures the maximum amount an intermediate curve must stretch during the homotopy. Homotopies minimizing this amount of stretch measure the homotopy height. Right: the width of a homotopy measures the maximum length of a "slice" of the homotopy connecting the two boundary curves. Homotopies minimizing the length of this slice measure the homotopy width, also known as the homotopic Fréchet distance.

A closely related problem also arises as a combinatorial question in lattice theory as a *b-northward migration* [13], where the authors leave monotonicity of such migrations as an open question. Our result [27] on the monotonicity of optimal homotopies also holds in more geometric settings. One such geometric setting concerns morphing across a polygonal domain in Euclidean space with weighted point obstacles. This setting is examined in Chapter 5.

Relations to graph searching and width parameters. The current chapter also connects to sweep and search parameters in graph theory; see for example [57] for a survey of this topic. In each variant, the goal is to find or isolate a fugitive using the minimum number of searchers. For example, in the node searching variant, the fugitive hides on edges. Initially, all edges are a contaminated in the sense that they are a candidate location for the fugitive, and the searchers clear an edge if two are on its two neighboring vertices. In this variant, edges can be recontaminated if they are connected to a contaminated edge by a path without searchers. If moving the searchers allows all edges to be decontaminated, the fugitive can be found and the number of searchers suffices.

One key issue in these graph searching problems is precisely that of monotonicity, or of determining whether in an optimal strategy, edges get recontaminated. In the node searching variant, monotonicity was established by Lapaugh [71], and the argument was simplified by Bienstock and Seymour [11]. One important corollary to monotonicity for these games is that it immediately shows the problem lies in NP, since a strategy can be certified by the list of edges cleared.

Our homotopy problem is quite similar to these graph parameters; sweeping a surface using a curve while keeping its length small is intuitively quite similar to blocking in a fugitive. However, our problem does display minor technical differences with the aforementioned variant – most notably, our setting is naturally edge-weighted and the cost is measured on the edges and not the vertices – the key difference is the one of *connectedness*, as node-searching games may allow for disconnected strategies. An important variant of node searching, called *connected* *node searching*, requires additionally that the decontaminated space remains connected, but makes no restriction on the contaminated space.

Connected variants of various graph parameters give rise to *connected pathwidth* [45] and *connected treewidth* [58], but in contrast to our homotopies, these parameters are only connected "on one side", which makes them incomparable. We believe that the "doubly-connected" aspect of homotopy height makes it a worthwhile new graph parameter which could offer insights to other parameters in this area.

For graph searching problems, the main argument to establish monotonicity does not maintain connectivity [11], and it was proven that for connected node searching, it may indeed be necessary to use a non-monotone strategy [91]. By contrast, Theorem 4.4 establishes monotonicity of the optimal homotopy in our setting, and the arguments differ radically from the ones of Lapaugh and Bienstock and Seymour. As such, we identify in this chapter a new variant of graph searching which is in NP.

In this chapter, we introduce the MINIMUM HEIGHT LINEAR LAYOUT problem which asks to draw an embedded graph with the same embedding such that the height of the drawing is minimized. This graph drawing problem is closely related to the minimum cut linear arrangement problem (also known as cut-width). However, in contrast to the MINIMUM HEIGHT LINEAR LAYOUT problem which we show to be equivalent to HOMOTOPY HEIGHT, the cut-width problem does not require any embedding to be preserved. While the cut-width problem is known to be NP-hard [74], the known reductions rely on changing the embedding of the input graph, and hence we have not found an immediate reduction to establish whether HOMOTOPY HEIGHT is NP-hard.

Outline of the chapter. After introducing the preliminaries in Section 4.1, we lay the foundations of our results by explaining the structural theorems we rely on in Section 4.2. In Section 4.3 we establish surgery lemmas based on the idea of *retractions*. Then, in Section 4.4 we prove that HOMOTOPY HEIGHT is in NP. In Section 4.5 we draw connections with HOMOTOPIC FRÉCHET DISTANCE, and we leverage these connections to provide an $O(\log n)$-approximation algorithm for HOMOTOPY HEIGHT.

4.1 Preliminaries

Homotopy and Isotopy. Let Σ be a surface, endowed with a cellularly embedded graph G with n vertices such as in Figure 4.2, and let γ_0 and γ_1 be two simple cycles on G bounding an annulus.

A *discrete homotopy* h between γ_0 and γ_1 is a sequence of cycles $h(t_i)$ with $0 = t_0 \leq \cdots \leq t_i \leq \cdots \leq t_m = 1$, with $h(t_0) = \gamma_0$ and $h(t_1) = \gamma_1$ and any two consecutive paths $h(t_i)$ and $h(t_{i+1})$ are one *move* apart. The intermediate curves $h(t)$ are called *level curves* or *intermediate curves*. A move is either a face-flip, an edge-spike or an edge-unspike (flip, spike or unspike, for short).

A *face-flip* for a face F replaces a single subpath p of $h(t_i) \cap \partial F$ with the path $\partial F \setminus p$ in $h(t_{i+1})$. An *edge-spike* for an edge $u \to v$ replaces a single occurrence of a vertex $u \in h(t_i)$ by the path $u \to v \to u$ consisting of two mirrored copies of that edge in $h(t_{i+1})$. Symmetrically, an *edge-unspike* replaces a path $u \to v \to u$ of $h(t_i)$ by the single vertex u in $h(t_{i+1})$. The *length* $\ell(h(i))$ of a path $h(i)$ is the sum of the weights of its edges (with multiplicity). The *height* of a homotopy h is the length of the longest path $h(t_i)$. An *optimal homotopy* is one that minimizes the height. The *homotopy height* between γ_0 and γ_1 is the height of an optimal homotopy between γ_0 and γ_1. Figure 4.3 illustrates an optimal homotopy that uses only face-flips for the instance of Figure 4.2.

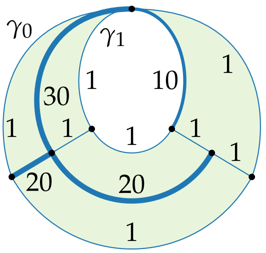


Figure 4.2: Example instance G, based on an example in [13].

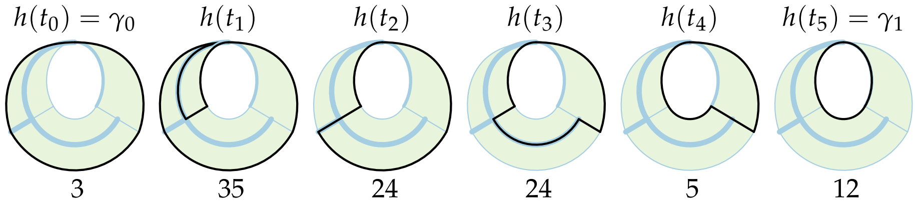


Figure 4.3: An optimal homotopy h of height 35 for the instance of Figure 4.2.

Cross Metric Surfaces. For most purposes, it is more convenient to think of this discrete model in a dual way, relying on the *cross-metric surfaces* [38] which are becoming increasingly used in the computational geometry and topology literature. In this dual setting, a cross-metric surface is a surface Σ endowed with a weighted (dual) graph G^*.

Assuming the primal surface is connected, we obtain this dual graph by gluing a disk to each boundary component, taking the dual graph, and puncturing the vertices corresponding to the added disks, without removing the adjacent edges. Such that these (dual) edges end at the boundary of the cross-metric surface instead of at a vertex, see Figure 4.4.

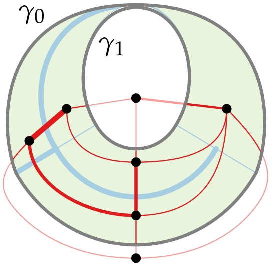


Figure 4.4: Dual representation of Figure 4.2.

For a curve γ on Σ with a finite number of crossings with G^*, its length $\ell(\gamma)$ is the weighted sum of the crossings $\gamma \cap G^*$. Now, a homotopy between γ_0 and γ_1 is a homotopy in the usual sense, that is, a continuous map $h\colon S^1 \times [0,1] \to \Sigma$ such that $h(\cdot, 0) = \gamma_0$ and $h(\cdot, 1) = \gamma_1$, except that we require that the values of t for which $h(\cdot, t)$ is not in general position with G^* are isolated, and each such curve has at most one such degeneracy[1] $h(x,t)$ with G^*. As before, the height of

---

[1] Any homotopy can be made so by a small perturbation without increasing the height, so we always

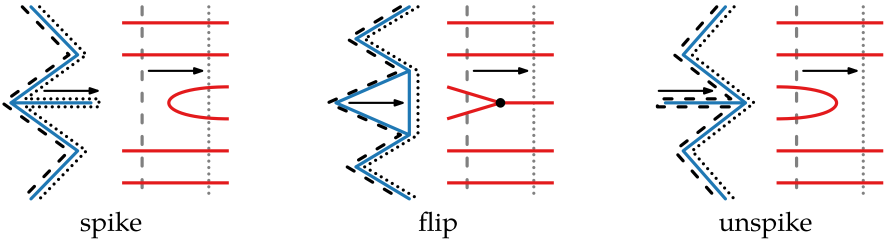


Figure 4.5: Three moves in the primal (left) and dual (right) representation.

a homotopy is defined as the maximal length of an intermediate non-degenerate level curve $h(t)$. A homotopy is an *isotopy* if all the intermediate curves are simple. Given a homotopy h^* in this setting, we obtain a discrete homotopy h on the primal graph G on Σ as follows. Pick a curve $h^*(t_i)$ in each maximal interval of non-degenerate curves in h^* (all curves in such interval have the same crossing pattern with G^*, and therefore the same length). Let $h(t_i)$ be the curve on G whose sequence of vertices and edges corresponds to the sequence of faces and edges of G^* visited by $h^*(t_i)$. This model is dual to the previous one, and Figure 4.5 illustrates how any move (flip, spike or unspike) connects two intermediate curves $h(t_i)$ and $h(t_{i+1})$. We say a discrete homotopy is an isotopy if it can be obtained from an isotopy in the dual setting.

4.2 Isotopies and monotonicity of optimal homotopies

We begin by restating and explaining the two structural results that we will rely on. Introducing the relevant Riemannian background lies outside of the scope of this chapter, so we will simply advise the uninitiated reader to picture a Riemannian surface as a surface embedded into $\mathbb{R}^3$, where the metric on the surface is the one induced by the usual Euclidean metric of $\mathbb{R}^3$. Thanks to the Nash-Kuiper embedding theorem (see [65]), this naive idea loses no generality. We refer to standard textbooks on the subject for more proper background on Riemannian geometry, for example do Carmo [49].

The first theorem shows that up to an arbitrarily small additive factor, the homotopy of minimal height between two simple closed curves can be assumed to be an isotopy.

Theorem 4.1 ([36, Theorem 1.1]). *Let Σ be two-dimensional Riemannian manifold with or without boundary, and let γ_0 and γ_1 be two non-contractible simple closed curves which are homotopic through curves bounded in length by L via a homotopy γ. Then for any $\varepsilon > 0$, there is an isotopy $\tilde{\gamma}$ from γ_0 to γ_1 through curves of length at most $L + \varepsilon$.*

Remark. The non-contractibility hypothesis is required because if M is not a sphere, contractible cycles with opposite orientations are homotopic but not iso-

---

consider this hypothesis fulfilled in the remainder of the chapter.

topic. However, if we disregard the orientations, the result holds in full generality.

This theorem has the following discrete analogue:

Theorem 4.2. *Let (Σ, G^*) be a cross-metric surface, and let γ_0 and γ_1 be two non-contractible simple closed curves on (Σ, G^*) which are homotopic through curves bounded in length by L via a homotopy γ.*

Then there is an isotopy $\widetilde{\gamma}$ from γ_0 to γ_1 through curves of length at most L.

The proof is exactly the same as the one of Theorem 4.1, except that it does not need the ε-slack: this was required to slightly perturb the curves so that they are simple but in the discrete setting it can be done with no overhead.

The second theorem shows that, *when the starting and finishing curves of a homotopy are the boundaries of the manifold*, there exists an optimal homotopy that is monotone, i.e., that never backtracks, once again up to an arbitrarily small additive factor. Formally, if γ is an isotopy (which we can assume the optimal homotopy to be, by Theorem 4.1) between γ_0 and γ_1, for $0 \leq t \leq 1$, the curves γ_t and γ_1 bound an annulus A_t. Then the isotopy γ is *monotone* if for $t < t' < 1$, $\gamma_{t'}$ is contained in A_t.

Theorem 4.3 ([27, Theorem 1.2 and the following paragraph]). *Let M be a Riemannian annulus with boundaries γ_0 and γ_1 such that there exists a homotopy between γ_0 and γ_1 of height less than L. Then there exists a monotone homotopy between γ_0 and γ_1 of height less than L.*

Note that the ε-slack of Theorem 4.1 is also present here but is hidden in the open upper bound of L. In this theorem, as was observed by Chambers and Rotman [37], crediting Liokumovitch, the hypothesis that the manifold is entirely comprised between both curves is necessary: see [37, Figure 5] for a counter-example.

In the discrete setting, the corresponding result is the following, where the definition of monotonicity is the same:

Theorem 4.4. *Let (Σ, G^*) be a cross-metric annulus with boundaries γ_0 and γ_1 such that there exists a homotopy between γ_0 and γ_1 of height L. Then there exists a monotone isotopy between γ_0 and γ_1 of height L.*

The proof is exactly identical to the one in the Riemannian setting and it yields a slightly stronger result, since the cross-metric setting removes the need for perturbations and thus the need of an ε-slack.

Remark. Observe that the discrete theorems are in some way more general than the Riemannian ones: not only do they bypass the need for some ε-slack, but they also directly imply their Riemannian converses by the following reduction. Starting with a Riemannian surface, and a (non-monotone) isotopy between two disjoint curves, one can find a triangulation of the surface allowing, at an ε-cost, to approximate the isotopy using only elementary moves. After making this isotopy monotone in the discrete setting, it can be translated back into a monotone isotopy in the Riemannian setting by interpolating between the face and edge moves.

4.3 Retractions and pausing at short cycles

In this section, we establish several technical lemmas which are necessary for our proofs in the next section. For simple closed curves β and γ that bound an annulus, denote that annulus by $A(\beta, \gamma)$. Let $\mathcal{S}(\beta, \gamma)$ be the set of closed curves in $A(\beta, \gamma)$ homotopic to boundaries β and γ, that do not intersect homotopic curves of shorter length. Then, for any point $p \in \alpha \in \mathcal{S}(\beta, \gamma)$, α is a shortest closed path through p in its homotopy class. Let $\mathcal{G}(\beta, \gamma)$ be the set of minimum length simple closed curves homotopic to the boundaries of $A(\beta, \gamma)$, then $\mathcal{G}(\beta, \gamma) \subseteq \mathcal{S}(\beta, \gamma)$.

We now introduce the concept of a retraction of a homotopy, which gives a way to shortcut a homotopy at a given curve, provided it is a curve of $\mathcal{S}(\beta, \gamma)$. This idea is implicit in Chambers and Rotman [37, Proof of Theorem 0.7], and we refer to their article for more details. For a monotone isotopy h between boundaries of an annulus A, and a homotopic annulus $A' \subset A$, define the *retraction* $h|^{A'}(t)$ of $h(t)$ to A' as the same curve with each arc of $h(t) \setminus A'$ replaced by the shortest homotopic path along the boundary of A'. Although paths along $\partial A'$ (dis)appear discontinuously as t varies, $h|^{A'}$ can be obtained in the form of a discrete homotopy by (un)spiking these paths as they (dis)appear. The resulting homotopy $h|^{A'}$ is a monotone isotopy.

Lemma 4.1. *If $\alpha \in \mathcal{S}(\alpha, \gamma)$ and $A(\alpha, \gamma) \subseteq A(\beta, \gamma)$, and h is a monotone isotopy from β to γ of height L, then $h|^{A(\beta,\alpha)}$ is a monotone isotopy from β to α with height at most L.*

Proof. The retraction $h' = h|^{A(\beta,\alpha)}$ is a monotone isotopy from $h'(0) = \beta$ to $h'(1) = \alpha$. Let t' be the maximum t for which $h(t)$ intersects $A(\beta, \alpha)$. For $t \geq t'$, we have $h'(t) = \alpha$ and therefore $|h'(t)| = |\alpha| \leq |h(t')| \leq L$. For $t \leq t'$, each arc a of $h(t) \setminus A(\beta, \alpha)$ is replaced in $h'(t)$ by a homotopic path b along α with $|b| \leq |a|$, and thus $|h'(t)| \leq |h(t)| \leq L$. Hence $\text{height}(h') \leq L$. ☐

Lemma 4.2. *If $\alpha \in \mathcal{S}(\beta, \gamma)$, and h is a monotone isotopy from β to γ of height L, then there is a monotone isotopy from β to γ of height at most L having α as a level curve.*

Proof. We have $\alpha \in \mathcal{S}(\alpha, \beta)$ and $\alpha \in \mathcal{S}(\alpha, \gamma)$. So by Lemma 4.1, the monotone isotopies $h|^{A(\beta,\alpha)}$ from β to α and $h|^{A(\alpha,\gamma)}$ from α to γ have height at most L and can be composed to obtain a monotone isotopy from β to γ of height at most L with α as a level curve. ☐

Lemma 4.3. *Let $\Pi = \{\pi_1, \dots, \pi_m\}$ be a set of paths from γ_0 to γ_1 without proper pairwise intersections, where each π_i is a shortest homotopic path in $A(\gamma_0, \gamma_1)$ between its endpoints. If h is a monotone isotopy from γ_0 to γ_1 of height L, then there exists a monotone isotopy of height at most L whose level curves all cross each π_i at most once (after infinitesimal perturbations).*

Proof. Denote by $c(a, b)$ the number of proper intersections of curves a and b, and by $c_\Pi(a) = \sum_{\pi \in \Pi} c(a, \pi)$ the total number of intersections of a with Π. Let $C_h = \max_t c_\Pi(h(t))$ be the maximum total number of intersections over all t, and let I_h

be the set of maximal intervals (τ_0, τ_1) with $c_\Pi(h(t)) = C_h$ if $t \in (\tau_0, \tau_1) \in I_h$. If $C_h = m$, each level curve of h crosses each π_i exactly once and we are done, thus we assume in the following that $c_\Pi(h(0)) = c_\Pi(h(1)) = m < C_h$.

If $C_h > m$, we obtain a homotopy h' from h with $C_{h'} < C_h$ by, for each interval $(\tau_0, \tau_1) \in I_h$, replacing subhomotopy $h|_{(\tau_0,\tau_1)}$ of h by some $h^* = h'|_{(\tau_0,\tau_1)}$ with $C_{h^*} < C_h$.

Consider a single interval $(\tau_0, \tau_1) \in I_h$ and let $A = A(h(\tau_0), h(\tau_1))$. Then $\Pi \cap A$ consists of C_h subarcs of Π, each connecting the two boundaries of A. For $t \in (\tau_0, \tau_1)$, $h(t)$ intersects each such arc exactly once, and each $h(t)$ intersects these arcs in the same order. Among the components of $A \setminus \Pi$, there is a disk D_0 bounded by one arc of $h(\tau_0)$ and two arcs of $\pi_i \cap A$, and a disk D_1 bounded by one arc of $h(\tau_1)$ and one arc of π_j, such that these disks contain no other arcs of Π. We can find $\alpha \in \mathcal{G}(h(\tau_0), h(\tau_1))$ that intersects any arc of $A \cap \Pi$ at most once (in the same order as $h(t)$), and does not intersect the interiors of D_0 and D_1 (because the two arcs of Π on their boundary form a shortest path). Then $c_\Pi(\alpha) < C_h$ and the retraction $h_0 = h|^{A(h(\tau_0),\alpha)}$ has $C_{h_0} < C_h$, since any arc $h_0(t)$ has fewer intersections than $h(t)$ has with Π (in particular with the boundary of D_1). Symmetrically, for $h_1 = h|^{A(\alpha,h(\tau_1))}$ we have $C_{h_1} < C_h$. Since the composition $h^* = h_0 h_1$ is a homotopy from $h(\tau_0)$ to $h(\tau_1)$ with $C_{h^*} < C_h$ and height at most L (by Lemma 4.2), we can use this as a replacement for $h|_{(\tau_0,\tau_1)}$ in h'. By induction, we obtain a homotopy of height at most L whose level curves all cross each π_i at most once. ☐

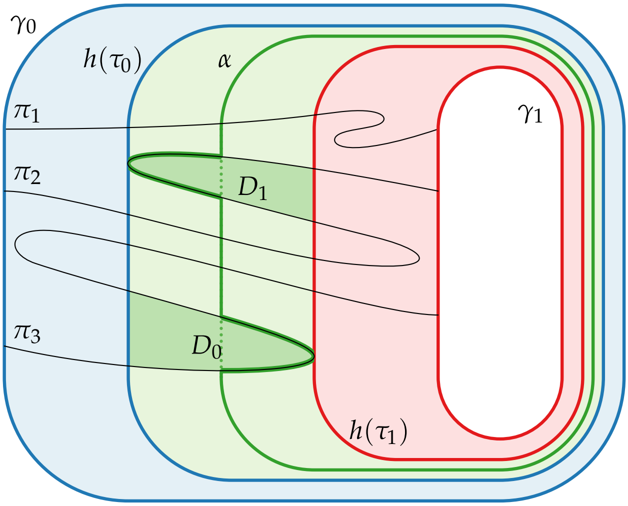


Figure 4.6: Choosing α such that $C_{h^*} < C_h$.

4.4 Computing homotopy height in NP

In this section, we show that in the discrete setting, there is an optimal homotopy with a polynomial number of moves. First, we show that there is a homotopy that flips each face exactly once.

Lemma 4.4. *For an annulus (Σ, G) bounded by γ_0 and γ_1, there is a homotopy of minimum height between γ_0 and γ_1 that flips each face of G exactly once.*

Proof. By Theorem 4.4, some homotopy h of minimum height is a monotone isotopy. For two consecutive level curves $h(t)$ and $h(t')$ in a monotone isotopy, the move between $h(t)$ and $h(t')$ flips face F if and only if F lies in $A(h(t'), \gamma_1)$ or $A(h(t), \gamma_1)$ but not both. Because $A(h(0), \gamma_1)$ contains all faces, and $A(h(1), \gamma_1)$ contains none, each face is flipped at least once. By monotonicity, we have for $0 \leq t' < t \leq 1$, that $A(h(t'), \gamma_1) \supseteq A(h(t), \gamma_1)$. So, if face F does not lie in $A(h(t), \gamma_1)$, it will not be flipped again in $h|_{(t,1]}$. Hence each face is flipped exactly once. □

It remains to show that each edge is involved in a polynomial number of spike and unspike moves; note that this does not directly follow from monotonicity, since a second spike of the same edge does not violate monotonicity (as can easily be seen in the dual setting).

Postponing spikes. Before we bound the number of spike moves, we transform an optimal monotone isotopy h into one where each spike move is delayed as much as possible, and each unspike move happens as soon as possible. We explain this transformation in the dual setting.

Suppose a spike move occurs for edge e between $h(t_i)$ and $h(t_{i+1})$, then denote by s the (unique) arc of $A(h(t_i), h(t_{i+1})) \cap G^*$ both of whose endpoints lie on $h(t_{i+1})$. This arc is a subarc of the dual edge e^*. Consider the maximum $j > i$, for which the component s_j of $e^* \cap A(\gamma_0, h(t_j))$ containing s has both endpoints on $h(t_j)$, and for all $t_i < t \leq t_j$, curve $h(t)$ has exactly two crossings with s_j (so the only action performed on arc s_j was the spike between $h(t_i)$ and $h(t_{i+1})$). Then s_j and $h(t_j)$ enclose a disk D_j. If the interior of D_j contains no edges of G^*, we can delay the spike of e at least until just before t_j, as illustrated in Figure 4.7 (a), where D_j is shaded.

Depending on what happens in the move between $h(t_j)$ and $h(t_{j+1})$, we may transform the isotopy further. This move is either (1) an unspike attached to s_j, or (2) a face-flip connected to one endpoint or (3) both endpoints[2] of s_j, or (4) a face-flip or spike inside D_{j+1}. In cases (1) and (2), we cancel the spike against the unspike or flip, as illustrated in Figure 4.7 (b) and (c). We do not postpone the spike in cases (3) and (4). Symmetrically, unspike moves can be made to happen earlier. Observe that these operations cannot increase the height of a homotopy since each level curve in the resulting homotopy crosses a subset of the edges of some curve in the original homotopy.

Call a homotopy *reduced* if it is the result of applying the above rules to h until no spike can be canceled or postponed until after a flip or unspike, and no unspike

---

[2] This happens only if the primal edge is adjacent to only one face of G.

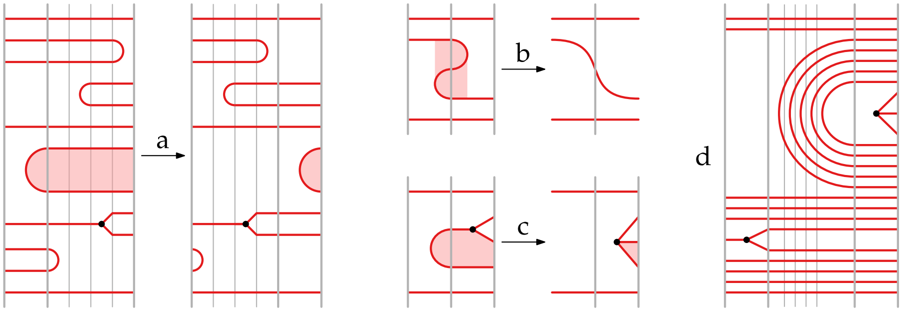


Figure 4.7: Delaying spikes (a). Canceling spikes against unspikes (b) or faces (c). Part of a reduced isotopy (d).

can be canceled or be made to happen before any prior flip or spike. Observe that starting from an optimal monotone isotopy, the reduced isotopy is also an optimal monotone isotopy. The structure of reduced homotopies is given in Lemma 4.5.

Lemma 4.5. *Between any two consecutive face-flips in a reduced isotopy lies a single (possibly empty) path of unspike moves followed by a (possibly empty) path of spiked moves.*

Proof. In a reduced homotopy, no unspike follows a spike move, and any spikes that remain "surround" the next face-flip (if any), see Figure 4.7 (d). Symmetrically, all unspikes between two consecutive face-flips surround the previous face-flip (if any). From the primal perspective, these unspike moves form a path from the previously flipped face and spike moves form a path towards the next flipped face. □

Any reduced homotopy starts with zero or more unspikes from γ_0, after which a possibly empty path of spikes to the first face-flip occurs, then that face is flipped, and a possibly empty path of unspikes enabled by this flip occurs. Subsequently, a spiked path, face-flip, and unspiked path occur for the remaining faces. Finally, a sequence of spikes towards γ_1 may occur. We may assume that on γ_0 and γ_1, any two consecutive edges are different, such that no immediate unspike moves are possible from γ_0, and no immediate spike moves are possible to γ_1. Otherwise we may by Lemma 4.1 perform those moves immediately without increasing the homotopy height.

Bounding spike moves. We are now ready to bound the number of spike and unspike moves in an optimal homotopy. Call a homotopy h *good* if it is a minimum-height reduced monotone isotopy and it has a minimum number of moves. By Theorems 4.2 and 4.4, the height of h is the homotopy height between γ_0 and γ_1.

Define an edge-spike of an edge e to be *between* existing copies of e, if the portion of the dual edge e^* crossed by the (dual) level curve, lies between two existing crossings of the level curve with e^*, such as in Figure 4.8. We show that such spikes never appear in h.

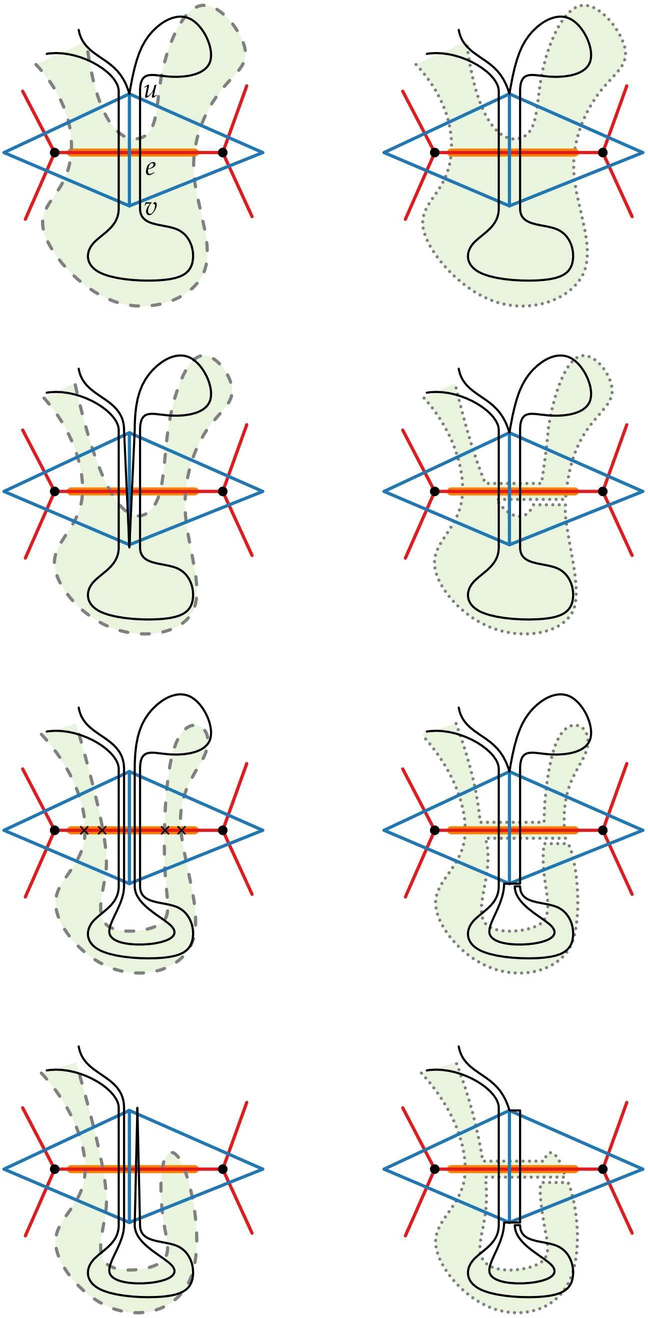


Figure 4.8: Left: (top to bottom) development of a spike between existing copies of e. Part of the graph in red (dual) and blue (primal) and the level curve in gray dashed (dual) and black (primal, perturbed). Right: a local surgery that avoids the spike between copies of e.

Lemma 4.6. *If homotopy h is good, there are no spikes between existing copies of any edge e.*

Proof. Suppose the move from $h(t_i)$ to $h(t_{i+1})$ is the last move between existing copies of the same edge, and assume this move is a spike of edge $e = (u, v)$ from u to v. In the dual setting, consider the component π of $e^* \cap A(h(t_i), \gamma_1)$ that is crossed by the spike move. This component is highlighted in Figure 4.8, which illustrates how a developing spike (in the left column) is removed (in the right column). Let $c(t)$ be the number of crossings of $h(t)$ with π, then for some τ_0 between t_i and t_{i+1}, $c(\tau_0) = 3$, and for some unique $\tau_1 > \tau_0$, $c(\tau_1) = 3$ again, and for $\tau_0 < t < \tau_1$, we have $c(t) = 4$ (because we assumed this was the last spike between existing copies of any edge).

For $\tau_0 < t < \tau_1$, label the four crossings of $h(t)$ with π by $p_1(t)$, $p_2(t)$, $p_3(t)$, and $p_4(t)$, in order along e^*, so the spike move at τ_0 creates p_2 and p_3. Consider the three components $C_1(t)$, $C_2(t)$ and $C_3(t)$ of $A(h(t), \gamma_1) \setminus \pi$, such that C_1 touches p_1 and p_2 from the dual face of u, and C_2 touches p_3 and p_4 from the dual face of u, and C_3 touches e^* in two segments from the dual face of v. Because C_3 lies between C_1 and C_2, h will first contract either component C_1 or C_2, namely at $h(\tau_1)$. Assume without loss of generality that C_2 contracts first.

We modify $h|_{[\tau_0, \tau_1]}$ such that any level curve crosses π at most twice by reconnecting the neighborhood of π, whose local structure evolves exactly as depicted in the top row of Figure 4.9. We essentially remove crossings p_2 and p_3, and reconnect $\partial C_1(t) \cap h(t)$ with $\partial C_2(t) \cap h(t)$ using a (zero-length) segment along π in face u^*. On the other side, consider the arc of $\partial C_3(t) \cap h(t) \cap v^*$ with $p_4(t)$ as endpoint. We cut this arc in two subarcs a and b, where a has $p_4(t)$ as endpoint, and connect the other endpoint to the arc of $\partial C_3(t) \cap h(t)$ at the endpoint at $p_2(t)$ using a segment along π in v^*. Similarly, we connect the endpoint of that at $p_3(t)$ to the loose end of b. These reconnections are depicted in the bottom row of Figure 4.9.

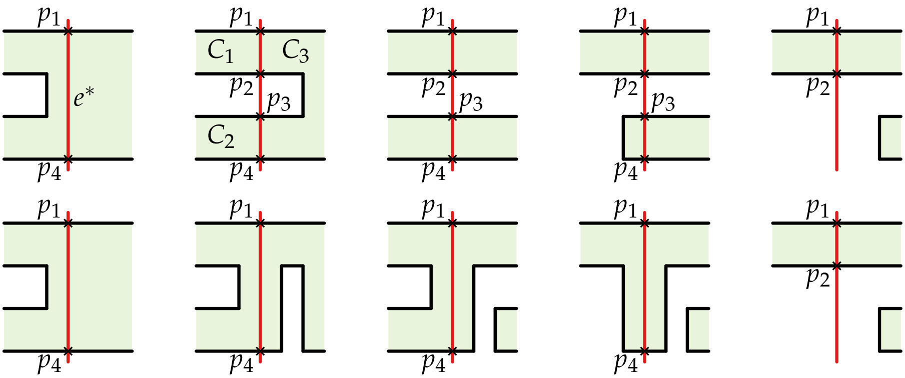


Figure 4.9: Top: the neighborhood of e^* throughout h. Bottom: the reconnected homotopy, reducing crossings with e^*. From left to right: the homotopy just before τ_0, just after τ_0, between τ_0 and τ_1, just before τ_1, and just after τ_1.

Observe that the reconnected curves can be made to appear continuously in such a way that they form a monotone isotopy. Because level curves only changed in the neighborhood of π, where they were shortened by avoiding the crossings with π, we have an isotopy whose height is at most that of h, and in which at least one spike is removed. So, because h was optimal, we have constructed an optimal monotone isotopy with fewer moves. Therefore, the corresponding reduced isotopy also has fewer moves, contradicting that h was good. $\square$

Our final step towards bounding the number of edge spikes is to derive a contradiction if for some interval $[\tau_0, \tau_1]$ without face-flips, an edge e is spiked (or unspiked) 5 times in $h|_{[\tau_0,\tau_1]}$. The proof is similar to that of Lemma 4.6.

Lemma 4.7. *For a good homotopy h, any subhomotopy $h|_{[\tau_0,\tau_1]}$ contains either a face-flip, or at most 4 spike (and at most 4 unspike) moves of the same edge.*

Proof. Suppose $h|_{[\tau_0,\tau_1]}$ contains no face-flip, then because h is reduced, the spike moves in $h|_{[\tau_0,\tau_1]}$ form a path σ of spike moves in G. Assume for a contradiction that some edge $e = (u, v)$ lies on σ at least 5 times. We say two spikes s_1 and s_2 are consecutive on e^* if no spike occurs on the arc of e^* between the first crossing of s_1 with e^* and the first crossing of s_2 with e^*.

Because by Lemma 4.6, h does not contain spikes between existing copies of edges, we can find three spikes s_1, s_2 and s_3 of e on σ where s_1 and s_2 as well as s_2 and s_3 are consecutive on e^*, and s_1 happens before s_2 and s_2 happens before s_3. Let σ_0, σ_1, σ_2 and σ_3 be the subpaths of σ such that $\sigma = \sigma_0 s_1 \sigma_1 s_2 \sigma_2 s_3 \sigma_3$, also labeled in Figure 4.10.

To get rid of spike s_2, we connect $\sigma_0 s_1 \sigma_1$ to $\sigma_2 s_3 \sigma_3$ in an alternative way. Figure 4.11 illustrates all possible ways s_1, s_2 and s_3 (in the dotted area) can be connected by σ, and how our method will reconnect σ without s_2. Formally, to decide where this reconnection takes place, we consider the components of $A(h(\tau_1), \gamma_1) \setminus \pi$, where π is the arc of e^* between its intersections with s_1 and s_3. There are three components, component C_1 touching π and σ_1, component C_2 touching π and σ_2, and component C_3 touching σ entirely, and touching π in two arcs. The component that h contracts first is either C_1 or C_2 (since C_3 lies between the other two).

First consider the case where C_1 is contracted first, then the path $\sigma_2 s_3 \sigma_3$ starts in the dual face of the endpoint of s_2. Note that there is a (zero-length) path between the start or endpoint of s_1 and the endpoint of s_2 because s_1 and s_2 are adjacent along e^*. Use this zero-length path to connect $\sigma_2 s_3 \sigma_3$ to $\sigma_0 s_1 \sigma_1$ at the start or endpoint of s_1 and call the resulting tree λ.

We claim we obtain an optimal monotone isotopy h' from h by replacing the spiked path σ by the spiked tree λ, and removing the unspike move of e^* following the contraction of C_1. Up until the creation of λ, the move sequence is the same as in h. Since λ contains a subset (all spikes except s_2) of the spikes of σ, the spiked tree can be created without surpassing the height of h. After the creation of σ in h and λ in h', locally, the level curves of h and h' differ only in a small neighborhood of π, so that all moves of h except those crossing π can still be performed in h'. Because s_2 is the only spike along e^* that lies between s_1 and s_3, the next move that crosses π is the unspike move, call it z, following the contraction of C_1. The level curve of h' just before z is identical to the level curve of h just after z, so it is

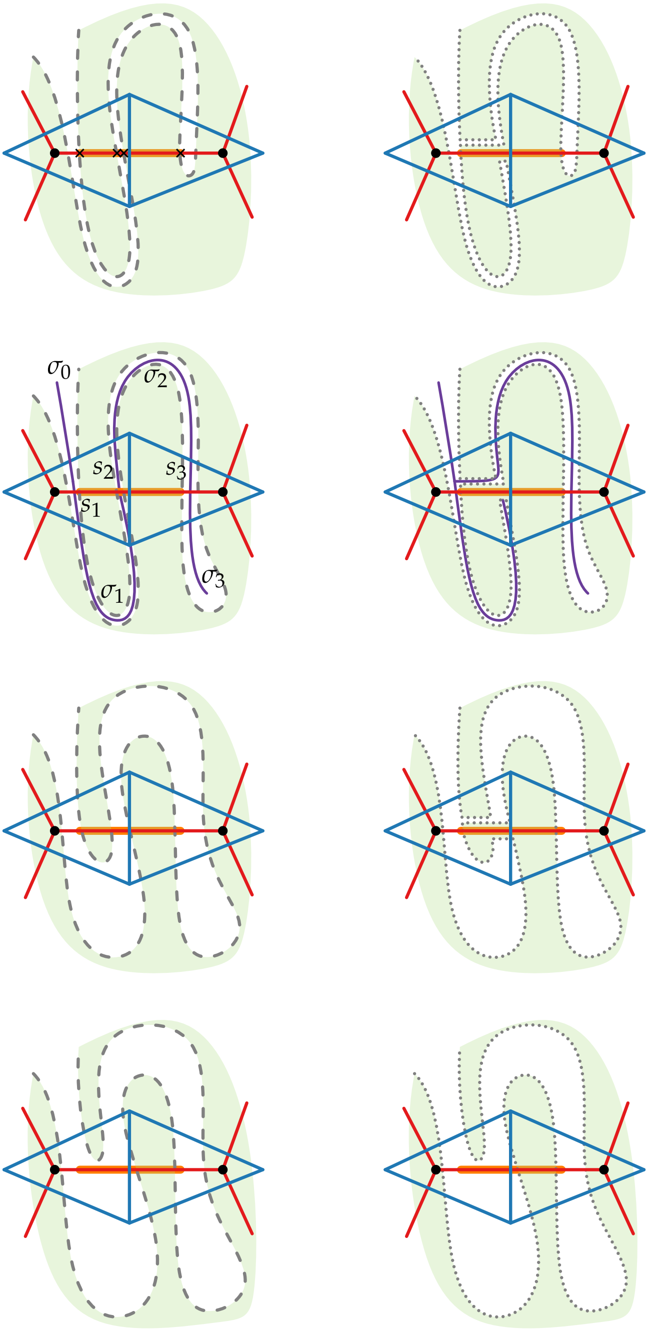

σ_0
σ_2
s_2
s_3
s_1
σ_1
σ_3

Figure 4.10: Left: A spiked path revisiting the same edge many times. Right: A local surgery to avoid five spikes of the same edge on a single spiked path.

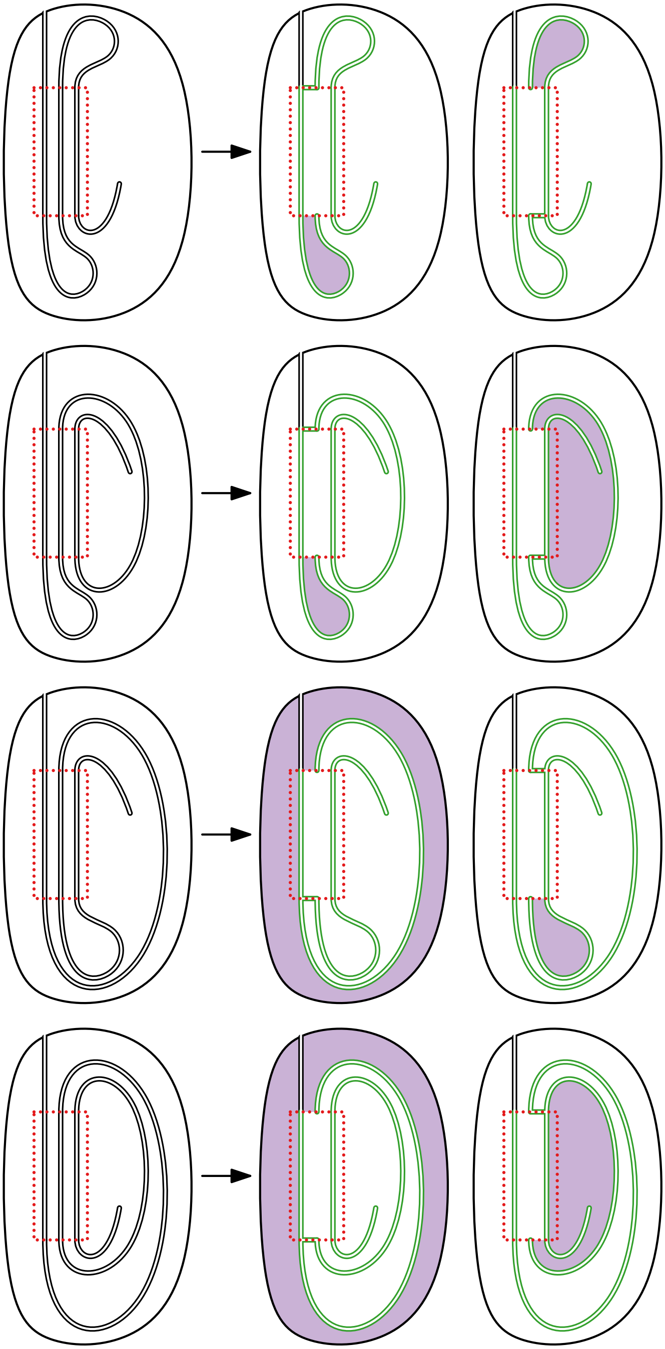


Figure 4.11: Cases for shortcutting spiked paths visiting the same edge often. The neighborhood of the repeated edge is dotted and the component contracted first is shaded.

safe to omit move z in h'. All subsequent level curves of h and h' are identical, so we conclude that h' is an optimal monotone isotopy (with fewer moves). Therefore, the reduced monotone isotopy of h' has fewer moves, contradicting that h is good.

The proof for the case where C_2 contracts first, is symmetrical, except that the spiked tree λ is created differently. In this case, we define λ to be $\sigma_0 s_1 \sigma_1$, whose endpoint is connected to $\sigma_2 s_3 \sigma_3$ at the start or endpoint of s_3. When spiking this tree, the direction of the spikes on σ_2 (and sometimes σ_3) is reversed, but this does not affect the proof.

Thus, in a good homotopy, no spiked path spikes the same edge five times. □

Theorem 4.5. *For γ_0 and γ_1 bounding an annulus with n faces and m edges, there is a homotopy of minimum height that has at most $O(mn)$ moves. Therefore, deciding whether their* Homotopy Height *is at most L is in NP.*

Proof. Let n be the number of faces, and m the number of edges in G. As a direct consequence of Lemmas 4.4 and 4.7, there is a good homotopy that spikes each edge at most $4(n+1)$ times and unspikes each edge at most $4(n+1)$ times. So there is a homotopy of minimum height that has at most $8m(n+1)+n = O(mn)$ moves. Testing whether this homotopy indeed has height at most L can be done by computing the maximum length over its (polynomially many) level curves, each containing a polynomial number of edges, and comparing this maximum with L. Given a good homotopy, all of this can be done in polynomial time assuming addition and comparisons of numbers takes polynomial time. □

We note that the Homotopy Height problem can also be defined in slightly different settings, for example

- γ_0 and γ_1 are two paths with common endpoints s and t, such that $\gamma_0 \cup \gamma_1$ is the boundary of a combinatorial disk. Then γ_0 is homotopic to γ_1 with fixed endpoints, and we are interested in computing the optimal height of this homotopy. This is the Homotopy Height problem considered by E. Chambers and Letscher [31].
- There is a single curve γ forming the boundary of a combinatorial disk. This curve is contractible, and we are interested in computing the optimal height of such a contraction. This is one of the settings considered in [27].

In both these cases, the Theorems 4.2 and 4.4 have analogues establishing that some optimal homotopy is an isotopy and is monotone. The rest of our proof techniques then readily apply, and prove that the Homotopy Height problem in these cases is also in NP. The next section investigates more distant variants.

4.5 Variants and approximation algorithms

4.5.1 Homotopic Fréchet distance

There is a strong connection between the problem of Homotopy Height and the problem of Homotopic Fréchet distance, which we now recall. As in [66], our setting is the one of a disk D with four points p_0, q_0, q_1 and p_1 on the boundary, connected by four disjoint boundary arcs γ_0, γ_1, P and Q, with γ_0 from p_0 to q_0; γ_1 from p_1 to q_1; P from p_0 to p_1; and Q from q_0 to q_1, see Figure 4.12, left. A homotopy between γ_0 and γ_1 is a series of elementary moves connecting curves of D with one endpoint on P and the other on Q, where the collection of curves starts at γ_0 and ends at γ_1. The HOMOTOPIC FRÉCHET DISTANCE between P and Q is the height of a homotopy between γ_0 and γ_1 of minimal height. The common intuition for this distance is that it is the minimal length of a leash needed for a man on P to walk his dog along Q, where the leash may stretch but cannot be lifted out of the underlying space.

We note that this is slightly different than the original setting for homotopic Fréchet distance in the original work [29], where an exact algorithm is presented for the plane minus a set of polygonal obstacles. In the original work, the start and end leashes are not fixed, and in fact the bulk of the work is in determining an optimal relative homotopy class in order to find the best homotopy.

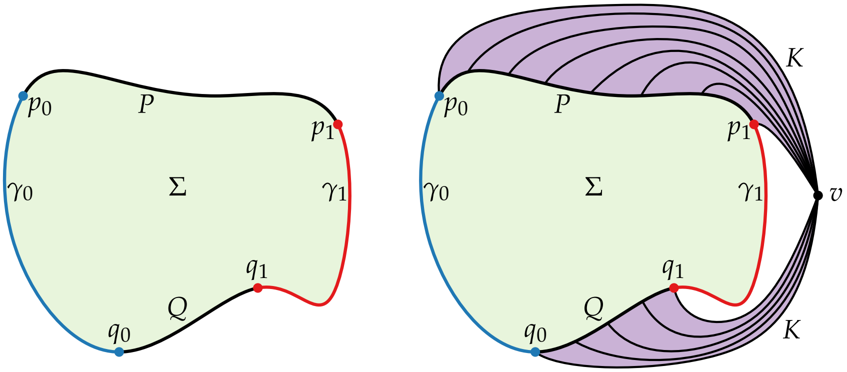


Figure 4.12: The setting of homotopic Fréchet distance.

Proposition 4.6. *The* HOMOTOPIC FRÉCHET DISTANCE *problem is in NP.*

Proof. We reduce HOMOTOPIC FRÉCHET DISTANCE to HOMOTOPY HEIGHT using the following construction. We add a vertex v and edges of weight K between this vertex and all the vertices of the paths P and Q, where K is a constant greater than the sum of the weights of the edges of the disk, as well as all the intermediate triangles, see Figure 4.12, right. This results in a pinched annulus A, with two boundaries γ'_0 and γ'_1 obtained from the paths γ_0 and γ_1, both completed into closed curves using the additional vertex v. We claim that an optimal homotopy between γ_0 and γ_1 translates into an optimal homotopy in A between γ'_0 and γ'_1, and vice-versa. Indeed, by Lemma 4.3, there exists an optimal homotopy in A such that any intermediate curve crosses the shortest path between γ'_0 and γ'_1 exactly once, and in our case the shortest path is the zero length path starting and ending at the vertex v. Furthermore, if the weight K is big enough, the level curves of an optimal homotopy between γ'_0 and γ'_1 will always use exactly two of the edges of weight K, since two are needed but any more would be too expensive. Thus, an optimal homotopy between γ_0' and γ_1' translates directly into an optimal homotopy between γ_0 and γ_1 after cutting on v and removing the edges linked to v and vice-versa. The homotopy height is increased by exactly $2K$ in this translation. □

Har-Peled et al. [66] provide an algorithm to compute in $O(n \log n)$ time a homotopy of height $O(d \log n)$, where d is a lower bound on the height of an optimal homotopy, and n is the complexity of Σ. In particular, one can set d to be the maximum of $\|\gamma_0\|$, $\|\gamma_1\|$, the diameter of Σ, and half of the total weight of the boundary of any face. This yields an $O(\log n)$-approximation for HOMOTOPIC FRÉCHET DISTANCE[3]. We show here that their algorithm can be adapted to yield an $O(\log n)$-approximation for HOMOTOPY HEIGHT.

Proposition 4.7. *One can compute in $O(n \log n)$ time an $O(\log n)$-approximation of HOMOTOPY HEIGHT.*

Proof. Starting with an annulus and two boundary curves γ_0 and γ_1, we first compute a shortest path $\mathcal{P}$ between the boundary curves γ_0 and γ_1 and cut along $\mathcal{P}$ to obtain a disk D. This brings us to the setting of HOMOTOPIC FRÉCHET DISTANCE, and we can apply the aforementioned algorithm and obtain a homotopy h. In order to recover a homotopy between γ_0 and γ_1, we glue back the disk along $\mathcal{P}$ into an annulus, and the level curves of h are completed into closed curves by using subpaths of $\mathcal{P}$, this gives us a homotopy h'. It remains to show that this is an $O(\log n)$-approximation of the optimal homotopy. By Lemma 4.3, some optimal homotopy between γ_0 and γ_1 has level curves cutting $\mathcal{P}$ exactly once. Thus, the height L of an optimal homotopy in the disk D is a lower bound for the height of an optimal homotopy in the annulus A. Furthermore, each level curve γ_t of h' consist of two subpaths, one being a level curve $h(t)$ of h and the other being a subpath $\mathcal{P}_t'$ of $\mathcal{P}$. Since $\mathcal{P}$ is a shortest path, $\mathcal{P}_t'$ is also a shortest path between its endpoints, so it is shorter than $h(t)$ since they have the same endpoints. By construction, the length of $h(t)$ is $O(L \log n)$, and thus the length of γ_t is $O(2L \log n) = O(L \log n)$. This concludes the proof. □

4.5.2 Minimum height linear layouts

We also show that a seemingly unrelated graph drawing problem is directly equivalent to the HOMOTOPY HEIGHT problem. A *linear layout* is an embedding of a planar graph where the edges have isolated tangencies with the vertical line, and all the vertices have distinct x coordinates. The MINIMUM HEIGHT LINEAR LAYOUT problem is the following one: Given a planar embedding of an edge-weighted graph G, find a homeomorphic linear layout of G in $\mathbb{R}^2$ such that the maximal weight of the vertical lines is minimized. Here, the weight of a vertical line is the sum of the weights of the edges that it crosses, and (similarly to the cross-metric setting), vertical lines crossing tangent to the edges or crossing vertices are not counted.

---

[3] This algorithm assumes triangular faces, but using our definition of d, one can extend the algorithm of [66] to also work with polygonal faces.

Theorem 4.8. *The* MINIMUM HEIGHT LINEAR LAYOUT *problem is equivalent to the* HOMOTOPY HEIGHT *problem.*

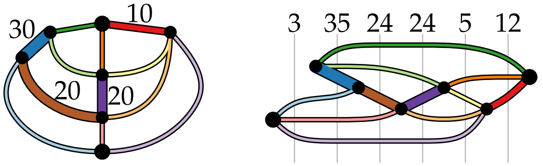


Figure 4.13: Dual representation of Figure 4.2 (left) and Figure 4.3 (right).

Proof. Indeed, a linear layout of a planar graph G naturally induces a discrete homotopy sweeping its dual graph G^*. More formally, we drill a small hole around the vertex dual to the outer face of G, and we view its complement as a disk D which is a cross-metric surface for the graph G. Since the hole was drilled in the middle of the face of G, its boundary has zero length. We pick two arbitrary vertices s and t on it, which cuts the boundary into two paths L and R. Then we claim that a minimum height linear layout of G is equivalent to a homotopy of minimum height between L and R (where the endpoints are fixed)[4]. Indeed, whenever the sweep of $\mathbb{R}^2$ induced by the vertical lines crosses an edge or passes a vertex, by the dual interpretation of homotopies with cross-metric surfaces outlined in the preliminaries, it amounts to doing a face or an edge move, and thus the whole vertical sweep defines a homotopy between the two paths L and R. Furthermore, this homotopy is an isotopy, since the vertical lines are simple, and a monotone one since they only go in a single direction. Conversely, a discrete homotopy of optimal height between L and R can be "straightened" into a linear layout: by Theorem 4.4, one can assume such a homotopy h to be an isotopy and to be monotone, and therefore the succession of dual moves of h with respect to G are homeomorphic to a sweep of G by vertical lines, as pictured in Figure 4.13. An optimal homotopy amounts, via this homeomorphism, to a minimum height linear layout. □

In particular, the MINIMUM HEIGHT LINEAR LAYOUT problem is in NP and admits an $O(\log n)$-approximation algorithm.

4.6 Discussion

Using the recent results on the monotonicity of optimal homotopies, we have shown that computing HOMOTOPY HEIGHT is in NP, which was an open problem posed by Har-Peled et al. [66]. The equivalent problem of MINIMUM HEIGHT LINEAR LAYOUT and the related problem of HOMOTOPIC FRÉCHET DISTANCE are also in NP. For the HOMOTOPIC FRÉCHET DISTANCE an efficient factor $O(\log n)$-approximation algorithm was known, and it turns out that this algorithm also provides a factor $O(\log n)$-approximation to the Homotopy Height and Minimum Height Linear Layout problems.

---

[4] The point of the somewhat artificial construction with L and R is to force the homotopy to go through the outer face of G at all times.

An interesting open question is whether any of these problems admit polynomial time constant-factor approximation algorithms. In fact, it remains open whether these problems are hard to compute exactly, despite the fact that related problems such as cut-width are known to be NP-hard.

5

Optimal Homotopies over a Spiked Polygonal Plane

In this chapter we study a problem similar to that of Chapter 4. The problem studied in this chapter is simpler in the sense that the surface over which our curves move is mostly flat. In particular, our surface no longer has a cellular representation, and consequently, the homotopies considered in this chapter require the curves move continuously rather than in discrete steps.

Our input is a polygonal domain with point obstacles and two curves on the boundary of the domain. Throughout the homotopy, the intermediate curves are allowed to pass over the point obstacles, albeit for a fixed cost for each obstacle. This is in contrast to [29], where the goal is to find a homotopy from one polyline to another, such that the intermediate curves of the homotopy avoid the point obstacles. In this chapter, we hence refer to the point obstacles as *spikes* to reinforce the intuition that encountering them is costly, but that they do not form impassable barriers. Our goal is to minimize the cost of the most expensive intermediate curve (also called a leash) in a homotopy between two polylines that are part of the boundary of our polygon. The cost of a leash is defined as its length plus the cost of the spikes it encounters. We consider both variable-cost spikes and unit-cost spikes, and describe several structural results as well as algorithms. In the following we first introduce some necessary definitions which allow us to state our our results more precisely.

Definitions and problem statement. Consider four polylines $f, g, \gamma_0, \gamma_1 \colon [0,1] \to \mathbb{R}^2$ (possibly of length 0) in the plane that are interior-disjoint, with $f(0) = \gamma_0(0)$, $f(1) = \gamma_1(0)$, $g(0) = \gamma_0(1)$, and $g(1) = \gamma_1(1)$, and whose union bounds a simple polygonal domain R of n vertices. Let $K \subset R$ be a finite set of k point-obstacles which we call *spikes* (see Figure 5.1). The *cost* of a spike is given by a function $w \colon K \to \mathbb{R}^{\geq 0}$. We assume that spikes lie in general position in the sense that no three spikes lie on the same geodesic in R. Standard perturbation techniques (as also described below) can lift this assumption.

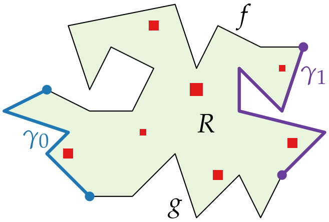


Figure 5.1: Polygonal domain R bounded by polylines $f \cup g \cup \gamma_0 \cup \gamma_1$. K in red squares.

Let Γ be the family of simple curves in R from f to g; i.e. all curves $\gamma\colon [0,1] \to R$ with $\gamma(0) \in \mathrm{Im}(f)$ and $\gamma(1) \in \mathrm{Im}(g)$. A *homotopy* on R from γ_0 to γ_1 is a continuous map $h\colon [0,1] \times [0,1] \to R$ with $h(0,\cdot) = \gamma_0$, $h(1,\cdot) = \gamma_1$. All homotopies we consider have $h(t,\cdot) \in \Gamma$ for all $t \in [0,1]$. With slight abuse of terminology, we refer to these simply as homotopies. We call a homotopy *monotone* if it is injective after infinitesimal perturbation. We refer to each curve $\gamma_t\colon p \mapsto h(t,p)$ as the *leash* of h at time t, and define the *cost* of a leash as its length plus the total cost of spikes in K (with multiplicity) it encounters. The *cost* of a homotopy is the cost of its maximum-cost leash. We are interested in the minimum-cost homotopy on R from γ_0 to γ_1 and refer to the cost of this homotopy as the *homotopy height* from γ_0 to γ_1 on (R, K, w).

For two curves γ and γ' from a monotone homotopy, denote the region between them by $R(\gamma,\gamma')$; that is, $R(\gamma,\gamma')$ is the (possibly degenerate) topological disk bounded by γ, γ', the arc of f between $\gamma(0)$ and $\gamma'(0)$, and the arc of g between $\gamma(1)$ and $\gamma'(1)$. For a simple curve $\gamma \in \Gamma$, define its *swept region* as $R(\gamma_0,\gamma)$ and symmetrically define its *unswept region* as $R(\gamma,\gamma_1)$. We call a region $D \subseteq R$ *convex relative to* R if for any two points in D, the shortest path in R connecting those points also lies in D [86]; for this definition we do not not charge the additional cost of any spikes. A curve $\gamma \in \Gamma$ is *forwards-convex* if it lies on the boundary of a region D convex relative to R, and D is contained in the swept region of γ. Symmetrically, γ is *backwards-convex* if it lies on the boundary of a region D convex relative to R, and D is contained in the unswept region of γ (see Figure 5.2). A curve may be both forwards- and backwards-convex; a shortest path is always both.

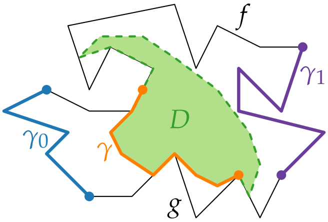


Figure 5.2: Disk D is convex relative to R. γ is backwards- but not forwards-convex.

Region R is convex relative to itself, and the intersection of any two convex sets relative to R is also convex relative to R. For $\gamma \in \Gamma$ define its convex hull (also known as geodesic hull) $\mathrm{gh}_R(\gamma)$ relative to R as the unique minimal region containing γ which is convex relative to R (see Figure 5.3).

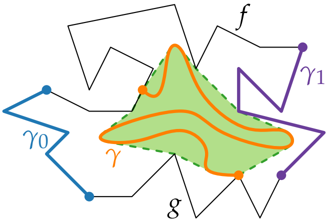


Figure 5.3: The convex hull $\mathrm{gh}_R(\gamma)$ of a curve $\gamma \in \Gamma$.

Consider a homotopy h that contains a leash γ_t that crosses several spikes. Infinitesimal perturbation of the leash at the spikes ensures that γ_t no longer crosses a spike, but forces the leash into some homotopy class. In particular, this has two implications: (1) as the perturbation tends to zero, this has no effect on the length of the leash and thus, strictly speaking, the optimal homotopy is an infimum rather than a minimum if the given leash is the maximum-cost one in the optimal homotopy; (2) we can decompose a single leash crossing a number of spikes into a homotopy crossing each spike separately, essentially holding the leash fixed. Hence an optimal homotopy exists that crosses spikes one by one.

Results and organization. We consider various settings of a spiked plane with polygonal boundary. In Section 5.1, we investigate the general setting where each spike may have a different cost. First, we consider the number of inflection points that the leash may need in an optimal homotopy and present a lower bound that is linear in the number of spikes, even if R is convex and only f has positive length. We then present a 2-approximation algorithm for the general case and an optimal algorithm for computing the homotopy height for the case that f and g have length 0 (i.e., γ_1 and γ_2 together form the boundary of R).

In Section 5.2, we consider the setting where all spikes have the same unit cost. Here we present our main result: an algorithm to compute the exact homotopy height in polynomial time. The algorithm combines structural properties of homotopies arising from the geometry with methodology for computing Fréchet distances. To the best of our knowledge, these are the first polynomial-time algorithms to compute the exact homotopy height in any setting.

Related work. As also used extensively in Chapter 4, we make use of our recent result [27] for the results of Section 5.2. The result—which readily transfers to our setting—states that there exists a minimum-cost homotopy between the boundaries of an annular surface that is an isotopy, and monotone: leashes of the homotopy never move backwards.

Homotopy height was introduced independently in the computational geometry community [31] and in the combinatorics community [13]. On triangulated surfaces, the best known algorithm gives an $O(\log n)$-approximation, where n is the complexity of the surface [66]. More recently, it has also been studied in more general settings, where instead of having point obstacles, obstacles are modeled by assigning a non-Euclidean metric to R [35, 36, 37], see also Chapter 4.

The classic algorithm for computing the Fréchet distance runs in $O(n^2 \log n)$ time [5], and many variants and approximation algorithms have been studied since. The geodesic and homotopic Fréchet distance are particularly related to our setting. The former is a variant where the leashes must stay inside a polygonal boundary and remain geodesic, but no obstacles are present inside the polygon [40]. For the latter, point or polygonal obstacles are given which no leash may cross: this case can be solved in polynomial time [29].

Optimal homotopies have also been studied in a variety of other settings. From the topology end, minimum-area homotopies can be computed for planar or surface embedded curves in polynomial time [34, 56, 79]. Minimum-area homologies are a closely related similarity measure on curves that can be computed very quickly using linear algebra packages [33, 47], but the resulting deformations are not as intuitive as homotopies.

In the computational geometry and graph drawing communities, much work has been done on computing morphs between inputs. In such morphs, the vertices of a graph tend to move while the edges remain straight. However, the present literature on morphing rarely incorporates obstacles. It is well known that any two drawings of the same planar graph can be morphed to each other. Optimal morphs between such graphs are still being studied, including work that bounds the complexity of the morph [6, 7]. However, none these morphs bound the length of any of the leashes tracing the paths of vertices during the morph.

Morphs based on geodesic width [53] force all intermediate curves to not cross the input curves (which are part of the boundary polygon in our setting). Dynamic time warping and related concepts [52] also consider ways to match and morph curves, but again do not naturally extend to more general settings.

5.1 Variable-cost spikes

In this section, we consider the setting where each spike may have a different cost. As we prove, the variable costs have a profound effect on the leash complexity. Nonetheless, we obtain a general approximation algorithm as well as an optimal algorithm for a special case.

Leash complexity. We define the *complexity* of a leash as the number of inflection points that are not caused by the boundary of R. In particular, the leash complexity needed for an optimal homotopy is defined by the number of spikes that cause an inflection point. Unfortunately, in the general case, the complexity may be linear, even when R is convex. Correspondingly, we do not have a polynomial-time algorithm even when R is convex.

Lemma 5.1. *In the worst case, the leash complexity for an optimal homotopy for* (R, K, w) *is* $\Omega(|K|)$*, even if* γ_0*,* γ_1 *and* g *have length* 0 *and* f *is convex.*

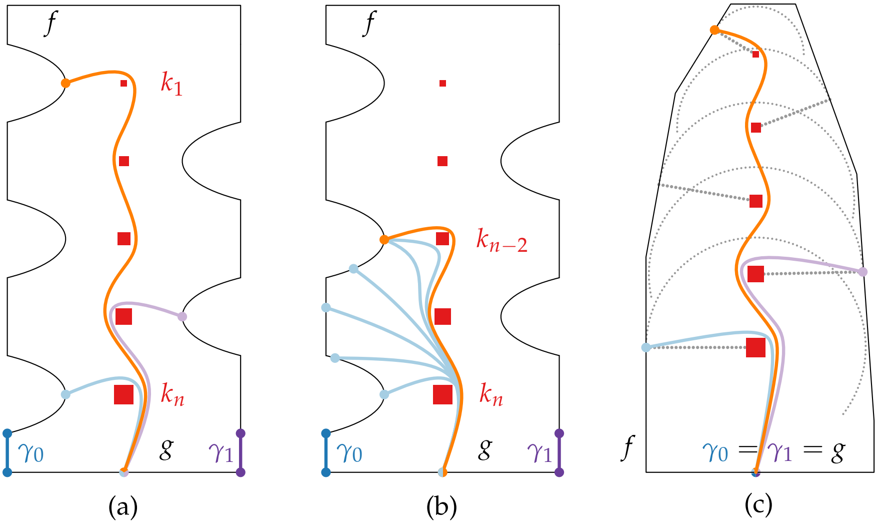


Figure 5.4: (a) A leash may need linear complexity when considering variable-cost spikes. (b) Part of the optimal homotopy, after crossing k_n up to crossing k_{n-2}. (c) This even holds in the convex case, with one boundary path and the two initial leashes having length 0. The closest point and corresponding distance circles are indicated for each spike.

Proof. For ease of exposition, we first argue the general case, using the construction illustrated in Figure 5.4(a). We have n spikes, $K = k_1, \ldots, k_n$, lined up in the middle, at vertical distance 1 from each other and f and g; k_1 is the highest- and k_n the lowest-positioned spike. Moreover, the odd-numbered spikes have a closest point on the first half of f and the even-numbered spikes have a closest point on the second half of f; these closest points are at distance 0.75. This implies that the optimal leash crossing k_i has cost $n - i + 1.75 + w(k_i)$. By setting $w(k_i) = c + i$, all these leashes have the same cost, for suitable constant $c > 0$, depending on the longest leash necessary that does not cross a spike. Now, the homotopy height of this instance is $1.75 + w(k_n) = 1.75 + c + n$. This requires the leashes to cross the spikes in the order of their closest points along f, that is, odd-numbered before even-numbered spikes (see Figure 5.4(b)). When crossing spike k_1, the leash has crossed all other odd-numbered spikes, but none of the even-numbered spikes. Thus, the leash has a linear number of inflection points, if we perturb all odd-numbered spikes slightly.

A similar construction can be made when we require that f is convex and γ_0, γ_1 and g have length 0. This is illustrated in Figure 5.4(c). The same principle applies: we position spikes such that their closest point is alternatingly on the left half and right half of f. By setting the weights appropriately, we can again ensure that the optimal homotopy must cross the spikes in some order along f and force a linear number of inflection points. □

Algorithms. If f and g collapse onto a point, we can compute the homotopy height in polynomial time with a greedy algorithm. Interestingly, this contrasts the potential complexity of the problem if γ_0 and γ_1 and even g collapse onto a point, as suggested by the lower bound in Lemma 5.1.

Lemma 5.2. *We can compute in polynomial time the homotopy height of (R, K, w), if f and g have length 0.*

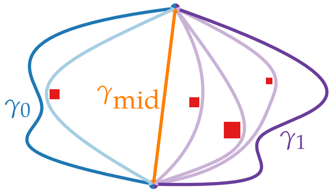


Figure 5.5: When f and g have length 0, a greedy strategy can be applied to shrink both γ_0 and γ_1 onto γ_{mid}.

Proof. Consider the geodesic leash γ_{mid} between $f(0)$ and $g(0)$, ignoring any spikes; see Figure 5.5. As described below, we greedily shrink γ_0 until we reach γ_{mid}. We first compute the geodesic leash γ_t between $f(0)$ and $g(0)$ in the same homotopy class as γ_0. By definition, γ_t cannot be longer than γ_0. Then, we cross the minimal-cost spike $k \in K \cap \gamma_t$, resulting in a cost $\|\gamma_t\| + w(k)$. We repeat the process from γ_t, until we reach γ_{mid}. Analogously, we shrink γ_1 to γ_{mid}. The maximum of $\|\gamma_0\|$, $\|\gamma_1\|$ and all intermediate $\|\gamma_t\| + w(k)$ is the homotopy height of (R, K, w).

As all intermediate leashes from γ_0 to γ_{mid} are backwards-convex, they never exceed the length of γ_0. In particular, this implies that we cannot make a leash on some spike k shorter, by first crossing other spikes that are not on the leash but in the unswept area. In other words, we cannot improve the cost by crossing a spike k on a geodesic by first crossing spikes that are not on the geodesic. Hence, the greedy choice is optimal. □

The above proof readily implies that the longest leash in the optimal homotopy is determined by the initial and final leash, and thus results in the following lemma. Note that we are interested here only in the geometric length, excluding any spikes. We capture this in a separate lemma as it supports the unit-cost case, detailed in the next section.

Lemma 5.3. *If $l \in \Gamma$ is backwards-convex and $r \in \Gamma$ is forwards-convex, with $l(0) = r(0)$, $l(1) = r(1)$, such that l and r together bound a region D convex relative to R, then there is a monotone homotopy from l to r consisting of only backwards- and forwards-convex leashes, and whose longest leash has length* $\max\{\|l\|, \|r\|\}$.

For the general case, there is also a simple 2-approximation achievable, by using the algorithm for the geodesic Fréchet distance.

Lemma 5.4. *We can compute in $O(|R|^2 \log^2 |R|)$ time a 2-approximation of the homotopy height of (R, K, w).*

Proof. The algorithm computes the geodesic Fréchet distance [40] in R, that is, ignoring the spikes. Consider the optimal geodesic Fréchet matching μ. We may extend μ into a homotopy h by infinitesimal perturbation to cross only one spike at once and by shortening γ_0 and γ_1 to the geodesics between $f(0)$ and $g(0)$ and between $f(1)$ and $g(1)$ respectively. We prove that h is a 2-approximation of the minimal-cost homotopy h^* including the spikes.

Let c be the longest leash in h and let $w_{\max}$ be the maximal cost of a spike. The cost of h is bounded by $c + w_{\max}$. Either c is defined by γ_0 or γ_1 (which must be in any homotopy) or c is defined by a leash in μ: in either case, c provides a lower bound on the maximal leash length in h'. Moreover, h' must also cross the maximal-cost spike. Hence, the cost of h' is at least $\max\{c, w_{\max}\}$. We now have that $c + w_{\max} \leq 2 \cdot \max\{c, w_{\max}\}$, thus h is a 2-approximation of h^*. ☐

5.2 Unit-cost spikes

In this section, we give an algorithm to compute the homotopy height in the case where all spikes have cost 1. We start by proving properties on the homotopy classes and lengths of curves in Γ. These properties allow us to construct for any homotopy, a homotopy of similar cost with a regular structure. Finally, we show how to decide the existence of such a regular homotopy cheaper than a given cost in polynomial time.

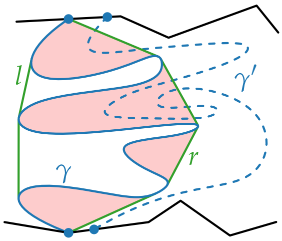
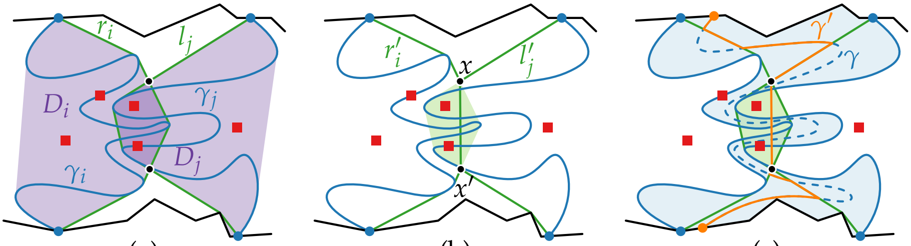


Figure 5.6: Pockets of γ with lids on r (shaded).

Shortcutting curves. Consider a curve $\gamma \in \Gamma$ and let $D = \mathrm{gh}(\gamma)$ be its convex hull. Let l and r respectively be the backwards- and forwards-convex arcs of ∂D between the endpoints of γ. Consider an arc φ of $\gamma \setminus l$ or $\gamma \setminus r$. Let $\bar{\varphi}$ be the corresponding arc of l or r between the endpoints of φ. Then we refer to the disk bounded by $\varphi \cup \bar{\varphi}$ as a *pocket* of γ, and refer to $\bar{\varphi}$ as its *lid*, see Figure 5.6. Each lid is a shortest path in R, and the pockets of γ partition D.



(a) (b) (c)

Figure 5.7: (a) The curves r_i and l_j on ∂D_i and ∂D_j, respectively. (b) The curves r'_i and l'_j obtained after replacing arcs with the geodesic between x and x'. (c) An example curve γ' in the homotopy class of r'_i that is shorter than γ (dashed).

Lemma 5.5. *Let γ and $\gamma' \in \Gamma$ be two non-crossing simple curves. Each arc ψ of $\gamma' \cap \mathrm{gh}(\gamma)$ has both endpoints on the same lid of the containing pocket of γ.*

Proof. Since the endpoints of γ' lie on the boundary of R, ψ must have both endpoints on the boundary of its pocket. Since ψ does not cross γ, and the pocket is bounded by an arc of γ and a lid, the endpoints of ψ lie on the lid. ☐

Consider a forwards- and a backwards-convex curve of Γ. If these curves intersect, they intersect in at most two points or geodesics, and occur in the same order and direction along the curves. As such, their first and last point of intersection are naturally well defined.

Lemma 5.6. *Let γ_i and $\gamma_j \in \Gamma$ be two non-intersecting simple curves, and assume γ_i lies in the swept region of γ_j. Let $D_i = \mathrm{gh}(\gamma_i)$ and $D_j = \mathrm{gh}(\gamma_j)$ be their convex hulls. Let $r_i \in \Gamma$ be the forwards-convex arc of ∂D_i and let $l_j \in \Gamma$ be the backwards-convex arc of ∂D_j. If r_i and l_j intersect, let x and x' be the first and last points of intersection of r_i and l_j. Let r'_i and l'_j be the curves obtained from r_i and l_j by replacing their arcs in $D_i \cap D_j$ by the geodesic between x and x'. If r_i and l_j do not intersect, let $r'_i = r_i$ and $l'_j = l_j$. Assume the region between γ_i and γ_j contains no spikes in its interior and consider a third simple curve γ in this region. There is a curve γ' with the same endpoints as γ and $\|\gamma'\| \leq \|\gamma\|$, such that γ' lies in the homotopy class of r'_i and l'_j.*

Proof. The setup is illustrated in Figure 5.7. We consider three cases illustrated in Figure 5.8, depending on the number of bends of r_i and l_j on $\partial(D_i \cap D_j)$ that are induced by γ_i and γ_j.

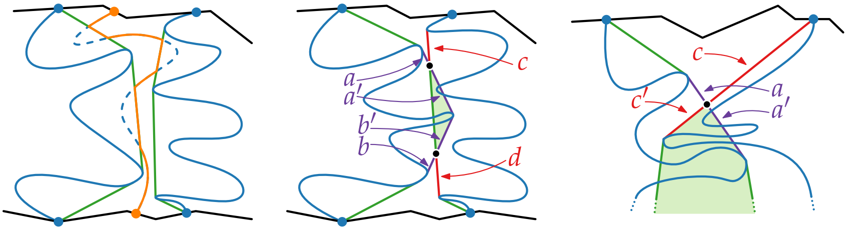
(a) In the first case, assume there are no such bends on r_i or l_j, then the interior of $D_i \cap D_j$ is empty. Since D_i and D_j are disjoint, so are the pockets of γ_i and γ_j. If we replace all arcs of γ that lie in pockets of γ_i or γ_j by the geodesic between the endpoints on the corresponding lid, then we obtain a curve γ' between r'_i and l'_j with $\|\gamma'\| \leq \|\gamma\|$. Since there are no spikes between r'_i and l'_j, γ' lies in the same homotopy class as r'_i.

(b) In the second case, assume either r_i or l_j has no such bend on $\partial(D_i \cap D_j)$, but the other has at least one bend. Without loss of generality, assume that r_i has at least one bend. If l_j has one, a symmetric argument applies. We replace γ by a curve that passes through both x and x'. If γ does not pass through x already, then let $\bar{\varphi}$ and $\bar{\psi}$ be the lids of pockets of γ_i and γ_j, respectively, that intersect in x. It is also possible that $\bar{\varphi}$ is an edge of γ_i; the proof is then similar. Denote by a the arc $\bar{\varphi} \setminus D_j$, by a' the arc $\bar{\varphi} \cap D_j$ and by c the arc $\bar{\psi} \setminus D_i$. For the lids $\bar{\varphi}'$ and $\bar{\psi}'$ of r_i and l_j crossing in x', denote by b the arc $\bar{\varphi}' \setminus D_j$, by b' the arc $\bar{\varphi}' \cap D_j$ and by d the arc $\bar{\psi}' \setminus D_i$, see Figure 5.8 (b).

We claim that γ crosses either c, or both a and a'. If γ crosses c, we are done, so assume it does not. As $a \cup c$ connects γ_i and γ_j, γ must cross $a \cup c$ (and therefore a) an odd number of times to connect f and g. Lemma 5.5 implies that γ crosses $a \cup a'$ an even number of times. Hence, a' is crossed an odd number of times. We can thus find an arc of γ with endpoints on a and a', and since this arc does not cross c, it lies in the pocket of $\bar{\varphi}$. Replace this arc by the arc of $\bar{\varphi}$ between those endpoints, which is a shortest path in R that passes through x. We now have a curve with the same endpoints as γ that passes through c or x, and this curve is not longer than γ. We allow the resulting curve to cross γ_i and γ_j along $\bar{\varphi}$, however the resulting curve contains a subcurve of γ that connects a, a' or c to g. Analogously, we can replace this subcurve by a curve that crosses either x' or d. This yields a curve from f to g that passes through c or x, and then through x' or d. Since l_j has no bends, c and d lie on the same lid, which passes through x and x'. Since this lid is a shortest path, we can replace the subpath between c or x and x' or d by a shortest path in R that passes through both x and x'.

The portions of the curve before x and after x' can be shortcut using the techniques of case (a) such that they lie between l_j and r_i and not in $D_i \cap D_j$. This yields a curve γ' in the homotopy class of r_i' with the same endpoints as γ, and $\|\gamma'\| \leq \|\gamma\|$.

(c) In the final case, both r_i and l_j have bends on $\partial(D_i \cap D_j)$. As before, we shortcut γ such that it first passes through x and then through x'. Let the arcs a, a', c be as before, and let c' be the arc $\bar{\psi} \cap D_i$. As $a \cup a'$, as well as $c \cup c'$ are crossed an even number of times, but $a \cup c$ is crossed an odd



(a) (b) (c)

Figure 5.8: The three cases of Lemma 5.6. $D_i \cap D_j$ shaded.

number of times, we have that either both a and a', or both c and c' are crossed by γ. Replacing the arc between the crossings by the shortest path on the corresponding lid yields a path through x with length at most that of γ. Similarly, we can replace the remainder of the resulting path to also pass through x'. From here, we can use the same technique as in case (b) to obtain a curve γ' in the homotopy class of r'_i with the same endpoints as γ, and $\|\gamma'\| \leq \|\gamma\|$. □

5.2.1 Regular homotopies

Let $G(a, b)$ denote the geodesic in R between $f(a)$ and $g(b)$. Given a homotopy class σ, $G_\sigma(a, b)$ denotes the geodesic between $f(a)$ and $g(b)$ in σ. For $\gamma \in \Gamma$ denote its homotopy class by $\sigma(\gamma)$. If γ is a geodesic in R, we say that $\sigma(\gamma)$ is a *straight* homotopy class.

For an optimal homotopy, we may without loss of generality start by shortening γ_0 into the geodesic $G_{\sigma(\gamma_0)}(0, 0)$ in its homotopy class, use an optimal homotopy h from $G_{\sigma(\gamma_0)}(0, 0)$ to $G_{\sigma(\gamma_1)}(1, 1)$, and end by lengthening the geodesic $G_{\sigma(\gamma_1)}(1, 1)$ into γ_1. The resulting homotopy has cost $\max\{\|\gamma_0\|, cost(h), \|\gamma_1\|\}$, and the computational challenge is to efficiently find an optimal homotopy h.

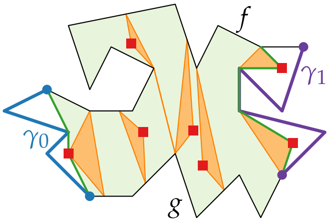


Figure 5.9: A decomposed homotopy. Subhomotopies S_i shaded green, and B_i shaded orange.

We call a homotopy from $G_{\sigma(\gamma_0)}(0, 0)$ to $G_{\sigma(\gamma_1)}(1, 1)$ a *regular* homotopy of order m if it can be decomposed into a sequence of homotopies S_0, B_1, S_1, $\ldots$, B_i, S_i, $\ldots$, B_m, S_m, subject to the following constraints, see also Figure 5.9.

- $S_i(1) = B_{i+1}(0)$ for $0 \leq i \leq m-1$ and $B_i(1) = S_i(0)$ for $1 \leq i \leq m$, that is, the last leash of a homotopy matches the first leash of the next homotopy.
- For each homotopy B_i, the leashes all have the same endpoints on f and g, but the leashes can move over spikes. Moreover, the longest leash in B_i is either $B_i(0)$ or $B_i(1)$.
- For each homotopy S_i, all leashes are geodesics in a single homotopy class σ_i, but the endpoints of leashes can move along f and g.
- The respective homotopy classes of leashes in S_0 and S_m are $\sigma_0 = \sigma(\gamma_0)$ and $\sigma_m = \sigma(\gamma_1)$.

In Lemma 5.7, we show that there exists a minimum-cost homotopy between $G_{\sigma(\gamma_0)}(0, 0)$ and $G_{\sigma(\gamma_1)}(1, 1)$ that is a regular homotopy of order at most k. In Lemma 5.8, we show that each σ_i (except possibly σ_0 and σ_m) can be assumed to be a straight homotopy class. For each homotopy B_i, some leash can be assumed be the geodesic in R between its endpoints.

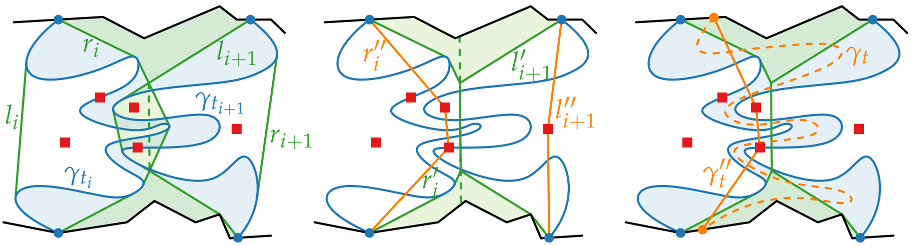
For a homotopy h, denote by $\alpha_h(t) \in [0,1]$, the position of the start of leash $h(t)$ in the parameter space of f (such that $f(\alpha_h(t)) = h(t,0)$). Symmetrically, denote by $\beta_h(t) \in [0,1]$ the position of the end of that leash in the parameter space of g (such that $g(\beta_h(t)) = h(t,1)$). If h is monotone, then α_h and $\beta_h \colon [0,1] \to [0,1]$ are continuous nondecreasing surjections.

Lemma 5.7. *Let h be a monotone homotopy from $G_{\sigma(\gamma_0)}(0,0)$ to $G_{\sigma(\gamma_1)}(1,1)$ of cost less than L, then there exists a regular homotopy of cost at most L.*

Proof. By monotonicity, as t increases, the number of spikes in the swept region of γ_t cannot decrease. Let $\{t_1, \dots, t_k\}$ be the minimum value t_i for which the swept region of γ_{t_i} contains at least i spikes. For each leash γ_{t_i} of h, let $D_i = \mathrm{gh}_R(\gamma_{t_i})$ be its geodesic hull, and let l_i and $r_i \in \Gamma$ respectively be the backwards-convex and forwards-convex curves on the boundary of D_i connecting $\gamma_{t_i}(0)$ with $\gamma_{t_i}(1)$. For $1 \leq i \leq k$, define r'_i and l'_{i+1} respectively to be the curves obtained by replacing the arcs of r_i and respectively l_{i+1} inside $D_i \cap D_{i+1}$ by the geodesic between the crossings of l_{i+1} and r_i (if any), dashed in Figure 5.10(a). Then l'_i lies in the swept region of r'_i, and r'_i lies in that of l'_{i+1}. Let r''_i and l''_i be the geodesics with the same endpoints and homotopy class as r'_i and l'_i, respectively (see Figure 5.10(b)).

Lemma 5.3 gives us a homotopy B_i from l''_i to r''_i whose leashes have length at most $\max\{\|l''_i\|, \|r''_i\|\}$. Including the cost of spikes, the cost of B_i is at most $\max\{\|l''_i\|, \|r''_i\|\} + 1 + \varepsilon$ (for any $\varepsilon > 0$) by perturbing leashes to each encounter at most one spike. Because l''_i is a geodesic in the same homotopy class as l'_i, we have $\|l''_i\| \leq \|l'_i\|$. Moreover, l'_i is a copy of l_i with a subpath replaced by a shortest path, so $\|l'_i\| \leq \|l_i\|$. Finally, because l_i lies on arcs of the convex hull of γ_{t_i}, we have $\|l_i\| \leq \|\gamma_{t_i}\|$. Therefore $\|l''_i\| \leq \|\gamma_{t_i}\|$, and by symmetry, $\|r''_i\| \leq \|\gamma_{t_i}\|$. Since γ_{t_i} encounters a spike, and the cost of h is less than L, we have $\|\gamma_{t_i}\| < L - 1$. Hence, $cost(B_i) \leq \max\{\|l''_i\|, \|r''_i\|\} + 1 + \varepsilon \leq \|\gamma_{t_i}\| + 1 + \varepsilon \leq L$.

Define $r''_0 = G_{\sigma(\gamma_0)}(0,0)$ and $l''_{k+1} = G_{\sigma(\gamma_1)}(1,1)$. Moreover, let $t_0 = 0$ and $t_{k+1} = 1$. It remains to construct homotopies S_i between r''_i and l''_{i+1} for $0 \leq i \leq k$. Since there are no spikes interior to the region between r'_i and l'_{i+1}, they lie in the



(a) (b) (c)

Figure 5.10: (a) The region between r_i and l_{i+1} obtained from the geodesic hulls of γ_{t_i} and $\gamma_{t_{i+1}}$. (b) The curves r'_i and l'_{i+1}, and the corresponding geodesics r''_i and l''_{i+1} in the same homotopy class. (c) For all $t_i \leq t \leq t_{i+1}$, the curve γ_t lies in the region between the curves γ_{t_i} and $\gamma_{t_{i+1}}$.

same homotopy class, which we denote by σ_i. To construct a homotopy from r''_i to l''_{i+1}, we consider the curves γ_t with $t_i \leq t \leq t_{i+1}$, and replace them by the geodesic $\gamma''_t = G_{\sigma_i}(\alpha_h(t), \beta_h(t))$ in homotopy class σ_i (see Figure 5.10(c)). These geodesics move continuously with t, so it remains to show that $\|\gamma''_t\| \leq \|\gamma_t\|$. This is not immediate since γ''_t may lie in a different homotopy class than γ_t. Instead, we use Lemma 5.6, which tells us that there is a curve with the same endpoints in the homotopy class of r'_i with length at most that of γ_t. Because γ''_t is a geodesic in the same homotopy class, its length is also at most that of γ_t. □

The straight homotopy classes in R can be enumerated by taking the geodesic between any two spikes in R, and extending it to the two points on the boundary of R. It is at these points where the geodesic hits the boundary of R that the homotopy class of the geodesic between points on the boundary of R changes: one can slide these points clockwise or counter-clockwise such the geodesic between them ends up in a different straight homotopy class. There are $O(k^2)$ straight homotopy classes.

Lemma 5.8. *For $1 \leq i \leq k-1$, we can assume σ_i to be a straight homotopy class without increasing its cost.*

Proof. Curve r_i is forwards-convex and l_{i+1} is backwards-convex, and the endpoints of r_i on f and g are not ahead of those of l_{i+1}. If r_i and l_{i+1} are disjoint, then we can find a geodesic in R separating r_i and l_{i+1}, and hence r''_i and l''_{i+1}. If they are not disjoint, then the shortest path between their points of intersection lies on a geodesic between f and g, separating r'_i and l'_{i+1}, and hence r''_i and l''_{i+1}. □

5.2.2 Computation

A tool that is commonly used to compute the Fréchet distance is the *free space diagram* [5]. This tool captures between which points of f and g the geodesic is sufficiently short to be used as a leash in a homotopy of a given cost L. Formally, the free space diagram is defined as $\mathcal{F}(L) = \{(a,b) \in [0,1] \times [0,1] \mid \|G(a,b)\| \leq L\}$. More generally, for a given homotopy class σ, we define $\mathcal{F}_\sigma(L) = \{(a,b) \in [0,1] \times [0,1] \mid \|G_\sigma(a,b)\| \leq L\}$ to capture the geodesics in σ of length at most L.

Let Σ be the set of homotopy classes consisting of $\sigma(\gamma_0), \sigma(\gamma_1)$, and all straight homotopy classes. There are $2 + O(k^2) = O(k^2)$ such homotopy classes, assuming $k \geq 1$. Let h be a regular homotopy of cost at most L, and let t_i and t'_i be the values of t in h at which the constituent homotopy S_i starts and stops, respectively. For all $t \in [t_i, t'_i]$, we have $\|h(t)\| \leq cost(S_i) \leq L$, so $(\alpha_h(t), \beta_h(t)) \in \mathcal{F}_{\sigma_i}(L)$, where σ_i is the homotopy class of the leashes in S_i. For $t \in [t'_{i-1}, t_i]$, the leashes $h(t)$ are part of homotopy B_i, and we even have $\|h(t)\| + 1 \leq L$, such that accounting for the spikes the leash passes over, we have $cost(B_i) \leq \max\{\|h(t'_{i-1})\| + 1, \|h(t_i)\| + 1\} \leq L$ by the construction of Lemma 5.7. Recall that the endpoints of leashes do not move throughout any homotopy B_i, so $\alpha_h(t) = \alpha_h(t_i)$ and $\beta_h(t) = \beta_h(t_i)$ for all $t \in [t'_{i-1}, t_i]$. Additionally, as the construction of Lemma 5.7 preserves monotonicity, we can assume α_h and β_h to both be continuous nondecreasing surjections. By Lemma 5.8, we can also assume that each σ_i lies in Σ. For the sake of presentation, since α_h and β_h are constant in the intervals $[t'_{i-1}, t_i]$, we assume from now on that $t'_{i-1} = t_i$, and prove that any homotopy with the structure imposed by Lemma 5.9 can be turned into a regular homotopy of cost L.

Lemma 5.9. *We can construct a regular homotopy of cost at most L if we can find appropriate α_h, β_h, t_i and t'_i and values of $\sigma_i \in \Sigma$, with the following conditions. Let α and $\beta\colon [0,1] \to [0,1]$ be continuous nondecreasing surjections. Let $\sigma_0 = \sigma(\gamma_0)$, $\sigma_m = \sigma(\gamma_1)$, and $\sigma_i \in \Sigma$ for $i \in \{1, \dots, m-1\}$. Let $0 = t_0 \leq t_1 \leq \cdots \leq t_{m+1} = 1$. Then, if $(\alpha(t), \beta(t)) \in \mathcal{F}_{\sigma_i}(L)$ for each $t \in [t_i, t_{i+1}]$, and additionally $(\alpha(t), \beta(t)) \in \mathcal{F}_{\sigma_i}(L-1) \cap \mathcal{F}_{\sigma_{i+1}}(L-1)$ for each t_i with $i \in \{1, \dots, m\}$, this corresponds to a regular homotopy of cost at most L.*

Proof. We use geodesics of σ_i for $t \in [t_i, t_{i+1}]$, and they move continuously. By Lemma 5.3, we can find a homotopy B_i of cost at most L if the geodesics of σ_i and σ_{i+1} based at $\alpha(t)$ and $\beta(t)$ both have length at most $L-1$. Furthermore, since $(\alpha(t), \beta(t)) \in \mathcal{F}_{\sigma_i}(L)$ for each $t \in [t_i, t_{i+1}]$, we can find a homotopy S_i of cost at most L between σ_i and σ_{i+1}. ☐

Computing free space diagrams. To compute the free space diagram in our setting, we subdivide the edges of f and g in such a way that for each pair of (subdivided) edges, the length of the geodesic can be described as a quadratic function in two parameters a and b. This subdivision is based on the lines through any pair of spikes and vertices of ∂R, and finding their projection onto f or g, if any. In total, this yields subdivided curves f' and g' of $O((n+k)^2)$ vertices. Using a standard rotating sweep around every spike and vertex, we can compute the projections in $O((n+k)^2 \log(n+k))$ time and sort them along every edge of f and g in the same time, giving the subdivided curves f' and g'.

Now, given any straight homotopy class, or the homotopy class of γ_0 or γ_1, we compute the quadratic function for each pair of edges $(e_{f'}, e_{g'})$ of f' and g'. To this end, we take a straight homotopy class σ and determine the induced partition of K into K_1 and K_2. Then we compute the convex hulls $\mathrm{gh}_R(K_1)$ and $\mathrm{gh}_R(K_2)$ of K_1 and K_2 relative to the domain R. Using the common inner tangents of $\mathrm{gh}_R(K_1)$ and $\mathrm{gh}_R(K_2)$ in R we can find all pairs of edges $(e_{f'}, e_{g'})$ of f' and g' for which the geodesic in σ is straight, and determine the corresponding quadratic functions (which are ellipses). For other pairs of edges of f' and g', the geodesic contains vertices of $\mathrm{gh}_R(K_1)$, $\mathrm{gh}_R(K_2)$, and R itself, and their lengths are determined by a hyperbolic part in a, a hyperbolic part in b, and a constant part (between the first and last vertices not on f' and g'). We fix an edge $e_{g'}$ of g' and traverse all edges of f' sequentially, updating the three parts of the quadratic function when needed. Updates of the constant part happen only at the ends of the geodesic, and amortized we can do all updates in time linear in the number of parts of f', that is, $O((n+k)^2)$.

Hence, we can compute all the quadratic functions for all straight homotopy classes in time $O(k^2)$ (for the straight homotopy classes) times $O((n+k)^2)$ (for the number of segments of g') times $O((n+k)^2)$ (for the number of segments of f'). In total, this is $O((n+k)^4 k^2) = O(n^4 k^2 + k^6)$ time. The $O(n^4 + k^4)$ cells of the free space diagram each have $O(k^2)$ quadratic functions, at most one for each of the homotopy classes.

Decision algorithm. For a parameter L, we define the *reachable free space* as the set of coordinates $(\sigma, a, b) \in \Sigma \times [0,1] \times [0,1]$, such that there exist continuous nondecreasing surjections $\alpha\colon [0,1] \to [0,a]$ and $\beta\colon [0,1] \to [0,b]$, a value m, values $0 = t_0 \leq t_1 \leq \cdots \leq t_{m+1} = 1$, and homotopy classes $\sigma_i \in \Sigma$ with $\sigma_0 = \sigma(\gamma_0)$ and $\sigma_m = \sigma$, such that for each $t \in [t_i, t_{i+1}]$, we have $(\alpha(t), \beta(t)) \in \mathcal{F}_{\sigma_i}(L)$ and for each $i \in \{1, \ldots, m\}$, we have $(\alpha(t_i), \beta(t_i)) \in \mathcal{F}_{\sigma_i}(L-1) \cap \mathcal{F}_{\sigma_{i+1}}(L-1)$. The reachable free space corresponds to the classes σ and points $f(a)$ and $g(b)$ that have a monotone regular homotopy from $G_{\sigma_0}(0,0)$ to $G_{\sigma}(f(a), g(b))$ of cost at most L. Deciding whether a regular homotopy of at most a certain cost L exists is then equivalent to testing whether $(\sigma(\gamma_1), 1, 1)$ lies in the reachable free space for parameter L. We can compute the reachable free space using dynamic programming. In contrast algorithms for most variants of the Fréchet distance, which need only information about the free space on the boundary of cells, we also need information about their interiors. In our dynamic program, we compute the reachable free space on the boundary of each cell.

The restriction of the free space to any cell and homotopy class is convex. Therefore, if a point (σ, a, b) lies in the reachable free space, then for all $a' \geq a$ and $b' \geq b$, in the same cell as (a, b), if $(a', b') \in \mathcal{F}_{\sigma}(e)(L)$, then (σ, a', b') also lies in the reachable free space. Moreover, for σ and $\sigma' \in \Sigma$, if $(a, b) \in \mathcal{F}_{\sigma}(L-1)$ and $(a, b) \in \mathcal{F}_{\sigma'}(L-1)$, then (σ, a, b) lies in the reachable free space if and only if (σ', a, b) lies in the reachable free space. Our dynamic program starts as follows: for a homotopy of cost at most L to exist, check whether $(0,0) \in \mathcal{F}_{\sigma_0}(L)$. If so, $(\sigma_0, 0, 0)$ lies in the reachable free space, and otherwise the reachable free space is empty. Now we propagate the reachable free space on a cell-by-cell basis, maintaining for each cell $[a, a'] \times [b, b']$ and for each homotopy class σ, two pieces of information. First, the minimum $a^* \in [a, a']$ for which (σ, a^*, b) lies in the reachable free space (if any); and second, the minimum $b^* \in [b, b']$ for which (σ, a, b^*) lies in the reachable free space (if any). The first piece of information can be propagated to a neighboring cell $[a, a'] \times [b', b'']$, and the second piece can be propagated to a neighboring cell $[a', a''] \times [b, b']$.

To propagate this information, we use a horizontal sweep line that maintains the reachable free space intersecting the sweep line for the cell $[a, a'] \times [b, b']$ in each of the $O(k^2)$ relevant homotopy classes, based on the coordinates (σ, a, b^*) and (σ, a^*, b) in each of those homotopy classes. Naively, we can propagate this information in $O(k^6)$ time per cell, using the coordinates of the $O(k^4)$ intersections of the boundary of free space cells from different homotopy classes as events for the sweep line.

After propagating the information through all $O(n^4 + k^4)$ cells of the free space in $O(n^4 k^6 + k^{10})$ time, we can return whether $(\sigma(\gamma_1), 1, 1)$ lies in the reachable free space to decide whether there exists a homotopy of cost at most L.

Exact computation. The candidate values for the minimum-cost regular homotopy depend on the values of L where the a- or b-coordinates of different intersections align. There are $O(((n+k)^2 k^4)^2)$ intersections that can align in this way, which yields $O(n^4 k^8 + k^{12})$ critical values, which we can enumerate in $O(1)$ time per value. We perform a binary search over these critical values, using linear-time median finding and running the decision procedure $O(\log nk)$ times to find the minimum cost of a regular homotopy. Doing so, we compute the homotopy height in $O(n^4k^6 \log n + n^4k^8 + k^{12})$ time.

5.3 Discussion

We have shown that in a spiked plane with polygonal boundary, we can compute the homotopy height between two curves on the boundary in polynomial time for various cases. In particular, this holds if all spikes have the same (unit) weight, or if the two curves together form the entire boundary of the domain. We also provide a 2-approximation algorithm for the general case. We have in addition shown that intermediate leashes may require many inflection points for an optimal homotopy, even if the polygonal domain is convex. The complexity of the leash has prevented us from developing a polynomial-time algorithm, thus it remains open whether the general case can be solved optimally in polynomial time.

Future work. Various other settings can also be studied. The case where f and g are not on the boundary of the polygonal domain is a natural first step. However, the monotonicity [27] that supports our results is not known to hold in this case, which is likely a premise for efficient optimal algorithms. The algorithm [40] upon which our approximation algorithm is based does not require f and g to lie on the boundary. Hence, our approximation algorithm (Lemma 5.4) extends to deal with the case where γ_0 and γ_1 are on the boundary, but f or g is not. If the initial leashes γ_0 and γ_1 are no longer specified, but can instead by chosen in such a way that the cost of an optimal homotopy is minimized, we readily get a 2-approximation algorithm by using the algorithm by Chambers et al. [29] to solve the decision variant combined with an appropriate search.

6

Concluding Remarks

In this thesis we studied the computational aspects of various similarity measures. In each case, we optimized an aspect of a continuous deformation between the compared objects, and measured the similarity of compared shapes based on this aspect. We considered two such aspects, namely the maximum distance that a point travels on its way from one shape to the other in Part I, and the size of the largest intermediate shape throughout the deformation in Part II.

The objects we considered in this thesis were either curves or surfaces, but the measures we studied for comparing them extend to other shapes as well. Examples of such shapes could be graphs, or higher-dimensional shapes. However, one important restriction that is imposed by continuous deformations is that the compared shapes must be homeomorphic.

In certain cases, it is desirable to meaningfully measure the similarity between shapes that are not homeomorphic. The field of topological data analysis studies the comparison of discrete point sets that do not necessarily contain the same number of points. Along the same lines, we want to be able to meaningfully compare shapes that are almost homeomorphic, and still strive for some (weaker) sense of continuity. To this end, we can allow small features that prevent a continuous deformation between compared shapes to be collapsed at a small cost. By collapsing such features, the compared shapes can ideally be transformed into shapes that are homeomorphic, and can thus be compared using continuous similarity measures. The contour tree distance as introduced in Chapter 2 is one approach that can be used to compare shapes that are not homeomorphic, namely trees.

The contour tree distance provides a lower bound on the Fréchet distance between surfaces of genus 0 equipped with a real-valued function. This lower bound is computed by computing the contour tree distance between the Reeb graphs of those functions. We leave it open whether the contour tree distance is actually equivalent to the Fréchet distance for real-valued functions on spheres. Whereas for surfaces of genus 0 the Reeb graph will form a tree, this is not necessarily the case for surfaces of higher genus.

Fortunately, the contour tree distance can be extended to graphs in several ways, such that it provides a lower bound on the Fréchet distance between real-valued functions on higher-genus surfaces. Loosely speaking, the freedom in generalizing the contour tree distance arises in how one allows cycles to be collapsed. The lower bound that we obtain from such a generalization would ideally be as close to the Fréchet distance as possible, using only the information stored in the Reeb graphs of the compared surfaces.

Moreover, for surfaces equipped with functions to $\mathbb{R}^d$, we can derive real-valued functions by projecting $\mathbb{R}^d$ to an arbitrary line in $\mathbb{R}^d$. For any such line, the contour tree distance (or its generalizations) form a lower bound on the Fréchet distance between the original surfaces. Since the Reeb graph views a real-valued surface simply as a graph, it seems that generalizations of the contour tree distance will be are lot easier to compute than the Fréchet distance. In fact, since computing the Fréchet distance between surfaces in $\mathbb{R}^d$ is an active area of research, the proposed lower bounds could provide a viable alternative for comparing surfaces.

Since projecting a surface in $\mathbb{R}^d$ onto a single line loses a lot of information, we propose to extend the contour tree distance further. For a set of multiple such lines, we wish to find matchings between their respective contour trees in such a way that the matchings are in some sense compatible. To explain what we mean by compatible, consider a pair of matched points on two contour trees. This pair captures (at least) one pair of matched points on a pair of contours on the corresponding surfaces. If we now consider contour trees for a second line, a compatible matching between them would also match a pair of points of the aforementioned contours (which arose from a different line).

This second extension would provide a more tight lower bound for the Fréchet distance, albeit at a computational cost. It is very appealing that, depending on how closely the generalization of the contour tree distance matches the Fréchet distance, Reeb graphs of non-homeomorphic surfaces can be compared this way.

Similar to how generalizations of the contour tree distance could be used to compare shapes that are not homeomorphic, we can ask how homotopy height can be generalized in order to compare different shapes. For this, we can use a different way of moving between the compared shapes. In the setting of curves on surfaces, we allow closed curves (or spheres in general settings) to emanate and disappear at points. Moreover, when two closed curves collide at a point, we can reconnect the two curves into a single one at the point of collision. Symmetrically, we allow splitting a single curve in two using reconnections at a point on the curve.

The moves described above give rise to the *homology height* parameter, which minimizes the total length of the longest intermediate set of curves throughout the motion. Homology height allows for the comparison of homologous cycles. Whereas it remains open whether computing the homotopy height between two curves bounding an annulus is NP-hard, it can be shown that computing homology height is NP-hard using a simple reduction from cut-width [74].

A second type of generalization of homotopy height concerns homotopies between higher dimensional shapes. The homotopies considered in this thesis involve the motion of curves over a surface. The simplest higher-dimensional analogue would be a motion of surfaces through a volume in such a way that the area of the largest intermediate surface is minimized. Specifically, the natural analogue to homotopies between the two boundaries of an annulus is that of a homotopy between the two boundaries of $S^2 \times [0,1]$. In this setting, homotopy height turns out to be an NP-hard problem by a simple reduction from homology height of curves on a surface [30]. Namely, a thickened version of a homology height instance can be embedded in $S^2 \times [0,1]$ in such a way that a minimum height homotopy between the boundaries of $S^2 \times [0,1]$ is in correspondence with the motion of the intersection pattern of the homology height instance with the intermediate surfaces of an optimal homotopy.

Despite the computational hardness of optimal homotopies of surfaces, results on the structure of optimal homotopies could lead to promising upper bounds on their complexity. First, we conjecture that there is always an optimal homotopy between the boundaries of $S^2 \times [0,1]$ that is monotone. This would also imply that an optimal homotopy is a monotone isotopy, which also remains an open problem.

References

[1] P. K. Agarwal, R. Ben Avraham, H. Kaplan, and M. Sharir. Computing the discrete Fréchet distance in subquadratic time. In *Proc. 24th Sympos. Discrete Algorithms (SODA)*, pages 156–167, 2013.

[2] P. K. Agarwal, K. Fox, A. Nath, A. Sidiropoulos, and Y. Wang. Computing the Gromov-Hausdorff distance for metric trees. In *Proc. 26th Internat. Sympos. Algorithms Comput. (ISAAC)*, pages 529–540, 2015.

[3] A. V. Aho, J. E. Hopcroft, and J. D. Ullman. *The Design and Analysis of Computer Algorithms*. Addison-Wesley, 1974.

[4] H. Alt and M. Buchin. Can we compute the similarity between surfaces? *Discrete Comput. Geom. (DCG)*, 43(1):78–99, 2010.

[5] H. Alt and M. Godau. Computing the Fréchet distance between two polygonal curves. *Internat. J. Comput. Geom. Appl. (IJCGA)*, 5(1–2):75–91, 1995.

[6] P. Angelini, G. Da Lozzo, G. Di Battista, F. Frati, M. Patrignani, and V. Roselli. Morphing planar graph drawings optimally. In *Internat. Colloq. Automata Lang. Program. (ICALP)*, LNCS 8572, pages 126–137, 2014.

[7] P. Angelini, F. Frati, M. Patrignani, and V. Roselli. Morphing planar graph drawings efficiently. In *Graph Drawing*, volume 8242 of *Lecture Notes in Computer Science*, pages 49–60. 2013.

[8] U. Bauer, X. Ge, and Y. Wang. Measuring distance between reeb graphs. In *Proc. 30th Sympos. Comput. Geom. (SoCG)*, pages 464–473, 2014.

[9] U. Bauer, E. Munch, and Y. Wang. Strong equivalence of the interleaving and functional distortion metrics for Reeb graphs. In *Proc. 31st Sympos. Comput. Geom. (SoCG)*, pages 461–475, 2015.

[10] K. Beketayev, D. Yeliussizov, D. Morozov, G. H. Weber, and B. Hamann. Measuring the distance between merge trees. In *Topological Methods in Data Analysis and Visualization III*, pages 151–165, 2014.

[11] D. Bienstock and P. Seymour. Monotonicity in graph searching. *Journal of Algorithms*, 12(2):239–245, 1991.

[12] S. Brakatsoulas, D. Pfoser, R. Salas, and C. Wenk. On map-matching vehicle tracking data. In *Proc. 31st Internat. Conf. on Very Large Data Bases (VLDB)*, pages 853–864, 2005.

[13] G. R. Brightwell and P. Winkler. Submodular percolation. *SIAM Journal on Discrete Mathematics*, 23(3):1149–1178, 2009.

[14] K. Bringmann. Why walking the dog takes time: Fréchet distance has no strongly subquadratic algorithms unless SETH fails. In *Proc. 55th Sympos. Found. Comput. Sci. (FOCS)*, pages 661–670, 2014.

[15] K. Bringmann and W. Mulzer. Approximability of the discrete Fréchet distance. In *Proc. 31st Sympos. Comput. Geom. (SoCG)*, pages 739–753, 2015.

[16] K. Buchin, M. Buchin, and J. Gudmundsson. Constrained free space diagrams: a tool for trajectory analysis. *Internat. J. of Geographical Information Science*, 24(7):1101–1125, 2010.

[17] K. Buchin, M. Buchin, W. Meulemans, and W. Mulzer. Four soviets walk the dog-with an application to Alt's conjecture. In *Proc. 25th Sympos. Discrete Algorithms (SODA)*, pages 1399–1413, 2014.

[18] K. Buchin, M. Buchin, and A. Schulz. Fréchet distance of surfaces: Some simple hard cases. In *Proc. 18th European Sympos. Algorithms (ESA)*, pages 63–74, 2010.

[19] K. Buchin, M. Buchin, R. van Leusden, W. Meulemans, and W. Mulzer. Computing the Fréchet distance with a retractable leash. In *Proc. 21st European Sympos. Algorithms (ESA)*, pages 241–252, 2013.

[20] K. Buchin, M. Buchin, and C. Wenk. Computing the Fréchet distance between simple polygons. *Comput. Geom. Theory Appl. (CGTA)*, 41(1–2):2–20, 2008.

[21] K. Buchin, E. W. Chambers, T. Ophelders, and B. Speckmann. Fréchet isotopies to monotone curves. In *Abstr. 33rd European Workshop on Computational Geometry (EuroCG)*, pages 41–44, 2017.

[22] K. Buchin, T. Ophelders, and B. Speckmann. Computing the similarity between moving curves. In *Proc. 23rd European Sympos. Algorithms (ESA)*, pages 928–940, 2015.

[23] K. Buchin, T. Ophelders, and B. Speckmann. Computing the Fréchet distance between real-valued surfaces. In *Proc. 28th Sympos. Discrete Algorithms (SODA)*, pages 2443–2455, 2017.

[24] M. Buchin. *On the Computability of the Fréchet Distance Between Triangulated Surfaces*. PhD thesis, Free University Berlin, Institute of Computer Science, 2007.

[25] B. Burton, E. W. Chambers, M. van Kreveld, W. Meulemans, T. Ophelders, and B. Speckmann. Computing optimal homotopies over a spiked plane with polygonal boundary. In *Proc. 25th European Sympos. Algorithms (ESA)*, pages 23:1 – 23:14, 2017.

[26] A. Calder and J. Siegel. On the width of homotopies. *Topology*, 19(3):209–220, 1980.

[27] E. W. Chambers, G. R. Chambers, A. de Mesmay, T. Ophelders, and R. Rotman. Constructing monotone homotopies and sweepouts. arXiv: 1704.06175, 2017.

[28] E. W. Chambers, A. de Mesmay, and T. Ophelders. On the complexity of optimal homotopies. In *Proc. 29th Sympos. Discrete Algorithms (SODA)*, pages 1121–1134, 2018.

[29] E. W. Chambers, É. de Verdière, J. Erickson, S. Lazard, F. Lazarus, and S. Thite. Homotopic Fréchet distance between curves or, walking your dog in the woods in polynomial time. *Comput. Geom. Theory Appl. (CGTA)*, 43(3):295–311, 2010.

[30] E. W. Chambers and D. Eppstein. personal communication, 2016.

[31] E. W. Chambers and D. Letscher. On the height of a homotopy. In *Proc. 21st Canadian Conference on Computational Geometry*, pages 103–106, 2009.

[32] E. W. Chambers, D. Letscher, T. Ju, and L. Liu. Isotopic Fréchet distance. In *Abstr. 23rd Canadian Conf. Comput. Geom. (CCCG)*, pages 59–64, 2011.

[33] E. W. Chambers and M. Vejdemo-Johansson. Computing minimum area homologies. *Computer Graphics Forum*, 34(6):13–21, 2015.

[34] E. W. Chambers and Y. Wang. Measuring similarity between curves on 2-manifolds via homotopy area. In *Proc. 29th Sympos. Comput. Geom. (SoCG)*, pages 425–434, 2013.

[35] G. R. Chambers. *Optimal Homotopies of Curves on Surfaces*. PhD thesis, University of Toronto, Canada, 2014.

[36] G. R. Chambers and Y. Liokumovich. Converting homotopies to isotopies and dividing homotopies in half in an effective way. *Geometric and Functional Analysis*, 24(4):1080–1100, 2014.

[37] G. R. Chambers and R. Rotman. Monotone homotopies and contracting discs on Riemannian surfaces. *Journal of Topology and Analysis*, pages 1–32, 2016.

[38] É. Colin de Verdière and J. Erickson. Tightening nonsimple paths and cycles on surfaces. *SIAM Journal on Computing*, 39(8):3784–3813, 2010.

[39] A. F. Cook IV, A. Driemel, S. Har-Peled, J. Sherette, and C. Wenk. Computing the Fréchet distance between folded polygons. In *Proc. 12th Algorithms and Data Structures (WADS)*, pages 267–278, 2011.

[40] A. F. Cook IV and C. Wenk. Geodesic Fréchet distance inside a simple polygon. *Transactions on Algorithms*, 7(1):Art. 9, 2010.

[41] M. de Berg, A. Mehrabi, and T. Ophelders. Data structures for fréchet queries in trajectory data. In *Abstr. 29th Canadian Conf. Comput. Geom. (CCCG)*, pages 214–219, 2017.

[42] M. De Biasi and T. Ophelders. The complexity of snake and undirected NCL variants. *Theoretical Computer Science*, 2017. To appear.

[43] J.-L. De Carufel, A. Dumitrescu, W. Meulemans, T. Ophelders, C. Pennarun, C. D. Tóth, and S. Verdonschot. On convex polygons in cartesian products. In *Abstr. 34th European Workshop on Computational Geometry (EuroCG)*, pages 39:1–39:6, 2018.

[44] M. Dehn. Transformation der Kurven auf zweiseitigen Flächen. *Mathematische Annalen*, 72:413–421, 1912.

[45] D. Dereniowski. From pathwidth to connected pathwidth. *SIAM Journal on Discrete Mathematics*, 26(4):1709–1732, 2012.

[46] T. K. Dey and S. Guha. Transforming curves on surfaces. *Journal of Computer and System Sciences*, 58:297–325, 1999.

[47] T. K. Dey, A. N. Hirani, and B. Krishnamoorthy. Optimal homologous cycles, total unimodularity, and linear programming. *SIAM Journal on Computing*, 40(4):1026–1044, 2011.

[48] B. Di Fabio and C. Landi. The edit distance for reeb graphs of surfaces. *Discrete Comput. Geom. (DCG)*, 55(2):423–461, 2016.

[49] M. P. do Carmo. *Riemannian geometry*. Birkhäuser, 1992.

[50] A. Driemel, S. Har-Peled, and C. Wenk. Approximating the Fréchet distance for realistic curves in near linear time. *Discrete Comput. Geom. (DCG)*, 48(1):94–127, 2012.

[51] H. Edelsbrunner and J. Harer. Persistent homology-a survey. *Contemporary mathematics*, 453:257–282, 2008.

[52] A. Efrat, Q. Fan, and S. Venkatasubramanian. Curve matching, time warping, and light fields: New algorithms for computing similarity between curves. *Journal of Mathematical Imaging and Vision*, 27(3):203–216, 2007.

[53] A. Efrat, L. J. Guibas, S. Har-Peled, J. S. B. Mitchell, and T. M. Murali. New similarity measures between polylines with applications to morphing and polygon sweeping. *Discrete Comput. Geom. (DCG)*, 28(4):535–569, 2002.

[54] M. J. Egenhofer. Reasoning about binary topological relations. In *Sympos. on Spatial Databases*, pages 141–160, 1991.

[55] J. Erickson and K. Whittlesey. Transforming curves on surfaces redux. In *Proc. 24th Sympos. Discrete Algorithms (SODA)*, pages 1646–1655, 2013.

[56] B. Fasy, S. Karakoç, and C. Wenk. On minimum area homotopies. In *Computational Geometry: Young Researchers Forum*, pages 49–50, 2016.

[57] F. V. Fomin and D. M. Thilikos. An annotated bibliography on guaranteed graph searching. *Theoretical computer science*, 399(3):236–245, 2008.

[58] P. Fraigniaud and N. Nisse. Connected treewidth and connected graph searching. In *Latin American Symposium on Theoretical Informatics*, pages 479–490, 2006.

[59] M. Fréchet. Sur quelques points du calcul fonctionnel. *Rendiconti Circ. Mat. Palermo*, 22:1–74, 1906.

[60] M. Fréchet. Sur la distance de deux surfaces. *Ann. Soc. Polon. Math.*, 3:4–19, 1924.

[61] M. R. Garey and D. S. Johnson. *Computers and intractability: a guide to the theory of NP-completeness*. 1979.

[62] L. Gattinoni and A. Protti. Ventilation in the prone position: For some but not for all? *Canadian Medical Association Journal*, 178(9):1174–1176, 2008.

[63] M. Godau. *On the complexity of measuring the similarity between geometric objects in higher dimensions*. PhD thesis, Freie Universität Berlin, Germany, 1998.

[64] M. Gromov. Quantitative homotopy theory. *Prospects in Mathematics: Invited Talks on the Occasion of the 250th Anniversary of Princeton University*, pages 45–49, 1999.

[65] Q. Han and J.-X. Hong. *Isometric embedding of Riemannian manifolds in Euclidean spaces*, volume 130 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2006.

[66] S. Har-Peled, A. Nayyeri, M. Salavatipour, and A. Sidiropoulos. How to walk your dog in the mountains with no magic leash. *Discrete Comput. Geom. (DCG)*, 55(1):39–73, 2016.

[67] D. Keysers and W. Unger. Elastic image matching is NP-complete. *Pattern Recognition Letters*, 24(1):445–453, 2003.

[68] M. Kleinhans, M. van Kreveld, T. Ophelders, W. Sonke, B. Speckmann, and K. Verbeek. Computing representative networks for braided rivers. In *Proc. 33rd Sympos. Comput. Geom. (SoCG)*, pages 48:1–48:16, 2017.

[69] G. Kokarev. On geodesic homotopies of controlled width and conjugacies in isometry groups. *Groups, Geometry, and Dynamics*, 7:911–929, 2013.

[70] M. Konzack, T. McKetterick, T. Ophelders, M. Buchin, L. Giuggioli, J. Long, T. Nelson, M. A. Westenberg, and K. Buchin. Visual analytics of delays and interaction in movement data. *Internat. J. of Geographical Information Science*, 32(2):320–345, 2017.

[71] A. S. LaPaugh. Recontamination does not help to search a graph. *Journal of the ACM (JACM)*, 40(2):224–245, 1993.

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

[73] F. Lazarus and J. Rivaud. On the homotopy test on surfaces. In *Proc. 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 440–449, 2012.

[74] B. Monien and I. Sudborough. Min cut is NP-complete for edge weighted trees. *Theoretical Computer Science*, 58(1):209 – 229, 1988.

[75] D. Morozov, K. Beketayev, and G. Weber. Interleaving distance between merge trees. *Discrete Comput. Geom. (DCG)*, 49:22–45, 2013.

[76] A. Nayyeri and A. Sidiropoulos. Computing the Fréchet distance between polygons with holes. In *Proc. 42nd Internat. Colloq. Automata Lang. Program. (ICALP)*, pages 997–1009, 2015.

[77] A. Nayyeri and H. Xu. On Computing the Fréchet Distance Between Surfaces. In *Proc. 32nd Sympos. Comput. Geom. (SoCG)*, volume 51, pages 55:1–55:15, 2016.

[78] A. Nayyeri and H. Xu. On the decidability of the Fréchet distance between surfaces. In *Proc. 29th Sympos. Discrete Algorithms (SODA)*, pages 1109–1120, 2018.

[79] Z. Nie. On the minimum area of null homotopies of curves traced twice. Computing Research Repository (arXiv):1412.0101, 2014.

[80] G. Reeb. Sur les points singuliers d'une forme de Pfaff complètement intégrable ou d'une fonction numérique. *C. R. de L'Académie des Sciences*, 222:847–849, 1946.

[81] M. Schaefer, E. Sedgwick, and D. Štefankovič. Recognizing string graphs in NP. *Journal of Computer and System Sciences*, 67(2):365 – 380, 2003.

[82] M. Schaefer and D. Štefankovič. Decidability of string graphs. In *Proc. 33rd Sympos. Theory Comput. (STOC)*, pages 241–246, 2001.

[83] D. R. Sheehy. Fréchet-stable signatures using persistence homology. In *Proc. 29th Sympos. Discrete Algorithms (SODA)*, pages 1100–1108, 2018.

[84] A. Sotiras, C. Davatzikos, and N. Paragios. Deformable medical image registration: A survey. *IEEE Transactions on Medical Imaging*, 32(7):1153–1190, 2013.

[85] J. Stillwell. *Classical topology and combinatorial group theory*. Springer-Verlag, New York, 1980.

[86] G. Toussaint. An optimal algorithm for computing the relative convex hull of a set of points in a polygon, 1986.

[87] J. van de Koppel, M. Löffler, and T. Ophelders. Computing wave impact in self-organised mussel beds. In *Abstr. 33rd European Workshop on Computational Geometry (EuroCG)*, pages 169–172, 2017.

[88] M. van Kreveld, R. van Oostrum, C. Bajaj, V. Pascucci, and D. Schikore. Contour trees and small seed sets for isosurface traversal. In *Proc. 13th Sympos. Comput. Geom. (SoCG)*, pages 212–220, 1997.

[89] D. Štefankovič. *Algorithms for simple curves on surfaces, string graphs, and crossing numbers*. PhD thesis, University of Chicago, 2005.

[90] B. White. Mappings that minimize area in their homotopy classes. *Journal of Differential Geometry*, 20(2):433–446, 1984.

[91] B. Yang, D. Dyer, and B. Alspach. Sweeping graphs with large clique number. In *Proc. 15th Internat. Sympos. Algorithms Comput. (ISAAC)*, pages 908–920, 2004.

Summary

Continuous Similarity Measures for Curves and Surfaces

Distinguishing objects is a crucial task for survival. Animals survive by distinguishing predators from prey, and poisonous plants from edible ones. As such, objects come in different types, and to recognize the type of an object, one needs a description of it. In theory, one would best describe it by listing all objects of that type. However, there tend to be many objects of a given type, and sometimes, objects of the same type may not have been discovered.

A more practical way to describe a type would be to provide a collection of objects of that type, complemented by a collection of objects not of that type. For an object of a given type, we expect that it is similar to the objects of the same type, and dissimilar from other objects. To test whether an object is of a given type, one can compare it to other objects that do and do not fit the description of that type. The similarity of objects can be measured in many different ways, and different applications tend to have different similarity measures that are suitable.

In this thesis, we investigate computational aspects of various similarity measures between curves, as well as surfaces. At the cost of accuracy, we give an intuition behind the similarity measures studied, and provide example applications for each of them. More formal definitions can be found in Chapter 1.

Image Deformation

Suppose two people take a picture of the same scene. It is very unlikely that the pictures look exactly the same: the position of the camera and the lighting could be different, one lens could distort one picture differently than the other, and perhaps some objects have moved slightly. However, one can expect that one picture looks very similar to a deformed version of the other picture. We investigate the computation of such deformations in Chapter 2.

Simple Curve Deformation

Curve-like objects such as snakes and rivers change shape over time. In Chapter 3, we study deformations between curves that prevent curves during the deformation from overlapping with themselves, since solid objects tend to not intersect themselves. Among such deformations, we aim to find one that uses as little movement as possible.

Gift Wrapping

Consider a roll of gift-wrapping paper. We want to determine whether the roll is wide enough to wrap a given object. More accurately, we want to cut out a region of paper such that we can tape one side of the paper to the other in such a way that no air is left trapped between the gift-wrapping paper and the object. Alternatively, one can consider a tubular object, and ask how much a rubber band must stretch to traverse the object. In Chapters 4 and 5, we essentially aim to compute the amount a rubber band needs to stretch, or the necessary width of a roll of gift-wrapping paper to wrap a given object.

Curriculum Vitæ

Tim Ophelders was born on the 28th of August 1991 in Geleen, the Netherlands. He finished his pre-university secondary education in 2009 at the Connect College in Echt, the Netherlands. In 2012, he received his Bachelor of Science degree in Computer Science from the Eindhoven University of Technology. He continued his studies at the same university. He received his Master's degree (*cum laude*) in Computer Science within the Algorithms group. During his studies, he participated in the Master's Honors Programme, consisting of a project on crowdsourced visualization and a project on detecting patterns in movement data, the results of which were later published. His Master's thesis on computing the similarity between moving curves was nominated for the TU/e academic awards, and the results were later published. In September 2014, he started his PhD project on computing similarity between complex objects at the Eindhoven University of Technology under the supervision of prof. dr. Bettina Speckmann. In addition to the results in this dissertation, the research conducted during his PhD has led to several results in geographic information science, combinatorial optimization, and the complexity of puzzles.

Titles in the IPA Dissertation Series since 2015

G. Alpár. *Attribute-Based Identity Management: Bridging the Cryptographic Design of ABCs with the Real World*. Faculty of Science, Mathematics and Computer Science, RU. 2015-01

A.J. van der Ploeg. *Efficient Abstractions for Visualization and Interaction*. Faculty of Science, UvA. 2015-02

R.J.M. Theunissen. *Supervisory Control in Health Care Systems*. Faculty of Mechanical Engineering, TU/e. 2015-03

T.V. Bui. *A Software Architecture for Body Area Sensor Networks: Flexibility and Trustworthiness*. Faculty of Mathematics and Computer Science, TU/e. 2015-04

A. Guzzi. *Supporting Developers' Teamwork from within the IDE*. Faculty of Electrical Engineering, Mathematics, and Computer Science, TUD. 2015-05

T. Espinha. *Web Service Growing Pains: Understanding Services and Their Clients*. Faculty of Electrical Engineering, Mathematics, and Computer Science, TUD. 2015-06

S. Dietzel. *Resilient In-network Aggregation for Vehicular Networks*. Faculty of Electrical Engineering, Mathematics & Computer Science, UT. 2015-07

E. Costante. *Privacy throughout the Data Cycle*. Faculty of Mathematics and Computer Science, TU/e. 2015-08

S. Cranen. *Getting the point — Obtaining and understanding fixpoints in model checking*. Faculty of Mathematics and Computer Science, TU/e. 2015-09

R. Verdult. *The (in)security of proprietary cryptography*. Faculty of Science, Mathematics and Computer Science, RU. 2015-10

J.E.J. de Ruiter. *Lessons learned in the analysis of the EMV and TLS security protocols*. Faculty of Science, Mathematics and Computer Science, RU. 2015-11

Y. Dajsuren. *On the Design of an Architecture Framework and Quality Evaluation for Automotive Software Systems*. Faculty of Mathematics and Computer Science, TU/e. 2015-12

J. Bransen. *On the Incremental Evaluation of Higher-Order Attribute Grammars*. Faculty of Science, UU. 2015-13

S. Picek. *Applications of Evolutionary Computation to Cryptology*. Faculty of Science, Mathematics and Computer Science, RU. 2015-14

C. Chen. *Automated Fault Localization for Service-Oriented Software Systems*. Faculty of Electrical Engineering, Mathematics, and Computer Science, TUD. 2015-15

S. te Brinke. *Developing Energy-Aware Software*. Faculty of Electrical Engineering, Mathematics & Computer Science, UT. 2015-16

R.W.J. Kersten. *Software Analysis Methods for Resource-Sensitive Systems*. Faculty of Science, Mathematics and Computer Science, RU. 2015-17

J.C. Rot. *Enhanced coinduction*. Faculty of Mathematics and Natural Sciences, UL. 2015-18

M. Stolikj. *Building Blocks for the Internet of Things*. Faculty of Mathematics and Computer Science, TU/e. 2015-19

D. Gebler. *Robust SOS Specifications of Probabilistic Processes*. Faculty of Sciences, Department of Computer Science, VUA. 2015-20

M. Zaharieva-Stojanovski. *Closer to Reliable Software: Verifying functional behaviour of concurrent programs*. Faculty of Electrical Engineering, Mathematics & Computer Science, UT. 2015-21

R.J. Krebbers. *The C standard formalized in Coq*. Faculty of Science, Mathematics and Computer Science, RU. 2015-22

R. van Vliet. *DNA Expressions – A Formal Notation for DNA*. Faculty of Mathematics and Natural Sciences, UL. 2015-23

S.-S.T.Q. Jongmans. *Automata-Theoretic Protocol Programming*. Faculty of Mathematics and Natural Sciences, UL. 2016-01

S.J.C. Joosten. *Verification of Interconnects*. Faculty of Mathematics and Computer Science, TU/e. 2016-02

M.W. Gazda. *Fixpoint Logic, Games, and Relations of Consequence*. Faculty of Mathematics and Computer Science, TU/e. 2016-03

S. Keshishzadeh. *Formal Analysis and Verification of Embedded Systems for Healthcare*. Faculty of Mathematics and Computer Science, TU/e. 2016-04

P.M. Heck. *Quality of Just-in-Time Requirements: Just-Enough and Just-in-Time*. Faculty of Electrical Engineering, Mathematics, and Computer Science, TUD. 2016-05

Y. Luo. *From Conceptual Models to Safety Assurance – Applying Model-Based Techniques to Support Safety Assurance*. Faculty of Mathematics and Computer Science, TU/e. 2016-06

B. Ege. *Physical Security Analysis of Embedded Devices*. Faculty of Science, Mathematics and Computer Science, RU. 2016-07

A.I. van Goethem. *Algorithms for Curved Schematization*. Faculty of Mathematics and Computer Science, TU/e. 2016-08

T. van Dijk. *Sylvan: Multi-core Decision Diagrams*. Faculty of Electrical Engineering, Mathematics & Computer Science, UT. 2016-09

I. David. *Run-time resource management for component-based systems*. Faculty of Mathematics and Computer Science, TU/e. 2016-10

A.C. van Hulst. *Control Synthesis using Modal Logic and Partial Bisimilarity – A Treatise Supported by Computer Verified Proofs*. Faculty of Mechanical Engineering, TU/e. 2016-11

A. Zawedde. *Modeling the Dynamics of Requirements Process Improvement*. Faculty of Mathematics and Computer Science, TU/e. 2016-12

F.M.J. van den Broek. *Mobile Communication Security*. Faculty of Science, Mathematics and Computer Science, RU. 2016-13

J.N. van Rijn. *Massively Collaborative Machine Learning*. Faculty of Mathematics and Natural Sciences, UL. 2016-14

M.J. Steindorfer. *Efficient Immutable Collections*. Faculty of Science, UvA. 2017-01

W. Ahmad. *Green Computing: Efficient Energy Management of Multiprocessor Streaming Applications via Model Checking*. Faculty of Electrical Engineering, Mathematics & Computer Science, UT. 2017-02

D. Guck. *Reliable Systems – Fault tree analysis via Markov reward automata*. Faculty of Electrical Engineering, Mathematics & Computer Science, UT. 2017-03

H.L. Salunkhe. *Modeling and Buffer Analysis of Real-time Streaming Radio Applications Scheduled on Heterogeneous Multiprocessors*. Faculty of Mathematics and Computer Science, TU/e. 2017-04

A. Krasnova. *Smart invaders of private matters: Privacy of communication on the Internet and in the Internet of Things (IoT)*. Faculty of Science, Mathematics and Computer Science, RU. 2017-05

A.D. Mehrabi. *Data Structures for Analyzing Geometric Data*. Faculty of Mathematics and Computer Science, TU/e. 2017-06

D. Landman. *Reverse Engineering Source Code: Empirical Studies of Limitations and Opportunities*. Faculty of Science, UvA. 2017-07

W. Lueks. *Security and Privacy via Cryptography – Having your cake and eating it too*. Faculty of Science, Mathematics and Computer Science, RU. 2017-08

A.M. Şutîi. *Modularity and Reuse of Domain-Specific Languages: an exploration with MetaMod*. Faculty of Mathematics and Computer Science, TU/e. 2017-09

U. Tikhonova. *Engineering the Dynamic Semantics of Domain Specific Languages*. Faculty of Mathematics and Computer Science, TU/e. 2017-10

Q.W. Bouts. *Geographic Graph Construction and Visualization*. Faculty of Mathematics and Computer Science, TU/e. 2017-11

A. Amighi. *Specification and Verification of Synchronisation Classes in Java: A Practical Approach*. Faculty of Electrical Engineering, Mathematics & Computer Science, UT. 2018-01

S. Darabi. *Verification of Program Parallelization*. Faculty of Electrical Engineering, Mathematics & Computer Science, UT. 2018-02

J.R. Salamanca Tellez. *Coequations and Eilenberg-type Correspondences*. Faculty of Science, Mathematics and Computer Science, RU. 2018-03

P. Fiterău-Broştean. *Active Model Learning for the Analysis of Network Protocols*. Faculty of Science, Mathematics and Computer Science, RU. 2018-04

D. Zhang. *From Concurrent State Machines to Reliable Multi-threaded Java Code*. Faculty of Mathematics and Computer Science, TU/e. 2018-05

H. Basold. *Mixed Inductive-Coinductive Reasoning Types, Programs and Logic*. Faculty of Science, Mathematics and Computer Science, RU. 2018-06

A. Lele. *Response Modeling: Model Refinements for Timing Analysis of Runtime Scheduling in Real-time Streaming Systems*. Faculty of Mathematics and Computer Science, TU/e. 2018-07

N. Bezirgiannis. *Abstract Behavioral Specification: unifying modeling and programming*. Faculty of Mathematics and Natural Sciences, UL. 2018-08

M.P. Konzack. *Trajectory Analysis: Bridging Algorithms and Visualization*. Faculty of Mathematics and Computer Science, TU/e. 2018-09

E.J.J. Ruijters. *Zen and the art of railway maintenance: Analysis and optimization* *of maintenance via fault trees and statistical model checking*. Faculty of Electrical Engineering, Mathematics & Computer Science, UT. 2018-10

F. Yang. *A Theory of Executability: with a Focus on the Expressivity of Process Calculi*. Faculty of Mathematics and Computer Science, TU/e. 2018-11

L. Swartjes. *Model-based design of baggage handling systems*. Faculty of Mechanical Engineering, TU/e. 2018-12

T.A.E. Ophelders. *Continuous Similarity Measures for Curves and Surfaces*. Faculty of Mathematics and Computer Science, TU/e. 2018-13