EMBEDDING, APPROXIMATING, AND VISUALIZING PERSISTENCE MODULES

By

ALEXANDER Y. WAGNER

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2020

I dedicate this to all the teachers I have had throughout my life.

ACKNOWLEDGEMENTS

I thank my advisor Peter Bubenik for his tireless dedication to his students and for always insisting that principled solutions to meaningful problems are within our reach and worth our effort. I also thank the members of my committee, the University of Florida Informatics Institute for awarding me a UF Informatics Institute Graduate Student Fellowship, and the Hausdorff Institute for Mathematics for hosting me in September 2017.

TABLE OF CONTENTS

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

EMBEDDING, APPROXIMATING, AND VISUALIZING PERSISTENCE MODULES

By

Alexander Y. Wagner

May 2020

Chair: Peter Bubenik
Major: Mathematics

Persistent homology applies homology to a nested sequence of spaces to obtain a graded module called a persistence module. The stability of persistence modules with respect to changes in the input supports their use as a signature of the underlying space for statistics and machine learning. This stability is with respect to a family of metrics on persistence modules parametrized by a constant p between one and infinity. Since every metric in this family is incompatible with an inner product, a non-trivial feature map is necessary in order to use kernel methods. It is natural to ask how much such maps necessarily distort the metric on persistence diagrams. We show that when p is strictly greater than two, the associated metric space does not coarsely embed into any Hilbert space. The nerve theorem allows for the computation of the persistent homology of certain continuous filtered spaces. This computation has a complexity which is cubic in the number of simplices of the associated simplicial filtration and may be prohibitively slow. To mitigate this, we approximate the persistent homology up to some pre-specified error using discrete Morse theory and prove the approximation is optimal among a restricted set of nearby filtrations. Finally, we propose a general technique for extracting a larger set of stable information from persistent homology computations than is currently done. These computations also produce other information of great interest to practitioners that is unfortunately unstable. We recast this information as discontinuous real-valued functions and observe that convolving such a function with a suitable function produces a Lipschitz function. The resulting stable function can be estimated by perturbing the input and averaging the output.

CHAPTER 1
INTRODUCTION

1.1 Overview

Persistent homology takes in a one-parameter family of topological spaces and outputs a signature, called the persistence diagram or barcode, of this family's changing homology. Persistence diagrams are one of the main tools in topological data analysis (TDA) [13, 33, 38, 40]. In combination with machine learning and statistical techniques, they have been used in a wide variety of real-world applications, including the assessment of road network reconstruction [3], neuroscience [7, 24], vehicle tracking [5], object recognition [46], protein compressibility [39], and protein structure [43]. Put briefly, these persistence diagrams are multi-sets of points in the extended plane, and they compactly describe some of the multi-scale topological and geometric information present in a high-dimensional point cloud or carried by a real-valued function on a domain. There is a natural metric on one-parameter families of topological spaces, called the interleaving distance, and a family of metrics on persistence diagrams, called the p-Wasserstein distances. Several theorems [17, 25, 27] state that persistence diagrams equipped with these metrics are stable with respect to certain variations in the point cloud or functional input. This stability supports the use of persistent homology for machine learning because it guarantees that small perturbations of the data, such as those caused by measurement noise, do not cause large changes in the associated features.

Kernel methods, such as support vector machines or principal components analysis, are machine learning algorithms that require an inner product on the data [64]. When the original data set X lacks an inner product or one would like a higher-dimensional representation of the data, a standard approach is to map the data into a Hilbert space $\mathscr{H}$. Such a mapping is called a feature map and kernel methods are implicitly performed in the codomain of the feature map. While specifying an explicit feature map may be difficult, it turns out to be equivalent to the often simpler task of constructing a positive definite kernel on the data. This equivalence is important for the practical success of kernel methods but should not obscure the fact that there is an underlying feature map $\varphi : X \to \mathscr{H}$ and that the associated learning algorithm works with $\varphi(X) \subseteq \mathscr{H}$. Because of this, when X represents stable signatures of input data, one would like a feature map φ that changes the original metric as little as possible. If one would like to apply kernel methods to persistence diagrams, a natural first question is whether the metrics on persistence diagrams can be induced by an inner product. More precisely, does there exist an isometric embedding of persistence diagrams into a Hilbert space? We show in Section 2.4.1 that the impossibility of such an isometric embedding follows from work of Turner and Spreemann [66] and classical results of Schoenberg [59, 60]. In other words, any feature map from persistence diagrams into a Hilbert space necessarily distorts the original metric.

Our first main result concerns the ∞-Wasserstein distance, also called the bottleneck distance. Among the p-Wasserstein distances on persistence diagrams, this is the only case for which persistent homology is 1-Lipschitz. Isometric embeddings require distances to be exactly preserved. More general are bi-Lipschitz embeddings which are allowed to distort distances at most linearly. Considerably more general are coarse embeddings, which need not be continuous and only require that distances be distorted in a uniform, but potentially non-linear, way. Coarse embeddings are an important notion in geometric group theory and coarse geometry [41, 58]. We show that the space of persistence diagrams with the bottleneck distance does not admit a coarse embedding into any Hilbert space (Theorem 3-2). In other words, the distortion caused by a feature map to the bottleneck distance is not uniformly controllable. In fact, even if one restricts to the subspace of (finite) persistence diagrams arising as the homology of a filtered finite simplicial complex, there still does not exist a coarse embedding of this subspace into a Hilbert space (Remark 3-1 and Lemma 3-1). This result about distortions of embeddings is something that people working with persistence diagrams have noticed in practice. Philosophically, this is to be expected because bottleneck distance is an ℓ^∞-type distance, and ℓ^∞ can only be embedded in ℓ^2 with distortion growing with dimension. Our result makes such an argument rigorous. As corollaries of Theorem 3-2, we obtain the generalized roundness, negative type, and asymptotic dimension of persistence diagrams with the bottleneck distance (Corollary 3-1, Remark 3-2, and Corollary 3-3). Toward our proof of Theorem 3-2, we show that any separable, bounded metric space isometrically embeds into the space of persistence diagrams with the bottleneck distance

(Theorem 3-1). Our proof of Theorem 3-2 combines Theorem 3-1 with ideas of Dranishnikov et al. [32] and Enflo [34].

We subsequently extend this result to $p > 2$. We begin by showing in Proposition 3-2 that every finite subset of $(\mathbb{R}^d, \|\cdot\|_p)$ isometrically embeds into the space of persistence diagrams with the p-Wasserstein metric. Equipped with Proposition 3-2, we can embed any finite subset of ℓ_p into the space of persistence diagrams with arbitrarily small metric distortion. If we suppose a coarse embedding of the space of persistence diagrams exists when $p > 2$, the associated distortion functions can be modified to bound the distortion of embedding any finite subset of ℓ_p into a Hilbert space. This implies by Theorem 3.4 of Nowak [55] that ℓ_p coarsely embeds into a Hilbert space, contradicting Theorem 1 of Johnson and Randrianarivony [42].

The computation of persistent homology is based on the construction of a filtered cell complex and scales with matrix multiplication coefficient in the number of cells [48]. Discrete Morse theory reduces the number of cells in a complex without changing its homology. Mischaikow and Nanda [49] use filtration-wise discrete Morse reductions to speed up certain persistent homology computations and implemented their algorithm `MorseReduce` in the software Perseus [53]. For many steps of certain filtrations, such as Čech and alpha filtrations, there is only a single cell added to the complex. These cells cannot be reduced by Nanda and Mischaikow's approach, as there is no possible matching cell.

In a data analysis setting, we might not be interested in an exact knowledge of the persistence diagram. In Chapter 4, we present and compare three different approaches, summarized below, to trade off correctness of the persistence diagram for a higher number of cell reductions. In all cases, we begin with a complex X and filtration $f : X \to \mathbb{R}$. Given an error bound δ, we approximate the persistent homology of f within δ in the bottleneck distance, w_∞. For definitions of these terms, see Sections 2.1 and 2.2.

1. (Binning) Find an approximate filtration, then use `MorseReduce` to construct and reduce by a filtered acyclic partial matching based on the approximate filtration.

2. (Induced Filtration) Construct an unfiltered acyclic partial matching and use this additional

information to find an approximate filtration whose induced filtered matching preserves as many matches as possible.

3. (Gradient Paths) Adapt the search of `MorseReduce` such that gradient paths are allowed to grow up to a δ filtration difference between start and end points, then reduce by this filtered acyclic partial matching.

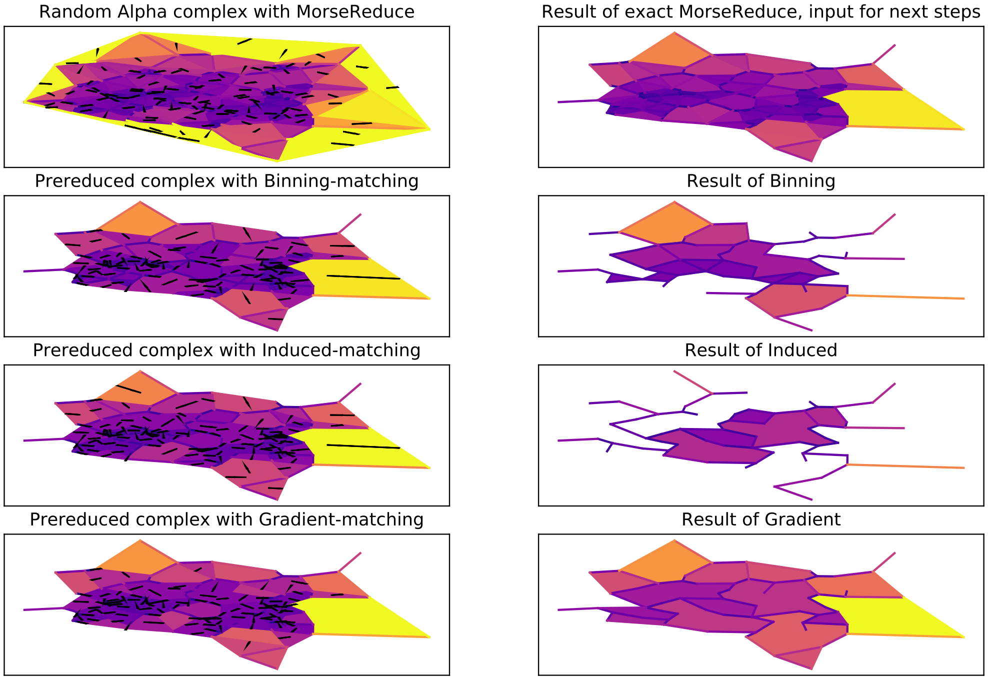
Random Alpha complex with MorseReduce
Result of exact MorseReduce, input for next steps
Prereduced complex with Binning-matching
Result of Binning
Prereduced complex with Induced-matching
Result of Induced
Prereduced complex with Gradient-matching
Result of Gradient

Figure 1-1. Reducing Alpha Complex by Three Methods. Alpha complex based on 100 standard normal samples from the plane, $\delta = 0.05$. Prereduced by `MorseReduce` and further reduced by three approximation algorithms. Lighter colors show higher filtration values; black lines are matches.

The first row of Figure 1-1 shows how an alpha complex constructed on 100 points drawn from the standard normal distribution on $\mathbb{R}^2$ is reduced with `MorseReduce` until it stabilizes. The lower rows show how the proposed approximation algorithms further reduce this complex. Figure 1-2 shows how this changes the persistence diagrams. The three algorithms considered here do not modify the appearance time of any simplex by more than a pre-specified error δ. This guarantees by Theorem 2-1 that the persistence diagram will not change by more than δ in the bottleneck distance.

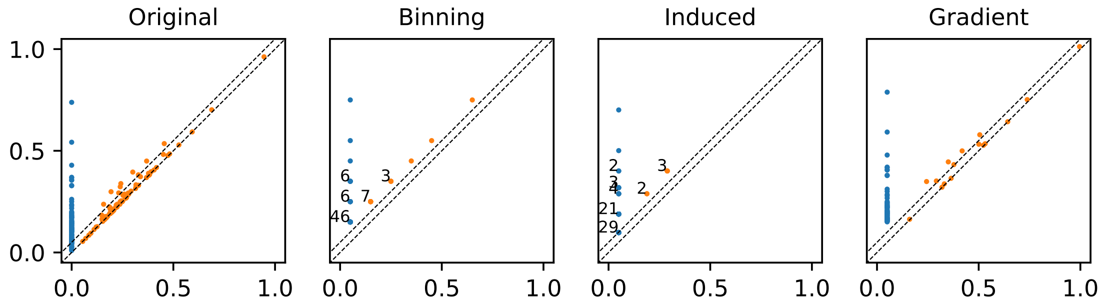

Figure 1-2. Approximate Persistence Diagrams. Persistence diagrams for the different algorithms with $\delta = 0.05$ band. Color shows homology degree, numbers show multiplicity greater than one.

There is additional, potentially very useful, but unstable information produced during the computation of persistence diagrams. For example, a point far from the diagonal in the degree-zero persistence diagram represents a connected component with high persistence. This component first appears somewhere and the computation that produces the persistence diagram can be used to find its location. However this location is not stable. As we will describe below, a small change in the input will cause only a small change in the persistence of this connected component, but it can radically alter the location of its birth. We summarize this as follows.

Definition 1-1 (Fundamental Conundrum of Topological Data Analysis)**.** *Users of topological data analysis would like to find the simplices or cycles corresponding to the birth of the most significant pairings of critical values. However, unlike the paired critical values, these simplices and cycles are unstable. In addition, persistent homology computations may rely on parameters such that the output persistence diagram is not stable with respect to changes of these parameters.*

In Chapter 5, we introduce a method for stabilizing desirable but unstable outputs of persistent homology computations. The main idea is the following. On the front end, we think of a persistent homology computation $\mathscr{C}$ as being parametrized by a vector $a = (a_1, \ldots, a_n)$ of real numbers. These parameters could specify the input to the computation, such as the coordinates of the vertices of a simplicial complex, or they could specify other values used in the computation, such as threshold parameters used in de-noising or bandwidths for smoothing. For a given choice of a, we get a persistence diagram. On the back end, we consider a function p that extracts a real-number summary from a persistence diagram. For example, p might extract the persistence of a homology class created by the addition of a specific edge in a filtered simplicial complex, or it might be an indicator function on whether or not the longest bar was born by the addition of a simplex contained in a fixed region of the input space, or it may indicate whether or not a chosen representative geometric cycle intersects a given region. The composite function h that maps the parameter vector to the real number need not be continuous, but it will in many cases be measurable. We convolve this function with a Gaussian, or indeed any Lipschitz function, to produce a new Lipschitz function that carries the persistence-based information we desire.

Our main theoretical results (Theorems 5-1, 5-2, and 5-3) give conditions on functions h and K, where K is a kernel, that guarantee that the convolution $h * K$ is Lipschitz with specified Lipschitz constant. From these we obtain the following theorem, where more precise statements are given as Corollaries 5-1, 5-2, and 5-3.

Theorem 1-1. *If h is locally essentially bounded then for the triangular and Epanechnikov kernels, $h * K$ is locally Lipschitz. If h is essentially bounded then for the Gaussian kernel, $h * K$ is Lipschitz.*

In practice, this can be translated to a simple procedure for stabilizing unstable persistent homology computations: perturb the input by adding, for example, Gaussian noise and redo the computation; repeat and average. By the law of large numbers, the result converges to the desired stable value.

Theorem 1-2. *Let $\varepsilon_1, \dots, \varepsilon_M$ be drawn independently from a kernel K. Then*

$$\frac{1}{M}\sum_{i=1}^{M} h(a-\varepsilon_i) \to (h*K)(a).$$

We summarize our computational pipeline in the following algorithm, also shown in Figure 1-3. Say we have performed a persistence computation and obtained an unstable output. For example, we have determined that the longest interval in the degree-one barcode of the Vietoris-Rips complex on points $X_1, \dots, X_N \in \mathbb{R}^d$ is born with the addition of the edge X_1X_2. We encode this output as a function $h : \mathbb{R}^n \to \mathbb{R}$ with input $a \in \mathbb{R}^n$. For example, the coordinates of the above points give us $a \in \mathbb{R}^n$ where $n = Nd$. We define the value of $h : \mathbb{R}^n \to \mathbb{R}$ to be the length of the longest interval in the barcode if it is born with the addition of the edge X_1X_2 and 0 otherwise. The choice of standard deviation σ, also called bandwidth, is discussed in Sections 5.4.1 and 5.4.2. In Section 5.3.6, we prove the algorithm in Figure 1-3 is stable with respect to this choice.

```
Input: $h : \mathbb{R}^n \to \mathbb{R}, a \in \mathbb{R}^n$
Parameters: $M \in \mathbb{N}, \sigma > 0$
  for $i \leftarrow 1, M$ do
      for $j \leftarrow 1, n$ do
          Sample $\varepsilon_j$ from $N(0, \sigma^2)$
      end for
      $y_i \leftarrow h(a + \varepsilon), \varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$
  end for
  return the average value of $y_1, \ldots, y_M$
```

Figure 1-3. Stabilizing Unstable Persistence Computations

1.2 Related Literature

Carrière and Bauer [16] have investigated bi-Lipschitz embeddings of persistence diagrams into separable Hilbert spaces. They have shown the impossibility of a bi-Lipschitz embedding into a finite-dimensional Hilbert space and that bi-Lipschitz embeddings into infinite-dimensional, separable Hilbert spaces only exist when restrictions are placed on the cardinality and spread of the persistence diagrams under consideration. Bell et al. [4] have shown that the space of persistence diagrams with the p-Wasserstein distance for $p < \infty$ has a discrete subspace that fails to have property A. The relevance of this result is that a discrete metric space with property A admits a coarse embedding into a Hilbert space [70]. Bubenik and Vergili [12] have shown that there exist cubes of arbitrary dimension with the ℓ^∞ distance which isometrically embed into the space of persistence diagrams with the bottleneck distance.

The standard persistent homology algorithm has a computational complexity that is cubic in the number of simplices. The fact that the number of simplices grows exponentially in the number of points for the Čech and Vietoris-Rips filtrations motivates the search for approximation algorithms. The seminal paper of Sheehy [61] addressed this problem by replacing the

Vietoris-Rips complex with a complex of size $O(n)$, where n is the number of points, that can be computed in $O(n\log(n))$ time. Classical persistent homology algorithms compute the persistent homology of a sequence of simplicial complexes with inclusions. Dey et al. [30] developed an algorithm to compute the persistent homology of a sequence of simplicial complexes connected by simplicial maps that may not be inclusions. Using this, they also provide an approximation algorithm for the Vietoris-Rips filtration similar to Sheehy.

Partial inspiration for the main idea in Chapter 5 comes from the trembling-hand equilibrium solution [51] to the non-uniqueness problem for Fréchet means of persistence diagrams. Our approach should also be compared with the topological reconstruction results of Niyogi, Smale, and Weinberger [54].

Several recent papers have advocated principled approaches for extracting features from persistence diagrams, including persistence landscapes [10], the stable multi-scale kernel [57], intensity functionals [23], persistence images [2], the stable topological signature [15], and the cover-tree entropy reduction [63]. Our result complements these ideas. Once one identifies some specific parts of the persistence diagram as having good classification power, one can then attempt to locate, in a robust way, the portions of the domain responsible for these parts. Other papers [1, 11, 18] have developed sophisticated schemes for data cleaning before persistent homology computation. These techniques are generally fragile to certain initial parameter choices, such as the m_0 parameter in Chazal et al. [18]. Again, we provide a complementary role. Any of these schemes can be run many times for several perturbations of an initial parameter choice, and the output can then be taken with confidence.

Dey and Wenger [31] have shown that the critical points of interval persistent homology are stable in the sense that they remain within some path-connected component. Zomorodian and Carlsson [71] use Mayer-Vietoris as inspiration in their technique for localizing (relative to a cover) homology classes within a given simplicial complex. However, this works only for a fixed simplicial complex, not a simplicial complex endowed with a filtration, and the results are certainly fragile to changes in this fixed complex. Weinberger [68] considers the sample complexity of some basic problems of topological inference. Specifically, he estimates the number of sample points necessary to determine the dimension, topological type, and to detect singularities for certain spaces.

Robust summaries of persistent homology are considered in the following papers; they do not consider the location of homology generators. Blumberg et al. [9] show that persistent homology on a metric measure space induces a stable empirical measure in the space of persistence diagrams. Taking the distance to a reference distribution or a reference barcode, they obtain robust statistics. Chazal et al. [19] derive limiting distributions and confidence sets for persistence diagrams based on the sub-level sets of the distance-to-a-measure.

Convolving with a kernel to obtain smoothness is a classical idea in statistics [62, 67]. It has been used to construct smooth estimators of discrete data as an initial step to computing persistent homology [10, 11, 35]. A related idea is to perform subsampling to obtain convergence results and confidence intervals for persistence diagrams and persistence landscapes [20, 21, 22, 35]. These papers use ideas related to ones presented here but to smooth initial data or to smooth stable outputs of persistence computations, not to stabilize unstable outputs of persistence computations.

CHAPTER 2
MATHEMATICAL BACKGROUND

2.1 Complexes and Persistence Homology

We use the following definition of complexes dating back to Tucker [65] and Lefschetz [44]. For concreteness, Chapter 5 deals with the special case of simplicial complexes. All homology groups are assumed to be computed over $\mathbb{Z}_2$. For a more thorough discussion of homology and persistence, we refer the reader to Munkres [52] and Oudot [56], respectively.

Definition 2-1. *Given a finite graded set* $X = \bigsqcup_k X_k$ *and a matrix* $\partial \in \mathbb{Z}_2^{X \times X}$ *we define a* complex (X, ∂) *to have the following properties:*

- *(Grading) For each a and a' in X,* $\partial[a,a'] \neq 0$ *implies* $\dim a = \dim a' + 1$.
- *(Boundary)* $\partial^2 = 0$.

An element $a \in X$ is called a *cell*, its grade $\dim a$ is called *dimension*. The *boundary matrix* ∂ induces a partial order $\preccurlyeq$ on X by the generating relation $a' \prec a \Leftrightarrow \partial[a,a'] \neq 0$. We frequently drop ∂ from the notation and refer to the complex simply as X. A subset $X' \subset X$ is a *subcomplex* of X if for all $a \in X'$ and $a' \preccurlyeq a$ it follows that $a' \in X'$. A subcomplex X' is a complex in its own right by restricting ∂ to X'. A *filtration* $f : X \to \mathbb{R}$ is a function such that $a' \preccurlyeq a$ implies $f(a') \leq f(a)$. Equivalently, f is a function such that $X^t := \{a \in X \mid f(a) \leq t\}$ is a subcomplex for every t. We refer to (X, ∂, f) as a *filtered complex*.

Let X be a complex, $f : X \to \mathbb{R}$ a filtration, and fix a homological dimension p. Suppose the distinct values of f are $r_1 < \ldots < r_m$. Whenever $i \leq j$, there is an inclusion $X^{r_i} \hookrightarrow X^{r_j}$, which induces a homomorphism

$$f_p^{i,j} : H_p(X^{r_i}) \to H_p(X^{r_j}).$$

A homology class $\alpha \in H_p(X^{r_i})$ is a *persistent homology class* that is *born* at level i if $\alpha \notin \operatorname{im} f_p^{i-1,i}$ and that *dies* entering level j if $f_p^{i,j}(\alpha) = 0$ but $f_p^{i,j-1}(\alpha) \neq 0$. If α never dies, we say that it dies entering level $j = \infty$ and $r_\infty = \infty$. The *persistence* of α is defined to be $\operatorname{pers}(\alpha) = r_j - r_i$. The set of classes which are born at i and die entering level j form a vector space, with rank denoted $\mu_p^{i,j}$. The degree-p *persistence diagram* of f, denoted $H_p(f)$, encodes these ranks. It is intuitively a multiset of points in the extended plane, with a point of multiplicity $\mu_p^{i,j}$ at each point (r_i, r_j).

2.2 The Space of Persistence Diagrams

In this section, we define persistence diagrams and a family of associated metric spaces. Persistence diagrams naturally arise as the output of persistent homology, which describes the changing homology of a one-parameter family of topological spaces. Persistence diagrams are usually defined to be multisets. We find it convenient to instead define them as indexed sets.

Definition 2-2. *Denote* $\{(x,y) \in \mathbb{R}^2 \mid x < y\}$ *by* $\mathbb{R}^2_<$. *A* persistence diagram *is a function from a countable set* I *to* $\mathbb{R}^2_<$, *i.e.* $D : I \to \mathbb{R}^2_<$.

To define the relevant metrics on persistence diagrams, we need two preliminary definitions.

Definition 2-3. *Suppose* $D_1 : I_1 \to \mathbb{R}^2_<$ *and* $D_2 : I_2 \to \mathbb{R}^2_<$ *are persistence diagrams. A* partial matching *between them is a triple* (I_1', I_2', f) *such that* $I_1' \subseteq I_1$, $I_2' \subseteq I_2$, *and* $f : I_1' \to I_2'$ *is a bijection.*

The p-Wasserstein distance between two persistence diagrams will be the minimal cost of a partial matching between them. More precisely, the p-cost of a partial matching is the ℓ^p norm of the sequence of ℓ^∞ distances between points paired by the partial matching and unpaired points with the diagonal in $\mathbb{R}^2$.

Definition 2-4. *Suppose* $D_1 : I_1 \to \mathbb{R}^2_<$ *and* $D_2 : I_2 \to \mathbb{R}^2_<$ *are persistence diagrams and* (I_1', I_2', f) *is a partial matching between them. Equip* $\mathbb{R}^2_<$ *with the norm* $\|a\|_\infty = \max(|a_x|, |a_y|)$. *The* p-cost *of* f *is denoted* $\text{cost}_p(f)$ *and defined as follows. If* $p < \infty$,

$$\text{cost}_p(f) = \left(\sum_{i \in I_1'} \|D_1(i) - D_2(f(i))\|_\infty^p + \sum_{i \in I_1 \setminus I_1'} \left(\frac{D_1(i)_y - D_1(i)_x}{2} \right)^p + \sum_{i \in I_2 \setminus I_2'} \left(\frac{D_2(i)_y - D_2(i)_x}{2} \right)^p \right)^{1/p}.$$

If $p = \infty$,

$$\text{cost}_\infty(f) = \max\left\{ \sup_{i \in I_1'} \|D_1(i) - D_2(f(i))\|_\infty, \sup_{i \in I_1 \setminus I_1'} \frac{D_1(i)_y - D_1(i)_x}{2}, \sup_{i \in I_2 \setminus I_2'} \frac{D_2(i)_y - D_2(i)_x}{2} \right\}.$$

If any of the terms in either expression are unbounded, we define the cost to be infinity.

Definition 2-5 ([25, 27])**.** *Let* $1 \le p \le \infty$. *If* D_1, D_2 *are persistence diagrams, define*

$$\tilde{w}_p(D_1, D_2) = \inf\{\text{cost}_p(f) \mid f \textit{ is a partial matching between } D_1 \textit{ and } D_2\}.$$

Let (Dgm_p, w_p) *denote the metric space of persistence diagrams* D *that satisfy* $\tilde{w}_p(D, \emptyset) < \infty$ *modulo the relation* $D_1 \sim D_2$ *if* $\tilde{w}_p(D_1, D_2) = 0$*, where* $\emptyset$ *is shorthand for the unique persistence diagram with empty indexing set. The metric* w_p *is called the* p-Wasserstein distance *and* w_∞ *is called the* bottleneck distance.

Note that the empty partial matching is the only one between $D : I \to \mathbb{R}^2_<$ and $\emptyset$. Hence, $\tilde{w}_p(D, \emptyset)$ is the ℓ_p norm of the sequence of distances between $\{D(i)\}_{i \in I}$ and the diagonal. In the following lemma, we prove that when $p < \infty$, two persistence diagrams D_1 and D_2 satisfying $\tilde{w}_p(D_1, \emptyset), \tilde{w}_p(D_2, \emptyset) < \infty$ belong to the same equivalence class if and only if they are equal up to a permutation of their indexing sets.

Lemma 2-1. *Let* $p < \infty$ *and suppose* $D_1 : I_1 \to \mathbb{R}^2_<$ *and* $D_2 : I_2 \to \mathbb{R}^2_<$ *are persistence diagrams satisfying* $\tilde{w}_p(D_1, \emptyset), \tilde{w}_p(D_2, \emptyset) < \infty$*. Then* D_1 *and* D_2 *satisfy* $\tilde{w}_p(D_1, D_2) = 0$ *iff there exists a bijection* $f : I_1 \to I_2$ *such that* $D_1(i) = D_2 f(i)$ *for every* $i \in I_1$.

Proof. If such a bijection exists, the partial matching (I_1, I_2, f) has zero p-cost. Conversely, suppose $\tilde{w}_p(D_1, D_2) = 0$. We will show that for any $p \in D_1(I_1)$, there exists a bijection f_p between $D_1^{-1}(p)$ and $D_2^{-1}(p)$ satisfying $D_1(i) = D_2 f_p(i)$ for any $i \in D_1^{-1}(p)$. Defining $f(i) = f_{D_1(i)}(i)$ then gives the desired bijection between I_1 and I_2.

Note that for $j = 1, 2$ the pre-image of any point in $\mathbb{R}^2_<$ under D_j must be finite since $\tilde{w}_p(D_j, \emptyset) < \infty$. For the same reason, $D_j(I_j)$ has no limit points in $\mathbb{R}^2_<$. Let $p \in D_1(I_1)$. Since p is not a limit point for either $D_1(I_1)$ or $D_2(I_2)$, there exists an $\varepsilon \in (0, (p_y - p_x)/2)$ such that $B(p, \varepsilon) \cap D_j(I_j) = \{p\}$. This implies that every partial matching between D_1 and D_2 has a cost of at least $|\text{card}(D_1^{-1}(p)) - \text{card}(D_2^{-1}(p))|\varepsilon$, which implies $\text{card}(D_1^{-1}(p)) = \text{card}(D_2^{-1}(p))$ since $\tilde{w}_p(D_1, D_2) = 0$. Letting f_p be an arbitrary bijection between $D_1^{-1}(p))$ and $D_2^{-1}(p)$ completes the proof. □

We end this section by recalling the diagram stability theorem which guarantees that persistence diagrams of nearby filtrations are close to one another in $(\text{Dgm}_\infty, w_\infty)$. An illustration of the theorem is given in Figure 2-1.

Theorem 2-1 ([17, 25]). *Suppose X is a complex and $f,g : X \to \mathbb{R}$ are filtrations. Let $H_p(f), H_p(g)$ be the degree-p persistence diagrams of f and g. Then $w_\infty(H_p(f), H_p(g)) \leq \|f-g\|_\infty$.*

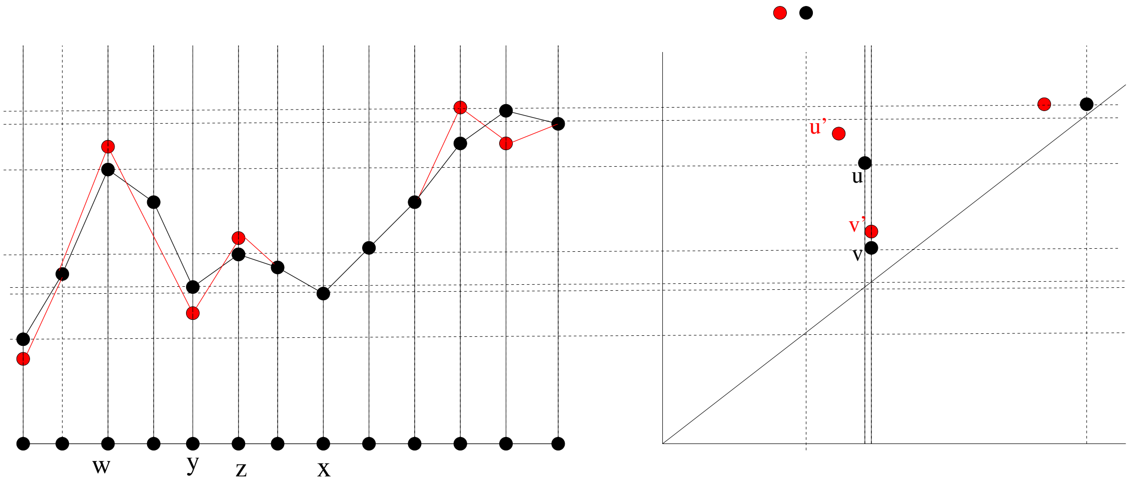

Figure 2-1. Stability Example. The graphs of functions f (black) and g (red), both on the same domain and the degree-0 persistence diagrams, $H_0(f)$ and $H_0(g)$, using the same color scheme.

2.3 Discrete Morse Theory

The definitions here are adapted from Mischaikow and Nanda [49].

Definition 2-6. *Let (X,∂) be a complex. A* partial matching *consists of a partition of X into sets A, Q, and K and a bijection $w : Q \to K$ such that $\partial[w(q), q] \neq 0$ for every $q \in Q$. Define $\ll$ on Q to be the transitive closure of the generating relation $q' \lhd q$ iff $q' \prec w(q)$. If the relation $\ll$ is a partial order, then the partial matching $(A, w : Q \to K)$ is said to be* acyclic.

We are interested in using discrete Morse theory to approximate the persistent homology of filtered complexes. As such, we now recall an extension of acyclic matchings to the filtered case.

Definition 2-7. *Let (X,∂,f) be a filtered complex. A* filtered acyclic partial matching *on (X,∂,f) is a family of acyclic partial matchings $(A_t, w_t : Q_t \to K_t)_{t\in\mathbb{R}}$ on $X^t := \{a \in X \mid f(a) \leq t\}$ such that $A_t \subset A_{t+s}$, $K_t \subset K_{t+s}$, $Q_t \subset Q_{t+s}$, and $w_t = w_{t+s}|_{Q_t}$ for all $s > 0$.*

We are now equipped to give the main theorem of discrete Morse theory due to Forman [36] and the extension by Mischaikow and Nanda [49] to the filtered setting.

Theorem 2-2 ([36, 49]). *Let (X, ∂) be a cell complex with an acyclic partial matching $(A, w : Q \to K)$. Then there exists an induced boundary map ∂^A such that $H_*(X, \partial) \cong H_*(A, \partial^A)$. We call the complex (A, ∂^A) the* Morse complex. *Moreover, if $(A_t, w_t : Q_t \to K_t)$ is a filtered acyclic partial matching on (X, ∂, f), then $H_*(f) \cong H_*(f|_A)$.*

2.4 Negative Type and Kernels

The following definition and theorem equate the problem of defining a feature map on a set to the frequently simpler problem of defining a positive definite kernel. Theorem 2-3 and the fact that kernel methods require access to only the inner products of elements is the content of the so-called kernel trick.

Definition 2-8. *Let X be a nonempty set. A symmetric function $k : X \times X \to \mathbb{R}$ is a* positive definite kernel *if for any $n \in \mathbb{N}$, $c_1, \ldots, c_n \in \mathbb{R}$, and $x_1, \ldots, x_n \in X$,*

$$\sum_{i,j=1}^{n} c_i c_j k(x_i, x_j) \geq 0.$$

Theorem 2-3 ([64]). *Let X be a nonempty set. A function $k : X \times X \to \mathbb{R}$ is a positive definite kernel iff there exists a Hilbert space $\mathscr{H}$ and a feature map $\varphi : X \to \mathscr{H}$ such that $\langle \varphi(x), \varphi(y) \rangle = k(x, y)$ for every $x, y \in X$.*

We now turn to the definition of negative type, which is closely related to positive definite kernels and to the embeddability of metric spaces into Hilbert spaces. Negative type played a central role in work of Schoenberg [59, 60] characterizing semi-metric spaces that admit an isometric embedding into a Hilbert space; see Theorems 2-4 and 2-5. Enflo [34] also implicitly used negative type to answer negatively the question of Smirnov on whether every separable metric space is uniformly homeomorphic to a subset of $L_2[0, 1]$. The equivalence between Enflo's notion of generalized roundness and the older notion of negative type was not proven until much later by Lennard et al. [45], giving a geometric characterization to the notion of negative type and, in particular, to the existence of isometric embeddings into Hilbert spaces. We refer the reader to Berg et al. [8] and Wells and Williams [69] for a more thorough treatment of the results referenced here. We remark that what we call a semi-metric space in the following definition is called a quasi-metric space in Wells and Williams [69].

Definition 2-9. *A* semi-metric space *is a nonempty set X together with a function $d : X \times X \to [0,\infty)$ such that $d(x,x) = 0$ and $d(x,y) = d(y,x)$ for every $x,y \in X$.*

Definition 2-10. *Let $q \geq 0$. A semi-metric space (X,d) is said to be of q*-negative type *if for any $n \in \mathbb{N}$, $x_1,\dots,x_n \in X$, and $a_1,\dots,a_n \in \mathbb{R}$ satisfying $\sum_{i=1}^{n} a_i = 0$, the following inequality is satisfied.*

$$\sum_{i,j=1}^{n} a_i a_j d(x_i,x_j)^q \leq 0$$

We define the negative type of a semi-metric space (X,d) to be the supremum of the set of $q \in [0,\infty)$ such that (X,d) is of q-negative type.

A relationship between positive definite kernels and negative type is given in the following.

Theorem 2-4 ([8])**.** *Let (X,d) be a semi-metric space. The following are equivalent.*

1. *(X,d) is of 1-negative type.*

2. *For any $x_0 \in X$, $k(x,y) = d(x,x_0) + d(y,x_0) - d(x,y)$ is a positive definite kernel.*

3. *$k(x,y) = e^{-td(x,y)}$ is a positive definite kernel for every $t > 0$.*

The negative type of a semi-metric space is closely related to questions regarding its embeddability into Hilbert spaces. An isometric embedding of a semi-metric space (X,d) into a Hilbert space $\mathscr{H}$ is a map $\varphi : X \to \mathscr{H}$ satisfying $d(x,y) = \|\varphi(x) - \varphi(y)\|_{\mathscr{H}}$ for every $x,y \in X$.

Theorem 2-5 ([69])**.** *A semi-metric space admits an isometric embedding into a Hilbert space iff it is of 2-negative type.*

Besides 2-negative type characterizing isometric embeddability into a Hilbert space, the following theorem states the important property that negative type is downward closed.

Theorem 2-6 ([69])**.** *Suppose (X,d) is a semi-metric space of q-negative type. Then it is of q'-negative type for any $0 \leq q' \leq q$.*

2.4.1 Isometric Embeddability of Diagram Space

It was shown by Turner and Spreemann [66] that (Dgm_p, w_p) is not of 1-negative type for any $1 \le p \le \infty$. This leads to the following negative result.

Theorem 2-7. *(Dgm_p, w_p) does not admit an isometric embedding into a Hilbert space for any $1 \le p \le \infty$.*

Proof. Let $1 \le p \le \infty$. Since (Dgm_p, w_p) is not of 1-negative type, by Theorem 2-6, (Dgm_p, w_p) is not of 2-negative type and so does not admit an isometric embedding into a Hilbert space by Theorem 2-5. □

2.4.2 Coarse Embeddings and Related Notions

If instead of demanding that distances be exactly preserved, we only require that distances be contracted or expanded a uniform amount, we arrive at the following definition.

Definition 2-11. *A map $f : (X, d) \to (Y, d')$ is a* coarse embedding *or* uniform embedding *if there exists non-decreasing $\rho_-, \rho_+ : [0, \infty) \to [0, \infty)$ such that*

1. *$\rho_-(d(x,y)) \le d'(f(x), f(y)) \le \rho_+(d(x,y))$ for all $x, y \in X$, and*
2. *$\lim_{t\to\infty} \rho_-(t) = \infty$.*

Note that if $\rho_-(x) = Ax$ and $\rho_+(x) = Bx$ for some $0 < A \le B$ then f is a bi-Lipschitz embedding. This definition was introduced by Gromov [41] where he posed the question of whether every separable metric space, of which (Dgm_p, w_p) are examples [9, 47], admits a coarse embedding into a Hilbert space. This question was answered negatively by Dranishnikov et al. [32].

Nowak [55] proved that a metric space can be coarsely embedded into a Hilbert space if and only if every finite subset can be embedded in ℓ_2 with the same distortion functions. Johnson and Randrianarivony [42] subsequently gave a sufficient condition for a metric space to not admit a coarse embedding into a Hilbert space; this condition is satisfied in particular by ℓ_p for $p > 2$.

Theorem 2-8 ([55])**.** *A metric space (X,d) admits a coarse embedding into a Hilbert space if and only if there exist non-decreasing functions $\rho_-, \rho_+ : [0,\infty) \to [0,\infty)$ such that $\lim_{t\to\infty} \rho_-(t) = \infty$ and for every finite subset $A \subseteq X$ there exists a map $f_A : A \to \ell_2$ satisfying*

$$\rho_-(d(x,y)) \leq \|f_A(x) - f_A(y)\|_2 \leq \rho_+(d(x,y))$$

for every $x,y \in A$.

Definition 2-12. *A basis $(e_n)_n$ for a Banach space X is a normalized symmetric basis if*

$$\|\sum_n \theta_n a_n e_{\sigma(n)}\| = \|\sum_n a_n e_n\|$$

for any choices of signs $\theta_n \in \{-1,+1\}$, permutation $\sigma : \mathbb{N} \to \mathbb{N}$, and $\sum_n a_n e_n \in X$.

Theorem 2-9 ([42])**.** *Suppose that a Banach space X has a normalized symmetric basis $(e_n)_n$ and that $\liminf_{n\to\infty} n^{-\frac{1}{2}} \left\| \sum_{i=1}^{n} e_i \right\| = 0$. Then X does not coarsely embed into a Hilbert space.*

Corollary 2-1 ([42])**.** *The space ℓ_p does not admit a coarse embedding into a Hilbert space for $p > 2$.*

The following definition gives a coarse analogue of covering dimension.

Definition 2-13. *Let n be a non-negative integer. A metric space (X,d) has* asymptotic dimension *$\leq n$ if for every $R > 0$ there exists a cover $\mathscr{U}$ of X such that every ball of radius R intersects at most $n+1$ elements of $\mathscr{U}$ and $\sup_{U\in\mathscr{U}} \sup\{d(x,y) \mid x,y \in U\} < \infty$.*

Theorem 2-10 ([58])**.** *If X is a metric space with finite asymptotic dimension, then there exists a coarse embedding of X into a Hilbert space.*

Property A is a simple condition for discrete metric spaces that also implies coarse embeddability into a Hilbert space.

Definition 2-14 ([70])**.** *A discrete metric space (X,d) has* property A *if for any $r > 0$, $\varepsilon > 0$ there is a family of finite subsets $\{A_x\}_{x\in X}$ of $X \times \mathbb{N}$ such that*

1. *$(x,1) \in A_x$ for all $x \in X$;*
2. *$\dfrac{|(A_x \setminus A_y)| + |(A_y \setminus A_x)|}{|A_x \cap A_y|} < \varepsilon$ whenever $d(x,y) \leq r$;*
3. *there exists $R > 0$ such that if $(x,m),(y,m) \in A_z$ for some $z \in X$, then $d(x,y) \leq R$.*

Theorem 2-11 ([70]). *If a discrete metric space X has property A, then X admits a coarse embedding into a Hilbert space.*

The following definition was introduced by Enflo [34] to answer negatively a question of Smirnov about uniform homeomorphisms into $L_2[0,1]$. Indeed, the negative answer to Gromov's question by Dranishnikov et al. was inspired by Enflo's negative answer to Smirnov's.

Definition 2-15. *Let $q \geq 0$. A metric space (X,d) has* generalized roundness q *if for any $n \in \mathbb{N}$ and $a_1,\ldots,a_n,b_1,\ldots,b_n \in X$, we have*

$$\sum_{i<j}(d(a_i,a_j)^q + d(b_i,b_j)^q) \leq \sum_{i,j} d(a_i,b_j)^q$$

Similarly to negative type, we define the generalized roundness of a metric space (X,d) to be the supremum of the set of $q \in [0,\infty)$ such that (X,d) has generalized roundness q.

CHAPTER 3
COARSE EMBEDDINGS OF PERSISTENCE DIAGRAMS INTO HILBERT SPACES

3.1 Coarse Embeddability of Persistence Diagrams with Bottleneck Distance

The main result of this section is that there does not exist a coarse embedding of $(\mathrm{Dgm}_\infty, w_\infty)$ into a Hilbert space. This implies that the generalized roundness and asymptotic dimension of $(\mathrm{Dgm}_\infty, w_\infty)$ are 0 and ∞, respectively. We also show that any separable, bounded metric space has an isometric embedding into the space of persistence diagrams with the bottleneck distance. The isometric embedding in question can be thought of as a shifted version of the Kuratowski embedding.

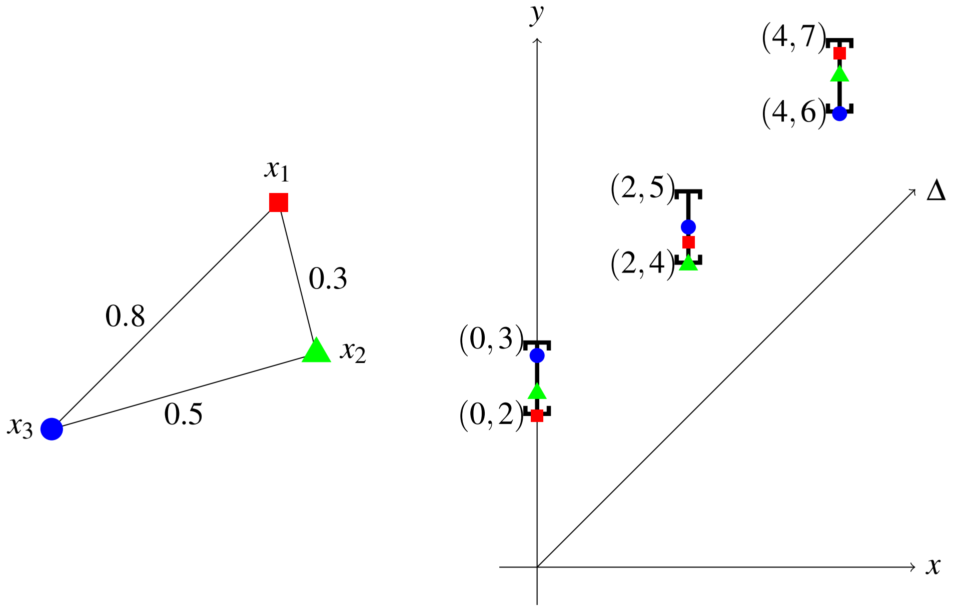

Figure 3-1. Example of Embedding into Persistence Diagrams with Bottleneck Metric. A metric space with three points and its image in $(\mathrm{Dgm}_\infty, w_\infty)$ under the map defined in Theorem 3-1 for $c = 1$, $x_i \mapsto \{(2(k-1), 2k + d(x_i, x_k)\}_{k=1}^{3}$.

Theorem 3-1. *Suppose (X,d) is a separable, bounded metric space. Then there exists an isometric embedding $\varphi : (X,d) \to (Dgm_\infty, w_\infty)$. Moreover, if $c > \sup\{d(x,y) \mid x,y \in X\}$, we may choose φ such that $\varphi(X) \subseteq B(\emptyset, \frac{3c}{2}) \setminus B(\emptyset, c)$, where $B(\emptyset, r) = \{D \in Dgm_\infty \mid w_\infty(D, \emptyset) < r\}$.*

Proof. Let $c > \sup\{d(x,y) \mid x,y \in X\}$. Let $\{x_k\}_{k=1}^{\infty}$ be a countable, dense subset of (X,d). Consider the following map.

$$\varphi : (X,d) \to (\mathrm{Dgm}_\infty, w_\infty)$$
$$x \mapsto \{(2c(k-1), 2ck + d(x,x_k)\}_{k=1}^{\infty}$$

Note that for any $x \in X$ and $k \in \mathbb{N}$,

$$d_\infty((2c(k-1), 2ck + d(x,x_k)), \Delta) = c + \frac{d(x,x_k)}{2} < \frac{3c}{2},$$

so $\tilde{w}_\infty(\varphi(x), \emptyset) < \infty$ for every $x \in X$ and thus φ is well-defined. Moreover, since

$$w_\infty(\varphi(x), \emptyset) = \sup_{1 \le k < \infty} d_\infty((2c(k-1), 2ck + d(x,x_k)), \Delta),$$

it follows that $\varphi(x) \in B(\emptyset, \frac{3c}{2}) \setminus B(\emptyset, c)$. A visualization of the image of φ for a metric space with three points is shown in Figure 3-1. We now show that for $y \in X$ an optimal partial matching of $\varphi(x)$ and $\varphi(y)$ matches points in each diagram with the same first coordinate, and the cost of this partial matching is $d(x,y)$.

For the equivalence classes $\varphi(x)$ and $\varphi(y)$, choose representative persistence diagrams $D_x : \mathbb{N} \to \mathbb{R}^2_<$ and $D_y : \mathbb{N} \to \mathbb{R}^2_<$. Consider the partial matching $(\mathbb{N}, \mathbb{N}, \mathrm{id}_\mathbb{N})$ between D_x and D_y, i.e. $(2c(k-1), 2ck + d(x,x_k))$ is matched with $(2c(k-1), 2ck + d(y,x_k))$ for every $k \in \mathbb{N}$. Observe that $d_\infty(D_x(k), D_y(k)) = |d(x,x_k) - d(y,x_k)|$ for every k, so the cost of this partial matching is $\sup_k |d(x,x_k) - d(y,x_k)|$. By the triangle inequality,

$$\sup_k |d(x,x_k) - d(y,x_k)| \le d(x,y).$$

Since $\{x_k\}_{k=1}^\infty$ is dense, for every $\varepsilon > 0$, there exists a k such that $d(x,x_k) < \varepsilon$, so

$$|d(x,x_k) - d(y,x_k)| \ge d(y,x_k) - d(x,x_k) \ge d(x,y) - 2d(x,x_k) > d(x,y) - 2\varepsilon.$$

This implies that $\sup_k |d(x,x_k) - d(y,x_k)| \ge d(x,y)$ and $\mathrm{cost}_\infty(\mathrm{id}_\mathbb{N}) = d(x,y)$.

We will now prove that the partial matching described above is optimal. Suppose $I, J \subseteq \mathbb{N}$ and (I, J, f) is a different partial matching between D_x and D_y. Then there exists a $k \in \mathbb{N}$ such that either $k \notin I$ or $k \in I$ and $f(k) \neq k$. If $k \notin I$, then

$$\mathrm{cost}_\infty(f) \ge d_\infty((2c(k-1), 2ck + d(x,x_k)), \Delta) \ge c.$$

If $k \in I$ and $f(k) = k' \neq k$, then

$$\mathrm{cost}_\infty(f) \geq \|(2c(k-1), 2ck + d(x,x_k)) - (2c(k'-1), 2ck' + d(y,x_{k'}))\|_\infty \geq 2c.$$

Therefore, $\mathrm{cost}_\infty(f) \geq c > d(x,y)$. Hence, $w_\infty(\varphi(x),\varphi(y)) = d(x,y)$, i.e. φ is an isometric embedding. □

We now apply Theorem 3-1 to show the generalized roundness of $(\mathrm{Dgm}_\infty, w_\infty)$ is 0. To do so, we embed a family of finite metric spaces, whose generalized roundness was observed by Enflo [34] to converge to 0, into $(\mathrm{Dgm}_\infty, w_\infty)$. One element of this family is shown in Figure 3-2.

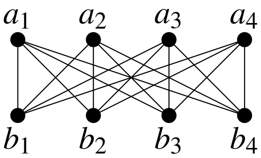

Figure 3-2. Complete Bipartite Graph $K_{n,n}$ for $n = 4$

Corollary 3-1. *The generalized roundness of* (Dgm_∞, w_∞) *is zero.*

Proof. Let $n \geq 2$. Define $K_{n,n} = \{a_1,\ldots,a_n,b_1,\ldots,b_n\}$ and equip this set with the metric $d(a_i,a_j) = d(b_i,b_j) = 2$ for any $i,j \in \{1,\ldots,n\}$ with $i \neq j$ and $d(a_i,b_j) = 1$. Enflo [34] remarks that X_n has generalized roundness that converges to 0 as $n \to \infty$. Indeed,

$$\sum_{i<j}(d(a_i,a_j)^q + d(b_i,b_j)^q) \leq \sum_{i,j} d(a_i,b_j)^q \iff$$
$$n(n-1)2^q \leq n^2 \iff$$
$$q \leq \log_2(1+(n-1)^{-1}).$$

Hence, X_n has generalized roundness at most $\log_2(1+(n-1)^{-1})$ which tends to 0 as n increases. By Theorem 3-1, we may isometrically embed X_n into $(\mathrm{Dgm}_\infty, w_\infty)$ for any n so the generalized roundness of $(\mathrm{Dgm}_\infty, w_\infty)$ must be zero. □

Our next result is that $(\mathrm{Dgm}_\infty, w_\infty)$ does not admit a coarse embedding into a Hilbert space. The proof relies on a construction of Dranishnikov et al. [32] based on ideas of Enflo [34].

Theorem 3-2. *(Dgm_∞, w_∞) does not admit a coarse embedding into a Hilbert space.*

Proof. Define $\mathbb{Z}_n$ to be the integers mod n with d_n, the metric induced by the standard metric $d(x,y) = |x-y|$ on $\mathbb{Z}$. Define $\mathbb{Z}_n^m$ to be the Cartesian product of m copies of $\mathbb{Z}_n$ with the following metric,

$$d_{n,m}(([k_1],\dots,[k_m]),([l_1],\dots,[l_m])) = \max_{1\le i\le m} d_n([k_i],[l_i]).$$

Let X be the disjoint union of $\mathbb{Z}_n^m$ for every $n,m \geq 1$ and suppose $\tilde{d}$ is a metric on X satisfying the following.

(1) The restriction of $\tilde{d}$ to each $\mathbb{Z}_n^m$ coincides with $d_{n,m}$.

(2) $\tilde{d}(x,y) \geq n+m+n'+m'$ if $x \in \mathbb{Z}_n^m$, $y \in \mathbb{Z}_{n'}^{m'}$, and $(n,m) \neq (n',m')$.

Proposition 6.3 of Dranishnikov et al. [32] shows that any such $(X,\tilde{d})$ does not admit a coarse embedding into a Hilbert space. Hence, it suffices to construct such an $(X,\tilde{d})$ and an isometric embedding of it into $(\mathrm{Dgm}_\infty, w_\infty)$, since a coarse embedding of $(\mathrm{Dgm}_\infty, w_\infty)$ into a Hilbert space would restrict to a coarse embedding of $(X,\tilde{d})$ into a Hilbert space.

Choose an enumeration $\{(n_i,m_i)\}_{i=1}^\infty$ of $\mathbb{N}\times\mathbb{N}$ such that $i<j$ implies $n_i+m_i \leq n_j+m_j$, for instance, $(1,1),(1,2),(2,1),(1,3),(2,2),(3,1),(1,4)$, etc. Define $c_1 = 1$ and for $i \geq 2$, $c_i = 4\max(c_{i-1}, n_i+m_i)$. For every (n_i,m_i), note that $c_i > n_i > \max\{d_{n_i,m_i}(x,y) \mid x,y \in \mathbb{Z}_{n_i}^{m_i}\}$. So by Theorem 3-1, there exists an isometry $\varphi_i : \mathbb{Z}_{n_i}^{m_i} \to (\mathrm{Dgm}_\infty, w_\infty)$ such that $\varphi_i(\mathbb{Z}_{n_i}^{m_i}) \subseteq B(\emptyset, \frac{3c_i}{2}) \setminus B(\emptyset, c_i)$.

Define $\varphi : X \to \mathrm{Dgm}_\infty$ by $\varphi(x) = \varphi_i(x)$ for $x \in \mathbb{Z}_{n_i}^{m_i}$ and define $\tilde{d}(x,y) = w_\infty(\varphi(x),\varphi(y))$ for any $x,y \in X$. By the definition of $\tilde{d}$, $\varphi : (X,\tilde{d}) \to (\mathrm{Dgm}_\infty, w_\infty)$ is an isometry. If $x,y \in \mathbb{Z}_{n_i}^{m_i}$, then $\tilde{d}(x,y) = w_\infty(\varphi_i(x),\varphi_i(y)) = d_{n_i,m_i}(x,y)$ so $\tilde{d}$ satisfies (1) above. It only remains to show $\tilde{d}$ satisfies (2).

Suppose $x \in \mathbb{Z}_{n_i}^{m_i}$, $y \in \mathbb{Z}_{n_j}^{m_j}$, and $(n_i,m_i) \neq (n_j,m_j)$. We may assume $i<j$. By construction, $\varphi(x) = \varphi_i(x) \in B(\emptyset, \frac{3c_i}{2}) \setminus B(\emptyset, c_i)$ and $\varphi(y) = \varphi_j(y) \in B(\emptyset, \frac{3c_j}{2}) \setminus B(\emptyset, c_j)$, which implies by the

triangle inequality for w_∞ that

$$\tilde{d}(x,y) = w_\infty(\varphi(x),\varphi(y)) \geq w_\infty(\varphi(y),\emptyset) - w_\infty(\varphi(x),\emptyset) > c_j - \frac{3c_i}{2}.$$

Additionally, we have $n_i + m_i \leq n_j + m_j$ and $c_j \geq 4\max(c_i, n_j + m_j) \geq 2(c_i + (n_j + m_j))$, so

$$\tilde{d}(x,y) > c_j - \frac{3c_i}{2} \geq 2(n_j + m_j) + 2c_i - \frac{3c_i}{2} > n_i + m_i + n_j + m_j.$$

We have shown that $\tilde{d}$ satisfies (2) which completes the proof. □

Remark 3-1. *For a finite metric space, the isometric embedding defined in Theorem 3-1 sends each point to a persistence diagram of finite cardinality in (Dgm_∞, w_∞). In particular, the map $\varphi_i : \mathbb{Z}_{n_i}^{m_i} \to (Dgm_\infty, w_\infty)$ given in the proof of Theorem 3-1 has an image consisting of finite persistence diagrams. Since X is the disjoint union of $\mathbb{Z}_n^m$ for every $n, m \geq 1$, it follows that $\varphi : (X, \tilde{d}) \to (Dgm_\infty, w_\infty)$ sends each point in the metric space X to a finite persistence diagram. Hence, the proof of Theorem 3-2 gives the slightly stronger result that the space of finite persistence diagrams with the bottleneck distance does not admit a coarse embedding into a Hilbert space.*

Theorem 3-2 and Remark 3-1 give the impossibility of coarsely embedding the space of finite persistence diagrams with the bottleneck distance into a Hilbert space. The primary motivation for this result was the application of kernel methods to persistent homology. In computational settings, the persistence diagrams of interest are frequently the result of applying homology to a filtered finite simplicial complex. We refer the interested reader to Oudot [56]. Hence, one may ask whether this more restricted space of persistence diagrams, i.e. the subspace arising from homology of filtered finite simplicial complexes, admits a coarse embedding into a Hilbert space. Unfortunately, this is easily seen to be false by the following.

Lemma 3-1. *Every finite persistence diagram is realizable as the persistent homology of a filtered finite simplicial complex.*

Proof. Suppose $D : \{1,\dots,n\} \to \mathbb{R}^2_<$ is a persistence diagram and define $(x_i, y_i) = D(i)$. Let V be the set $\{a_i, b_i, c_i\}_{i=1}^n$. Consider the simplicial complex on V that is the disjoint union of n 2-simplices and has the filtration given by assigning the value x_i to $\{a_i\}, \{b_i\}, \{c_i\}, \{a_i, b_i\}, \{a_i, c_i\}, \{b_i, c_i\}$ and the value y_i to $\{a_i, b_i, c_i\}$. Applying the simplicial homology functor $H_1(-, \mathbb{Z}_2)$ recovers the persistence diagram D. To see this, note that the persistent homology of the filtration on V is the direct sum of the persistent homology of the filtration on each individual triangle since the triangles are mutually disjoint. For triangle i, the 1-skeleton appears at time x_i giving rise to one degree-1 homology generator that vanishes when the 2-simplex is added at time y_i. □

Property A is a concrete condition satisfiable by a discrete metric space that implies it can be coarsely embedded into a Hilbert space. In the following proposition, we show that a semi-metric space being of q-negative type for some positive q is similarly a concrete condition that implies coarse embeddability into a Hilbert space.

Proposition 3-1. *A semi-metric space (X, d) of q-negative type for some $q > 0$ admits a coarse embedding into a Hilbert space.*

Proof. Suppose there exists a $q > 0$ such that the metric space (X, d) has q-negative type. Define $f(t) = t^{q/2}$ and observe that $fd(x,x) = 0^{q/2} = 0$ and $fd(x,y) = fd(y,x)$ so (X, fd) is a semi-metric space. Let $x_1, \dots, x_n \in X$ and $a_1, \dots, a_n \in \mathbb{R}$ such that $\sum_{i=1}^n a_i = 0$. Then

$$\sum_{i,j=1}^{n} a_i a_j (fd(x_i, x_j))^2 = \sum_{i,j=1}^{n} a_i a_j d(x_i, x_j)^q \le 0,$$

so (X, fd) is a semi-metric space of 2-negative type. By Theorem 2-5, there exists an isometric embedding φ from (X, fd) into a Hilbert space $\mathscr{H}$. Define $\rho_+ = \rho_- = f$. It follows that φ satisfies the requirements of a coarse embedding of (X, d) into $\mathscr{H}$, i.e.

$$\rho_+(d(x,y)) = \rho_-(d(x,y)) = fd(x,y) = \|\varphi(x) - \varphi(y)\|_{\mathscr{H}}. \qquad \square$$

Remark 3-2. *Since (Dgm_∞, w_∞) does not admit a coarse embedding into a Hilbert space, Proposition 3-1 implies that (Dgm_∞, w_∞) is of 0-negative type. This also follows from Corollary 3-1 and the result of Lennard et al. [45] on the equivalence of negative type and generalized roundness. Finally, we state two corollaries of Theorem 3-2 that answer Questions 3.10 and 3.11 of Bell et al. [4].*

Corollary 3-2. *(Dgm_∞, w_∞) contains a discrete subspace that fails to have property A.*

Proof. In the proof of Theorem 3-2, we consider a discrete metric space $(X, \tilde{d})$ and prove it embeds in $(\mathrm{Dgm}_\infty, w_\infty)$ via an isometry φ. Dranishnikov et al. [32] have shown that $(X, \tilde{d})$ does not admit a coarse embedding into a Hilbert space so by Theorem 2-11, $\varphi(X)$ fails to have property A. □

Corollary 3-3. *(Dgm_∞, w_∞) has infinite asymptotic dimension.*

Proof. If $(\mathrm{Dgm}_\infty, w_\infty)$ had finite asymptotic dimension, then it would admit a coarse embedding into a Hilbert space by Theorem 2-10, which contradicts Theorem 3-2. □

3.2 Coarse Embeddability of Persistence Diagrams with $p > 2$ Wasserstein Metric

In this section, we will show (Dgm_p, w_p) does not coarsely embed into a Hilbert space for $p > 2$. As in the case when $p = \infty$, the proof relies on the fact that w_p is the infimum of the ℓ_p norm over all partial matchings.

Proposition 3-2. *Let $d \in \mathbb{N}$. Every finite subset of $(\mathbb{R}^d, \|\cdot\|_p)$ isometrically embeds into (Dgm_p, w_p).*

Proof. Let $A = \{a^1, \ldots, a^n\}$ be a finite subset of $\mathbb{R}^d$. Let $c > \max\{\|a^i\|_\infty, \|a^i - a^j\|_p \mid 1 \le i, j \le n\}$ and consider the following map.

$$\varphi : A \to (\mathrm{Dgm}_p, w_p)$$
$$a^i \mapsto D_i : [d] \to \mathbb{R}^2_<,\ k \mapsto \{(2c(k-1), 2c(k+1) + a^i_k)\}_{k=1}^d$$

Formally, $\varphi(a^i)$ and $\varphi(a^j)$ are equivalence classes of persistence diagrams $D_i : [d] \to \mathbb{R}^2_<$ and $D_j : [d] \to \mathbb{R}^2_<$. Consider the partial matching $([d],[d],\mathrm{id}_{[d]})$ between D_i and D_j, i.e. $(2c(k-1), 2c(k+1)+a^i_k)$ is matched with $(2c(k-1), 2c(k+1)+a^j_k)$ for every $k \in [d]$. Observe that $\|D_i(k) - D_j(k)\|_\infty = |a^i_k - a^j_k|$ for every k, so the cost of this partial matching is $(\sum\limits_{k=1}^{d} |a^i_k - a^j_k|^p)^{\frac{1}{p}} = \|a^i - a^j\|_p$.

Suppose $I, J \subseteq [d]$ and (I, J, f) is a different partial matching between D_i and D_j. Then there exists a $k \in [d]$ such that either $k \notin I$ or $k \in I$ and $f(k) \neq k$. If $k \notin I$, then

$$\mathrm{cost}_p(f) \geq \frac{D_i(k)_y - D_i(k)_x}{2} = 2c + \frac{a^i_k}{2} > \frac{3c}{2}.$$

If $k \in I$ and $f(k) = k' \neq k$, then

$$\mathrm{cost}_p(f) \geq \|(2c(k-1), 2ck + a^i_k) - (2c(k'-1), 2ck' + a^j_{k'})\|_\infty \geq 2c.$$

Therefore, $\mathrm{cost}_p(f) > c > \|a^i - a^j\|_p$. Hence, $([d],[d],\mathrm{id}_{[d]})$ is the optimal partial matching and $w_p(\varphi(a^i), \varphi(a^j)) = \|a^i - a^j\|_p$, i.e. φ is an isometric embedding. □

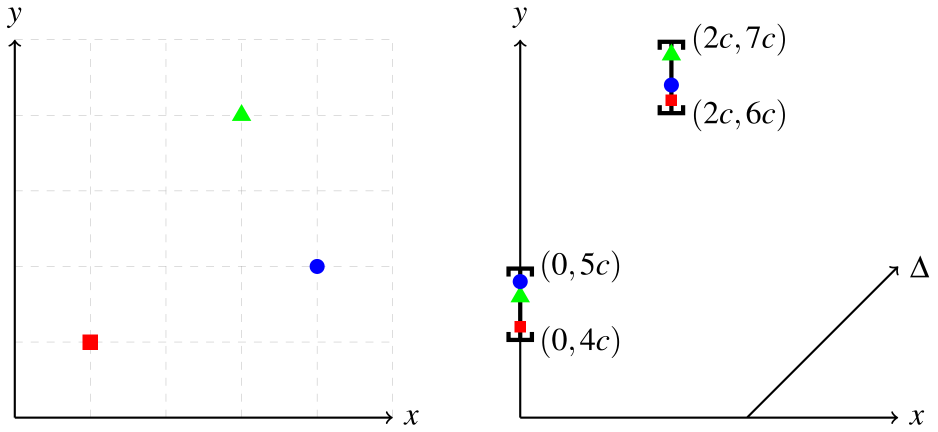

Figure 3-3. Embedding Finite Subset of $\mathbb{R}^2$ into (Dgm_2, w_2). Each point in a finite subset of $\mathbb{R}^2$ is mapped to a persistence diagram consisting of 2 points. The points lie on intervals in $\mathbb{R}^2_<$ chosen sufficiently far apart to guarantee the cost of the optimal matching equals the p-norm distance of the original points.

Theorem 3-3. *(Dgm_p, w_p) does not coarsely embed into a Hilbert space for $2 < p < \infty$.*

Proof. Fix $p > 2$ and suppose (Dgm_p, w_p) admits a coarse embedding into a Hilbert space. Then by Theorem 2-8, there exist non-decreasing functions $\rho_-, \rho_+ : [0,\infty) \to [0,\infty)$ such that $\lim_{t\to\infty} \rho_-(t) = \infty$ and for every finite subset $S \subseteq (\mathrm{Dgm}_p, w_p)$ there exists a map $f_S : S \to \ell_2$ satisfying

$$\rho_-(w_p(x,y)) \le \|f_S(x) - f_S(y)\|_2 \le \rho_+(w_p(x,y)) \tag{3-1}$$

for every $x, y \in S$. Define $\tilde{\rho}_- : [0,\infty) \to [0,\infty)$ by $\tilde{\rho}_-(t) = \rho_-(\max(t-1, 0))$ and observe that $\tilde{\rho}_-$ is non-decreasing and $\lim_{t\to\infty} \tilde{\rho}_-(t) = \infty$. We will show that for any finite subset $A \subseteq \ell_p$ there exists a map $g_A : A \to \ell_2$ satisfying

$$\tilde{\rho}_-(\|x-y\|_p) \le \|g_A(x) - g_A(y)\|_2 \le \rho_+(\|x-y\|_p) \tag{3-2}$$

for every $x, y \in A$. This implies by Theorem 2-8 that ℓ_p coarsely embeds into a Hilbert space which contradicts Corollary 2-1.

For any $m \in \mathbb{N}$, define $\pi_m, \rho_m : \ell_p \to \ell_p$ by $\pi_m(x) = (x_1, \dots, x_m, 0, 0, \dots)$ and $\rho_m(x) = (0, \dots, 0, x_{m+1}, x_{m+2}, \dots)$. Let $A = \{a^1, \dots, a^n\}$ be a finite subset of ℓ_p and choose $m \in \mathbb{N}$ sufficiently large such that $\|\rho_m(a^i)\|_p \le \frac{1}{2}$ for every $i = 1, \dots, n$. Then

$$\begin{aligned} \|a^i - a^j\|_p \ge \|\pi_m(a^i) - \pi_m(a^j)\|_p &\ge \|a^i - a^j\|_p - \|\rho_m(a^i) - \rho_m(a^j)\|_p \\ &\ge \|a^i - a^j\|_p - (\|\rho_m(a^i)\|_p + \|\rho_m(a^j)\|_p) \\ &\ge \|a^i - a^j\|_p - 1. \end{aligned}$$

Since $\pi_m(A)$ is isometric to a finite subset of $(\mathbb{R}^m, \|\cdot\|_p)$, there exists an isometric embedding $\varphi : \pi_m(A) \to (\mathrm{Dgm}_p, w_p)$ by Proposition 3-2. Let $S = \varphi(\pi_m(A))$ and f_S be as in (3-1). Define $g_A : A \to \ell_2$ by $g_A = f_S \circ \varphi \circ \pi_m$.

The following two series of inequalities show g_A satisfies (3-2). In both cases, the first inequality follows from (3-1), the equality follows from φ being an isometric embedding, and the second inequality follows from the monotonicity of ρ_+ and $\tilde{\rho}_-$, respectively.

$$\|g_A(a^i) - g_A(a^j)\|_2 \leq \rho_+(w_p(\varphi\pi_m(a^i), \varphi\pi_m(a^j))) = \rho_+(\|\pi_m(a^i) - \pi_m(a^j)\|_p) \leq \rho_+(\|a^i - a^j\|_p)$$

$$\|g_A(a^i) - g_A(a^j)\|_2 \geq \rho_-(w_p(\varphi\pi_m(a^i), \varphi\pi_m(a^j))) = \rho_-(\|\pi_m(a^i) - \pi_m(a^j)\|_p) \geq \tilde{\rho}_-(\|a^i - a^j\|_p)$$

□

CHAPTER 4
APPROXIMATION OF PERSISTENCE MODULES WITH DISCRETE MORSE THEORY

Definition 4-1. *Let f and f' be filtrations on a cell complex (X,∂) and $\delta > 0$. We say f' is a* δ-approximation *of f if $\|f - f'\|_\infty < \delta$. A δ-approximation f' is* monotone *if there exists a monotone map $g : f(X) \to \mathbb{R}$ such that $f' = g \circ f$.*

The distinction between these two concepts is relevant in Section 4.1.2 where we prove the optimality of a particular approximate filtration upon restricting the search space to monotone δ-approximations. Obviously, all monotone δ-approximations are δ-approximations. If (X,∂,f) is a filtered complex and $(A, w : Q \to K)$ is an acyclic partial matching, there is an induced filtered acyclic partial matching.

Definition 4-2. *Consider a filtered cell complex (X,∂,f) and an acyclic partial matching $(A, w : Q \to K)$. Then $Q_t = \{q \in Q \cap X^t \mid f(q) = f(w(q))\}$, $K_t = \{k \in K \cap X^t \mid f(k) = f(w^{-1}(k))\}$, $A_t = X^t \setminus (K_t \cup Q_t)$, and $w_t = w|_{Q_t} : Q_t \to K_t$ defines a filtered acyclic partial matching, called the* induced filtered acyclic partial matching. *We write w_f as a shorthand and call it an* induced matching. *We define $\|w_f\|$ to be the number of $q \in Q$ such that $q \in Q_t$ for some t, which equals $|Q_{\max(f(X))}|$.*

Suppose you are given a filtered cell complex (X,∂,f) and want to efficiently approximate the persistence diagram of the sub-level set filtration of f, denoted $H_*(f)$, within δ in the bottleneck distance. As a proxy for efficient computation, the algorithms in the following section are designed to find a filtered acyclic partial matching of a δ-approximation of f, denoted f', minimizing the number of cells in the Morse complex associated to the approximate filtration, denoted (A,∂^A). Since $H_*(f'|_A) \cong H_*(f')$ by Theorem 2-2 and $w_\infty(H_*(f'),H_*(f)) < \delta$ by Theorem 2-1, we can approximate $H_*(f)$ by computing $H_*(f'|_A)$. This procedure is illustrated in Figure 1-1.

Parts of this work were done in collaboration with Arkadi Schelling, University of Bremen.

4.1 Approximation Algorithms

4.1.1 Binning

Recall Definition 2-7 of filtered acyclic partial matchings and observe that cells with distinct filtration values cannot be matched even if they are very close. This suggests binning of filtration values as an obvious approach to find an approximate filtration that allows further reduction of the complex.

The most naive approach is to round up all filtration values to lie in $\delta\mathbb{Z}$. This can be improved by rounding up all values to lie in $2\delta\mathbb{Z}$ and subtracting δ from the result. This maps the intervals $(2k\delta, 2(k+1)\delta]$ to the single value $(2k+1)\delta$ for $k \in \mathbb{Z}$. As these intervals are twice as long as the intervals resulting from a simple ceiling function, this has the potential to double the number of possible matches compared to the naive δ-approximation.

However, in most filtrations on a point cloud, all 0-cells appear at time 0. The procedure above would prevent any matches between 0-cells and 1-cells. Hence, in computations, we shift the ceiling function by $\varepsilon \ll \delta$ such that an interval starts before the minimal filtration value. This defines the map $x \mapsto \min\{z \in (2\delta\mathbb{Z} - \varepsilon) \mid x \leq z\} - \delta$.

4.1.2 Induced Filtration

This approach constructs an unfiltered acyclic partial matching, like `MorseReduce` with a single filtration step. We sort the cells in their filtration order to prefer matches between cells of zero or small filtration differences. Afterwards, we find an optimal monotone approximation for this unfiltered acyclic partial matching.

Definition 4-3. *Let (X, ∂, f) be a filtered complex and $(A, w : Q \to K)$ an acyclic partial matching. A δ-approximation f' of f is* optimal with respect to *w if $\|w_{f'}\|$ is maximal among all δ-approximations of f.*

Definition 4-4. *Let (X, ∂, f) be a filtered complex. A δ-*discretization *of f is a finite subset $D = \{d_1 < d_2 < \cdots < d_n\}$ of $\mathbb{R}$ such that $d_i - d_{i-1} < \delta$ and $f(X) \subseteq [d_1, d_n)$.*

Let $(A, w : Q \to K)$ be an acyclic partial matching of X. For any $Y \subset \mathbb{R}$, define $P(Y)$ to be the set of $q \in Q$ for which Y does not intersect the interval $(f(q), f(w(q))]$, i.e.

$P(Y) = \{q \in Q \mid Y \cap (f(q), f(w(q))] = \emptyset\}$. Define $C(Y)$ to be the set of $q \in Q$ for which Y intersects the interval $(f(q), f(w(q))]$, i.e. $C(Y) = \{q \in Q \mid Y \cap (f(q), f(w(q))] \neq \emptyset\}$. Note that $P(Y) \cup C(Y) = Q$ since every $\{q, w(q)\}$-pair is either preserved or cut by Y. We say a δ-discretization D^* of f is optimal with respect to *w if the following inequality is satisfied for any other δ-discretization D of f.*

$$|C(D^*)| \leq |C(D)| \iff |P(D^*)| \geq |P(D)|$$

The following theorem states that a 2δ-discretization of a filtration f that is optimal with respect to w induces a δ-approximation of f that is optimal with respect to w among monotone approximations. In other words, finding optimal filtrations in this restricted setting can be reduced to finding optimal discretizations.

Theorem 4-1. *Let $\delta > 0$ and (X, ∂, F) be a filtered complex with an acyclic matching w. If $D^* = \{d_1, \ldots, d_n\}$ is an optimal 2δ-discretization of F with respect to w, then the map $f^* : F(X) \to \mathbb{R}$ given by*

$$f^*(s) = \frac{d_i + d_{i+1}}{2} 1_{[d_i, d_{i+1})}(s)$$

induces a monotone δ-approximation $F^* = f^* \circ F$ that is optimal with respect to w among monotone δ-approximations.

Proof. Suppose $F' = fF$ is an arbitrary monotone δ-approximation of F. Our goal is to show $\|w_{F^*}\| \geq \|w_{F'}\|$. We prove this in two steps.

1. Construct a 2δ-discretization D of F satisfying $\|w_{F'}\| \leq |P(D)|$.

2. Prove $|P(D^*)| \leq \|w_{F^*}\|$.

By the optimality of D^*, this completes the proof. Note that since $fF = F'$ is a δ-approximation of F, we have $|f(x) - x| < \delta$ for every $x \in F(X)$. Let $x_1 < x_2 < \cdots < x_N$ be an ordering of $F(X)$ and let $i := \max\{j = 1, \ldots, N \mid f(x_j) = f(x_1)\}$. Then

$$|x_1 - x_i| \leq |x_1 - f(x_1)| + |x_i - f(x_1)| = |x_1 - f(x_1)| + |x_i - f(x_i)| < 2\delta.$$

Choose $a_1 < x_1$ and $b_1 \in (x_i, x_{i+1})$ such that $b_1 - a_1 < 2\delta$. Note that $f(x_j) = f(x_1)$ if and only if $x_j \in (a_1, b_1)$ by monotonicity of f. By repeating this process we obtain a sequence $a_1 < b_1 < a_2 < b_2 \cdots < a_m < b_m$ such that $f(x_i) = f(x_j)$ if and only if $x_i, x_j \in (a_k, b_k)$ for some $k = 1, \ldots, m$. The set $\{a_1, b_1, \ldots, a_m, b_m\}$ is almost a 2δ-discretization of F because $F(X) \subseteq [a_1, b_m)$ and $b_i - a_i < 2\delta$ for every i. However, it may be the case that $a_{i+1} - b_i \geq 2\delta$. Extend the set $\{a_1, b_1, \ldots, a_m, b_m\}$ to a 2δ-discretization D of F by adding $\cup_{i=1}^{m-1}\{b_i + k\delta \mid k \in \mathbb{N}, b_i + k\delta < a_{i+1}\}$. Suppose $f(F(q)) = f(F(w(q)))$. By construction, $F(q), F(w(q)) \in (a_k, b_k)$ for some k, so $D \cap (F(q), F(w(q))] = \emptyset$. This completes step one.

For step two, suppose $q \in Q$ satisfies $D^* \cap (F(q), F(w(q))] = \emptyset$. Since $F(X) \subseteq [d_1, d_n)$, there must exist i such that $F(q), F(w(q)) \in [d_i, d_{i+1})$ so $f^*(F(q)) = f^*(F(w(q)))$. This completes step two. □

```
def SolveMatchingDiagram(start, stop, Is, delta):
  # Is is a multi-set of left-open, right-closed intervals.
  D = {start}
  Ds = {0: (D, {})}
  while True:
    best_k = min(Ds.keys())
    D, cuts = Ds.pop(best_k)
    if D[-1] > stop
      return D
    end_pt = D[-1] + delta
    rel_ints = Is[D[-1], end_pt]
    rel_pts = [iv.begin for iv in rel_ints]
    exts = {end_pt} ∪ {p for p in rel_pts if D[-1] < p < end_pt}
    for ext in exts:
      new_D = D ∪ {ext}
      new_cuts = Is[ext] ∪ cuts
      new_k = len(new_cuts)
      if new_k not in Ds or new_D[-1] > Ds[new_k][-1]:
        Ds[new_k] = (new_D, new_cuts)
```

Figure 4-1. Algorithm to Compute Optimal Discretization.

If we relax the definition of a δ-discretization to allow $d_i - d_{i-1} = \delta$, the filtration F^* defined in Theorem 4-1 may no longer be a δ-approximation of F. However, a useful algorithm to find an optimal such relaxed δ-discretization using an interval tree structure for a set of intervals `Is` and dynamic programming has been proposed by Orson L. Peters.[1] An interval tree is a data structure built from a set of intervals; see Section 14.3 of Cormen et al. [29]. For a set of n intervals, initial creation of an interval tree requires $O(n\log(n))$ time and intersection queries require $O(\log(n)+m)$ time, where m is the number of intervals returned. We provide the algorithm in Figure 4-1 and describe it below.

Let (X,∂,F) be a filtered complex, $(A,w:Q\to K)$ an acyclic partial matching, and $\delta>0$. Let `start` $=\min(F(X))$, `stop` $=\max(F(X))$, `Is` $=\{(F(q),F(w(q))] \mid q\in Q\}$, and `delta` $=\delta$. The variable `start` is used to initialize the output `D` in Line 3. The `while` loop on Line 5 repeats until the last element of `D` is strictly larger than `stop`. This is the stopping condition on Line 8 and the output on Line 9 is `D`. The variable `Is` is an interval tree data structure and `delta` is the upper-bound on the distance between consecutive points of `D`. Interval queries of the form `Is[a,b]` return the multi-set of intervals in `Is` which intersect $[a,b]$. Similarly, point queries of the form `Is[x]` return the multi-set of intervals in `Is` that contain the point x. A dictionary data structure is a collection of (key, value)-pairs such that each key appears at most once. The variable `Ds` is a dictionary data structure. Its values are pairs consisting of a partial solution `D` and the multi-subset of intervals in `Is` cut by `D`. The key corresponding to each value in `Ds` is the number of intervals cut. `Ds` is initialized on Line 4 to contain the partial solution corresponding to $\{$`start`$\}$, which cuts no intervals. The main part of the algorithm is the `while` loop on Line 5. On Lines 6-7, `D` is set to the partial solution in `Ds` cutting the fewest number of intervals, `cuts` is set to the multi-subset of `Is` cut by `D`, and the (`best_k`, `Ds[best_k]`)-pair is removed from `Ds`. Lines 10-13 construct a set of possible extensions called `exts` of `D` consisting of

$$\{\max(D)+\delta\}\cup\{F(q)\in F(Q) \mid \max(D)<F(q)<\max(D)+\delta\}.$$

For each such possible extension `ext`, Lines 15-17 extend `D` by `ext` and store this new `D` as `new_D`. The number of intervals cut by `new_D` is stored in `new_k`. The `if` statement on Line 18 checks if either `Ds` lacks the key `new_k`, i.e. `new_k not in Ds`, or else if `new_D` extends farther than the

[1] https://stackoverflow.com/questions/57250782

value in Ds corresponding to new_k, i.e. new_D[-1] > Ds[new_k][-1]. If either condition is satisfied, new_D is added to Ds with the key new_k. The purpose of Line 18 is to add a partial solution new_D to Ds if either Ds does not contain a partial solution cutting the same number of intervals as new_D or if Ds contains such a partial solution but its maximum value is less than that of new_D. The steps of the algorithm are visualized in Figure 4-2. Note that, as in Figure 4-2, intervals may appear in Is with multiplicity.

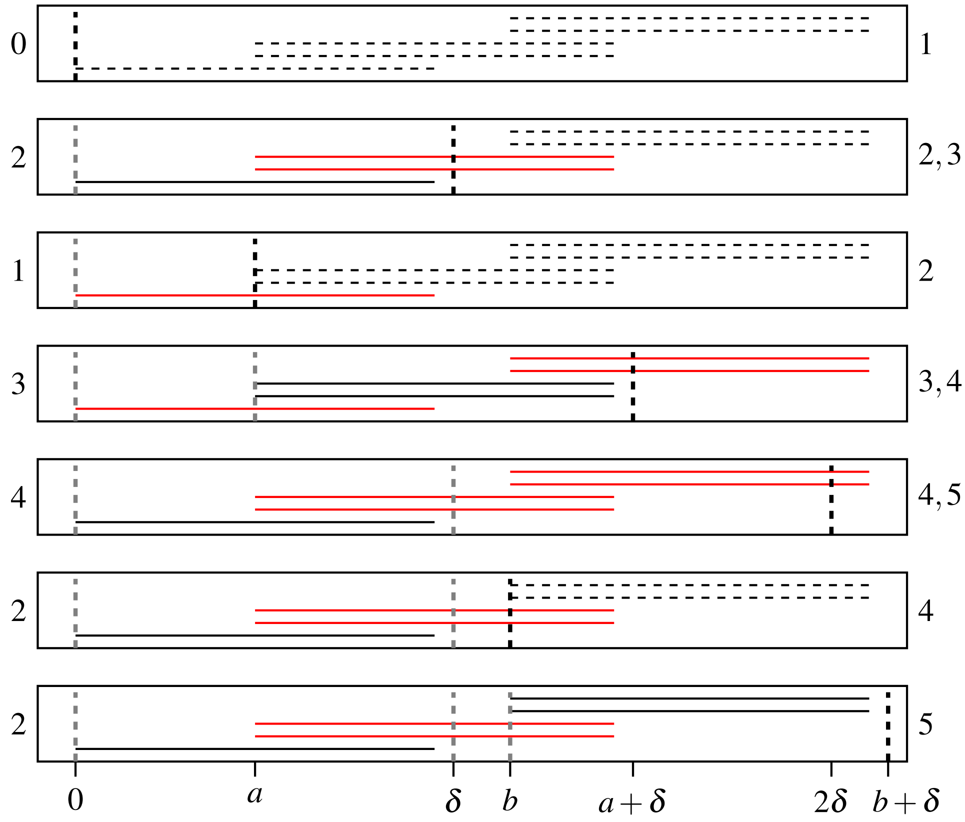

Figure 4-2. Finding an Optimal Approximation. Vertical lines show a subsolution, with the furthest point marked in black. Intervals cut by the subsolution are marked in red, uncut intervals in black, and unseen intervals are dashed. The number of cuts k for a given subsolution is on the left. On the right are the iterations of the while loop at the beginning of which the given subsolution is part of Ds. New subsolutions are built from the existing one with minimal k.

We now prove the correctness of the algorithm in Figure 4-1. Recall that $C(Y) = \{q \in Q \mid Y \cap (F(q), F(w(q))] \neq \emptyset\}$ in the following.

Proposition 4-1. *Let (X, ∂, F) be a filtered complex and $(A, w : Q \to K)$ be an acyclic partial matching. The output of*

$$\textit{SolveMatchingDiagram}(\min(F(X)), \max(F(X)), \{(F(q), F(w(q))] \mid q \in Q\}, \delta),$$

denoted $S = \{s_1 < s_2 < \cdots < s_n\}$, satisfies the following properties.

(1) $F(X) \subset [s_1, s_n)$ and $s_{i+1} - s_i \leq \delta$ for every i.

(2) If $Z = \{z_1 < \cdots < z_m\}$ satisfies the previous property then $|C(S)| \leq |C(Z)|$.

Proof. We first check that the returned set S satisfies (1). It is initialized at $\min(F(X))$ and every step of the algorithm extends a subsolution by at most δ; see Lines 10-13. Hence, it suffices to check that $\max(F(X)) < s_n$. Every subsolution is extended either to an element of $F(Q)$ or by δ, so by the finiteness of $F(X)$, the algorithm reaches the stopping condition. Since S was output, it must have satisfied the stopping condition, i.e. $\max(F(X)) < s_n$, so when the algorithm terminates, the returned set satisfies (1).

It remains to show the returned set S satisfies (2). Define D_i to be the value of the variable `D` on the i-th iteration of Line 7, e.g. $D_1 = \{\min(F(X))\}$. Suppose $Z = \{z_1 < \cdots < z_m\}$ satisfies (1). We prove by induction that for every $t = 1, \ldots, m$, there exists $i \in \mathbb{N}$ such that D_i satisfies $|C(D_i)| \leq |C(\{z_1, \ldots, z_t\})|$ and either $\max(D_i) \geq z_t$ or $\max(D_i) > \max(F(X))$. When $t = m$, this implies that S satisfies (2). If $t = 1$, this condition is satisfied for $i = 1$ when $D_1 = \{\min(F(X))\}$ because Z satisfying (1) implies $z_1 \leq \min(F(X)) = \max(D_1)$ and $|C(D_1)| = 0 = |C(\{z_1\})|$.

Suppose $t > 1$. By induction, there exists $i \in \mathbb{N}$ such that $D_i = \{x_1 < x_2 < \cdots < x_l\}$ satisfies $|C(D_i)| \leq |C(\{z_1, \ldots, z_{t-1}\})|$ and either $\max(D_i) \geq z_{t-1}$ or $\max(D_i) > \max(F(X))$. If $\max(D_i) > \max(F(X))$ or $\max(D_i) \geq z_t$, we are done, so we may assume $\max(D_i) \leq \max(F(X))$ and $\max(D_i) < z_t$. Hence, $z_{t-1} \leq x_l < z_t$, which implies $z_t \in (x_l, x_l + \delta]$. Define

$$x_{l+1} = \min\{x \in F(Q) \cup \{x_l + \delta\} \mid z_t \leq x \leq x_l + \delta\}.$$

Let $q \in C(\{x_{l+1}\}) \setminus C(D_i)$. Then $x_{l+1} \in (F(q), F(w(q))]$ and $x_l \notin (F(q), F(w(q))]$, which implies

$x_l \leq F(q)$. Since $z_{t-1} \leq x_l$, we have $q \notin C(\{z_1, \ldots, z_{t-1}\})$. Since $F(q) < x_{l+1} \leq x_l + \delta$, if $z_t \leq F(q)$ then $F(q) \in \{x \in F(Q) \cup \{x_l + \delta\} \mid z_t \leq x \leq x_l + \delta\}$, but this is a contradiction since x_{l+1} is by definition the minimum of this set. Hence, $F(q) < z_t$ and we have $F(q) < z_t \leq x_{l+1} \leq F(w(q))$, so $q \in C(\{z_t\}) \setminus C(\{z_1, \ldots, z_{t-1}\})$. Together with the inductive hypothesis, this gives the inequality in the series below. Recall that $D_i = \{x_1, \ldots, x_l\}$.

$$\begin{aligned} |C(\{x_1, \ldots, x_l, x_{l+1}\})| &= |C(D_i)| + |C(\{x_{l+1}\}) \setminus C(D_i)| \\ &\leq |C(\{z_1, \ldots, z_{t-1}\})| + |C(z_t) \setminus C(z_1, \ldots, z_{t-1})| \\ &= |C(\{z_1, \ldots, z_t\})| \end{aligned}$$

Since $\max(D_i) \leq \max(F(X))$, the stopping condition on Line 8 is not met on the i-th iteration. Note that x_{l+1} above is one of the extensions in `exts` on Line 13 because every element of $\{x \in F(Q) \cup \{x_l + \delta\} \mid z_t \leq x \leq x_l + \delta\}$ is one of the extensions in `exts`. Indeed, $x_l + \delta = \max(D_i) + \delta$ and if $z_t \leq F(q) \leq x_l + \delta$ then $x_l < z_t \leq F(q) < x_l + \delta$. If `Ds` contains the key $|C(\{x_1, \ldots, x_{l+1}\})|$ at the beginning of the i-th iteration of the `for` loop on Line 14, the associated partial solution will only be replaced if one of the considered extensions extends farther; see Line 18. Otherwise, the partial solution $\{x_1, \ldots, x_{l+1}\}$ or another partial solution extending farther will be added to `Ds` with the key $|C(\{x_1, \ldots, x_{l+1}\})|$. This implies that at the end of the i-th iteration of the `for` loop on Line 14, `Ds` will contain a partial solution D with the key $|C(\{x_1, \ldots, x_{l+1}\})|$ satisfying $\max(D) \geq x_{l+1} \geq z_t$ and $|C(D)| = |C(\{x_1, \ldots, x_{l+1}\})| \leq |C(\{z_1, \ldots, z_t\})|$.

If the algorithm terminates after the i-th iteration with a `D` that cuts fewer than $|C(\{x_1, \ldots, x_{l+1}\})|$ intervals, then the inductive step is complete because $|C(D)| < |C(\{x_1, \ldots, x_{l+1}\})| \leq |C(\{z_1, \ldots, z_t\})|$ and $\max(D) > \max(F(X))$. Each iteration of Line 7 sets `D` to be the partial solution that cuts the fewest number of intervals, and a key can only be removed from `Ds` by Line 7. Hence, if the algorithm does not terminate with a solution cutting fewer than $|C(\{x_1, \ldots, x_{l+1}\})|$ intervals, then for some $j > i$, we have $|C(D_j)| = |C(\{x_1, \ldots, x_{l+1}\})|$. The `if` statement on Line 18 only replaces a partial solution in `Ds` if a new one is found cutting the same number of intervals and extending farther. Hence, $|C(D_j)| = |C(\{x_1, \ldots, x_{l+1}\})|$ and $\max(D_j) \geq x_{l+1} \geq z_t$, which completes the inductive step and the proof. □

Remark 4-1. *Note that since $\{(F(q), F(w(q))] \mid q \in Q\}$ is finite and the intervals are right-closed, if $S = \{s_1, \ldots, s_n\}$ is the output of SolveMatchingDiagram then $S' = \{s_1, s_2 - \varepsilon, \ldots, s_n - n\varepsilon\}$, for $\varepsilon > 0$ sufficiently small, is an optimal δ-discretization with respect to w in the sense of Definition 4-4.*

4.1.3 Gradient Paths

There is an adaptation of `MorseReduce` that constructs a non-monotone δ-approximation to the filtration. The main idea in `MorseReduce` to construct an acyclic matching consists of iteratively tagging a random simplex as critical and growing gradient paths towards this critical cell. In the original algorithm this growth is strictly limited to paths along the same filtration value, such that it constructs a filtered acyclic partial matching. Instead, the approximation algorithm we consider grows gradient paths up to a filtration difference of δ over the filtration value of the initial critical cell. Afterwards, it decreases all filtration values of cells along the gradient paths to match the filtration value of the critical cell. This new filtration is a δ-approximation from below in which the constructed gradient paths are a filtered acyclic partial matching.

4.2 Reduction Sizes in a Synthetic Experiment

When comparing the three approximation approaches, we are mainly interested in the reduction sizes. `MorseReduce` already reduces complexes by a substantial fraction. We compare how many of the remaining cells we are able to reduce by approximate solutions. Given a filtered complex (X, f) we first apply `MorseReduce` until stabilization to get a prereduced complex (X', f'), which is further reduced using one of the three approximation approaches in Section 4.1.

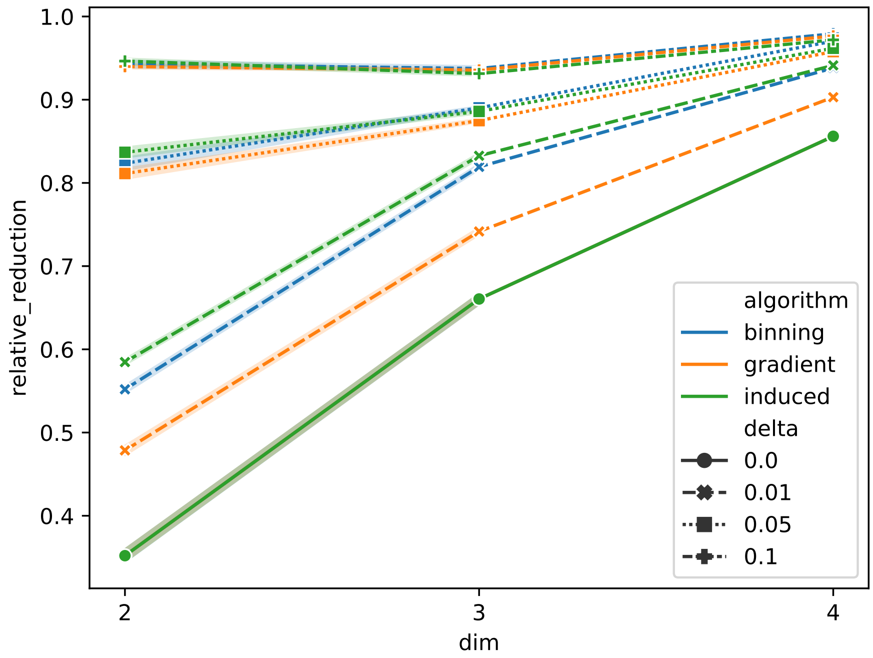

Figure 4-3. Reduction Size Experiment.

We apply this pipeline to the alpha complex of 20 independent samples of 200 standard normally distributed points in $\mathbb{R}^n$. We measure reduction by the *relative reduction* $r = \frac{|A|}{|X|}$. The possible reduction size of a filtered complex depends on the distribution of filtration differences between faces and cofaces. In the case of a Vietoris-Rips filtration, exact `MorseReduce` can already shrink the complex by a substantial fraction because higher dimensional simplices are added at the same filtration time as their last edge. Approximate solutions become more useful in a case like Čech or alpha filtrations.

Figure 4-3 shows the results for different values of n and δ. Note that $\delta = 0$ for all three approaches behaves like exact `MorseReduce` and already reduces about 35% of the complex when $n = 2$. All approaches behave similarly in that they quickly increase reductions to over 80% with $\delta = 0.05$. The method in Section 4.1.2 picks the optimal solution for a given acyclic matching w, but this w may be suboptimal despite constructing it in filtration order. Binning can theoretically reduce more cells than this induced filtration. However, this did not happen in any of our experimental runs.

CHAPTER 5
STABILIZING THE UNSTABLE OUTPUT OF PERSISTENT HOMOLOGY COMPUTATIONS

We propose a general technique for extracting a larger set of stable information from persistent homology computations than is currently done. The persistent homology algorithm is usually viewed as a procedure that starts with a filtered complex and ends with a persistence diagram. This procedure is stable, at least to certain types of perturbations of the input. This supports the use of the diagram as a signature of the input and the use of features derived from it in statistics and machine learning. However, these computations also produce other information of great interest to practitioners that is unfortunately unstable. For example, each point in the diagram corresponds to a simplex whose addition in the filtration results in the birth of the corresponding persistent homology class, but this correspondence is unstable. In addition, the persistence diagram is not stable with respect to other procedures that are employed in practice, such as thresholding a point cloud by density. We recast these problems as real-valued functions which are discontinuous but measurable and then observe that convolving such a function with a suitable function produces a Lipschitz function. The resulting stable function can be estimated by perturbing the input and averaging the output. We illustrate this approach with a number of examples, including a stable localization of a persistent homology generator from brain imaging data.

5.1 Instability of Auxiliary Information

Theorem 2-1 tells us that the persistence diagram obtained in the output of a persistent homology computation is stable with respect to certain perturbations of the input used to construct a filtered abstract simplicial complex. However, other outputs of persistent homology computations are not stable. This includes the simplices and cycles that generate persistent homology classes. These are of great interest to practitioners hoping to interpret persistence calculations more directly. In addition, many persistence computations rely on choices of parameters and the resulting persistence diagrams may be unstable with respect to these choices.

Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature Journal of Applied and Computational Topology [6]

5.1.1 Instability of Generating Cycles or Simplices

Persistence diagrams are useful and robust measures of the size of topological features. What they are less good at, on the other hand, is robustly pinpointing the location of important topological features. We use Figure 2-1 to illustrate this problem. Suppose that we have the fixed domain X and we observe the function f. One of the most prominent points in $H_0(f)$ is u, which corresponds to the pair of values $f(x)$ and $f(w)$. We might thus be tempted to say that f has an important feature, a component of high-persistence, at x. However, consider the nearby function g. Its degree-0 diagram $H_0(g)$ has a point u' that is very close to u, but this point corresponds to the pair of values $f(y)$ and $f(w)$. There is still a component born at $g(x)$, but it corresponds to the much smaller persistence point v'. And so while the persistence of the point u is a stable summary of the function f, the actual location x of the topological feature it corresponds to is not.

This is unfortunate. Several recent works [5, 7] have shown that the presence of points in certain regions of the persistence diagram has strong correlation with covariates under study. For example, each diagram in Bendich et al. [7] came from a filtration of the brain artery tree in a specific patient's brain, and it was found that the density of points in a certain middle-persistence range gave strong correlations with patient age. It would of course be tempting to hold specific locations in the brain responsible for these points with high distinguishing power. In Section 5.3, we both rigorously define this non-robustness and give a method for addressing it.

5.1.2 Outliers and Instability of Parameter Choices

Theorem 2-1 guarantees the persistence diagrams associated to two Hausdorff-close point clouds will themselves be close. However, it says nothing about the outlier problem. For example, consider the point cloud X in the top left of Figure 5-1 to which we apply the Vietoris-Rips construction. Its degree-1 persistence diagram (top right of same figure) has one high-persistence point, which corresponds to the circle that we qualitatively see when looking at the points. On the other hand, consider the point cloud X' on the bottom left, which consists of X and three outlier points spread across the interior of the circle. The degree-1 persistence diagram of X' is not close to that of X. There is still one point of fairly high persistence, but it is much closer to the diagonal.

In practice, this problem is often addressed by first de-noising the point cloud in some way. For example, Carlsson et al. [14] first thresholded by density before computing Vietoris-Rips filtrations when they discovered a Klein bottle in the space of natural images. There are no guarantees that a different, nearby choice of density threshold parameter would not give a qualitatively different persistence diagram. Section 5.3 addresses this by introducing a general method for handling parameter choice in persistence computations.

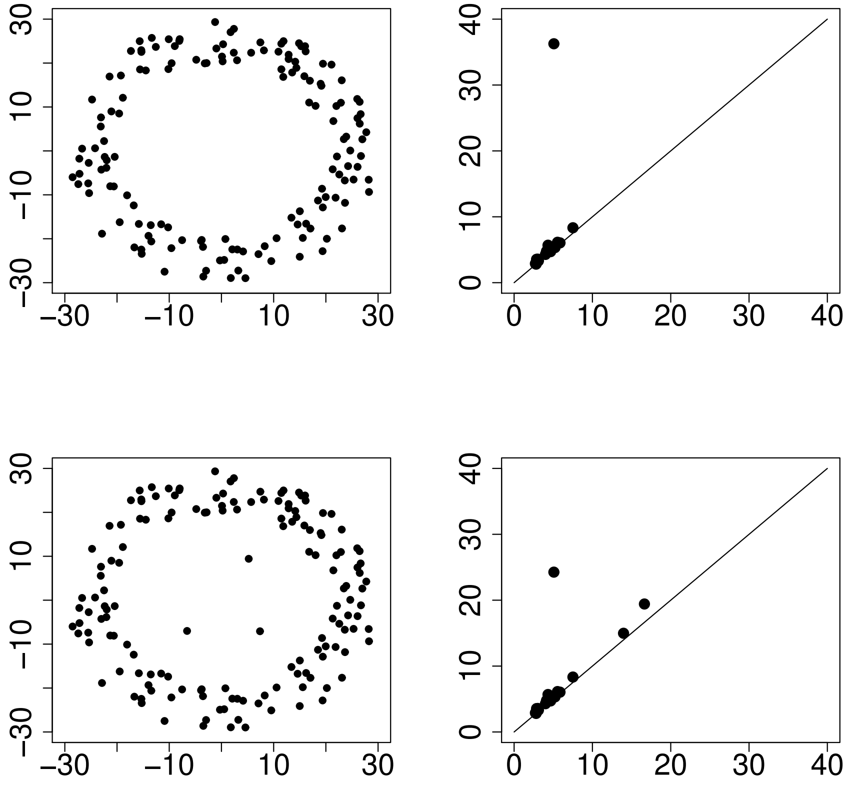

Figure 5-1. Outlier Problem. Illustration of the outlier problem for the persistent homology of the Vietoris-Rips complex of a point cloud.

5.2 Three Motivating Examples

Before developing the theory, we begin by providing three examples, the first and third to synthetic data and the second to brain imaging data.

5.2.1 A Cycle Generating a Persistent Homology Class

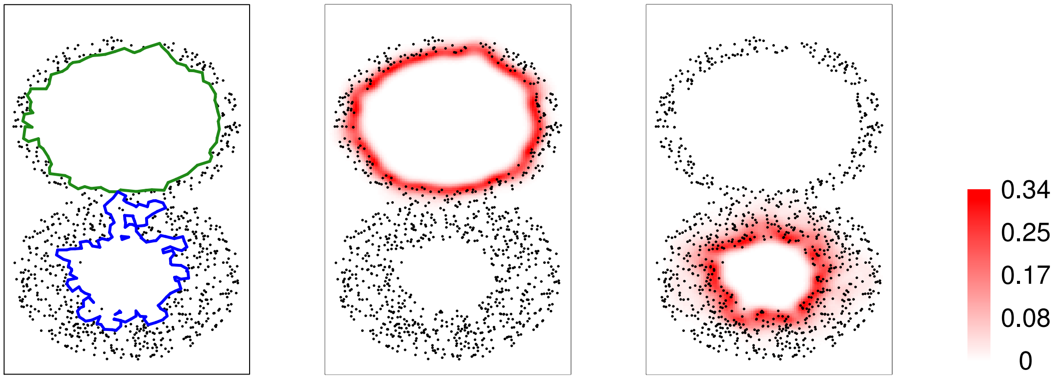

Figure 5-2. Finding the First and Second Longest Bars. The first panel shows the original sample and the representative cycles produced by Dionysus [50] for the first and second longest bars. The latter two panels show the proportion of perturbations for which each square in a grid intersected the representative geometric cycle of the first or second longest bar, respectively.

We sample 1000 points uniformly from two conjoined annuli of inner and outer radii $(20,50)$ and $(40,50)$. Using Dionysus [50], we compute the 1-dimensional persistent homology of the alpha filtration of our sample and obtain a representative cycle for the longest and second-longest bars. See Figure 5-2, left panel. However, the embedded location of these cycles is unstable. We would like to quantify the uncertainty of this location. To do so, we consider a square grid with edge-length 1. Our function $h : \mathbb{R}^{2000} \to \mathbb{R}$ has input the coordinates of the sampled points and has output 1 if the geometric cycle produced by Dionysus intersects a given square in our grid and otherwise has output 0.

We perturb the sampled points 10,000 times by adding Gaussian noise with standard deviation 3. For each square, we find the proportion of trials in which the representative geometric cycle for the longest or second-longest bar produced by Dionysus intersects the square. By performing this procedure simultaneously for every square in the grid, we obtain the second and third panels in Figure 5-2. To see the effect of varying the choice of bandwidth, see Section 5.4.2.

5.2.2 Location of a Persistent Homology Generator in Brain Imaging Data

Bendich et al. [7] apply topological data analysis to brain arteries extracted from magnetic resonance images. Mathematically, each of these brain arteries is a graph embedded in three-dimensional Euclidean space. Using the height (the z-coordinate) one obtains a filtration on this graph, which may be used to compute degree-zero persistent homology.

To facilitate statistical analysis of the resulting persistence diagrams, they convert each persistence diagram to a vector consisting of the lengths of the 100 longest bars in decreasing order. In their analysis, the length of the 28th longest bar is a numerical feature that yields a correlation with age that is near-optimal among vector features consisting of the lengths of the ith through jth longest bars for any $1 \leq i \leq j \leq 100$.

If one wants to find a biological interpretation of this result, it is obvious to ask for the location of the generator of the 28th longest bar for each subject. It is easy to locate the generator responsible for the birth of the 28th longest bar. It will be a particular vertex of the graph, whose image is a point in space. However, the location of this point is unstable. As explained in Section 5.1.1, small perturbations of the spatial coordinates of the vertices of the graph can lead to large changes of this location.

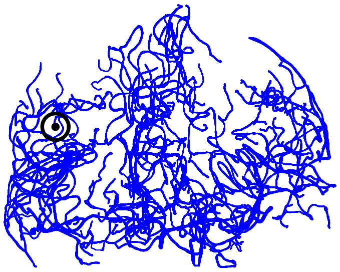

Figure 5-3. Location of Generating Simplex. The black dot is the location of the generator of the 28th longest bar in degree-zero persistent homology. We consider the indicator function on the location of this generator with respect to the given sphere.

We choose a ball centered at this point and consider the function whose value is 1 if the location of the generator of the 28th longest bar is located in this ball and is 0 otherwise. The resulting function $h : \mathbb{R}^{3V} \to \mathbb{R}$ (where V is the number of vertices in the graph) is unstable, but it may be stabilized using the method summarized in Figure 1-3. Applying the algorithm in Figure 1-3 with $M = 1000$ and $\sigma = 0.1$, we obtain an estimate of the stable value of $h * K$ evaluated at the observed input, equal to 0.637. This shows that under small perturbations of the input, over half of the time the generator of the 28th longest bar is located in the chosen ball. This result holds for a large range of sizes of balls. See Section 5.4.3 for some further discussion.

We remark that this approach provides a resolution of the conflict between TDA theorists and TDA users expressed in the Fundamental Conundrum of Topological Data Analysis in the introduction. We can provide TDA users with a location of a generator of a persistent homology class together with an estimate of a stable real value of how often this location lies in a given region under certain perturbations.

5.2.3 Persistence of a Homology Class Born in a Region

Consider the function f on the square in Figure 5-4. This induces a function $\bar{f}$ on the torus since $f(x,y) = 0$ on the boundary of the square. Suppose we are only given a finite sample of this induced function and we are interested in the presence of long-lived bars which are born in the region of the torus corresponding to the second quadrant of the square.

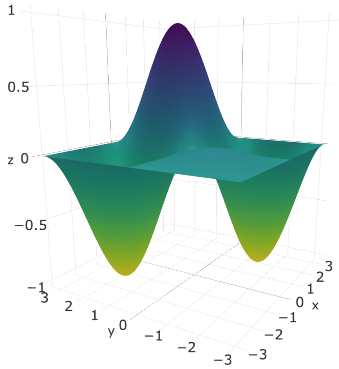

Figure 5-4. Torus Underlying Function. The graph of the function on the square $[-\pi,\pi]^2$ given by $f(u,v) = \sin(u)\sin(v)(1-0.9^*1_{\{u<0,v<0\}}) : [-\pi,\pi]^2 \to \mathbb{R}$. It induces a function on the torus, $\bar{f} : T^2 \to \mathbb{R}$, with two global minima with value -1, one global maximum with value 1, one local maximum with value 0.1, and four saddle points with value 0.

To be concrete, we start with a sample X of N points from the graph of $\bar{f}$, by sampling u_i, v_i independently from the uniform distribution on $[-\pi,\pi]$ and letting $z_i = f(u_i,v_i)$. Note that X is a random variable. We use X to construct a filtered simplicial complex approximating the unknown function $\bar{f}$ as follows. From the points $\{(u_i, v_i)\}$ we construct a Delaunay triangulation of the torus. We filter this simplicial complex by assigning the vertex (u_i, v_i) the value z_i and assigning edges and triangles the maximum value of their vertices.

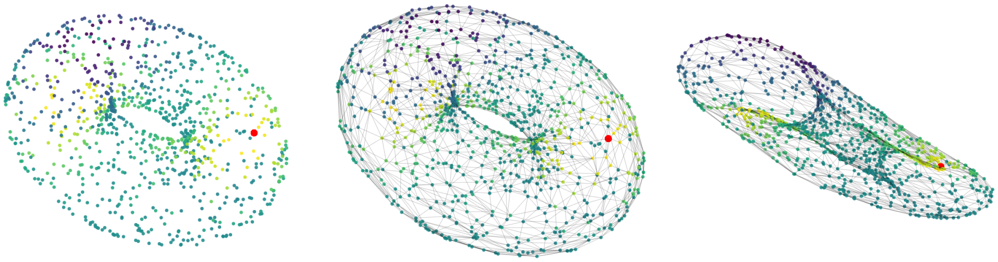

Figure 5-5. Torus Experiment. Sample of 1000 points from the graph $\{(x, f(x)) : x \in T^2\}$, where the function values are indicated using the same color scale as in Figure 5-4. The points on the torus are used to construct a Delaunay triangulation, which is filtered using the function values. On the right we indicate the filtration values by moving the points in the normal direction.

We compute the 0-dimensional extended[1] persistence diagram of this filtered simplicial complex. Let $h(X)$ be the length of the longest bar if that bar was born in the region corresponding to the second quadrant (see Figure 5-4) and 0 otherwise. This process defines a function $h : \mathbb{R}^{3N} \to \mathbb{R}$, but h is unstable. Consider the sample $X = x$ in Figure 5-5. We have $h(x) = 0$ since the global minimum, highlighted in red, is born outside the region corresponding to the second quadrant. Because of the symmetry of f, the random variable $h(X)$ is 0 approximately half the time and about 2 approximately half the time. Let K denote the $3N$-variate Gaussian with mean 0 and standard deviation 0.2. For $M \geq 1$, sample $\varepsilon_1, \ldots, \varepsilon_M$ independently from K. Compute $\frac{1}{M}\sum_{i=1}^{M} h(x - \varepsilon_i)$. See Figure 5-6. As M increases, this quantity converges to $g(x)$, where $g := h * K$ is the stabilized version of h.

[1] Extended persistent homology follows the homology of increasing sublevel sets with the relative homology of the whole space relative to decreasing superlevel sets [26]. In the case considered here, it pairs the global minimum with the global maximum.

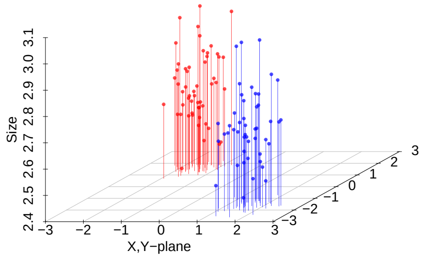

Figure 5-6. Locations and Sizes of 100 Longest Bars from the Trials. Averaging the lengths of the red bars over 1000 trials we get 1.291, which is consistent with the fact that the random variable $h(X)$ is 0 or about 2 with equal probability. We should not expect $\lim_{M\to\infty}\frac{1}{M}\sum_{i=1}^{M}h(x+\varepsilon_i)$ to converge to 1 because unlike f, a particular sample $X=x$ is not symmetric with respect to the second and fourth quadrants.

5.3 Stability from Convolutions

In this section we show how functions may be stabilized by convolving them with a kernel. First, we give three general results with various assumptions on the function and the kernel. Next, we apply them in three particular cases: the simple triangular kernel and the commonly used Epanechnikov and Gaussian kernels.

5.3.1 Lipschitz Functions and Convolution

Let us start by recalling a few definitions. For $C \geq 0$, a function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be *C-Lipschitz* if for all $u, v \in \mathbb{R}^n$, $|f(u) - f(v)| \leq C|u - v|$, where $|x|$ denotes the Euclidean norm. We will call a function *Lipschitz* if it is C-Lipschitz for some $C \geq 0$. The support of f, denoted $\mathrm{supp}(f)$, is the closure of the subset of $\mathbb{R}^n$ where f is non-zero.

Let $h, g : \mathbb{R}^n \to \mathbb{R}$ be (Lebesgue) measurable functions that are defined almost everywhere. The *1-norm* of h is given by $\|h\|_1 = \int_{\mathbb{R}^n} |h(t)|dt$, if it exists. The *essential supremum* of h, denoted by $\|h\|_\infty$, is the smallest number a such that the set $\{x \mid |f(x)| > a\}$ has measure 0. If it exists, the *convolution product* of h and g, is given by

$$(h * g)(t) = \int_{\mathbb{R}^n} h(s)g(t-s)ds = \int_{\mathbb{R}^n} h(t-s)g(s)ds.$$

It exists everywhere, for example, if one function is essentially bounded and the other is integrable or if one function is bounded and compactly supported and the other is locally integrable [37]. Throughout this section we assume that $h : \mathbb{R}^d \to \mathbb{R}$ is defined almost everywhere, $K : \mathbb{R}^d \to \mathbb{R}$ and that that the convolution product $h * K$ exists almost everywhere.

5.3.2 Stability Theorems

We now give several conditions on a pair of functions which imply that their convolution product is (locally) Lipschitz. The first result appears in Fremlin [37], but the proof is included here for completeness.

Theorem 5-1. *If $\|h\|_1 = a$ and K is b-Lipschitz, then $h * K$ is ab-Lipschitz.*

Proof. Let $g = h * K$. First we have, $g(u) - g(v) = \int_{\mathbb{R}^n} h(s)\,(K(u-s) - K(v-s))\,ds$. Then, $|g(u) - g(v)| \leq \int_{\mathbb{R}^n} |h(s)||K(u-s) - K(v-s)|ds \leq \int_{\mathbb{R}^n} |h(s)|b|u-v|ds \leq ab|u-v|$. □

Let $B_\alpha(x)$ denote the closed ball of radius α centered at $x \in \mathbb{R}^d$, and let V_d denote the volume of the d-dimensional ball of radius 1.

Theorem 5-2. *Let $x \in \mathbb{R}^d$ and let $\alpha > 0$. If $\|h\|_\infty \leq M$ on $B_{2\alpha}(x)$, K is b-Lipschitz and* $\operatorname{supp}(K) \subseteq B_\alpha(0)$, *then $h * K$ is $2Mb\alpha^d V_d$-Lipschitz in $B_\alpha(x)$.*

Proof. Let $g = h * K$. Let $u, v \in B_\alpha(x)$. As in the previous proof, $|g(u) - g(v)| \leq \int_{\mathbb{R}^n} |h(s)||K(u-s) - K(v-s)|ds \leq \int_{B_\alpha(u) \cup B_\alpha(v)} |h(s)|b|u-v|\,dx \leq 2Mb\alpha^d V_d |u-v|$. □

Theorem 5-3. *If $\|h\|_\infty \leq M$ and $\int |K(s+t) - K(s)|\,ds \leq b|t|$ for all $t \in \mathbb{R}^d$, then $h * K$ is Mb-Lipschitz.*

Proof. Let $g = h * K$. Again,
$|g(u) - g(v)| \leq \int_{\mathbb{R}^n} |h(s)||K(u-s) - K(v-s)|ds \leq \int M|K(u-v+x) - K(x)|\,dx \leq Mb|u-v|$. □

5.3.3 Application to Kernels

We now apply the above theorems to smooth a function h, obtaining a Lipschitz function. That is, we will take K to be a *kernel*, a non-negative integrable real-valued function on $\mathbb{R}^n$ satisfying $\int K(x)dx = 1$, $\int xK(x)dx = 0$ and $\int x^2 K(x)dx < \infty$. For example, we can choose K to be the *triangular kernel*, $K(x) = c\max(1-\|x\|, 0)$, for appropriate normalization constant c (see Figure 5-7). The most common choices are the Gaussian kernel and the Epanechnikov kernel, which are described below (see Figure 5-7). Notice that if K is a kernel, then so is $K_\alpha(x) = \frac{1}{\alpha^n}K(\frac{x}{\alpha})$.[2] The parameter α is called the *bandwidth* and allows one to control the amount of smoothing.

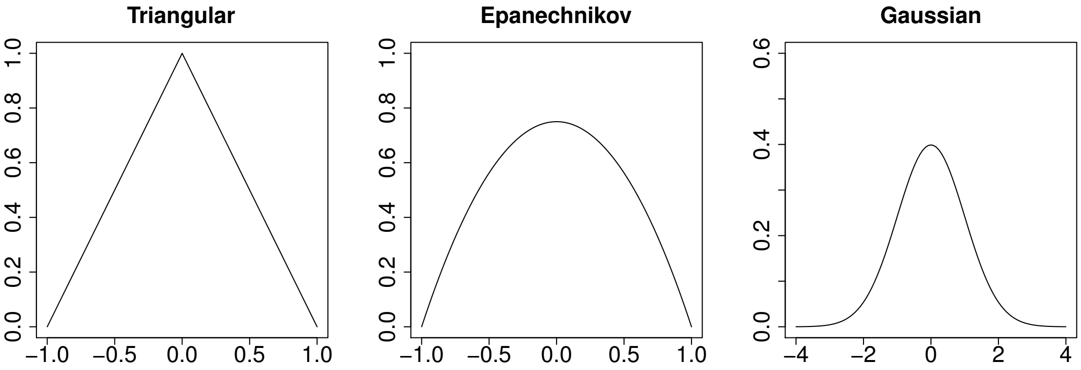

Figure 5-7. Graphs of Three Common Kernels.

5.3.3.1 The triangular kernel

Let $\alpha > 0$. Let V_d denote the volume of the n-dimensional ball of radius 1. For $A \subseteq \mathbb{R}^d$, let I_A denote the indicator function on A. That is, $I_A(x) = 1$ if $x \in A$ and 0 otherwise. The *triangular kernel* is given by

$$K_\alpha(x) = \frac{d+1}{\alpha^d V_d}\left(1 - \frac{|x|}{\alpha}\right) I_{B_\alpha(0)}.$$

Note that $\operatorname{supp}(K_\alpha) = B_\alpha(0)$ and K_α is $\frac{d+1}{\alpha^{d+1}V_d}$-Lipschitz. Applying Theorem 5-2, we have the following.

Corollary 5-1. *Let $x \in \mathbb{R}^d$. If $\|h\|_\infty \le M$ on $B_{2\alpha}(x)$ then $h * K_\alpha$ is $\frac{2M(d+1)}{\alpha}$-Lipschitz in $B_\alpha(x)$.*

Note that it follows that if the bound on h is global then so is the Lipschitz bound.

5.3.3.2 The Epanechnikov kernel

Let $\alpha > 0$. The *Epanechnikov kernel* is given by

$$K_\alpha(x) = \frac{d+2}{2\alpha^d V_d}\left(1 - \frac{|x|^2}{\alpha^2}\right) I_{B_\alpha(0)}.$$

[2] More generally, we can choose the bandwidth to be a symmetric positive definite matrix H and let $K_H(x) = \frac{1}{\sqrt{\det H}}K(H^{-1/2}x)$.

Now $\text{supp}(K_\alpha) = B_\alpha(0)$ and K_α is $\frac{d+2}{\alpha^{d+1}V_d}$-Lipschitz. Applying Theorem 5-2, we have the following.

Corollary 5-2. *Let $x \in \mathbb{R}^d$. If $\|h\|_\infty \leq M$ on $B_{2\alpha}(x)$ then $h * K_\alpha$ is $\frac{2M(d+2)}{\alpha}$-Lipschitz in $B_\alpha(x)$.*

5.3.3.3 The Gaussian kernel

Let $\alpha > 0$. The *Gaussian kernel* is given by

$$K_\alpha(x) = \frac{1}{\alpha^d (2\pi)^{d/2}} e^{-|x|^2/2\alpha^2}.$$

Lemma 5-1. *For the Gaussian kernel K_α, let $f(t) = \int |K_\alpha(s+t) - K_\alpha(s)|\, ds$. Then $f(t) \leq \frac{2}{\alpha\sqrt{2\pi}}|t|$ for all $t \in \mathbb{R}^d$.*

Proof. Change coordinates so that $s = -\frac{|t|}{2}e_1$ and $s+t = \frac{|t|}{2}e_1$. Then by symmetry

$$\begin{aligned}
f(t) &= 2\left[\int_{x_1 \geq -\frac{|t|}{2}} K_\alpha(x)\, dx - \int_{x_1 \geq \frac{|t|}{2}} K_\alpha(x)\, dx\right] \\
&= 4\int_{0 \leq x_1 \leq \frac{|t|}{2}} K_\alpha(x)\, dx \\
&= \frac{4}{\alpha^d (2\pi)^{d/2}} \int_0^{\frac{|t|}{2}} e^{-x_1^2/2\alpha^2}\, dx_1 \int_{-\infty}^{\infty} e^{-x_2^2/2\alpha^2}\, dx_2 \cdots \int_{-\infty}^{\infty} e^{-x_d^2/2\alpha^2}\, dx_d \\
&= \frac{4}{\alpha\sqrt{2\pi}} \int_0^{\frac{|t|}{2}} e^{-x_1^2/2\alpha^2}\, dx_1
\end{aligned}$$

It follows that $f(t) \leq \frac{4}{\alpha\sqrt{2\pi}} \int_0^{|t|/2} dx_1 = \frac{2}{\alpha\sqrt{2\pi}}|t|$. □

Thus by Theorem 5-3 we have the following corollary.

Corollary 5-3. *If $\|h\|_\infty \leq M$ then $h * K_\alpha$ is $\frac{2M}{\alpha\sqrt{2\pi}}$-Lipschitz.*

In practice, the function h will rarely be essentially bounded, but this can be arranged by setting it to be 0 outside a closed ball centered at a specified configuration.

5.3.4 Sharpness of the Lipschitz Constants

Assume that $K_\alpha : \mathbb{R}^d \to \mathbb{R}$ is symmetric in the first variable. Let $h : \mathbb{R}^d \to \mathbb{R}$ be defined by $h(x) = 1$ if $x_1 \geq 0$ and -1 otherwise. Let $g(t) = h * K_\alpha(te_1) - h * K_\alpha(-te_1)$. Then we calculate

$$\begin{aligned} h * K_\alpha(te_1) &= \int_{\mathbb{R}^d} h(te_1 - x)K_\alpha(x)\,dx \\ &= \int_{x_1 \leq t} K_\alpha(x)\,dx - \int_{x_1 > t} K_\alpha(x)\,dx \\ &= \operatorname{sign}(t) \int_{-|t| \leq x_1 \leq |t|} K_\alpha(x)\,dx. \end{aligned}$$

So

$$g(t) = 2\operatorname{sign}(t) \int_{-|t| \leq x_1 \leq |t|} K_\alpha(x)\,dx = 4\operatorname{sign}(t) \int_{0 \leq x_1 \leq |t|} K_\alpha(x)\,dx.$$

Let K_α be the Gaussian kernel. Then

$$\begin{aligned} g(t) &= \frac{4}{\alpha^d (2\pi)^{d/2}} \int_0^t e^{-x_1^2/2\alpha^2}\,dx_1 \int_{-\infty}^{\infty} e^{-x_2^2/2\alpha^2}\,dx_2 \cdots \int_{-\infty}^{\infty} e^{-x_d^2/2\alpha^2}\,dx_d \\ &= \frac{4}{\alpha\sqrt{2\pi}} \int_0^t e^{-x_1^2/2\alpha^2}\,dx_1. \end{aligned}$$

It follows that $g(t)$ converges to $\frac{4t}{\alpha\sqrt{2\pi}}$ as t approaches 0 by the first fundamental theorem of calculus. Hence, the Lipschitz constant given in Corollary 5-3 is optimal.

When $d = 1$ and K_α is the triangular kernel, $g(t) = \frac{4 \cdot 2}{\alpha V_1} \int_0^t (1 - \frac{|x|}{\alpha})\,dx \to \frac{4t}{\alpha}$ as $t \to 0$. So the Lipschitz constant of $h * K_\alpha$ is at least $\frac{2}{\alpha}$. Hence, the Lipschitz constant given in Corollary 5-1 is optimal up to at most a factor of 2.

When $d = 1$ and K_α is the Epanechnikov kernel, $g(t) = \frac{4 \cdot 3}{2\alpha V_1} \int_0^t (1 - \frac{x^2}{\alpha^2})\,dx \to \frac{3t}{\alpha}$ as $t \to 0$. So the Lipschitz constant of $h * K_\alpha$ is at least $\frac{3}{2\alpha}$. Hence, the Lipschitz constant given in Corollary 5-2 is optimal up to at most a factor of 4.

5.3.5 Stable Computations in Practice

Suppose that we can compute $h(x)$ for values of x for which it is defined, we can sample from K, and that for a fixed $a \in \mathbb{R}^d$ we want to compute $g(a) = (h * K)(a) = \int_{\mathbb{R}^d} h(a-x)K(x)dx$. In practice, we will not be able to evaluate this integral analytically. We approximate $g(a)$ as follows. Let V be a random variable with probability distribution given by the kernel K, i.e.

$V \sim K$. Let W be the random variable given by $h(a-V)$. Then the expected value of W is given by $E[W] = \int_{\mathbb{R}^d} h(a-x)K(x)dx = g(a)$. We will approximate $E[W]$ by drawing a sample $\varepsilon_1, \ldots, \varepsilon_M$ where $\varepsilon_i \sim K$ are independent. Then $E[W]$ can be approximated by $\overline{W}_M = \frac{1}{M}\sum_{i=1}^{M} h(a-\varepsilon_i)$. By the law of large numbers, $\overline{W}_M \to E[W]$, where the convergence may be taken to be in probability (the weak law) or almost surely (the strong law). Let us record this result.

Theorem 5-4. *Let $a \in \mathbb{R}^d$ and $\varepsilon_1, \ldots, \varepsilon_M$ be drawn independently from K. Then*

$$\frac{1}{M}\sum_{i=1}^{M} h(a-\varepsilon_i) \to g(a).$$

5.3.6 Stability of the Choice of Kernel

As should be clear, the value of $(h*K)(a)$, for fixed h and a, will certainly depend on K. However, there is no fragility of output with respect to this choice, as shown by the following fact.

Theorem 5-5. *Let $h : \mathbb{R}^d \to \mathbb{R}$ be an essentially bounded function. Then the map $K \to h*K$ is Lipschitz from $L^1(\mathbb{R}^d)$ to $L^\infty(\mathbb{R}^d)$.*

Proof. Let $\phi : L^1(\mathbb{R}^d) \to L^\infty(\mathbb{R}^d)$ be given by $\phi(K) = h*K$. For $x \in \mathbb{R}^d$,

$$|\left[\phi(K) - \phi(K')\right](x)| \leq \int |h(x-t)|\,|K(t) - K'(t)|\,dt \leq \|h\|_\infty \|K - K'\|_1. \qquad \square$$

5.4 Experimental Considerations

5.4.1 Theoretical Considerations for Bandwidth Selection

After choosing a family of kernels, such as the Gaussian kernels K_α described in Section 5.3.3, the most important choice in implementing the method described here is the choice of bandwidth α. Choosing the amount of smoothing is a well-studied problem in nonparametric regression, where increasing the bandwidth decreases the estimation variance, but increases the squared bias. Both of these terms contribute to the error. A bandwidth which optimizes this trade-off may be estimated using cross-validation. A proper understanding of this problem in our situation requires analysis that goes beyond the scope of the present work.

However, we offer some heuristics for the choice of bandwidth. First, it may be chosen to obtain a desired amount of smoothness of $h * K_\alpha$. For example, we may want $h * K_\alpha$ to be 1-Lipschitz. Second, it seems reasonable to choose the bandwidth to (at least) equal the level of estimated noise of the input data. One may combine these two to find the minimum bandwidth that satisfies both requirements.

5.4.2 Practical Considerations for Bandwidth Selection

Our procedures depend on a free parameter, the bandwidth. For example, we chose a bandwidth of 3 in Section 5.2.1. Here we consider a slightly simpler example and consider the effect of varying the bandwidth.

We sample 1000 points uniformly from an annulus of inner and outer radius 20 and 40. Using Dionysus [50], we compute the 1-dimensional persistent homology of the alpha filtration of our sample and obtain a representative cycle for the longest bar. However, the embedded location of this cycle is unstable. We would like to quantify and visualize the uncertainty of this location. To do so, we consider a grid of squares with edge-length 1. We perturb the sampled points 1000 times by adding Gaussian noise and find the proportion of trials in which the representative cycle produced by Dionysus intersects each square. By performing this procedure simultaneously for every square in the grid, we obtain Figure 5-8. When the standard deviation of the Gaussian noise is very small, the representative cycle barely changes between perturbations, resulting in a small number of squares with high probability of intersecting the cycle. As the standard deviation increases, the picture becomes more diffuse.

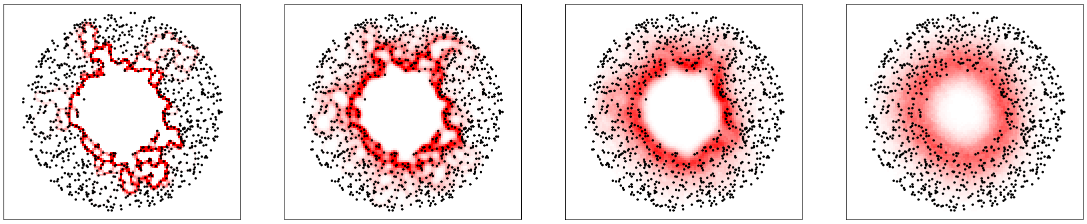

Figure 5-8. Bandwidth Effects. The probability of intersecting a representative cycle for each square in a grid. The bandwidths of the Gaussian kernel are 0.2, 1, 3, and 10. The color scale is given in Figure 5-2.

5.4.3 Possible Choices for the Location of the Generator in the Brain Imaging Data

Section 5.2.2 applies our method to real data. We can obtain the estimate of $h * K$ for the observed data in Section 5.2.2 for a ball of any radius by considering the distance from the location of the generator of the 28th longest bar in one of the iterations of the algorithm in Figure 1-3 to the location of the generator of the 28th longest bar in the observed data. The empirical cumulative distribution function of this distance is shown in Figure 5-9. We see that there is a competitor for the location of the generator of the 28th longest bar.

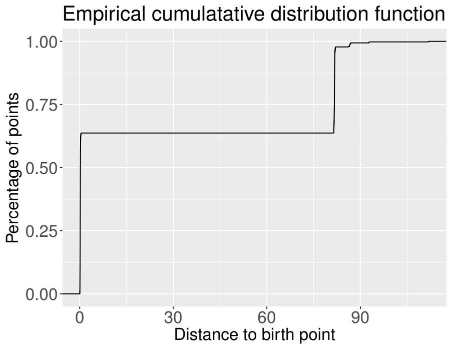

Figure 5-9. Empirical Cumulative Distribution Function of Distance to Generator.

5.4.4 Computational Considerations

Our methods applying Theorem 5-4 require repeated computation of persistence diagrams for similar filtrations. The computational cost may be considerable. In Sections 5.2.2 and 5.2.3 we repeat $M = 1000$ times. In Section 5.2.1 we repeat $10,000$ times. Note that we do not have convergence results at this time. For repeated persistent homology calculations it is important to have efficient software. In Section 5.2.1 we use Dionysus [50], in Section 5.2.2 we calculate persistent homology using a union-find data structure [33], and in Example 5.2.3 we use Perseus [53]

However, our methods are trivially parallelizable. With access to many cores, our repeated computations can be computed in parallel without increasing the running time. Note that for small perturbations, much of the persistent homology computation may be the same. In this case, there may be considerable computational savings by using vineyard updates [28]. Let us also remark that our methods combine nicely with subsampling, which is crucial for allowing persistent homology computations in the big data setting [20].

CHAPTER 6
CONCLUSIONS

The stability of persistence diagrams with Wasserstein metrics to changes in the input data motivates the search for stable and discriminative feature maps on persistence diagrams. In Chapter 3, we showed that the space of persistence diagrams with the p-Wasserstein metric does not admit a coarse embedding into a Hilbert space when $2 < p \leq \infty$ (Theorems 3-2 and 3-3). In other words, the distortion caused by a feature map to these Wasserstein metrics is not uniformly controllable. In fact, even if one restricts to the subspace of (finite) persistence diagrams arising as the homology of a filtered finite simplicial complex, there still does not exist a coarse embedding of this subspace into a Hilbert space. The proof when $p = \infty$ involves embedding a construction of Dranishnikov et al. [32] and Enflo [34]. The proof when $2 < p < \infty$ makes use of a characterization of coarse embeddability into Hilbert spaces due to Nowak [55] to show a coarse embedding of persistence diagrams would imply a coarse embedding of ℓ_p, contradicting a theorem of Johnson and Randrianarivony [42]. In future work, we will investigate whether or not persistence diagrams with the p-Wasserstein metric admit a coarse embedding when $p = 1, 2$.

We address the question of approximating persistent homology in Chapter 4. We investigate three approximation techniques that replace a filtration of a complex with a nearby filtration that allows for a greater number of cell reductions using the filtered discrete Morse theory of Mischaikow and Nanda [49]. The main theoretical result of this chapter is Theorem 4-1, which shows that the optimal solution to a certain combinatorial problem induces a nearby filtration that is optimal over a restricted search space. We compared the approximation algorithm associated to Theorem 4-1 to a simple binning and to relaxation of the original algorithm of Mischaikow and Nanda [49].

In Chapter 5, we introduced a general technique for stabilizing unstable information produced during persistent homology computations. Examples of such instability were given in Section 5.1. Our approach is to encode the unstable information as a discontinuous but measurable real-valued function and stabilize it via convolution with a kernel function. The Lipschitz constants for the stabilized function are given in Corollaries 5-1, 5-2, and 5-3. This procedure is stable to the amount of smoothing done (Theorem 5-5), and the value of the stabilized function can be approximated via simulation (Theorem 5-4). The procedure is applied to synthetic and real datasets in Section 5.2.

REFERENCES

[1] Henry Adams, Atanas Atanasov, and Gunnar Carlsson, *Nudged elastic band in topological data analysis*, Topol. Methods Nonlinear Anal. **45** (2015), no. 1, 247–272. MR 3365014

[2] Henry Adams, Tegan Emerson, Michael Kirby, Rachel Neville, Chris Peterson, Patrick Shipman, Sofya Chepushtanova, Eric Hanson, Francis Motta, and Lori Ziegelmeier, *Persistence images: a stable vector representation of persistent homology*, J. Mach. Learn. Res. **18** (2017), Paper No. 8, 35. MR 3625712

[3] Mahmuda Ahmed, Brittany Terese Fasy, and Carola Wenk, *Local persistent homology based distance between maps*, SIGSPATIAL, ACM, Nov. 2014.

[4] Greg Bell, Austin Lawson, C. Neil Pritchard, and Dan Yasaki, *The space of persistence diagrams fails to have Yu's property A*, arXiv e-prints (2019), arXiv:1902.02288.

[5] P. Bendich, S. P. Chin, J. Clark, J. Desena, J. Harer, E. Munch, A. Newman, D. Porter, D. Rouse, N. Strawn, and A. Watkins, *Topological and statistical behavior classifiers for tracking applications*, IEEE Transactions on Aerospace and Electronic Systems **52** (2016), no. 6, 2644–2661.

[6] Paul Bendich, Peter Bubenik, and Alexander Wagner, *Stabilizing the unstable output of persistent homology computations*, Journal of Applied and Computational Topology (2019).

[7] Paul Bendich, J. S. Marron, Ezra Miller, Alex Pieloch, and Sean Skwerer, *Persistent homology analysis of brain artery trees*, Ann. Appl. Stat. **10** (2016), no. 1, 198–218. MR 3480493

[8] Christian Berg, Jens Peter Reus Christensen, and Paul Ressel, *Harmonic analysis on semigroups: Theory of positive definite and related functions*, Graduate Texts in Mathematics, vol. 100, Springer-Verlag, New York, 1984. MR 747302

[9] Andrew J. Blumberg, Itamar Gal, Michael A. Mandell, and Matthew Pancia, *Robust statistics, hypothesis testing, and confidence intervals for persistent homology on metric measure spaces*, Found. Comput. Math. **14** (2014), no. 4, 745–789. MR 3230014

[10] Peter Bubenik, *Statistical topological data analysis using persistence landscapes*, J. Mach. Learn. Res. **16** (2015), 77–102. MR 3317230

[11] Peter Bubenik, Gunnar Carlsson, Peter T. Kim, and Zhi-Ming Luo, *Statistical topology via Morse theory persistence and nonparametric estimation*, Algebraic methods in statistics and probability II, Contemp. Math., vol. 516, Amer. Math. Soc., Providence, RI, 2010, pp. 75–92. MR 2730741

[12] Peter Bubenik and Tane Vergili, *Topological spaces of persistence modules and their properties*, J. Appl. Comput. Topol. **2** (2018), no. 3-4, 233–269. MR 3927353

[13] Gunnar Carlsson, *Topology and data*, Bull. Amer. Math. Soc. (N.S.) **46** (2009), no. 2, 255–308. MR 2476414

[14] Gunnar Carlsson, Tigran Ishkhanov, Vin de Silva, and Afra Zomorodian, *On the local behavior of spaces of natural images*, Int. J. Comput. Vis. **76** (2008), no. 1, 1–12. MR 3715451

[15] M. Carriere, S. Y. Oudot, and M. Ovsjanikov, *Stable topological signatures for points on 3d shapes*, Proc. Sympos. on Geometry Processing, 2015.

[16] Mathieu Carrière and Ulrich Bauer, *On the metric distortion of embedding persistence diagrams into separable Hilbert spaces*, 35th International Symposium on Computational Geometry, LIPIcs. Leibniz Int. Proc. Inform., vol. 129, Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2019, pp. Art. No. 21, 15. MR 3968607

[17] Frédéric Chazal, David Cohen-Steiner, Marc Glisse, Leonidas J. Guibas, and Steve Y. Oudot, *Proximity of persistence modules and their diagrams*, Proceedings of the Twenty-fifth Annual Symposium on Computational Geometry (New York, NY, USA), SCG '09, ACM, 2009, pp. 237–246.

[18] Frédéric Chazal, David Cohen-Steiner, and Quentin Mérigot, *Geometric Inference for Measures based on Distance Functions*, Foundations of Computational Mathematics **11** (2011), no. 6, 733–751 (Anglais), RR-6930 RR-6930.

[19] Frédéric Chazal, Brittany Fasy, Fabrizio Lecci, Bertrand Michel, Alessandro Rinaldo, and Larry Wasserman, *Robust topological inference: distance to a measure and kernel distance*, J. Mach. Learn. Res. **18** (2017), Paper No. 159, 40. MR 3813808

[20] Frédéric Chazal, Brittany Terese Fasy, Fabrizio Lecci, Bertrand Michel, Alessandro Rinaldo, and Larry Wasserman, *Subsampling methods for persistent homology*, Proceedings of the 32nd International Conference on Machine Learning, Lille, France, vol. 37, JMLR: W&CP, 2015.

[21] Frédéric Chazal, Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, Aarti Singh, and Larry Wasserman, *On the bootstrap for persistence diagrams and landscapes*, Modeling and Analysis of Information Systems **20** (2014), no. 6, 96–105.

[22] Frédéric Chazal, Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, and Larry Wasserman, *Stochastic convergence of persistence landscapes and silhouettes*, J. Comput. Geom. **6** (2015), no. 2, 140–161. MR 3323391

[23] Yen-Chi Chen, Daren Wang, Alessandro Rinaldo, and Larry Wasserman, *Statistical Analysis of Persistence Intensity Functions*, arXiv e-prints (2015), arXiv:1510.02502.

[24] Moo K. Chung, Peter Bubenik, and Peter T. Kim, *Persistence diagrams in cortical surface data*, Information Processing in Medical Imaging (IPMI) 2009, Lecture Notes in Computer Science, vol. 5636, 2009, pp. 386–397.

[25] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer, *Stability of persistence diagrams*, Discrete Comput. Geom. **37** (2007), no. 1, 103–120. MR 2279866

[26] ______, *Extending persistence using Poincaré and Lefschetz duality*, Found. Comput. Math. **9** (2009), no. 1, 79–103. MR 2472288

[27] David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Yuriy Mileyko, *Lipschitz functions have L_p-stable persistence*, Found. Comput. Math. **10** (2010), no. 2, 127–139. MR 2594441

[28] David Cohen-Steiner, Herbert Edelsbrunner, and Dmitriy Morozov, *Vines and vineyards by updating persistence in linear time*, Computational geometry (SCG'06), ACM, New York, 2006, pp. 119–126. MR 2389318

[29] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein, *Introduction to algorithms*, third ed., MIT Press, Cambridge, MA, 2009. MR 2572804

[30] Tamal K. Dey, Fengtao Fan, and Yusu Wang, *Computing topological persistence for simplicial maps [extended abstract]*, Computational geometry (SoCG'14), ACM, New York, 2014, pp. 345–354. MR 3382315

[31] Tamal K. Dey and Rephael Wenger, *Stability of critical points with interval persistence*, Discrete Comput. Geom. **38** (2007), no. 3, 479–512. MR 2352705

[32] A. N. Dranishnikov, G. Gong, V. Lafforgue, and G. Yu, *Uniform embeddings into Hilbert space and a question of Gromov*, Canad. Math. Bull. **45** (2002), no. 1, 60–70. MR 1884134

[33] Herbert Edelsbrunner and John L. Harer, *Computational topology*, American Mathematical Society, Providence, RI, 2010, An introduction. MR 2572029

[34] Per Enflo, *On a problem of Smirnov*, Ark. Mat. **8** (1969), 107–109. MR 0415576

[35] Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, Larry Wasserman, Sivaraman Balakrishnan, and Aarti Singh, *Confidence sets for persistence diagrams*, Ann. Statist. **42** (2014), no. 6, 2301–2339. MR 3269981

[36] Robin Forman, *Morse theory for cell complexes*, Adv. Math. **134** (1998), no. 1, 90–145. MR 1612391

[37] D. H. Fremlin, *Measure theory. Vol. 4*, Torres Fremlin, Colchester, 2006, Topological measure spaces. Part I, II, Corrected second printing of the 2003 original. MR 2462372

[38] P. Frozini and B. Landi, *Size theory as a topological tool for computer vision*, Pattern Recognition and Image Analysis **9** (1999), 596–603.

[39] Marcio Gameiro, Yasuaki Hiraoka, Shunsuke Izumi, Miroslav Kramar, Konstantin Mischaikow, and Vidit Nanda, *A topological measurement of protein compressibility*, Jpn. J. Ind. Appl. Math. **32** (2015), no. 1, 1–17. MR 3318898

[40] Robert Ghrist, *Barcodes: the persistent topology of data*, Bull. Amer. Math. Soc. (N.S.) **45** (2008), no. 1, 61–75. MR 2358377

[41] M. Gromov, *Asymptotic invariants of infinite groups*, Geometric group theory, Vol. 2 (Sussex, 1991), London Math. Soc. Lecture Note Ser., vol. 182, Cambridge Univ. Press, Cambridge, 1993, pp. 1–295. MR 1253544

[42] William B. Johnson and N. Lovasoa Randrianarivony, l_p $(p > 2)$ *does not coarsely embed into a Hilbert space*, Proc. Amer. Math. Soc. **134** (2006), no. 4, 1045–1050. MR 2196037

[43] Violeta Kovacev-Nikolic, Peter Bubenik, Dragan Nikolić, and Giseon Heo, *Using persistent homology and dynamical distances to analyze protein binding*, Stat. Appl. Genet. Mol. Biol. **15** (2016), no. 1, 19–38. MR 3464008

[44] Solomon Lefschetz, *Algebraic Topology*, American Mathematical Society Colloquium Publications, v. 27, American Mathematical Society, New York, 1942. MR 0007093

[45] C. J. Lennard, A. M. Tonge, and A. Weston, *Generalized roundness and negative type*, Michigan Math. J. **44** (1997), no. 1, 37–45. MR 1439667

[46] Chunyuan Li, M. Ovsjanikov, and F. Chazal, *Persistence-based structural recognition*, Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, June 2014, pp. 2003–2010.

[47] Yuriy Mileyko, Sayan Mukherjee, and John Harer, *Probability measures on the space of persistence diagrams*, Inverse Problems **27** (2011), no. 12, 124007, 22. MR 2854323

[48] Nikola Milosavljević, Dmitriy Morozov, and Primož Škraba, *Zigzag persistent homology in matrix multiplication time*, Computational geometry (SCG'11), ACM, New York, 2011, pp. 216–225. MR 2919613

[49] Konstantin Mischaikow and Vidit Nanda, *Morse theory for filtrations and efficient computation of persistent homology*, Discrete Comput. Geom. **50** (2013), no. 2, 330–353. MR 3090522

[50] Dimitriy Morozov, *Dionysus: a C++ library with various algorithms for computing persistent homology*, Software available at http://www.mrzv.org/software/dionysus/, 2012.

[51] Elizabeth Munch, Katharine Turner, Paul Bendich, Sayan Mukherjee, Jonathan Mattingly, and John Harer, *Probabilistic Fréchet means for time varying persistence diagrams*, Electron. J. Stat. **9** (2015), no. 1, 1173–1204. MR 3354335

[52] James R. Munkres, *Elements of algebraic topology*, Addison-Wesley Publishing Company, Menlo Park, CA, 1984. MR 755006

[53] Vidit Nanda, *Perseus: the persistent homology software*, Software available at http://www.math.rutgers.edu/~vidit/perseus/index.html, 2013.

[54] P. Niyogi, S. Smale, and S. Weinberger, *A topological view of unsupervised learning from noisy data*, SIAM J. Comput. **40** (2011), no. 3, 646–663. MR 2810909

[55] Piotr W. Nowak, *Coarse embeddings of metric spaces into Banach spaces*, Proc. Amer. Math. Soc. **133** (2005), no. 9, 2589–2596. MR 2146202

[56] Steve Y. Oudot, *Persistence theory: from quiver representations to data analysis*, Mathematical Surveys and Monographs, vol. 209, American Mathematical Society, Providence, RI, 2015. MR 3408277

[57] Jan Reininghaus, Stefan Huber, Ulrich Bauer, and Roland Kwitt, *A stable multi-scale kernel for topological machine learning*, Proc. CVPR (2015), 4741–4748.

[58] John Roe, *Lectures on coarse geometry*, University Lecture Series, vol. 31, American Mathematical Society, Providence, RI, 2003. MR 2007488

[59] I. J. Schoenberg, *Remarks to Maurice Fréchet's article "Sur la définition axiomatique d'une classe d'espace distanciés vectoriellement applicable sur l'espace de Hilbert" [MR1503246]*, Ann. of Math. (2) **36** (1935), no. 3, 724–732. MR 1503248

[60] ______, *Metric spaces and positive definite functions*, Trans. Amer. Math. Soc. **44** (1938), no. 3, 522–536. MR 1501980

[61] Donald R. Sheehy, *Linear-size approximations to the Vietoris-Rips filtration*, Discrete Comput. Geom. **49** (2013), no. 4, 778–796. MR 3068574

[62] B. W. Silverman, *Density estimation for statistics and data analysis*, Monographs on Statistics and Applied Probability, Chapman & Hall, London, 1986. MR 848134

[63] Abraham Smith, Paul Bendich, John Harer, Alex Pieloch, and Jay Hineman, *Supervised Learning of Labeled Pointcloud Differences via Cover-Tree Entropy Reduction*, arXiv e-prints (2017), arXiv:1702.07959.

[64] Ingo Steinwart and Andreas Christmann, *Support vector machines*, Information Science and Statistics, Springer, New York, 2008. MR 2450103

[65] A. W. Tucker, *Cell spaces*, Ann. of Math. (2) **37** (1936), no. 1, 92–100. MR 1503271

[66] Katharine Turner and Gard Spreemann, *Same but Different: distance correlations between topological summaries*, arXiv e-prints (2019), arXiv:1903.01051.

[67] M. P. Wand and M. C. Jones, *Kernel smoothing*, Monographs on Statistics and Applied Probability, vol. 60, Chapman and Hall, Ltd., London, 1995. MR 1319818

[68] Shmuel Weinberger, *The complexity of some topological inference problems*, Found. Comput. Math. **14** (2014), no. 6, 1277–1285. MR 3273679

[69] J. H. Wells and L. R. Williams, *Embeddings and extensions in analysis*, Springer-Verlag, New York-Heidelberg, 1975, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 84. MR 0461107

[70] Guoliang Yu, *The coarse Baum-Connes conjecture for spaces which admit a uniform embedding into Hilbert space*, Invent. Math. **139** (2000), no. 1, 201–240. MR 1728880

[71] Afra Zomorodian and Gunnar Carlsson, *Localized homology*, Comput. Geom. **41** (2008), no. 3, 126–148. MR 2442490

BIOGRAPHICAL SKETCH

Alexander received his bachelor’s and master’s degrees from Vanderbilt University. He received his doctorate in mathematics from the University of Florida in 2020.