Approximating Continuous Functions on Persistence Diagrams Using Template Functions

Jose A. Perea   joperea@msu.edu
Elizabeth Munch   muncheli@egr.msu.edu
Department of Computational Mathematics, Science, and Engineering; and
Department of Mathematics

Firas A. Khasawneh   khasawn3@egr.msu.edu
Department of Mechanical Engineering

Michigan State University
East Lansing, MI 48824, USA

Abstract

The persistence diagram is an increasingly useful tool from Topological Data Analysis, but its use alongside typical machine learning techniques requires mathematical finesse. The most success to date has come from methods that map persistence diagrams into $\mathbb{R}^n$, in a way which maximizes the structure preserved. This process is commonly referred to as *featurization*. In this paper, we describe a mathematical framework for featurization using template functions. These functions are general as they are only required to be continuous and compactly supported. We discuss two realizations: tent functions, which emphasize the local contributions of points in a persistence diagram, and interpolating polynomials, which capture global pairwise interactions. We combine the resulting features with classification and regression algorithms on several examples including shape data and the Rossler system. Our results show that using template functions yields high accuracy rates that match and often exceed those of existing featurization methods. One counter-intuitive observation is that in most cases using interpolating polynomials, where each point contributes globally to the feature vector, yields significantly better results than using tent functions, where the contribution of each point is localized. Along the way, we provide a complete characterization of compactness in the space of persistence diagrams.

Keywords: Topological Data Analysis, Persistent Homology, Machine Learning, Featurization, Bottleneck Distance

1. Introduction

Many machine learning tasks can be reduced to the following problem: Approximate a continuous function defined on a topological space, the "ground truth," given the function values (or approximations thereof) on some subset of the points. This task has been well studied for data sitting in Euclidean space; however, more work is necessary to extend these ideas to arbitrary topological spaces. In this paper, we focus on the task of classification and regression on the space of persistence diagrams endowed with the bottleneck distance, $(\mathcal{D}, d_B)$. These objects arise in the field of Topological Data Analysis (TDA) as signatures

giving insight into the underlying structure of a data set. The issue is that the geometry of $(\mathcal{D}, d_B)$ is not directly amenable to the application of existing machine learning theories.

Existing methods for applying statistics and machine learning methods to persistence diagrams can be loosely divided into two categories. The first attempts to work in the space of persistence diagrams directly. For example, Mileyko et al. (2011), Turner et al. (2014), and Munch et al. (2015) investigate the Fréchet mean for collections of diagrams. However, $\mathcal{D}$ is not Cat(0), and non-uniqueness of geodesics leads to non-uniqueness of the Fréchet mean, even when looking at something as simple as the mean of a set of diagrams. Another option within the first category of methods is given by Fasy et al. (2014) which derived confidence sets for persistence diagrams. This enables the separation of points with small persistence, commonly considered to be topological noise, from points with long persistence which are considered topological signals. However, that theory does not immediately extend to investigating collections of diagrams arising from unrelated point clouds. Finally, Li et al. (2014) use the Wasserstein and bottleneck distances on persistence diagrams in combination with distance metric learning for classification tasks.

The second collection of methods maps the space of persistence diagrams into another, more well-behaved space where available mathematical machinery can be readily applied. Our work fits into this category. We next give an overview of the extensive list of available featurization methods.

Adcock et al. (2016) identify an appropriate subring of algebraic (i.e. polynomial) functions on persistence diagrams, and a convenient system of free generators. Unfortunately the resulting functions are not continuous with respect to the bottleneck distance, and in particular do not extend from the space of finite diagrams, denoted $\mathcal{D}_0$, to the case of infinite diagrams $\mathcal{D}$. Infinite diagrams are relevant, for instance, in capturing the shape of spaces with fractal structure. Carlsson and Verovsek (2016), and Kališnik (2018) propose a change in the algebraic structure using ideas from tropical geometry, and describe a collection of rational tropical functions which are Lipschitz continuous on $\mathcal{D}_0$, separate points, and are generated by a countable set of tropical polynomials. Lipschitz continuity implies that these rational tropical functions have unique continuous extensions to $\mathcal{D}$, but it is unclear if the resulting collection is dense in $C(\mathcal{D}, \mathbb{R})$, the space of continuous real-valued functions on $\mathcal{D}$. Further, Kališnik (2018) points out that the resulting vectors can be very high-dimensional and it is not clear how to automatically select the needed weights of the coordinate functions.

Di Fabio and Ferri (2015) explored obtaining an algebraic representation of persistence diagrams using complex polynomials. These polynomials are obtained through mapping the persistence diagrams to the complex plane, and then representing the original persistence diagram using polynomials whose roots are the mapped, complex points. This representation bypasses the need for the computationally expensive bottleneck distances between diagrams in favor of more tractable metrics that utilize the polynomials' coefficients. Di Fabio and Ferri (2015) propose using this representation as a preprocessing step to eliminate objects that are too far, thus reducing the set of objects for which a more accurate bottleneck distance needs to be computed. As an example, Di Fabio and Ferri (2015) apply two filtering functions to a database of 3D-surface mesh models. They then obtain the corresponding 0-dimensional persistence diagrams, and use the distances between their complex vector representations to construct precision/recall graphs.

Functional summaries—which map persistence diagrams into continuous functions—can be used to featurize the persistence diagrams space (Berry et al. (2018)). Examples of these functional summaries include arguably the most commonly used featurization method, the persistence landscape of Bubenik (2015), which turns the persistence diagram into a function $\mathbb{N} \times \mathbb{R} \to \mathbb{R}$. Bubenik (2015) showed that persistence landscapes provide a 1-Lipschitz embedding of $\mathcal{D}$ into the Banach space $L^\infty(\mathbb{N} \times \mathbb{R})$, and Chazal et al. (2014) investigated the statistical properties of landscapes such as weak convergence and convergence of the bootstrap. For certain subsets of $\mathcal{D}$, one gets embeddings into $L^p(\mathbb{N} \times \mathbb{R})$ for $1 \leq p < \infty$, but the value of p depends on the particular subset. The closely related generalized landscape functions and weighted silhouettes are given in Chazal et al. (2014), and the multiscale version is given in Padellini and Brutti (2017).

Another functional summary, the persistence images of Adams et al. (2017) (closely related to the persistence intensity functions of Chen et al. (2015)) turn the persistence diagram into a sum of Gaussians centered at the points of the diagram, and then take a histogram of the image for featurization. This is closely related to the work of Donatini et al. (1998) and Ferri et al. (1998) doing the same on the size function, the precursor to persistence. Persistence images are also related to a less stable version based on binning the persistence diagram given by Rouse et al. (2015).

Chevyrev et al. (2018) utilized persistence paths and signature features to obtain a feature map for statistical learning. They constructed a feature map by composing a persistence path embedding of the persistence diagram with the path signature features. One example of persistence path embedding is the persistence landscape of Bubenik (2015). By composing the persistence path with its signature features, Chevyrev et al. (2018) were able to bypass the difficulty in choosing a feature map for paths, which is a critical but generally is not an obvious choice that determines the statistical learning guarantees.

Reproducing Kernel Hilbert Space approaches have also been pursued. In particular, the persistence scale space kernel given by Reininghaus et al. (2015) is defined by treating the persistence diagram as a sum of Dirac deltas at each point in the persistence diagram (plus negative Dirac deltas for the mirror images across the diagonal), and using this as the initial condition for a heat diffusion problem. They further provide a closed form solution for this kernel. Kwitt et al. (2015) modify this formulation to provide a universal kernel. The persistence weighted Gaussian kernel is proposed in Kusano et al. (2016, 2017); Kusano (2018) which generalizes the persistence scale space kernel to accept a weighting function allowing for tuning the importance of persistent classes. Carrire et al. (2017) give a related kernel that is stable with respect to an approximation of the Wasserstein distance. These kernels have been extended to multiparameter persistent homology by Corbet et al. (2018). Mathematical properties of these kernels were studied in Carrière and Bauer (2018).

Another class of kernels based on Riemannian geometry have been proposed (Anirudh et al. (2016); Le and Yamada (2018)) by similarly treating the persistence diagram as a sum of Dirac delta masses. Kernels defined on the persistence landscapes of Bubenik (2015) are given in Zhu et al. (2016).

Beyond these, quite a few ad hoc methods of featurization exist in the literature. One such example is Bendich et al. (2016), who extracted a vector-based summary of sorted lifetimes from 0- and 1-dimensional persistence diagrams of brain artery trees. Singh et al. (2014) used a similar approach where they used as features the first few sorted persistence

lifetimes derived from treating the cell nuclei in histology images as point clouds. Chung et al. (2009) computed 0, 1, and 2-d persistence for cortical surface data, and obtained a pairing concentration by fixing a circle of a certain radius at a point in the persistence diagram and then computing the number of pairings within that circle. Pachauri et al. (2011) used kernel density estimation (KDE) to approximate the concentration of points in the persistence diagram while ignoring the diagonal behavior, and used a KDE similarity measure to perform several tasks including kernel regression and classification. Carrière et al. (2015) featurize a persistence diagram by encoding the following. For any two points in the persistence diagram the minimum of (i) the separation ℓ^∞ distance between the two points and (ii) the points' ℓ^∞ distances to the diagonal is computed. The log of the obtained values (in descending order) are truncated and used as a vector signature. Finally, Zieliński et al. (2018a,b) cluster points within the persistence diagrams themselves to utilize a bag of words approach.

1.1 Our contribution

Mathematically, we are working with the following framework. Let $\mathcal{D}$ be the space of infinite (also known as generalized) persistence diagrams with topology induced by the bottleneck distance d_B. Suppose one has a set $\mathcal{S} \subset \mathcal{D}$, typically compact, and a continuous function $F : \mathcal{S} \longrightarrow \mathbb{R}$. The problem at hand is to devise provably-correct and computationally feasible approaches to approximating F, given a finite sample $D_1, \ldots, D_n \in \mathcal{S}$ and their values $F(D_1), \ldots, F(D_n) \in \mathbb{R}$. This encompasses, for instance, regression and classification tasks from labeled training data. The problem at hand is thus (1) to characterize compactness in $(\mathcal{D}, d_B)$, (2) to construct dense subsets of the space of continuous functions from $\mathcal{D}$ to $\mathbb{R}$, and (3) to devise algorithms using said families to approximate real valued functions on compact subsets of $\mathcal{D}$.

The first contribution of this paper is a complete characterization of (relative) compactness for subsets of $\mathcal{D}$ with respect to d_B (see Figure 2 and Theorem 11). This provides an actionable criterion with which to determine the problem-specific relevance for our approximation framework. Our characterization also comes with some unexpected consequences: (1) Every compact subset of $\mathcal{D}$ has empty interior (hence $\mathcal{D}$ is not locally compact); (2) $\mathcal{D}$ cannot be written as a countable union of compact subsets; and if $C(\mathcal{D}, \mathbb{R})$ denotes the set of continuous functions from $\mathcal{D}$ to $\mathbb{R}$, then (3) the compact-open topology on $C(\mathcal{D}, \mathbb{R})$—which captures approximations on compact subsets of $\mathcal{D}$—is not metrizable.

Next, we turn our attention to the problem of finding dense subsets of $C(\mathcal{D}, \mathbb{R})$ with respect to the compact-open topology. Ideally, the elements of these sets should be succinctly represented (e.g., with a few parameters) and efficiently searched (e.g., via appropriate optimization schemes), in order to devise general computational schemes. Our second contribution is a methodology for constructing infinitely many examples of said families (see Theorems 29 and 30). The strategy we employ goes as follows: We continuously embed $\mathcal{D}$ in an appropriate topological vector space V—in a way which extends the monoidal structure of $\mathcal{D}$ given by disjoint union of multisets (see Theorem 26)—and then restrict the continuous $\mathbb{R}$-linear maps on V to yield continuous real-valued functions on $\mathcal{D}$. We show that these maps coincide exactly (see Theorem 27 and discussion thereafter) with taking compactly supported continuous function from $\mathbb{W} = \{(x, y) \in \mathbb{R}^2 : y > x \geq 0\}$ to $\mathbb{R}$, and integrating

them against the measure associated to each persistence diagram. It is in this sense that compactly supported maps $f : \mathbb{W} \longrightarrow \mathbb{R}$ are interpreted as *template functions*. We then provide a procedure (Theorem 30) for constructing countable sets of template functions, such that the algebra they generate in $C(\mathcal{D}, \mathbb{R})$ is dense with respect to the compact-open topology (Theorem 29).

As the final contribution of this paper, we provide two explicit families of template functions—called respectively tent functions and interpolating polynomials (see Section 6)—so that the algebras they generate in $C(\mathcal{D}, \mathbb{R})$ are dense. We then provide algorithms to perform regularized regression and classification using template functions (in Section 7), and finally, we compare tent functions and interpolating polynomials in several tasks including shape classification and inference in dynamical systems (Section 8).

1.2 Outline

We go over the background needed for understanding persistence diagrams in Section 2. In Section 3, we give a full characterization of compact sets in persistence diagram space with the bottleneck distance (Theorem 11). We provide the mathematical justification for the template functions in Section 4, and fit this into a function approximation scheme in Section 5. In Section 6 we give two options for template functions: tent functions and (Chebyshev) interpolating polynomials. In Section 7 we fit these into a regression framework. We give results of our experiments in Section 8 and discuss implications and future directions in Section 9.

2. Basics

Traditionally, persistence diagrams arise in the course of the following procedure. Given a real valued function on a topological space $f : \mathbb{X} \to \mathbb{R}$, denote the sublevel set by $\mathbb{X}_a = f^{-1}(-\infty, a]$. For example, given a point cloud $\chi \subseteq \mathbb{R}^d$, this function could be defined to be $f : \mathbb{R}^d \to \mathbb{R}$, $y \to \inf_{x \in \chi} \|x - y\|$. A function induces a filtration of the space, $\mathbb{X}_a \subseteq \mathbb{X}_b$ for all $a \leq b$. Applying k-dimensional homology to this filtration yields the k-dimensional persistence module $(H_k(\mathbb{X}_a), \phi_a^b)$, namely the collection of vector spaces[1] $H_k(X_a)$ with induced maps $\phi_a^b : H_k(X_a) \to H_k(X_b)$ for all $a \leq b$. In full generality, persistence modules can simply be viewed as a collection of vector spaces and linear maps $\mathcal{V} = (V_a, \phi_a^b)$ where $\phi_a^b : V_a \to V_b$, $\phi_a^a = \mathbb{1}_{U_a}$, and $\phi_b^c \phi_a^b = \phi_a^c$.

Given a sufficiently well-behaved function, this persistence module can be decomposed uniquely. The pieces of the decomposition are *interval modules*, defined to be a persistence module $\mathcal{I}_{[r,s)} = (I_a, i_a^b)$ where $I_a = \mathbf{k}$ if $a \in [r, s)$ and 0 otherwise. The maps i_a^b are identities whenever possible; i.e., when $r \leq a \leq b < s$. A persistence module $\mathcal{V} = (V_a, \phi_a^b)$ is called q-tame if the rank of ϕ_a^b is finite for all $a < b$. Every q-tame persistence module decomposes uniquely as a direct sum of interval modules, $\mathcal{V} = \bigoplus_{[r,s) \in \mathcal{A}} \mathcal{I}_{[r,s)}$. This decomposition is often visualized as a persistence diagram as seen in center of Fig. 1. The diagram consists of a point at $(r, s) \in \mathbb{R}^2$ for each $[r, s) \in \mathcal{A}$. The diagonal is drawn on the diagram to indicate that all points in the diagram are above it. In this paper, we will also work with the birth-

---

1. The restriction from group to vector space occurs because we assume that we compute homology with coefficients in a field $\mathbf{k}$.
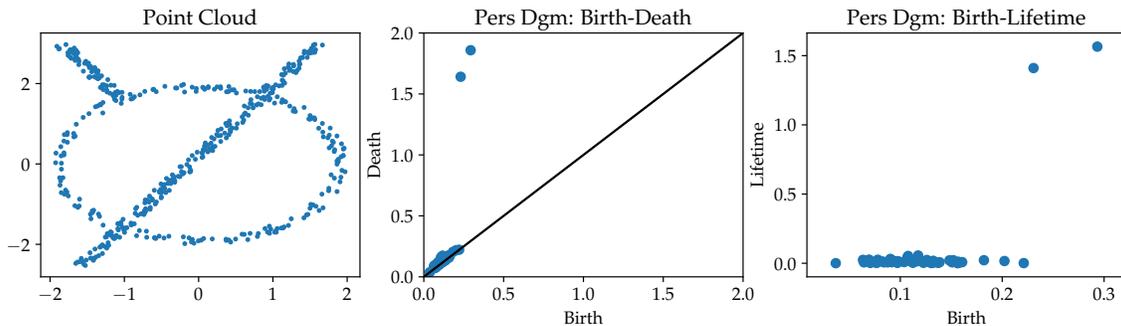
Figure 1: An example point cloud is shown at left, with its persistence diagram shown in the middle. At right, we show the conversion of the persistence diagram into the birth-lifetime plane which is used throughout this paper.

lifetime visualization of the persistence diagram, consisting of a point at $(r, s - r) \in \mathbb{R}^2$ for each $[r, s) \in \mathcal{A}$. See the right of Fig. 1 for an example.

2.1 The space of persistence diagrams

A persistence diagram D, then, can be thought of as a collection of points

$$S \subset \{(x, y) \in \mathbb{R}^2 : 0 \leq x < y\}$$

with a notion of multiplicity, which we write as a function

$$\mu : S \to \mathbb{N} = \{1, 2, \ldots\}.$$

For this reason, we will often write $D = (S, \mu)$ for a persistence diagram. However, in order to make statements about the structure of the space of persistence diagrams, we need to put a few restrictions on these collections.

Definition 1 *Given $D = (S, \mu)$ and $U \subset \mathbb{R}^2$, the **multiplicity of** D **in** U is*

$$\mathsf{Mult}(D, U) = \begin{cases} \sum_{\mathbf{x} \in S \cap U} \mu(\mathbf{x}) & \text{if this is finite, or} \\ \infty & \text{else.} \end{cases}$$

Some other common notions for persistence diagrams which we will use repeatedly are as follows. The diagonal is denoted $\Delta := \{(x, x) \in \mathbb{R}^2\}$. The wedge is denoted $\mathbb{W} := \{(x, y) \in \mathbb{R}^2 : 0 \leq x < y\}$. Note that the boundary of $\mathbb{W}$ is Δ, but it is not included in $\mathbb{W}$. The persistence of a point $\mathbf{x} = (x, y) \in \mathbb{W}$ is $\mathsf{pers}(\mathbf{x}) = y - x$. The portion of the wedge with persistence greater than ϵ is denoted

$$\mathbb{W}^\epsilon := \{\mathbf{x} \in \mathbb{W} \mid \mathsf{pers}(\mathbf{x}) > \epsilon\}.$$

6

Note that the lower boundary is not included. When we wish to include the boundary, we write

$$\overline{\mathbb{W}^\epsilon} = \{\mathbf{x} \in \mathbb{W} : \mathsf{pers}(\mathbf{x}) \geq \epsilon\}.$$

If we want to just work with the portion of the points of a diagram $D = (S, \mu)$ in a given region $U \subseteq \mathbb{R}^2$, we write

$$D \cap U := (S \cap U, \mu|_{S \cap U}).$$

We further abuse notation by writing $D \subset U$ if $S \subset U$. If $S = \emptyset$, we follow the convention $\mu = \emptyset$ and denote by $\varnothing = (\emptyset, \emptyset)$ the resulting (empty) persistence diagram. Sometimes, we think of a persistence diagram $D = (S, \mu)$ instead as a set S_μ, obtained by replicating the elements of S and decorating them with integer labels:

$$S_\mu := \big\{(\mathbf{x}, k) : \mathbf{x} \in S \text{ and } 1 \leq k \leq \mu(\mathbf{x})\big\}. \tag{1}$$

For the sake of figures, we sometimes plot persistence diagrams in the birth-lifetime plane. That is, we draw the point $\mathbf{x} = (x, y) \in \mathbb{W}$ at the point $(x, y - x) = (x, \mathsf{pers}(\mathbf{x}))$. In this representation, Δ gets mapped to the x-axis. It should be noted that this transformation is different from the rotation used by Bubenik (2015) and Adams et al. (2017), but there is no reason to prefer one over the other.

With this notation, we define the set of persistence diagrams as follows.

Definition 2 *The **space of persistence diagrams**, denoted $\mathcal{D}$, is the collection of pairs $D = (S, \mu)$ where:*

1. *$S \subset \mathbb{W}$, called the **underlying set of** D, so that for any $\epsilon > 0$, $\mathsf{Mult}(D, \mathbb{W}^\epsilon)$ is finite.*

2. *μ is a function from S to the set of natural numbers $\mathbb{N} = \{1, 2, \ldots\}$. In particular, $\mu(\mathbf{x}) \in \mathbb{N}$ is the **multiplicity** of $\mathbf{x} \in S$.*

*The **space of finite persistence diagrams** is $\mathcal{D}_0 := \{(S, \mu) \in \mathcal{D} : S \text{ is finite}\}$.*

Note that the finite persistence diagrams, $\mathcal{D}_0$, were the first to appear in the literature and many current papers implicitly assume finiteness.

Since we will be interested in studying subsets of $\mathcal{D}$, we will extend Definition 1 as follows.

Definition 3 *Given $\mathcal{S} \subset \mathcal{D}$ and $U \subset \mathbb{R}^2$, then the total multiplicity of $\mathcal{S}$ in U is*

$$\mathrm{Mult}(\mathcal{S}, U) = \begin{cases} \displaystyle\sum_{D \in \mathcal{S}} \sum_{\mathbf{x} \in D \cap U} \mu_D(\mathbf{x}) & \text{if this is finite, or} \\ \infty & \text{else.} \end{cases}$$

In particular, for each $D \in \mathcal{D}$ we write $\mathrm{Mult}(D, U)$ instead of $\mathrm{Mult}(\{D\}, U)$.

2.2 Bottleneck distance

The space of persistence diagrams can be given a metric as defined next. A **partial matching** between two persistence diagrams $S_\mu, T_\alpha \in \mathcal{D}$ (with notation as in Eqn. 1) is a bijection between a subset of S_μ and a subset of T_α

$$
\begin{array}{ccc}
\mathsf{M}: & S'_\mu & \longrightarrow & T'_\alpha \\
 & \cap & & \cap \\
 & S_\mu & & T_\alpha.
\end{array}
$$

If $(\mathbf{y}, n) = \mathsf{M}(\mathbf{x}, k)$ we say that $(\mathbf{x}, k)$ is matched with $(\mathbf{y}, n)$ and, conversely, that $(\mathbf{y}, n)$ is matched with $(\mathbf{x}, k)$. If $(\mathbf{z}, m)$ is in either $S_\mu \setminus S'_\mu$ or $T_\alpha \setminus T'_\alpha$, then we call it unmatched.

Given $\delta > 0$, a partial matching M between (S, μ) and (T, α) is a δ-**matching** if two things happen:

1. If $(\mathbf{x}, k) \in S'_\mu$ is matched with $(\mathbf{y}, n) = \mathsf{M}(\mathbf{x}, k)$, then we have that $\|\mathbf{x} - \mathbf{y}\|_\infty < \delta$, where $\|(x_1, x_2)\|_\infty = \max\{|x_1|, |x_2|\}$ denotes the L^∞ norm on $\mathbb{R}^2$.

2. If $(\mathbf{z}, m) \in S_\mu \cup T_\alpha$ is unmatched then $\mathsf{pers}(\mathbf{z}) < 2\delta$.

Definition 4 *The **bottleneck distance**, $d_B : \mathcal{D} \times \mathcal{D} \longrightarrow [0, \infty)$, is given by*

$$d_B(D_1, D_2) := \inf \left\{ \delta > 0 : \text{there is a } \delta\text{-matching between } D_1 \text{ and } D_2 \right\}$$

It has been shown that d_B defines a metric on $\mathcal{D}$ (Cohen-Steiner et al. (2007)), and that $\mathcal{D}$ is the metric completion of $\mathcal{D}_0$ (Blumberg et al. (2014)).

We assume in this paper that all persistence points are finite. That is, we discount the existence of any homology classes that live forever of the form $[a, \infty)$. However, we do allow for infinitely many points in our persistence diagrams. The assumptions on $\mathcal{D}$ make it so that the bottleneck distance is still finite between diagrams in $\mathcal{D}$. In particular, this comes from the fact that for every $D = (S, \mu) \in \mathcal{D}$,

$$d_B(D, \varnothing) = \tfrac{1}{2} \max\{\mathsf{pers}(\mathbf{x}) : \mathbf{x} \in S\}$$

and the triangle inequality provides finiteness.

3. Compactness in $(\mathcal{D}, d_B)$

Our first contribution is Theorem 11, which gives a criterion for characterizing a compact set in $\mathcal{D}$ topologized with the bottleneck distance. This work can be viewed in parallel to (Mileyko et al., 2011, Thm. 21), which does the same using the related Wasserstein distance for persistence diagrams. For other structural properties of families of persistence modules see Bubenik and Vergili (2018).

Definition 5 *A subspace of a topological space is **relatively compact** if its closure is compact.*

As in Mileyko et al. (2011), we will actually give a criterion to check that a subset $\mathcal{S} \subseteq \mathcal{D}$ is relatively compact. Specifically, we show that a set of diagrams satisfies three properties if and only if it is relatively compact.
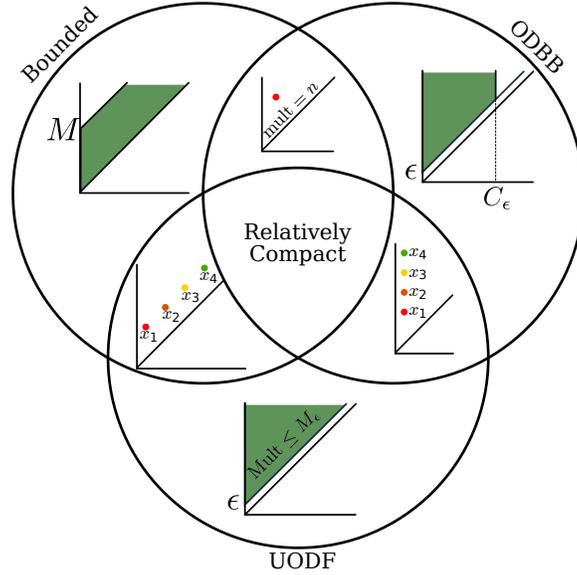
Figure 2: The three criterion for compact sets with examples given on intersections.

3.1 Bounded

The first property of interest is boundedness. A subset of a metric space is said to be bounded if it is contained in an open ball of finite radius. Let $B_C(D) := \{D' \in \mathcal{D} \mid d_B(D, D') < C\}$ denote the ball of radius $C > 0$ about the diagram D. In particular, it can be seen from the definition that if $\mathcal{S} \subset \mathcal{D}$ is bounded then there exists $C > 0$ so that $\mathcal{S} \subseteq B_C(\varnothing)$.

Proposition 6 *Relatively compact subsets of $\mathcal{D}$ are bounded.*

Proof Let $\mathcal{S} \subseteq \mathcal{D}$ be relatively compact. To see that $\mathcal{S}$ is in fact bounded, consider the cover $\{B_1(D) \mid D \in \overline{\mathcal{S}}\}$ by open balls of radius 1 and let $B_1(D_1), \ldots, B_1(D_N)$ be a finite subcover. If
$$C = \max\{d_B(D_i, D_j) : 1 \leq i, j \leq N\}$$
it follows that $\mathcal{S} \subset \overline{\mathcal{S}} \subset B_{C+1}(D_1)$, as claimed. ∎

Note that this proof works for a general metric space, but we work in $\mathcal{D}$ for clarity.

3.2 Off-diagonally birth bounded

The second property of interest controls the points in the set with birth going to ∞.

Definition 7 (Off-diagonally birth bounded) *A set $\mathcal{S} \subset \mathcal{D}$ is said to be **off-diagonally birth bounded (ODBB)** if for every $\epsilon > 0$ there exists a constant $C_\epsilon \geq 0$ so that if $\mathbf{x} \in S \cap \overline{\mathbb{W}^\epsilon}$ (that is, if $\mathsf{pers}(\mathbf{x})) \geq \epsilon$) for $(S, \mu) \in \mathcal{S}$ then $\mathsf{birth}(\mathbf{x}) \leq C_\epsilon$.*

See Fig. 2 for a visualization of the notation.

Proposition 8 *Relatively compact subsets of $\mathcal{D}$ are ODBB.*

Proof By way of contradiction, assume $\mathcal{S} \subseteq \mathcal{D}$ is relatively compact but not ODBB. Then there exist

- $\epsilon > 0$;

- a sequence $\{D_n\}_{n \in \mathbb{N}} \subseteq \mathcal{S}$ with $D_n = (S_n, \mu_n)$;

- a fixed diagram $D = (S, \mu) \in \overline{\mathcal{S}}$ such that

$$\lim_{n \to \infty} D_n = D,$$

 which exists because $\overline{\mathcal{S}}$ is compact; and

- a chosen point in each diagram D_n, namely $\mathbf{x}_n \in S_n$, for every $n \in \mathbb{N}$

so that for all $n \in \mathbb{N}$ we have $\mathsf{pers}(\mathbf{x}_n) \geq \epsilon$, and $\lim_{n \to \infty} \mathsf{birth}(\mathbf{x}_n) = \infty$. Let $\delta < \frac{\epsilon}{2}$ and let $N \in \mathbb{N}$ be large enough so that $d_B(D_n, D) < \delta$ for every $n \geq N$. Then for each $n \geq N$ there exists a δ-matching

$$\gamma_n : \begin{array}{ccc} (S_n)'_{\mu_n} & \longrightarrow & S'_\mu \\ \cap & & \cap \\ (S_n)_{\mu_n} & & S_\mu. \end{array}$$

Because $\mathbf{x}_n \in S_n$ has $\mathsf{pers}(\mathbf{x}_n) \geq \epsilon$ and $\delta < \epsilon/2$, $(\mathbf{x}_n, 1) \in (S_n)'_{\mu_n}$. Let $\mathbf{y}_n \in S$ be such that $\gamma_n(\mathbf{x}_n, 1) = (\mathbf{y}_n, i)$. As γ_n is a delta matching, this has the property that $\|\mathbf{x}_n - \mathbf{y}_n\|_\infty < \delta$ which means, in particular, that $\mathsf{pers}(\mathbf{y}_n) \geq \delta$. But then $\{\mathbf{y}_n\}_{n \in \mathbb{N}} \subseteq S$ is an infinite set in $\mathbb{W}^\delta$, contradicting that $D \in \mathcal{D}$. ∎

3.3 Uniformly off-diagonally finite

The final property of interest controls the multiplicity of points across all diagrams in the set.

Definition 9 (Uniformly off-diagonally finite) *A set $\mathcal{S} \subset \mathcal{D}$ is said to be **uniformly off-diagonally finite (UODF)** if for every $\epsilon > 0$ there exists $M_\epsilon \in \mathbb{N}$ so that*

$$\mathsf{Mult}\left(D, \overline{\mathbb{W}^\epsilon}\right) \leq M_\epsilon$$

for all $D \in \mathcal{S}$

Again, see Fig. 2 for a visualization of the notation.

Proposition 10 *Relatively compact subsets of $\mathcal{D}$ are uniformly off-diagonally finite.*

Proof By the contrapositive, if $\mathcal{S} \subset \mathcal{D}$ is not UODF then there exist $\epsilon > 0$ and a sequence $\{D_n\}_{n \in \mathbb{N}} \subset \mathcal{S}$ so that

$$\mathsf{Mult}\left(D_n, \overline{\mathbb{W}^\epsilon}\right) < \mathsf{Mult}\left(D_{n+1}, \overline{\mathbb{W}^\epsilon}\right)$$

for all $n \in \mathbb{N}$. In particular, using the pigeonhole principle, any partial matching between D_n and D_{n+1} must have at least one point in D_{n+1} unmatched. As this point has persistence greater than ϵ, it follows that $d_B(D_n, D_{n+1}) \geq \epsilon$ for every $n \in \mathbb{N}$ and therefore $\{D_n\}_{n \in \mathbb{N}}$ cannot have a convergent subsequence. This shows that $\overline{\mathcal{S}}$ is not compact. ∎

3.4 Helpful counterexamples

The three conditions bounded, off-diagonally birth bounded and uniformly off-diagonally finite are independent. Indeed, here are three examples of sets which satisfy only two out of the three conditions. See Fig. 2.

1. Bounded and ODBB, but not UODF.

   $\mathcal{S} = \{D_n : n \in \mathbb{N}\}$, $D_n = \{(0, 1)\}$ with $\mu_{D_n}(0, 1) = n$.

2. Bounded and UODF, but not ODBB.

   $\mathcal{S} = \{D_n : n \in \mathbb{N}\}$, $D_n = \{(n, n + 1)\}$ with $\mu_{D_n}(n, n + 1) = 1$.

3. UODF and ODBB, but not bounded.

   $\mathcal{S} = \{D_n : n \in \mathbb{N}\}$, $D_n = \{(0, n)\}$ with $\mu_{D_n}(0, n) = 1$.

3.5 Compact sets

With these definitions, we can now state our main compactness theorem.

Theorem 11 (Characterization of compactness in $(\mathcal{D}, d_B)$) *A set $\mathcal{S} \subset \mathcal{D}$ is relatively compact if and only if it is bounded, off-diagonally birth bounded (ODBB) and uniformly off-diagonally finite (UODF).*

Note that one direction is already provided by Propositions 6, 8 and 10, so our main job is to show that a set which satisfies the three conditions is relatively compact. Before we prove this, however, we will need to build a bit of machinery. First, notice that if $\mathcal{S} \subset \mathcal{D}$ is bounded ($\mathcal{S} \subset B_C(\varnothing)$) and ODBB, then there exist a collection of finite "boxes" $\mathbb{B}_k$ in $\mathbb{W}$ whose union contain all points in the diagrams. Specifically, if $\mathcal{S}$ is bounded and ODBB, there is a $C > 0$ and $\{C_k\}_{k \in \mathbb{N}} \subset \mathbb{R}_{>0}$ non-decreasing, so that if

$$\mathbb{B}_k = \left\{ \mathbf{x} \in \mathbb{W} \ : \ 0 \leq \mathsf{birth}(\mathbf{x}) \leq C_k \quad \text{and} \quad \frac{C}{k+1} < \mathsf{pers}(\mathbf{x}) \leq \frac{C}{k} \right\} \tag{2}$$

then for all $(S, \mu) \in \overline{\mathcal{S}}$, $S \subset \bigcup_{k \in \mathbb{N}} \mathbb{B}_k$. While these are parallelograms in the birth-death plane, they become rectangles in the birth-lifetime plane, hence the moniker "box", see Fig. 3.

Second, we can control the multiplicity of the diagrams in these boxes. If a set $\mathcal{S} \subset \mathcal{D}$ is bounded, ODBB, and UODF, then there exists a sequence $\{M_k\}_{k \in \mathbb{N}} \subset \mathbb{N}$ so that for every $D \in \overline{\mathcal{S}}$ and every $k \in \mathbb{N}$

$$\mathsf{Mult}(D, \mathbb{B}_k) \leq M_k \tag{3}$$

We can use this to prove the following useful, technical lemma, remembering that for a set $U \subset \mathbb{R}^2$ and a diagram $D = (S, \mu)$, we write $D \cap U := (S \cap U, \mu|_{S \cap U})$; and we write $D \subset U$ if $S \subset U$.
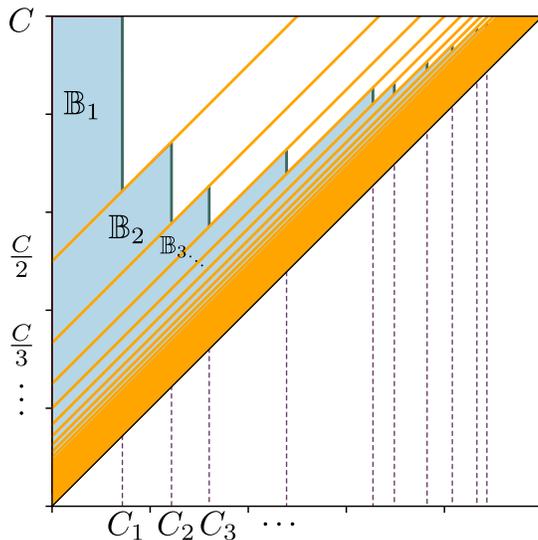
Figure 3: An example to show the notation used in Eq. (2).

Lemma 12 *Let $R \subset \mathbb{W}$ be a relatively compact subset of $\mathbb{R}^2$. If $\{D_n\}_{n \in \mathbb{N}} \subset \mathcal{D}$ is so that*

$$\sup_{n \in \mathbb{N}} \; \mathsf{Mult}(D_n, R) < \infty$$

then the restricted sequence $\{D_n \cap R\}_{n \in \mathbb{N}}$ has a convergent subsequence. Specifically, there exists a diagram $A \in \mathcal{D}$ with $A \subset \overline{R}$ and a strictly increasing function $\varphi : \mathbb{N} \to \mathbb{N}$ so that

$$\lim_{n \to \infty} \left(D_{\varphi(n)} \cap R \right) = A.$$

Proof If the total, combined multiplicity $\mathsf{Mult}(\{D_n\}_{n \in \mathbb{N}}, R) < \infty$, then there is an $N \in \mathbb{N}$ so that $n \geq N$ implies $\mathsf{Mult}(D_n, R) = 0$, and therefore

$$\lim_{n \to \infty} (D_n \cap R) = \varnothing.$$

Thus, we can assume that $\mathsf{Mult}(\{D_n\}_{n \in \mathbb{N}}, R) = \infty$. Let $M \in \mathbb{N}$ be so that $\mathsf{Mult}(D_n, R) \leq M$ for every $n \in \mathbb{N}$. For each $n \in \mathbb{N}$, let $\mathbf{m}_n \in \{0, \ldots, M\}^M$ be the vector having as entries the integers $\mu_n(\mathbf{x})$ for $\mathbf{x} \in D_n \cap R$, sorted in descending order and padded with zeros at the end as necessary. Since $\mathsf{Mult}(\{D_n\}_{n \in \mathbb{N}}, R) = \infty$, the pigeonhole principle implies that there is a non-zero vector $\mathbf{m} \in \{0, \ldots, M\}^M$ and a strictly increasing function $\phi : \mathbb{N} \longrightarrow \mathbb{N}$ such that $\mathbf{m} = \mathbf{m}_{\phi(n)}$ for all $n \in \mathbb{N}$. Let J be the index of the last non-zero entry of $\mathbf{m}$; i.e.

$$\mathbf{m} = (m_1, \ldots, m_J, 0, \ldots, 0)$$

where $M \geq m_1 \geq m_2 \geq \cdots \geq m_J > 0$.

For each $j = 1, \ldots, J$ and $n \in \mathbb{N}$, write the underlying set of $D_{\phi(n)} \cap R$ in order as $\{\mathbf{x}_1^n, \cdots, \mathbf{x}_J^n\}$ so that $\mu_{\phi(n)}(\mathbf{x}_j^n) = m_j$. Then the collection $\{(\mathbf{x}_1^n, \ldots, \mathbf{x}_J^n)\}_{n \in \mathbb{N}}$ is an infinite

sequence in R^J which, by compactness of $\overline{R}$, has an accumulation point $(\mathbf{x}_1, \ldots, \mathbf{x}_J) \in \overline{R}^J$. Thus, let $\psi : \mathbb{N} \longrightarrow \mathbb{N}$ be strictly increasing with the property that

$$\lim_{n \to \infty} \left(\mathbf{x}_1^{\psi(n)}, \ldots, \mathbf{x}_J^{\psi(n)} \right) = (\mathbf{x}_1, \ldots, \mathbf{x}_J).$$

Set A as the disjoint union

$$A = \bigsqcup_{j=1}^{J} (\{\mathbf{x}_j\}, m_j)$$

where $(\{\mathbf{x}_j\}, m_j)$ is the diagram having one off-diagonal point at $\mathbf{x}_j \in \overline{R}$ with multiplicity m_j.

We contend that the subsequence of $\{D_n \cap R\}_{n \in \mathbb{N}}$ defined by $\Phi = \phi \circ \psi$ converges to A. Indeed, given $\epsilon > 0$ let $N \in \mathbb{N}$ be so that $n \geq N$ implies

$$\sup_{1 \leq j \leq J} \left\| \mathbf{x}_j^{\psi(n)} - \mathbf{x}_j \right\|_\infty < \epsilon. \tag{4}$$

Since the diagram $D_{\Phi(n)} \cap R$ has the collection $\{\mathbf{x}_1^{\psi(n)}, \ldots, \mathbf{x}_J^{\psi(n)}\}$ as underlying set, the set function

$$\begin{aligned} \gamma_n : \quad D_{\varphi(n)} \cap R \quad &\longrightarrow \quad A \\ \mathbf{x}_j^{\psi(n)} \quad &\mapsto \quad \mathbf{x}_j \end{aligned}$$

is a bijection. As no points are unmached, Eq. (4) implies that this is an ϵ-matching. Thus $d_B(D_{\Phi(n)} \cap R, A) < \epsilon$ for all $n \geq N$. ∎

With this lemma in place, we can return to the proof of the main theorem.

Proof [Theorem 11] ($\Rightarrow$) If $\mathcal{S} \subset \mathcal{D}$ is relatively compact, then it being bounded, ODBB and UODF follow from Propositions 6, 8 and 10, respectively.

($\Leftarrow$) Let $\mathcal{S} \subset \mathcal{D}$ be bounded, ODBB, and UODF. Fix $\{M_k\}_{k \in \mathbb{N}} \subset \mathbb{N}$ and $\mathbb{B}_k \subset \mathbb{W}$, $k \in \mathbb{N}$, as in Eqs. (2) and (3), and let $\{D_n\}_{n \in \mathbb{N}} \subset \overline{\mathcal{S}}$ be arbitrary. We will use Lemma 12 inductively to construct a sequence $\varphi_1, \varphi_2, \ldots, \varphi_k, \ldots$ of strictly increasing functions $\varphi_k : \mathbb{N} \longrightarrow \mathbb{N}$ so that if $\Phi_k = \varphi_1 \circ \cdots \circ \varphi_k$, then the subsequence of restricted diagrams

$$\left\{ D_{\Phi_k(n)} \cap \mathbb{B}_k \right\}_{n \in \mathbb{N}}$$

converges to a diagram A^k, with $A^k \subset \overline{\mathbb{B}}_k$, for each $k \geq 1$. Once we have built this, we set

$$\begin{aligned} \varphi : \quad \mathbb{N} \quad &\longrightarrow \quad \mathbb{N} \\ n \quad &\mapsto \quad \Phi_n(n) \end{aligned}$$

and the main task for the proof is to show that $\{D_{\varphi(n)}\}_{n \in \mathbb{N}}$ converges to

$$A = \bigsqcup_{k=1}^{\infty} A^k.$$

We proceed inductively in k. The base case follows from applying Lemma 12 to the sequence $\{D_n\}_{n \in \mathbb{N}}$ and the relatively compact set $\mathbb{B}_1$. This results in a strictly increasing function $\varphi_1 : \mathbb{N} \to \mathbb{N}$ and a diagram A^1 for which $A^1 \subset \overline{\mathbb{B}}_1$, and

$$\lim_{n \to \infty} D_{\varphi_1(n)} \cap \mathbb{B}_1 = A^1$$

which finishes the base case.

Now the inductive step. Let $k \geq 1$ and assume that $\varphi_1, \ldots, \varphi_k : \mathbb{N} \to \mathbb{N}$ and $A^1, \cdots, A^k \in \mathcal{D}$ have been constructed so that if $1 \leq j \leq k$ and $\Phi_j = \varphi_1 \circ \cdots \circ \varphi_j$ then $A^j \subset \overline{\mathbb{B}}_j$ and

$$\lim_{n \to \infty} D_{\Phi_j(n)} \cap \mathbb{B}_j = A^j.$$

The sequence $\{D_{\Phi_k(n)}\}_{n \in \mathbb{N}}$ and the set $\mathbb{B}_{k+1}$ satisfy the hypotheses of Lemma 12. Thus, there exist a strictly increasing function $\varphi_{k+1} : \mathbb{N} \longrightarrow \mathbb{N}$ and a diagram A^{k+1} with $A^{k+1} \subset \overline{\mathbb{B}}_{k+1}$ so that if $\Phi_{k+1} = \Phi_k \circ \varphi_{k+1}$ then

$$\lim_{n \to \infty} D_{\Phi_{k+1}(n)} \cap \mathbb{B}_{k+1} = A^{k+1}.$$

Now, we need to show that $\{D_{\varphi(n)}\}_{n \in \mathbb{N}}$ converges to $A = \bigsqcup_{k=1}^{\infty} A^k$. Fix $\epsilon > 0$ and let $K \in \mathbb{N}$ be large enough so that $\frac{C}{K} < \frac{\epsilon}{2}$. For each $1 \leq k \leq K$, let $N_k \in \mathbb{N}$ be so that $n \geq N_k$ implies

$$d_B \left(D_{\Phi_k(n)} \cap \mathbb{B}_k, A^k \right) < \frac{\epsilon}{2}$$

and let $N = \max\{K, N_1, \ldots, N_K\}$. The first thing to notice is that if $n > N$ and $1 \leq k \leq K$ then

$$\varphi(n) := \Phi_n(n) = \Phi_k(\varphi_{k+1} \circ \cdots \circ \varphi_n(n)).$$

Since $\varphi_{k+1} \circ \cdots \circ \varphi_n(n) \geq n > N_k$ then $d_B \left(D_{\varphi(n)} \cap \mathbb{B}_k, A^k \right) < \frac{\epsilon}{2}$. Thus, we can assume we have an $\epsilon/2$-matching

$$\gamma_n^k : D_{\varphi(n)} \cap \mathbb{B}_k \longrightarrow A^k.$$

As the $\mathbb{B}_k$'s are disjoint, the union of the γ_n^ks yields a bijection of multisets

$$\Gamma_n^K : \; D_{\varphi(n)} \cap \bigcup_{k \leq K} \mathbb{B}_k \; \longrightarrow \; \bigsqcup_{k \leq K} A^k$$

Moreover, all points in $D_{\varphi(n)} \cap \bigcup_{k > K} \mathbb{B}_k$ have persistence at most $\epsilon/2$, so it follows that Γ_n^K extends to an $\epsilon/2$-matching

$$\Gamma_n : D_{\varphi(n)} \longrightarrow A.$$

Hence $d_B(D_{\varphi(n)}, A) < \epsilon$ and the result follows. $\blacksquare$

3.6 Consequences of Theorem 11

We note a few immediate consequences of the theorem characterizing compactness.

Theorem 13 *Relatively compact subsets of $(\mathcal{D}, d_B)$ have empty interior.*

Proof Let $\mathcal{S} \subset \mathcal{D}$ be relatively compact, and let

$$\mathbb{B}_k = \left\{ \mathbf{x} \in \mathbb{W} : 0 \leq \mathsf{birth}(\mathbf{x}) < C_k \text{ and } \frac{C}{k+1} \leq \mathsf{pers}(\mathbf{x}) < \frac{C}{k} \right\} , \quad k \in \mathbb{N}$$

be a sequence of boxes (as defined in Eq. (2)) so that $S \subset \bigcup_k \mathbb{B}_k$ for every $(S, \mu) \in \mathcal{S}$. Fix $D \in \mathcal{S}$. We will show that any open ball around D contains a persistence diagram whose underlying set is not in the union of these boxes. Indeed, given $\epsilon > 0$, there exists $k \in \mathbb{N}$ so that $C/k < \epsilon/2$, and if D' is the persistence diagram obtained from D by adding the point $\left(C_k + 1, C_k + 1 + \frac{C}{k} \right)$ with multiplicity one, then it follows that $d_B(D, D') < \epsilon$ but $D' \notin \mathcal{S}$. ∎

Recall that a topological space is *locally compact* if every point has an open neighborhood contained in a compact set; said open set is called a compact neighborhood of the point. The following corollary is a direct consequence of Theorem 13.

Corollary 14 *The space of persistence diagrams $(\mathcal{D}, d_B)$ is not locally compact. Moreover, no diagram $D \in \mathcal{D}$ has a compact neighborhood.*

Proof Let $D \in \mathcal{D}$ and suppose, by way of contradiction, that D has a compact neighborhood. That is, that there exist an open set $\mathcal{U} \subset \mathcal{D}$ and a compact set $\mathcal{S} \subset \mathcal{D}$ so that $D \in \mathcal{U} \subset \mathcal{S}$. Since taking interiors preserves the order of inclusions, then

$$\mathcal{U} = \mathsf{int}(\mathcal{U}) \subset \mathsf{int}(\mathcal{S}) = \emptyset$$

which contradicts $D \in \mathcal{U}$. ∎

As we will see next, the lack of enough compact regions in the space of persistence diagrams also has global implications. Recall that a subset of a topological space is called *nowhere dense* if its closure has empty interior. It follows from Corollary 14 that in the space of persistence diagrams all compact sets are nowhere dense. Moreover, since $(\mathcal{D}, d_B)$ is complete (Blumberg et al. (2014)), then the Baire Category Theorem (Baire (1899)) — which contends that no complete metric space can be written as the countable union of nowhere dense subsets — implies the following.

Corollary 15 *$\mathcal{D}$ cannot be written as the countable union of compact subsets.*

Definition 16 *Let X, Y be topological spaces and let $C(X, Y)$ denote the set of continuous functions from X to Y. Given $K \subset X$ compact and $V \subset Y$ open, let*

$$U(K, V) = \{ f \in C(X, Y) : f(K) \subset V \}.$$

The collection

$$\{ U(K, V) : K \subset X \text{ compact}, \ V \subset Y \text{ open} \}$$

forms a subbase for a topology on $C(X, Y)$, called the compact-open topology.

When Y is a metric space, a sequence of continuous functions $\{f_n : X \longrightarrow Y\}_{n \in \mathbb{N}}$ converges to f in the compact-open topology, if and only if $\{f_n|_K\}_{n \in \mathbb{N}}$ converges uniformly to $f|_K$ for each compact set $K \subset X$. Since $\mathcal{D}$ cannot be written as a countable union of compact sets (thus it is not *hemicompact*), then we have the following.

Corollary 17 *The compact-open topology on $C(\mathcal{D}, \mathbb{R})$ is not metrizable.*

Proof See Example 2.2, Chapter IV, of Conway (2013). ■

4. Linearizing $\mathcal{D}$

The fact that the compact-open topology on $C(\mathcal{D}, \mathbb{R})$ is not metrizable, implies that the problem of finding small compact-open dense subsets needs to be handled with care. The goal of this section is to provide methods for doing this. We begin with the following definition.

Definition 18 *A coordinate system for $\mathcal{D}$ is a collection $\mathcal{F} \subset C(\mathcal{D}, \mathbb{R})$ which separates points. That is, if $D, D' \in \mathcal{D}$ are distinct then there exists $F \in \mathcal{F}$ for which $F(D) \neq F(D')$.*

Of course one could take $\mathcal{F}$ to be the space of all real-valued continuous functions on $\mathcal{D}$, but this is an extreme case; the quality of a coordinate system is determined by its size—the smaller the better. The metaphor to keep in mind is Euclidean space, $\mathbb{R}^n$. In this case, an oblique coordinate system (e.g. Cartesian coordinates) is uniquely determined by a linear basis for the space $\mathcal{L}(\mathbb{R}^n)$ of (continuous) linear functions from $\mathbb{R}^n$ to $\mathbb{R}$.

Our goal is to coordinatize the space of persistence diagrams by finding a continuous embedding of $\mathcal{D}$ into an appropriate topological vector space V, and taking the restriction to $\mathcal{D}$ of elements from $\mathcal{L}(V)$. In order to choose V we will use two principles. First, that persistence diagrams can be interpreted as (rectangular) measures on $\mathbb{W}$ (Oudot (2017); Chazal et al. (2016)), which suggests embedding $\mathcal{D}$ into the dual space of some set of continuous real-valued functions on $\mathbb{W}$. Second, that $\mathcal{D}$ is a topological monoid: the sum of two persistence diagrams $D, D' \in \mathcal{D}$ is their disjoint union $D \sqcup D'$ as multisets, the empty diagram $\varnothing$ is the identity, i.e. $D \sqcup \varnothing = D$, and the operation $\sqcup : \mathcal{D} \times \mathcal{D} \longrightarrow \mathcal{D}$ is associative and continuous. In what follows we will construct an embedding $\nu : \mathcal{D} \hookrightarrow V$ which recovers the measure-theoretic interpretation of persistence diagrams, and preserves the monoidal structure of $\mathcal{D}$ (see Theorem 26). In addition, we will show that appropriate subsets of $\mathcal{L}(V)$ will yield coordinate systems for $\mathcal{D}$ (see Theorem 30), and these in turn will generate dense subsets of $C(\mathcal{D}, \mathbb{R})$ with respect to the compact-open topology (see Theorem 29).

4.1 Topological vector spaces, duals and their topologies

We will first review some basics of topological vector spaces, following Conway (2013). Let V be a topological vector space; that is, a vector space endowed with a topology so that addition and scalar multiplication are continuous functions. Its (topological) dual is the vector space

$$V' = \{T : V \longrightarrow \mathbb{R} \text{ , so that } T \text{ is linear and continuous}\}$$

In particular, if the topology on V comes from a norm $\| \cdot \|_V$, then we write V^* instead of V'. If V^* is endowed with the operator norm

$$\|T\|_* = \sup_{\|\mathbf{v}\|_V = 1} |T(\mathbf{v})|,$$

then V^* is in fact a Banach space. There are three standard topologies on V^*:

Strong: The strong topology is the topology generated by the operator norm $\| \cdot \|_*$. A basis for open neighborhoods of a point $T \in V^*$ is given by sets of the form

$$B_\epsilon(T) = \left\{ T' \in V^* \; : \; \sup_{\|\mathbf{v}\|_V = 1} |T(\mathbf{v}) - T'(\mathbf{v})| < \epsilon \right\}$$

where $\epsilon > 0$. In particular, a sequence $\{T_n\}_{n \in \mathbb{N}} \subset V^*$ converges to $T \in V^*$ in the strong topology if and only if $\{T_n(\mathbf{v})\}_{n \in \mathbb{N}}$ converges to $T(\mathbf{v})$ uniformly in $\mathbf{v} \in V$.

Weak: If V^{**} denotes the dual of the normed space $(V^*, \| \cdot \|_*)$, then the weak topology on V^* is the smallest topology so that every $\mathcal{T} \in V^{**}$ is continuous. A basis for open neighborhoods of a point $T \in V^*$ is given by sets of the form

$$N(\mathcal{T}_1, \ldots, \mathcal{T}_I; \epsilon)(T) = \left\{ T' \in V^* : \max_{1 \le i \le I} \left| \mathcal{T}_i(T') - \mathcal{T}_i(T) \right| < \epsilon \right\}$$

where $\mathcal{T}_1, \ldots, \mathcal{T}_I \in V^{**}$ and $\epsilon > 0$. In particular, $\{T_n\}_{n \in \mathbb{N}} \subset V^*$ converges to $T \in V^*$ in the weak topology if and only if $\{\mathcal{T}(T_n)\}_{n \in \mathbb{N}}$ converges to $\mathcal{T}(T)$ for all $\mathcal{T} \in V^{**}$.

Weak-*: The weak-* topology is the smallest topology so that for each $\mathbf{v} \in V$, the resulting evaluation function

$$
\begin{array}{cccc}
e_{\mathbf{v}} : & V^* & \longrightarrow & \mathbb{R} \\
& T & \mapsto & T(\mathbf{v})
\end{array}
$$

is continuous. A basis for open neighborhoods of $T \in V^*$ is given by sets of the form

$$N(\mathbf{v}_1, \ldots, \mathbf{v}_I; \epsilon)(T) = \left\{ T' \in V^* : \max_{1 \le i \le I} |T'(\mathbf{v}_i) - T(\mathbf{v}_i)| < \epsilon \right\}$$

where $\mathbf{v}_1, \ldots, \mathbf{v}_I \in V$ and $\epsilon > 0$. A sequence $\{T_n\}_{n \in \mathbb{N}} \subset V^*$ converges to $T \in V^*$ in the weak-* topology if and only if $\{T_n(\mathbf{v})\}_{n \in \mathbb{N}}$ converges to $T(\mathbf{v})$ for each $\mathbf{v} \in V$. The convergence, however, need not be uniform in $\mathbf{v}$.

One can check that the weak-* topology is weaker than the weak topology, which in turn is weaker than the strong topology.

4.2 Linearizing the set of finite diagrams

It is useful to first illustrate some of the difficulties associated to finding embeddings for the set of finite diagrams $\mathcal{D}_0$. In what follows we will prove several negative results which will inform the choices in embedding $\mathcal{D}$. Indeed, the first thing to notice is that the set of compactly supported continuous functions from $\mathbb{W}$ to $\mathbb{R}$, denoted $C_c(\mathbb{W})$, is a normed vector space if endowed with the sup norm $\| \cdot \|_\infty$.

The Dirac mass centered at $\mathbf{x} \in \mathbb{W}$ is the linear function

$$\begin{array}{rccc} \delta_{\mathbf{x}}: & C_c(\mathbb{W}) & \longrightarrow & \mathbb{R} \\ & f & \mapsto & f(\mathbf{x}) \end{array}$$

and since $|\delta_{\mathbf{x}}(f)| \leq \|f\|_\infty$ for each $f \in C_c(\mathbb{W})$, it follows that $\delta_{\mathbf{x}} \in C_c(\mathbb{W})^*$. Let

$$\begin{array}{rccc} \nu_0: & \mathcal{D}_0 & \longrightarrow & C_c(\mathbb{W})^* \\ & \varnothing & \mapsto & 0 \\ & \varnothing \neq (S, \mu) & \mapsto & \displaystyle\sum_{\mathbf{x} \in S} \mu(\mathbf{x})\delta_{\mathbf{x}}. \end{array}$$

It is not hard to see that

Proposition 19 $\nu_0 : \mathcal{D}_0 \longrightarrow C_c(\mathbb{W})^*$ *is injective, and satisfies* $\nu_0(D \sqcup D') = \nu_0(D) + \nu_0(D')$ *for every* $D, D' \in \mathcal{D}_0$.

Deciding weather or not ν_0 is continuous depends on the topology with which $C_c(\mathbb{W})^*$ is endowed. We start with the coarser topologies, but immediately have the following negative results.

Proposition 20 *If* $C_c(\mathbb{W})^*$ *is endowed with the weak topology, then* ν_0 *is discontinuous at every point.*

Proof Fix $D \in \mathcal{D}_0$, and let $D_n \in \mathcal{D}_0$ be the diagram obtained from D by adding the point $(1, 1 + 1/n)$ with multiplicity n. It follows that $\{D_n\}_n$ converges to D with respect to the bottleneck distance. We contend that $\nu_0(D_n)$ does not converge to $\nu_0(D)$ with respect to the weak topology; in other words, we will show that there exist $\epsilon_0 > 0$ and a linear operator $\mathcal{T} : C_c(\mathbb{W})^* \longrightarrow \mathbb{R}$, continuous with respect to the strong topology, so that $|\mathcal{T}(\nu_0(D_n)) - \mathcal{T}(\nu_0(D))| \geq \epsilon_0$ for infinitely many values of n.

Indeed, the first thing to notice is that since $\mathbb{W}$ is not compact, $C_c(\mathbb{W})$ is not complete. Its completion is the space $C_0(\mathbb{W})$ of continuous functions on $\mathbb{W}$ which vanish at the diagonal and at infinity. Let $\iota : C_c(\mathbb{W}) \hookrightarrow C_0(\mathbb{W})$ be the inclusion and let $\iota^* : C_0(\mathbb{W})^* \longrightarrow C_c(\mathbb{W})^*$ be the induced homomorphism. Since $C_c(\mathbb{W})$ is dense in $C_0(\mathbb{W})$, then ι^* is an isometric isomorphism; its inverse $\jmath^* : C_c(\mathbb{W})^* \longrightarrow C_0(\mathbb{W})^*$ sends a continuous linear map $T : C_c(\mathbb{W}) \longrightarrow \mathbb{R}$ to its unique continuous linear extension $\jmath^*(T) : C_0(\mathbb{W}) \longrightarrow \mathbb{R}$.

Now, let $\varphi : \mathbb{R}^2 \longrightarrow [0, 1]$ be a continuous (bump) function so that

$$\varphi(x, y) = \begin{cases} 1 & \text{if } \max\{x, y - x\} < 2 \\ \\ 0 & \text{if } \max\{x, y - x\} \geq 3. \end{cases}$$

It follows that $f(x, y) = (y - x) \cdot \varphi(x, y) \in C_0(\mathbb{W})$, and hence the evaluation function $e_f : C_0(\mathbb{W})^* \longrightarrow \mathbb{R}$ is a bounded linear operator. Let $\mathcal{T} : C_c(\mathbb{W})^* \longrightarrow \mathbb{R}$ be the composition $e_f \circ \jmath^*$. Then, for each $n > 1$ we have that

$$\mathcal{T}(\nu_0(D_n)) = \mathcal{T}(\nu_0(D)) + 1,$$

and letting $\epsilon_0 = 1$ completes the proof. $\blacksquare$

Corollary 21 *If $C_c(\mathbb{W})^*$ is endowed with the strong topology, then ν_0 is discontinuous at every point.*

Proof This follows directly from the fact that the strong topology contains the weak topology. ∎

It is not until we pass to the weakest of the three standard topologies that we approach a useful result.

Proposition 22 *If $C_c(\mathbb{W})^*$ is endowed with the weak-* topology, then $\nu_0 : \mathcal{D}_0 \longrightarrow C_c(\mathbb{W})^*$ is continuous.*

Before presenting the proof, we have the following useful lemma.

Lemma 23 *For each $f \in C_c(\mathbb{W})$, the function*

$$
\nu_f : \quad
\begin{array}{ccc}
\mathcal{D} & \longrightarrow & \mathbb{R} \\
(S, \mu) & \mapsto & \sum\limits_{\mathbf{x} \in S} \mu(\mathbf{x}) f(\mathbf{x})
\end{array}
\tag{5}
$$

is continuous.

Proof The first observation is that the sum defining ν_f is always finite, since the support of $f \in C_c(\mathbb{W})$ intersects the underlying set of any persistence diagram at only finitely many points. Let $D = (S, \mu) \in \mathcal{D}$ and fix $\epsilon > 0$. Since $\mathsf{supp}(f) \subset \mathbb{W}$ is compact, then f is uniformly continuous. Further, there exists $\delta > 0$ for which $\mathsf{supp}(f) \subset \mathbb{W}^{2\delta}$, and

$$
|f(\mathbf{x}) - f(\mathbf{y})| < \frac{\epsilon}{\sum\limits_{\mathbf{z} \in S \cap \mathbb{W}^{\delta}} \mu(\mathbf{z})}
\tag{6}
$$

whenever $\|\mathbf{x} - \mathbf{y}\|_\infty < \delta$.

Let $D' = (T, \alpha) \in \mathcal{D}$ be given with $d_B(D, D') < \delta$. We will show that $|\nu_f(D) - \nu_f(D')| < \epsilon$. Fix a δ-matching

$$
\mathsf{M} : \quad
\begin{array}{ccc}
S'_\mu & \overset{\cong}{\longrightarrow} & T'_\alpha \\
\cap & & \cap \\
S_\mu & & T_\alpha.
\end{array}
$$

This means that if $(\mathbf{x}, k) \in S'_\mu$ and $(\mathbf{y}, n) = \mathsf{M}(\mathbf{x}, k)$ are matched, then $\|\mathbf{x} - \mathbf{y}\|_\infty < \delta$; and if $(\mathbf{z}, m) \in (S_\mu \smallsetminus S'_\mu) \cup (T_\alpha \smallsetminus T'_\alpha)$ is unmatched, then $\mathsf{pers}(\mathbf{z}) < 2\delta$. In this case $\mathbf{z} \notin \mathsf{supp}(f)$, which implies $f(\mathbf{z}) = 0$. Hence

$$
\begin{aligned}
\nu_f(D) &= \sum_{\mathbf{x} \in S} \mu(\mathbf{x}) f(\mathbf{x}) \\
&= \sum_{(\mathbf{x}, k) \in S_\mu} f(\mathbf{x}) \\
&= \sum_{(\mathbf{x}, k) \in S'_\mu} f(\mathbf{x})
\end{aligned}
$$

19

and similarly,

$$\nu_f(D') \;=\; \sum_{(\mathbf{y},n)\in T'_\alpha} f(\mathbf{y}).$$

Therefore

$$
\begin{aligned}
|\nu_f(D) - \nu_f(D')| \;&=\; \left| \sum_{(\mathbf{x},k)\in S'_\mu} f(\mathbf{x}) - \sum_{(\mathbf{y},n)\in T'_\alpha} f(\mathbf{y}) \right| \\
&\leq\; \sum_{\substack{(\mathbf{y},n)=\mathsf{M}(\mathbf{x},k) \\ (\mathbf{x},k)\in S'_\mu}} |f(\mathbf{x}) - f(\mathbf{y})|
\end{aligned}
$$

where each term $|f(\mathbf{x}) - f(\mathbf{y})|$ is potentially nonzero only when $\mathbf{x}$ or $\mathbf{y}$ are in $\mathsf{supp}(f) \subset \mathbb{W}^{2\delta}$. Since in this case $\|\mathbf{x} - \mathbf{y}\|_\infty < \delta$, we would get $\mathbf{x}, \mathbf{y} \in \mathbb{W}^\delta$. Combining this observation with equation (6) completes the proof. ∎

Proof [Proposition 22] Let $D = (S, \mu) \in \mathcal{D}_0$, and fix a weak-* basic neighborhood $N(f_1, \ldots, f_I; \epsilon)$ for $\nu_0(D)$. Notice that for each $i = 1, \ldots, I$ we have $\nu_0(D)(f_i) = \nu_{f_i}(D)$. Since ν_{f_i} is continuous at D, then given $\epsilon > 0$ there exists $\delta_i > 0$ so that $d_B(D, D') < \delta_i$ implies $|\nu_{f_i}(D) - \nu_{f_i}(D')| < \epsilon$. If we let $\delta = \min\{\delta_1, \ldots, \delta_I\}$, it follows that whenever $d_B(D, D') < \delta$ then for all $i = 1, \ldots, I$

$$|\nu_0(D)(f_i) - \nu_0(D')(f_i)| = |\nu_{f_i}(D) - \nu_{f_i}(D')| < \epsilon.$$

This shows that $\nu_0(D') \in N(f_1, \ldots, f_I; \epsilon)$ and hence ν_0 is continuous. ∎

These results imply that out of the three standard topologies on $C_c(\mathbb{W})^*$, the weak-* topology is the only one for which ν_0 yields a continuous embedding of $\mathcal{D}_0$ into $C_c(\mathbb{W})^*$. The question now is whether this embedding can be extended to $\mathcal{D}$. The answer, as it turns out, is negative.

Proposition 24 *If $C_c(\mathbb{W})^*$ is endowed with the weak-* topology, then $\nu_0 : \mathcal{D}_0 \longrightarrow C_c(\mathbb{W})^*$ cannot be continuously extended to any $D \in \mathcal{D} \smallsetminus \mathcal{D}_0$.*

Proof Assume, by way of contradiction, that ν_0 extends continuously to some $D = (S, \mu) \in \mathcal{D} \smallsetminus \mathcal{D}_0$. If for each $n \in \mathbb{N}$ we let D_n be the restriction of D to $\mathbb{W}^{1/n}$, then $D_n \in \mathcal{D}_0$ for all $n \in \mathbb{N}$, and the sequence $\{D_n\}_{n\in\mathbb{N}}$ converges to D with respect to the bottleneck distance. By the continuity assumption of ν_0 at D, we have that

$$\nu_0(D) = \lim_{n\to\infty} \nu_0(D_n)$$

where convergence is with respect to the weak-* topology. In other words,

$$\nu_0(D)(f) = \lim_{n\to\infty} \nu_0(D_n)(f)$$

for every $f \in C_c(\mathbb{W})$. It follows that, given $f \in C_c(\mathbb{W})$, there exists $N_f \in \mathbb{N}$ so that $\mathsf{supp}(f) \subset \mathbb{W}^{1/n}$ for all $n \geq N_f$, and therefore the sequence $\nu_0(D_n)(f)$ becomes constant and equal to

$$\sum_{\mathbf{x} \in S} \mu(\mathbf{x}) f(\mathbf{x}). \tag{7}$$

We claim that if $C_c(\mathbb{W})$ is endowed with the sup norm $\| \cdot \|_\infty$, then the linear function

$$\nu_0(D): \begin{array}{ccc} C_c(\mathbb{W}) & \longrightarrow & \mathbb{R} \\ f & \mapsto & \sum_{\mathbf{x} \in S} \mu(\mathbf{x}) f(\mathbf{x}) \end{array}$$

is discontinuous at every point. To this end, we will show that given $f \in C_c(\mathbb{W})$ there exists a sequence $\{f_n\}_{n \in \mathbb{N}} \subset C_c(\mathbb{W})$ which converges to f with respect to $\| \cdot \|_\infty$, but for which $\{\nu_0(D)(f_n)\}_{n \in \mathbb{N}}$ does not converge to $\nu_0(D)(f)$. This would contradict $\nu_0(D) \in C_c(\mathbb{W})^*$.

Indeed, since $(S, \mu) = D \notin \mathcal{D}_0$, then there exists a sequence $\{\mathbf{x}_n\}_{n \in \mathbb{N}} \subset S \smallsetminus \mathsf{supp}(f)$ so that $\mathsf{pers}(\mathbf{x}_n)$ is strictly decreasing as n goes to infinity. Therefore, it is possible to construct a sequence $\{r_n\}_{n \in \mathbb{N}}$ of positive real numbers, so that the balls $B_{r_n}(\mathbf{x}_n) \subset \mathbb{W}$ are all disjoint and disjoint with the support of f. Let $\phi_n : \mathbb{R}^2 \longrightarrow [0, \infty)$ be the bump function

$$\phi_n(\mathbf{x}) = \frac{\max\{0, \ r_n - 2\|\mathbf{x} - \mathbf{x}_n\|\}}{r_n}$$

supported on the closure of $B_{\frac{r_n}{2}}(\mathbf{x}_n)$, and let

$$f_n = f + \frac{\phi_1 + \cdots + \phi_n}{n}.$$

It follows that $\{f_n\}_{n \in \mathbb{N}}$ is a sequence of continuous and compactly supported functions on $\mathbb{W}$, so that $\|f_n - f\|_\infty < \frac{1}{n}$ for all $n \in \mathbb{N}$, and for which

$$\begin{aligned} \nu_0(D)(f_n) - \nu_0(D)(f) & = & \frac{1}{n} \nu_0(D)(\phi_1 + \cdots + \phi_n) \\[2mm] & \geq & \frac{1}{n} \big(\mu(\mathbf{x}_1)\phi_1(\mathbf{x}_1) + \cdots + \mu(\mathbf{x}_n)\phi_n(\mathbf{x}_n) \big) \\[2mm] & \geq & 1. \end{aligned}$$

Hence the sequence $\{\nu_0(D)(f_n)\}_{n \in \mathbb{N}}$ does not converge to $\nu_0(D)(f)$, showing that $\nu_0(D)$ is discontinuous at f. ∎

4.3 Linearizing infinite diagrams

There are two main lessons to draw from the previous results: First, that even though there is a candidate for extending ν_0 to infinite diagrams, namely Eq. (7), the topology on $C_c(\mathbb{W})$ induced by the sup norm is inadequate as it does not have enough open sets. The second lesson is that a weak-* topology on the dual of $C_c(\mathbb{W})$ is the most likely to ensure continuity when embedding $\mathcal{D}$. In what follows we will describe a (locally convex)

topology on $C_c(\mathbb{W})$, and a corresponding weak-* topology on the topological dual $C_c(\mathbb{W})'$ with the required properties. We will utilize the theory of *locally convex topological vector spaces*, which generalize Banach spaces, and provide a rich framework in which to study weak topologies. For a more detailed account we direct the interested reader to Chapters IV and V of Conway (2013).

Let $\{K_n\}_{n\in\mathbb{N}}$ be a sequence of compact subsets of $\mathbb{W}$ so that $K_n \subset K_{n+1}$ for all $n \in \mathbb{N}$, and for which

$$\mathbb{W} = \bigcup_{n\in\mathbb{N}} K_n.$$

It follows that each vector space

$$C_c(K_n) = \{f \in C(\mathbb{W}) : \mathsf{supp}(f) \subset K_n\}$$

is a Banach space if endowed with the sup norm; in particular it is a *locally convex space*.

Definition 25 *A locally convex space is a topological vector space V, whose topology is generated by a family $\mathscr{P}$ of seminorms on V which separate points. More specifically, $\mathscr{P}$ is a collection $\{\rho_\alpha\}_{\alpha\in\Gamma}$ of continuous functions $\rho_\alpha : V \longrightarrow [0,\infty)$ so that*

1. $\rho_\alpha(\mathbf{u} + \mathbf{v}) \leq \rho_\alpha(\mathbf{u}) + \rho_\alpha(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in V$,

2. $\rho_\alpha(\lambda\mathbf{u}) = |\lambda|\rho_\alpha(\mathbf{u})$ for all scalars λ,

3. $\rho_\alpha(\mathbf{u}) = 0$ for all $\alpha \in \Gamma$ implies $\mathbf{u} = \mathbf{0}$

and the topology of V is the weakest for which all the ρ_α's are continuous.

In particular, all normed spaces are locally convex. Notice also that each inclusion

$$C_c(K_n) \subset C_c(K_{n+1}) \qquad n \in \mathbb{N}$$

is continuous and that

$$C_c(\mathbb{W}) = \bigcup_{n\in\mathbb{N}} C_c(K_n).$$

The **strict inductive limit topology** on $C_c(\mathbb{W})$ is the finest locally convex topology so that each inclusion $C_c(K_n) \hookrightarrow C_c(\mathbb{W})$ is continuous. In this topology, a linear map $T : C_c(\mathbb{W}) \longrightarrow Y$ to a locally convex space Y is continuous if and only if the restriction of T to each $C_c(K_n)$ is continuous.

Let $C_c(\mathbb{W})'$ denote the topological dual of $C_c(\mathbb{W})$ with respect to the strict inductive limit topology, and endow $C_c(\mathbb{W})'$ with the weakest topology so that for each $f \in C_c(\mathbb{W})$ the resulting evaluation function

$$\begin{array}{rccc} e_f: & C_c(\mathbb{W})' & \longrightarrow & \mathbb{R} \\ & T & \mapsto & T(f) \end{array}$$

is continuous. This is the corresponding weak-* topology. It follows that a basis for neighborhoods of a point $T \in C_c(\mathbb{W})'$ is given by sets of the form

$$N(f_1,\ldots,f_I;\epsilon)(T) = \left\{ \tilde{T} \in C_c(\mathbb{W})' : \left|(T - \tilde{T})(f_i)\right| < \epsilon \,,\, i = 1,\ldots,I \right\}$$

where $f_1,\ldots,f_I \in C_c(\mathbb{W})$ and $\epsilon > 0$. We have the following theorem.

Theorem 26 *Given a persistence diagram $D = (S, \mu) \in \mathcal{D}$ and a function $f \in C_c(\mathbb{W})$, define*

$$\nu_D(f) := \sum_{\mathbf{x} \in S} \mu(\mathbf{x}) f(\mathbf{x}). \tag{8}$$

If $C_c(\mathbb{W})$ is endowed with the strict inductive limit topology, and $C_c(\mathbb{W})'$ is its topological dual endowed with the corresponding weak-* topology, then

$$
\begin{array}{rccc}
\nu : & \mathcal{D} & \longrightarrow & C_c(\mathbb{W})' \\
& D & \mapsto & \nu_D
\end{array}
$$

is continuous, injective and satisfies $\nu(D \sqcup D') = \nu(D) + \nu(D')$ for all $D, D' \in \mathcal{D}$.

Proof First, we ensure that ν_D is well defined. Fix $D = (S, \mu) \in \mathcal{D}$ and $n \in \mathbb{N}$. Then $S \cap K_n$ is a finite set and hence for each $f \in C_c(K_n)$ it follows that

$$\nu_D(f) = \sum_{\mathbf{x} \in S} \mu(\mathbf{x}) f(\mathbf{x}) < \infty.$$

As for continuity of ν_D, fix $f_0 \in C_c(K_n)$, let $\epsilon > 0$, and let

$$\delta < \frac{\epsilon}{\sum\limits_{\mathbf{z} \in S \cap K_n} \mu(\mathbf{z})}.$$

If $f \in C_c(K_n)$ is so that $\|f - f_0\|_\infty < \delta$, then

$$
\begin{aligned}
\left| \nu_D(f) - \nu_D(f_0) \right| &= \left| \sum_{\mathbf{x} \in S} \mu(\mathbf{x}) \big(f(\mathbf{x}) - f_0(\mathbf{x}) \big) \right| \\
&\leq \sum_{\mathbf{z} \in S \cap K_n} \mu(\mathbf{z}) \|f - f_0\|_\infty \\
&< \epsilon.
\end{aligned}
$$

Therefore ν_D is a real-valued continuous linear function on $C_c(K_n)$ for each n. This shows that $\nu(D) = \nu_D \in C_c(\mathbb{W})'$ for all $D \in \mathcal{D}$.

To see that $\nu : \mathcal{D} \longrightarrow C_c(\mathbb{W})'$ is continuous, we proceed exactly as in the proof of Proposition 22. Indeed, let $D \in \mathcal{D}$, and fix a basic neighborhood $N(f_1, \ldots, f_I; \epsilon)$ for $\nu(D)$. For each $i = 1, \ldots, I$ the function $\nu_{f_i} : \mathcal{D} \longrightarrow \mathbb{R}$, $\nu_{f_i}(D) = \nu_D(f_i)$, is continuous by Lemma 23, and hence there exists $\delta > 0$ such that

$$|\nu(D)(f_i) - \nu(D')(f_i)| = |\nu_{f_i}(D) - \nu_{f_i}(D')| < \epsilon$$

for all $i = 1, \ldots, I$, whenever $d_B(D, D') < \delta$.

Injectivity of ν is deduced from the following observation. If $(S, \mu), (T, \alpha) \in \mathcal{D}$ are distinct, then we can assume without loss of generality that there exists $\mathbf{x} \in S$ such that either: $\mathbf{x} \notin T$, or $\mathbf{x} \in T$ and $\mu(\mathbf{x}) \neq \alpha(\mathbf{x})$. Let $f \in C_c(\mathbb{W})$ be such that $\mathsf{supp}(f) \cap (S \cup T) = \{\mathbf{x}\}$, and for which $f(\mathbf{x}) = 1$. If $\mathbf{x} \notin T$, then

$$\nu(S, \mu)(f) = \mu(\mathbf{x}) \neq 0 = \nu(T, \alpha)(f).$$

Similarly, for the case where $\mathbf{x} \in T$ we have

$$\nu(S, \mu)(f) = \mu(\mathbf{x}) \neq \alpha(\mathbf{x}) = \nu(T, \alpha)(f),$$

which completes the proof. ∎

The Riesz-Markov representation theorem—see for instance Theorem 2.14 of Rudin (2006)—contends that if $T : C_c(\mathbb{W}) \longrightarrow \mathbb{R}$ is linear and satisfies $T(f) \geq 0$ whenever $f(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathbb{W}$, then there exists a unique positive (regular Baire) measure η on $\mathbb{W}$ so that

$$\int_{\mathbb{W}} f d\eta = T(f)$$

for all $f \in C_c(\mathbb{W})$, and so that $\eta(K) < \infty$ for every compact set $K \subset \mathbb{W}$. Applying this theorem to elements in the image of $\nu : \mathcal{D} \longrightarrow C_c(\mathbb{W})'$, implies that $\nu_D = \nu(D)$ is a regular Baire measure on $\mathbb{W}$ for each $D \in \mathcal{D}$. The resulting measure is in fact the rectangular measure introduced by Chazal et al. (2009).

5. Approximating Continuous Functions on Persistence Diagrams

As we saw in Theorem 26, the function $\nu : \mathcal{D} \longrightarrow C_c(\mathbb{W})'$ provides a continuous embedding so that $\nu(D \sqcup D') = \nu(D) + \nu(D')$ for all $D, D' \in \mathcal{D}$. We can now proceed to the task of finding coordinate systems for $\mathcal{D}$ (see Definition 18). Indeed, the first thing to note is that composing ν with elements from $C_c(\mathbb{W})''$, the topological dual of $C_c(\mathbb{W})'$, yields real-valued continuous functions on $\mathcal{D}$. By construction, these functions also respect the monoidal structure of $\mathcal{D}$. The following theorem characterizes exactly what the elements of $C_c(\mathbb{W})''$ are.

Theorem 27 *Let V be a locally convex space, and endow its topological dual V' with the associated weak-* topology. That is, the smallest topology such that all the evaluations*

$$\begin{array}{cccc} e_{\mathbf{v}} & V' & \longrightarrow & \mathbb{R} \\ & T & \mapsto & T(\mathbf{v}) \end{array}$$

for $\mathbf{v} \in V$, are continuous. Then the function

$$\begin{array}{cccc} e : & V & \longrightarrow & V'' \\ & \mathbf{v} & \mapsto & e_{\mathbf{v}} \end{array}$$

is an isomorphism of locally convex spaces.

Proof See Theorem 1.3 in Chapter V of Conway (2013). ∎

Applying this theorem to the locally convex space $C_c(\mathbb{W})$, topologized with the strict inductive limit topology, implies that the elements of $C_c(\mathbb{W})''$ are evaluations e_f, with $f \in C_c(\mathbb{W})$ uniquely determined. Composing e_f with ν yields a continuous function $e_f \circ \nu : \mathcal{D} \longrightarrow \mathbb{R}$ which preserves the monoidal structure of $\mathcal{D}$. Moreover, given $D \in \mathcal{D}$ we have that

$$e_f \circ \nu(D) = \nu_D(f) = \nu_f(D)$$

24

where $\nu_f : \mathcal{D} \longrightarrow \mathbb{R}$ is defined by Eq. (5). We saw in Lem. 23 that these types of functions are indeed continuous, but now we have a more complete picture: they arise exactly as the continuous linear functions on a linearization of $\mathcal{D}$. The goal now is to construct coordinate systems for $\mathcal{D}$ by selecting appropriate subsets of $C_c(\mathbb{W})$. Indeed, we define these subsets as follows.

Definition 28 *A template system for $\mathcal{D}$ is a collection $\mathcal{T} \subset C_c(\mathbb{W})$ so that*

$$\mathcal{F}_{\mathcal{T}} = \{\nu_f : f \in \mathcal{T}\}$$

is a coordinate system (see Defn. 18) for $\mathcal{D}$. The elements of $\mathcal{T}$ are called template functions.

The point of working with these template systems is that they can be used to approximate continuous functions on persistence diagrams, as given by the following theorem.

Theorem 29 *Let $\mathcal{T} \subset C_c(\mathbb{W})$ be a template system for $\mathcal{D}$, let $\mathcal{C} \subset \mathcal{D}$ be compact, and let $F : \mathcal{C} \longrightarrow \mathbb{R}$ be continuous. Then for every $\epsilon > 0$ there exist $N \in \mathbb{N}$, a polynomial $p \in \mathbb{R}[x_1, \ldots, x_N]$ and template functions $f_1, \ldots, f_N \in \mathcal{T}$ so that*

$$\left| p\big(\nu_D(f_1), \ldots, \nu_D(f_N)\big) - F(D) \right| < \epsilon$$

for every $D \in \mathcal{C}$. That is, the collection of functions of the form $D \mapsto p\big(\nu_D(f_1), \ldots, \nu_D(f_N)\big)$, is dense in $C(\mathcal{D}, \mathbb{R})$ with respect to the compact-open topology.

Proof Let $\mathcal{T} \subset C_c(\mathbb{W})$ be a template system for $\mathcal{D}$ and let $\mathcal{F} = \{\nu_f : f \in \mathcal{T}\} \subset C(\mathcal{D}, \mathbb{R})$ be the corresponding coordinate system. Let $\mathcal{A} \subset C(\mathcal{D}, \mathbb{R})$ denote the algebra generated by $\mathcal{F}$. In other words, $\mathcal{A}$ is the set of finite linear combinations of finite products of elements from $\mathcal{F}$. It follows that every element of $\mathcal{A}$ can be written as

$$p(\nu_{f_1}, \ldots, \nu_{f_N})$$

for some collection of templates $f_1, \ldots, f_N \in \mathcal{T}$ and some polynomial $p \in \mathbb{R}[x_1, \ldots, x_N]$. Let $\iota : \mathcal{C} \hookrightarrow \mathcal{D}$ be the inclusion and let $\iota^* : C(\mathcal{D}, \mathbb{R}) \longrightarrow C(\mathcal{C}, \mathbb{R})$ be the corresponding restriction homomorphism. Now, since $\mathcal{F}$ separates points in $\mathcal{D}$ and $\mathcal{F} \subset \mathcal{A}$, then it follows that $\iota^*(\mathcal{A})$ is an algebra which separates points in $C(\mathcal{C}, \mathbb{R})$ and contains the nonzero constant functions. The result follows from the Stone-Weierstrass theorem, which contends that any such algebra is dense with respect to the sup norm. ∎

The main question now is how to construct template systems in practice. The next theorem gives one possible method for constructing a countable template system for $\mathcal{D}$ by translating and re-scaling the support of any nonzero $f \in C_c(\mathbb{W})$. In particular, it shows that there are plenty of coordinate systems for the space of persistence diagrams, and helps explain why we refer to these functions as templates.

Theorem 30 *Let $f \in C_c(\mathbb{W})$, $n \in \mathbb{N}, \mathbf{m} \in \mathbb{Z}^2$ and define*

$$f_{n,\mathbf{m}}(\mathbf{x}) = f\left(n\mathbf{x} + \frac{\mathbf{m}}{n}\right).$$

25

If f is nonzero, then

$$\mathcal{T} = \left\{ f_{n,\mathbf{m}} \mid n \in \mathbb{N}, \mathbf{m} \in \mathbb{Z}^2 \right\} \cap C_c(\mathbb{W})$$

is a template system for $\mathcal{D}$. Moreover, if f is Lipschitz, then the elements of the associated coordinate system

$$\left\{ \nu_{f_{n,\mathbf{m}}} = f_{n,\mathbf{m}} \circ \nu \mid f_{n,\mathbf{m}} \in \mathcal{T} \right\} = \mathcal{F}_{\mathcal{T}}$$

are Lipschitz on any relatively compact set $\mathcal{S} \subset \mathcal{D}$. In other words, the coordinate system associated to a nonzero Lipschitz template function, is stable on relatively compact subsets of $\mathcal{D}$.

Proof In order to show that $\mathcal{F}_{\mathcal{T}}$ separates points in $\mathcal{D}$, let $(S, \mu), (T, \alpha) \in \mathcal{D}$ be distinct diagrams, and assume without loss of generality that there exists $\mathbf{y} = (y_1, y_2) \in S$ so that either: $\mathbf{y} \notin T$; or $\mathbf{y} \in T$ and $\mu(\mathbf{y}) \neq \alpha(\mathbf{y})$. We will find a $f_{n,\mathbf{m}} \in \mathcal{T}$ so that $f_{n,\mathbf{m}}(\mathbf{y}) \neq 0$, and $f_{n,\mathbf{m}}(\mathbf{x}) = 0$ for all other $\mathbf{x} \in S \cup T$.

Let $\mathbf{z} = (z_1, z_2) \in \mathbb{W}$ be so that $f(\mathbf{z}) \neq 0$. By continuity of f with respect to the Euclidean norm $\| \cdot \|$, which is equivalent to the sup norm $\| \cdot \|_\infty$, there exists $r > 0$ so that

$$B_r^\infty(\mathbf{z}) := \left\{ \mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x} - \mathbf{z}\|_\infty < r \right\} \subset \mathsf{supp}(f).$$

Moreover, since $\mathsf{supp}(f) \subset \mathbb{W}$ is compact, there exists $s > r$ so that $\mathsf{supp}(f) \subset B_s^\infty(\mathbf{z})$. Putting this together, we have $r < s$ so that

$$B_r^\infty(\mathbf{z}) \subset \mathsf{supp}(f) \subset B_s^\infty(\mathbf{z}).$$

Fix $\epsilon > 0$ small enough so that $B_\epsilon^\infty(\mathbf{y}) \subset \mathbb{W}$ and $B_\epsilon^\infty(\mathbf{y}) \cap (T \cup S) = \{\mathbf{y}\}$. What we will show now is that it is possible to find $n \in \mathbb{N}$ and $\mathbf{m} \in \mathbb{Z}^2$ so that $n\mathbf{y} + \frac{\mathbf{m}}{n} \in B_r^\infty(\mathbf{z})$, and so that $\mathbf{x} \notin B_\epsilon^\infty(\mathbf{y})$ implies $n\mathbf{x} + \frac{\mathbf{m}}{n} \notin B_s^\infty(\mathbf{z})$.

Fix $n \in \mathbb{N}$ large enough so that $n \geq \max\left\{ \frac{1}{r}, \frac{2s}{\epsilon} \right\}$. Define $L_j(t) = nt + (z_j - ny_j)$ for $j = 1, 2$ with $\mathbf{z} = (z_1, z_2)$ and $\mathbf{y} = (y_1, y_2)$. This function has the property that $L_j(y_j) = z_j$. Further, if $|t - y_j| > \epsilon$, then

$$|L_j(t) - z_j| = |(nt + z_j) - (ny_j + z_j)| = n|t - y_j| > \frac{2s}{\epsilon}\epsilon = 2s.$$

Let $k_j \in \mathbb{Z}$ be the unique integer so that

$$k_j \leq z_j - ny_j < k_j + 1.$$

By dividing the interval $[k_j, k_j + 1)$ into n subintervals of length $\frac{1}{n}$, we have that there exists a unique integer $0 \leq \ell_j < n$ so that

$$k_j + \frac{\ell_j}{n} \leq z_j - ny_j < k_j + \frac{\ell_j + 1}{n}.$$

Let $m_j = nk_j + \ell_j$, and $\mathbf{m} = (m_1, m_2)$. It follows that $\left\| \mathbf{z} - \left(n\mathbf{y} + \frac{\mathbf{m}}{n} \right) \right\|_\infty < \frac{1}{n} < r$, and therefore $f_{n,\mathbf{m}}(\mathbf{y}) = f\left(n\mathbf{y} + \frac{\mathbf{m}}{n} \right) \neq 0$. Moreover, if $\mathbf{x} \notin B_\epsilon^\infty(\mathbf{y})$ and $j \in \{1, 2\}$ is so that $|x_j - y_j| \geq \epsilon$, then

$$|nx_j - ny_j| = |L_j(x_j) - z_j| > 2s,$$

26

and therefore

$$
\begin{aligned}
\left\| \left(n\mathbf{x} + \frac{\mathbf{m}}{n} \right) - \mathbf{z} \right\|_\infty
&\geq \left| \left(nx_j + \frac{m_j}{n} \right) - z_j \right| \\
&= \left| nx_j - ny_j - \left(z_j - \left(ny_j + \frac{m_j}{n} \right) \right) \right| \\
&\geq |nx_j - ny_j| - \left| z_j - \left(ny_j + \frac{m_j}{n} \right) \right| \\
&> 2s - r \\
&\geq s,
\end{aligned}
$$

showing that $n\mathbf{x} + \frac{\mathbf{m}}{n} \notin B_s^\infty(\mathbf{z})$, which in turn implies $f_{n,\mathbf{m}}(\mathbf{x}) = 0$.

Let us see that the support of $f_{n,\mathbf{m}}$ is a bounded subset of $\mathbb{W}$. To this end, let $\mathbf{x} \in \mathsf{supp}(f_{n,\mathbf{m}})$. Hence $n\mathbf{x} + \frac{\mathbf{m}}{n} \in \mathsf{supp}(f)$, and $\|(n\mathbf{x} + \frac{\mathbf{m}}{n}) - \mathbf{z}\|_\infty < s$. Then for $j = 1, 2$,

$$
\begin{aligned}
|nx_j - ny_j| &\leq \left| (nx_j - ny_j) + z_j - \left(ny_j + \frac{m_j}{n} \right) \right| + \left| z_j - \left(ny_j + \frac{m_j}{n} \right) \right| \\
&= \left| nx_j + z_j - \frac{m_j}{n} \right| + \left| z_j - \left(ny_j + \frac{m_j}{n} \right) \right| \\
&\leq s + r < 2s,
\end{aligned}
$$

and thus $|x_j - y_j| < \frac{2s}{n} < \epsilon$. Therefore $\mathbf{x} \in B_\epsilon^\infty(\mathbf{y}) \subset \mathbb{W}$, and so $f_{n,\mathbf{m}} \in C_c(\mathbb{W})$.

Thus far we have that $f_{n,\mathbf{m}}(\mathbf{y}) \neq 0$, and that if $\mathbf{x} \notin B_\epsilon^\infty(\mathbf{y})$ then $f_{n,\mathbf{m}}(\mathbf{x}) = 0$. This observation, paired with $B_\epsilon^\infty(\mathbf{y}) \cap (S \cup T) = \{\mathbf{y}\}$, implies that

$$
\nu_{f_{n,\mathbf{m}}}(S, \mu) = \mu(\mathbf{y}) f_{n,\mathbf{m}}(\mathbf{y}) \neq 0.
$$

If $\mathbf{y} \notin T$ then we have that $\nu_{f_{n,\mathbf{m}}}(T, \alpha) = 0$; and if $\mathbf{y} \in T$ then

$$
\nu_{f_{n,\mathbf{m}}}(T, \alpha) = \alpha(\mathbf{y}) f_{n,\mathbf{m}} \neq \mu(\mathbf{y}) f_{n,\mathbf{m}} = \nu_{f_{n,\mathbf{m}}}(S, \mu),
$$

showing that $\mathcal{F}_{\mathcal{T}}$ separates points in $\mathcal{D}$.

Let us now show that if f is Lipschitz and $\mathcal{S} \subset \mathcal{D}$ is relatively compact, then the elements of $\mathcal{F}_{\mathcal{T}}$ are Lipschitz on $\mathcal{S}$. Indeed, let $D, D' \in \mathcal{S}$, and let $\delta > 0$ be so that $d_B(D, D') < \delta$. Moreover, fix a δ-matching $\mathsf{M} : S'_\mu \longrightarrow T'_\alpha$ between $D = (S, \mu)$ and $D' = (T, \alpha)$. Since $\mathcal{S}$ is relatively compact, then it is uniformly off-diagonally finite (see Def. 9), and hence there exists a uniform upper bound $\beta > 0$ for the multiplicity in $\mathsf{supp}(f)$ of any diagram in $\mathcal{S}$—see Definition 1. If $L > 0$ is the Lipschitz constant of f with respect to the sup norm $\|\cdot\|_\infty$ on $\mathbb{W}$, then for every $n \in \mathbb{N}$ and all $\mathbf{m} \in \mathbb{Z}^2$

$$
\begin{aligned}
|\nu_{f_{n,\mathbf{m}}}(D) - \nu_{f_{n,\mathbf{m}}}(D')|
&\leq \sum_{\substack{(\mathbf{y},n)=\mathsf{M}(\mathbf{x},k) \\ (\mathbf{x},k) \in S'_\mu}} |f(\mathbf{x}) - f(\mathbf{y})| \\
&\leq \sum_{(\mathbf{x},k) \in S'_\mu} L \cdot \delta \\
&\leq \beta \cdot L \cdot \delta.
\end{aligned}
$$

Since this inequality holds for any $\delta > d_B(D, D')$, it readily follows that

$$
|\nu_{f_{n,\mathbf{m}}}(D) - \nu_{f_{n,\mathbf{m}}}(D')| \leq \beta \cdot L \cdot d_B(D, D'),
$$

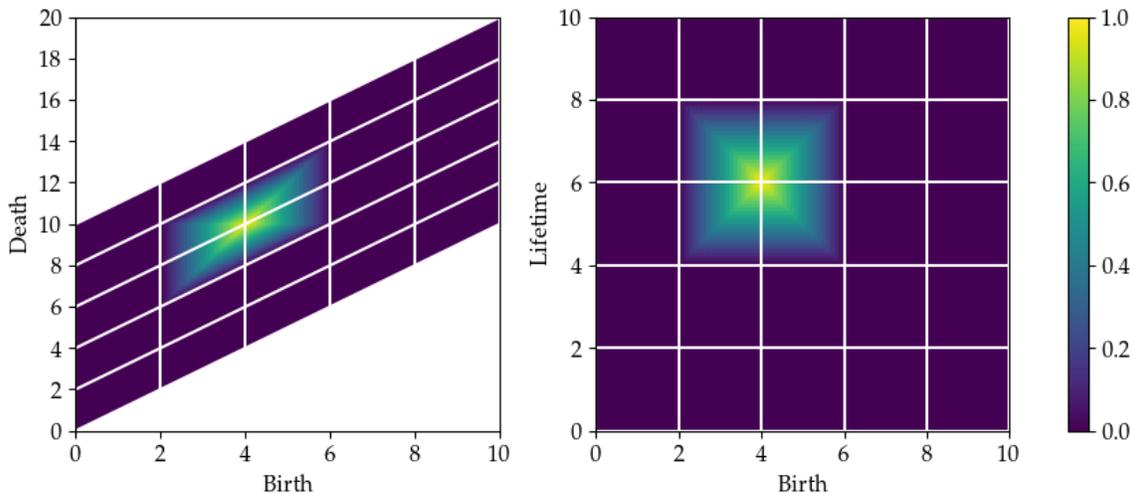and hence $\nu_{f_{n,\mathbf{m}}}$ is Lipschitz on $\mathcal{S}$. ∎

Figure 4: Tent function $g_{2,3}$ drawn in the birth-death plane (left) and in the birth-lifetime plane (right) for $d = 5$, $\delta = 2$, and $\epsilon = 0$.

6. Example template functions

The mathematical framework built to this point leaves open the choice of template system. In our experiments, we use two collections of functions, but we have no reason to suspect that these are the only or even the best available options. The first, which we call tent functions, are described in Section 6.1. The second are interpolating polynomials, traditionally used for approximating functions, which are described in Section 6.2.

For the entirety of this section, we will define functions on the birth-lifetime plane as this simplifies notation substantially. We use the tilde to denote the portions that are defined in this plane to emphasize the change from the birth-lifetime plane. So, let $\widetilde{\mathbb{W}} = \{(x, y) \mid x \in \mathbb{R}, y \in \mathbb{R}_{<0}\}$; that is, the conversion of $\mathbb{W}$ to the birth-lifetime plane. Likewise, let $\widetilde{\mathbb{W}^\epsilon} = \{(x, y) \in \widetilde{\mathbb{W}} \mid y > \epsilon\}$ so that it is the conversion of $\mathbb{W}^\epsilon$ to the birth-lifetime plane. Given a point $\mathbf{x} = (a, b) \in \mathbb{W}$, we write $\tilde{\mathbf{x}} = (a, b - a) \in \widetilde{\mathbb{W}}$ for the converted point. Given a diagram $D = (S, \mu)$, we write $\widetilde{D} = (\widetilde{S}, \mu)$ where $\widetilde{S} = \{\tilde{\mathbf{x}} \mid \mathbf{x} \in S\}$ and $\tilde{\mu}(\tilde{\mathbf{x}}) = \mu(\mathbf{x})$.

6.1 Tent functions

We first define a template system in the birth-lifetime plane which we call tent functions. Given a point $\mathbf{a} = (a, b) \in \widetilde{\mathbb{W}}$ and a radius $\delta \in \mathbb{R}_{>0}$ with $0 < \delta < b$, define the tent function on $\widetilde{\mathbb{W}}$ to be

$$g_{\mathbf{a},\delta}(x, y) = \left| 1 - \frac{1}{\delta} \max\{|x - a|, |y - b|\} \right|_+$$

where $|\cdot|_+$ is the positive part of the function, and 0 otherwise. As $\delta < b$, this function has support in the compact box $[a - \delta, a + \delta] \times [b - \delta, b + \delta] \subset \widetilde{\mathbb{W}}$.

28

Given a persistence diagram $D = (S, \mu)$, the tent function is defined to be the sum over the evaluation on the points in the diagram, namely

$$G_{\mathbf{a}, \delta}(D) = \widetilde{G}_{\mathbf{a}, \delta}(\tilde{D}) = \sum_{\tilde{\mathbf{x}} \in \tilde{S}} \widetilde{\mu}(\mathbf{x}) \cdot g_{\mathbf{a}, \delta}(\tilde{\mathbf{x}}).$$

We use G or $\widetilde{G}$ depending on whether we want our input to be a diagram in the birth-death or birth-lifetime plane, respectively, but all subscript notation is written in the birth-lifetime plane for ease of notation.

We then have the following theorem.

Theorem 31 *The collection of tent functions*

$$\left\{ G_{\mathbf{a}, \delta} \mid \mathbf{a} = (a, b) \in \widetilde{\widetilde{\mathbb{W}}}, 0 < \delta < b \right\}$$

separates points in $\mathcal{D}$.

Proof We are given two persistence diagrams $D_1 = (S_1, \mu_1)$ and $D_2 = (S_2, \mu_2) \in \mathcal{D}$, with $D_1 \neq D_2$. WLOG there is an $\mathbf{x} \in S_1$ for which either (i) $\mathbf{x} \notin S_2$ or (ii) $\mathbf{x} \in S_2$ but $\mu_1(\mathbf{x}) > \mu_2(\mathbf{x})$. For ease of notation, assume in case (ii) that $\mathbf{x} \in S_2$ and $\mu_2(\mathbf{x}) = 0$. Then we always have $\mathbf{x} \in S_2$ and $\mu_1(\mathbf{x}) > \mu_2(\mathbf{x})$.

Let $\tilde{\mathbf{x}} = (a, b)$. For any δ, define $\widetilde{B}_\delta = [a - \delta, a + \delta] \times [b - \delta, b + \delta]$ and note that this is the support of $g_{\mathbf{x}, \delta}$. As D_1 and D_2 are in $\mathcal{D}$, both diagrams have finite multiplicity in $\widetilde{\mathbb{W}^{b/2}}$. So, there exists a $\delta < b/2$ so that $\widetilde{S_1} \cap \widetilde{B_\delta} = \{\mathbf{x}\} = \widetilde{S_2} \cap B$. As $\mathbf{x}$ is the only point in either diagram in the support of $g_{\mathbf{x}, \delta}$,

$$G_{\mathbf{x}, \delta}(D_1) = \mu_1(\mathbf{x}) > \mu_2(\mathbf{x}) = G_{\mathbf{a}, \delta}(D_2).$$

Thus, the collection of tent functions separates points. ∎

For practical purposes, we pick a subset of these tent functions. Let $\delta > 0$ be the partition scale, let d the number of subdivisions along the diagonal (resp. y axis), and let $\epsilon > 0$ be the upward shift. In our experiments described in Section 8, we use the collection of tent functions given by

$$\left\{ G_{(\delta i, \delta j + \epsilon), \delta} \mid 0 \leq i \leq d, 1 \leq j \leq d \right\}. \tag{9}$$

That is, these are the tent functions centered at a regular grid shifted up by ϵ to ensure that g is supported on a compact set in $\widetilde{\mathbb{W}}$. See Fig. 4 for an example.

6.2 Interpolating polynomials

Say we are given a nonempty, finite set of distinct mesh values $\mathcal{A} = \{a_i\}_{i=0}^m \subset \mathbb{R}$ and a collection of evaluation values $\{c_i \in \mathbb{R}\}$, the first goal is to build a polynomial such that $f(a_i) = c_i$ for all i. The Lagrange polynomial $\ell_j^{\mathcal{A}}(x)$ corresponding to node a_j is defined as

$$\ell_j^{\mathcal{A}}(x) = \prod_{i \neq j} \frac{x - a_i}{a_j - a_i}. \tag{10}$$

Note that this function satisfies

$$\ell_j^{\mathcal{A}}(a_k) = \begin{cases} 1, & j = k, \\ 0, & \text{otherwise,} \end{cases} \qquad \text{and} \qquad \sum_{j=0}^{m} \ell_j^{\mathcal{A}}(x) = 1.$$

The Lagrange interpolation polynomial is then simply $f(x) = \sum_{j=0}^{m} c_j \ell_j^{\mathcal{A}}(x)$. Note that for numerical stability, one must work with the barycentric form of Lagrange interpolation formula described by Berrut and Trefethen (2004) and shown in Appendix A.

We will now use these polynomials to create functions on $\widetilde{\mathbb{W}}$. Assume we have two collections of mesh points $\mathcal{A} = \{a_i\}_{i=0}^{m} \subset \mathbb{R}$ and $\mathcal{B} = \{b_i\}_{i=0}^{n} \subset \mathbb{R}_{>0}$ so that $(a_i, b_j) \in \widetilde{\mathbb{W}}$ for all i, j. Then, given a collection of evaluation points $\mathcal{C} = \{c_{i,j}\}_{i,j} \subset \mathbb{R}$, we want to build a polynomial for which $f(a_i, b_j) = c_{i,j}$. Note that in general the evaluation points $\mathcal{C}$ are not part of the given persistence diagrams; nevertheless, these values are not needed in our construction but we do keep track of their coefficients. We define the 2D interpolating polynomial for the collection $\mathcal{A}, \mathcal{B}, \mathcal{C}$ to be

$$f(x, y) = \sum_{i=0}^{m} \sum_{j=0}^{n} c_{i,j} \cdot g\left(\ell_i^{\mathcal{A}}(x) \cdot \ell_j^{\mathcal{B}}(y)\right). \tag{11}$$

where $g(\cdot)$ is either the identity function or $g(\cdot) = |\cdot|$; in our experiments we used the latter for the simple reason that it seemed to give better results. We now evaluate f at each of the N query points which are the points of a persistence diagram in $\widetilde{\mathbb{W}}$ to obtain N equations that we can write as

$$\mathbf{f} = \mathbf{L}\,\mathbf{c}, \tag{12}$$

where $\mathbf{L}$ is an $N \times ((m+1) \times (n+1))$ matrix, and $\mathbf{f}$ is an $(m+1) \times (n+1)$ vector obtained by concatenating a 2D mesh, similar to the one shown in Fig. 6, row-wise.

Each column of matrix $\mathbf{L}$ in Eq. (12) represents a vector that describes the contributions of all the query points to the corresponding entry in $\mathbf{f}$.

Renumbering the entries in vector $\mathbf{f}$ according to $ij \to i(n+1) + j + 1 = r$ where $r \in \{0, \ldots, (m+1)(n+1)\}$, we can now assign a score for each point in the mesh using the map $S_r : \mathbb{R}^N \to \mathbb{R}$, i.e., by operating on the rows of matrix $\mathbf{L}$ according to

$$S_r = \sum_{j=0}^{N-1} L_{j,r}. \tag{13}$$

Choosing a larger base mesh implies using a higher degree polynomial in the interpolation. Therefore, the role of increasing the degree of the polynomial is similar to the role of increasing the number of tent functions. A larger mesh leads to more features which gives a tool for either increasing or reducing the number of features. The former improves the fit to the training set, while the latter reduces the number of features which allows mitigating overfitting effects. While any class of interpolating polynomials can be used, in this study we chose Chebyshev interpolating polynomials due to their excellent approximation properties, see Trefethen (2012). Appendix A describes how to use the interpolation matrices separately obtained for each of the birth times and lifetimes of a given persistence diagram to construct $L_{j,r}$.
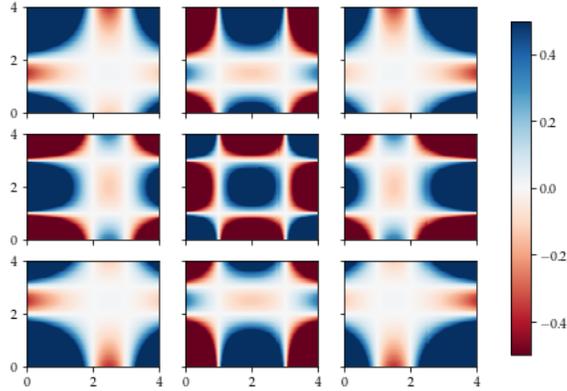
Figure 5: An example of interpolating polynomials for the mesh $\mathcal{A} = \mathcal{B} = \{1, 2, 3\}$. For evaluation values $\mathcal{C}_{i,j} = \{c_{i,j} = 1, c = 0 \text{ else}\}$, $p_{i,j}$ is drawn at location (i, j) in the figure, with $(1, 1)$ at the bottom left.

Fix a compact region $K \subset \widetilde{\mathbb{W}}$; a $\epsilon \in \mathbb{R}_{<0}$ such that the compact set

$$\overline{K^\epsilon} = \{\mathbf{x} \in \mathbb{R}^2 \mid \|\mathbf{x} - K\| \leq \epsilon\} \subset \widetilde{\mathbb{W}};$$

and a collection of mesh points $\{(a_i, b_j)\} \subset K$ given by $\mathcal{A}$ and $\mathcal{B}$ as above. Define $h_{K,\epsilon}$ to be a continuous function on $\widetilde{\mathbb{W}}$ such that

$$h_{K,\epsilon}(\mathbf{x}) = \begin{cases} 1 & \mathbf{x} \in K \\ 0 & \|\mathbf{x} - K\| \geq \epsilon \end{cases}$$

given by Urysohn's lemma (Munkres (2000)). Note that the support of this function is contained in the compact set $\overline{K^\epsilon} \subset \widetilde{\mathbb{W}}$.

Let $\mathcal{C}_{i,j}$ be the collection of evaluation values which are entirely 0 except for $c_{i,j} = 1$. Define $f_{i,j}^{\mathcal{A},\mathcal{B}} = f_{i,j}$ to be the interpolating polynomial (Eq. (11)) for this setup. Then the function on diagrams is defined to be

$$F_{i,j}^{\mathcal{A},\mathcal{B},K,\epsilon}(D) = \widetilde{F}_{i,j}^{\mathcal{A},\mathcal{B},K,\epsilon}(\widetilde{D}) := \sum_{\tilde{\mathbf{x}} \in \tilde{S}} \mu(\tilde{\mathbf{x}}) \cdot f_{i,j}^{\mathcal{A},\mathcal{B}}(\tilde{\mathbf{x}}) \cdot h_{K,\epsilon}(\tilde{\mathbf{x}}).$$

We have the following theorem to show that these interpolating polynomials can be used as template functions.

Theorem 32 *The collection of interpolating polynomials*

$$\left\{ F_{i,j}^{\mathcal{A},\mathcal{B},K,\epsilon} \right\}$$

separates points, where the collection varies over all choices of compact $K \subset \widetilde{\mathbb{W}}$, $\delta \in \mathbb{R}_{>0}$, and of mesh $\mathcal{A}$, $\mathcal{B}$ as specified above.
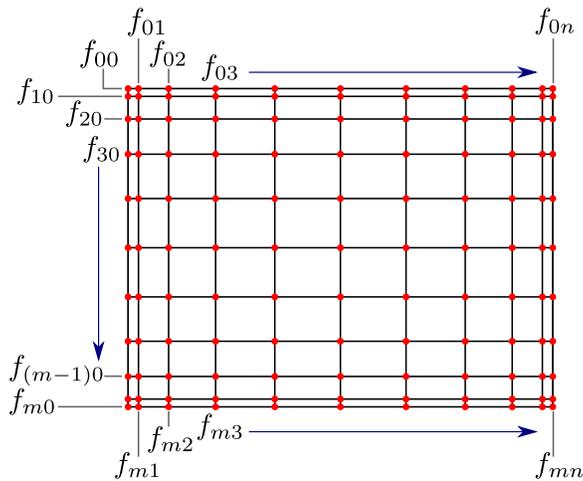
Figure 6: An example 11×11 2D grid with $m = n = 10$ defined using the Chebyshev points of the second kind.

Proof We are given two persistence diagrams $D_1 = (S_1, \mu_1)$ and $D_2 = (S_2, \mu_2) \in \mathcal{D}$, with $D_1 \neq D_2$. WLOG there is an $\mathbf{x} = (a, b) \in S_1$ for which either (i) $\mathbf{x} \notin S_2$ or (ii) $\mathbf{x} \in S_2$ but $\mu_1(\mathbf{x}) > \mu_2(\mathbf{x})$. To avoid case checking, we assume as before that $\mathbf{x} \in S_2$ with $\mu_2(\mathbf{x}) = 0$ in the later case. This way, in both cases have $\mathbf{x} \in S_2$ and $\mu_1(\mathbf{x}) > \mu_2(\mathbf{x})$.

Choose a compact set $K \ni \mathbf{x}$ and ϵ both small enough so that $\overline{K^\epsilon} \cap S_1 = \mathbf{x}$ and $\overline{K^\epsilon} \cap S_2 = \mathbf{x}$. If $\widetilde{\mathbf{x}} = (a, b)$, set $\mathcal{A} = \{a\}$ and $\mathcal{B} = \{b\}$. Note that in this overly simplistic setup, $\ell^{\mathcal{A}}(x) = x/a$ and $\ell^{\mathcal{B}}(y) = y/b$, so the only interpolating polynomial is $f(x, y) = g((xy)/(ab))$. Whether g is the identity or the absolute value function, f evaluates to 1 at $\widetilde{\mathbf{x}}$. Because the only point in either diagram inside $\overline{K^\epsilon}$ is $\mathbf{x}$, $h_{K,\epsilon}(\widetilde{\mathbf{x}}) = 1$ and $h_{K,\epsilon}(\widetilde{\mathbf{y}}) = 0$ for every other $\mathbf{y} \in S_1 \cup S_2$, so

$$F(D_i) = \sum_{\widetilde{\mathbf{x}} \in \tilde{S}} \mu(\widetilde{\mathbf{x}}) \cdot f(\widetilde{\mathbf{x}}) \cdot h_{K,\epsilon}(\widetilde{\mathbf{x}}) = \mu_i(\mathbf{x})$$

for $i = 1, 2$. Thus F separates the two diagrams. ∎

In our experiments, we set K to be a box $[A, A'] \times [B, B']$ with $B > 0$, ϵ to be either machine precision or $B/2$, and use the non-uniform Chebyshev mesh as seen in Fig. 6 (Trefethen (2012)). See Appendix A for an explanation of the vectorization used to implement the code.

7. Implementing regularized regression/classification

The results thus far imply that template systems on $\mathcal{D}$ can be used to vectorize collections of persistence diagrams; and that these vectorizations, in turn, can be used as inputs to machine learning algorithms for classification and regression tasks. We describe next one avenue for implementing these ideas in practice. Indeed, given a finite collection of labeled

persistence diagrams

$$\{(D_m, \ell_m)\}_{m=1}^{M} \subset \mathcal{D} \times L$$

with $L \subset \mathbb{R}$, and a template system $\mathcal{T} \subset C_c(\mathbb{W})$, the goal is to find $N \in \mathbb{N}$, template functions $f_1, \ldots, f_N \in \mathcal{T}$, and a polynomial $p \in \mathbb{R}[x_1, \ldots, x_N]$, such that the function

$$
\begin{aligned}
P : \quad \mathcal{D} &\longrightarrow \mathbb{R} \\
D &\mapsto p\big(\nu_{f_1}(D), \ldots, \nu_{f_N}(D)\big)
\end{aligned}
$$

satisfies $P(D_m) \approx \ell_m$ for $m = 1, \ldots, M$. It follows from Theorem 29 that this process results in arbitrarily accurate approximations on compact subsets of $\mathcal{D}$, provided the labels ℓ_m vary continuously. In practice the integer N and the template functions $f_1, \ldots, f_N \in C_c(\mathbb{W})$ are user-provided parameters, though adaptive and data-driven approaches can be implemented. This direction is out of the scope of the current work. Given $N \in \mathbb{N}$ and template functions $f_1, \ldots, f_N \in C_c(\mathbb{W})$, the optimal polynomial $p \in \mathbb{R}[x_1, \ldots, x_N]$ is uniquely determined by its vector of coefficients, $\mathbf{a} \in \mathbb{R}^k$. We will make this explicit with the notations $p_{\mathbf{a}}$ and $P_{\mathbf{a}}$, and an optimization will be set up in order to determine $\mathbf{a} \in \mathbb{R}^k$ from the available labeled data.

The error of fit $P_{\mathbf{a}}(D_m) \approx \ell_m$ is measured in the usual way via a loss function

$$\mathcal{E} : \mathbb{R} \times L \longrightarrow \mathbb{R}$$

where common choices include:

Square is given by $\mathcal{E}_{sq}(t, \ell) = (t - \ell)^2$, $L = \mathbb{R}$, and yields a least-squares regression. Can handle multi-class classification.

Hinge is given by $\mathcal{E}_{hg}(t, \ell) = \max\{0, 1 - \ell \cdot t\}$, with $L = \{-1, 1\}$, and appears in the soft-margin classifier of support vector machine.

Logistic is given by the log-loss $\mathcal{E}_{log}(t, \ell) = \ln\big(1 + e^{-\ell \cdot t}\big)$, with $L = \{-1, 1\}$, and yields logistic regression.

Meanwhile, the complexity of the model can be measured, for instance, via a regularization function

$$\Omega : \mathbb{R}^k \longrightarrow [0, \infty).$$

The regularized optimization scheme looking to minimize the regularized mean loss is therefore

$$\mathbf{a} = \operatorname*{argmin}_{\mathbf{v} \in \mathbb{R}^k} \frac{1}{M} \sum_{m=1}^{M} \mathcal{E}\big(P_{\mathbf{v}}(D_m), \ell_m\big) + \lambda \Omega(\mathbf{v})$$

where $\lambda > 0$ is the regularization parameter, often chosen from the set $\{10^n\}_{n \in \mathbb{Z}}$.

7.1 Visualization of Coefficients

Our collections of template functions have a uniquely 2d geometric flavor. In particular, for both tent functions and Lagrange polynomials on our formulation, we have a function for each a_i, b_j location on a mesh. This means that we can pull the coefficients $\mathbf{v}$ determined in the optimization back to the grid which built them for visualization by drawing a heat map with $v_{i,j}$ drawn at (i, j) to more fully understand the model. Examples of this are shown in Figs. 10 and 12.

8. Experiments

8.1 Code

Code for doing classification and regression using tent and interpolating polynomial functions is available in the python `teaspoon` package[2]. Classification and regression were done using the `RidgeClassifierCV` and `RidgeCV` functions from the `sklearn` package, respectively. Scores for classification experiments are reported using the percent that were correctly classified. Scores for regression experiments are reported using the coefficient of determination, R^2. Note that this latter score can potentially take negative values; perfect regression would score 1, and a method which returns the constant prediction of the expected value is given a score of 0.

8.2 Off-diagonal, normally distributed points

We generate diagrams from the following procedure. Given μ and σ, draw n points from the gaussian $N(\mu, \sigma)$ on $\mathbb{R}^2$. Retain all points which are are in $\mathbb{W}$. For our simulations, we fixed $\sigma = 1$ and varied μ. Examples of two overlaid example diagrams are shown in Fig. 7.
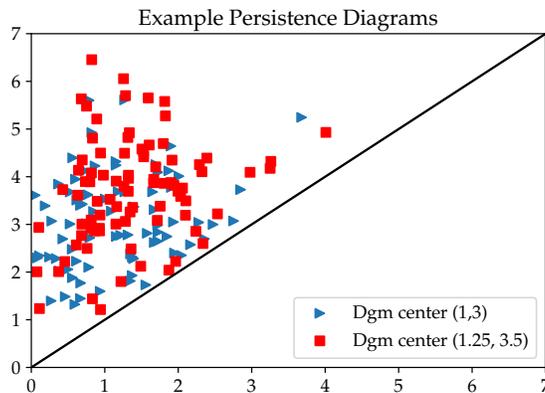

Figure 7: Example diagrams randomly generated by the procedure described in Sec. 8.2. Two diagrams drawn from a distributions with different choices of μ are shown.

Classification. We tested our classification accuracy with the following experiment. We chose two collections, A and B, of 500 persistence diagrams each generated by drawing $n = 20$ points (note that this means there are at most n points). The means μ were different: $\mu_A = (1, 3)$ and μ_B was varied along the line $(1, 3)t + (2, 5)(1 - t)$ for $t \in [0, 1]$. Classification accuracy using tent and interpolating polynomial functions is shown in Fig. 8. As expected, the correct classification percentage for the test set is around 50% when $\mu_A \sim \mu_B$, and improves as they move farther apart. In particular, by the time the means are at distance apart equal to the standard deviation used for the normal distribution ($\sigma = 1$), classification is well above 90%. For this particular experiment, we do not see any difference between the choice of template function used.
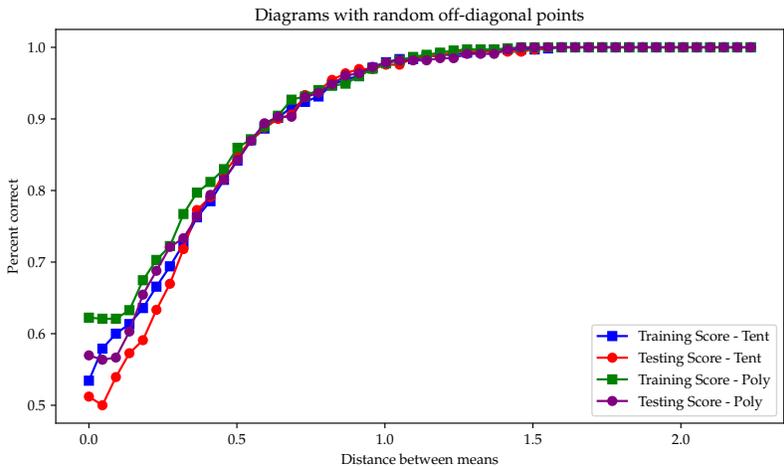
---

2. `https://github.com/lizliz/teaspoon`

Figure 8: Results from classification test for pairs of choices of μ with $\mu_A = (1,3)$ and μ_B chosen on the line from $(1,3)$ to $(2,5)$.

	Tents - Train	Tents - Test	Polynomials - Train	Polynomials - Test
Line	0.977 ± 0.002	0.970 ± 0.004	0.979 ± 0.001	0.970 ± 0.002
Ball	0.823 ± 0.023	0.782 ± 0.023	0.832 ± 0.021	0.786 ± 0.021

Table 1: The R^2 results of the regression tests described in Section 8.2.

Regression. We further ran two regression versions of the experiment as follows. In the first test, we drew 500 diagrams from the above procedure with a choice of center μ drawn uniformly on the line segment $t(1,3) + (1-t)(6,8)$, $t \in [0,1]$. Then, we predicted the distance from μ to $(1,3)$. We call this the "line" experiment. Second, we drew μ from the normal distribution $N((1,3),1)$ and again predicted the distance of μ from the point $(1,3)$. We call this the "ball" experiment.

Each of these experiments was run 10 times, and the results can be seen in Table 1. Example predictions for single runs can be seen in Fig. 9, and the coefficients for these examples are in Fig. 10. Note that the coefficients are drawn at the location of their index. In particular, the interpolating polynomials are determined using a non-uniform mesh, so the heatmap for these coefficients does not align with the location of the associated point.

8.3 Manifold experiment

Following an experiment run in Adams et al. (2017), we generated collections of point clouds drawn from different manifolds embedded in $\mathbb{R}^2$ or $\mathbb{R}^3$. Each point cloud has $N = 200$ points. The categories are as follows:

Annulus. Points drawn uniformly from an annulus with inner radius 1 and outer radius 2.

3 clusters. The 200 points are drawn from one of three different normal distributions, with means $(0,0)$, $(0,2)$ and $(2,0)$ respectively, and all with standard deviation 0.05.
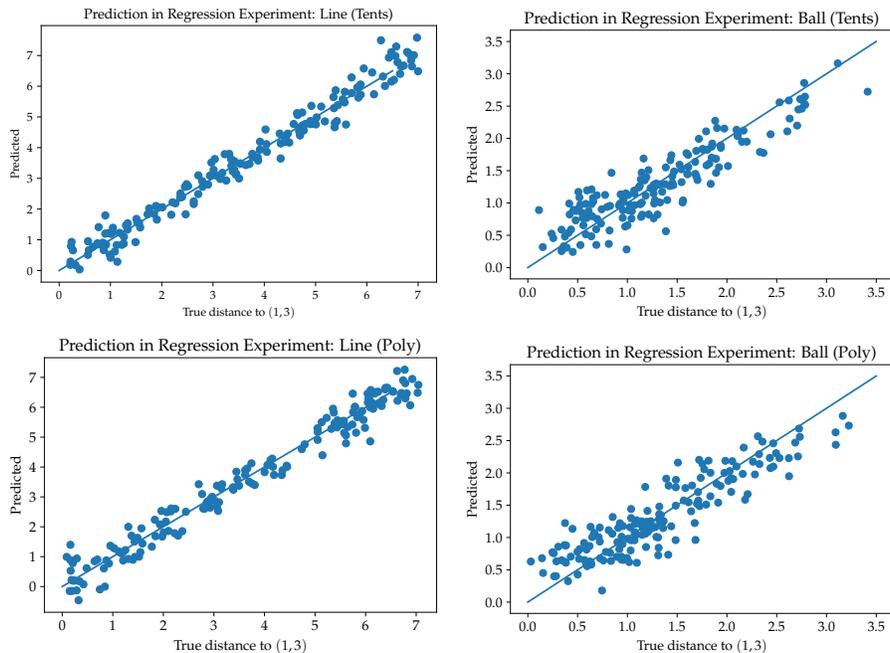
Figure 9: True vs predicted of distance to starting mean for the ball and line experiments described in Section 8.2.

3 clusters of 3 clusters. The points are drawn from normal distributions with standard deviation 0.05 centered at the points $(0,0)$, $(0, 1.5)$, $(1.5, 0)$, $(0, 4)$, $(1, 3)$, $(1, 5)$, $(3, 4)$, $(3, 5.5)$, $(4.5, 4)$.

Cube. Points drawn uniformly from $[0, 1]^2 \subseteq \mathbb{R}^2$.

Torus. Points drawn uniformly from a torus thought of as rotating a circle of radius 1 in the xz-plane centered at $(2,0)$ around the z-axis. The generation of the points is done using the method from Diaconis et al. (2013).

Sphere. Points drawn from a sphere of radius 1 in $\mathbb{R}^3$ with uniform noise in $[-0.05, 0.05]$ added to the radius.

Examples of each of these can be seen in Fig. 11. Code for generation of these point clouds as well as the full dataset can be found in the `teaspoon` package at `teaspoon.MakeData.PointCloud.testSetManifolds`.

The choice of tent function parameters was done as follows. We determined the bounding box necessary to enclose the training set diagrams in the (birth, lifetime) plane and added padding of 0.05. We fixed $d = 10$; ϵ was chosen to be half the minimum lifetime over all training set diagrams; then δ was chosen to ensure coverage of the bounding box.

We reserved 33% of the data for testing and trained a regression model on the remaining data. The results of this experiment averaged over 10 runs can be seen in Table 2. In this
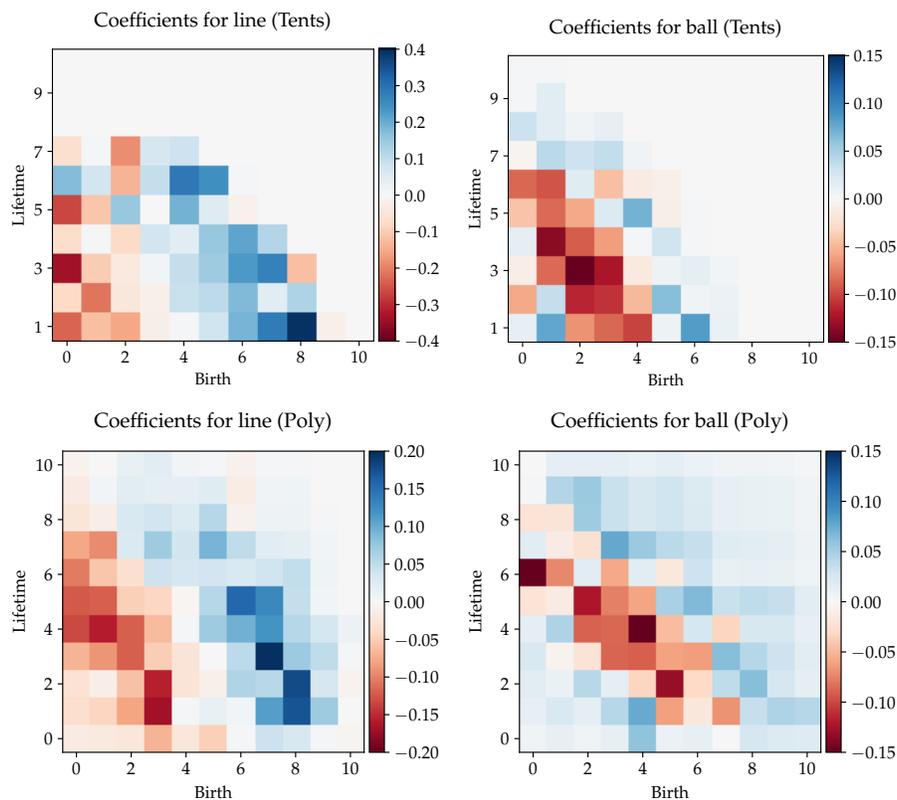
Figure 10: Coefficients for regression experiment with mean drawn from a line (left) and from a ball (right). The top row uses tent functions, the bottom uses interpolating polynomials. The x and y coordinates correspond to the index of the test function used.

experiment, particularly when we have 50 or more diagrams per class, we see excellent ($\geq 99\%$) classification.

8.4 Shape Data

We compared our results to the kernel method results reported in Reininghaus et al. (2015) by applying feature functions to the same data set from that paper. In particular, the synthetic SHREC 2014 data set (Pickup et al. (2014)) consists of 3D meshes of humans in different poses. The people are labeled as male, female, and child (five each); and each person assumes one of 20 poses. Reininghaus et al. defined a function on each mesh using the heat kernel signature (Sun et al. (2009)) for 10 parameters and computed the 0- and 1-dimensional diagrams of each.

We start with this data set of 300 pairs of persistence diagrams (0- and 1-dimensional) for each of the 10 parameter values, and predicted the human model; i.e. which of the 15 people were represented by each mesh. A comparison of the results reported in Reininghaus
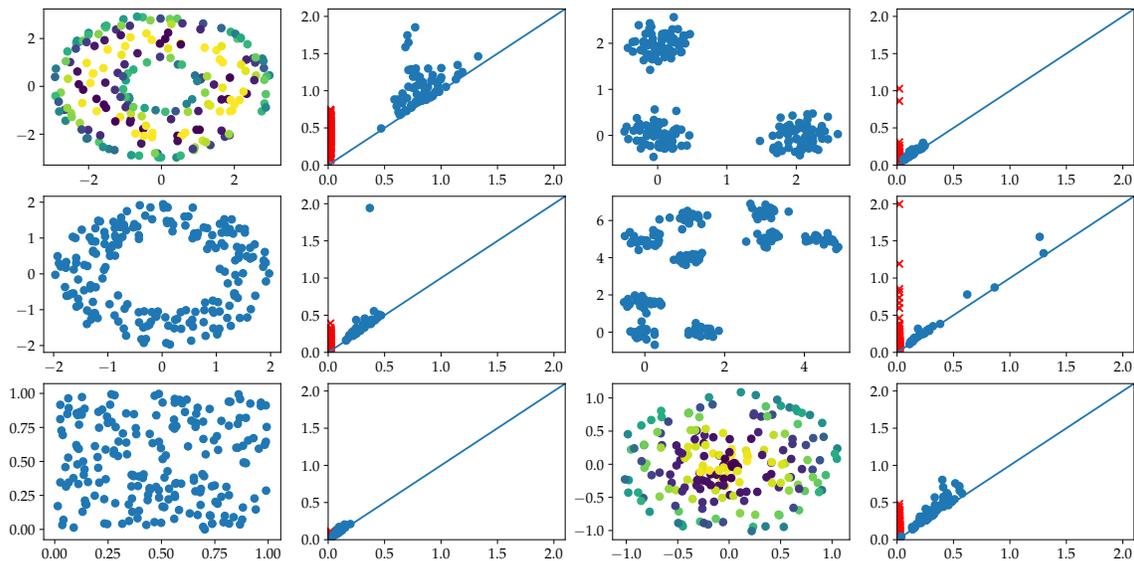
Figure 11: Example point clouds from the experiment described in Sec. 8.3. From top left reading across rows, the point clouds are a torus (in $\mathbb{R}^3$ but drawn with color as the third coordinate), 3 clusters, annulus, three clusters of three clusters, uniform box, and a sphere (again drawn with third coordinate as color). The associated persistence diagrams are shown to the right of the point clouds; the 1-dimensional diagram is given by blue dots and the 0-dimensional diagram is shown by red x's. Diagrams are drawn with the same axis values.

| | Tents | | Polynomials | |
No. Dgms	Train	Test	Train	Test
10	$99.8\% \pm 0.9$	$96.5\% \pm 3.2$	$99.8\% \pm 0.9$	$95.0\% \pm 3.9$
25	$99.9\% \pm 0.3$	$99.0\% \pm 1.0$	$99.7\% \pm 0.5$	$97.6\% \pm 1.5$
50	$99.9\% \pm 0.2$	$99.9\% \pm 0.3$	$100\% \pm 0$	$99.2\% \pm 0.9$
100	$99.8\% \pm 0.1$	$99.7\% \pm 0.4$	$99.6\% \pm 0.2$	$99.3\% \pm 0.5$
200	$99.5\% \pm 0.1$	$99.5\% \pm 0.3$	$99.2\% \pm 0.2$	$98.9\% \pm 0.5$

Table 2: Results from the manifold test described in Sec. 8.3 for different numbers of examples drawn for each type of manifold. The reported results are averaged over 10 experiments each.

et al. (2015) with our method using polynomial functions is shown in Table 3. Additional results using tent functions are provided in the appendix (Table 4).

For this experiment, polynomial features (Table 3) were considerably more successful than the tent functions. Further, using the 0- and 1-dimensional persistence diagrams
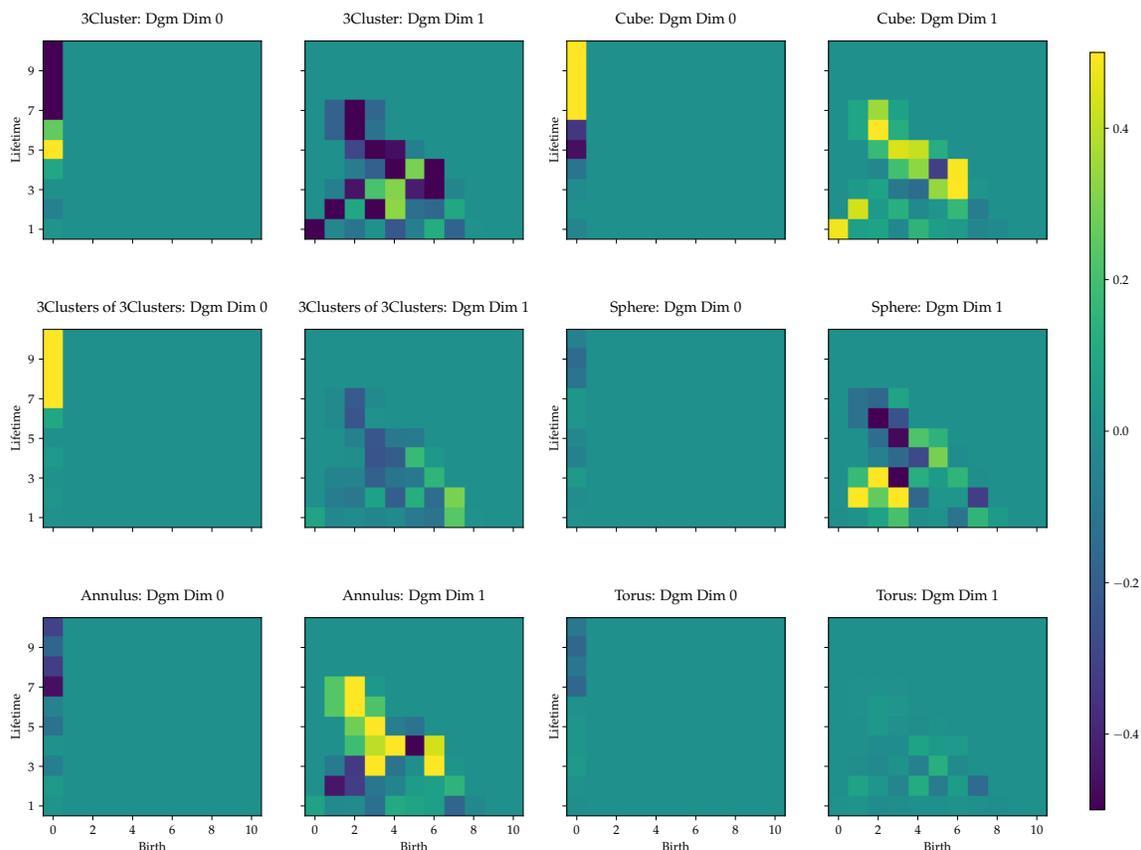
Figure 12: Coefficients for the manifold experiment using tent functions run with 100 diagrams each.

together was largely better than the 1-dimensional diagram alone, and considerably better than the 0-dimensional diagram alone. The average classification rates were improved in four out of the ten parameter choices; results with intersecting confidence intervals occurred in an additional four out of ten parameter choices.

8.5 Rossler Periodicity

We tested our machine learning approach on time series simulated from the Rossler system (McCullough et al. (2015))

$$
\begin{aligned}
\dot{x} &= -y - z, \\
\dot{y} &= x + \alpha y, \\
\dot{z} &= \beta + z, (x - \gamma),
\end{aligned}
\tag{14}
$$

where the overdot denotes a derivative with respect to time. We used an explicit Runge-Kutta (4,5) formula to solve the Rossler system for $\beta = 2$, $\gamma = 4$, and 1201 evenly spaced

| freq | MSK | Dim 0 | | Dim 1 | | Dim 0 & Dim 1 | |
		Train	Test	Train	Test	Train	Test
1	94.7% ± 5.1	94.3% ± 0.5	67.1% ± 4.7	99.1% ± 0.3	85.4% ± 3.0	99.8% ± 0.3	90.4% ± 5.3
2	99.3% ± 0.9	92.1% ± 1.4	60.8% ± 6.3	99.9% ± 0.3	89.9% ± 1.5	100.% ± 0.0	95.1% ± 2.4
3	96.3% ± 2.2	83.4% ± 2.4	45.1% ± 2.9	99.6% ± 0.5	88.9% ± 3.0	99.7% ± 0.5	90.0% ± 2.0
4	97.3% ± 1.9	74.7% ± 2.0	37.4% ± 4.7	99.1% ± 0.7	85.2% ± 2.5	98.6% ± 0.9	84.8% ± 3.9
5	96.3% ± 2.5	65.3% ± 2.9	27.8% ± 5.0	99.2% ± 0.7	93.0% ± 2.2	99.7% ± 0.4	93.3% ± 2.2
6	93.7% ± 3.2	67.2% ± 2.5	36.5% ± 3.6	99.2% ± 0.5	93.4% ± 2.8	98.8% ± 0.5	92.9% ± 1.8
7	88.0% ± 4.5	71.5% ± 2.8	40.9% ± 4.1	98.3% ± 0.7	96.6% ± 0.7	99.0% ± 0.4	95.6% ± 1.4
8	88.3% ± 6.0	84.2% ± 3.3	63.0% ± 4.5	99.0% ± 0.5	93.0% ± 1.8	99.6% ± 0.4	94.0% ± 2.2
9	88.0% ± 5.8	83.5% ± 2.7	62.4% ± 5.0	98.4% ± 1.2	92.9% ± 1.5	98.5% ± 1.3	92.6% ± 2.1
10	91.0% ± 4.0	79.8% ± 2.7	59.0% ± 4.6	96.9% ± 0.6	92.1% ± 1.7	97.7% ± 1.1	89.5% ± 4.6

Table 3: Results of classification of shape data discussed in Section 8.4. The function used was the Chebyshev polynomial of the second kind. The MSK column gives the original results from Reininghaus et al. (2015); the subsequent columns use the 0-dimensional diagrams only, the 1-dimensional diagrams only, and both, respectively. Scores highlighted in blue give best average score MSK vs. template functions; scores highlighted in orange have overlapping intervals with the best score. Compare this to the results with tent functions, Table 4.
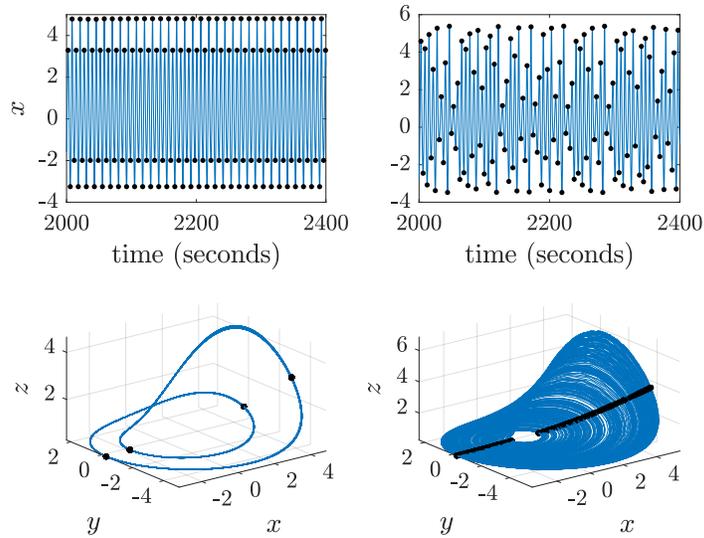

Figure 13: Time series for the x value in a Rossler system (top row) and the corresponding phase portrait in the (x, y, z) space (bottom row) for a periodic case with $\alpha = 0.37$ (left column), and a chaotic case with $\alpha = 0.42$ (right column). The superimposed black dots correspond to the extrema of x.

values of the bifurcation parameter α where $0.37 \leq \alpha \leq 0.43$. For each value of α a set of initial conditions was sampled from uniformly distributed values in $[0, 1]$. We simulated 2×10^4 points using a time step of 0.2 seconds. Half of the simulated points were discarded, and only the second half of the x variable data was used in the current analysis. The left and right columns of Fig. 13 show two examples of the resulting time series: one periodic with $\alpha = 0.37$, and one chaotic with $\alpha = 0.42$, respectively. The first row of the figure shows the time series after dropping the first half of the simulated data, and the bottom row shows the corresponding phase space. The black dots in Fig. 13 represent the extrema of x which were accurately computed using a modified version of Henon's algorithm (Henon (1982); Palaniyandi (2009)). These dots are basically the Poincaré points obtained by finding all the intersections of the x trajectory with the surface $\dot{x} = 0$.

The two examples in Fig. 13 show how the bifurcation parameter α can influence the system behavior. This dependence on α is further illustrated in the top graph of Fig. 15 which depicts the maximum Lyapunov exponent computed using the algorithm described by Benettin et al. (1980); Eckmann and Ruelle (1985); and Sandri (1986). The bottom graph of Fig. 15 shows the score of the zero-one test for chaos (Gottwald and Melbourne (2004, 2009, 2016)): a binary test that yields a score of 0 for regular dynamics, and 1 for chaotic dynamics. In order to avoid the failure of the 0-1 test due to oversampling, the test was applied to the subsampled data which was obtained by retaining every sixth point from the original signal. The periodic windows shown in Figs. 14 and 15 were identified by examining the plot of the bifurcation diagram in Fig. 14 as well as plots of the maximum Lyapunov exponent and the 0-1 test scores in Fig. 15.

We applied the feature function method on the resulting data set using the tent functions. In this experiment, we set $d = 10$, $\delta = 0.4$, and ϵ to be machine precision. For this test, we got a score of 98.9% on the training set, and 97.2% on the test set. The misclassified time series are superimposed on the bifurcation diagram in Fig. 14. In this figure, green circles show the values of α that were tagged chaotic but that the algorithm identified as periodic. Similarly, purple diamonds indicate the α values that were identified as chaotic even though they were tagged as periodic. It can be seen, unsurprisingly, that misclassification occurs near the transitions of the system behavior from chaotic to periodic or vice versa.

We note that tagging of the data used for both training and testing was performed by inspecting the bifurcation diagram, the maximum Lyapunov exponent plot, and the 0-1 test. Therefore, for the very few misclassified α values we would actually conjecture that our approach can provide a check for the correctness of tagging in the testing and training sets especially for the boundary cases.

9. Discussion

In this paper, we have provided a new method for the featurization of persistence diagrams through the use of template functions; that is, collections of functions compactly supported on the upper half plane away from the diagonal whose induced functions on diagrams separate points. To do this, we further gave a complete description of compact sets in persistence diagram space endowed with the bottleneck distance.
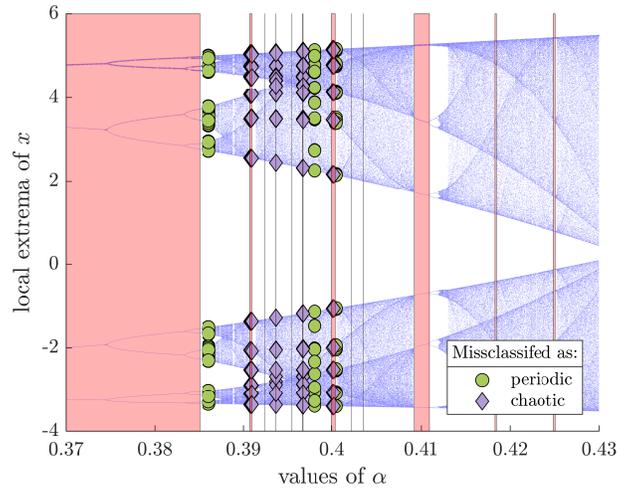
Figure 14: The bifurcation diagram for the Rossler system with α as the bifurcation parameter and the extrema of x as the response parameter. The shaded windows indicate the regions that were tagged as periodic. The misclassified points are superimposed with diamonds indicating the points that were incorrectly identified as chaotic, while dots indicate that the algorithm incorrectly identified chaotic points as periodic.
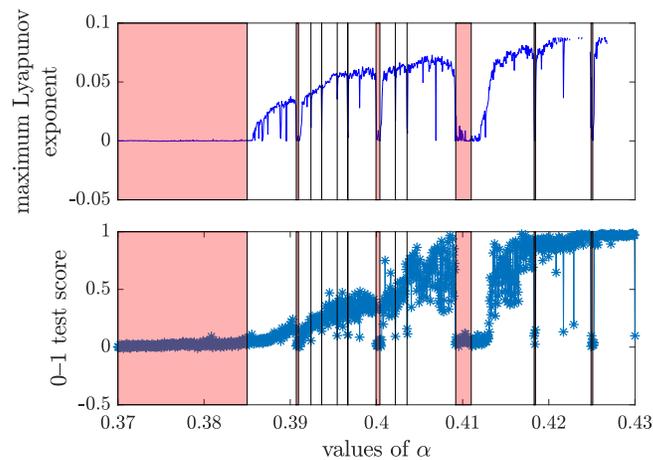

Figure 15: **Top:** The maximum Lyapunov exponent for the Rossler system as a function of the bifurcation parameter α. **Bottom:** The score of the 0-1 test for chaos where 0 indicates periodicity, while 1 indicates chaos. The shaded windows denote the regions that were tagged as periodic.

This method of featurization allows for a great deal of flexibility for the end user. In particular, we have provided two options for template functions, tent functions and interpolating polynomials, but surely there are many other collections of functions which could be tested for optimizing classification and regression tasks. We showed these two functions worked quite well on standard experiments, as well as in comparison to other methods available in the literature.

We find the particular results of the SHREC data set (Section 8.4) to be quite fascinating due to the vast improvement seen from tent functions to interpolating polynomials. The usual knee-jerk reaction to setting up these featurization methods for persistence diagrams is that localization is key. This was the impetus for creation of the tent functions as they have support contained in a small box, so each tent function truly only sees a small window of the diagram. Meanwhile, the interpolating polynomials are nonzero well away from their chosen "basepoint" so the fact that these functions work at all is surprising to say the least.

We hope to understand this behavior further in the course of future work.

Acknowledgments

Appendix A. Implementation of the interpolating polynomials algorithm

In this appendix, we give more details on the implementation of the interpolating polynomials described in Section 6.2. The barycentric formula for Lagrange interpolation described by Berrut and Trefethen (2004) is given by

$$f(x) := \sum_{j=0}^{m} \ell_j^{\mathcal{A}}(x) c_j = \frac{\sum_{j=0}^{m} \frac{w_j}{x - \tilde{x}_j} c_j}{\sum_{j=0}^{k} \frac{w_j}{x - \tilde{x}_j}}; \text{ where } w_j = \frac{1}{\ell'(a_j)}; \quad \ell'(a_j) = \prod_{i=0, i \neq j}^{m} (a_j - a_i), \quad (15)$$

while $\mathcal{A} = \{a_i\}_{i=0}^{m} \subset \mathbb{R}$ is a finite set of distinct mesh values, and $\{c_i \in \mathbb{R}\}$ is a collection of evaluation values. The function in Eq. (15) has the property that $f(a_i) = c_i$ for all i, and it also satisfies the partition of unity condition $\sum_{j=0}^{m} f(x) = 1, \forall x$.

Barycentric Lagrange interpolation is often used for approximating $\mathbb{R}$-valued functions and there are efficient algorithms for obtaining the weights associated with it. However, in our formulation we need to an interpolating polynomial over an $\mathbb{R}^2$-valued function. Therefore, we next describe how to expand the algorithm for interpolating a scalar valued function to interpolating a function on the plane. Note that the notation used here is self-contained from Section 6.2.

We assume that our planar mesh is the outer product of $m + 1$ mesh points along the birth time x-axis, and $n + 1$ points along the lifetime y-axis. We also assume that the persistence diagram has N pairs of (birth, lifetime) points.

1. Get $\tilde{\gamma}$ and ϕ which correspond to the interpolation matrices along the x-mesh and the y-mesh, respectively. These are the matrices that describe the linear transformation from the $m+1$ mesh points of birth times ($n+1$ mesh of lifetimes) to the corresponding interpolated values of the N query birth times (N query lifetimes) for a given diagram. This step is equivalent to separately obtaining the interpolation matrices for the birth times and the lifetimes.

2. Set $\gamma = \tilde{\gamma}^T$.

3. (a) Replicate each column in γ $n + 1$ times to obtain Γ whose dimensions are $(m + 1) \times (N \times (n + 1))$.

   (b) Unravel ϕ row-wise into a row vector, then replicate each row $m + 1$ times to obtain Φ whose dimensions are $(m + 1) \times (N \times (n + 1))$.

4. Use element-wise multiplication to obtain $\tilde{\Psi} = \Gamma \cdot \Phi$, where $\cdot$ means element-wise multiplication, and $\tilde{\Psi}$ has dimension $(m + 1) \times (N \times (n + 1))$.

5. (a) Split $\tilde{\Psi}$ into N chunks of $(m + 1) \times (n + 1)$ matrices along the columns axis.

   (b) Concatenate the split pieces row-wise to obtain an $(N \times (m+1)) \times (n+1)$ matrix Ψ.

6. Reshape Ψ by concatenating each $(m + 1) \times (n + 1)$ piece row-wise to obtain an $N \times ((m + 1) \times (n + 1))$ matrix Ξ.

7. Let the 2D base mesh be given as

$$\begin{bmatrix} f_{00} & f_{01} & \cdots & f_{0n} \\ f_{10} & f_{11} & \cdots & f_{1n} \\ \vdots & & & \vdots \\ f_{m0} & f_{m1} & \cdots & f_{mn} \end{bmatrix},$$

where $f_{ij} = f(x_i, y_j)$ and (x_i, y_j) is a unique point in the 2D mesh. Define the vector $[f_{00}\ f_{01}\ \ldots\ f_{mn}]$ which is obtained by unraveling the 2D mesh row-wise.

8. We can interpolate the query points (x_q, y_q) using

$$p(x_q, y_q) = \begin{bmatrix} \ell_0(x_0)\ell_0(y_0) & \cdots & \ell_m(x_0)\ell_n(y_0) \\ \ell_0(x_1)\ell_0(y_1) & \cdots & \ell_m(x_1)\ell_n(y_1) \\ \vdots & & \vdots \\ \ell_0(x_{N-1})\ell_0(y_{N-1}) & \cdots & \ell_m(x_{N-1})\ell_n(y_{N-1}) \end{bmatrix} \begin{bmatrix} f_{00} \\ f_{01} \\ \vdots \\ f_{mn} \end{bmatrix}.$$

Here is a sketch of the resulting matrices:

$$\tilde{\gamma} = \begin{bmatrix} \ell_0(x_0) & \ell_1(x_0) & \dots & \ell_m(x_0) \\ \vdots & & & \vdots \\ \ell_0(x_{N-1}) & \ell_1(x_{N-1}) & \dots & \ell_m(x_{N-1}) \end{bmatrix}_{N \times (m+1)},$$

$$\phi = \begin{bmatrix} \ell_0(y_0) & \ell_1(y_0) & \dots & \ell_n(y_0) \\ \vdots & & & \vdots \\ \ell_0(y_{N-1}) & \ell_1(y_{N-1}) & \dots & \ell_n(y_{N-1}) \end{bmatrix}_{N \times (n+1)},$$

$$\gamma = \tilde{\gamma}^T = \begin{bmatrix} \ell_0(x_0) & \ell_0(x_1) & \dots & \ell_0(x_{N-1}) \\ \ell_1(x_0) & \ell_1(x_1) & \dots & \ell_1(x_{N-1}) \\ \vdots & & & \vdots \\ \ell_m(x_0) & \ell_m(x_1) & \dots & \ell_m(x_{N-1}) \end{bmatrix}_{(m+1) \times N},$$

$$\Gamma = \begin{bmatrix} \ell_0(x_0) & \ell_0(x_0) & \dots & \ell_0(x_0) & \dots & \ell_0(x_{N-1}) & \ell_0(x_{N-1}) & \dots & \ell_0(x_{N-1}) \\ \ell_1(x_0) & \ell_1(x_0) & \dots & \ell_1(x_0) & \dots & \ell_1(x_{N-1}) & \ell_1(x_{N-1}) & \dots & \ell_1(x_{N-1}) \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ \ell_m(x_0) & \ell_m(x_0) & \dots & \ell_m(x_0) & \dots & \ell_m(x_{N-1}) & \ell_m(x_{N-1}) & \dots & \ell_m(x_{N-1}) \end{bmatrix}$$

where Γ has dimension $(m+1) \times (N \times (n+1))$.

$$\Phi = \begin{bmatrix} \ell_0(y_0) & \ell_1(y_0) & \dots & \ell_n(y_0) & \dots & \ell_0(y_{N-1}) & \ell_1(y_{N-1}) & \dots & \ell_n(y_{N-1}) \\ \ell_0(y_0) & \ell_1(y_0) & \dots & \ell_n(y_0) & \dots & \ell_0(y_{N-1}) & \ell_1(y_{N-1}) & \dots & \ell_n(y_{N-1}) \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ \ell_0(y_0) & \ell_1(y_0) & \dots & \ell_n(y_0) & \dots & \ell_0(y_{N-1}) & \ell_1(y_{N-1}) & \dots & \ell_n(y_{N-1}) \end{bmatrix}$$

where Φ has dimension $(m+1) \times (N \times (n+1))$.

We can now compute the elementwise product $\Psi = \Gamma \cdot \Phi$, which has the dimension $(m+1) \times (N \times (n+1))$.

We then need to apply the following operations: (i) reshaping Ψ to obtain $\hat{\Psi}_1$ given by

$$\hat{\Psi}_1 = \begin{bmatrix} \ell_0(x_0)\ell_0(y_0) & \ell_0(x_0)\ell_1(y_0) & \dots & \ell_0(x_0)\ell_n(y_0) \\ \ell_1(x_0)\ell_0(y_0) & \ell_1(x_0)\ell_1(y_0) & \dots & \ell_1(x_0)\ell_n(y_0) \\ \vdots & & & \vdots \\ \ell_m(x_0)\ell_0(y_0) & \ell_m(x_0)\ell_1(y_0) & \dots & \ell_m(x_0)\ell_n(y_0) \\ \vdots & & & \vdots \\ \ell_0(x_{N-1})\ell_0(y_{N-1}) & \ell_0(x_{N-1})\ell_1(y_{N-1}) & \dots & \ell_0(x_{N-1})\ell_n(y_{N-1}) \\ \ell_1(x_{N-1})\ell_0(y_{N-1}) & \ell_1(x_{N-1})\ell_1(y_{N-1}) & \dots & \ell_1(x_{N-1})\ell_n(y_{N-1}) \\ \vdots & & & \vdots \\ \ell_m(x_{N-1})\ell_0(y_{N-1}) & \ell_m(x_{N-1})\ell_1(y_{N-1}) & \dots & \ell_m(x_{N-1})\ell_n(y_{N-1}) \end{bmatrix}.$$

(ii) unraveling $\hat{\Psi}_1$ into an $N \times ((m+1) \times (n+1))$ matrix $\hat{\Psi}_2$ given by

$$\hat{\Psi}_2 = \begin{bmatrix} \ell_0(x_0)\ell_0(y_0) & \dots & \ell_0(x_0)\ell_n(y_0) & \dots & \ell_m(x_0)\ell_n(y_0) \\ \vdots & & \vdots & & \\ \ell_0(x_k)\ell_0(y_k) & \dots & \ell_0(x_k)\ell_n(y_k) & \dots & \ell_m(x_k)\ell_n(y_k) \\ \vdots & & \vdots & & \\ \ell_0(x_{N-1})\ell_0(y_{N-1}) & \dots & \ell_0(x_{N-1})\ell_n(y_{N-1}) & \dots & \ell_m(x_{N-1})\ell_n(y_{N-1}) \end{bmatrix}. \qquad (16)$$

The collection of all the scores constitutes the feature vector corresponding to the chosen base mesh point and to the query points where the latter are the persistence diagram points. In this study we summed the rows of $\hat{\Psi}_2$ after taking the absolute value of each entry. The resulting number represents the score at each base mesh point. If the persistence diagram contains the mesh points and we want to find the interpolated values at query points p_{interp}, then we would compute $p_{\text{interp.}} = \hat{\Psi}_2 f$.

The implementation of this algorithm can be found in the teaspoon package at `teaspoon.ML.feature_functions.interp_polynomial`.

Appendix B. Additional shape data results

This appendix gives additional results for the SHREC data set described in Section 8.4 using tent functions instead of interpolating polynomials. Table 4 should be compared to the results of Table 3.

freq	MSK	Dim 0		Dim 1		Dim 0 & Dim 1	
		Train	Test	Train	Test	Train	Test
1	$94.7\% \pm 5.1$	$8.3\% \pm 0.5$	$3.4\% \pm 1.1$	$8.1\% \pm 0.2$	$3.7\% \pm 0.5$	$8.2\% \pm 0.3$	$3.5\% \pm 0.5$
2	$99.3\% \pm 0.9$	$8.3\% \pm 0.3$	$3.4\% \pm 0.7$	$8.2\% \pm 0.5$	$3.5\% \pm 1.1$	$8.56\% \pm 0.4$	$3.0\% \pm 1.0$
3	$96.3\% \pm 2.2$	$66.5\% \pm 2.7$	$31.8\% \pm 4.8$	$50.6\% \pm 2.1$	$31.1\% \pm 4.0$	$80.5\% \pm 1.3$	$44.4\% \pm 4.3$
4	$97.3\% \pm 1.9$	$46.2\% \pm 2.5$	$27.0\% \pm 3.8$	$83.1\% \pm 1.6$	$63.5\% \pm 4.6$	$89.1\% \pm 1.5$	$69.0\% \pm 4.9$
5	$96.3\% \pm 2.5$	$28.5\% \pm 1.4$	$18.9\% \pm 4.0$	$75.2\% \pm 2.6$	$58.3\% \pm 4.6$	$76.8\% \pm 2.7$	$58.4\% \pm 7.9$
6	$93.7\% \pm 3.2$	$25.4\% \pm 1.8$	$19.0\% \pm 2.4$	$96.5\% \pm 1.1$	$88.7\% \pm 2.4$	$96.8\% \pm 0.67$	$89.9\% \pm 1.7$
7	$88.0\% \pm 4.5$	$19.4\% \pm 2.6$	$10.0\% \pm 3.4$	$98.2\% \pm 0.5$	$93.6\% \pm 1.9$	$98.3\% \pm 0.6$	$94.1\% \pm 2.5$
8	$88.3\% \pm 6.0$	$10.8\% \pm 2.6$	$3.6\% \pm 2.4$	$91.9\% \pm 0.9$	$88.8\% \pm 2.7$	$91.9\% \pm 1.2$	$89.7\% \pm 3.3$
9	$88.0\% \pm 5.8$	$10.6\% \pm 2.7$	$4.3\% \pm 2.2$	$63.8\% \pm 2.7$	$53.3\% \pm 5.9$	$64.9\% \pm 2.3$	$53.7\% \pm 3.8$
10	$91.0\% \pm 4.0$	$9.2\% \pm 2.3$	$3.6\% \pm 1.7$	$27.0\% \pm 3.9$	$16.2\% \pm 3.2$	$27.3\% \pm 3.4$	$18.6\% \pm 5.6$

Table 4: Results of classification of shape data discussed in Section 8.4. The functions used are the tent functions with $d = 20$, and a ridge regression classifier. The MSK column gives the original results from Reininghaus et al. (2015); the subsequent columns use the 0-dimensional diagrams only, the 1-dimensional diagrams only, and both, respectively. Scores highlighted in blue give best average score MSK vs. template functions; scores highlighted in orange have overlapping intervals with the best score.

References

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. *Journal of Machine Learning Research*, 18(8):1–35, 2017.

Aaron Adcock, Erik Carlsson, and Gunnar Carlsson. The ring of algebraic functions on persistence bar codes. *Homology, Homotopy and Applications*, 18(1):381–402, 2016. doi: 10.4310/HHA.2016.v18.n1.a21.

Rushil Anirudh, Vinay Venkataraman, Karthikeyan Natesan Ramamurthy, and Pavan Turaga. A Riemannian framework for statistical analysis of topological persistence diagrams. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2016.

René Baire. Sur les fonctions de variables réelles. *Annali di Matematica Pura ed Applicata (1898-1922)*, 3(1):1–123, 1899.

Paul Bendich, J. S. Marron, Ezra Miller, Alex Pieloch, and Sean Skwerer. Persistent homology analysis of brain artery trees. *The Annals of Applied Statistics*, 10(1):198–218, mar 2016. doi: 10.1214/15-aoas886.

Giancarlo Benettin, Luigi Galgani, Antonio Giorgilli, and Jean-Marie Strelcyn. Lyapunov characteristic exponents for smooth dynamical systems and for hamiltonian systems; a method for computing all of them. part 2: Numerical application. *Meccanica*, 15(1): 21–30, Mar 1980. ISSN 1572-9648. doi: 10.1007/BF02128237.

J. Berrut and L. N. Trefethen. Barycentric Lagrange interpolation. *SIAM Review*, 46(3): 501–517, 2004.

Eric Berry, Yen-Chi Chen, Jessi Cisewski-Kehe, and Brittany Terese Fasy. Functional summaries of persistence diagrams. *arXiv:1804.01618*, 2018.

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. *Foundations of Computational Mathematics*, 14(4):745–789, 2014.

Peter Bubenik. Statistical topological data analysis using persistence landscapes. *Journal of Machine Learning Research*, 16:77–102, 2015.

Peter Bubenik and Tane Vergili. Topological spaces of persistence modules and their properties. *Journal of Applied and Computational Topology*, dec 2018. doi: 10.1007/ s41468-018-0022-4.

Gunnar Carlsson and Sara Kalisnik Verovsek. Symmetric and r-symmetric tropical polynomials and rational functions. *Journal of Pure and Applied Algebra*, pages –, 2016. ISSN 0022-4049. doi: 10.1016/j.jpaa.2016.05.002.

Mathieu Carrière and Ulrich Bauer. On the metric distortion of embedding persistence diagrams into reproducing kernel hilbert spaces. *arXiv:1806.06924*, 2018.

Mathieu Carrière, Steve Y. Oudot, and Maks Ovsjanikov. Stable topological signatures for points on 3d shapes. *Computer Graphics Forum*, 34(5):1–12, aug 2015. doi: 10.1111/cgf. 12692.

Mathieu Carrire, Marco Cuturi, and Steve Oudot. Sliced wasserstein kernel for persistence diagrams. *arXiv:1706.03358*, 2017.

Frédéric Chazal, David Cohen-Steiner, Marc Glisse, Leonidas J. Guibas, and Steve Y. Oudot. Proximity of persistence modules and their diagrams. In *Proceedings of the 25th annual symposium on Computational geometry - SCG '09*. ACM Press, 2009. doi: 10.1145/1542362.1542407.

Frédéric Chazal, Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, and Larry Wasserman. Stochastic convergence of persistence landscapes and silhouettes. In *Proceedings of the Thirtieth Annual Symposium on Computational Geometry*, SOCG'14, pages 474:474–474:483, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2594-3. doi: 10.1145/2582112.2582128. URL http://doi.acm.org/10.1145/2582112.2582128.

Frédéric Chazal, Vin de Silva, Marc Glisse, and Steve Oudot. *The Structure and Stability of Persistence Modules*. Springer International Publishing, 2016. doi: 10.1007/ 978-3-319-42545-0.

Yen-Chi Chen, Daren Wang, Alessandro Rinaldo, and Larry Wasserman. Statistical analysis of persistence intensity functions. *arXiv:1510.02502*, 2015.

Ilya Chevyrev, Vidit Nanda, and Harald Oberhauser. Persistence paths and signature features in topological data analysis. *arXiv:1806.00381*, 2018.

Moo K. Chung, Peter Bubenik, and Peter T. Kim. Persistence diagrams of cortical surface data. In Jerry L. Prince, Dzung L. Pham, and Kyle J. Myers, editors, *Information Processing in Medical Imaging*, volume 5636 of *Lecture Notes in Computer Science*, pages 386–397. Springer Berlin Heidelberg, 2009. ISBN 978-3-642-02497-9. doi: 10.1007/978-3-642-02498-6_32.

David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete Comput. Geom.*, 37(1):103–120, January 2007. ISSN 0179-5376. doi: 10.1007/s00454-006-1276-5. URL http://dx.doi.org/10.1007/s00454-006-1276-5.

John B Conway. *A course in functional analysis*, volume 96. Springer Science & Business Media, 2013.

Ren Corbet, Ulderico Fugacci, Michael Kerber, Claudia Landi, and Bei Wang. A kernel for multi-parameter persistent homology. *arXiv:1809.10231*, 2018.

Barbara Di Fabio and Massimo Ferri. Comparing persistence diagrams through complex vectors. In *Image Analysis and Processing — ICIAP 2015*, pages 294–305. Springer International Publishing, 2015. doi: 10.1007/978-3-319-23231-7_27.

Persi Diaconis, Susan Holmes, and Mehrdad Shahshahani. Sampling from a manifold. In *Advances in Modern Statistical Theory and Applications: A Festschrift in honor of Morris L. Eaton*, pages 102–125. Institute of Mathematical Statistics, 2013. doi: 10.1214/12-imscoll1006.

Pietro Donatini, Patrizio Frosini, and Alberto Lovato. Size functions for signature recognition. In Robert A. Melter, Angela Y. Wu, and Longin J. Latecki, editors, *Vision Geometry VII*. SPIE, oct 1998. doi: 10.1117/12.323253.

J. P. Eckmann and D. Ruelle. Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.*, 57:617–656, Jul 1985. doi: 10.1103/RevModPhys.57.617.

Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, Larry Wasserman, Sivaraman Balakrishnan, and Aarti Singh. Confidence sets for persistence diagrams. *Annals of Statistics*, 42(6):2301–2339, 12 2014. doi: 10.1214/14-AOS1252.

Massimo Ferri, Patrizio Frosini, Alberto Lovato, and Chiara Zambelli. Point selection: A new comparison scheme for size functions (with an application to monogram recognition). In *ACCV*, 1998.

Georg A. Gottwald and Ian Melbourne. A new test for chaos in deterministic systems. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 460(2042):603–611, 2004. ISSN 1364-5021. doi: 10.1098/rspa.2003.1183.

Georg A Gottwald and Ian Melbourne. On the validity of the 01 test for chaos. *Nonlinearity*, 22(6):1367, 2009.

Georg A. Gottwald and Ian Melbourne. *The 0-1 Test for Chaos: A Review*, pages 221–247. Springer Berlin Heidelberg, Berlin, Heidelberg, 2016. ISBN 978-3-662-48410-4. doi: 10.1007/978-3-662-48410-4_7.

M. Henon. On the numerical computation of Poincaré maps. *Physica D: Nonlinear Phenomena*, 5(2):412 – 414, 1982. ISSN 0167-2789. doi: 10.1016/0167-2789(82)90034-3.

Sara Kališnik. Tropical coordinates on the space of persistence barcodes. *Foundations of Computational Mathematics*, Jan 2018. ISSN 1615-3383. doi: 10.1007/s10208-018-9379-y.

Genki Kusano. Persistence weighted gaussian kernel for probability distributions on the space of persistence diagrams. 2018.

Genki Kusano, Yasuaki Hiraoka, and Kenji Fukumizu. Persistence weighted gaussian kernel for topological data analysis. In *International Conference on Machine Learning*, pages 2004–2013, 2016.

Genki Kusano, Kenji Fukumizu, and Yasuaki Hiraoka. Kernel method for persistence diagrams via kernel embedding and weight factor. 2017.

Roland Kwitt, Stefan Huber, Marc Niethammer, Weili Lin, and Ulrich Bauer. Statistical topological data analysis - a kernel perspective. In C. Cortes, N.D. Lawrence, D.D. Lee, M. Sugiyama, R. Garnett, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 3052–3060. Curran Associates, Inc., 2015.

Tam Le and Makoto Yamada. Persistence Fisher kernel: A Riemannian manifold kernel for persistence diagrams. In *32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.*, 2018.

Chunyuan Li, Maks Ovsjanikov, and Frederic Chazal. Persistence-based structural recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1995–2002, 2014.

Michael McCullough, Michael Small, Thomas Stemler, and Herbert Ho-Ching Iu. Time lagged ordinal partition networks for capturing dynamics of continuous dynamical systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25(5):053101, 2015. doi: 10.1063/1.4919075.

Yuriy Mileyko, Sayan Mukherjee, and John Harer. Probability measures on the space of persistence diagrams. *Inverse Problems*, 27(12):124007, 2011.

Elizabeth Munch, Katharine Turner, Paul Bendich, Sayan Mukherjee, Jonathan Mattingly, and John Harer. Probabilistic fréchet means for time varying persistence diagrams. *Electron. J. Statist.*, 9:1173–1204, 2015. doi: 10.1214/15-EJS1030.

James R. Munkres. *Topology.* Prentice Hall, 2000.

Steve Y. Oudot. *Persistence Theory: From Quiver Representations to Data Analysis (Mathematical Surveys and Monographs).* American Mathematical Society, 2017. ISBN 978-1-4704-3443-4.

Deepti Pachauri, Chris Hinrichs, Moo K. Chung, Sterling C. Johnson, and Vikas Singh. Topology-based kernels with application to inference problems in alzheimer's disease. *IEEE Transactions on Medical Imaging*, 30(10):1760–1770, oct 2011. doi: 10.1109/tmi. 2011.2147327.

Tullia Padellini and Pierpaolo Brutti. Persistence flamelets: multiscale persistent homology for kernel density exploration. *arxiv:1709.07097*, 2017.

P. Palaniyandi. On computing Poincaré map by Hénon method. *Chaos, Solitons & Fractals*, 39(4):1877 – 1882, 2009. ISSN 0960-0779. doi: 10.1016/j.chaos.2007.06.118.

David Pickup, X Sun, Paul L Rosin, RR Martin, Z Cheng, Z Lian, M Aono, A Ben Hamza, A Bronstein, M Bronstein, et al. SHREC14 track: Shape retrieval of non-rigid 3D human models. In *Proceedings of the 7th Eurographics workshop on 3D Object Retrieval*, volume 1, page 6. Eurographics Association, 2014.

Jan Reininghaus, Stefan Huber, Ulrich Bauer, and Roland Kwitt. A stable multi-scale kernel for topological machine learning. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

David Rouse, Adam Watkins, David Porter, John Harer, Paul Bendich, Nate Strawn, Elizabeth Munch, Jonathan DeSena, Jesse Clarke, Jeffrey Gilbert, Peter Chin, and Andrew

Newman. Feature-aided multiple hypothesis tracking using topological and statistical behavior classifiers. In Ivan Kadar, editor, *Signal Processing, Sensor/Information Fusion, and Target Recognition XXIV*. SPIE, may 2015. doi: 10.1117/12.2179555.

Walter Rudin. *Real and complex analysis*. Tata McGraw-Hill Education, 2006.

Marco Sandri. Numerical calculation of lyapunov exponents. *The Mathematica Journal*, 6 (3):78–84, 1986.

Nikhil Singh, Heather D. Couture, J. S. Marron, Charles Perou, and Marc Niethammer. *Topological Descriptors of Histology Images*, pages 231–239. Springer International Publishing, Cham, 2014. ISBN 978-3-319-10581-9. doi: 10.1007/978-3-319-10581-9_29.

Jian Sun, Maks Ovsjanikov, and Leonidas Guibas. A concise and provably informative multi-scale signature based on heat diffusion. In *Proceedings of the Symposium on Geometry Processing*, SGP '09, pages 1383–1392, Aire-la-Ville, Switzerland, Switzerland, 2009. Eurographics Association. URL `http://dl.acm.org/citation.cfm?id=1735603.1735621`.

Lloyd N. Trefethen. *Approximation Theory and Approximation Practice (Applied Mathematics)*. Society for Industrial and Applied Mathematics, 2012. ISBN 9781611972399.

Katharine Turner, Yuriy Mileyko, Sayan Mukherjee, and John Harer. Fréchet means for distributions of persistence diagrams. *Discrete & Computational Geometry*, 52(1):44–70, 2014. ISSN 0179-5376. doi: 10.1007/s00454-014-9604-7. URL `http://dx.doi.org/10.1007/s00454-014-9604-7`.

Xiaojin Zhu, Ara Vartanian, Manish Bansal, Duy Nguyen, and Luke Brandl. Stochastic multiresolution persistent homology kernel. In *IJCAI*, pages 2449–2457, 2016.

Bartosz Zieliński, Mateusz Juda, and Matthias Zeppelzauer. Persistence codebooks for topological data analysis. *arXiv:1802.04852*, 2018a.

Bartosz Zieliński, Micha Lipiński, Mateusz Juda, Matthias Zeppelzauer, and Paweł Dłotko. Persistence bag-of-words for topological data analysis. *arXiv:1812.09245*, 2018b.