Modes of Gaussian Mixtures and an Inequality for the Distance Between Curves in Space

by

Brittany Terese Fasy

Department of Computer Science
Duke University

Date: _____

Approved:

_____
Herbert Edelsbrunner, Advisor

_____
Hubert L. Bray

_____
John Harer

_____
Carlo Tomasi

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Computer Science
in the Graduate School of Duke University
2012

Abstract

(Computational Topology)

Modes of Gaussian Mixtures and an Inequality for the Distance Between Curves in Space

by

Brittany Terese Fasy

Department of Computer Science
Duke University

Date: _____

Approved:

_____
Herbert Edelsbrunner, Advisor

_____
Hubert L. Bray

_____
John Harer

_____
Carlo Tomasi

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Computer Science in the Graduate School of Duke University
2012

Copyright © 2012 by Brittany Terese Fasy
All rights reserved except the rights granted by the
Creative Commons Attribution-Noncommercial Licence

Abstract

This dissertation studies high-dimensional problems from a low-dimensional perspective. First, we bound the difference between two rectifiable curves in high-dimensional space by using the Fréchet distance between and total curvatures of the two curves. We create this bound by mapping the curves into $\mathbb{R}^3$ while preserving the lengths of the curves and increasing neither the total curvature of the curves nor the Fréchet distance between them. The bound is independent of the dimension of the ambient Euclidean space, it improves upon a bound by Cohen-Steiner and Edelsbrunner for dimensions greater than three and it generalizes a result by Fáry and Chakerian.

In the second half of the dissertation, we analyze Gaussian mixtures. In particular, we consider the sum of $n+1$ identical isotropic Gaussians, where each Gaussian is centered at the vertex of a regular n-simplex. We prove that all critical points are located on one-dimensional lines (axes) connecting barycenters of complementary faces of the simplex. Fixing the width of the Gaussians and varying the diameter of the simplex from zero to infinity by increasing a parameter that we call the scale factor, we find the window of scale factors for which the Gaussian mixture has more modes, or local maxima, than components. We analyze these modes using the one-dimensional axes that contain the critical points. We see that the extra mode created is subtle, but becomes more pronounced as the dimension increases.

I dedicate this dissertation to my grandmother, Joan C. Fasy, as well as to my niece, Peyton Gertrude Grady.

Contents

Abstract **iv**

List of Tables **ix**

List of Figures **x**

Acknowledgements **xiii**

1 Introduction **1**

1.1 Contributions 2

2 Preliminaries **5**

2.1 Computational Topology 5

2.1.1 Preliminaries 6

2.1.2 Homology 12

2.1.3 Persistent Homology 13

2.2 Metrics 20

2.2.1 Direct Metrics 20

2.2.2 Approximation of the Fréchet Distance 22

2.2.3 Metrics on Persistence Diagrams 30

2.3 Volumes of Spheres and Balls 35

2.3.1 The Slab Technique 36

2.3.2 Volume of the Unit Ball 38

2.3.3 Volume of the Sphere 39

2.3.4 Ratio of Volumes of Spheres 40

3 Lengths and Distances **41**

3.1 History 41

3.2 Curves and Inscribed Polygons 43

3.2.1 Angles and Distances 43

3.2.2 Length in the Limit 44

3.2.3 Total Curvature 45

3.2.4 Fréchet Distance in the Limit. 46

3.2.5 Integral Geometry Interpretations 47

3.3 Original Length Bound 48

3.4 Dimension-Independent Length Bound 51

3.5 Discussion 55

4 Geometry and Topology of Gaussian Mixtures **56**

4.1 Curve Analysis 57

4.2 Scale Space 64

4.2.1 A Fixed Scale 64

4.2.2 An Axiomatic View of Scale Space 71

4.2.3 No Ghost Modes in $\mathbb{R}^1$ 75

4.2.4 Iterative Algorithm 77

5 Ghosts in Gaussian Mixtures **81**

5.1 Simplex Design 81

5.2 Analysis 84

5.2.1 Lines of Critical Points 85

5.2.2 Transitions 89

5.2.3 One-Dimensional Sections 94

5.2.4 n-Dimensional Sections 95
5.2.5 Resilient Modes 103
5.3 Many Ghosts 106
5.4 Numerical Evidence 108
5.4.1 Persistence of the Ghost Features 108
5.4.2 Other Kernels 110
5.5 Discussion 111
A Notations **114**
B Scale Factor Look-up Tables **117**
Bibliography **117**
Biography **128**

List of Tables

2.1 Volumes of Low Dimensional Balls and Spheres. 37

A.1 Mathematical Symbols Introduced in Chapter 2. 115

A.2 Mathematical Symbols Introduced in Chapter 3. 115

A.3 Mathematical Symbols Introduced in Chapter 4. 116

A.4 Mathematical Symbols Introduced in Chapter 5. 116

B.1 Upper Transition Scale Factor 117

B.2 Lower Transition Scale Factor 118

B.3 Balancing Scale Factor 119

List of Figures

2.1 Chain Groups of a Triangle 10

2.2 Persistence Diagram 15

2.3 Polygonal Chains with Small Hausdorff Distance 22

2.4 Free Space Diagram for Two Polygonal Curves 24

2.5 Discrete Fréchet Distance Bounds Fréchet Distance 28

2.6 Slab Technique 36

2.7 Shelling Technique 39

3.1 2D Caliper 43

3.2 Inscribed Polygon with Turning Angle 45

3.3 Hyperplane Subdivides Curve 48

3.4 Indicator Function 49

3.5 Inductively Adding Vertices 51

3.6 Visualization of Induction 53

4.1 Unit Gaussian Kernel 58

4.2 Ratio Function 61

4.3 Sums of Gaussian Kernels 63

4.4 Disappearing Bridge 70

4.5 Epanechnikov Ghosts 71

4.6 Neighborhood 78

5.1 2-Design 82

5.2 Distance Coordinates 86
5.3 Strict Convexity 88
5.4 Ghost Lifespan 90
5.5 Transitions 91
5.6 Area Between Curves 91
5.7 Chandelier 101
5.8 Transitions 103
5.9 Persistence of Gaussian Mixtures 109
5.10 n-Design with Various Kernels 112

List of Algorithms

1 Left-Right Persistence Algorithm 19
2 Spectral Sequence Persistence Algorithm 19
3 Computing the Discrete Fréchet Distance for Polygonal Chains 26

Acknowledgements

The person to whom I owe the most gratitude is my future husband, David L. Millman. He has selflessly supported me from the first day of graduate school. Thank you, Dave!, for all of your love and support.

The next person I would like to thank is my adviser, Herbert Edelsbrunner. He has supported me in my endeavors, introduced me to many researchers in our field, and always provided honest feedback on my writing, on my presentations, as well as on my research. I thank you, Herbert, for all of your support and guidance.

I also thank my other committee members: Hubert Bray, John Harer and Carlo Tomasi. Thank you for your discussions that have helped to develop my dissertation and for the time that you have committed to serving on my committee. I am grateful for the support and resources made available to me from Duke University and from IST Austria. I thank Pankaj Agarwal, Patrizio Frosini, Sayan Mukherjee, Caroline Uhler and Jeff Phillips for their valuable discussions.

I thank my undergraduate institution, Saint Joseph's University, for preparing me for graduate school. Thank you to Steve Cooper for encouraging me to study computer science. Thank you to Sam Smith for introducing me to topology and patiently working with me on my undergraduate thesis. Thank you to Paul Klingsberg for your constant words of encouragement and for all of those dim-sum lunches in Chinatown.

I thank my family, the *crazy Fasys*, who lovingly support me, regardless of where

in the world I am. I especially thank my grandparents: Edward J. Ryan, Jr., Gertrude T. Ryan, Jack Fasy and Joan C. Fasy. I would not be where I am today without them. I thank my parents, Joseph F. X. and Terese G. Fasy for their love and encouragement. To my siblings, Joey, Dana and Devon: thank you for always believing in me. I thank my godfather Bill Fasy for his constant encouragement. I also thank my future family: Ron Millman, Merri Millman and Lisa Millman. Thank you for welcoming me into your family with open arms.

Finally, I thank my friends and colleagues who have helped me develop academically as well as personally. Those who have supported me are not few in numbers, so I apologize in advance if your name is not mentioned here. Natalie Soroka has been my best friend since we were five years old. She proudly displays my papers (as well as Dave!'s) in her office. I also thank Natalie's parents, *Ma and Pa Soroka*, who have always treated me like a daughter. I thank Diana Zalewski, who brings enthusiasm wherever she goes. I thank Jessica Klusek, who is always ready for a study date. I am indebted to Kathleen Vaccaro, who is always smiling, regardless of how much sleep she has had. Thank you to my friends at Duke and at UNC–Robby Cochran, Eduardo Cuervo, Herodotos Herodotou, Laura Johnson, Srinivas Krishnan, Alana Libonati, Justin Manweiler, Elizabeth Munch, Catie Welsh and Brian Williams–who have not forgotten me, despite me living across the ocean. Thank you to Darrell Bethea for always ensuring that, at any hour of the day, there was someone awake and available to talk with me. Thank you to Frank Miller for not doubting my ability to ski down a mountain, even when every direction is up. Amit Patel became my best "European" friend when we moved to Vienna together in 2009. He has inspired me to work harder and to ski more often. Bei Wang has taught me the ropes of networking at conferences. I thank Chao Chen, who is always enthusiastic about research, even at the gym. I thank Michael Kerber and his wife Michaela. To them, I have just one word: *böse*. I thank Olga Symonova for allowing me to think

out loud in the office. I thank Susanna Ricco for not complaining when our office suddenly became a social gathering room. To my friends in Austria–Filip Korč, Novi Quadrianto, Viktoriia Sharmanska and the girls of the Vienna Cherokees–thank your for making me feel at home in a foreign land. I am grateful for my friendship with Morten Bojsen-Hansen. I am in awe of his enthusiasm for learning. Thank you, Morten, for always being ready for an adventure, whether it be learning topology or running to another country.

1

Introduction

In the introduction to [39], Istvan Fáry writes, *Tous ces problèmes peuvent être ramenés aux questions concernant des courbes planes*, meaning, *All of these problems* [*regarding the total curvature of knots in* $\mathbb{R}^n$] *can be reduced to questions concerning planar curves.* While all high-dimensional problems cannot be reduced to problems in the plane, this dissertation is modeled after Fáry's insight. The theme seen throughout this dissertation is the reduction of a high-dimensional problem to a low-dimensional one. By reducing the dimensionality, we may exploit the more well-known and easier to comprehend theorems of lower-dimensional space.

After introducing persistent homology, the Fréchet distance, and select theorems regarding the volumes of spheres in Chapter 2, the first high-dimensional problem that we investigate is finding a dimension-independent bound for the difference of lengths of curves. We first introduce and prove a dimension-dependent inequality; see Theorem 19. Then, we map a curve embedded in $\mathbb{R}^n$ down to $\mathbb{R}^3$, while preserving certain properties, allowing us to employ Theorem 19 in $\mathbb{R}^3$.

The second high-dimensional problem that we investigate is the geometry and topology of Gaussian mixtures defined over $\mathbb{R}^n$. Specifically, we analyze the modality

of mixtures, beginning with a curve analysis on the sum of two Gaussians in $\mathbb{R}^1$. Not surprisingly, we find that a Gaussian mixture in $\mathbb{R}^1$ cannot have more modes, or local maxima, than components; see Theorem 49. For this reason, we say that Gaussian mixtures in $\mathbb{R}^1$ behave nicely; on the other hand, Gaussian mixtures in $\mathbb{R}^n$ can have many more modes than components. We seek to better understand the phenomenon of having more modes than components, as it is in general not well understood. Associating a unique mode to a component of the mixture by assigning the mode closest to component center, we call any unassociated modes *ghost modes* as they seem to appear from nowhere. Chapter 5 aims at analyzing a family of Gaussian mixtures in which a ghost mode appears. In particular, we give a detailed analysis of the sum of $n+1$ identical isotropic Gaussian kernels placed at the vertices of a regular n-simplex in $\mathbb{R}^n$. We prove that all critical points lie on the 1-dimensional symmetry axes of the n-simplex. Restricting the n-dimensional Gaussian mixture to one of these axes creates a 1-dimensional Gaussian mixture; hence, we can use our curve analysis of 1-dimensional Gaussian mixtures to determine the location of the critical points in $\mathbb{R}^n$, leading up to Theorem 57, where we count and classify the critical points and modes of our Gaussian mixture.

1.1 Contributions

We articulate the new contributions of this dissertation, contrasting the contributions to the related work in the field.

Chapter 2. Section 2.1 defines persistent homology and highlights a few known lemmas that will be used later in the dissertation. In Section 2.2, we define the Hausdorff distance, the Gromov-Hausdorff distance, the Fréchet distance, the bottleneck distance and the Wasserstein distance. We provide an extended discussion of the Fréchet distance, in which we also define the discrete Fréchet distance from [37].

The new contribution of this section is Lemma 6, in which we prove an upper bound on the difference between the Fréchet distance between two curves and the discrete Fréchet distance between polygons inscribed in the curves. This lemma is a generalization of a result in [37] which bounds the difference between the Fréchet distance and the discrete Fréchet distance between polygons. Finally, in Section 2.3, we derive the formulas for the volume of the n-ball and of the n-sphere. We follow the proof presented in [74], using slightly different terminology to ease the exposition. We conclude this chapter with Theorem 17, which states that the ratio of the volume of the $(n-1)$-sphere to the volume of the n-sphere grows like the square root of n. Although the proof of this theorem falls directly from the formula for the volume of a sphere, we believe that the result is not well known, as we have not found the proof of nor a statement of the theorem elsewhere.

Chapter 3. This focus of this chapter is the proof of Theorem 20, which bounds the difference of lengths of curves by a function of the Fréchet distance between the curves and the total curvatures of the curves. This result was published in [43]. A similar result was proven in [20] but depends on the dimension of the ambient space in which the curves are embedded. We prove both theorems in this chapter, noting that we simplified the proof of [20]. Theorem 20 removes the dependency on the ambient dimension and improves upon the result in [20] for all dimensions greater than three. To prove this theorem, we use the fact that the lengths of, the total curvatures of and the Fréchet distance between two rectifiable curves can be approximated by the lengths of, the total curvatures of and the Fréchet distance between of inscribed polygons; see Lemmas 23, 24 and 25, respectively. The first two lemmas are well-known in the literature, see for example [75], and Lemma 25 follows from the Fréchet distance being a metric.

Chapter 4. Chapter 4 defines Gaussian kernels and Gaussian mixtures as well as provides a few results necessary for Chapter 5. There are no new results in this section, but we do provide our own proofs of the conditions for which a mixture of two Gaussians has one or two modes in Section 4.1. The conditions were first given in [12] for the balanced sum of Gaussians in $\mathbb{R}^1$ and in [9] for the weighted sum of Gaussians.

Chapter 5. In this chapter, we analyze the conditions for which a Gaussian mixture has more modes than components. For the most part, the entirety of this chapter is a new contribution based on the results that we presented in [31]. The fact that a Gaussian mixture can have more modes than components was first observed in [14]. In Section 5.1, we formalize the design that they used, calling it the standard n-design. We expand upon the observation that an extra mode exists: We prove that all critical points are located on the axes of the n-simplex in Lemma 52, find the scale factors for which the axes witness two 1-dimensional local maxima in Subsection 5.2.2, characterize for which of these maxima we have modes of the mixture Lemma 53, give conditions for which the barycenter of the n-simplex has a mode at the barycenter in Lemma 54, characterize the indices of all critical points in Subsection 5.2.4, and prove that the resilience of the extra mode (interval of scale factors for which it exists) grows like the square root of the dimension in Subsection 5.2.5. In Section 5.4, we investigate the persistence of the extra mode empirically and look at examples of the 2-design using different kernels, highlighting the limits of our proof technique in extending to other kernels.

2

Preliminaries

This chapter covers three topics: computational topology, metrics and properties of spheres. As these two topics will play important roles throughout this dissertation, we present the relevant concepts and results in this chapter.

First, I give a brief review of the necessary background of computational topology, but refer the reader to [32, 33, 50] for more details. In Chapter 3, we use persistence diagrams in the proof of the Length Theorem of [20]. This theorem is then used in the proof of the Improved Length Theorem originally published in [43]. In Chapter 4, I use the persistence of (ghost) features to quantify the importance the feature plays in a probability distribution. Appendix A contains a table summarizing the notation used in this document.

2.1 Computational Topology

Observing patterns and features in data sets is a common goal in many disciplines. However, extracting the key features from a noisy data set can be an ambiguous task and often involves simplifying and finding the best view of the data. Computational topology, and more specifically persistent homology, is a set of tools used

for data analysis. Persistent homology studies topological spaces endowed with a function. Persistence describes the function with a multiset of points in the plane, each corresponding to a topological feature of the function.

2.1.1 Preliminaries

Before defining persistence homology, we define manifolds and simplicial complexes, which can be used to approximate manifolds in computations.

Manifolds. A *paracompact* space is a topological space such that every open cover has a locally finite refinement. A *Hausdorff* space is a topological space such that for every two points $a \neq b$, there exist disjoint open sets A and B, where A contains a and excludes b, and B contains b and excludes a. An *m-manifold* $\mathbb{M}$ is a paracompact Hausdorff space such that each point is contained in a neighborhood homeomorphic to $\mathbb{R}^m$. A *Riemannian manifold* is a manifold $\mathbb{M}$ equipped with an inner product on the tangent space at every point p in $\mathbb{M}$:

$$\langle \cdot, \cdot \rangle_p \colon T_p\mathbb{M} \times T_p\mathbb{M} \to \mathbb{R} \tag{2.1}$$

that satisfies the following properties given smooth injective maps x_α, x_β from $\mathbb{R}^m$ into $\mathbb{M}$:

1. $x_\beta^{-1} x_\alpha \colon \mathbb{R}^m \to \mathbb{R}^m$ is smooth.

2. Let dx_α denote the linear map induced on the tangent spaces by x_α. Given differentiable maps $\varphi_i, \varphi_j \colon \mathbb{R}^m \to \mathbb{R}^m$, the map $g \colon \mathbb{R}^m \to \mathbb{R}$ defined by

$$g(p) \quad = \quad \langle dx_\alpha \varphi_i, dx_\alpha \varphi_j \rangle_{x_\alpha(p)}$$

   has defined and continuous partial derivatives of all orders.

We summarize the relationships among $\mathbb{M}$, $\mathbb{R}^m$, and $T\mathbb{M} = \cup_{p\in\mathbb{M}} T_p\mathbb{M}$ in the following diagram:

$$\begin{array}{ccc} \mathbb{M} & & T\mathbb{M} \\ {\scriptstyle x_\alpha, x_\beta}\uparrow & & \uparrow{\scriptstyle dx_\alpha} \\ \mathbb{R}^m & \xrightarrow{\varphi_i, \varphi_j} & \mathbb{R}^m \end{array}$$

When these conditions are satisfied, the inner product of (2.1) is called the *Riemannian metric*, and for each $p \in \mathbb{M}$, there exists an *essential matrix* or *fundamental matrix* such that

$$\langle v, w \rangle_p = v^T E w. \tag{2.2}$$

The Riemannian metric can be used to define lengths, areas, and angles on the manifold; see [28]. We focus on the length of a *path* γ, which is defined as a continuous map from the unit interval $[0, 1]$ to $\mathbb{M}$, i.e., $\gamma\colon [0, 1] \to \mathbb{M}$. The length of γ is found by integrating the magnitude of the velocity vector:

$$\begin{aligned} \ell(\gamma) &= \int_0^1 ||\gamma'(t)||\, dt, \text{ where} \\ ||\gamma'(t)||^2 &= \langle \gamma'(t), \gamma'(t) \rangle_{\gamma(t)}. \end{aligned} \tag{2.3}$$

The *geodesic distance* between any two points $x, y \in \mathbb{M}$, denoted $d_{\mathbb{M}}(x, y)$, is the length of the shortest path connecting x to y in $\mathbb{M}$:

$$d_{\mathbb{M}}(x, y) = \inf_{\alpha\colon I \to \mathbb{M}} \ell(\alpha), \tag{2.4}$$

where $\alpha(0) = x$ and $\alpha(1) = y$. With this distance function, every connected Riemannian manifold $\mathbb{M}$ becomes a *metric space*, which we denote by $(\mathbb{M}, d_{\mathbb{M}})$.

Definition 1 (Metric Space)**.** The function $d_{\mathbb{M}}\colon \mathbb{M} \times \mathbb{M} \to [0, \infty]$ is a *distance function* or *metric* on a set $\mathbb{M}$ if for all $x, y, z \in \mathbb{M}$, the following three properties hold:

1. Definiteness: $d_{\mathbb{M}}(x, y) = 0$ if and only if $x = y$.

2. Symmetry: $d_{\mathbb{M}}(x, y) = d_{\mathbb{M}}(y, x)$.

3. Triangle Inequality: $d_{\mathbb{M}}(x, y) \leq d_{\mathbb{M}}(x, z) + d_{\mathbb{M}}(z, y)$.

A *metric space* $(\mathbb{M}, d_{\mathbb{M}})$ is a set $\mathbb{M}$ with a distance function $d_{\mathbb{M}}$ defined between any pair of elements [70].

Many real-world objects can be modeled by manifolds or manifolds with boundaries. Manifolds give a clean and concise characterization of a space, such as a 3D object. However, it is often difficult to obtain a mathematical description of the manifold, let alone to perform computations on it. Thus, we use simplicial complexes to approximate the sometimes unknown manifolds or manifolds with boundary.

Curves and Polygons. We investigate a subclass of one-manifolds, loops and polygons. Following the notation used in [75], a *loop* in m-dimensional Euclidean space is a continuous map from the unit circle $\mathbb{S}^1$ to $\mathbb{R}^m$:

$$\gamma \colon \mathbb{S}^1 \to \mathbb{R}^m.$$

In general, we think of a loop as the parametrized image of the circle, but we use the unit interval whenever it is more convenient to parametrize with $I = [0, 1]$. Recall from above that a path is a map $\gamma \colon I \to \mathbb{R}^m$; it is a loop if $\gamma(0) = \gamma(1)$. A *curve* is the image of a loop or a path in Euclidean space.

A *polygonal chain* is a piecewise linear path. A polygonal chain can be specified by a mapping from the discrete interval of length τ to Euclidean m-dimensional space:

$$P \colon I_\tau \to \mathbb{R}^m.$$

The discrete interval I_τ is the set of $\tau + 1$ rational numbers: $\{0, \frac{1}{\tau}, \frac{2}{\tau}, \ldots, 1\}$. The polygonal chain P comprises $\tau + 1$ vertices and τ edges in $\mathbb{R}^m$. Each edge e_i defines a linear interpolation between $P(\frac{i}{\tau})$ and $P(\frac{i+1}{\tau})$. A *polygon* is a closed polygonal chain, i.e., a polygonal chain such that $P(0) = P(1)$.

We say that a curve γ is *k-bounded* if for every $0 < \tau_i < \tau_j < 1$, the sub-curve from $p_i = \gamma(\tau_i)$ to $p_j = \gamma(\tau_j)$ is contained within the union of the balls centered at p_i and p_j with radius k times one half the distance between p_i and p_j:

$$\gamma_{p_i,p_j} \subseteq B_{\frac{k}{2}||p_i-p_j||}(p_i) \cup B_{\frac{k}{2}||p_i-p_j||}(p_j).$$

By this definition, if p_i and p_j coincide, then the path between p_i and p_j necessarily remains at p_i. We will see in Section 2.2.1 that the Fréchet distance between two k-bounded curves is upper bounded by $k+1$ times the Hausdorff distance between the two vertex sets.

A curve γ is called *c-packed* if the total length of the sub-curves contained inside any ball is bounded by c times the radius of that ball. Unlike k-bounded curves, c-packed curves can self intersect and are closed under concatenation [29]. A k-bounded curve does not need to be c-packed, as a k-bounded curve can be arbitrarily long in a ball of a given diameter.

Simplicial Complexes. An *n-simplex* is the convex hull of $n+1$ vertices such that the affine space spanned by the set of difference vectors is an n-dimensional plane. For example, a 1-simplex is a line segment and a 2-simplex is a triangle; see Figure 2.1. If σ is an n-simplex, then any subset of the vertices of σ defines a *face* of σ; we use $\tau \leq \sigma$ to denote this relationship. If τ contains exactly n vertices, then it is a *co-dimension* 1 *face*. A *simplicial complex* K is a collection of simplices such that each face of simplex $\sigma \in K$ is also in K and the intersection of two simplices in K is either a common face or empty. The dimension of a simplicial complex is the maximum dimension of any simplex contained in it. The underlying topological space of K, denoted $|K|$, is the union of all simplices under the subspace topology. We say that a topological space $\mathbb{M}$ is triangulated by $|K|$ if there exists a homeomorphism between $\mathbb{M}$ and $|K|$.

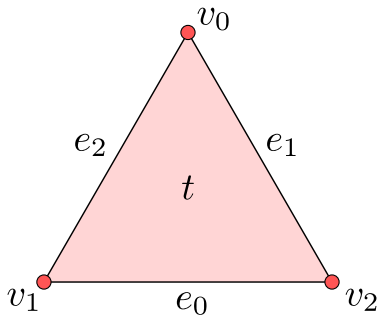


FIGURE 2.1: The chain groups of a triangle are as follows: C_0 is generated by the set of vertices $\{v_0, v_1, v_2\}$, C_1 is generated by the set of edges $\{e_0, e_1, e_2\}$, and C_2 is generated by the single element t. The boundary operator $\partial_p(\alpha)$ returns the boundary of the p-chain α. For example, $\partial_2(t) = e_1 + e_2 + e_3$ and $\partial_1(e_0) = v_1 + v_2$.

We define the *mesh* of a simplicial complex K as the maximum Euclidean distance between any two points that belong to the same simplex. We use $N(r)$ to denote the minimum number of simplices needed to triangulate $\mathbb{M}$ with mesh at most r.

Functions. Often, the space of interest is equipped with a function $f\colon \mathbb{M} \to \mathbb{R}$. If $\mathbb{M}$ is a curve that goes monotonically from left to right, then we can think of f as the height function of its graph, as depicted in Figure 2.2. In the case of images, the function is usually an intensity function. For Riemannian 2-manifolds, we can define functions that are stable with respect to perturbations of the manifold; see e. g. [17] where the authors define a class of maps that are stable with respect to Gromov-Hausdorff perturbations.

There are three special types of functions that we will use: Lipschitz functions, Morse functions, and tame functions. If there exists a real number λ such that for all x, y in $\mathbb{M}$,

$$|f(x) - f(y)| \quad \leq \quad \lambda \cdot d_{\mathbb{M}}(x, y),$$

then f is said to be *Lipschitz*. In this case, λ is called the *Lipschitz constant* of f. If f is a smooth real-valued function over $\mathbb{M}$, then a point $p \in \mathbb{M}$ is a *critical point* if the derivative of f vanishes. The value $f(p)$ is then a *critical value* of f. A critical

point p is *degenerate* if the Hessian matrix, the matrix of second derivatives in local coordinates, is either singular or undefined at p. A *Morse function* is a smooth function, such that no two critical points share a critical value and every critical point is nondegenerate; see [62].

The *sublevel set* of a function is defined by

$$\mathbb{M}_s(f) \quad := \quad f^{-1}((-\infty, s]). \tag{2.5}$$

We sometimes write $\mathbb{M}_s$ if it is understood which function we are using. Similarly, the *superlevel set* is $\mathbb{M}^s(f) = f^{-1}([s, \infty))$. The function f is said to be a *tame function* if there are a finite number of critical values and if the homology groups, which we will define in (2.6), of $\mathbb{M}_s$ have finite rank for all s; see [32].

While we like to think about the continuous case, computations must be done on the discrete representation K. Assuming that a function is given on the vertices of the complex, we extend the function piecewise linearly over $|K|$. In order to define a function on K, a simplex σ can be assigned the value $f(\sigma)$ by taking the supremum of the function values attained over points in σ. However, we are not always given the function on the vertices; sometimes, we must define it ourselves. For example, a function of interest might be the *diameter function* $f_{\mathcal{R}} \colon K \to \mathbb{R}$, where we use the diameter of a simplex as its function value. A simplex is said to have *diameter* 2α if the maximum distance between two points in the simplex is at most 2α.

Filtrations. A *filtration* is a sequence of nested subspaces. For example, suppose that we have a simplicial complex K and that there exists a function f on the simplices of K. To obtain a filtration of K, we order the simplices by the following two rules:

1. If $f(\sigma_1) < f(\sigma_2)$, then σ_1 appears before σ_2.

2. If τ is a face of σ, then τ appears before σ.

We observe that the second rule does not contradict the first, as $f(\tau) \leq f(\sigma)$ whenever τ is a face of σ. An ordering of the simplices following these rules is called a *filter*. We notice that these rules alone do not guarantee uniqueness of the filter; for example, edges e_0 and e_1 have the same function value in Figure 2.1 if the function evaluated at a simplex is the largest index of any vertex comprising that simplex. The implications of changing the order of the simplices in the filter are discussed in [24]. However, the two rules above do imply that every initial subsequence of the filter defines a subcomplex of K. Growing this initial subsequence until it equals the entire filter gives a sequence of simplicial complexes called the *induced filtration*. To construct a filter requires sorting the n simplices of the complex. Thus, the time complexity of constructing a filter is lower bounded by the time complexity of sorting the simplices, which in the case of a comparison-based sorting algorithm is $O(n \log n)$.

For example, we define the *Vietoris-Rips filtration* over the complete simplicial complex $K_{\mathcal{S}}$ on a set of vertices $\mathcal{S}$. For a given real number α, the Vietoris-Rips Complex $\mathcal{R}_\alpha(\mathcal{S})$ is the set of simplices σ in $K_{\mathcal{S}}$ such that $f_{\mathcal{R}}(\sigma) \leq 2\alpha$. The Vietoris-Rips filtration is the sequence of complexes $\{\mathcal{R}_\alpha(\mathcal{S})\}_\alpha$ created by increasing α from zero to infinity; see [82] for the first introduction of this complex and [49] for an extension of the idea.

2.1.2 Homology

Let X be a simplicial complex of dimension d. For $p \in \mathbb{N}$ and $p \leq d$, the symbol X_p denotes the power set of all p-simplices in X. Each set of X_p is called a *p-chain*. The *chain group* C_p is defined by the set X_p under the symmetric difference operation. This operation can be interpreted as addition modulo two. More generally, the group C_p can be defined as the free group on X_p, where the simplices in X_p are oriented and the coefficients are an arbitrary group. To compute persistence, we will need

the group coefficients to be a field; see [87]. For historical reasons and for simplicity, we use $\mathbb{Z}_2$ coefficients. Consider the boundary homomorphism :

$$\partial_p : C_p \to C_{p-1},$$

which maps the p-chain $\alpha \in C_p$ to the boundary of α, a chain in C_{p-1}; see e. g. [50]. Two important subsets of C_p and C_{p-1} are respectively the kernel and the image of ∂_p. The *kernel* of ∂_p is the set of elements in the domain that are evaluated to zero (the empty set) and the image of ∂_{p+1} is the set of elements of the form $\partial_{p+1}(x)$, where x is in the domain C_{p+1}:

$$\mathrm{Ker}(\partial_p) = \{\alpha \in C_p |\ \partial_p(\alpha) = \emptyset\} \text{ and}$$

$$\mathrm{Im}(\partial_{p+1}) = \{\alpha \in C_p |\ \exists \alpha' \in C_{p+1} \text{ with } \partial_{p+1}(\alpha') = \alpha\}.$$

The p^{th} homology group of X, denoted $H_p(X)$, is the kernel of ∂_p modulo the image of ∂_{p+1}:

$$H_p(X) = \mathrm{Ker}(\partial_p)/\mathrm{Im}(\partial_{p+1}). \tag{2.6}$$

As the chain groups are abelian groups, the homology groups are as well. The p^{th} *Betti number*, β_p, is the rank of the p^{th} homology group of X. The *rank* of a group is the smallest number of generators needed to define the group up to isomorphism. Since we are concerned with groups with $\mathbb{Z}_2$ coefficients, the rank uniquely defines the group up to isomorphism. For example, a group with three generators is isomorphic to $\mathbb{Z}_2^3 = \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2$ and a group with k generators is isomorphic to $\mathbb{Z}_2^k$.

2.1.3 Persistent Homology

We define persistent homology for functions from a manifold $\mathbb{M}$ to $\mathbb{R}$. Let f be a tame real valued function defined on a manifold $\mathbb{M}$. We characterize the topology of the sublevel set $\mathbb{M}_s(f)$, and we monitor how the homology groups change as the parameter s grows from negative infinity to positive infinity. Noticing that the

homology of the sublevel sets changes only at critical values of f, say at $c_0 < c_1 < \cdots < c_n$, we choose interleaving values $\{s_i\}$:

$$s_0 < c_0 < s_1 < c_1 < \cdots s_n < c_n < s_{n+1}.$$

Then, the following sequence is a filtration of $\mathbb{M}$:

$$\mathbb{M}_{s_0} \subset \mathbb{M}_{s_1} \subset \cdots \subset \mathbb{M}_{s_{n+1}}.$$

If we apply the homology functor $H_p(\cdot)$ to this sequence, then we obtain a sequence of homology groups connected by homomorphism:

$$0 = H_p(\mathbb{M}_{s_0}) \to H_p(\mathbb{M}_{s_1}) \to \cdots \to H_p(\mathbb{M}_{s_{n+1}}) = H_p(\mathbb{M}).$$

Persistent homology is the study of $H_p(\mathbb{M}_s)$ as s increases.

The zeroth persistent homology group, denoted $H_0(\mathbb{M}_s)$, changes at critical values corresponding to maxima, minima, and (some) index-one saddles of the function f. For a Morse function f, the critical points are the set of $r \in \mathbb{M}$ with $s := f(r)$, such that the sum of Betti numbers changes by exactly one from $\mathbb{M}_{s-\epsilon}$ to $\mathbb{M}_{s+\epsilon}$ for every sufficiently small value of $\epsilon > 0$. If the sum of the Betti numbers increases, we call r a *creating critical point*. If the sum decreases, then r is a *destroying critical point*. As s increases from negative infinity, we label each new component with the value of the creating critical point that introduces the component. We pair each destroying critical point with value s with the most recently discovered unpaired creating critical point representing the components joined at s. This will lead to the diagram $\mathrm{Dgm}_0(f)$.

We demonstrate this construction in Figure 2.2. Consider one pair: the creating critical value at $s = 3.2$ and the destroying critical value at $s = 6.2$. This pair is represented in $\mathrm{Dgm}_0(f)$ as the point $(3.2, 6.2)$. The *persistence* of that pair is equal to the absolute difference in function values: $6.2 - 3.2 = 3$. In Figure 2.2, there

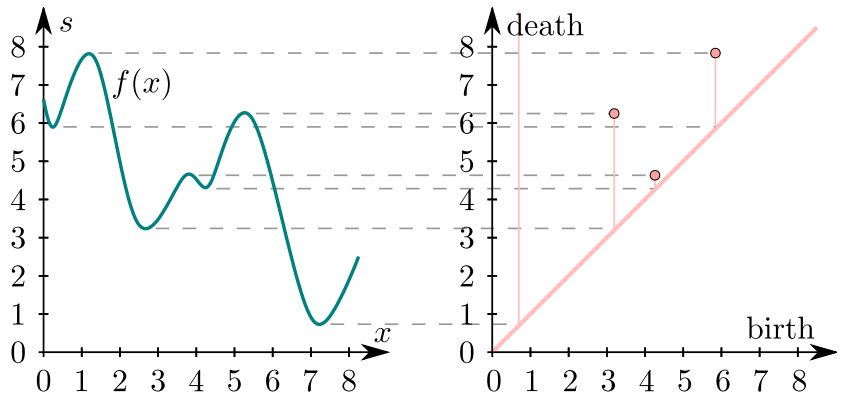


FIGURE 2.2: On the left, we see the graph of a function $f\colon \mathbb{R} \to \mathbb{R}$. On the right is the corresponding 0-dimensional persistence diagram $\mathrm{Dgm}_0(f)$. Each point is drawn with multiplicity one. The birth at the critical point at $x = 7.1$ remains unpaired.

are three points and one creating critical value that remains unmatched. This value represents an *essential* homology class. An essential homology class is nontrivial in the topological space $\mathbb{M}$. It would be paired if we were to consider the extended persistence diagram presented in [22].

We have just discussed the construction of the 0^{th} persistence diagram, which records the births and deaths of homology classes in $H_0(\mathbb{M}_s)$ as s increases. Similarly, the births and deaths of homology classes in $H_p(\mathbb{M}_s)$ are recorded in the p^{th} persistence diagram $\mathrm{Dgm}_p(f)$. While $\mathrm{Dgm}_0(f)$ monitors the birth and death of components of the sublevel sets, $\mathrm{Dgm}_p(f)$ monitors the p-dimensional holes of the sublevel sets. We let $\mathrm{Dgm}(f)$ be the overlay of all persistence diagrams for f.

Although persistence is inspired by Morse Theory [62], it grew into an area of research in the 1990s. Edelsbrunner, Letcher, and Zomorodian defined persistent homology in [34]. Independently, Frosini [46] looked at size functions of submanifolds, which uses different terminology to study the zero-dimensional persistent homology. Cagliari, Ferri, and Pozzi [13] and Robins in [67] extended Frosini's results for general homology groups. In [87], Zomorodian and Carlsson extended the original theory behind persistent homology to a more general framework for field coefficients (as

opposed to $\mathbb{Z}_2$ coefficients). A thorough summary of persistent homology, including results on its stability, can be found in [32, 33]. Today, persistent homology continues to be an active area of research in computational geometry and topology.

Total Persistence. The *degree k total persistence* is the sum of the k^{th} powers of persistence over all points in the persistence diagram:

$$\text{Pers}_k(f) = \sum_{(x,y)\in \text{Dgm}(f)} (y-x)^k \tag{2.7}$$

Sometimes we are interested in the total persistence over all points with persistence exceeding a threshold ϵ, for some $\epsilon \geq 0$. In this case, we use $\text{Pers}_k(f, \epsilon)$ to denote the sum in Equation (2.7) restricted to the points with persistence greater than ϵ. We justify this simplification by interpreting points with persistence less than or equal to ϵ as noise. Similarly, Edelsbrunner, Morozov and Pascucci [35] call a function g a *simplification* of a function f if $||f-g||_\infty$ is small and $\text{Dgm}(g)$ is equal to $\text{Dgm}(f)$ minus the low-persistence points.

Removing too many low-persistence points can significantly effect the total persistence of the diagram. For this reason, we seek to bound the number of points removed from the diagram. For a Lipschitz function f, the number of these points is in fact bounded:

Lemma 2. (Persistent Cycle Lemma [23]). *The number of points in* $\text{Dgm}(f)$ *whose persistence exceeds* ϵ *is at most* $N(\epsilon/\lambda)$*, where* λ *is the Lipschitz constant.*

Recall that $N(\epsilon)$ is the minimum number of simplices needed to triangulate $\mathbb{M}$ with mesh at most ϵ. With these low persistence pairs removed, we turn to bounding the total persistence. The amplitude of f is the maximum difference in function values:

$$\text{Amp}(f) = \max_{(x,y)\in \mathbb{M}\times\mathbb{M}} |f(x) - f(y)|. \tag{2.8}$$

We use the Persistent Cycle Lemma to obtain an upper bound on the degree k total persistence:

Lemma 3. (Moment Lemma [23]).

$$\mathrm{Pers}_k(f,r) \leq r^k N\left(\tfrac{r}{\lambda}\right) + k\int_{\epsilon=r}^{\mathrm{Amp}(f)} N\left(\tfrac{\epsilon}{\lambda}\right)\epsilon^{k-1}\,d\epsilon.$$

Suppose that the size of the triangulation grows only polynomially with the reciprocal of the mesh; that is, there are constants C_0 and M such that $N(r) \leq C_0/r^M$. Applying the Moment Lemma for $k > M$ results in the following inequality:

$$\mathrm{Pers}_k(f,r) \;\leq\; C_0\lambda^M \mathrm{Amp}(f)^{k-M}\frac{2k-M}{k-M}.$$

To interpret this inequality, notice that the right hand side does not change with respect to r. Therefore, we can conclude

$$\mathrm{Pers}_k(f) \;\leq\; C_0\lambda^M \mathrm{Amp}(f)^{k-M}\frac{2k-M}{k-M}.$$

In words, for $k > M$¡ the degree k total persistence is bounded away from infinity.

Left-Right Algorithm. We conclude this discussion by giving two algorithms to compute persistence. The first one is the standard persistence algorithm, originally presented in [34] using a sparse matrix implementation and summarized in [32]. We assume that we are given a filtration on K: $\emptyset = K_0 \subset \cdots \subset K_n = K$, where K_i contains only one more simplex than K_{i-1}. Then, we compute the persistence diagrams, just as we did for the sublevel sets above.

We explain the persistence algorithm in terms of matrix reduction. Then, we label the columns and rows of a matrix D with the simplices in the order in which they appear in the filter. We use σ_i to represent the simplex that is represented in column (and row) i. We create the *boundary matrix* D defined by $D(i,j) = 1$ iff simplex σ_i

is on the boundary of (is a co-dimension one face of) simplex σ_j and all other entries are zero. For example, the boundary matrix corresponding to Figure 2.2 is

$$D = \begin{array}{c} \\ v_0 \\ v_1 \\ v_2 \\ e_2 \\ e_1 \\ e_0 \\ t \end{array} \begin{array}{c} \begin{array}{ccccccc} v_0 & v_1 & v_2 & e_2 & e_1 & e_0 & t \end{array} \\ \left(\begin{array}{ccccccc} 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & \textcircled{1} & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & \textcircled{1} & \textcircled{1} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & \textcircled{1} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right) \end{array}.$$

For clarity, the lowest 1 of each column in D has been circled. We reduce D until each row contains at most one lowest 1. Notice that e_1 and e_0 share a common lowest 1 in the row corresponding to v_2. In the end, if the lowest 1 in the column corresponding to e_2 is in the row corresponding to v_1, then this represents a persistence pairing (v_1, e_2).

In Algorithm 1, we scan the columns from left to right, incrementing through the filter. Let $\text{low}(j)$ denote the index of the lowest 1 in column j. If column j is a zero column, then $\text{low}(j) = -1$. A zero column corresponds to a simplex that represents the birth of a homology class. If the lowest 1 in the column has only zero entries to the left, then we have found a persistence pair. In this case, the column corresponds to the death of a homology class. Otherwise, there exists a column to the left with a lowest 1 in the same row. To resolve this conflict, we add the second column to the first using $\mathbb{Z}_2$ coefficients. Then, we check to the left again. We repeat this until either we find a lowest 1 that is the first one in the row, or we have a column with only zero entries.

Spectral Sequence Algorithm. In Algorithm 2, we scan parallel to the diagonal of the matrix instead of left to right. At stage r, we consider all entries such that the

Algorithm 1 Left-Right Persistence Algorithm

Input: An $n \times n$ boundary matrix D.
Output: Reduced form of D.

```
for j = 1 to n do
  while ∃j′ < j with low(j′) =low(j) ≠ −1 do
    add column j′ to column j
  end while
end for
```

index of the column is equal to r plus the index of the row. For each of these entries $D(i, j)$, we look to see if the value is zero or one and proceed as in the previous algorithm. We ignore zero entries, and for a nonzero entry that is the lowest 1 in column j, we check to see if this entry corresponds to a death. If the row is currently unpaired, i.e. only zeros to the left, then we have found a persistence pair. Otherwise, we add the second column to the first using $\mathbb{Z}_2$ coefficients. Since we have just moved the lowest 1 up (or removed all ones), we do not continue with this column as we will revisit it in a subsequent iteration. Once every column j has been checked, we increment r and repeat. Although this algorithm will result in the same persistence pairings as Algorithm 1, the Spectral Sequence Persistence Algorithm has the advantage that it can be parallelized.

Algorithm 2 Spectral Sequence Persistence Algorithm

Input: An $n \times n$ boundary matrix D.
Output: Normal form of D.

```
for r = 1 to n − 1 do
  for all j ∈ [r + 1, n] do
    if D(j − r, j) = 1 then
      if ∃i′ < i with low(i′) =low(i) ≠ 0 then
        add column i′ to column i
      end if
    end if
  end for
end for
```

Both the Left-Right Algorithm and the Spectral Sequence Algorithm have a running time of $O(n^3)$, where n is the number of simplices in the filtration; however, the algorithms often exhibit near-linear behavior. However, improvements to the algorithms presented here have been made. The zero-dimensional homology can be

computed in $O(n\alpha(n))$ time using union-find, where $\alpha(n)$ is the inverse Ackerman function. Recently, Milosavljevic, Morozov, and Skraba [63] presented an algorithm that runs in matrix-multiplication time, i.e. $O(n^{2.376})$, and Chen and Kerber [19] presented an output-sensitive algorithm to compute persistent homology.

2.2 Metrics

Recalling Definition 1, we give several examples of metrics between topological spaces. In particular, we are interested in computing the distance between two closed curves in n-dimensional Euclidean space. One of these metrics, the Fréchet distance, plays a central role in the proofs presented in Chapter 3.

2.2.1 Direct Metrics

We discuss three metrics: the Hausdorff Distance, the Gromov-Hausdorff Distance, and the Fréchet Distance. We consider these metrics to be *direct* because they measure the distance between spaces based on the inherent shapes. In Section 2.2.3, we look at distance metrics that measure the distance between two functions based on the associated persistence diagrams.

Hausdorff Distance. Let Z be a metric space, with $\mathcal{S}_1, \mathcal{S}_2 \subseteq Z$. One way to measure the distance between these sets is to take the *Hausdorff distance*, denoted $\mathcal{H}(\mathcal{S}_1, \mathcal{S}_2)$, which is the maximum distance (measured in Z) from a point in one set to the closest point in the other set:

$$\mathcal{H}(\mathcal{S}_1, \mathcal{S}_2) \quad = \quad \max\{\max_{p\in\mathcal{S}_1}\min_{q\in\mathcal{S}_2} d_Z(p,q),\ \max_{p\in\mathcal{S}_2}\min_{q\in\mathcal{S}_1} d_Z(p,q)\}. \tag{2.9}$$

When the ambient space Z is not obvious from the context, we will note it in a subscript: $\mathcal{H}_Z(\mathcal{S}_1, \mathcal{S}_2)$. Intuitively, this is the maximum distance between a point in one set to the closest point in the other set when $\mathcal{S}_1$ and $\mathcal{S}_2$ are embedded in Z.

Gromov-Hausdorff Distance. The Hausdorff distance is limited, as we must choose our embedding space Z before computing the distance. The *Gromov-Hausdorff distance* d_{GH} finds the minimum Hausdorff distance over all possible embedding spaces:

$$d_{GH}(\mathcal{S}_1, \mathcal{S}_2) = \inf_{\substack{Z \\ f\colon \mathcal{S}_1 \to Z \\ g\colon \mathcal{S}_2 \to Z}} \mathcal{H}_Z(f(\mathcal{S}_1), g(\mathcal{S}_2)) \tag{2.10}$$

where f, g are isometric embeddings of $\mathcal{S}_1$ and $\mathcal{S}_2$, and Z is a metric space; see [17]. This allows the distance to be computed between metric spaces $\mathcal{S}_1$ and $\mathcal{S}_2$ up to embedding.

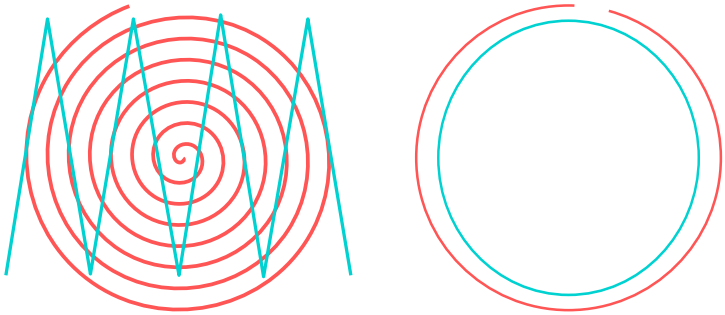
Fréchet Distance. Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be two homeomorphic spaces embedded in a metric space Z with parametrizations $\mathcal{S}_1(t)$ and $\mathcal{S}_2(s)$. The *Fréchet distance* between $\mathcal{S}_1$ and $\mathcal{S}_2$ is the minimum of the maximum distances between two corresponding points over all homeomorphisms:

$$\mathcal{F}(\mathcal{S}_1, \mathcal{S}_2) \quad = \quad \inf_{f:\mathcal{S}_1 \to \mathcal{S}_2} \max_{t \in \mathcal{S}_1} \|\mathcal{S}_1(t) - \mathcal{S}_2(f(t))\|. \tag{2.11}$$

This distance function was introduced in Maurice Fréchet's doctoral dissertation, wherein he defines *l'écart* (the gap) between elements of a set [45]. As a specific example, he considers l'écart between two rectifiable curves in Euclidean space. The relationship between Hausdorff and the Fréchet distances for planar curves is explored in [5]. We can use the Hausdorff distance as an upper bound for the Fréchet distance: for k-bounded planar polygons P_1 and P_2,

$$\mathcal{F}(P_1, P_2) \leq (k+1)\mathcal{H}(P_1, P_2). \tag{2.12}$$

Furthermore, the Hausdorff and Fréchet distances coincide for two convex planar curves. The extension of (2.12) to k-bounded curves in $\mathbb{R}^n$ is still an open question, as are the necessary and sufficient conditions for the Hausdorff and Fréchet distances to coincide. We do know, however, that (2.12) does not hold for all planar curves,



(a) Large Fréchet Distance (b) Non-homeomorphic

FIGURE 2.3: Both pairs of curves have a small Hausdorff distance. The figure on the left has a large Fréchet distance, and the Fréchet distance is not defined between the pair of curves on the right.

as is demonstrated in Figure 2.3a, which shows two curves with a small Hausdorff distance but a large Fréchet distance. This indicates that, in some cases, Fréchet distance is a more appropriate way to measure the similarity between curves. On the other hand, the requirement that f be a homeomorphism between $\mathcal{S}_1$ and $\mathcal{S}_2$ is quite limiting, as the Fréchet distance is not even defined between two spaces that are not homeomorphic, as in Figure 2.3b. Nonetheless, the Fréchet distance is a useful metric and will be explored further as it seems to be a more natural way to describe the distance between shapes, considering not only the position of the shapes within a metric space, but also the structure of the shapes.

2.2.2 Approximation of the Fréchet Distance

The Fréchet distance measures the distance between homeomorphic metric subspaces. In Corollary 4, we will see that there exists a homeomorphism that realizes an ϵ-approximation of the Fréchet distance. Knowing this, we will then turn to algorithms for computing the Fréchet distance and some of its variants.

Fréchet Distance for Closed Space Curves. The *Fréchet distance* between closed curves γ_1 and γ_2 is defined as the infimum over all homeomorphisms of the maximum distance between corresponding points. Recalling that a closed curve is a map

from the unit circle $\mathbb{S}^1$ to $\mathbb{R}^n$, we can take the infimum over homeomorphisms of $\mathbb{S}^1$ onto itself.

$$\mathcal{F}(\gamma_1, \gamma_2) = \inf_{\varphi:\mathbb{S}^1\to\mathbb{S}^1} \max_{t\in\mathbb{S}^1} \|\gamma_1(t) - \gamma_2(\varphi(t))\|. \tag{2.13}$$

The Fréchet distance has earned the name *dog-leash distance*. If we imagine a man walking on the path γ_1 and his dog walking along γ_2, then the Fréchet distance is the length of the shortest leash needed so that the man can traverse γ_1 and the dog traverse γ_2 continuously and without backtracking. If we allow the dog and/or the man to walk backward as well as forward, then the length of the shortest leash is called the *weak Fréchet distance*. We note here an important corollary to the definition of the Fréchet distance:

Corollary 4 (Near-Optimal Homeomorphism)**.** *For every $\epsilon > 0$, there exists a homeomorphism $\varphi_\epsilon \colon \mathbb{S}^1 \to \mathbb{S}^1$ such that $\max_{t\in\mathbb{S}^1} \|\gamma_1(t) - \gamma_2(\varphi_\epsilon(t))\| \leq \mathcal{F}(\gamma_1, \gamma_2) + \epsilon$.*

Any homeomorphism that satisfies the inequality of the above corollary is said to *realize* an ϵ-approximation of the Fréchet distance. If such a homeomorphism does not exist, then let γ_1 and γ_2 be two curves for which this corollary does not hold. Then, for all homeomorphisms φ, there exists a $t \in \mathbb{S}^1$ such that $\|\gamma_1(t) - \gamma_2(\varphi(t))\| > \mathcal{F}(\gamma_1, \gamma_2)$, which cannot happen by the definition of the Fréchet Distance. To make this corollary concrete, consider two closed polygons P_1 and P_2 in $\mathbb{R}^n$, then:

Lemma 5 (Realizing the Fréchet Distance)**.** *There exists a piecewise linear homeomorphism between P_1 and P_2 such that the Fréchet distance is realized at a vertex of P_1 or of P_2.*

Proof. Let $f : P_1 \to P_2$ be a homeomorphism. Suppose $v_0, v_1, \ldots, v_{m-1}$ is a sequence of points in P_1 that includes all vertices of P_1 and all preimages of vertices in P_2. Using the v_i and the $f(v_i)$, we construct a piecewise linear homeomorphism $g : P_1 \to P_2$

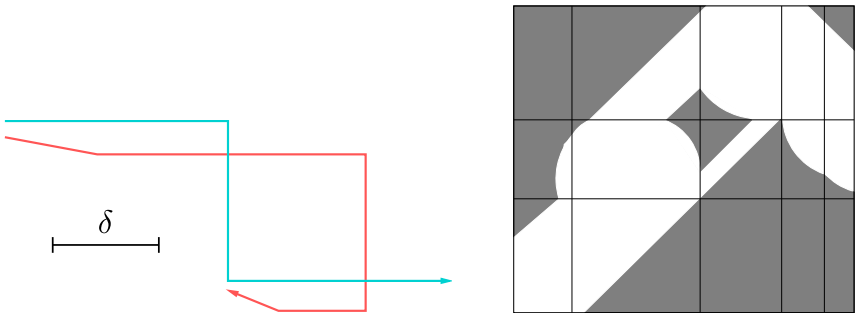
by defining $g(v_i) = f(v_i)$ for all vertices v_i and linearly interpolating in between the vertices. We claim that this implies

$$\max_{s \in P_1} \|s - g(s)\| \quad \leq \quad \max_{s \in P_1} \|s - f(s)\|.$$

Indeed, the maximum distance on the left occurs for s equal to one of the v_i. To see this, let $s = (1-t)v_i + tv_{i+1}$ for some $t \in [0,1]$, and note that

$$\begin{aligned} \|s - g(s)\| &= \|(1-t)v_i + tv_{i+1} - (1-t)g(v_i) - tg(v_{i+1})\| \\ &= \|(1-t)(v_i - g(v_i)) + t(v_{i+1} - g(v_{i+1}))\| \\ &\leq \max\{\|v_i - g(v_i)\|, \|v_{i+1} - g(v_{i+1})\|\}. \end{aligned}$$

This implies that the Fréchet distance between two polygonal chains is determined by a piecewise linear homeomorphism given by its values at the vertices of P_1 and at the preimages of the vertices of P_2. □



(a) Two Polygonal Curves. (b) Free Space Diagram

FIGURE 2.4: We draw the free space diagram for two open polygons. In this diagram, the curves are parametrized by arc length. The horizontal and vertical lines correspond to the vertices of the polygons. The regions that the lines create are called the *grid cells* of the free space diagram.

Exact Computation for Polygonal Chains. The fundamental tool used when computing–and, as we will see next, approximating–the Fréchet distance is the *free space diagram* [4]. Given a parametrization of two curves $\gamma_1, \gamma_2 \colon I \to \mathbb{R}^n$ and a choice of $\delta > 0$, we create a black and white image with domain $D = I \times I$. This is the free

space diagram, as illustrated in Figure 2.4[1], in which a point $(x, y) \in D$ is white if $d(\gamma_1(x), \gamma_2(y)) \leq \delta$ and is black otherwise. All white points are *feasible* or *free* in the domain; that is, the point (x, y) is white iff $\gamma_1(x)$ and $\gamma_2(y)$ have distance less than δ, indicating that they could be paired by a homeomorphism that realizes the Fréchet distance. The set of feasible points is called the *free space*, denoted $F_\delta(\gamma_1, \gamma_2)$:

$$F_\delta(\gamma_1, \gamma_2) \quad = \quad \left\{(s, t) \in [0, 1]^2 \mid d(\gamma_1(s), \gamma_2(t)) \leq \delta\right\}. \tag{2.14}$$

On the other hand, the points that are marked black represent *infeasible* pairings of points, since the Euclidean distance between the points is greater than δ.

The Fréchet distance is less than or equal to δ if there exists a bi-monotone path within the white (feasible) region connecting $(0, 0)$ and $(1, 1)$. Thus, in Figure 2.4, the Fréchet Distance is greater than the δ used to compute the free space diagram. If the curves are open polygons with p and q points respectively, then there are a finite number of possible values for the Fréchet Distance. Thus, we can construct an algorithm to compute the exact Fréchet Distance between polygonal chains. Alt and Godau [4] present an $O(pq \log(pq))$ time algorithm for polygonal chains ($O(pq \log^2(pq))$ in the case of closed polygons) using parametric search, improving upon a previous $O((pq^2 + p^2q) \log(pq))$ result that used binary search [47]. The Alt and Godau algorithm was successfully implemented by van Oostrum and Veltkamp for the Shape Matching Environment (SHAME) library [81], but it is a cumbersome library as it requires a complicated and delicate data structure. Rote extended the result of [4] to work for piecewise smooth curves [68]; however, this too is difficult to implement.

In general, computing the Fréchet distance between surfaces is NP-Hard [48], but it is semi-computable, i.e., there exists "a non-halting Turing machine which produces a monotone decreasing sequence of rationals converging to the result" [3].

[1] This image has been put into the public domain by Günter Rote. http://en.wikipedia.org/wiki/File:Free-space-diagram.png

For this reason, we limit our study to curves and we turn to approximating the Fréchet distance. There are two alternatives to computing the Fréchet distance using parametric search: to use a randomized algorithm or to approximate the distance. Cook and Wenk [25] explore a randomized algorithm for the Fréchet distance between two bounded planar curves. We look into the possibility of approximating the Fréchet distance.

Discrete Fréchet Distance for Polygonal Chains. Above, we compute the exact Fréchet distance between two polygonal chains P and Q. The natural question that arises is: What happens if we only consider correspondences between the vertex sets Vert(P) and Vert(Q)? Eiter and Mannila [37] answer this question by defining the *discrete Fréchet distance* or the *coupling distance* as follows:

$$\widetilde{\mathcal{F}}(P,Q) \;=\; \inf_{C\in\mathcal{O}(P,Q)} \max_{(i,j)\in C} \|P(i)-Q(j)\|,$$

where $\mathcal{O}(P,Q)$ is the set of ordered correspondences between Vert(P) and Vert(Q). A *correspondence* is a subset C of Vert(P) $\times$ Vert(Q) such that for each $p \in$ Vert(P) and each $q \in$ Vert(Q), there exists $p' \in P$ and q' in Q such that (p,q') and (p',q) are in C. We say that the correspondence is *ordered* if for all $(p_1,q_1),(p_2,q_2) \in C$, the inequality $p_1 < p_2$ implies $q_1 \leq q_2$ and symmetrically $q_1 < q_2$ implies $p_1 \leq p_2$. The

Algorithm 3 Computing the Discrete Fréchet Distance for Polygonal Chains

Input: Two polygonal chains P and Q.
Output: The discrete Fréchet distance between P and Q.

```
1 p ← |Vert(P)| − 1
2 q ← |Vert(Q)| − 1
3 for j = 0 to p do
4    for j = 0 to q do
5       c(i, j) = max{min{ c(i − 1, j − 1), c(i − 1, j), c(i, j − 1)}, ‖P(i/p) − Q(j/q)‖ }
6    end for
7 end for
8 return c(p, q)
```

discrete Fréchet distance can be computed using dynamic programming in $O(pq)$ time

and space. A straightforward implementation of the idea is given in Algorithm 3. This algorithm is both asymptotically faster than and easier to implement than the algorithms mentioned above. Agarwal et.al. [1] present a sub-quadratic algorithm for computing the discrete Fréchet distance by subtly exploiting the geometry of the problem.

The discrete Fréchet distance is a powerful tool, as it can be used to approximate the Fréchet distance between curves, as we will see next. First, we observe that

$$\mathcal{F}(P,Q) \;\leq\; \widetilde{\mathcal{F}}(P,Q) \tag{2.15}$$

since a correspondence between the vertices induces a piecewise linear homeomorphism between the polygonal curves. For polygonal chains, Eiter and Mannila prove the following inequality:

$$\widetilde{\mathcal{F}}(P,Q) \;\leq\; \mathcal{F}(P,Q) + \max_{e\in \mathrm{Edges}(P)\cup\mathrm{Edges}(Q)} \ell(e), \tag{2.16}$$

where $\ell(e)$ is the length of the edge e. This inequality justifies using $\widetilde{\mathcal{F}}(P,Q)$ to approximate $\mathcal{F}(P,Q)$, as $\widetilde{\mathcal{F}}(P,Q)$ can be arbitrarily close to $\mathcal{F}(P,Q)$ by splitting longer edges into shorter ones. We strengthen (2.16) in the following lemma:

Lemma 6 (Discrete Fréchet Distance for Curves)**.** *Let $\gamma_1, \gamma_2 \colon I \to \mathbb{R}^n$ be rectifiable curves, and let P and Q be inscribed in γ_1 and γ_2 respectively, parametrized so that the vertices occur at the same time in the inscribed polygon as in the original curve. Then,*

$$\left|\widetilde{\mathcal{F}}(P,Q) - \mathcal{F}(\gamma_1,\gamma_2)\right| \;\leq\; \mathcal{F}(\gamma_1,P) + \mathcal{F}(\gamma_2,Q) + \max_{e\in Edges(P)} \ell(e) + \max_{e\in Edges(Q)} \ell(e).$$

Proof. Let $\alpha \colon [0,1] \to [0,1]$ be a homeomorphism. Let parameter values $0 = u_0 < u_1 < \ldots < u_k = 1$ define P and parameter values $0 = v_0 < v_1 < \ldots < v_\ell = 1$ define

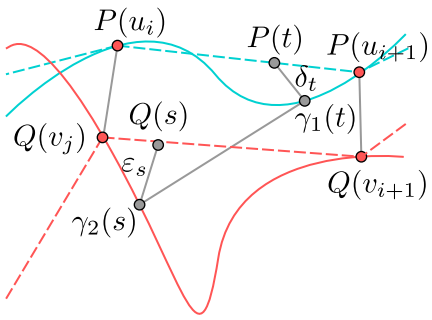


FIGURE 2.5: The inequality of Lemma 6 can be seen by bounding the difference between the lengths of vector $\gamma_1(t) - \gamma_2(s)$ and vector $P(u_i) - Q(v_j)$ over all α.

Q. We define $V(P)$ and $V(Q)$ as the vertex sets of P and Q:

$$\begin{aligned} V(P) &= \{\gamma_1(u_i)\}_{0 \le i \le k}, \\ V(Q) &= \{\gamma_2(v_i)\}_{0 \le i \le \ell}. \end{aligned}$$

Without loss of generality, we assume P and Q are parametrized so that $\gamma_1(u_i) = P(u_i)$ and $\gamma_2(v_j) = Q(v_j)$. We define

$$\begin{aligned} \delta_t &:= \|P(t) - \gamma_1(t)\|, \\ \varepsilon_s &:= \|Q(s) - \gamma_2(s)\|. \end{aligned}$$

Since γ_1 and γ_2 are rectifiable, $\delta = \max_t \delta_t$ and $\varepsilon = \max_s \varepsilon_s$ are finite.

Let $t \in [0, 1]$, then there exists an i such that $t \in [u_i, u_{i+1}]$. Similarly, there exists a j such that $\alpha(t) \in [v_j, v_{j+1}]$. We look at bounding the difference between the lengths of vector $\gamma_1(t) - \gamma_2(\alpha(t))$ and vector $\gamma_1(u_i) - \gamma_2(v_j)$; see Figure 2.5. Letting $s = \alpha(t)$, the difference between the lengths of the vector $\gamma_1(t) - \gamma_2(s)$ and vector $\gamma_1(u_i) - \gamma_2(v_j)$ is at most the sum of the differences between their endpoints,

$$\delta_t + \varepsilon_s + \|P(t) - P(u_i)\| + \|Q(s) - Q(v_j)\|.$$

Taking the maximum over t, we let r denote the sum of the longest edge in P and the longest edge in Q and we obtain the following inequality for a given homeomor-

phism α:

$$\left|\max_t \|\gamma_1(t) - \gamma_2(s)\| - \max_t \|\gamma_1(u_i) - \gamma_2(v_j)\|\right| \;\leq\; \delta + \varepsilon + r.$$

Since α was an arbitrary homeomorphism and the Fréchet distance is invariant with respect to reparametrization, we can now conclude:

$$\left|\widetilde{\mathcal{F}}(P,Q) - \mathcal{F}(\gamma_1,\gamma_2)\right| \;\leq\; \mathcal{F}(\gamma_1,P) + \mathcal{F}(\gamma_2,Q) + r,$$

as desired. ◻

Although the above lemma has been proven for curves, the result also holds for loops. To understand this claim, let $\gamma_1, \gamma_2 \colon \mathbb{S}^1 \to \mathbb{R}^n$ with P and Q inscribed in γ_1 and γ_2. Let C be an ordered correspondence between $V(P)$ and $V(Q)$ that realizes $\widetilde{\mathcal{F}}(P,Q)$. Then, choose $(p,q) \in C$ to be the starting points of the open polygonal chains. The proof above can then be followed exactly.

Approximation. Recall from Section 2.1.1 that a curve is c-packed if the total length of the curve contained inside any ball is bounded by c times the radius of the ball. In [29], Driemel, Har-Peled and Wenk present an algorithm to approximate the Fréchet distance that runs in near linear time for c-packed curves. We give a brief overview of this algorithm.

Let P_1 and P_2 be two c-packed polygons. The first step of the algorithm is to find a μ-simplifications of the polygons; that is, to create a polygon $\text{simp}_\mu(P_i)$ from a subset of the vertices of P_i such that the Fréchet distance between P_i and $\text{simp}_\mu(P_i)$ is at most μ. The basic idea of their algorithm is to first search over a (finite) set of approximate Fréchet distances in order to define an interval $[\alpha, \beta]$ that contains the exact Fréchet distance between the simplified polygons. Then, perform a binary search on the interval until a tolerable approximation of the actual Fréchet distance is found.

Given $\delta > 0$, the *reachable free space* $\mathcal{R}_\delta(P_1, P_2)$ is the set of all points in the free space diagram $F_\delta(P_1, P_2)$ that are reachable from the origin by a bi-monotone path. The Fréchet distance is at most δ if the upper left corner $(1, 1) \in \mathcal{R}_\delta(P_1, P_2)$. The *complexity* of the reachable free space $\mathcal{N}_\delta(P_1, P_2)$ is the maximum number of grid cells with non-empty intersection with $\mathcal{R}_\delta(P_1, P_2)$ over all μ-simplifications of P_1 and P_2. The algorithm described in [29] is $O(\mathcal{N}_\delta(P_1, P_2) \log n)$. Since $\mathcal{N}_\delta(P_1, P_2)$ is linear for c-packed curves, the approximation algorithm for c-packed curves is $O(n \log n)$.

Application. The Fréchet distance can be used to measure the distance between the motion of jointed objects. As an object with n joints moves in time, we keep track of the angle at each joint. Then, we represent the motion of the object as a curve in $\mathbb{R}^n$:

$$\gamma(t) = (\alpha_1(t), \alpha_2(t), \ldots, \alpha_n(t)).$$

Each function $\alpha_i(t)$ for $1 \leq i \leq n$ represents the continuous motion of the i^{th} joint by recording the angle. This function is the *joint-angle trajectory*, and has been explored in [77] to classify the gaits of people walking. Given two such curves, γ_1 and γ_2, we can use the Fréchet distance between them as a way to measure similarity (or difference) between the two motions. The Fréchet distance can measure the distance up to speed invariance. The challenge with this approach is that the start time of the motion must be well chosen. In addition, care must be taken when interpreting the angles, since an angle of 2π is equal to an angle of zero.

2.2.3 Metrics on Persistence Diagrams

As above, let $\mathcal{S}_1$ and $\mathcal{S}_2$ be two topological spaces. Suppose further that there are functions $f\colon \mathcal{S}_1 \to \mathbb{R}$ and $g\colon \mathcal{S}_2 \to \mathbb{R}$. Now, we will use the persistence diagrams to measure the distance between $\mathcal{S}_1$ and $\mathcal{S}_2$. We say that these metrics are *indirect* since they use only the information in the persistence diagrams in order to compute a distance between the original spaces.

We first describe the general matching problem, where we match the elements of a set $A \subseteq \mathbb{R}^2$ with elements of a second set $B \subseteq \mathbb{R}^2$. A *matching* between A and B is a set of edges (a, b) with a in A and b in B such that no vertex is incident to two edges. A matching is *maximal* if the addition of any edge would result in a graph that is no longer a matching. A matching is *perfect* if every vertex is incident on exactly one edge. In other words, it is a matching where there does not exist an unmatched vertex.

The goal is to obtain a perfect matching between $\mathrm{Dgm}_p(f)$ and $\mathrm{Dgm}_p(g)$ that minimizes the cost associated with the matching. To resolve the issue where the number of off-diagonal points in both diagrams is not equal or the diagrams are dissimilar, we allow an off-diagonal point to be matched to a point on the line $y = x$. Hence, we let A be the set of off-diagonal points in $\mathrm{Dgm}_p(f)$ along with the orthogonal projection of the off-diagonal points in $\mathrm{Dgm}_p(g)$ onto the line $y = x$. Symmetrically, we let B be the set of off-diagonal points in $\mathrm{Dgm}_p(g)$ along with the orthogonal projection of the off-diagonal points in $\mathrm{Dgm}_p(f)$ onto the line $y = x$. The cost of a matching is a function of a cost of the edges in the matching. The cost $d^\infty(a, b)$ between a and b is the L_∞ distance:

$$d^\infty(a, b) = \max\{|a_x - b_x|, |a_y - b_y|\}. \tag{2.17}$$

As we will see next, the Wasserstein distance minimizes the sum of the pairwise costs; whereas, the Bottleneck distance minimizes the maximum pairwise cost.

Wasserstein Matching of Persistence Diagrams. We define the Wasserstein cost function as follows:

Definition 7 (Wasserstein Cost)**.** The *degree q Wasserstein cost* $C_q(P)$ of a perfect matching P is the q^{th} root of the sum of the q^{th} powers of the edge costs over all

edges in the matching:

$$C_q(P) = \left(\sum_{(a,b)\in P} d^{\infty}(a,b)^q \right)^{1/q}.$$

The *Wasserstein distance* $W_q(A,B)$ is the minimum degree q Wasserstein cost over all perfect matchings:

$$W_q(A,B) \;=\; \min_P C_q(P). \tag{2.18}$$

The *Wasserstein matching* is the matching that attains the Wasserstein distance.

The Hungarian method computes the Wasserstein matching in $O(n^4)$ computational complexity [56], where $|A| = |B| = n$. In [79], Vaidya maintains weighted Voronoi diagrams for an $O(n^{2.5} \log n)$ computation of this matching. Further improvements were made by Agarwal, Efrat, and Sharir [2]. They utilize a data structure that improves the running time to $O(n^{2+\epsilon})$.

Bottleneck Matching of Persistence Diagrams. Analogous to finding the Wasserstein matching, the bottleneck matching minimizes the bottleneck cost function, which is defined as follows:

Definition 8 (Bottleneck Cost)**.** The *bottleneck cost* of a perfect matching P is the maximum edge cost in the matching:

$$C_{\infty}(P) = \max_{(a,b)\in P} d^{\infty}(a,b).$$

We minimize this cost over all perfect matchings of A and B to obtain the Bottleneck Distance between the sets:

$$W_{\infty}(A,B) \;=\; \min_P C_{\infty}(P). \tag{2.19}$$

The matching that attains the bottleneck distance is called the *bottleneck matching*. Although for small values of q, the bottleneck and the Wasserstein matchings may differ, the Wasserstein cost approaches the Bottleneck cost as q approaches infinity.

The computational complexity of the bottleneck matching is slightly smaller than that of the Wasserstein matching. A maximal matching can be found in $O(n^{2.5})$ using the Hopcroft-Karp algorithm [51]. If we mimic the thresholding approach of the Hungarian method [56], then the bottleneck solution can be found in $O(n^{2.5} \log n)$. Since A and B are sets of points in the plane, we can improve the computational complexity of determining the matching under the bottleneck distance. Efrat, Itai, and Katz developed a geometric improvement to the Hopcroft-Karp algorithm with a running time of $O(n^{1.5} \log^2 n)$ [36].

Since the running time of computing the bottleneck distance between finite point sets is polynomial, approximating computationally intensive metrics by the Bottleneck Distance is often desired. For example, we can lower bound d_{GH} of Equation (2.10) by the bottleneck distance W_∞ of persistence diagrams corresponding to Rips filtrations.

Theorem 9 (Lower Bound for the Gromov-Hausdorff Distance [17])**.** *For all finite metric spaces $\mathcal{S}_1$ and $\mathcal{S}_2$ and for all $k \in \mathbb{N}$, the bottleneck distance between diagrams of Rips filtrations is upper bounded by the Gromov-Hausdorff Distance between the vertex sets:*

$$W_\infty \left(Dgm_k(\mathcal{R}(\mathcal{S}_1)), Dgm_k(\mathcal{R}(\mathcal{S}_2)) \right) \leq d_{GH}(\mathcal{S}_1, \mathcal{S}_2).$$

In words, the bottleneck distance is stable with respect to the Gromov-Hausdorff distance. Since computing the bottleneck distance is more feasible than computing the NP-Hard Gromov-Hausdorff distance, Chazal et al. successfully use the bottleneck distance as a shape signature in [17].

Editing Distances. Neither the Wasserstein nor the bottleneck distances between persistence diagrams measures the cost of transforming the function f into the function g. Both measure the difference between the two fixed functions. An *editing distance* sums the cost of edit operations necessary to transform one object into another [76]. In the original context, the objects are trees; however, for us, the objects are functions on simplicial complexes. In [27], the authors define a (stable) edit distance between Reeb graphs that closely relates to the results of [21]. The Fréchet distance $\mathcal{F}$ is also an edit distance, where $\mathcal{F}$ is the largest distance traveled when deforming one curve into another one. In [42], I defined the *vineyard distance*, which uses the vineyard arising from a homotopy between the functions f and g; see [24] for a definition of the vineyard. Although the vineyard distance and other edit distances have the potential to be insightful, we limit our discussion of distances between persistence diagrams to the Wasserstein and the bottleneck distances.

Stability Theorems. A distance metric (and the corresponding matching) is stable if a small change in the input sets A and B produces a small change in the measured distance between the sets. The property of stability is not easy to verify and sometimes not even true.

Let f, g be tame functions. The L_∞-distance $||f-g||_\infty$ is the maximum difference between the function values:

$$||f-g||_\infty = \sup_{x \in \mathcal{S}_1} |f(x) - g(x)|.$$

The L_∞-distance between f and g is an upper bound for the bottleneck distance between the corresponding persistence diagrams:

Theorem 10 (Bottleneck Stability Theorem [21])**.** *The bottleneck distance between the persistence diagrams of tame functions is bounded from above by the L_∞ distance*

between them

$$W_\infty(Dgm_p(f), Dgm_p(g)) \leq ||f - g||_\infty. \tag{2.20}$$

The proof of the Bottleneck Stability Theorem presented in [24] uses the following lemma:

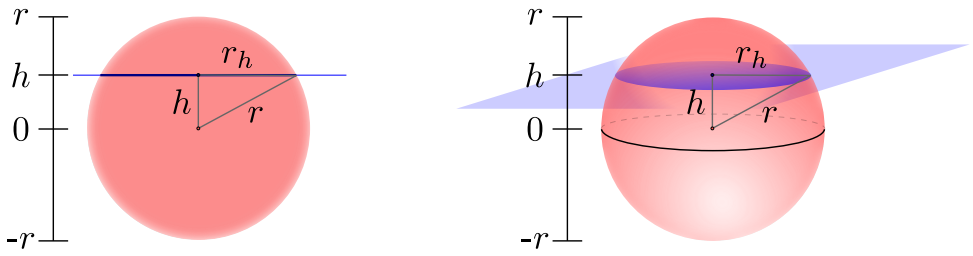
Lemma 11 (Stability of the Straight Line Homotopy)**.** *Given the straight line homotopy from f to g, there exists a perfect matching P of the persistence diagrams for f and g such that the bottleneck cost of P is upper bounded by the distance between f and g:*

$$\max_{(a,b)\in P} C(a,b) \leq ||f - g||_\infty.$$

Although (2.20) does not hold for the Wasserstein Distance, the Wasserstein Distance is stable for Lipschitz functions with bounded degree k total persistence [23], and potentially unstable if we relax either of these two conditions (Lipschitz or bounded total persistence).

2.3 Volumes of Spheres and Balls

Properties of the n-sphere and n-ball arise several times throughout this dissertation. In Chapter 3, we eliminate a factor of $\frac{\text{Vol}(\mathbb{S}^{n-1})}{\text{Vol}(\mathbb{S}^n)}$ from an inequality in order to create an upper bound for the difference of lengths of two curves that is independent of the dimension in which the curves are embedded. At the end of this subsection, we will show that this factor grows to infinity as n increases to infinity. Then, in the discussion of Gaussian mixtures, we see spheres as the level sets of isotropic Gaussian kernels. These spheres are important in our analysis, since for small enough radii r, the Gaussian kernel is concave over a ball of radius r centered at the Gaussian mean.



(a) Slab technique applied to a 2-Ball. (b) Slab technique applied to a 3-Ball.

FIGURE 2.6: The slab technique decomposes the n-ball into an infinite collection of $(n-1)$-balls by intersecting the n sphere with $(n-1)$-dimensional planes. On the left, the 2-ball is decomposed into 1-balls, or line segments. On the right, the 3-ball is decomposed into 2-balls, or disks. We compute the radius of the slab from the radius of the n-ball and the distance between the centers of the balls; see Lemma 12.

2.3.1 The Slab Technique

One method of computing the volume of a ball $B \subset \mathbb{R}^n$ is to decompose it into horizontal slabs by projecting B onto its n^{th} Euclidean coordinate; this projection is denoted by $\pi_n(B)$. If B has radius r, then the image $\pi_n(B)$ has length $2r$; see Figure 2.6. As the volume of a ball is translation invariant, we consider balls centered at the origin. We denote the ball of radius r centered at the origin in $\mathbb{R}^n$ by $\mathbb{B}_r^n$. For points on a ball centered at the origin, we call the n^{th} (last) Euclidean coordinate the *height* of the point.

For every $h \in (-r, r)$, the preimage of h under the projection map, denoted $\pi_n^{-1}(h)$, is exactly the $(n-1)$-ball defined by the intersection of $\mathbb{B}_r^{n-1}$ with the horizontal plane $\pi_n(x) = h$. As shown in Figure 2.6, we use the Pythagorean theorem to compute the radius of this ball:

Lemma 12 (Radius of a Slab)**.** *The radius of* $\mathbb{B}_r^n$ *at height* h *is* $r_h = \sqrt{r^2 - h^2}$.

Knowing this lemma, we then integrate over all $h \in (-r, r)$ to obtain a recursive formula for the volume of the n-ball.

Table 2.1: We record the volume of the n-ball and of the $(n-1)$-sphere for dimension one through five. Notice that the exponent of the radius increases linearly and $\mathrm{Vol}(\mathbb{S}_r^{n-1})$ is the derivative of $\mathrm{Vol}(\mathbb{B}_r^n)$.

	1	2	3	4	5
$\mathrm{Vol}(\mathbb{B}_r^n)$	$2r$	πr^2	$\frac{4}{3}\pi r^3$	$\frac{1}{2}\pi^2 r^4$	$\frac{8}{15}\pi^2 r^5$
$\mathrm{Vol}(\mathbb{S}_r^{n-1})$	2	$2\pi r$	$4\pi r^2$	$2\pi^2 r^3$	$\frac{8}{3}\pi^2 r^4$

Lemma 13 (Slab Technique for the n-Ball)**.**

$$\mathrm{Vol}(\mathbb{B}_r^n) = 2\int_0^r \mathrm{Vol}(\mathbb{B}_{r_h}^{n-1})\, dh.$$

In this lemma, we have used the symmetry of the n-ball in order to simplify the integral. In particular, we used the insight is that the $(n-1)$-ball at height $-h$ is isometric to the $(n-1)$-ball at height h.

In low-dimensions, the volume of $\mathbb{B}_r^n$ is well-known. For example, the ball $\mathbb{B}_r^1$ is the line segment extending from $-r$ to r. The one-dimensional volume of this line segment is its length, $2r$. In two-dimensions, the volume of a ball, i.e. a disk, is the area, πr^2. The volume of $\mathbb{B}_r^n$ for $n \leq 5$ is given in Table 2.1. In this table, we observe that the exponent of r is n, thus we hypothesize:

Lemma 14 (General Form of the Volume of a Ball)**.** *The volume of a ball can be written as*

$$\mathrm{Vol}(\mathbb{B}_r^n) = v_n r^n, \tag{2.21}$$

where $v_n \in \mathbb{R}$ depends only on n.

Proof. We prove this lemma by induction on n and, at the same time, determine a recursive formula for v_n. Recall from Table 2.1 that $\mathrm{Vol}(\mathbb{B}_r^1) = 2r$ and $\mathrm{Vol}(\mathbb{B}_r^2) = \pi r^2$ both satisfy (2.21). Now, assume that (2.21) holds for $n \leq k$.

Let $n = k+1$. Then, by Lemma 13 we have:

$$\mathrm{Vol}(\mathbb{B}_r^{k+1}) \;=\; 2\int_0^r \mathrm{Vol}(\mathbb{B}_{r_x}^k)\, dx. \tag{2.22}$$

Then, we convert to polar coordinates and use our inductive assumption to obtain:

$$\begin{aligned}\text{Vol}(\mathbb{B}_r^{k+1}) &= 2\int_0^r \int_0^{2\pi} \frac{1}{2}\text{Vol}(\mathbb{B}_{r_x}^{k-1}) x \, d\theta \, dx \\ &= \int_0^r \int_0^{2\pi} v_{k-1} r_x^{k-1} x \, d\theta \, dx.\end{aligned}$$

From Lemma 12, we know $r_x = \sqrt{r^2 - x^2}$. Thus, we can write:

$$\begin{aligned}\text{Vol}(\mathbb{B}_r^{k+1}) &= v_{k-1} \int_0^{2\pi} d\theta \int_0^r r_x^{k-1} x \, dx \\ &= \frac{2\pi v_{k-1}}{k+1} r^{k+1}.\end{aligned}$$

We conclude that $\text{Vol}(\mathbb{B}_r^{k+1}) = v_{k+1} r^{k+1}$, where v_{k+1} is a real number that depends only on k, namely $v_{k+1} = \frac{2\pi v_{k-1}}{k+1}$. ◻

2.3.2 Volume of the Unit Ball

If we let $r = 1$ in Lemma 13, then we find that $\text{Vol}(\mathbb{B}_1^n) = v_n$. Hence, if we know the volume of the unit n-ball, then we can compute the volume the ball of any given radius. Plugging 2.21 into 2.22 and converting to polar coordinates as shown in [74, p.135-6], we obtain the closed-form expression for v_n:

$$v_n = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2} + 1)} \tag{2.23}$$

where $\Gamma(\cdot)$ is Euler's Gamma Function, defined as follows:

Definition 15 (Euler Gamma Function)**.** For $z \in \mathbb{C}$ with $\text{Re}(z) > 0$, the Gamma Function is defined by the second Eulerian integral:

$$\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} \, dt$$

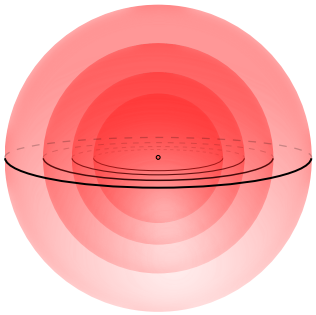


FIGURE 2.7: The shelling technique decomposes the n-ball into an infinite collection of $(n-1)$-spheres. For each $x \in (0, r)$, there is a distinct sphere of radius x contained in the ball of radius r.

A more detailed coverage of this function can be found in [55, §21.4], but we note two special cases:

$$\Gamma(z) = \begin{cases} n! & \text{for } z = n+1 \\ \frac{\sqrt{\pi}}{2^n}(2n-1)!! & \text{for } z = n+\frac{1}{2}. \end{cases}$$

In the second case, the double factorial $(2n-1)!!$, is the product of every other positive integer less than or equal to $2n-1$ with the same parity as $2n-1$. Plugging the above formula into (2.23), allows us to compute the volume of the unit ball in $\mathbb{R}^n$.

2.3.3 Volume of the Sphere

To compute the volume of a Sphere, we can take two approaches:

1. *Slab Approach.* We can use the same technique as presented in the previous subsection to decompose the sphere into lower dimensional spheres by looking at the intersection of a sphere with a horizontal hyperplane.

2. *Shell Approach.* We decompose the n-ball into a 1-parameter family of $(n-1)$-dimensional concentric spheres, as shown in Figure 2.7.

We use the Shell approach. The n-ball is in fact an infinite collection of concentric $(n-1)$-spheres; see Figure 2.7. Thus, we are able to write the volume of the n-ball

as an integral over the volume of $(n-1)$-dimensional spheres:

$$\mathrm{Vol}(\mathbb{B}_r^n) = \int_0^r \mathrm{Vol}(\mathbb{S}_x^{n-1})\, dx. \tag{2.24}$$

From Lemma 14, we know that the volume of the n-ball is $r^n v_n$, hence we have the following by taking the derivative of both sides of (2.24):

$$\mathrm{Vol}(\mathbb{S}_r^{n-1}) = n v_n r^{n-1}.$$

From this equation, we plug in the formula for v_n from (2.23) to obtain the volume of the unit sphere for even and odd dimensions:

$$\mathrm{Vol}(\mathbb{S}^n) = \begin{cases} \frac{(2\pi)^{n/2}}{(n-2)!!}, & n \text{ even} \\ \frac{2(2\pi)^{(n-1)/2}}{(n-2)!!}, & n \text{ odd.} \end{cases} \tag{2.25}$$

This formula immediately leads to the following corollary.

Corollary 16. *There is a recursive relationship between volumes of spheres:*

$$\mathrm{Vol}(\mathbb{S}^n) = \frac{2\pi}{n-2}\mathrm{Vol}(\mathbb{S}^{n-2}).$$

2.3.4 Ratio of Volumes of Spheres

We consider the ratio of volumes of co-dimension one spheres, which is given by

$$r(n) \quad = \quad \frac{\mathrm{Vol}(\mathbb{S}^{n-1})}{\mathrm{Vol}(\mathbb{S}^n)} \quad = \quad c_n \cdot \frac{(n-1)!!}{(n-2)!!}. \tag{2.26}$$

The constant c_n depends only on the parity of n: $c_n = 1/2$ if n is even and $c_n = 1/\pi$ if n is odd. We conclude this section with the following theorem, which to our knowledge has not been published elsewhere:

Theorem 17. *The ratio of the volume of spheres $r(n)$ grows like $\sqrt{n}$.*

Proof. This lemma is simple to prove given (2.25). Considering the product $r(n) \cdot r(n+1) = n/2\pi$, we note that $r(n)$ grows like $\sqrt{n}$ as claimed. ◻

3

Lengths and Distances

We are interested in the Fréchet distance $\mathcal{F}(\gamma_1, \gamma_2)$ between two closed curves. The main result of this chapter is a theorem that bounds the difference in length of two curves in terms of their total curvatures and the Fréchet distance. This result was originally published in [43]. The bound is independent of the dimension of the ambient Euclidean space, it improves upon a bound by Cohen-Steiner and Edelsbrunner [20] in dimensions four and higher, and it generalizes a result by Fáry and Chakerian [40, 16].

3.1 History

In 1950, Fáry [40] showed that given a closed curve γ contained in a disk of radius r in the plane, the length of the curve is bounded from above by the total curvature of the curve:

$$\ell(\gamma) \quad \leq \quad r\kappa(\gamma). \tag{3.1}$$

Fáry also found a similar inequality for a closed curve γ contained in a ball of radius r in 3-dimensional Euclidean space:

$$\ell(\gamma) \quad \leq \quad \frac{4}{\pi} r\kappa(\gamma). \tag{3.2}$$

Of particular interest, we see that for a planar curve γ contained in the unit disk, $\ell(\gamma) \leq \kappa(\gamma)$. While equality can be achieved in (3.1) when γ is the unit circle, the bound in (3.2) is not tight. In fact, ten years later, Chakerian [16] proved that (3.1) also holds for curves in $\mathbb{R}^n$:

Theorem 18 (Fáry-Chakerian)**.** *If γ is a closed curve contained in a ball of radius r in $\mathbb{R}^n$, then $\ell(\gamma) \leq r\kappa(\gamma)$.*

Now that we have a tight bound for the length of a single curve, we explore what is known for two curves. Cohen-Steiner and Edelsbrunner [20] provide an upper bound for the difference:

Theorem 19 (Original Length Bound [20])**.** *If γ_1 and γ_2 are two rectifiable closed curves in $\mathbb{R}^n$, then $|\ell_1 - \ell_2| \leq \frac{2\text{Vol}(\mathbb{S}^{n-1})}{\text{Vol}(\mathbb{S}^n)}(\kappa_1 + \kappa_2 - 2\pi)\mathcal{F}(\gamma_1, \gamma_2)$.*

In this theorem, ℓ_j and κ_j are the length and the total curvature of γ_j and $\mathcal{F}(\gamma_1, \gamma_2)$ is the Fréchet distance, as defined in (2.13). The ratio of sphere volumes $\text{Vol}(\mathbb{S}^{n-1})/\text{Vol}(\mathbb{S}^n)$ is an increasing function that reaches $2/\pi$ at $n = 3$ and diverges as n goes to infinity. More precisely, the ratio grows like $\sqrt{n}$ as shown in Theorem 17. Thus, an upper bound that avoids this diverging factor is an improvement on the Length Bound. Such a bound is the main result of this chapter:

Theorem 20 (Improved Length Bound)**.** *If γ_1 and γ_2 are rectifiable closed curves in $\mathbb{R}^n$, then $|\ell_1 - \ell_2| \leq \frac{4}{\pi}(\kappa_1 + \kappa_2)\mathcal{F}(\gamma_1, \gamma_2)$.*

The remainder of this chapter presents the proofs of these two theorems.

3.2 Curves and Inscribed Polygons

The results in this section address questions about approximating a smooth curve by a polygon. We prove that length, total curvature, and Fréchet distance are the same in the limit as a polygon approaches the curve. We also provide an integral geometric interpretation of both the length and the total curvature. We begin this section with a proof of the 3D Caliper Lemma.

3.2.1 Angles and Distances

Let $\triangle abc$ and $\triangle ABC$ be triangles with two congruent edges: $\|a-c\| = \|A-C\|$ and $\|b-c\| = \|B-C\|$. Proposition 24 of Euclid's Elements says that $\|a-b\| \leq \|A-B\|$ implies that the angle at c is less than or equal to the angle at C, where all angles are measured between 0 and π [38, p. 296]. This is also known as the Caliper Lemma [73]; see Figure 3.1 for an illustration of this lemma. The distance between a and b is maximized when the angle at c is straight. We generalize this insight to higher dimensions. The three-dimensional case is important so we discuss it first. Consider two tetahedra hinged on a face: $axyz$ and $xyzb$; see Figure 3.5 with

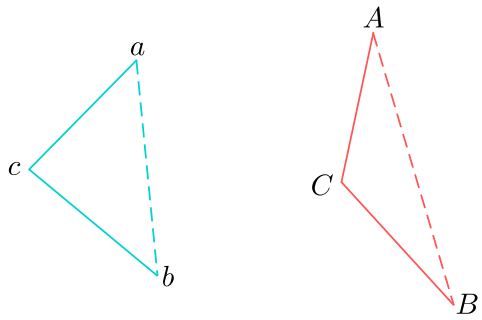


FIGURE 3.1: The Caliper Lemma states that $\|a-b\| \leq \|A-B\|$ implies that the angle at c is less than or equal to the angle at C.

$a = p_{k-1}$, $b = p_{k+1}$, $x = p_k$, $y = q_k$, $z = q_{k-1}$.

Lemma 21 (3D Caliper)**.** *The distance between a and b is maximized when $axyz$ and $xyzb$ lie in a common three-dimensional hyperplane, and a and b are on opposite*

sides of the plane containing $\triangle xyz$.

Proof. We begin by isometrically embedding $axyzb$ in $\mathbb{R}^4$. Let the orbit of b be the set of points b' such that $xyzb'$ is isometric to $xyzb$. The set of points with constant distance to three points in $\mathbb{R}^4$ is a circle, and the two-plane spanned by those three points is the orthogonal axis of rotation. Hence, the orbit of b is a circle orthogonal to the three-plane H defined by $axyz$. There are two points where this circle intersects H. Let b_0 be the point that causes the two tetrahedra $axyz$ and $xyzb$ to overlap, and b_1 the reflection of b_0 across the two-plane xyz. Consider the line segment that connects a and b_1. Let c be the intersection of the line and the plane xyz. By construction, the distance cb is constant throughout the orbit of b. Since $b = b_1$ corresponds to the angle at c being straight, we know by the Caliper Lemma that $b = b_1$ maximizes the distance between a and b. ◻

This proof directly extends to higher dimensions. We state the generalized from of Lemma 21:

Lemma 22 (n-Dimensional Caliper)**.** *Let* Δ *be an* n*-simplex in* $\mathbb{R}^{n+1}$*. Let* $a, b \in \mathbb{R}^{n+1}$*. If we are free to move* a *and* b *around while maintaining the distance between* a *and any vertex of* Δ *and between* b *and any vertex of* Δ*, then* $\|a - b\|$ *is maximized when* a, b, Δ *lie in a common* n*-plane such that* a *and* b *are on opposite sides of the* $(n-1)$*-plane defined by* Δ*.*

3.2.2 Length in the Limit

In general, we think of a closed curve as the image of the circle, $\mathbb{S}^1$, but we use the unit interval whenever it is more convenient to parametrize with $I = [0, 1]$. Given a closed curve $\gamma : I \to \mathbb{R}^n$ and a choice of parameter values $0 = t_0 < t_1 < t_2 < \cdots < t_m = 1$, we let P be the *inscribed polygon* consisting of the vertices $v_i = \gamma(t_i)$ and the edges $v_i v_{i+1}$, where we take indices modulo m. Figure 3.2 shows an example of

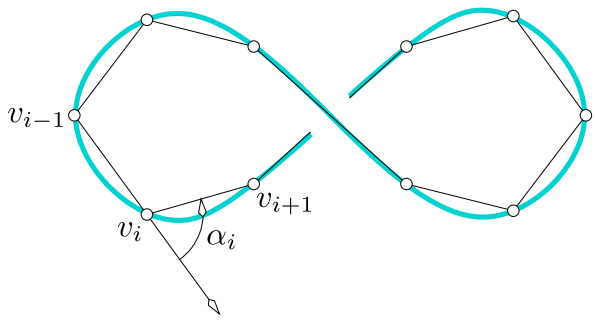


FIGURE 3.2: A closed curve in space, an inscribed polygon, and the turning angle at the vertex v_i.

an inscribed polygon.

We write $P < \gamma$ to denote that P is inscribed in γ. If γ is smooth, then we can get good polygonal approximations by choosing the values without leaving large gaps. To make this precise, we define the *mesh* of P as the maximum distance between two consecutive values in the parametrized interval: $\text{Mesh}(P) = \max_{0 \le i < m}\{t_{i+1} - t_i\}$. Using the trapezoidal rule for approximating the integral of (2.3), we arrive at the following result:

Lemma 23 (Length in the Limit)**.** *If P^k is a sequence of polygons inscribed in a rectifiable closed curve γ such that* $\text{Mesh}(P^k)$ *goes to zero as k goes to infinity, then* $\ell(\gamma) = \lim_{k\to\infty} \ell(P^k)$.

3.2.3 Total Curvature

As shown in Figure 3.2, the *turning angle* α_i at the vertex v_i is the complementary angle to the angle between vectors $v_{i-1} - v_i$ and $v_{i+1} - v_i$. Note that $0 \le \alpha_i \le \pi$. Milnor [62] defines the *total curvature* of a closed polygon P in $\mathbb{R}^n$ as the sum of the turning angles: $\kappa(P) = \sum_{i=0}^{m-1} \alpha_i$. The total curvature of a closed curve γ in $\mathbb{R}^n$ is the supremum over all inscribed polygons:

$$\kappa(\gamma) \quad = \quad \sup_{P<\gamma} \kappa(P). \tag{3.3}$$

Another notion of total curvature was studied by Fenchel [44]. Letting $\kappa(s)$ at a point s on a smooth curve be one over the radius of the locally best approximating circle, he gets the *total curvature* by integration:

$$\kappa(\gamma) \quad = \quad \int_{s\in\gamma} \kappa(s)\,\mathrm{d}s. \tag{3.4}$$

Milnor proved that these two definitions of total curvature are equivalent, and using his definition, we find the total curvature of a smooth curve as the limit of the total curvature of inscribed polygons [39]:

Lemma 24 (Total Curvature in the Limit)**.** *If P^k is a sequence of polygons inscribed in a smooth closed curve γ such that* $\mathrm{Mesh}(P^k)$ *goes to zero as k goes to infinity, then $\kappa(\gamma) = \lim_{k\to\infty} \kappa(P^k)$.*

3.2.4 Fréchet Distance in the Limit.

Recall the definition of the Fréchet distance between two curves in $\mathbb{R}^n$ from (2.13). In this subsection, we show that the Fréchet distance between inscribed polygons can be arbitrarily close to the Fréchet distance between the corresponding rectifiable curves. Given two curves γ_1 and γ_2 in $\mathbb{R}^n$ with $P_1 < \gamma_1$ and $P_2 < \gamma_2$, the corresponding values Fréchet distances $\mathcal{F}(P_j^k, \gamma_j)$ also goes to zero. Furthermore, since the Fréchet distance is a metric, we have:

$$\left|\mathcal{F}(P_1^k, P_2^k) - \mathcal{F}(\gamma_1, \gamma_2)\right| \quad \leq \quad \mathcal{F}(\gamma_1, P^k) + \mathcal{F}(\gamma_2, P_2^k).$$

Letting k increase to infinity, we the Fréchet distance between the polygons P_1 and P_2 becomes arbitrarily close to the Fréchet distance between the curves γ_1 and γ_2. We formalize this in the following lemma:

Lemma 25 (Fréchet Distance in the Limit)**.** *If P_1^k and P_2^k are sequences of polygons inscribed in the smooth closed curves γ_1 and γ_2 in $\mathbb{R}^n$ such that* $\mathrm{Mesh}(P_1^k)$ *and* $\mathrm{Mesh}(P_2^k)$ *both go to zero, then $\mathcal{F}(\gamma_1, \gamma_2) = \lim_{k\to\infty} \mathcal{F}(P_1^k, P_2^k)$.*

3.2.5 Integral Geometry Interpretations

Let γ be a curve embedded in $\mathbb{R}^n$ and let $u \in \mathbb{S}^n$. Then, there is a family of $(n-1)$-planes H_z in $\mathbb{R}^n$ with normal vector u, parametrized by the distance z from the origin. For any $x \in \mathbb{R}^n$, the inner product $x \cdot u$ is the height of x in the direction u, hence $x \in H_{x \cdot u}$. We will abuse notation and use u for both the vector and the inner product. Then, $u \circ \gamma$ is a map from $[0,1]$ to $\mathbb{R}$ and is called the *height function*. The preimage $(u \circ \gamma)^{-1}$ is generically a set of points. We call $(u \circ \gamma)^{-1}(z)$ a *level-set* of the height function. Let $\sharp(\mathcal{S})$ count the points in the set $\mathcal{S}$, then $\sharp\left((u \circ \gamma)^{-1}(z)\right)$ counts the number of intersections between γ and the hyperplane H_z. The following equation gives an integral-geometric interpretation of the length of a curve:

Lemma 26 (Cauchy-Crofton Formula [20, 69])**.** *The length of a closed curve is equal to π over the volume of the n-dimensional sphere times the integral over all $(n-1)$-planes H in $\mathbb{R}^n$ of the number of intersections between γ and H.*

$$\ell(\gamma) \;=\; \frac{\pi}{\mathrm{Vol}(\mathbb{S}^n)} \int_{\mathbb{S}^n - 1} \int_{-\infty}^{\infty} \sharp\left((u \circ \gamma)^{-1}(z)\right)\, dz\, du. \tag{3.5}$$

This equation is called the *Cauchy-Crofton formula*; see [20, 69]. A similar interpretation can be found for the total curvature.

Lemma 27 (Total Curvature Interpretation [20, 69])**.** *The total curvature of a closed curve equals π times the average number of critical points of $u \circ \gamma \colon \mathbb{S}^1 \to \mathbb{R}$.*

If c is the average number of critical points of $u \circ \gamma$, then we can summarize the previous theorem by the following equation:

$$\kappa(\gamma) = c\pi.$$

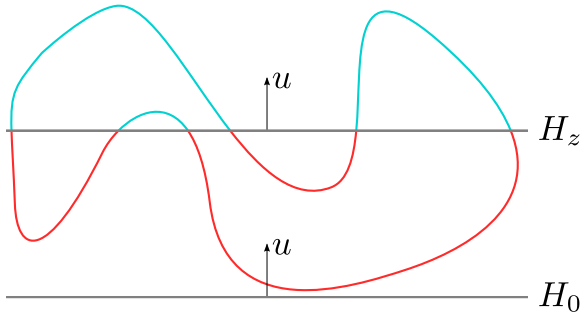


FIGURE 3.3: For almost all values $z \in \mathbb{R}$, the hyperplane H_z with normal vector u divides the curve γ into an equal number of components above and below the hyperplane. In this curve, we see a line that divides the curve into four arcs, two above and two below H_z.

3.3 Original Length Bound

In this section, we prove Theorem 19. The proof presented here is based on, but differs from, the original proof from [20]. We begin by restating the theorem here:

(Original Length Bound Theorem). *If γ_1 and γ_2 are two rectifiable closed curves in $\mathbb{R}^n$, then $|\ell_1 - \ell_2| \leq \frac{2\text{Vol}(\mathbb{S}^{n-1})}{\text{Vol}(\mathbb{S}^n)}(\kappa_1 + \kappa_2 - 2\pi)\mathcal{F}(\gamma_1, \gamma_2)$.*

Proof. We prove The Original Length Bound Theorem in three steps. First, we reformulate the Cauchy-Crofton formula 3.5 in terms of sublevel sets instead of level sets. Second, we express the inner integral of the Cauchy-Crofton formula for $\ell(\gamma_i)$ in terms of points in the persistence diagram of a height function of γ_i. Third, we apply the Bottleneck Stability Theorem to upper bound the difference of lengths of curves.

Given two closed curves γ_1 and γ_2 in $\mathbb{R}^n$, denote the length and total curvature of γ_i by ℓ_i and κ_i, respectively. For the first step, let u be a unit vector in $\mathbb{R}^n$ and let H_z denote the hyperplane with normal vector u whose signed distance from the origin is z. Notice that almost all height functions $u \circ \gamma_i$ are Morse. Assume $u \circ \gamma_i$ is Morse, which implies that the hyperplanes with normal vector u that break γ into an odd number of components form a measure zero subset, exactly correspond to the critical values of $u \circ \gamma_i$. Assuming z is a regular value, $\sharp((u \circ \gamma)^{-1}(z))$ is equal

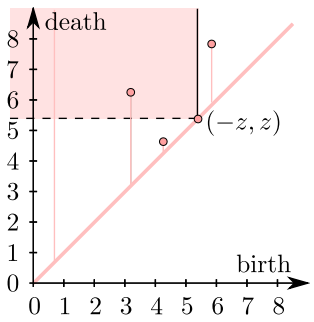


FIGURE 3.4: The indicator function $\mathbb{1}_a(z)$ evaluates to one if a is in the shaded region $[-\infty, z] \times (z, \infty]$.

to twice the number of components $c_i(u, z)$ of $(u \circ \gamma_i)^{-1}(-\infty, z]$, which means that we can rewrite (3.5) for the length of γ_i in terms of the number of components in sublevel sets:

$$\ell(\gamma_i) \quad = \quad \frac{2\pi}{\mathrm{Vol}(\mathbb{S}^n)} \int_{\mathbb{S}^n - 1} \int_{-\infty}^{\infty} c_i(u, z)\, dz\, du. \tag{3.6}$$

For the second step, we wish to express $c_i(u, z)$ in terms of the number of points in the persistence diagram for $u \circ \gamma_i$. Each component in $(u \circ \gamma_i)^{-1}(-\infty, z]$ corresponds to a 0-cycle that is born before z and dies after z. To help count these points in the persistence diagram, we define the indicator function $\mathbb{1}_a$ for $a = (a_1, a_2)$ as:

$$\mathbb{1}_a(z) \quad = \quad \begin{cases} 1, & a_1 \leq z \leq a_2, \\ 0, & \text{otherwise.} \end{cases} \tag{3.7}$$

In Figure 3.4, the function $\mathbb{1}_a(z)$ evaluates to 1 for all z such that a is contained in the region $[-\infty, z] \times (z, \infty]$ shaded in pink. Therefore, the number of components of the sublevel set $(u \circ \gamma_i)^{-1}(-\infty, z]$ is

$$c_i(u, z) = \sum_{a \in \mathrm{Dgm}(u_i)} \mathbb{1}_a(z). \tag{3.8}$$

For the third step of this proof, let $\epsilon > 0$. By Corollary 4, we can reparametrize γ_2 by $\varphi \colon \mathbb{S}^1 \to \mathbb{S}^1$ such that $\|\gamma_1(t) - \gamma_2(\varphi(t))\| \leq \mathcal{F}(\gamma_1, \gamma_2) + \epsilon$ for all $t \in \mathbb{S}^1$. We

apply Theorem 10 and find

$$\begin{aligned} W_\infty(\mathrm{Dgm}_0(\gamma_1), \mathrm{Dgm}_0(\gamma_2)) &= W_\infty(\mathrm{Dgm}_0(\gamma_1), \mathrm{Dgm}_0(\gamma_2 \circ \varphi)) \\ &\leq \mathcal{F}(\gamma_1, \gamma_2) + \epsilon \end{aligned}$$

Let $\psi \colon \mathrm{Dgm}_0(\gamma_1) \to \mathrm{Dgm}_0(\gamma_2)$ be the bijection that realizes the bottleneck distance between the persistence diagrams. Plugging (3.8) into (3.6) and taking the difference between lengths of curves results in

$$\begin{aligned} |\ell(\gamma_1) - \ell(\gamma_2)| &\leq \frac{2\pi}{\mathrm{Vol}(\mathbb{S}^n)} \int_{\mathbb{S}^n - 1} \int_{-\infty}^{\infty} \sum_{a \in \mathrm{Dgm}_0(u_1)} \left| \mathbb{1}_a(z) - \mathbb{1}_{\psi(a)}(z) \right| \, dz \, du \\ &= \frac{2\pi}{\mathrm{Vol}(\mathbb{S}^n)} \int_{\mathbb{S}^n - 1} \sum_{a \in \mathrm{Dgm}_0(u_1)} \int_{-\infty}^{\infty} \left| \mathbb{1}_a(z) - \mathbb{1}_{\psi(a)}(z) \right| \, dz \, du. \end{aligned}$$

In the worst case, each term of the above sum is $2(\mathcal{F}(\gamma_1, \gamma_2) + \epsilon)$. If both a and $\psi(a)$ are diagonal points, then $\mathbb{1}_a(z) = \mathbb{1}_{\psi(a)}(z) = 0$. Additionally, there are two points a_∞ and b_∞ of $\mathrm{Dgm}_0(u_1)$ and $\mathrm{Dgm}_0(u_2)$ respectively that have one infinite coordinate each. In this case, $\psi(a_\infty) = b_\infty$ and $\int_{-\infty}^{\infty} |\mathbb{1}_{a_\infty}(z) - \mathbb{1}_{b_\infty}(z)| \, dz \leq \mathcal{F}(\gamma_1, \gamma_2) + \epsilon$. This implies that each critical point can contribute at most $\mathcal{F}(\gamma_1, \gamma_2) + \epsilon$ to the sum and the two global maxima do not contribute to the sum. Hence, we have

$$\sum_{a \in \mathrm{Dgm}_0(u_1)} \int_{-\infty}^{\infty} \left| \mathbb{1}_a(z) - \mathbb{1}_{\psi(a)}(z) \right| \, dz \leq (c_1 + c_2 - 2)(\mathcal{F}(\gamma_1, \gamma_2) + \epsilon),$$

where c_i is the the number of critical points of $u \circ \gamma_i$. Finally, we use Lemma 27 to obtain

$$|\ell(\gamma_1) - \ell(\gamma_2)| \leq \frac{2\mathrm{Vol}(\mathbb{S}^{n-1})}{\mathrm{Vol}(\mathbb{S}^n)} (\kappa_1 + \kappa_2 - 2\pi)(\mathcal{F}(\gamma_1, \gamma_2) + \epsilon).$$

Letting ϵ approach zero in the above inequality, we arrive at the desired result. ◻

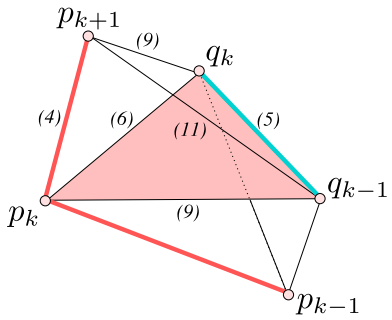


FIGURE 3.5: The first half of the inductive step in which we add p_{k+1} to the construction. In the next step, we will use $\triangle p_k q_k p_{k+1}$ to place q_{k+1} at the furthest distance from q_{k-1}.

3.4 Dimension-Independent Length Bound

We return to the main result of this chapter, the proof of Theorem 20. We begin by restating the theorem:

(Improved Length Bound Theorem)**.** *If γ_1 and γ_2 are rectifiable closed curves in $\mathbb{R}^n$, then $|\ell_1 - \ell_2| \leq \frac{4}{\pi}(\kappa_1 + \kappa_2)\mathcal{F}(\gamma_1, \gamma_2)$.*

Following Chakerian, we first turn polygonal chains in $\mathbb{R}^n$ into similar polygonal chains in $\mathbb{R}^3$, namely polygonal chains that have the same lengths and whose Fréchet distance and total curvatures are at most that of the corresponding polygonal chains in $\mathbb{R}^n$. Then, we apply the case $n = 3$ of Theorem 19 in which twice the volume ratio is $4/\pi$. We use the limit lemmas of the previous section in order to extend the result from polygons to closed curves.

Polygonal Chains. We begin with the case in which the two curves are (open) polygonal chains in $\mathbb{R}^n$, which we denote by P_1 and P_2. Letting $\epsilon > 0$, there exists a piecewise linear homeomorphism $f : P_1 \to P_2$ such that $\|x - f(x)\| \leq \mathcal{F}(P_1, P_2) + \epsilon$ for all $x \in P_1$; see Lemma 5. Writing V_j for the set of vertices of P_j, we are interested in $V = V_1 \cup f^{-1}(V_2)$. Assuming $m = |V|$, we write $v_0, v_1, \ldots, v_{m-1}$ for the m points in order along P_1. Since f is a homeomorphism, $f(v_0), f(v_1), \ldots, f(v_{m-1})$ are the corresponding points in order along P_2.

We now map V and $f(V)$ into $\mathbb{R}^3$. We start by isometrically embedding the first two vertices of both polygons, mapping v_0 to p_0, v_1 to p_1, $f(v_0)$ to q_0, and $f(v_1)$ to q_1. We map the remaining vertices inductively. Suppose v_0 through v_k and $f(v_0)$ through $f(v_k)$ have been mapped to $\mathbb{R}^3$ such that the following eight conditions are satisfied:

$$\|v_i - v_{i-1}\| = \|p_i - p_{i-1}\|, \text{ for } 1 \leq i \leq k, \tag{3.9}$$

$$\|f(v_i) - f(v_{i-1})\| = \|q_i - q_{i-1}\|, \text{ for } 1 \leq i \leq k, \tag{3.10}$$

$$\|v_i - f(v_i)\| = \|p_i - q_i\|, \text{ for } 0 \leq i \leq k, \tag{3.11}$$

$$\|v_i - v_{i-2}\| \leq \|p_i - p_{i-2}\|, \text{ for } 2 \leq i \leq k, \tag{3.12}$$

$$\|f(v_i) - f(v_{i-2})\| \leq \|q_i - q_{i-2}\|, \text{ for } 2 \leq i \leq k. \tag{3.13}$$

$$\|v_i - f(v_{i-1})\| = \|p_i - q_{i-1}\|, \text{ for } 1 \leq i \leq k, \tag{3.14}$$

$$\|v_{i-1} - f(v_i)\| = \|p_{i-1} - q_i\|, \text{ for } 1 \leq i \leq k, \tag{3.15}$$

$$\|v_i - f(v_{i-2})\| = \|p_i - q_{i-2}\|, \text{ for } 2 \leq i \leq k, \tag{3.16}$$

The first two equalities ensure that the total length of each curve is preserved. Equation (3.11) ensures that the Fréchet distance at the vertices, and hence the edges, does not increase. The next two inequalities guarantee that turning angles do not increase. Along with (3.9), (3.10), and (3.11), the final three equalities are needed for the inductive arguments.

Next, we explain how to construct p_{k+1} and q_{k+1}. Generally speaking, we map p_{k+1} (and q_{k+1}) to a vertex that maximizes its distance to p_{k-1} by choosing a point on the opposite side of $\triangle p_k q_k q_{k-1}$ ($\triangle p_{k+1} q_k p_k$). We use (3.10), (3.11) and (3.14) for $i = k$, which we get by inductive assumption, to obtain $\triangle p_k q_k q_{k-1}$. We can therefore map v_{k+1} to a point p_{k+1} such that (3.9), (3.14) and (3.16) are satisfied for $i = k+1$; see the solid blue line segments ending at p_{k+1} in Figure 3.6. There are two such points p_{k+1} and we choose the one such that the new tetrahedron, $p_{k+1} q_k p_k q_{k-1}$,

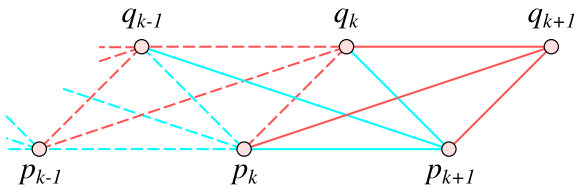


FIGURE 3.6: When adding the next vertex, the distance to the antecedent three vertices is fixed. These distances are depicted by line segments connecting vertices.

and the old tetrahedron, $q_k p_k q_{k-1} p_{k-1}$, overlap only on the shared face defined by $q_k p_k q_{k-1}$. By Lemma 21, this is also the configuration that maximizes the distance between p_{k-1} and p_{k+1} without distorting either tetrahedra. This gives us (3.12) for $i = k+1$. Finally, we use the same method to construct q_{k+1} using (3.9) and (3.14) for $i = k+1$ and (3.11) for $i = k$; see the solid red line segments ending at q_{k+1} in Figure 3.6. We choose q_{k+1} so that (3.10), (3.11), (3.13) and (3.15) are satisfied for $i = k+1$. This completes the inductive step and thus the construction of the two open polygonal curves in $\mathbb{R}^3$.

Closed Polygons. Assume now that P_1 and P_2 are (closed) polygons in $\mathbb{R}^n$. This does not imply that their images in $\mathbb{R}^3$ are closed. Since we aim at using Theorem 19, which is for closed curves, we must artificially close the images, but there is an advantage to first develop the curves some more. Specifically, we choose an integer $N \geq 1$ and iterate the construction of polygonal chains in $\mathbb{R}^3$ by going around the two polygons N times. Writing $v_j = v_{j \bmod N}$ for all $0 \leq j \leq Nm$, we construct p_j and q_j from v_j and $f(v_j)$, as before. The length of the polygonal chain from p_0 to p_{Nm} is therefore N times the length of P_1. Similarly, the length of the polygonal chain from q_0 to q_{Nm} is N times the length of P_2. Furthermore, the sums of turning angles at the interior vertices is at most N times the sums for P_1 and for P_2. Finally, because of (3.11) and Lemma 5, the Fréchet distance between the polygonal chains in

$\mathbb{R}^3$ is less than or equal to $\mathcal{F}(P_1, P_2) + \epsilon$. To close the polygons in $\mathbb{R}^3$, we double each open curve, traversing the second copy from back to front. This operation doubles the lengths, preserves the Fréchet distance, doubles the sums of turning angles, and adds 2π to each sum to account for the two full turns at the first and the last vertices. Using Theorem 19 for $n = 3$, we thus get

$$|2N\ell_1 - 2N\ell_2| \leq \frac{4}{\pi}|2N\kappa_1 + 2\pi + 2N\kappa_2 + 2\pi - 2\pi|(\mathcal{F}(P_1, P_2) + \epsilon).$$

Dividing by $2N$, letting N go to infinity and letting ϵ go to zero, we obtain Theorem 20 for polygons in $\mathbb{R}^n$.

Smooth Closed Curves. We finally consider the case of two smooth curves, γ_1 and γ_2, in $\mathbb{R}^n$. Let $\delta > 0$. By Lemmas 23, 24 and 25, there are inscribed polygons $P_1 < \gamma_1$ and $P_2 < \gamma_2$ such that

$$\begin{aligned} |\ell(P_j) - \ell(\gamma_j)| &\leq \delta, \\ |\kappa(P_j) - \kappa(\gamma_j)| &\leq \delta, \\ |\mathcal{F}(P_1, P_2) - \mathcal{F}(\gamma_1, \gamma_2)| &\leq \delta, \end{aligned}$$

for $j = 1, 2$. Writing $\ell_j = \ell(\gamma_j)$ and $\kappa_j = \kappa(\gamma_j)$ and using Theorem 20 for the polygons, we thus obtain

$$\begin{aligned} |\ell_1 - \ell_2| &\leq |\ell(P_1) - \ell(P_2)| + 2\delta \\ &\leq \frac{4}{\pi}|\kappa(P_1) + \kappa(P_2)|\mathcal{F}(P_1, P_2) + 2\delta \\ &\leq \frac{4}{\pi}|\kappa_1 + \kappa_2 + 2\delta|(\mathcal{F}(\gamma_1, \gamma_2) + \delta) + 2\delta. \end{aligned}$$

Since $\kappa_1 + \kappa_2$ and $\mathcal{F}(\gamma_1, \gamma_2)$ are both finite, Theorem 20 for smooth closed curves follows by letting δ go to zero.

3.5 Discussion

For closed curves in $\mathbb{R}^n$, Theorem 20 provides an upper bound for the length difference without including the diverging ratio of volumes of spheres. This upper bound is an improvement over the best previous result in all dimensions $n > 3$. However, it is still an open question as to whether or not this is a tight bound on the difference of curve lengths. Is it possible that $|\ell_1 - \ell_2| \leq (\kappa_1 + \kappa_2 - 2\pi)\mathcal{F}(\gamma_1, \gamma_2)$ in all dimensions?

The Fáry-Chakerian Theorem (Theorem 18) can be found as a consequence of the Original Length Bound (Theorem 19) for $n = 2$. To see this, let γ_1 and γ_2 be two curves contained in a ball with diameter r. If γ_1 is a circle with radius ϵ, then we know that $\ell_1 = 2\pi\epsilon$, $\kappa_1 = 2\pi$ and the Fréchet distance between γ_1 and γ_2 is at most r. Plugging this into the Original Length Bound, we have:

$$\|2\pi\epsilon - \ell_2\| \quad \leq \quad r\kappa_2.$$

Letting ϵ approach zero, we arrive at Fáry-Chakerian Theorem, which is known to be sharp. Plugging these equations into the Improved Length bound gives us:

$$\ell_2 \quad \leq \quad r(\kappa_2 + 2\pi),$$

which is a weaker formulation of the Fáry-Chakerian Theorem. We ask: *What is the infimum over $c \in \mathbb{R}$ such that $|\ell_1 - \ell_2| \leq c(\kappa_1 + \kappa_2)\mathcal{F}(\gamma_1, \gamma_2)$ in all dimensions?* This remains an open question. Applying the Fáry-Chakerian Theorem for two curves contained in the disk of radius r gives us an upper bound on c: $c \leq \frac{r}{\mathcal{F}(\gamma_1, \gamma_2)}$ since

$$\begin{aligned} \|\ell_1 - \ell_2\| &\leq \|\ell_1 + \ell_2\| \\ &\leq r(\kappa_1 + \kappa_2). \end{aligned}$$

This bound on c may be surprising, on account of c increasing as the Fréchet distance decreases. The value c is not unbounded, however, as we have shown $c \leq \frac{4}{\pi}$.

4

Geometry and Topology of Gaussian Mixtures

The process of diffusing chemical substances, such as hormones, and physical quantities, such as temperature, is a phenomenon. Assuming a uniform medium, the process is described by the solution to the heat equation. In Euclidean space, solving this equation is synonymous to convolving with a Gaussian kernel. Convolution is also a popular computational method, in particular in computer vision, where the one-parameter family of convolutions of a given image is known as its *scale space*; see [61, 85].

We are interested in the quantitative analysis of diffusion and Gaussian convolution. In particular, we study the evolution of the critical points of a function that is convolved with a progressively wider Gaussian kernel. If the function is one-dimensional, i.e. from $\mathbb{R}$ to $\mathbb{R}$, then Gaussian convolution does not create new critical points [7, 54, 66, 86]. As a consequence, the diffusion of m point masses (a sum of m Dirac delta functions) cannot have more than m modes (local maxima); see [9, 15, 72]. For two- or higher-dimensional functions, more modes than point masses can be created through Gaussian convolution; see [59] for a two-dimensional function for which diffusion creates modes and [26, 65] for a mathematical analysis of

the unfolding events that cause this effect. It has been observed that these events are rare in practice [57, 58] and confirmed that the ability to create critical points with non-negligible persistence deteriorates rapidly [18]. In this chapter and the next, we focus our discussions on Gaussian convolution applied to Dirac delta functions and to weighted sums of Dirac delta functions. As we will see, this setting is equivalent to studying the sum of Gaussian kernels.

4.1 Curve Analysis

In this subsection, we introduce Gaussian kernels and discuss some of their fundamental properties.

Gaussian Kernels and Derivatives. We call a real-valued function $f\colon \mathbb{R}^{n+1} \to \mathbb{R}$ of the form $f(x) = We^{-C\|x-\mu\|^2}$ an $(n+1)$-*dimensional isotropic Gaussian kernel*, where W and C are real constants, $\mu \in \mathbb{R}^{n+1}$ is a point and $\|x-\mu\|$ denotes the Euclidean distance between the two points. We call the kernel *normalized* if it integrates to unity, in which case it can be written as:

$$\mathcal{N}_{n+1}(x-\mu,\sigma^2) \;=\; \frac{1}{(2\pi\sigma^2)^{\frac{n+1}{2}}} \cdot e^{-\frac{\|x-\mu\|^2}{2\sigma^2}}. \tag{4.1}$$

The normalization constant $-1/(2\pi\sigma^2)^{\frac{n+1}{2}}$ ensures that the integral of the Gaussian kernel over the domain is unity:

$$\int_{x\in\mathbb{R}^{n+1}} \mathcal{N}_{n+1}(x-\mu,\sigma^2)\,dx = 1. \tag{4.2}$$

For convenience, we define $g_\mu(x,\sigma) := \mathcal{N}_{n+1}(x-\mu,\sigma^2)$ when n is fixed. We call $\mu \in \mathbb{R}^{n+1}$ the *center* or *mean* and $\sigma > 0$ the *width* (or *standard deviation* if $n = 0$). For an isotropic kernel, we call the square width the *variance* of the kernel. Finally, g_μ is *unit* if it is normalized with *height*, or maximum value, one. Independent of the

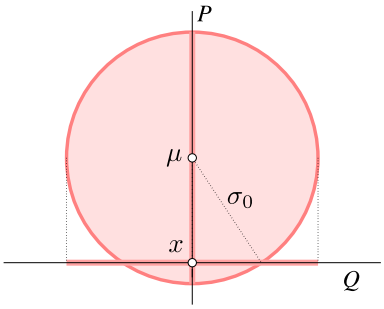


FIGURE 4.1: A unit Gaussian kernel g_μ with center μ in $\mathbb{R}^2$ represented by the disk with radius $\sigma_0 = 1/\sqrt{2\pi}$. The restriction of g_μ to the lines P and Q are called 1-dimensional sections of the Gaussian kernel. Both sections are 1-dimensional Gaussian kernels. The restriction to P has the additional property that it is normalized.

dimension, the width and the formula of the unit Gaussian kernel are:

$$\begin{aligned} \sigma_0 &= 1/\sqrt{2\pi}, \\ g_\mu(x, \sigma_0) &= e^{-\pi\|x-\mu\|^2}. \end{aligned} \tag{4.3}$$

We use $g(x)$ to denote the unit kernel centered at the origin, which will simplify our computations. The unit Gaussian centered at the origin can be transformed into every other Gaussian by a translation (to change the center), a scaling of the domain (to change the width), and a scaling of the range (to change the height). Specifically, we go between variances $t_0 \leq t_1$ by first stretching the distribution along each basis direction b_i of the domain by a factor of $\sqrt[n+1]{t_1/t_0}$ and second shrinking it along the range by t_0/t_1. The result is an isotropic function such that the integral over the entire domain is one, as in (4.1). Indeed, we have

$$\mathcal{N}_{n+1}(x, t_1) = \sqrt[n+1]{\frac{t_0}{t_1}} g\left(\frac{x}{t_1/t_0}, t_0\right).$$

For the case $n = 0$, we can write formulas of its first few derivatives:

$$g'(x) = [-2\pi x] \cdot g(x), \tag{4.4}$$

$$g''(x) = [4\pi^2 x^2 - 2\pi] \cdot g(x). \tag{4.5}$$

$$g^{(3)}(x) = [-4\pi^2 x][2\pi x^2 + 3] g(x) \tag{4.6}$$

Note that the first derivative of the Gaussian kernel is written in a convenient form: as a function of the Gaussian kernel itself. In fact, the Gaussian kernel is the only function that satisfies $g'(x) = [-2\pi x] \cdot g(x)$:

Lemma 28 (Picard-Lindelöf)**.** *Let $a, c \in \mathbb{R}$. Then, the unique solution to $f'(x) = cxf(x)$ with $f(0) = a$ is $f(x) = ae^{\frac{c}{2}x^2}$.*

This lemma is a direct consequence of Picard iteration as presented in [64], but can also be found in many ordinary differential equation textbooks; for example, see [10]. We also observe that, in the one-dimensional case, g'' is negative in the interior of $[-\sigma_0, \sigma_0]$ and positive outside this interval. Accordingly, in one and in higher dimensions, g is concave at every point in the interior of the closed ball with center at the origin and radius σ_0, and it fails to be concave at every point outside this ball. This ball is a convenient illustration of the kernel; see Figure 4.1. Furthermore, we notice that $(-x)^m g^{(m)}(x)$ is positive for large enough values of $|x|$. In fact, this end behavior is true regardless of μ and σ. The following lemma quantifies the limiting behavior as σ approaches zero.

Lemma 29 (Limiting to Delta Functions)**.** *In the limit as the width approaches zero, the Gaussian kernel approaches a Dirac delta function:*

$$\lim_{\sigma \to 0} \mathcal{N}_{n+1}(x - \mu, \sigma^2) = \delta_\mu(x).$$

Proof. We prove this lemma point-wise. First notice, that $\lim_{\sigma \to 0} \mathcal{N}_{n+1}(0, \sigma^2) = \infty$ since $\mathcal{N}_{n+1}(0, \sigma^2) = (2\pi\sigma^2)^{-(n+1)/2}$. Additionally, for all $x \neq \mu$, the Gaussian kernel approaches zero as σ decreases to zero; that is, $\lim_{\sigma \to 0} \mathcal{N}_{n+1}(x - \mu, \sigma^2) = 0$. ◻

Sections. A *k-section* of a Gaussian kernel is the restriction to a k-dimensional plane P. Assuming a unit Gaussian kernel, we can write this as

$$\begin{aligned} g_\mu|_P(x,\sigma_0) &= \frac{1}{(2\pi\sigma_0^2)^{\frac{n+1}{2}}} \cdot e^{-\frac{\|x-\mu\|^2}{2\sigma_0^2}} \\ &= \frac{1}{(2\pi\sigma_0^2)^{\frac{n-k+1}{2}}} \cdot e^{-\frac{\|y-\mu\|^2}{2\sigma_0^2}} \cdot \frac{1}{(2\pi\sigma_0^2)^{\frac{k}{2}}} \cdot e^{-\frac{\|x-y\|^2}{2\sigma_0^2}} \\ &= \mathcal{N}_{n-k+1}(y-\mu,\sigma_0) \cdot \mathcal{N}_k(x-y,\sigma_0). \end{aligned}$$

where y is the orthogonal projection of μ onto P. We call the first factor $\mathcal{N}_{n-k+1}(y-\mu,\sigma_0)$ the *weight* since it is constant for all x in P and we call $g_\mu|_P$ a *weighted* unit Gaussian kernel. Importantly, we see that the section has the same width as the original kernel. The restriction is a unit kernel itself iff P passes through μ. In this case, $g_\mu|_P(x)$ is the weight of the $(n-k+1)$-section defined by the plane Q that intersects P orthogonally at x; see Figure 4.1. Iterating this construction, we write the $(n+1)$-dimensional unit Gaussian kernel as a product of one-dimensional kernels:

$$g_\mu(x) = \prod_{i=0}^{n} g(x_i - \mu_i), \tag{4.7}$$

where the x_i and μ_i are the Cartesian coordinates of x and μ respectively. In words, the high-dimensional unit kernel can be *separated* into mutually orthogonal one-dimensional unit kernels. The separability of the Gaussian kernel is one of the Axioms of scale space, which we will discuss in Subsection 4.2.2.

Balanced Sums. Consider the sum of two unit one-dimensional Gaussian kernels. For symmetry, we choose their centers at distance $\mu \geq 0$ to the left and right of the origin. As proven in [12], $G = g_{-\mu} + g_\mu$ has either one or three critical points and no other number is possible. More specifically, G has one maximum iff $\mu \leq \sigma_0$, G has two maxima and one minimum iff $\mu > \sigma_0$ and G has a degenerate maximum iff $\mu = \sigma_0$. We present our own proof of this result, as we need the concepts it uses.

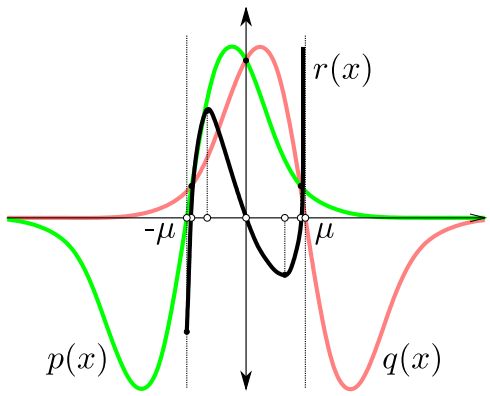


FIGURE 4.2: The graphs of the derivatives of $-g_{-\mu}$ and of g_μ intersect above the roots of the (black) ratio function.

Since $G' = g'_{-\mu} + g'_\mu$, a point $x \in \mathbb{R}$ is a critical point of G iff the graphs of $p = -g'_{-\mu}$ and $q = g'_\mu$ intersect above x. Since g' is an odd function, the number of intersections (counted with multiplicities) must be odd, and it is visually plausible that the number can only be one or three. To be sure, we introduce the *ratio function*, $r = \frac{p}{q} - 1$, with formula

$$r(x) \quad = \quad \frac{\mu + x}{\mu - x} \cdot e^{-4\pi\mu \cdot x} - 1; \tag{4.8}$$

see Figure 4.2. Setting $r(x) = 0$ gives us the intersections of p and q and thus the critical points of G. For simple geometric reasons, the roots of r are necessarily in $[-\mu, \mu]$. Independent of μ, we have $r(0) = 0$ since $p(0) = q(0)$. To see whether there are additional roots, we take the derivative:

$$r'(x) \quad = \quad \frac{2\mu - 4\pi\mu(\mu^2 - x^2)}{(\mu - x)^2} \cdot e^{-4\pi\mu \cdot x}.$$

Setting $r'(x) = 0$, we have $x^2 = \mu^2 - \frac{1}{2\pi}$, which has two real solutions if $\mu > \sigma_0$ and no solution if $\mu < \sigma_0$. Observing that $r(-\mu) = -1$ and $r(\mu)$ goes to infinity as x approaches μ, we conclude that r has one root if $\mu \leq \sigma_0$ and three roots if $\mu > \sigma_0$. As anticipated, G has one critical point if $\mu \leq \sigma_0$ and it has three non-degenerate critical points if $\mu > \sigma_0$.

Unbalanced Sums. Next, we study sums $G_w = g_{-\mu} + wg_\mu$, where $w \geq 0$ is the *weight* of the second term. The number of critical points of G_w is at most three, but in contrast to the balanced case, it can also be two, as we will see. More specifically, Behboodian [9] gives necessary and sufficient conditions for all three cases (one, two, or three critical points), but they are not as easy to state as in the balanced case. As before, we present our own proof since we need the concepts it uses.

Recalling that the critical points of G_w correspond to the intersections between the graphs of $p = -g'_{-\mu}$ and $wq = wg'_\mu$, we consider their ratio function, which is

$$r_w \;=\; \frac{p}{wq} - 1 \;=\; \frac{r}{w} - \left(1 - \frac{1}{w}\right). \tag{4.9}$$

Notice that r_w vanishes at x iff $p(x) = wq(x)$, in which case $w = r(x) + 1$. Its derivative is $r'_w = \frac{r'}{w}$, which has at most two roots. Similarly, $r_w(-\mu) = -1$ and $r_w(\mu)$ goes to infinity as x approaches μ, so we see that r_w has at most three roots. It follows that G_w has at most three critical points, just like G.

A new phenomenon is the possibility of two critical points. To see when this case arises, we set $w = r(x) + 1$ and note that for this choice of weight $r_w(x) = 0$. In words, p and wq intersect above x and, equivalently, x is a critical point of G_w. If x has the additional property of being critical for r, the intersection between p and wq is degenerate. As computed above, the critical points of r_w are given by $x^2 = \mu^2 - \sigma_0^2$. Let $x_1 = -\sqrt{\mu^2 - \sigma_0^2}$ and $x_2 = \sqrt{\mu^2 - \sigma_0^2}$ be the two solutions, and note that x_1 gives a weight $w_1 = r(x_1) + 1$ that is larger than one, while x_2 gives a weight $w_2 = r(x_2) + 1$ between 0 and one. We call w_1 and w_2 the *transition weights* for μ, remembering that they exist iff $\mu \geq \sigma_0$.

We illustrate the case analysis by observing how the shape of G_w changes for a fixed weight and moving centers. We get qualitatively the same behavior for every positive weight, so we take $w = \frac{1}{2}$. Let 2ζ be the distance between the two centers

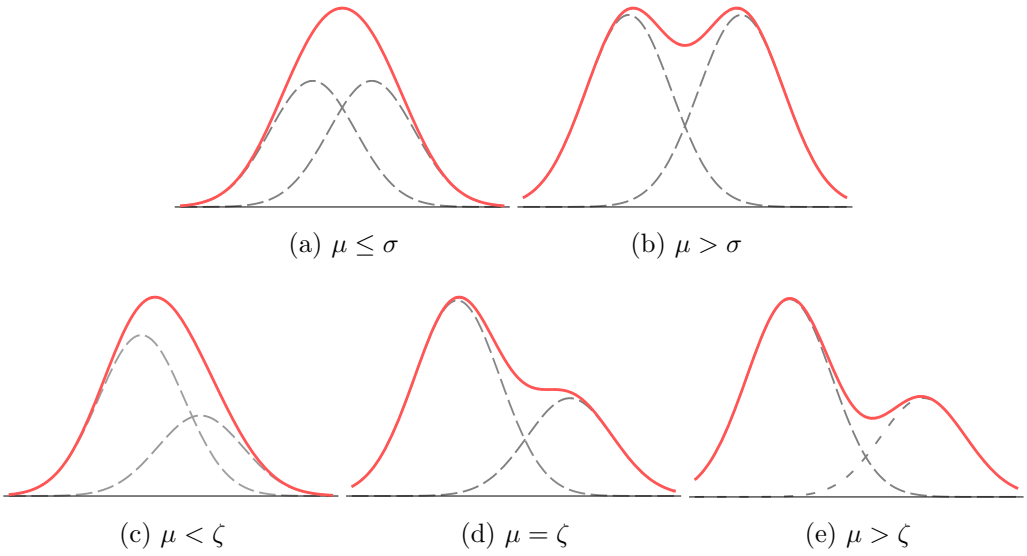


FIGURE 4.3: In the first row we see the balanced sum of two Gaussian kernels and in the second row we see the weighted sum of two Gaussian kernels. Progressing from left to right, we see the evolution of the sum as μ increases. The component kernels are shown in gray and the resulting Gaussian mixture is in red.

for which w is a transition weight. We find ζ by solving $w = r(x_2) + 1$ for μ:

$$\zeta = \sqrt{\frac{2\pi\sigma_0^2 + \sqrt{2^2\pi^2\sigma_0^4 + \ln^2(1-w)}}{2^3\pi}}.$$

Starting with $-\mu = \mu = 0$, we see the evolution sketched in Figure 4.3. For $\mu < \zeta$, the function G_w has only one maximum. At $\mu = \zeta$, we have a degenerate critical point forming a shoulder on the right. For $\mu > \zeta$, this shoulder turns into a min-max pair. This is a generic event in the one-parameter evolution of a Morse function, known as an anti-cancellation.

Whether balanced or unbalanced sums are taken, we have seen that the sum of two one-dimensional Gaussians has either one or two local maxima. In general, the number of local maxima does not exceed the number of Gaussian kernels in a Gaussian mixture over $\mathbb{R}^1$. We postpone the proof of this claim to the end of the next section, as the proof utilizes terminology from scale space.

4.2 Scale Space

The *deep structure*, or *scale space*, of an image f is a family of images, parametrized by t, that describe f at different scales. With $t = 0$, we have the original image f, and as t increases, f appears to blur. The blurring is governed by the heat equation, computed by convolving the image with the Gaussian kernel.

4.2.1 A Fixed Scale

A grayscale image h_0 is a function from a rectangular domain $\mathbb{M} \subset \mathbb{R}^n$ to $\mathbb{Z}_{256} = \{0, 1, 2, \ldots, 255\}$. One can also color images, i.e., maps $\mathbb{M} \to \mathbb{Z}_{256}^3$, but we will limit our discussion to images with range $\mathbb{Z}_{256}$. Thinking of the value $h_0(x)$ at a point $x \in \mathbb{M}$ as *heat*, we consider the diffusion of heat over time. We call this the study of the *deep structure* or the *scale space* of the image h_0. This process creates a family of parametrized functions, which can be used to understand the image. In [41], we defined the heat equation homotopy, which uses the scale space of the difference between images in order to create a discrete homotopy between the images. Still, there exists a gap between human perception and computer vision. Scale space to provides a mathematical description of an image at all scales, attempting to quantify the natural way that humans view the world; see [86] for a brief discussion of the psychophysics behind human vision and the connection to computer vision.

The Heat Equation. We describe the heat equation as it applies to a continuous function. A more detailed account can be found in [6, 11]. In Section 4.2.4, we present a discretization of the heat equation homotopy.

Let $\{b_1, b_2, \ldots, b_m\}$ be an orthonormal basis for $T_p\mathbb{M}$, where $\mathbb{M}$ is a closed subspace of $\mathbb{R}^m$ and $p \in \mathbb{M}$. Let f be a real-valued, twice differentiable function on $\mathbb{M}$. We define the Laplace-Beltrami operator as the following second order elliptic operator:

$$\Delta_{\mathbb{M}}(f) = \operatorname{div}(\nabla(f)). \tag{4.10}$$

In words, the Laplace-Beltrami operator applied to a function f is the divergence of the gradient of f. In Euclidean space, the Laplace-Beltrami operator is often referred to as simply the *Laplace* operator and we can work out an explicit formula:

$$\triangle\ f(x) = \sum_{i=0}^{m-1} \frac{\partial^2 f}{\partial e_i^2}(x),$$

where e_i are the standard basis elements for $\mathbb{R}^m$. To understand this equation as it applies to a manifold, we must first define the divergence and the gradient on the Riemannian manifold. Using local coordinates, we define the gradient on the manifold:

$$\triangledown(f) = E_{ij}^{-1} \frac{\partial f}{\partial b^i} \frac{\partial}{\partial b^j}, \tag{4.11}$$

where E is the essential matrix at p; see (2.2). Writing $\triangledown(f) = a^i \frac{\partial}{\partial b^j}$, then the divergence is defined by:

$$\mathrm{div}(X) = \frac{1}{\sqrt{|E|}} \frac{\partial(\sqrt{|E|}a^j)}{\partial b^j} \tag{4.12}$$

Now, we have fully defined (4.10) for a Riemannian manifold. The general form of the heat equation satisfies the following two conditions:

$$\frac{\partial u}{\partial t}(x,t) - \triangle_{\mathbb{M}}\ u(x,t) \quad = \quad 0 \tag{4.13}$$

and

$$u(x,0) \quad = \quad f(x), \quad \forall x \in \mathbb{M}. \tag{4.14}$$

If we can solve this partial differential equation (PDE), we obtain $u(x,t)$ defined for all $x \in \mathbb{M}$ and $t \geq 0$. Typically, the heat equation has the additional constraint that $u(x,t)$ is constant with respect to t for all $x \in \partial\mathbb{M}$. This is an example of the *Dirichlet boundary condition* that fixes the values in $\partial\mathbb{M}$. However, we are interested in the case

where $u(x,t)$ is *heat conserving*; that is, we would like $\text{avg}_x(u(x,t_1)) = \text{avg}_x(u(x,t_2))$ for all t_1, t_2. As we will show, in order to obtain this goal, the values on the boundary will depend on the interior values.

If we think of $u(x,t)$ as the temperature at location x at time t, then (4.13) governs how the temperature changes with respect to time and space. We note that if we impose the condition $\frac{\partial u}{\partial t}(x,t) = 0$, then the heat equation will not change with respect to time and (4.13) becomes Laplace's equation $\Delta u(x,t) = 0$. This setting is known as the steady-state heat equation and has a unique solution. The iterative methods that we look at in Section 4.2.4 seeks an approximation of $u(x,t)$ for this problem. We are interested in studying the behavior of heat equation as it approaches the solution to the steady-state heat equation. As we will see next, the solution to the heat equation at time t is equivalent to convolving a function with the Gaussian kernel defined in (4.1). Thus, for all $t > 0$, the function $u(x,t)$ is smooth and singularities are isolated.

Gaussian Convolution. Scale space dates back to the 1960s with a publication in Japanese by Taizo Iijima [53]. His work, however, did not reach the western world until it was summarized in English in [83]. Thus, the concept of scale space was introduced to the vision community by Witkin [84, 85] and was connected to the heat equation by Koenderink [54]. Scale space is constructed by blurring an image with a one-parameter family of Gaussian kernels. We then blur an image $h_0 \colon \mathbb{R}^n \to [0, 255]$ at a point $x \in \mathbb{R}^n$ by integrating the values of h_0, weighting each point by its value under the Gaussian kernel centered at x:

$$\mathcal{T}_t h_0(x) = \int_{y \in \mathbb{R}^n} \mathcal{N}_n(y - x, t) h_0(y)\, dy. \tag{4.15}$$

This process is called *Gaussian convolution*. To allow for a more general framework, we replace $\mathcal{N}_n(\cdot, t)$ with a different kernel $K_t(\cdot)$ to obtain a general definition

of *convolution*:

$$[K_t * h_0](x) = \int_{y \in \mathbb{R}^n} K_t(y - x) h_0(y)\, dy. \tag{4.16}$$

For now, we focus on Gaussian convolution. The variance t of the Gaussian kernel is the *scale parameter* of the scale space. Let $f = \mathcal{T}_t h_0$. We call $\mathcal{T}_t^{-1}$ the *deconvolution operator* if applying it to the convolved function f restores the original function: $\mathcal{T}_t^{-1} f = h_0$. In general, the problem of finding the deconvolution operator is an ill-posed problem [78]. A stable version of deconvolution for polynomials of a fixed finite degree is presented in [52] and for images using wavelets is presented in [80]. In some formulations, the deconvolution operator can be defined by convolution with the deconvolution kernel $D(x, t)$:

$$\mathcal{T}_t^{-1} f(x) = \int_{y \in \mathbb{R}^n} D(y - x, t) h_0(y)\, dy.$$

We simplify the setting by requiring f to be a convolved function. For $t > s > 0$, we define:

$$\mathcal{T}_s^{-1} \mathcal{T}_t f(x) := \mathcal{T}_{t-s} f(x).$$

In words, deconvolution reverses the process of convolution.

Let $g(x, t) = \mathcal{N}_n(x, t)$. We now wish to verify that $g(x, t)$ is a solution to the heat equation; that is, we will show $\partial g / \partial t - \triangle\, g = 0$. First, we compute the derivative of

$g(x,t)$ with respect to t, then we will compute the Laplacian operator.

$$\begin{aligned}
\frac{\partial g}{\partial t}(x,t) &= \frac{1}{(\sqrt{2\pi t})^n}\frac{\partial\left(e^{-\frac{|x|^2}{2t}}\right)}{\partial t} + \frac{\partial\left((\sqrt{2\pi t})^{-n}\right)}{\partial t}e^{-\frac{|x|^2}{2t}} \\
&= \frac{1}{(\sqrt{2\pi t})^n}\left(\frac{\|x\|^2}{2t^2}\right)e^{-\frac{\|x\|^2}{2t}} + (-n\pi)(2\pi t)^{-n/2-1}e^{-\frac{\|x\|^2}{2t}} \\
&= g(x,t)\left(\frac{\|x\|^2}{2t^2}\right) + (-n\pi)(2\pi t)^{-1}g(x,t) \\
&= g(x,t)\left(\frac{\|x\|^2}{2t^2}\right) - \frac{n}{2t}g(x,t) \\
&= g(x,t)\left(\frac{\|x\|^2 - nt}{2t^2}\right). \qquad (4.17)
\end{aligned}$$

Now, we take the first partial derivative of $g(x,t)$ with respect to a basis vector b_i:

$$\begin{aligned}
\frac{\partial g}{\partial b_i}(x,t) &= -\frac{1}{(\sqrt{2\pi t})^n}\frac{\partial\left(e^{-\frac{|x|^2}{2t}}\right)}{\partial b_i} - \frac{\partial\left((\sqrt{2\pi t})^{-n}\right)}{\partial b_i}e^{-\frac{|x|^2}{2t}} \\
&= -\frac{1}{(\sqrt{2\pi t})^n}e^{-\frac{|x|^2}{2t}}\cdot\frac{\partial\left(-\frac{|x|^2}{2t}\right)}{\partial b_i} \\
&= g(x,t)\cdot\frac{-x_i}{t}.
\end{aligned}$$

We take the derivative of this result to obtain one summand of (4.10):

$$\begin{aligned}
\frac{\partial^2 g}{\partial b_i^2}(x,t) &= \frac{\partial\left(g(x,t)\cdot\frac{-x_i}{t}\right)}{\partial b_i} \\
&= g(x,t)\frac{\partial\left(\frac{-x_i}{t}\right)}{\partial b_i} + \frac{-x_i}{t}\frac{\partial g(x,t)}{\partial b_i} \\
&= g(x,t)\frac{-1}{t} + \frac{x_i^2}{t^2}g(x,t) \\
&= g(x,t)\cdot\frac{x_i^2 - t}{2t^2}. \qquad (4.18)
\end{aligned}$$

We use (4.17) and (4.18) for $i = 1, 2, \ldots, n$ to obtain

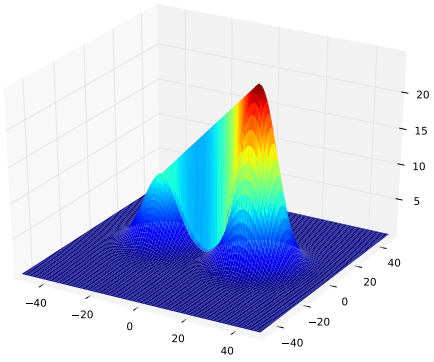
$$\frac{\partial g}{\partial t}(x,t) - \sum_{i=1}^{n} \frac{\partial^2 u}{\partial b_i^2}(x,t) = 0. \tag{4.19}$$

Comparing (4.19) with (4.13), we see that the heat equation is satisfied. In fact, the Gaussian kernel is the *fundamental solution* to the heat equation. General solutions to the heat equation with a specified initial condition are found by convolution with the fundamental solution.

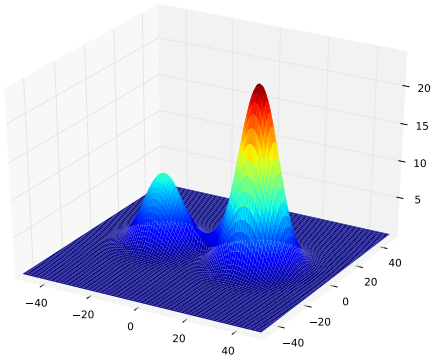
Intuitively, we hope that increasing the scale parameter t will blur away any unimportant details of the image. Since heat diffuses, one might expect regions where the heat exceeds a particular threshold would either grow or shrink. When two regions meet, it seems plausible the regions would merge and continue to grow or shrink as a single component. This intuition, however, is false [26]. During the course of the heat equation, regions (of a particular value or range of values) can merge and split. Furthermore, new regions can be created, as demonstrated in the following example:

Example 30 (Mountain and Bridge [59])**.** Consider a mountain range with exactly two peaks at different heights. Suppose we connect these peaks by a thin ramp-like bridge with steep sides, as shown in Figures 4.4a and 4.4c and consider the height function $h(x)$ defined by the height of the mountain if there is no bridge above x and the height of the bridge otherwise. There is one index-two critical point, the location of the higher peak, as the lower mountain peak is not a maximum of h. If we apply the heat equation to this function, there will soon be two local maxima as the bridge quickly collapses; see Figures 4.4b and 4.4d.

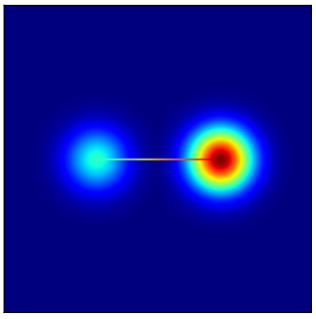
Example 31 (Epanechinikov Kernel)**.** For $x \in \mathbb{R}^n$, the Epanechinikov kernel with



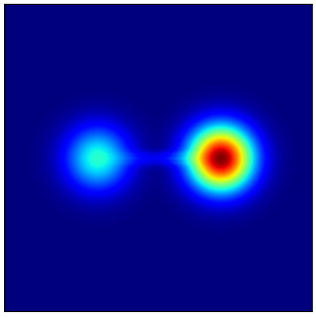
(a) Initial Surface



(b) Convolved Surface



(c) Initial Function



(d) Convolved Function

FIGURE 4.4: On the top left, we see the surface described in Example 30: two mountain peaks connected by a narrow bridge. Below the surface, the height function is given as a heat map. On the right, we show both the surface and the 2D function after convolution with a Gaussian kernel. Notice that the bridge has collapsed, resulting in two local maxima.

width σ is defined by:

$$\mathcal{E}(x-\mu,\sigma) \quad = \quad \begin{cases} \frac{3}{4\sigma}\left(1-\frac{\|x-\mu\|^2}{\sigma^2}\right), & \|x-\mu\| \leq \sigma \\ 0, & \text{otherwise.} \end{cases} \tag{4.20}$$

In $\mathbb{R}^1$, the sum of two Epanechinikov kernels can result in three local maxima. The third local maxima occurs half way between the two kernel centers, as shown in Figure 4.5. In Section 4.1, we proved that the Gaussian kernel does not exhibit this behavior in $\mathbb{R}^1$. The desire to have sums of kernels, or more generally, convolutions of functions, to behave in a predictable manner led to the axiomatic interpretation

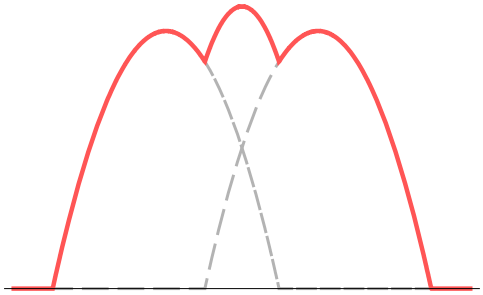


FIGURE 4.5: The sum of the two gray Epanechnikov kernels $\mathcal{E}(x+\mu,\sigma)$ and $\mathcal{E}(x-\mu,\sigma)$ is shown in red. Although the Epanechnikov kernels are unimodal, the sum has three local maxima.

of scale space, which we address next.

4.2.2 An Axiomatic View of Scale Space

The creation of additional critical values, especially of maxima, seems to go against our intuition of scale space: scale space should *blur* or *simplify* the initial function, not *enhance* or *create* additional features. The phenomenon described in Example 30 is not an anomaly. In fact, Damon [26] proves that the creation of *singularities*, points where the gradient vector field vanishes, is a generic event in the heat equation.

Under the axiomatic interpretation of scale space, a set of axioms is defined that regulate the behavior of the kernel used in (4.16). The idea behind this is to align the mathematical definition of scale space with an intuitive description of blurring functions. Several authors have independently developed their own set of axioms of scale space [7, 53, 54, 60, 86]. The Gaussian kernel is the unique kernel that satisfies all of the axioms of scale space. Alternative scale space theory has been developed by weakening some of these axioms; for example, [30] considers a one-parameter family of kernels. In this subsection, we list a non-exhaustive set of scale space axioms. Then, we preset a few key results.

Let $\{K_t\}$ be a one-parameter family of kernels $K_t\colon \mathbb{R}^n \to \mathbb{R}$ and let H be the set of all bounded functions $h\colon \mathbb{R}^n \to \mathbb{R}$. We define the convolution of a function

$h_0 \in H$ by the kernel K_t as in (4.16); denoting it as $[K_t * h_0](x)$. Then, the following are some of the potential axioms of scale space:

Axiom 32 (Bandwidth)**.** *The parameter t is the square inverse of a* bandwidth parameter *for K_t; that is, there exists a function $\kappa\colon \mathbb{R}^n \to \mathbb{R}$, called the* inner kernel*, such that for $y = \frac{1}{\sqrt{t}}$, we have $K_t(x) = y^n \kappa(xy)$ for all $t > 0$.*

Axiom 33 (Causality)**.** *All nondegenerate local minima of the height function of the surface $(K_t * h, t)$ occur in the plane $t = 0$.*

Axiom 34 (Continuity)**.** *The convolved function $K_t * h_0$ is continuous for all $t > 0$.*

Axiom 35 (Flat Tails)**.** *As $\|x\|$ approaches infinity, the kernel becomes flat:*

$$\lim_{\|x\| \to \infty} K_t(x) \quad = \quad 0.$$

Axiom 36 (Isometry Invariance)**.** *If $T\colon H \to H$ is an orthogonal transformation, such as a translation or a rotation of the domain of a function $h \in H$, then T commutes with K_t:*

$$T\left([K_t * h_0]\right) = [K_t * T(h_0)].$$

Axiom 37 (Locality)**.** *The filter recovers the whole image at sufficiently small scales:*

$$\lim_{t \to 0} [K_t * h_0](x) = h_0(x).$$

Axiom 38 (Non-Negativity)**.** *Given $h_0 \in H$ and a kernel $K_t(x)$, if h_0 is non-negative, then the convolved function $K_t * h_0$ is also non-negative.*

Axiom 39 (Normalization)**.** *For all $t > 0$, the kernel $K_t(\cdot)$ integrates to unity:*

$$\int_{x \in \mathbb{R}^n} K_t(x)\, dx \quad = \quad 1$$

Axiom 40 (Semi-Group Structure)**.** *The convolution of two kernels with variance t_a and t_b is a kernel with variance $t_a + t_b$:*

$$[K_{t_a} * K_{t_b}] \quad = \quad K_{t_a+t_b}$$

Axiom 41 (Separability)**.** *There exists a family of one-dimensional kernels $k_i \colon \mathbb{R} \to \mathbb{R}$ such that we can write: $K_t(x) = \prod_{i=0}^{n-1} k_i(x_i)$, where $x = (x_0, \ldots, x_{n-1})$ in some basis for $\mathbb{R}^n$. Typically, the basis is taken to be the Cartesian basis.*

If $n = 1$, then the following axioms may apply:

Axiom 42 (Inner Kernel)**.** *There exists a positive integer p such that $\kappa^{(2p)}(0) \neq 0$, where κ is the inner kernel defined in Axiom 32.*

The last axiom that we will mention is implied by the more general Axiom 36, but it is worth stating as its own axiom:

Axiom 43 (Symmetry Invariance)**.** *The kernel K_t is symmetrical in x, i.e.,*

$$K_t(-x) \quad = \quad K_t(x).$$

Implications of the Axioms. Assuming different subsets of the axioms above, different authors have developed theories of scale space. All of these theories, however, have one thing in common: the Gaussian kernel is always in the family of allowable kernels. In fact, sometimes it is the only allowable kernel, as we will see in the next two theorems:

Theorem 44 (Gaussian Kernel [30])**.** *The only differentiable kernel $K_t \colon \mathbb{R}^n \to \mathbb{R}$ which is both separable and isotropic must be Gaussian.*

Proof. If K_t is separable, then we can write $K_t(x, y) = k_1(x)k_2(y)$. If the kernel is isotropic, then it is invariant with respect to rotations. Let R be a rotation matrix in $\mathbb{R}^n$, then rotational invariance means $K_t(R(x, y)) = K_t(x, y)$. Converting

to radial coordinates and taking the derivative with respect to the radius results in: $(y\partial_x - x\partial_y)K_t(x,y) = 0$, which is true if and only if the following holds:

$$yk_2(y)k_1'(x) \;=\; xk_1(x)k_2'(y),$$

which implies

$$\frac{xk_1(x)}{k_1'(x)} \;=\; \frac{yk_2(y)}{k_2'(y)} \;=\; c, \tag{4.21}$$

where $c \in \mathbb{R}$ is constant for all choices of x and y. Thus, we can use Lemma 28 to obtain the isotropic Gaussian kernel $k_i(x) = k_i(0)e^{\frac{x^2}{2c}}$. □

Theorem 45 (Monotonically Decreasing Maxima [7])**.** *If $\{g_t(\cdot)\}_{t>0}$ is a family of kernels from $\mathbb{R}$ to $\mathbb{R}$ such that $g_t(x)$ decreases with respect to t and satisfies Axioms 32, 39, 42 and 43, then the maxima of $[g_t * f]$ decrease monotonically as t increases if and only if g_t is the Gaussian kernel $g_t(x) = \frac{1}{\sqrt{2\pi t}}e^{\frac{-1}{2t}x^2}$.*

In Theorem 49, we prove that Theorem 45 holds for f a weighted sum of Gaussian kernels. For the full proof, we refer the reader to Section II.B. of [7]. In a similar fashion, the Gaussian kernel satisfies all of the axioms listed above. We state the theorem here, but note that proofs are scattered throughout the literature [7, 30, 53, 54, 60, 83, 86].

Theorem 46 (Gaussian Kernel is Sufficient)**.** *The Gaussian kernel satisfies Axiom 32 through Axiom 43.*

We may be interested in other kernels that do not satisfy all of the axioms. For example, the Epanechnikov kernel violates the causality axiom. In Subsection 5.4.2, we briefly discuss how our results extend to other kernels. For now, however, we focus on the Gaussian kernel.

4.2.3 No Ghost Modes in $\mathbb{R}^1$

We now prove a strengthening of the results in Section 4.1: a Gaussian Mixture in $\mathbb{R}^1$ can have no more modes than components. This statement has been proven in several papers; see [7, 54, 86]. Following [72], we use a scale space argument to prove a slightly generalized result. First, we need a lemma by Schoenberg:

Lemma 47 (Diminishing Variation [71])**.** *If $h_0(t)$ is measurable and continuous, then the number of sign changes of $\mathcal{T}_t h_0(x)$ decreases as t increases.*

We refer the reader to Section 2 of [71] for the proof of the this lemma, but we will use its result in the proof of the next lemma. Consider the following Gaussian mixture: $G_1(x,t) = \sum_{i=0}^{n} \mathcal{N}_1(x-\mu_i, t)$, then we have:

Lemma 48 (Maxima of Mixture Partials [72])**.** *For each $m \geq 0$, the number of maxima of $\frac{\partial^m G_1}{\partial x^m}$ is a right continuous decreasing function of t.*

Proof. Let $v_{m+1}(t)$ denote the number of variations of sign of

$$G_1^{(m+1)}(\cdot, t) \;:=\; \frac{\partial^m G_1}{\partial x^m}(\cdot, t).$$

We must show that v_{m+1} is decreasing and right continuous for all $m \geq 0$. First, we prove that v_{m+1} is a decreasing function of t. Let $t_2 > t_1 > 0$. Then, $G_1(\cdot, t_2)$ is the convolution of $G_1(\cdot, t_1)$ with the kernel $\mathcal{N}_1(x, t_2 - t_1)$ by the semi-group structure of convolution; see Axiom 40. We note here that the m+1-st derivative of the convolved kernel is the convolution of the m+1-st derivative of the Gaussian kernel. The number of variations of sign v_{m+1} is a decreasing function of t follows from Lemma 47, since $G_1^{(m+1)}(\cdot, t_1)$ is continuous and bounded (and hence measurable).

To finish this proof, we must show that v_{m+1} is right continuous. Let $\{a_i\}$ and $\{b_i\}$ be sequences of interleaving values

$$a_1 < b_1 < \ldots < a_r < b_r$$

such that $G_1(a_i, t) > 0$ and $G_1(b_i, t) < 0$ for all i. Since $G_1(., t)$ is continuous for all $t > 0$ by Axiom 34, there exists an ε_0 small enough such that $G_1^{(m+1)}(a_i, t + \varepsilon) > 0$ $G_1^{(m+1)}(b_i, t + \varepsilon) < 0$ for all $\varepsilon < \varepsilon_0$ and for all i. Hence, $v_{m+1}(t + \varepsilon) \geq v_{m+1}(t)$. Recalling that v_{m+1} is decreasing, right continuity follows since $v_{m+1}(t + \varepsilon) = v_{m+1}(t)$ for all $\varepsilon < \varepsilon_0$. ◻

Consequently, the number of maxima of the m^{th} derivative of a convolved one-dimensional function $\mathcal{T}_t h_0(x)$ at any scale $t > 0$ is at most the number of maxima of $\frac{\partial^m \mathcal{T}_0 h_0}{\partial x^m}(x)$. In particular, the number of maxima of a Gaussian mixture in $\mathbb{R}^1$ is bounded by the number of components, as we see in the following theorem:

Theorem 49 (Ghost Buster [15])**.** *The weighted sum of Gaussian kernels*

$$G_1^{\omega}(x, t) \quad = \quad \sum_{i=0}^{n} \omega_i \mathcal{N}_1(x - \mu_i, t)$$

has at most $n + 1$ *modes.*

Proof. If $n = 0$, then $G_1^{\omega}(x, t)$ is the Gaussian kernel centered at μ_0 with variance t and weight ω_0; thus, the theorem holds for $n = 0$. Now, assume that this theorem holds for the weighted sum of $n - 1$ Gaussian kernels. Apply Gaussian deblurring of variance t. Then, we obtain a new mixture:

$$\mathcal{T}_t^{-1} G_1^{\omega}(x, t) \quad = \quad \sum_{i=0}^{n} \omega_i \mathcal{N}_1(x - \mu_i, 0).$$

By Lemma 29, the sum on the right hand side is the weighted sum of Dirac delta functions:

$$\mathcal{T}_t^{-1} G_1^{\omega}(x, t) \quad = \quad \sum_{i=0}^{n} \omega_i \delta_{\mu}(x),$$

which has exactly $n + 1$ modes. Applying Lemma 48 gives the desired result. ◻

We contrast Theorem 49 with the behavior of modes in Gaussian mixtures in higher

dimensions: Gaussian mixtures in $\mathbb{R}^n$ for $n > 1$ *can* have more modes than mixture components; see Theorem 57.

4.2.4 Iterative Algorithm

We conclude this chapter with an algorithm to compute an approximation of the scale space of a function defined over a bounded two dimensional domain. Computing a solution to a PDE is not a simple task. There are two issues that can arise when using the continuous formulation of the heat equation described in Section 4.2.1:

1. We need to solve the PDE presented in Equations (4.13) and (4.14).
2. The partial derivatives $\frac{\partial^2 u_t}{\partial b_i^2}(x)$ and $\frac{\partial u_t}{\partial t}(x)$ are not well defined over a discrete domain.

In order to resolve these issues, we defer to numerical methods to estimate this solution, which require spatial and temporal discretization [11]. The goal is to obtain a discretization of the heat equation solution $u(x,t)$. The method for approximating a solution to the PDE that we describe in this section is the *finite difference method*.

Spatial Discretization. In the following computations, we use a regular grid decomposition of $\mathbb{M} = [0,1]^2$, writing $x_i = (i-1)h$ and $y_j = (j-1)h$, with $h = 1/(m-1)$ for some fixed integer $m \geq 2$. There are m^2 vertices in this decomposition. The vertices are connected by vertical and horizontal edges. The neighborhood $\mathrm{Nhd}(i,j)$ of the point (x_i, y_j) is defined to be the set of points directly connected to by an edge to (x_i, y_j). For $1 < i < m$ and $1 < j < m$, the neighborhood is

$$\mathrm{Nhd}(i,j) = \{(x_i, y_{j\pm1}), (x_{i\pm1}, y_j)\}. \tag{4.22}$$

More generally, it is possible to define a neighborhood of a vertex, $\mathrm{Nhd}_G(v)$, to be the set of vertices connected to v by an edge in the graph G. Other approaches to spatial discretization may assign a weight for each $w \in \mathrm{Nhd}_G(v)$.

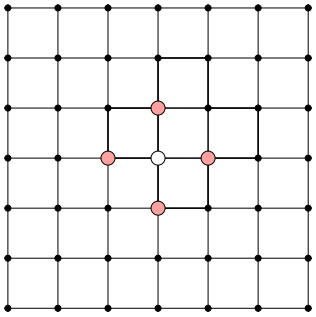


FIGURE 4.6: We represent $[0,1]^2$ by 49 vertices connected by vertical and horizontal edges. In the center, the white dot represents the vertex $x = (i, j)$. The vertices highlighted in pink are those whose values contribute to the estimate of the heat equation at x.

Mathematical Description. We focus on the steady-state heat equation over $\mathbb{R}^2$:

$$\frac{\partial^2 u}{\partial e_1^2}(x,t) + \frac{\partial^2 u}{\partial e_2^2}(x,t) = 0. \tag{4.23}$$

In this equation, we are using $\{e_1, e_2\}$ as the standard basis vectors for $\mathbb{R}^2$. At each mesh point $p_{ij} = (x_i, y_j)$, we employ the Taylor polynomial in the variable b_1 to obtain an approximation of the second derivative with respect to b_1:

$$\frac{\partial^2}{\partial e_1^2}u(p_{ij},t) = \frac{u(p_{i+1,j},t) - 2u(p_{ij},t) + u(p_{i-1,j},t)}{h^2}, \tag{4.24}$$

where h is the spatial step size. Similarly, we also have an approximation for the second derivative with respect to b_2:

$$\frac{\partial^2}{\partial e_2^2}u(p_{ij},t) = \frac{u(p_{i,j+1},t) - 2u(p_{ij},t) + u((p_{i,j-1}),t)}{h^2}. \tag{4.25}$$

If we plug (4.24) and (4.25) into (4.23), then we obtain the following equation:

$$4u(x_{i,j},t) - u(x_{i+1,j},t) - u(x_{i-1,j},t) - u(x_{i,j+1},t) - u(x_{i,j-1},t) = 0 \tag{4.26}$$

Solving this equation for $u(x_{ij}, t)$, an approximation of the heat equation can be made by looking at local neighborhoods for each vertex $x_{i,j}$. Figure 4.6 highlights

the four neighbors of the mesh needed to compute an approximation to the heat equation. Consider the equations of the form presented in (4.26), of which we have one for each vertex $x_{i,j}$ in the interior of the $m \times m$ grid. We generalize this equation so that it may hold for vertices on the boundary of the mesh:

$$4u(x_{ij}, t) - \sum_{y \in \mathrm{Nhd}(x_{ij})} u(y, t) = 0,$$

where we assume that $\mathrm{Nhd}(x_{i,j})$ contains four vertices. The elements in $\mathrm{Nhd}(x_{i,j})$ for a boundary vertex $x_{i,j}$ dictate the *boundary conditions*. We relabel the grid in column-major order to allow the use of one index instead of two, using $v_{\{(j-1)m+i\}}$ to represent $u(x_{i,j}, t)$. Then, we may express the m^2 linear equations in matrix-vector form by $Av = 0$. Solving the discrete heat equation finds a solution v to the linear system of equations $Av = 0$. This solution is dependent on the choice of boundary conditions.

Algorithm. The matrix A is equal to L_m, the *Poisson matrix* of order m. We note here that this matrix is $m^2 \times m^2$. We can write $L_m = D - N$, where D is the diagonal matrix $4I$ and N is a matrix with 0's on the diagonal and with only 1 as the non-zero entries of the matrix. Sometimes we refer to D as the *valency matrix*, since it expresses the degree of each mesh point. The matrix N is symmetric and we call it the *neighborhood matrix* since the non-zero entries in row i correspond to the neighbors of the mesh point v_i [8]. That is, $N(i, j) = 1$ iff v_i and v_j are adjacent and $N(i, j) = 0$ otherwise.

The iterative algorithm for finding a solution to the heat equation can be defined by these matrices. We want a solution that satisfies $Av = 0$, which means $(D-N)v = 0$. We can re-write this as $Dv = Nv$. The iterative algorithm can immediately be seen:

$$v_{\mathrm{new}} \quad = \quad (D^{-1}N)v.$$

In the original formulation, this translates to:

$$u_{t+1}(x) \;=\; \frac{1}{4} \sum_{y \in \mathrm{Nhd}(x)} u_t(y).$$

where the neighborhood $\mathrm{Nhd}(x)$ is the neighborhood of x defined by Equation (4.22). This iterative method is known as *Jacobi iteration* [11]. Notice that this technique generalizes to non-regular meshes by computing u_t at a vertex by averaging the values of u_{t-1} over the vertices in $\mathrm{Nhd}_G(v)$.

5

Ghosts in Gaussian Mixtures

While the existence of spurious modes in scale space was observed as early as 1986 in [7], it has been an open question whether the sum of finitely many isotropic Gaussian kernels in $n \geq 2$ dimensions can have more modes than kernels, until in 2003 Carreira-Perpiñán and Williams exhibited $n+1$ isotropic Gaussian kernels in $\mathbb{R}^n$ with $n+2$ modes [14]. We give a detailed analysis of this example, showing that it has exponentially many critical points and that the resilience of the extra mode grows like $\sqrt{n}$. In addition, we exhibit finite configurations of isotropic Gaussian kernels with super-linearly many modes. This chapter is largely based on [31]. In addition, this chapter includes numerical evidence that helps to provide an intuition for conjectures we have yet to prove about the persistence of the ghost feature and extensions to other kernels.

5.1 Simplex Design

In this section, we design a sum of Gaussian kernels in $\mathbb{R}^{n+1}$ that has the symmetry group of the regular n-simplex. We begin with a geometric study of the simplex, whose shape properties will play a central role in our design.

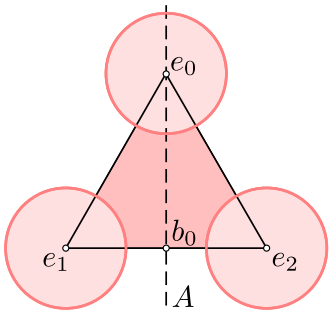


FIGURE 5.1: A 2-design, which is the sum of three unit Gaussian kernels placed at the vertices of an equilateral triangle. The axis A connects e_0 to b_0, the barycenter of $\{e_1, e_2\}$.

Standard Simplex. A convenient model is the *standard n-simplex*, defined as the convex hull of the $n+1$ unit coordinate vectors in $\mathbb{R}^{n+1}$: $\Delta^n = \text{conv}\,\{e_0, e_1, \ldots, e_n\}$. Each subset of $k+1$ vectors defines a *k-face* of Δ^n, which is itself a standard k-simplex. The *barycenter* of Δ^n is the point whose $n+1$ coordinates are all equal to $\frac{1}{n+1}$. Let $0 \le k \le \ell$ with $k+\ell = n-1$ and consider a k-face and the unique disjoint ℓ-face. Their barycenters use complementary subsets of the $n+1$ coordinates, which makes it easy to compute the distance between them as

$$D_{k,\ell} \;=\; \sqrt{\frac{1}{k+1} + \frac{1}{\ell+1}}. \tag{5.1}$$

For example, the height of the n-simplex, which is defined as the distance between a vertex to the opposite $(n-1)$-face, is $D_{0,n-1} = \sqrt{(n+1)/n}$. Similarly, we can compute the radius of the circumsphere:

$$R_n \;=\; \sqrt{\frac{n}{n+1}}. \tag{5.2}$$

We note that R_n is the distance between the barycenter of Δ^n and a vertex of Δ^n.

Standard Design. We turn Δ^n into a function by placing a unit Gaussian kernel at every vertex:

$$G = g_{e_0} + g_{e_1} + \ldots + g_{e_n}.$$

which we call an *n-design*; see Figure 5.1. We are interested in changing the widths of the $(n+1)$-dimensional kernels uniformly. Equivalently, we scale the n-simplex by moving the centers of the unit Gaussian kernels closer to or further from each other, without changing their widths and heights. To do this, we introduce the *scaled n-design*. Writing g_i for the unit Gaussian kernel with center se_i, we obtain $G_s = g_0 + g_1 + \ldots + g_n$ Here, we call s the *scale factor*, and we write $s\Delta^n$ for the *scaled n-simplex* whose vertices are the se_i. We are interested in the evolution of the critical points in the one-parameter family of scaled n-design $G_s : \mathbb{R}^{n+1} \to \mathbb{R}$, as s goes from zero to infinity. Note that this corresponds to deconvolution, going from $t = \infty$ back to $t = 0$.

The symmetry group of the scaled n-design is that of the n-simplex with the additional reflection across the n-dimensional plane spanned by the n-simplex. It is therefore isomorphic to $\Sigma_{n+1} \oplus \mathbb{S}^0$, where Σ_{n+1} is the symmetric group on $n+1$ elements and $\mathbb{S}^0$ consists of two points representing the reflection through the n-plane. To argue about the symmetries, we use lines that connect barycenters of complementary faces of the n-simplex. We call these lines *axes*. Suppose for example that $0 \leq k \leq \ell$ are integers with $k + \ell = n - 1$, and that A is the axis that connects the barycenter of the k-face spanned by se_0 to se_k with the barycenter of the ℓ-face spanned by se_{k+1} to se_n. Except for the barycenter of Δ^n, every point x of A has two distinct barycentric coordinates, one with multiplicity $k+1$ and the other with multiplicity $\ell + 1$, where the barycentric coordinates are defined as follows:.

Definition 50 (Barycentric Coordinates)**.** The *barycentric coordinates* of a point x in the affine space defined by the vertices of $s\Delta^n$ are the values $x_0, x_1, \ldots, x_n$ such that

$$x \;=\; x_0(se_0) + x_1(se_1) + \cdots + x_n(se_n).$$

We write the coordinates as a vector: $(x_0, x_1, \ldots, x_n)_B$.

Applying the symmetries of the n-simplex, we get $(n+1)!$ permutations of the $n+1$ barycentric coordinates as the *orbit* of x; however, some of these permutations give the same point. In this particular case, the orbit of x consists of $\binom{n+1}{k+1}$ points. Recall that the section defined by the axis is the restriction of the function to the line: $G|_A$. The restrictions of $g_0, g_1, \ldots, g_k$ to A are all identical, namely a weighted one-dimensional unit Gaussian kernel whose center is the barycenter of the k-face. Similarly, the restrictions of $g_{k+1}, \ldots, g_{n-1}, g_n$ are all identical, and we can write the one-section as the sum of two kernels:

$$G|_A \;=\; (k+1)\cdot g_0|_A + (\ell+1)\cdot g_n|_A, \tag{5.3}$$

which are one-dimensional weighted unit Gaussian kernels, with weights $(k+1)g(R_k)$ and $(\ell+1)g(R_\ell)$, and distance $D_{k,\ell}$ between their centers.

5.2 Analysis

We begin this section by proving that all critical points lie on the axes of the n-simplex. Thereafter, we analyze each axis, characterizing for which scales we see one, two, or three critical points. To decide which of the one-dimensional maxima are modes (that is, $n+1$-dimensional maxima), we analyze the n-sections orthogonal to the axes. As it turns out, all modes lie on axes that pass through vertices of the n-simplex. Most interesting is the critical point at the barycenter, which changes from unique mode during an initial interval of scales, to $(n+2)$-nd mode during a non-empty intermediate interval, to a saddle of index one during a final interval. We call the length of the intermediate interval the resilience of the extra mode and show that it grows like the square root of the dimension. Finally, we construct sums of isotropic Gaussian kernels with a super-linear number of modes.

5.2.1 Lines of Critical Points

In this subsection, we prove that all critical points of a scaled n-design lie on the axes of the scaled n-simplex. We begin by introducing coordinates that are more natural for the n-design, and we show how they relate to the barycentric coordinates.

Distance Coordinates. Write $v_i = se_i$, for $0 \leq i \leq n$, and let x be a point of the corresponding scaled n-simplex $s\Delta^n$. Setting $r_i = \|x - v_i\|$, we note that x is uniquely defined by the vector of $n+1$ distances since x lies on the hyperplane spanned by $\{v_i\}$. We express this by writing $x = (r_0, r_1, \ldots, r_n)_D$, and by calling the r_i the *distance coordinates* of x. Recalling from Definition 50 that $(x_0, x_1, \ldots, x_n)_B$ is the representation of the same point in barycentric coordinates, we compute the barycentric from the distance coordinates via the Coordinate Transformation below.

Lemma 51 (Coordinate Transformation)**.** *For $0 \leq i \leq n$, the i^{th} barycentric coordinate is given by:*

$$x_i = \frac{1}{n+1} + \frac{1}{2(n+1)s^2}\left(\sum_{j=0}^{n} r_j^2 - (n+1)r_i^2\right). \tag{5.4}$$

Proof. Let $(r_0, r_1, \ldots, r_n)_D$ be the distance coordinates of a point x in the scaled n-simplex $s\Delta^n$. Let $i \neq j$ and consider the edge connecting v_i with v_j, recalling that v_i and v_j are two vertices of $s\Delta^n$. The length of the edge is $s\sqrt{2}$. Let x be a point of $s\Delta^n$ and set x_{ij} to be the distance between v_j and the orthogonal projection y_{ij} of x onto the edge, normalized by dividing with $s\sqrt{2}$; see Figure 5.2. We first show that

$$x_{ij} = \frac{1}{2} + \frac{1}{4s^2}(r_j^2 - r_i^2). \tag{5.5}$$

Indeed, if $x = (1-t)v_j + tv_i$, then we have $r_j = s\sqrt{2}t$ and $r_i = s\sqrt{2}(1-t)$. Furthermore, $x_{ij} = t$, which agrees with the equation we get by plugging the values of r_i

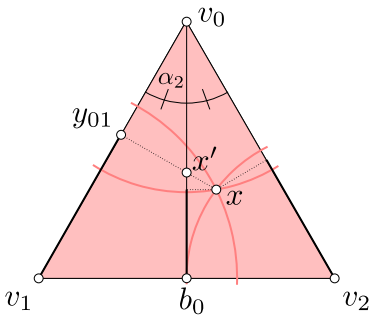


FIGURE 5.2: The radii of the three circles are the distance coordinates of the point x. The orthogonal projections onto the vertical axis and the left and right edges of the triangle give x_0, x_{01}, and x_{02}. Notice that $\Delta v_0 b_0 v_1$ and $\Delta v_0 y x'$ are similar triangles.

and r_j into (5.5). Realizing that $r_j^2 - r_i^2$ is constant along hyperplanes orthogonal to the edge, we satisfy (5.5) for all points of the n-simplex.

For the next step, let b_i be the barycenter of the $(n-1)$-face complementary to v_i. Set α_n to the angle between the edges that connect v_i to v_j and to b_i. Because $s\Delta^n$ is regular, this angle does not depend on the choice of i and j. Suppose x lies on the latter edge, which connects v_i and b_i. Then $x = (1 - x_i)b_i + x_i v_i$ and we have two expressions for $\cos \alpha_n$. Setting these two expression equal we arrive at

$$\frac{D_{0,n-1}}{\sqrt{2}} \quad = \quad \frac{\sqrt{2}(1 - x_{ij})}{D_{0,n-1}(1 - x_i)},$$

for every $0 \leq j \leq n$ and $j \neq i$. Adding the n equations gives

$$nD_{0,n-1}^2(1 - x_i) \quad = \quad 2n - \sum_{i \neq j} x_{ij}. \tag{5.6}$$

Similar to before, we notice that the two sides of the equation are constant along hyperplanes orthogonal to the axis defined by v_i and b_i. Hence, (5.6) holds for all points x of the scaled n-simplex. It remains to plug (5.1) and (5.5) into (5.6), which gives

$$(n+1)(1 - x_i) \quad = \quad n + \frac{1}{2s^2}\left(nr_i^2 - \sum_{i \neq j} r_j^2\right).$$

The equation simplifies to the claimed equation. ◻

Non-Zero Gradients. Recall that $G_s : \mathbb{R}^{n+1} \to \mathbb{R}$ is the scaled n-design formed by taking the sum of the $n+1$ unit Gaussian kernels whose centers are the vertices of $s\Delta^n$. We use the Coordinate Transformation to show that G_s has no critical points away from the axes of the scaled n-simplex:

Lemma 52 (Axes Lemma)**.** *Every critical point of G_s lies on an axis of the scaled n-simplex $s\Delta^n$.*

Proof. By the previous lemma, a point x belongs to an axis of $s\Delta^n$ iff it has at most two distinct barycentric coordinates. We will show that if x has three distinct barycentric coordinates then the gradient of G_s at x is non-zero.

$$\nabla G_s(x) \;=\; -2\pi G_s(x) \cdot x + 2\pi \sum_{i=0}^{n} g_i(x) \cdot v_i.$$

Setting the gradient to zero, we solve for x

$$x \;=\; \sum_{i=0}^{n} \frac{g_i(x)}{G_s(x)} \cdot v_i. \tag{5.7}$$

We will show that Equation (5.7) can hold only if x has at most two distinct barycentric coordinates. To this end, we write x in distance coordinates: $x = (r_0, r_1, \ldots, r_n)_D$. Similar to the barycentric coordinates, x lies on an axis of $s\Delta^n$ iff there are at most two distinct distance coordinates. Transforming x into barycentric coordinates, we have $x = (x_0, x_1, \ldots, x_n)_B$, in which

$$x_i \;=\; \frac{1}{n+1} + \frac{1}{2(n+1)s^2}\left(\sum_{j=0}^{n} r_j^2 - (n+1)r_i^2\right),$$

for $0 \le i \le n$. Assume now that x does not lie on any of the axes. If follows that x has three distinct distance coordinates: $r_k < r_\ell < r_m$. Subtracting the m-th

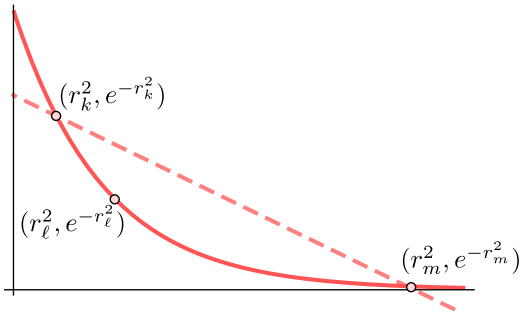


FIGURE 5.3: The solid curve is the graph of $f(x) = e^{-x}$. Since $f(x)$ is strictly convex and $r_k < r_\ell < r_m$, we know that the dashed secant line through $(r_k^2, e^{-r_k^2})$ and $(r_m^2, e^{-r_m^2})$ must be above the point $(r_\ell^2, e^{-r_\ell^2})$.

barycentric coordinate from the other two , we obtain

$$x_k - x_m = \frac{1}{2s^2}(r_m^2 - r_k^2), \tag{5.8}$$

$$x_\ell - x_m = \frac{1}{2s^2}(r_m^2 - r_\ell^2). \tag{5.9}$$

Assuming a zero gradient, the barycentric coordinates of x have the form given in (5.7). Hence, $x_k - x_m$ and $x_\ell - x_m$ are equal to

$$\frac{g_k(x) - g_m(x)}{G_s(x)} = \frac{e^{-\pi r_k^2} - e^{-\pi r_m^2}}{G_s(x)}, \tag{5.10}$$

$$\frac{g_\ell(x) - g_m(x)}{G_s(x)} = \frac{e^{-\pi r_\ell^2} - e^{-\pi r_m^2}}{G_s(x)}, \tag{5.11}$$

respectively. Since the right hand sides of (5.8) and (5.10) are equal, as well as the right hand sides of (5.9) and (5.11), we have

$$\frac{r_m^2 - r_k^2}{e^{-r_m^2} - e^{-r_k^2}} = \frac{r_m^2 - r_\ell^2}{e^{-r_m^2} - e^{-r_\ell^2}}.$$

But this is impossible because $r_m^2 - r_k^2 > r_m^2 - r_\ell^2$, by assumption, and because the function $f(t) = e^{-t}$ is strictly convex; see Figure 5.3. □

5.2.2 Transitions

Given by (5.3), the restriction of G_s to an axis of $s\Delta^n$ is a sum of two weighted Gaussian kernels. As we saw in Section 4.1, this sum has two local maxima for a range of scale factors. In this subsection, we analyze this range.

Computing Transitions. Recall that the n-design consists of $n+1$ unit Gaussian kernels placed at the vertices of the standard n-simplex. Consider the 1-section defined by the line that connects the barycenter of a k-face with the barycenter of the complementary ℓ-face, with $k+\ell = n-1$, and vary the construction by scaling the design with $s \geq 0$. We call a value a *transition* if the number of critical points of the 1-section changes as s passes the value. It is easy to compute the transition for $k = \ell = \frac{n-1}{2}$ because the corresponding 1-section is balanced for all scale factors s. The distance between the two centers is $sD_{\ell,\ell} = 2s/\sqrt{n+1}$ and we find the transition by setting the distance equal to $2\sigma_0$, which gives s equal to

$$U_n \;=\; \sqrt{\frac{n+1}{2\pi}}. \tag{5.12}$$

Consider next the case $k < \ell$. Equation (5.3) gives the weights of the two kernels in the decomposition of the 1-section as $(k+1)g(sR_k)$ and $(\ell+1)g(sR_\ell)$. Using (5.2) and taking the ratio of the weights, the *weight function* is computed by the following function:

$$\omega_{k,\ell}(s) \;=\; \frac{\ell+1}{k+1} \cdot e^{-\pi s^2\left(\frac{1}{k+1}-\frac{1}{\ell+1}\right)}. \tag{5.13}$$

We compare this with the two *transition functions*, which we get by setting $\mu = \frac{s}{2}D_{k,\ell}$ and plugging the two solutions of $x^2 = \mu^2 - \sigma_0^2$ into $r(x)+1$, using $r(x)$ as given

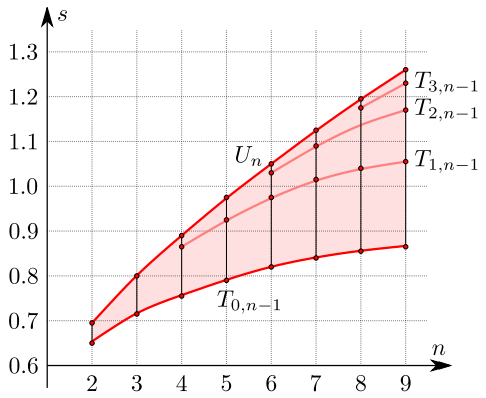


FIGURE 5.4: The vertical interval bounded from below by $T_{0,n-1}$ and above by U_n contains the scale factors for which an extra mode appears.

by (4.8). The result is two functions of s:

$$\tau_{k,\ell}(s) = \frac{\mu - \sqrt{\mu^2 - \sigma_0^2}}{\mu + \sqrt{\mu^2 - \sigma_0^2}} \cdot e^{4\pi\mu\sqrt{\mu^2-\sigma_0^2}}, \tag{5.14}$$

$$\upsilon_{k,\ell}(s) = \frac{\mu + \sqrt{\mu^2 - \sigma_0^2}}{\mu - \sqrt{\mu^2 - \sigma_0^2}} \cdot e^{-4\pi\mu\sqrt{\mu^2-\sigma_0^2}}. \tag{5.15}$$

Note that $\upsilon_{k,\ell}(s) = 1/\tau_{k,\ell}(s)$. We find the first transition $T_{k,\ell}$ by solving $\omega_{k,\ell}(s) = \tau_{k,\ell}(s)$, and the second transition U_n by solving $\omega_{k,\ell}(s) = \upsilon_{k,\ell}(s)$; see Tables B.1 and B.2 in Appendix B for sample values of these transition scale factors. While we have no analytic expression for $T_{k,\ell}$, we will derive one for an upper bound in Section 5.2.5. In fact, we prove next that both transitions are well defined, also showing that the second transition, which depends on n but not on k and ℓ, is given by (5.12).

Curve Analysis. We give a detailed analysis of the intersections between the weight function, $\omega_{k,\ell}$, and the two transition functions $\tau_{k,\ell}$ and $\upsilon_{k,\ell}$. We recall that the two transitions, $T_{k,\ell}$ and U_n, are the solutions to $\omega_{k,\ell}(s) = \tau_{k,\ell}(s)$ and to $\omega_{k,\ell}(s) = \upsilon_{k,\ell}(s)$ respectively. We discuss the graphs of the three functions to convey a feeling for how they intersect. To begin, we note that $s_0 = 2\sigma_0/D_{k,\ell}$ is the smallest scale factor for which the transition functions are defined. Writing $\mu = \frac{s}{2}D_{k,\ell}$, we have $\mu^2 - \sigma_0^2 = 0$

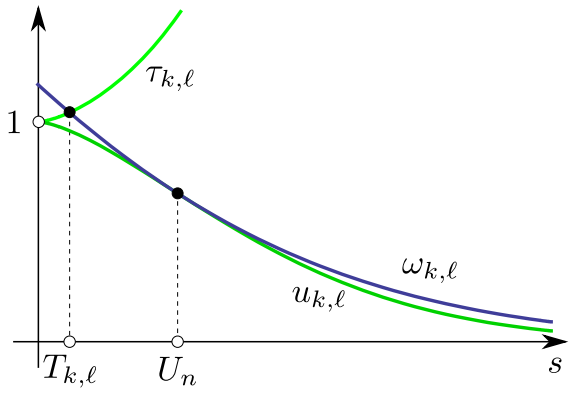


FIGURE 5.5: The graphs of the weight function and the two transition functions. The intersections define the transitions.

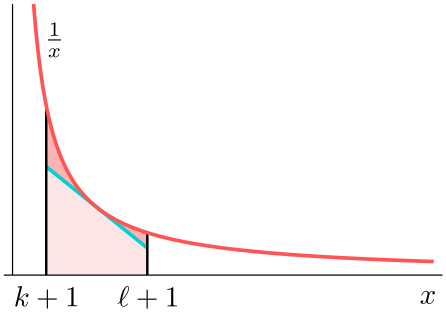


FIGURE 5.6: The right hand side of (5.16) is equal to the dark pink shaded area between the curve $\frac{1}{x}$ and the tangent line shown.

for $s = s_0$. It follows that $\tau_{k,\ell}(s_0) = \upsilon_{k,\ell}(s_0) = 1$. To compare this with the weight function at the same value, we compute

$$\begin{aligned} \omega_{k,\ell}(s_0) &= \frac{\ell+1}{k+1} \cdot e^{-\frac{2(\ell-k)}{n+1}}, \\ \ln \omega_{k,\ell}(s_0) &= \ln \frac{\ell+1}{k+1} - \frac{2(\ell-k)}{n+1}. \end{aligned} \tag{5.16}$$

We interpret the right hand side of (5.16) as the difference between the area below the graph of $\frac{1}{x}$ for $k+1 \le x \le \ell+1$ and the area below the tangent line that touches the graph at the midpoint of the interval, again for $k+1 \le x \le \ell+1$. Since $\frac{1}{x}$ is a convex function, the second area is smaller, which implies $\ln \omega_{k,\ell}(s_0) > 0$ and therefore $\omega_{k,\ell}(s_0) > 1$. This implies (5.17), which is the first of the two pairs of inequalities that describe the relation between the three functions on the left and

the right:

$$\tau_{k,\ell}(s_0) \;=\; \upsilon_{k,\ell}(s_0) \;<\; \omega_{k,\ell}(s_0), \tag{5.17}$$

$$\upsilon_{k,\ell}(s) \;<\; \omega_{k,\ell}(s) \;<\; \tau_{k,\ell}(s), \tag{5.18}$$

as in Figure 5.5. To compare the functions on the right for sufficiently large s, we look at the exponents. Writing $\mu = \frac{s}{2}D_{k,\ell}$, as before, the exponent of $\omega_{k,\ell}(s)$ is $-4\pi\mu^2\frac{\ell-k}{n+1}$, which is clearly smaller than the exponent of $\tau_{k,\ell}(s)$, which is $4\pi\mu\sqrt{\mu^2-\sigma_0^2}$. Less obvious is the comparison with the exponent of $\upsilon_{k,\ell}(s)$, which is $-4\pi\mu\sqrt{\mu^2-\sigma_0^2}$. After dividing by $-4\pi\mu$ and squaring, we obtain $\mu^2(\ell-k)^2/(n+1)^2 < \mu^2-\sigma_0^2$. It follows that the exponent of $\omega_{k,\ell}(s)$ is larger than that of $\upsilon_{k,\ell}(s)$, for s large, which implies (5.18).

From (5.17) and (5.18), we conclude that the number of intersections between the graphs of $\omega_{k,\ell}$ and $\tau_{k,\ell}$ (with multiplicity) is odd. Since both functions are monotonic, with slopes of opposite signs, we have exactly one intersection. It follows that $T_{k,\ell}$ is well defined.

Double Intersection. We use (5.17) and (5.18) to conclude that the number of intersections between the graphs of $\omega_{k,\ell}$ and $\upsilon_{k,\ell}$ (again counting with multiplicity) is even. We will establish that there is only one double intersection, namely at $s = U_n$. To prove that U_n is a solution to the equation $\omega_{k,\ell}(s) = \upsilon_{k,\ell}(s)$, we write $A = (\ell+1)+(k+1)$ and $B = (\ell+1)-(k+1)$, noting that $A^2-B^2 = 4(k+1)(\ell+1)$. Setting

$$s \;=\; \sqrt{\frac{n+1}{2\pi}+\frac{y}{2\pi(n+1)}},$$

and noting that $A = n + 1$, we have

$$\begin{aligned}
\mu &= \sqrt{\frac{A^2 + y}{2\pi(A^2 - B^2)}}, \\
\sqrt{\mu^2 - \sigma_0^2} &= \sqrt{\frac{B^2 + y}{2\pi(A^2 - B^2)}}, \\
\frac{\mu + \sqrt{\mu^2 - \sigma_0^2}}{\mu - \sqrt{\mu^2 - \sigma_0^2}} &= \frac{\sqrt{A^2 + y} + \sqrt{B^2 + y}}{\sqrt{A^2 + y} - \sqrt{B^2 + y}}, \\
-4\pi\mu\sqrt{\mu^2 - \sigma_0^2} &= -\frac{2\sqrt{A^2 + y}\sqrt{B^2 + y}}{A^2 - B^2}.
\end{aligned}$$

Using the definitions of the weight function in (5.13) and the second transition function in (5.15), we arrive at

$$\begin{aligned}
\omega_{k,\ell}(s) &= \frac{(A + B)^2}{A^2 - B^2} \cdot e^{-\frac{2AB + 2\frac{B}{A}y}{A^2 - B^2}}, \\
\upsilon_{k,\ell}(s) &= \frac{\left[\sqrt{A^2 + y} + \sqrt{B^2 + y}\right]^2}{A^2 - B^2} \cdot e^{-\frac{2\sqrt{A^2+y}\sqrt{B^2+y}}{A^2 - B^2}}.
\end{aligned}$$

Clearly, $\omega_{k,\ell}(s) = \upsilon_{k,\ell}(s)$ if $y = 0$, which shows that $s = U_n$ is indeed a solution to the equation. We continue by showing that for small but non-zero y, we have $\upsilon_{k,\ell}(s) < \omega_{k,\ell}(s)$. This is equivalent to showing that the natural logarithm of $\upsilon_{k,\ell}(s)$ over $\omega_{k,\ell}(s)$ is smaller than 0. Equivalently, LHS $<$ RHS, where

$$\begin{aligned}
\text{LHS} &= \ln \frac{\sqrt{A^2 + y} + \sqrt{B^2 + y}}{A + B}, \\
\text{RHS} &= \frac{\sqrt{A^2 + y}\sqrt{B^2 + y} - AB - \frac{B}{A}y}{A^2 - B^2}.
\end{aligned}$$

To prove this inequality for small values of y, we use the Taylor expansions of the square root and the natural logarithm functions:

$$\begin{aligned}\sqrt{A^2+y} &= A+\frac{y}{2A}-\frac{y^2}{8A^3}+\frac{y^3}{16A^5}-\ldots,\\ \ln(1+z) &= z-\frac{z^2}{2}+\frac{z^3}{3}-\frac{z^4}{4}+\ldots.\end{aligned}$$

With this, we can re-write the two sides of the inequality: LHS $= l_1y+l_2y^2+\ldots$, and RHS $= r_1y+r_2y^2+\ldots$. Computing the coefficients, in turn, we find

$$l_1 = \frac{1}{2AB} = r_1,$$

$$l_2 = -\frac{A^2+B^2}{8A^3B^3} < -\frac{A^2-B^2}{8A^3B^3} = r_2.$$

In words, $s=U_n$ is a double solution,and $\upsilon_{k,\ell}(s)<\omega_{k,\ell}(s)$ for values of s chosen in a small neighborhood but different from U_n.

For large values of y (and hence large values of s), we notice that LHS is $O(\ln(y^{1/2}))$ and RHS is $O(y)$. As a result, for large enough y, LHS is strictly less than RHS. For intermediate values of y, we use numerical evidence to support our claim; see Figure 5.5.

5.2.3 One-Dimensional Sections

We follow the 1-section defined by an axis of the n-simplex as the scale factor s goes from zero to infinity. By construction, we have qualitative changes at the transitions, which we now summarize.

Lemma 53 (Maxima of One-Sections)**.** *Let $0\le k\le\ell$ with $k+\ell=n-1$, and let A be the axis passing through the barycenters of a k-face and its complementary ℓ-face of $s\Delta^n$. Then $G_s|_A$ has one maximum whenever $s<T_{k,\ell}$, and two maxima whenever $T_{k,\ell}<s$ and $s\neq U_n$.*

Indeed, the double intersection is responsible for the special evolution of the 1-section. In particular, we go from one maximum for $s < T_{k,\ell}$ to two maxima for $T_{k,\ell} < s < U_n$, of which one is the barycenter of the n-simplex. After the second transition at the double intersection, we still have two maxima, but now the separating minimum is the barycenter of the n-simplex. Figure 5.4 shows all transition scale factors in a single picture for small values of k and ℓ. First, we look at $T_{k,\ell}$ for a fixed value of n. We observe that $T_{k,\ell}$ increases with growing k. This implies that for constant n, the axes defined for small values of k spawn a second maximum earlier than do axes defined by large values of k. For $k = \ell$, the two transitions coincide and the corresponding 1-section does not witness the extra maximum at all. Second, we fix k and observe that $T_{k,\ell}$ increases with growing ℓ. This implies that the k-faces of a low-dimensional simplex spawn second maxima earlier than do the k-faces in high-dimensional simplices.

Next, we look at the second transition U_n. As we have observed, it depends only on n. This implies that all 1-sections lose the maximum at the barycenter at the same scale factor. For $k = \ell$, the two transitions coincide, collapsing the interval to a single point.

5.2.4 n-Dimensional Sections

In this subsection, we show that most maxima of the 1-sections are not modes. We begin with the analysis of the barycenter of $s\Delta^n$, which belongs to every axis of the scaled n-simplex.

Barycenter of n-Simplex. The scaled n-design has the symmetry group of the n-simplex, which implies that the barycenter b_G of $s\Delta^n$ is a critical point of G_s. Indeed, if b_G is not a critical point then it has a non-zero gradient, which contradicts the symmetry.

Lemma 54 (Barycenter Lemma)**.** *Let $n \geq 1$. Then the barycenter of $s\Delta^n$ is a mode of G_s for $s < U_n$ and it is a saddle of index 1 for $s > U_n$.*

Proof. We compute the Hessian of G_s at b_G by taking partial derivatives with respect to the Cartesian basis of $\mathbb{R}^{n+1}$. Because of the symmetry, we have

$$\begin{aligned} d &= \frac{\partial^2 G_s}{\partial x_0^2}(b_G) &&= \frac{\partial^2 G_s}{\partial x_i^2}(b_G), \\ c &= \frac{\partial^2 G_s}{\partial x_0 \partial x_1}(b_G) &&= \frac{\partial^2 G_s}{\partial x_i \partial x_j}(b_G), \end{aligned}$$

for all $0 \leq i \leq n$ and all $j \neq i$. The characteristic polynomial of the Hessian is therefore

$$\det \begin{bmatrix} d-\xi & c & \dots & c \\ c & d-\xi & \dots & c \\ \vdots & \vdots & \ddots & \vdots \\ c & c & \dots & d-\xi \end{bmatrix}. \tag{5.19}$$

Its roots are the eigenvalues of the Hessian, which are $\xi = d + nc$, with multiplicity one, and $\xi = d - c$, with multiplicity n. For simple geometric reasons, $d + nc$ is negative and corresponds to the eigenvector in the diagonal direction of $\mathbb{R}^{n+1}$. To compute $d - c$, it suffices to consider a single 1-section through the barycenter, hence we choose the line B that passes through the vertex $v_0 = se_0$ and the barycenter of the complementary $(n-1)$-face, which we denote by b_0. The distances from the barycenter of the n-simplex to v_0 and b_0 are $\|b_G - v_0\| = sR_n$ and $\|b_G - b_0\| = \frac{s}{n}R_n$ respectively. Furthermore, the common distance of the vertices $v_i = se^i$ from B is $\|v_i - b_0\| = sR_{n-1}$, for $1 \leq i \leq n$. Plugging these distances into (4.5), we compute the second derivatives at b_G of the 1-sections defined by the $n+1$ unit Gaussian

kernels as

$$[4\pi^2 s^2 R_n^2 - 2\pi] \cdot e^{-\pi s^2 R_n^2}, \tag{5.20}$$

$$[4\pi^2 \frac{s^2}{n^2} R_n^2 - 2\pi] \cdot e^{-\pi s^2 (R_{n-1}^2 + R_n^2/n^2)}, \tag{5.21}$$

where the first line applies for $i = 0$ and the second line for $1 \leq i \leq n$. Note that $R_{n-1}^2 + R_n^2/n^2 = R_n^2$ and $R_n^2(1 + \frac{1}{n}) = 1$. Adding (5.20) and n times (5.21), we obtain a root of (5.19):

$$d - c \;\; = \;\; [4\pi^2 s^2 - 2\pi(n+1)] \cdot e^{-\pi s^2 R_n^2}.$$

This eigenvalue has the same sign as $s^2 - \frac{n+1}{2\pi}$, indicating that b_G is a maximum of G_s for $s < U_n$ and a saddle of index 1 for $s > U_n$ as claimed. □

We note here that the barycenter is an index-1 saddle for $s > U_n$, as opposed to a minimum, because we place the n-simplex in $\mathbb{R}^{n+1}$. If we were to consider the restriction of G_s to the hyperplane containing $s\Delta$, then we would find a local minimum at the barycenter of the scaled simplex for all $s > U_n$. At the transition, when $s = U_n$, the barycenter of the n-simplex is a degenerate critical point that corresponds to a maximum of the mixture.

Orthogonal Sections. We generalize the analysis of the barycenter. Let $1 \leq k \leq \ell$ with $k+\ell = n-1$, and consider a k-face of the n-simplex as well as the complementary ℓ-face. Writing G_s as the sum of the g_i, for $0 \leq i \leq n$, we assume that the centers of g_0 to g_k span the k-face, and that the centers of g_{k+1} to g_n span the ℓ-face. Hence, $G_s = K_s + L_s$, where $K_s = \sum_{i=0}^{k} g_i$ and $L_s = \sum_{i=k+1}^{n} g_i$. Writing b_K and b_L for the barycenters of the two faces, we let A be the axis defined by $A(t) = (1-t)b_K + tb_L$. We are interested in the Hessian of G_s at $x = A(t)$. For symmetry reasons, this matrix has at most four distinct eigenvalues, each a second derivative along pairwise orthogonal lines. One line is the axis, another is orthogonal to the n-simplex, a third

line is parallel to the k-face, and a fourth line is parallel to the ℓ-face. The latter two eigenvalues have multiplicity k and ℓ. We write κ for the length parameter along the third eigenvector and λ for the length parameter along the fourth eigenvector.

Lemma 55 (Partial Derivatives Along n-Section)**.** *Let $1 \leq k \leq \ell$ with $k+\ell = n-1$. The second order partial derivatives of G_s at $x = A(t)$ along lines parallel to the complementary k- and ℓ-faces of $s\Delta^n$ are*

$$\frac{\partial^2 G_s}{\partial \kappa^2}(x) = -2\pi G_s(x) + 4\pi^2 s^2 g_0(x), \tag{5.22}$$

$$\frac{\partial^2 G_s}{\partial \lambda^2}(x) = -2\pi G_s(x) + 4\pi^2 s^2 g_n(x). \tag{5.23}$$

Proof. Recall $G_s = \sum_{i=0}^{n} g_i$ and $g_i(x) = e^{-\pi \|x - s e_i\|^2}$. The derivative with respect to the i-th coordinate direction is

$$\frac{\partial G_s}{\partial x_i}(x) = -2\pi x_i G_s(x) + 2\pi s g_i(x).$$

Deriving again, with respect to the same and a different coordinate direction, we have

$$\frac{\partial^2 G_s}{\partial x_i^2}(x) = [-2\pi + 4\pi^2 x_i^2] G_s(x) - 4\pi^2 (2 s x_i - s^2) g_i(x), \tag{5.24}$$

$$\frac{\partial^2 G_s}{\partial x_i \partial x_j}(x) = 4\pi^2 [x_i x_j G_s(x) - s x_i f_j(x) - s x_j g_i(x)]. \tag{5.25}$$

The point at which we take the second derivative has only two distinct coordinates, $\frac{(1-t)s}{k+1}$, repeated $k+1$ times, and $\frac{ts}{\ell+1}$, repeated $\ell+1$ times. We can therefore substitute x_0 and x_1 for any two among the first $k+1 \geq 2$ coordinate directions, and we can substitute x_n and x_{n-1} for any two among the last $\ell + 1 \geq 2$ coordinate directions.

The Hessian at the point x is

$$H(x) \;=\; \begin{bmatrix} d & \dots & c & \gamma & \dots & \gamma \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ c & \dots & d & \gamma & \dots & \gamma \\ \gamma & \dots & \gamma & D & \dots & C \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \gamma & \dots & \gamma & C & \dots & D \end{bmatrix},$$

where

$$d \;=\; \frac{\partial^2 G_s}{\partial x_0^2}(x), \qquad c \;=\; \frac{\partial^2 G_s}{\partial x_0 \partial x_1}(x), \tag{5.26}$$

$$D \;=\; \frac{\partial^2 G_s}{\partial x_n^2}(x), \qquad C \;=\; \frac{\partial^2 G_s}{\partial x_n \partial x_{n-1}}(x), \tag{5.27}$$

$$\gamma \;=\; \frac{\partial^2 G_s}{\partial x_0 \partial x_{n-1}}(x). \tag{5.28}$$

We obtain the eigenvalues of H as the roots of the characteristic polynomial, which we find by subtracting the variable ξ from each diagonal element and taking the determinant, as in (5.19). In particular, $d-c$ is the k-fold eigenvalue that corresponds to the k-face, and $D-C$ is the ℓ-fold eigenvalue that corresponds to the ℓ-face. Plugging (5.24) and (5.25) into (5.26) and (5.27), we arrive at

$$\begin{aligned} d - c &= -2\pi G_s(x) + 4\pi^2 s^2 f_0(x), \\ D - C &= -2\pi G_s(x) + 4\pi^2 s^2 f_n(x). \end{aligned}$$

These are the two claimed second derivatives of (5.22) and (5.23). □

Sign Change. A point $x = A(t)$ is a mode of $G_s : \mathbb{R}^{n+1} \to \mathbb{R}$ iff it is a maximum of the 1-section defined by A as well as of the n-section defined by the hyperplane at x orthogonal to A. Focusing on the latter, we compute the values of the parameter t at which the second derivatives with respect to κ and with respect to λ vanish.

Beginning with κ, we set (5.22) to zero and find

$$[4\pi^2 s^2 - 2\pi(k+1)]g_0(x) \;=\; 2\pi(\ell+1)g_n(x). \tag{5.29}$$

We note that the natural logarithm of $g_n(x)/g_0(x)$ is $-\pi$ times the following difference of squared distances:

$$\|x - se_n\|^2 - \|x - se_0\|^2 = 2s^2 \cdot \frac{(\ell+1) - t(n+1)}{(k+1)(\ell+1)}. \tag{5.30}$$

Plugging (5.30) into (5.29) gives us

$$\frac{2\pi s^2 - (k+1)}{\ell+1} \;=\; e^{-2\pi s^2 \frac{(\ell+1)-t(n+1)}{(k+1)(\ell+1)}}.$$

Solving this equation, we get t as a function of the scale parameter. We call this function t_K. Doing the symmetric computations for λ, we find a second function $t_L : \mathbb{R} \to \mathbb{R}$, both defined by

$$t_K(s) = \frac{k\ell+n}{2\pi s^2(n+1)} \cdot \ln \frac{2\pi s^2 - k - 1}{\ell+1} + \frac{\ell+1}{n+1}, \tag{5.31}$$

$$t_L(s) = \frac{k\ell+n}{2\pi s^2(n+1)} \cdot \ln \frac{k+1}{2\pi s^2 - \ell - 1} + \frac{\ell+1}{n+1}. \tag{5.32}$$

For example, for $s = U_n$, we get $t_K = t_L = \frac{\ell+1}{n+1}$, which is consistent with the Barycenter Lemma, where $s = \frac{\ell+1}{n+1}$ is identified as the scale factor at which the barycenter of the n-simplex changes from a maximum to a minimum. Note also that t_K is undefined for $s = U_k$, and t_L is undefined for $s = U_\ell$.

Chandelier. To get a feeling for the situation, we draw the trajectories of the critical points of G_s, and in particular those of the modes. We call this set in $\mathbb{R}^{n+1} \times \mathbb{R}$ the *chandelier* of the one-parameter family of functions. Letting s increase from bottom to top, Figure 5.7 sketches the chandelier for $n = 1, 2$. The most

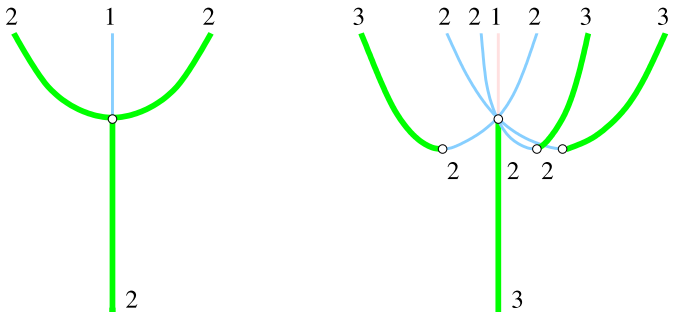


FIGURE 5.7: The chandelier for $n = 1$ on the the left and for $n = 2$ on the right. Each curve is labeled by the index of its critical points.

prominent feature is the *base point*, which we use to decompose the chandelier into curves. Two of these curves are vertical, both swept out by the barycenter of the n-simplex, which changes from index $n+1$ to index 1 when it passes through the base point. For each curve, we define the *height function* by mapping $(x, s) \in \mathbb{R}^{n+1} \times \mathbb{R}$ to s, and we further subdivide the curves so that the height function is injective. In other words, we cut each curve at the local minima and maxima of the height function. The benefit of this subdivision is that now each curve is swept out by a critical point of G_s with constant index. While the total number of curves in the chandelier grows exponentially with the dimension, the number of curves that correspond to modes grows only by one for each dimension. To count the curves, we compute the number of complementary face pairs of the n-simplex:

$$p_n \quad = \quad \frac{1}{2} \sum_{k=0}^{n-1} \binom{n+1}{k+1} \quad = \quad 2^n - 1.$$

For each pair, two branches emanate from the base point. Adding the two vertical lines, we count $2p_n + 2 = 2^{n+1}$ branches. For each complementary face pair with $0 \leq k < \ell$, the height function of one of the two corresponding branches has a local minimum and is therefore subdivided into two curves. The number of local minima

is therefore

$$l_n = \begin{cases} p_n & \text{if } n \text{ is even,} \\ p_n - \frac{1}{2}\binom{n+1}{(n+1)/2} & \text{if } n \text{ is odd.} \end{cases}$$

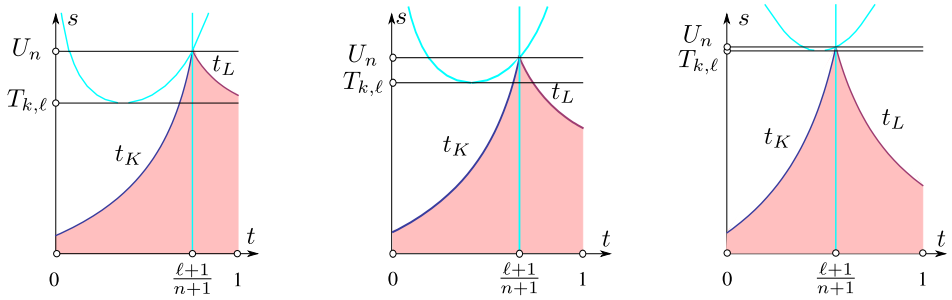
The total number of curves is $2p_n + 2 + l_n$. Of these, only $n+2$ correspond to modes.

Indices. The index of the critical point sweeping out a curve in the chandelier is easy to determine numerically, but at this time, we lack analytic proofs. We first state the result and then explain the numerical evidence that supports it.

$0 \leq k < \ell$: There are $\binom{n+1}{k+1}$ complementary face pairs of k- and ℓ-faces. Besides the barycenter, the corresponding axes witness critical points of index $\ell + 2$ and $\ell + 1$ for $s \in (T_{k,\ell}, U_n)$ and two critical points of index $k + 2$ for $s > U_n$.

$k = \ell = \frac{n-1}{2}$: There are $\frac{1}{2}\binom{n+1}{k+1}$ complementary pairs of k-faces. Besides the barycenter, the corresponding axes witness two critical points of index $k + 2$ for $s > T_{k,k} = U_n$.

To explain the numerical evidence, we consider $t_K(s)$ and $t_L(s)$, which are given by (5.31) and (5.32). We make t_K injective by restricting it to the range $[0, \frac{\ell+1}{n+1}]$ and we make t_L injective by restricting it to the range $[\frac{\ell+1}{n+1}, 1]$. Figure 5.8 plots the inverses of the restricted functions for both even and odd values of n. Drawing the horizontal line for a value of s, we note that the portion below the graphs of t_K and t_L consists of the points x at which the n-section orthogonal to the axis has a maximum at x. For each scale factor s, there is either one or two modes witnessed by $A_{k,\ell}$, drawn in cyan. We notice empirically that the mode at the barycenter, given by $t = \frac{\ell+1}{n+1}$, is the only mode under the piecewise defined curve for $0 < k < \ell$. This means that the only mode is at the barycenter and the other critical points are saddles of index $\ell + 1$ and $\ell + 2$.



(a) $k = 1$ and $\ell = 5$. (b) $k = 2$ and $\ell = 5$. (c) $k = 3$ and $\ell = 4$.

FIGURE 5.8: Graphs of t_K^{-1} and t_L^{-1} for select values of k and ℓ.

5.2.5 Resilient Modes

We have seen that the sum of Gaussian kernels can have a ghost mode. In this subsection, we analyze the significance of this mode, showing that it lasts for an interval of scale factors whose length increases with the dimension.

Balancing Scales. To get started, we need more information on the transition at which the extra minima appear. We get an upper bound on $T_{k,\ell}$ by studying the scale factor at which the weights of the two one-dimensional kernels in the decomposition of G_s restricted to a relevant axis are balanced. For $k = \ell$, the two one-dimensional kernels in the decomposition are always balanced. For $k < \ell$, the *balancing scale factor* is

$$B_{k,\ell} \;=\; \sqrt{\frac{\ln(\ell+1) - \ln(k+1)}{\pi\left(\frac{1}{k+1} - \frac{1}{\ell+1}\right)}}. \tag{5.33}$$

Of particular importance is the balancing scale factor for $k = 0$: $B_{0,n-1} = \sqrt{\frac{n \ln n}{\pi(n-1)}}$. Recomputing the weights using (5.33) gives $(k+1)g(B_{k,\ell}R_k) = (\ell+1)g(B_{k,\ell}R_\ell)$. Similar to $T_{k,\ell}$ and U_n, the balancing scale factor increases with respect to k and ℓ. Numerically, we observed that $B_{k,\ell}$ is not very different, but consistently larger than $T_{k,\ell}$; see Table B.3. We prove that this relationship is not accidental.

Lemma 56 (Transition Lemma)**.** *We have $T_{k,\ell} < B_{k,\ell} < U_n$ for all integers $0 \leq k < \ell$ with $k + \ell = n - 1$.*

Proof. We prove the claim indirectly, by showing that $s = B_{k,\ell}$ gives two maxima in the 1-section along any axis connecting the barycenter of a k-face with the barycenter of the complementary ℓ-face. For balanced weights, we have two maxima iff the centers of the two one-dimensional kernels are further apart than twice the width; see Section 4.1. To prove the latter property, we compute

$$\frac{B_{k,\ell} D_{k,\ell}}{2\sigma_0} \quad = \quad \sqrt{\frac{n+1}{2(\ell-k)} \cdot \ln \frac{\ell+1}{k+1}}, \tag{5.34}$$

using (4.3), (5.1) and (5.33). Recall the logarithmic inequality:

$$\frac{x}{1+\frac{x}{2}} \quad < \quad \ln(1+x),$$

for $x > 0$. Setting $x = \frac{\ell-k}{k+1}$, we see that the right hand side of (5.34) exceeds one for all choices of $0 \leq k < \ell$. This implies that we have two maxima along the axis, which implies that the balancing scale factor lies between the first and second transitions, as claimed. ◻

Resilience. We define the *resilience* of a mode as the length of the interval of scale values at which it exists. This definition is not satisfactory for a general one-parameter family of smooth functions; however, it will suffice in our context, in which we know enough about the modes to follow them through the family parametrized by the scale s. Specifically, we have a single mode for $0 \leq s \leq T_{0,n-1}$, and we have $n+1$ modes for $U_n \leq s$. The picture is more interesting in the interval $T_{0,n-1} < s < U_n$, in which we have $n + 2$ modes. One of these modes is the barycenter of the n-simplex, and we study the resilience of this *extra* mode. The upper endpoint of the

interval is defined in (5.12), and an upper bound for the lower endpoint is given in the Transition Lemma, with the definition of the bound in (5.33):

$$T_{0,n-1} \quad < \quad \sqrt{\frac{\ln n}{\pi(1-1/n)}}.$$

As n goes to infinity, U_n grows roughly like the square root of n, and $T_{0,n-1}$ grows roughly like the square root of the logarithm of n. The gap between the two widens, so that the resilience of the mode at the barycenter of the n-simplex grows roughly like $\sqrt{n}$; see Figure 5.4.

Summary. We are now ready to summarize the findings in regard to the critical points and the modes of the one-parameter family of functions $G_s : \mathbb{R}^{n+1} \to \mathbb{R}$. For values $s < T_{0,n-1}$, we have a single critical point with index $n+1$. Thereafter, we pick up $2\binom{n+1}{k+1}$ critical points at every $T_{k,\ell}$, for $0 \leq k < \ell$, until we accumulated $2l_n + 1$ critical points right before reaching U_n. The barycenter has index $n+1$, and the other critical points come in pairs, with indices $\ell + 2$ and $\ell + 1$, for $\frac{n-1}{2} < \ell \leq n-1$. For $U_n < s$, we have $2p_n + 1$ critical points. The barycenter has index one, and the other critical points come in pairs with indices $\ell + 2$ and $k + 2$, for $\frac{n-1}{2} \leq \ell \leq n-1$. As a sanity check, we consider the Euler-Poincaré formula, which states that the alternating sum of critical points is equal to the Euler characteristic of $\mathbb{R}^{n+1}$:

$$\sum_{i=0}^{n+1} (-1)^i c_i \quad = \quad (-1)^{n+1}, \tag{5.35}$$

where c_i counts the critical points with index i. We also write $c = \sum_{i=0}^{n+1} c_i$. Trivially, (5.35) holds in the first case. Thereafter, we pick up the critical points in pairs whose contribution to the alternating sum cancel, so (5.35) is maintained. Finally, for $U_n < s$, we have a bijection between the critical points and the faces of the n-simplex such that the index is $n+1$ minus the dimension of the face. Since the

n-simplex is a closed ball, its Euler characteristic is one, which again implies (5.35). We thus have a complete description of the critical points of the n-design as the scale factor increases from zero to infinity.

Theorem 57 (Ghost Modes)**.** *Let $n \geq 1$ and consider the sum of $n+1$ unit Gaussian kernels placed at the vertices of the scaled standard n-simplex $s\Delta^n$.*

(1) For $s < T_{0,n-1}$, we have one critical point which is also a mode.

(2) For $T_{0,n-1} < s < U_n$, we have gradually more critical points after passing each $T_{k,\ell}$, until we accumulate $2l_n+1$ critical points right before U_n. Of these critical points, $n+2$ are modes and they exist during the entire interval.

(3) For $U_n < s$, we have $2p_n + 1$ critical points, of which $n+1$ are modes.

The resilience of the extra mode in Case (2) is $U_n - T_{0,n-1}$, which grows like $\sqrt{n}$.

5.3 Many Ghosts

In this subsection, we construct a finite configuration of isotropic Gaussian kernels with a super-linear number of modes. While there is a family of such constructions, it will suffice to explain one. This construction was first shown in [14]. Prior to this, 5/3 times the number of components was the greatest number of modes that could be reached using Gaussian kernels in Euclidean space. The arrangement that was used to obtain this result was attributed to David MacKay. Obtaining that many modes was a limit behavior, using Gaussian kernels placed at the vertices of the Kekule lattice.

Products of Simplices. The basic building block of our construction is the standard 2-simplex. Let the dimension be $3n$ and write the $3n$-dimensional Euclidean space as

the Cartesian product of n three-dimensional planes: $\mathbb{R}^{3n} = H_1 \times H_2 \times \ldots \times H_n$, in which H_i is spanned by the three coordinate vectors e_{3i-2}, e_{3i-1}, e_{3i}, for $1 \leq i \leq n$. Let Δ_i^2 be the standard 2-simplex in H_i, with vertices $v_{i0} = e_{3i-2}$, $v_{i1} = e_{3i-1}$, $v_{i2} = e_{3i}$. Correspondingly, we write $g_{ij} : H_i \to \mathbb{R}$ for the 3-dimensional unit Gaussian kernel with center v_{ij}, for $0 \leq j \leq 2$, and $G_i : H_i \to \mathbb{R}$ defined by

$$G_i(x) \quad = \quad g_{i0}(x) + g_{i1}(x) + g_{i2}(x)$$

for their sum. Next, we construct a $3n$-dimensional sum of Gaussian kernels by taking products. To begin, we let $P \subseteq \mathbb{R}^{3n}$ be the largest subset of points whose orthogonal projection to H_i is $\{v_{i0}, v_{i1}, v_{i2}\}$, for $1 \leq i \leq n$. This is the set of 3^n points formed by taking the Cartesian product of the n triplets of points. For each point $p \in P$, let $f_p : \mathbb{R}^{3n} \to \mathbb{R}$ be the unit Gaussian kernel with center p. Adding these kernels, we get $F : \mathbb{R}^{3n} \to \mathbb{R}$, defined by

$$F(x) \quad = \quad \sum_{p \in P} f_p(x).$$

To understand F, we recall that f_p can be written as the product of $3n$ 1-dimensional unit Gaussian kernels; see (4.7). Collecting the terms in sets of three, we write

$$f_p(x) \quad = \quad \prod_{i=1}^{n} g_{ij}(x),$$

where j is chosen such that v_{ij} is the orthogonal projection of p onto H_i. Substituting the sum of the three kernels for the singletons, we obtain

$$F(x) \quad = \quad \prod_{i=1}^{n} G_i(x).$$

In words, the sum of the 3^n $3n$-dimensional unit Gaussian kernels is the product of n sums of three 3-dimensional unit Gaussian kernels.

Counting Modes. We arrive at the final construction by reintroducing the scale factor, writing $F_s : \mathbb{R}^{3n} \to \mathbb{R}$ for the product of the $G_{is} : H_i \to \mathbb{R}$, where G_{is} is of course the sum of the three unit Gaussian kernels with centers se_{3i-2}, se_{3i-1}, se_{3i}. We have seen in Section 5.2 that s can be chosen such that G_{is} has 4 modes. Since F_s is the product of the G_{is}, its sets of modes is the largest subset of $\mathbb{R}^{3n}$ whose orthogonal projection to H_i is the set of four modes of G_{is}, for $1 \leq i \leq n$. Its size is $4^n = 3^{(1+\log_3 \frac{4}{3})n} > 3^{1.261n}$. This shows that the number of modes is roughly the number of kernels to the power 1.261.

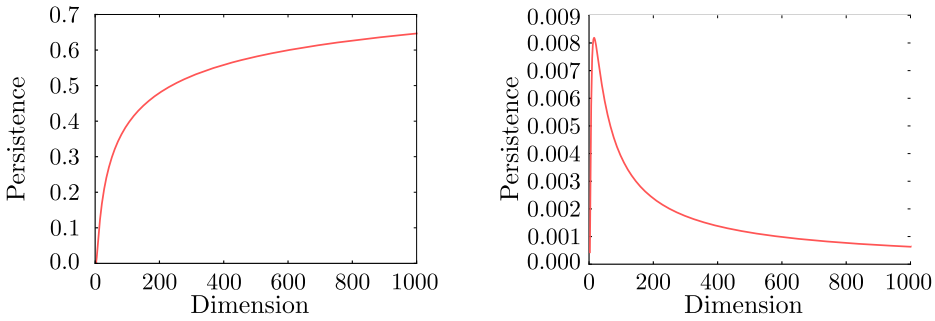
There is an entire family of similar constructions. The one presented here neither maximizes the number nor the resilience of the extra modes. Indeed, we can increase the exponent by improving the ratio of modes over kernels in each H_i (see [9]), and we can improve the resilience by using higher-dimensional simplices.

5.4 Numerical Evidence

This section focuses on the empirical results surrounding mixtures. First, we look at the relationship between the persistence of the ghost feature and the number of components (and hence dimension) of the Gaussian mixture. Second, we investigate the number and location of modes of the n-design when we replace the Gaussian kernel with another kernel.

5.4.1 Persistence of the Ghost Features

We are interested in investigating the persistence of the Gaussian mixture. In particular, we are interested in the persistence of the ghost mode. In other words, how large is the difference in function value between an extra mode and the highest saddle separating it from another mode? We derive the persistence pairings from our empirical knowledge of the indices of the critical points of the scaled n-design. Let $s \in (T_{0,n-1}, U_n)$ and consider the mixture G_s. We will call the mode closest to



(a) Gaussian Mixture (b) Gaussian Probability Mixture

FIGURE 5.9: We plot the persistence of the ghost feature with respect to dimension. On the left, we see the graph for the Gaussian mixture G_s and on the right we see this graph divided by $n+$, which is the graph for the probability density obtained by normalizing G_s.

vertex v_i of $s\Delta^n$ the i^{th} *mode* and the $(n+2)$-nd mode the *ghost mode*. Let $\mathcal{A}_k$ be the set of axes connecting the barycenter of a k-face with the barycenter of the complementary face. All modes (index $n+2$ critical values) of G_s lie on an axis in $\mathcal{A}_0$. Furthermore, all critical points of index n lie either in $\mathcal{A}_0$ or $\mathcal{A}_1$. For geometric reasons, the barycenter must be paired with a point on an axis in $\mathcal{A}_0$ or with the point at infinity. If paired with a point on an axis, then the *persistence of the ghost feature* is the height distance between the barycenter and the point with which it is paired. If the barycenter is paired with the point at infinity, then for each A in $\mathcal{A}_0$, the minimum along A is paired with the non-barycenter maximum along A. In this case, the *persistence of the ghost feature* is defined to be the height difference between the minimum and the non-barycenter maximum on A. In either case, the persistence of the ghost feature is the height difference between the minimum and the lower maximum along A. Empirically, we have observed that the maximum persistence occurs for the balancing scale factor. Thus, we compare the persistence of the ghost feature at the balancing scale factor as the dimension increases.

By (5.3), we can compute the values of the Gaussian mixture along a one-section, then compute the persistence of that one-section in order to find the persistence of

the ghost feature. We do this for dimensions three through 1000 and find that the persistence grows like $\ln n$; see Figure 5.9a. We observe that by summing $n+1$ Gaussian kernels, the integral under the graph of the sum is $n+1$. To force the graph of the sum to be a probability distribution, we divide by $n+1$ to obtain Figure 5.9b. We call this normalized sum the *normalized Gaussian mixture*. Empirically, we have found that the persistence of the ghost feature is maximized when n is 16.

5.4.2 Other Kernels

The arguments presented so far in this chapter are specific to the Gaussian kernel. We ask, *how general are the results that we have presented?* Recalling Theorem 44, we analyze the necessity of using the Gaussian kernel to prove Lemma 52, which requires the kernel to be separable and isotropic in order to define the restriction to the axes and differentiable in order to define the critical points. Thus, the only kernel for which we have proven that the critical points lie on the axes of the scaled n-simplex is the Gaussian kernel. To determine where other kernels would fail for non-separable kernels, we turn to the proof of Lemma 52, which uses the convexity of e^x. More generally, the proof holds for all differentiable kernels $K_t(z) = f_t(\|z\|)$ such that f_t' is strictly convex. The first example that we look at is

$$f_t(x) \quad = \quad \frac{-x^2 t}{x^4 + 4x^2 t + 6t^2}.$$

We notice that $f_t(x)$ is bi-modal, yet symmetric. Hence, in $\mathbb{R}^2$, the graph of $K_t(z) = f_t(\|z\|)$ looks like a volcano. For this reason, we call this kernel the *volcano kernel*. The n-design with the volcano kernel can have twice as many modes as kernels as seen in Figure 5.10b. In addition, the critical points are not restricted to the axes. This kernel, however, is rather unsatisfying, as it is bi-modal. Thus, we look next at

$$h_t(x) \quad = \quad \frac{t^2}{x^4 + 6t}.$$

The kernel defined by $K_t(z) = h_t(\|z\|)$ is unimodal and, as shown in Figure 5.10c, can have more modes than components. The most surprising fact here is that these modes occur off of the axes. The location of the modes in this figure leads us to the following conjecture:

Conjecture 58 (Quadratic Number of Modes)**.** *The n-design mixture of kernels of the form $K_t(z - \mu_i) = h_t(\|z - \mu_i\|)$ can have $O(n^2)$ modes.*

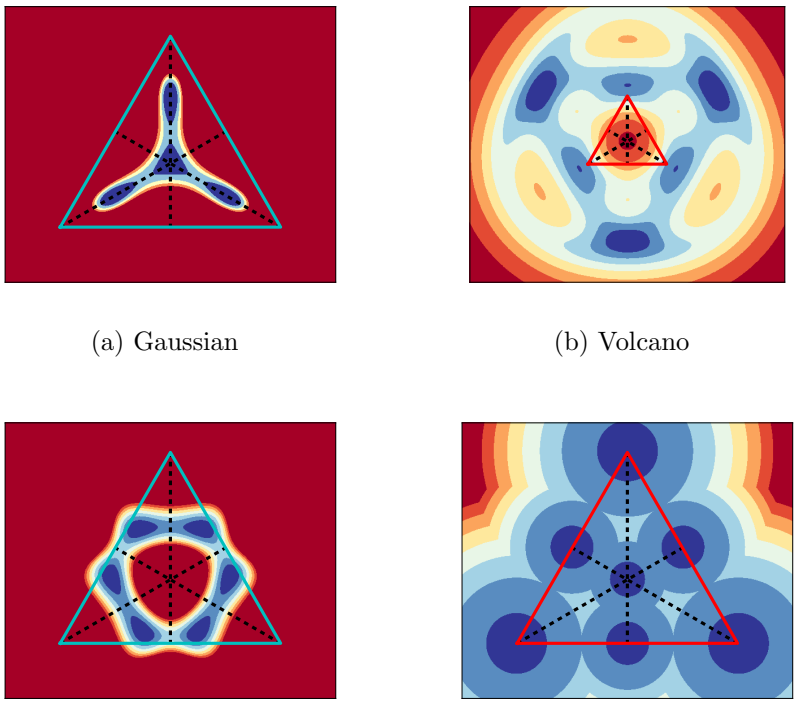
Finally, we consider the Epanechinikov kernel. While all critical points seem to be located on the axes, the n-design mixture formed with this kernel, as shown in Figure 5.10d, can have seven modes in the 2-design. This is the most that we have seen yet. In fact, we suspect the following conjecture is true:

Conjecture 59 (Exponential Number of Modes)**.** *The n-design mixture of kernels of the form $K_t(z - \mu_i) = \mathcal{E}(x - \mu_i, t)$ can have $2^{n+1} - 1$ modes.*

In sum, for almost all types of kernels, we find that there can be more modes than components. The problem is that that these modes are intractable; that is, we cannot define a one-manifold on which they must occur, unlike the Gaussian case where we know that all modes and all critical values must lie on the axes of the n-simplex.

5.5 Discussion

The main contribution of this chapter is a cautionary message about the sum of Gaussian kernels. Giving a detailed analysis of the n-design by examining restrictions of a high-dimensional mixture to a line, we show that there is indeed exactly one extra mode, but that its resilience increases like the square root of the dimension. We also exhibit configurations of finitely many identical isotropic Gaussian kernels whose sums have super-linearly many modes. We saw empirically that the persistence of the



(a) Gaussian

(b) Volcano

(c) Rational Polynomial

(d) Epanechinikov

FIGURE 5.10: We use the standard n-design with various kernels and observe the behavior. The modes of these mixtures are located inside the dark blue regions. Starting at the top left, we see the Gaussian kernel with four modes, the volcano kernel with six modes, the Epanechinikov kernel producing seven modes, and a rational polynomial kernel producing six kernels, all of which are off the axes.

ghost features increases logarithmically with dimension. However, when the Gaussian mixture is normalized to produce a probability distribution, then the persistence with respect to dimension increases for dimensions one through 16 and decreases after dimension 16. We conjecture that the persistence decreases like $\ln n/n$, but at this time, we do not have a proof of this rate of diminishing persistence.

Our intuition for scale space must originate from what we are most familiar: low dimensional examples. In $\mathbb{R}^1$, scale space does not create spurious features. Looking at $\mathbb{R}^2$, we recall the mountain and bridge mode creation seen in Example 30. While mode creation is a generic (but surprising) event in scale space, the creation of a ghost mode is especially surprising. The ghost mode seen in our n-design grows

from a void (when running our design for a normalized sum backwards); whereas, the second mountain peak of the mountain and bridge example is formed when the bridge erodes. The mountain does not rise, but is carved out. This process aligns with our intuition, formalized by Axiom 33 that local extrema should not be enhanced in the diffusion process. Thus, our example must cause a degenerate critical point in the levelsurfaces of scale space. For this reason, caution must be taken when using Gaussians for both parametric and nonparametric density estimation, especially for dimensions around 16, where the effects of the ghost appear to be maximized.

Appendix A

Notations

In this appendix, we summarize the mathematical symbols used throughout this dissertation. Where appropriate, equation numbers are given for definitions.

Table A.1: Mathematical Symbols Introduced in Chapter 2.

	Definition
$\mathrm{Amp}(f)$	The maximum difference in function values of f (2.8).
$d^{\infty}(a,b)$	The L_∞ distance between points a and b (2.17).
$d_{GH}(\mathcal{S}_1,\mathcal{S}_2)$	The Gromov-Hausdorff distance between $\mathcal{S}_1$ and $\mathcal{S}_2$ (2.10).
$d_{\mathbb{M}}(x,y)$	The geodesic distance between points x and y in $\mathbb{M}$ (2.4).
$\mathrm{Dgm}_p(f)$	The p^{th} persistence diagram of f.
E	The essential matrix or the fundamental matrix (2.2).
$f_{\mathcal{R}}(\sigma)$	The diameter of simplex σ.
$F_\delta(\gamma_1,\gamma_2)$	Set of (i,j) such that $\|\gamma_1(i)-\gamma_2(j)\| \le \delta$ (2.14).
$\mathcal{H}(\mathcal{S}_1,\mathcal{S}_2), \mathcal{H}_Z(\mathcal{S}_1,\mathcal{S}_2)$	The Hausdorff Distance between $\mathcal{S}_1$ and $\mathcal{S}_2$ (2.9).
$H_p(\mathbb{M})$	The p^{th} homology group of $\mathbb{M}$ (2.6).
$\ell(\gamma)$	The length of curve γ.
$\mathbb{M}$	An m-manifold.
$\mathbb{M}_s(f), \mathbb{M}_s$	A sublevel set of the function f (2.5).
$\mathbb{M}^s(f), \mathbb{M}^s$	A superlevel set of the function f (2.5).
$N(r)$	The size of the triangulation with mesh size at most r.
$\mathcal{O}$	The set of ordered correspondences between $\mathrm{Vert}(P)$ and $\mathrm{Vert}(Q)$.
$\mathrm{Pers}_k(f)$	The degree k total persistence (2.7).
$r(n)$	Ratio of the volume of $\mathbb{S}^{n-1}$ to the volume of $\mathbb{S}^n$ (2.26).
$\mathcal{R}_\delta(P_1,P_2)$	Reachable free space in $F_\delta(P_1,P_2)$.
v_n	The volume of the unit n-ball (2.23).
$W_q(A,B)$	The Wasserstein matching cost between sets A and B (2.18).
$W_\infty(A,B)$	The bottleneck matching cost between sets A and B (2.19).
λ	A Lipschitz constant.
σ	A simplex.

Table A.2: Mathematical Symbols Introduced in Chapter 3.

	Definition
$\mathcal{F}(\gamma_1,\gamma_2)$	The Fréchet distance between curves γ_1 and γ_2.
$\kappa(\gamma)$	The total curvature of γ (3.3).
$\mathrm{Mesh}(P)$	Maximum parametrized distance between two vertices of P.

Table A.3: Mathematical Symbols Introduced in Chapter 4.

	Definition
$g(x)$	The Gaussian kernel centered at the origin with width σ_0.
$\mathcal{E}(x-\mu,\sigma)$	The Epanechinikov kernel (4.20).
H	The set of all bounded functions $h\colon \mathbb{R}^n \to \mathbb{R}$.
K_t	A kernel used to define scale space.
$\mathcal{N}_{n+1}(x-\mu,\sigma^2)$	The Gaussian kernel in $\mathbb{R}^{n+1}$ centered at μ with width σ (4.1).
$r(x), r_w(x)$	The ratio function and the weighted ratio function (4.8), (4.9).
σ_0	The width of the unit kernel, $\sigma_0 = 1/\sqrt{2\pi}$.
$\Delta_{\mathbb{M}}$	The Laplace-Beltrami operator (4.10).

Table A.4: Mathematical Symbols Introduced in Chapter 5.

	Definition
U_n	The upper transition factor (5.12).
$\omega_{k,\ell}(s)$	The ratio of weights of restriction to an axis (5.13).
$\tau_{k,\ell}(s)$	A transition functions (5.14).
$\upsilon_{k,\ell}(s)$	A transition functions (5.15).

Appendix B

Scale Factor Look-up Tables

The values in Table B.1 were computed with (5.12). Since we do not have a closed form of the lower transition T_{kl}, Newton's method was implemented in Mathematica to find an approximation of the roots of $\omega_{k,\ell}(s) - \tau_{k,\ell}$ in Table B.2.

Table B.1: The upper transition scale factor U_n.

n	2	3	4	5	6	7	8
U_n	0.690988	0.797885	0.892062	0.977205	1.0555	1.12838	1.19683
n	9	10	11	12	13	14	15
U_n	1.26157	1.32314	1.38198	1.43841	1.49271	1.5451	1.59577

Table B.2: The lower transition scale factor $T_{k,\ell}$ for small values of k, ℓ.

	k=0	1	2	3	4	5	6	7	8	9	10	11
$l = 0$												
1	0.6610											
2	0.7157	0.8785										
3	0.7532	0.9349	1.0474									
4	0.7814	0.9778	1.1014	1.1913								
5	0.8038	1.10122	1.1450	1.2424	1.3190							
6	0.8224	1.0652	1.1814	1.2853	1.3674	1.4352						
7	0.8382	1.0864	1.2125	1.3221	1.4091	1.4812	1.5425					
8	0.8519	1.1051	1.2397	1.3543	1.4457	1.5216	1.5864	1.6427				
9	0.8639	1.1218	1.2637	1.3828	1.4781	1.5576	1.6255	1.6847	1.7371			
10	0.8747	1.1368	1.2852	1.4083	1.5073	1.5899	1.6607	1.7226	1.7774	1.8266		
11	0.8844	1.1505	1.3046	1.4315	1.5337	1.6192	1.6927	1.7570	1.8141	1.8654	1.9119	
12	0.8932	1.1505	1.3223	1.4526	1.5578	1.6461	1.7220	1.7886	1.8477	1.9010	1.9493	1.9935
13	0.9013	1.1631	1.3385	1.4719	1.5800	1.6707	1.7490	1.8176	1.8788	1.9339	1.9839	2.0297
14	0.9088	1.1747	1.3534	1.4898	1.6004	1.6936	1.7739	1.8446	1.9076	1.9664	2.0160	2.0633

Table B.3: The balancing scale factor for small values of k, ℓ.

	k=0	1	2	3	4	5	6	7	8	9	10	11
$l = 0$												
1	0.664282											
2	0.724257	0.879989										
3	0.767047	0.939437	1.04827									
4	0.800234	0.986009	1.10431	1.19188								
5	0.827286	1.02425	1.15057	1.24449	1.31949							
6	0.85008	1.05667	1.18993	1.2894	1.36905	1.43556						
7	0.86975	1.08477	1.22417	1.32856	1.41236	1.48247	1.5428					
8	0.88703	1.10955	1.25445	1.36327	1.45081	1.52419	1.58741	1.64298				
9	0.902426	1.1317	1.28158	1.39443	1.48538	1.56173	1.62761	1.68557	1.73734			
10	0.916298	1.15172	1.30614	1.42268	1.51677	1.59587	1.66419	1.72436	1.77816	1.8268		
11	0.928912	1.16996	1.32856	1.44851	1.54551	1.62715	1.69775	1.75998	1.81565	1.86603	1.91205	
12	0.940471	1.18671	1.34919	1.47231	1.57201	1.65602	1.72874	1.79289	1.85033	1.90234	1.94987	1.99365
13	0.951134	1.20219	1.36828	1.49435	1.59658	1.68282	1.75753	1.82349	1.88258	1.93613	1.98509	2.03019
14	0.961026	1.21658	1.38605	1.51489	1.61949	1.70782	1.7844	1.85207	1.91273	1.96772	2.01802	2.0644

Bibliography

[1] Agarwal, P. K., Avraham, R. B., Kaplan, H., and Sharir, M. Computing the discrete Fréchet distance in subquadratic time. ArXiv preprint at http://arxiv.org/abs/1204.5333, Apr. 2012. Referenced on 27.

[2] Agarwal, P. K., Efrat, A., and Sharir, M. Vertical decomposition of shallow levels in 3-dimensional arrangements and its applications. pp. 39–50. 1995. Referenced on 32.

[3] Alt, H., and Buchin, M. Semi-computability of the Fréchet distance between surfaces. pp. 45–48. 2005. Referenced on 25.

[4] Alt, H., and Godau, M. Computing the Fréchet distance between two polygonal curves. *Internat. J. Comput. Geom. Appl. 5*, 1 (Mar. 1995), 75–91. Referenced on 24, 25.

[5] Alt, H., Knauer, C., and Wenk, C. Comparison of distance measures for planar curves. *Algorithmica 38*, 1 (Oct. 2003), 45–58. Referenced on 21.

[6] Alt, H. W. Analysis III. Online., 2007. Lecture Winter Semester 2001–2002 at the Inst. of Applied Math., Univ. of Bonn. Referenced on 64.

[7] Babaud, J., Witkin, A. P., Baudin, M., and Duda, R. O. Uniqueness of the Gaussian kernel for scale-space filtering. *IEEE Trans. Pattern Anal. Mach. Intell. 8*, 1 (Jan. 1986), 26–32. Referenced on 56, 71, 74, 75, 81.

[8] Babić, D., Klein, D. J., Lukovits, I., Nikolić, S., and Trinajstić, N. Resistance-distance matrix: A computational algorithm and its application. *Internat. J. Quantum Chem. 90*, 1 (2002), 166–176. Referenced on 79.

[9] Behboodian, J. On the modes of a mixture of two normal distributions. *Technometrics 12*, 1 (1970), 131–139. Referenced on 4, 56, 62, 108.

[10] Brauer, F., and Nohel, J. A. *Ordinary Differential Equations: a First Course*, 2 ed. W. A. Benjamin, Menlo Park, CA, 1973. Referenced on 59.

[11] Burden, R. L., and Faires, J. D. *Numerical Analysis*, eighthẽd. Thompson, 2005. Referenced on 64, 77, 80.

[12] Burke, P. J. Solution of problem 4616 [1954, 718], proposed by A. C. Cohen, Jr. *Amer. Math. Monthly 63*, 2 (Feb. 1956), 129. Referenced on 4, 60.

[13] Cagliari, F., Ferri, M., and Pozzi, P. Size functions from a categorical viewpoint. *Acta Appl. Math. 67*, 3 (2001), 225–235. Referenced on 15.

[14] Carreira-Perpinán, M., and Williams, C. An isotropic gaussian mixture can have more modes than components. Informatics Research Report EDI-INF-RR-0185, Institute for Adaptive and Neural Computation, University of Edinbugh, Dec. 2003. Referenced on 4, 81, 106.

[15] Carreira-Perpinán, M., and Williams, C. On the number of modes of a Gaussian mixture. *Scale Space Methods in Computer Vision, Lecture Notes in Computer Science 2695* (2003), 625–640. Referenced on 56, 76.

[16] Chakerian, G. D. An inequality for closed space curves. *Pacific J. Math. 12*, 1 (1962), 53–57. Referenced on 41, 42.

[17] Chazal, F., Cohen-Steiner, D., Guibas, L. J., Mémoli, F., and Oudot, S. Y. Gromov-Hausdorff stable signatures for shapes using persistence. In *Proc. of the Seventh Eurographics Symp. Geom. Processing* (2009), vol. 28.5, Eurographics, pp. 1393–1403. Symposium held in Berlin, Germany. Referenced on 10, 21, 33.

[18] Chen, C., and Edelsbrunner, H. Diffusion runs low on persistence fast. In *Proc. of the Thirteenth Internat. Conf. Comput. Vis.* (2011), pp. 423–430. Referenced on 57.

[19] Chen, C., and Kerber, M. An output sensitive algorithm for persistent homology. In *Proc. of the* 27^{th} *Annu. Symp. Comput. Geom.* (New York, Jun. 2011), ACM, pp. 207–215. Symposium held in Paris, France. Referenced on 20.

[20] Cohen-Steiner, D., and Edelsbrunner, H. Inequalities for the curvature of curves and surfaces. *Found. Comput. Math. 7*, 4 (Nov. 2007), 391–404. Referenced on 3, 5, 41, 42, 47, 48.

[21] Cohen-Steiner, D., Edelsbrunner, H., and Harer, J. Stability of persistence diagrams. pp. 263–271. 2005. Referenced on 34.

[22] Cohen-Steiner, D., Edelsbrunner, H., and Harer, J. Extending persistence using Poincaré and Lefschetz duality. *Found. Comput. Math. 9*, 1 (Feb. 2009), 79–103. Referenced on 15.

[23] Cohen-Steiner, D., Edelsbrunner, H., Harer, J., and Mileyko, Y. Lipschitz functions have l_p-stable persistence. *Found. Comput. Math. 10*, 2 (Apr. 2010), 127–139. Referenced on 16, 17, 35.

[24] Cohen-Steiner, D., Edelsbrunner, H., and Morozov, D. Vines and vineyards by updating persistence in linear time. pp. 119–126. 2006. Referenced on 12, 34, 35.

[25] Cook IV, A. F., and Wenk, C. Geodesic Fréchet distance inside a simple polygon. pp. 193–204. 2008. Referenced on 26.

[26] Damon, J. Local morse theory for solutions to the heat equation and Gaussian blurring. *J. Differential Equations 115*, 2 (Jan. 1995), 368–401. Referenced on 56, 69, 71.

[27] Di Fabio, B., and Landi, C. Stability of reeb graphs under function perturbations: the case of closed curves. http://arxiv.org/abs/1003.4610, April 2010. University of Bologna. Referenced on 34.

[28] Do Carmo, M. P. *Differential Geometry of Curves and Surfaces*, vol. 2. Prentice Hall, Englewood Cliffs, NJ, 1976. Published originally in Portuguese and translated to English. Referenced on 7.

[29] Driemel, A., Har-Peled, S., and Wenk, C. Approximating the Fréchet distance for realistic curves in near linear time. In *Proc. of the* 26^{th} *Annu. Symp. Comput. Geom.* (New York, Jun. 2010), ACM. Symposium held in Snowbird, UT. Referenced on 9, 29, 30.

[30] Duits, R., Florack, L., Graaf, J. D., and Romeny, B. T. H. On the axioms of scale space theory. *J. Math. Imaging Vision* (2004), 267–298. Referenced on 71, 73, 74.

[31] Edelsbrunner, H., Fasy, B. T., and Rote, G. Add isotropic Gaussian mixtures at own risk: More and more resilient modes in higher dimensions. In *Proc. of the* 27^{th} *Annu. Symp. Comput. Geom.* (Jun. 2012), ACM. Symposium held in Chapel Hill, NC. Referenced on 4, 81, 128.

[32] Edelsbrunner, H., and Harer, J. Persistent homology–a survey. In *Surveys on Discrete and Computational Geometry–Twenty Years Later*, vol. 453 of

Contemp. Math. Amer. Math. Soc., 2006, pp. 257–282. AMS–IMS–SIAM Joint Summer Research Conference at Snowbird, UT. Referenced on 5, 11, 16, 17.

[33] Edelsbrunner, H., and Harer, J. *Computational Topology. An Introduction.* Amer. Math. Soc., Providence, RI, 2010. Referenced on 5, 16.

[34] Edelsbrunner, H., Letscher, D., and Zomorodian, A. Topological persistence and simplification. *Discrete Comput. Geom. 28*, 4 (Jul. 2002), 511–533. Referenced on 15, 17.

[35] Edelsbrunner, H., Morozov, D., and Pascucci, V. Persistence-sensitive simplifcation of functions on 2-manifolds. In *Proc. of the* 22^{nd} *Annu. Symp. Comput. Geom.* (New York, Jun. 2006), ACM. Symposium held in Sedona, AZ. Referenced on 16.

[36] Efrat, A., Itai, A., and Katz, M. J. Geometry helps in bottleneck matching and related problems. *Algorithmica 31*, 1 (Dec. 2001), 1–28. Referenced on 33.

[37] Eiter, T., and Mannila, H. Computing discrete Fréchet distance. Tech. Rep. CD-TR 94/64, Technische Universität Wien, Vienna, Austria, Apr. 1994. Referenced on 2, 3, 26.

[38] Euclid. *The Thirteen Books of the Elements*, secondẽd., vol. 1. Dover, New York, 1956. Books I and II, unabridged. Referenced on 43.

[39] Fáry, I. Sur la courbure totale dune courbe gauche faisant un noeud. *Bull. Soc. Math. France 77* (1949), 128–138. Referenced on 1, 46.

[40] Fáry, I. Sur certaines inégalités géométriques. *Acta Sci. Math. (Szeged) 12*, aa (1950), 117–124. Referenced on 41.

[41] Fasy, B. T. Persistence diagrams and the heat equation homotopy. http://arxiv.org/abs/1002.1937, 2009. Research Initiation Project, Duke University Comput. Sci. Dept. Referenced on 64.

[42] Fasy, B. T. Discovering metrics and scale space, Jul. 2010. Preliminary Examination, Duke University Comput. Sci. Dept. Referenced on 34.

[43] Fasy, B. T. The difference of length in curves in $\mathbb{R}^n$. *Acta Sci. Math. (Szeged) 77* (2011), 359–367. Referenced on 3, 5, 41, 128.

[44] Fenchel, W. Über Krümmung und Windung geschlossener Raumkurven. *Math. Ann. 101*, 1 (Dec. 1929), 238–252. Referenced on 46.

[45] Fréchet, M. Sur quelques points du calcul fonctionnel. *Rend. Circ. Mat. Palermo (1884–1940) 22*, 1 (Dec. 1906), 1–72. Referenced on 21.

[46] Frosini, P. Discrete computation of size functions. *Journal of Combinatorics, Information and System Sciences 17*, 3-4 (1992), 232–250. Referenced on 15.

[47] Godau, M. A natural metric for curves–computing the distance for polygonal chains and approximation algorithms. pp. 127–136. 1991. Referenced on 25.

[48] Godau, M. *On the Complexity of Measuring the Similarity between Geometric Objects in Higher Dimensions*. PhD thesis, Freie Univ. Berlin, 1998. Referenced on 25.

[49] Gromov, M. Hyperbolic groups. In *Essays in Group Theory*, S. M. Gersten, Ed., vol. 8 of *Math. Sci. Res. Inst. Publ.* Springer-Verlag, 1987, pp. 75–263. Referenced on 12.

[50] Hatcher, A. *Algebraic Topology*. Cambridge Univ. P., 2002. Electronic Version. Referenced on 5, 13.

[51] Hopcroft, J. E., and Karp, R. M. An $n^{5/2}$ algorithm for maximum matchings in bipartite graphs. *SIAM J. Comput. 2*, 4 (1973), 225–231. Referenced on 33.

[52] Hummel, R. A., Kimia, B., and Zucker, S. W. Deblurring Gaussian blur. *Computer Vision, Graphics, and Image Processing 38* (1987), 66–80. Referenced on 67.

[53] Iijima, T. Basic theory on normalization of a pattern. *Bull. Electr. Lab. 26* (1962), 368–388. Referenced on 66, 71, 74.

[54] Koenderink, J. J. The structure of images. *Biol. Cybernet. 50*, 5 (Aug. 1984), 363–370. Referenced on 56, 66, 71, 74, 75.

[55] Korn, G. A., and Korn, T. M. *Mathematical Handbook for Scientists and Engineers: Definitions, Theorems, and Formulas for Reference and Review*. Dover Publ., 2000. Reprint. Referenced on 39.

[56] Kuhn, H. W. The hungarian method for the assignment problem. *Naval Res. Logist. Quarterly 2*, 1 (Mar. 1955), 83–97. Referenced on 32, 33.

[57] Kuijper, A., and Florack, L. M. J. The application of catastrophe theory to image analysis. Tech. Rep. UU-CS-2001-23, Dept. Comput. Sci. Utrecht Univ., Utrecht, the Netherlands, 2001. Referenced on 57.

[58] Kuijper, A., and Florack, L. M. J. The relevance of non-generic events in scale space models. *Internat. J. Comput. Vis. 57* (2004), 67–84. Referenced on 57.

[59] Lifshitz, L. M., and Pizer, S. M. A multiresolution hierarchical approach to image segmentation basedo n intensity extrema. *IEEE Trans. Pattern Anal. Mach. Intell. 12*, 6 (Jun. 1990). Referenced on 56, 69.

[60] Lindeberg, T. Scale-space for discrete signals. *IEEE Trans. Pattern Anal. Mach. Intell. 12*, 3 (Mar. 1990). Referenced on 71, 74.

[61] Lindeberg, T. *Scale-Space Theory in Computer Vision.* Kluwer Academic Publ., Dortrecht, the Netherlands, 1994. Referenced on 56.

[62] Milnor, J. *Morse Theory.* No. 51 in Ann. of Math. Stud. Princeton Univ. P., Princeton, 1963. Referenced on 11, 15, 45.

[63] Milosavljevic, N., Morozov, D., and Skraba, P. Zigzag persistent homology in matrix multiplication time. In *Proc. of the* 27^{th} *Annu. Symp. Comput. Geom.* (New York, Jun. 2011), ACM. Symposium held in Paris, France. Referenced on 20.

[64] Picard, É. *Traité d'analyse*, 2 ed. Gauthier-Villars et fils, Paris, 1908. Referenced on 59.

[65] Rieger, J. Generic evolutions of edges on families of diffused greyvalue surfaces. *J. Math. Imaging Vision 5* (1995), 207–217. Referenced on 56.

[66] Roberts, S. J. Parametric and non-parametric unsupervised cluster analysis. *Pattern Recognition 30*, 2 (1997), 261–272. Referenced on 56.

[67] Robins, V. Towards computing homology from finite approximations. In *Topology Proc.* (1999), vol. 24, pp. 503–532. Referenced on 15.

[68] Rote, G. Computing the Fréchet distance between piecewise smooth curves. *Comput. Geom. 37*, 3 (Aug. 2007), 162–174. Referenced on 25.

[69] Santaló, L. *Integral Geometry and Geometric Probability*. Addison-Wesley, England, 1976. Reprinted by Cambridge Univ. Press, England, 2004. Referenced on 47.

[70] Schilling, R. L. *Measures, Integrals and Martingales*. Cambridge Univ. P., New York, 2005. Referenced on 8.

[71] Schoenberg, I. J. On Póly frequency functions. ii: Variation-diminishing integral operators of the convolution type. *Acta Sci. Math. (Szeged) 12B* (1950), 97–106. Referenced on 75.

[72] Silverman, B. W. Using kernel density estimates to investigate multimodality. *J. R. Stat. Soc. Ser. B. Stat. Methodol. 43* (1981), 97–99. Referenced on 56, 75.

[73] Singer, D. A. *Geometry: Plane and Fancy*. Springer-Verlag, New York, NY, 1998. Referenced on 43.

[74] Sommerville, D. M. Y. *Introduction to the Geometry of N Dimensions*. Dover Publ., New York, NY, 1958. Reprint. Referenced on 3, 38.

[75] Sullivan, J. Curves of finite total curvature. In *Discrete Differential Geometry*, A. I. Bobenko, P. Shröder, J. M. Sullivan, and G. M. Ziegler, Eds., vol. 38 of *Oberwolfach Seminars*. Birkhäuser Verlag Basel / Switzerland, Boston, 2008, pp. 137–161. In Part II: Curvatures of Discrete Curves and Surfaces. Referenced on 3, 8.

[76] Tai, K.-C. The tree-to-tree correction problem. *J. ACM 26*, 3 (Jul. 1979), 422–433. Referenced on 34.

[77] Tanawongsuwan, R., and Bobick, A. F. Gait recognition from time-normalized joint-angle trajectories in the walking plane. pp. 726–731. 2001. Referenced on 30.

[78] Tikhonov, A. N., and Arsenin, V. Y. *Solutions of Ill-Posed Porblems*. Winston Publ., Washington, D. C. 1977. Referenced on 67.

[79] Vaidya, P. Geometry helps in matching. *SIAM J. Comput. 18*, 6 (1989), 1201–1225. Referenced on 32.

[80] Vairy, M., and Venkatesh, Y. Deblurring Gaussian blur using a wavelet transform. *Pattern Recognit. 28*, 7 (Jul. 1995), 965–976. Referenced on 67.

[81] Van Oostrum, R., and Veltkamp, R. C. Parametric search made practical. pp. 1–9. 2002. Referenced on 25.

[82] Vietoris, L. Über den höheren Zusammenhang kompakter Räume und eine Klasse von zusammenhangstreuen Abbildungen. *Math. Ann. 97* (1929), 454–472. Referenced on 12.

[83] Weickert, J., Ishikawa, S., and Imiya, A. On the history of Gaussian scale-space axiomatics. In *Gaussian Scale-Space Theory*, vol. 8 of *Computational Imaging and Vision*. Kluwer Academic Publ., Dortrecht, the Netherlands, 2010, pp. 45–59. Chapter 4. Referenced on 66, 74.

[84] Witkin, A. P. Scale-space filtering. In *Proc. of the Eighth Internat. Joint Conf. Artif. Intell.* (Aug. 1983), vol. 2, Morgan Kaufmann Publ., pp. 1019–1022. Conference held in Karlshuhe, West Germany. Referenced on 66.

[85] Witkin, A. P. Scale-space filtering. *Readings in Computer Vision: Issues, Problems, Principles, and Paradigms* (1987), 329–332. Referenced on 56, 66.

[86] Yuille, A., and Poggio, T. Scaling theorems for zero crossings. *IEEE Trans. Pattern Anal. Mach. Intell.* (1986), 15–25. Referenced on 56, 64, 71, 74, 75.

[87] Zomorodian, A., and Carlsson, G. Computing persistent homology. *Discrete Comput. Geom. 33*, 2 (2005), 249–274. Referenced on 13, 15.

Biography

Brittany Terese Fasy was born on 25 February 1985 in Philadelphia, PA. In May 2007, she earned the Bachelor of Science degrees in Mathematics and Computer Science from Saint Joseph's University of Philadelphia, PA. In June 2012, Brittany completed the Doctor of Philosophy degree in Computer Science from Duke University under the direction of Dr. Herbert Edelsbrunner.

In Fall 2009, Brittany moved to Klosterneuburg, Austria to work at IST Austria (Institute of Science and Technology Austria). In Austria, Brittany enjoyed meeting people from around the world and skiing in the Austrian Alps. In April 2011, David L. Millman proposed to Brittany on a ski trip in St. Anton. Their wedding is planned for 1 June 2013 and will be held in Philadelphia, PA. Brittany became a member of the Vienna Cherokees lacrosse team and even played a few games in the Austrian National lacrosse team.

At the time of writing, Brittany has published a single author journal paper "The difference in length in curves in $\mathbb{R}^n$" [43] and she has co-authored a paper that appeared in the proceedings of 27^{th} Annual Symposium on Computational Geometry 2012 [31].

For the first four years of her Graduate studies, Brittany was funded by the Graduate Aid in Areas of National Need (GAANN) fellowship. Her fifth year was funded by IST Austria.