PRITCHARD, NEIL, Ph.D. Injective and coarse embeddings of persistence diagrams and Wasserstein space. (2023)
Directed by Dr. Thomas Weighill. 39 pp.

In this dissertation we will examine questions related to two fields of mathematics, topological data analysis (TDA) and optimal transport (OT). Both of these fields center on complex data types to which one often needs to apply standard machine learning or statistical methods. Such application will typically mandate that these data types are embedded into a vector space. It has been shown that for many natural metrics such embeddings necessarily have high distortion, i.e. are not even coarse embeddings. Whether coarse embeddings exist with respect to the $p$-Wasserstein distance for $1 \leq p \leq 2$ remains an open question, however, both for persistence diagrams (from TDA) and planar distributions (from OT). In this first part of this dissertation, we use coarse geometric techniques to show that the TDA and OT sides of this open question are equivalent for $p > 1$. In the second, we study an embedding of persistence diagrams, and show that under mild conditions it is injective, i.e. distinguishes between distinct diagrams.

# INJECTIVE AND COARSE EMBEDDINGS OF PERSISTENCE DIAGRAMS AND WASSERSTEIN SPACE

by

Neil Pritchard

A Dissertation Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Greensboro
2023

Approved by

Committee Chair Thomas Weighill

APPROVAL PAGE

This dissertation written by Neil Pritchard has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair _____
Thomas Weighill

Committee Members _____
Greg Bell

_____
Talia Fernós

_____
Michael Hull

_____
Alexander Wagner

_____
Date of Acceptance by Committee

_____
Date of Final Oral Examination

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Thomas Weighill and my committee members, Dr. Greg Bell, Dr. Talia Fernós, Dr. Michael Hull, and Dr. Alexander Wagner for their assistance and guidance towards the completion of my dissertation.

# Table of Contents

# List of Figures

# Chapter 1: Introduction

## 1.1 Motivation

In this dissertation we will be primarily focused on spaces arising in two fields of mathematics, namely topological data analysis (TDA) and optimal transport (OT). Loosely speaking, TDA seeks to answers questions about complex data sets by looking at their "shape". OT on the other hand has its roots in answering questions about optimal resource allocation. In both fields one often seeks to make use of standard machine learning and statistical techniques. In order to do this it is first necessary to transform the objects that arise in these spaces into real-valued vectors. For such a process to be effective one hopes that the mappings preserve some of the original structure of the spaces.

The structure of the dissertation is as follows. In this chapter we will provide some of the basic structure necessary to understand the scope of TDA, OT, and coarse geometry. For a more in-depth background on these fields the interested reader can appeal to, [15] for TDA, [21, 26] for OT, and [22] for coarse geometry. In Chapter 2 we connect the questions of coarse embeddability for Wasserstein space and the space of persistence diagrams by showing that if the space of persistence diagrams with the $p$-Wasserstein metric embeds into Hilbert space, for $1 \leq p < \infty$, then so does Wasserstein space on $\mathbb{R}^2$ with the $p$-Wasserstein metric; the converse holds if $p > 1$ (Theorem 2.4.8). In Chapter 3 we show that unweighted Gaussian persistence curves, a functional summary of persistence diagrams, are injective under mild restrictions. We note that Gaussian persistence curves may be viewed as a map into $L^2(\mathbb{R})$, notably a Hilbert space.

## 1.2 Topological Data Analysis

Topological data analysis is a field of mathematics which seeks to apply the power of topology to certain data spaces. There are a number of techniques and spaces one might be interested in. Of particular focus are **point clouds**, which are finite subsets of a Euclidean space, and **gray-scale images**, which are functions $I\colon [m] \times [n] \to [256]$, where $([n] = \{0, 1, \cdots, n-1\})$. In order to use topological techniques on these data spaces some work needs to be done to transform these spaces into a "nice" sequence of topological spaces. We

proceed by introducing some necessary definitions for this procedure.

**Definition 1.2.1.** Suppose that the $k+1$ points $\{u_0, u_1, \cdots, u_k\}$ are affinely independent, i.e. $u_1 - u_0, u_2 - u_0, \cdots, u_k - u_0$ are linearly independent. The $k$**-simplex (simplex)** determined by these points is the set of points

$$S = \left\{ \alpha_0 u_0 + \cdots + \alpha_k u_k \mid \sum_{i=0}^{k} \alpha_i = 1 \text{ and } \alpha_i \geq 0 \text{ for } i = 0, \cdots, k \right\}$$

If $s$ is a nonempty subset of the points $\{u_0, u_1, \cdots, u_k\}$, then the convex hull of $s$ is called a **face** of the simplex, S.

Note that a $0$-simplex is simply a point, a $1$-simplex is a line segment, and a $2$-simplex is a triangle. In general one commonly thinks of simplices as $k$-dimensional generalizations of triangles.

**Definition 1.2.2.** A **simplicial complex** $\mathcal{K}$ is a set of simplices which satisfies the following criterion:

1. Every face of a simplex in $\mathcal{K}$ is also in $\mathcal{K}$
2. The nonempty intersection of any two simplices $\sigma_1, \sigma_2 \in \mathcal{K}$ is a face of both $\sigma_1$ and $\sigma_2$.

One manner in which point clouds can viewed as topological spaces is by constructing a sequence of simplicial complexes from the point cloud. There are two common ways to do this, both require an additional parameter $\epsilon > 0$ to be chosen. For a point cloud $X$ with an associated metric $d$, the **Čech complex**, denoted $C_\epsilon(X)$ is constructed in the following manner. Let the elements of the point cloud $X$ be the vertex set of $C_\epsilon(X)$. Further, say that for each subset $\sigma \subset X$ let $\sigma$ be a simplex of $C_\epsilon(X)$ if $\cap_{x \in \sigma} B_\epsilon(x) \neq \varnothing$. Here $B_\epsilon(X)$ denotes the closed ball of radius $\epsilon$ centered at $x$. Note that checking all such intersections can be quite computationally expensive. In light of this a relaxation of this idea called the **Vietoris-Rips** complex, denoted $V_\epsilon(X)$ is often used. Here, with the same notation and vertex set as above, $\sigma \in V_\epsilon(X)$ if $d(x_i, x_j) \leq \epsilon$ for all $x_i, x_j \in \sigma$.

In either case, one transforms the data set into a family of simplicial complexes parameterized by $\epsilon$. In addition, one should note that these constructions have the property that if $\epsilon < \epsilon'$ then $C_\epsilon(x) \subseteq C_{\epsilon'}(X)$ or $V_\epsilon(x) \subseteq V_{\epsilon'}(X)$, in other word the family is in fact a nested family of simplicial complexes.

**Definition 1.2.3.** A **filtered complex**, $\mathcal{F}$, is an indexed family $(F_i)_{i \in I}$ of subcomplexes satisfying $F_i \subseteq F_j$ whenever $i < j$. Here $I$ is some totally ordered indexing set.

So in fact both the Čech complex and the Vietoris-Rips complex create filtered simplicial complexes parameterized by $\epsilon > 0$. To obtain such a filtered sequence of complexes for gray-scale images, which preserves the structure in some manner, one must appeal to a different type of complex.

| | | | |
|---|---|---|---|
| 1 | 4 | 4 | 2 |
| 1 | 5 | 5 | 2 |
| 1 | 5 | 5 | 2 |
| 1 | 4 | 4 | 2 |

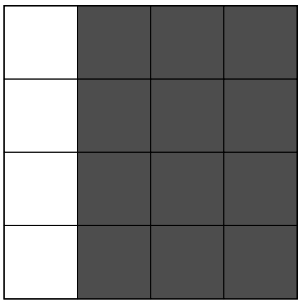


Figure 1.1. A $4 \times 4$ gray-scale image (values range from $0$ (black) to $255$ (white)) and its associated binary image from a filtration value of $t = 1$.

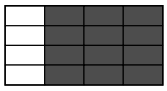
**Definition 1.2.4.** A set $X \subseteq \mathbb{R}^n$ is said to be a **cubical complex** if it can be written as a union of elementary cubes. Here an **elementary cube** is a finite product of intervals of the form $[l, l+1]$ or $[l, l]$ for some $l \in \mathbb{Z}$.

In order to construct the filtration for gray-scale images we must first consider **binary images**, which are defined as a functions $I\colon [m] \times [n] \to \{0, 1\}$. Note the only difference here is the codomain is restricted for binary images to that of only two possibilities. The output value of $1$ represents a white pixel and the output value of $0$ represents a black pixel. While gray-scale images have some additional complexity it is quite natural to consider a binary image as a cubical complex on the white pixels. Namely, the cubical complex associated to a binary image is the union of elementary cubes and their boundary elements for each white pixel. To proceed, for a gray-scale image $I$, let $F'_t$ denote the binary image obtained by setting $I(x, y) = 1$ if $I(x, y) \leq t$ and $I(x, y) = 0$ if $I(x, y) > t$. Then the family $(F_t)_{t\in[256]}$ of cubical complexes associated to $F'_t$ is a filtered cubical complex. This is in fact a special case of a sublevel set filtration.

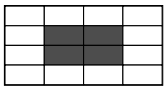
**Definition 1.2.5.** Let $f\colon X \to \mathbb{R}$ be a function then the family $(F_t)_{t\in\mathbb{R}}$, where $F_t = f^{-1}((-\infty, t])$ is a filtration called the **sublevel set filtration**.

Given a filtered complex $(F_t)_{t\in I}$ there are natural inclusion maps for each $i \leq j$. These maps induce homomorphisms on the homology groups $H_k(F_i)$ for each dimension $k$. The images of these maps are defined to be the $k$-th persistent homology groups and the ranks of these groups are defined to be the $k$-th persistent Betti numbers. Informally, one may think of the $k$-th persistent Betti number as tracking the number of $k$-dimensional holes (or in dimension $0$, the number of connected components) which persistent along the filtration from $F_i$ to $F_j$. For more details on the persistent homology see [15]

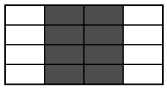
The appearance of a topological feature in a persistence homology group is referred to as the **birth** of that feature and the disappearance of a feature or the merging with an older feature is referred to as the **death** of that feature. It is beneficial to have a concise tool to represent the entirety of this information.



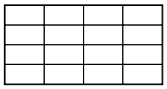
$\beta_0 = 1, \beta_1 = 0$, t=1



$\beta_0 = 1, \beta_1 = 1$, t=4



$\beta_0 = 2, \beta_1 = 0$, t=2



$\beta_0 = 1, \beta_1 = 0$, t=5

Figure 1.2. Betti numbers associated to filtered cubical complex constructed from the gray-scale image in Figure 1.1. Note the appearance of distinct connected component across the filtration as well as the birth and death of a hole.

**Definition 1.2.6.** Denote by $\mathbb{R}^2_<$ the set $\{(b,d) \in \mathbb{R}^2 \mid b < d\}$. A **persistence diagram** is a finite multiset $D \subseteq \mathbb{R}^2_<$.

It is of course obvious that a topological feature must be born before it can die and hence persistence diagrams are natural ways to store the information obtained by the above procedure. In reality, there are a number of different ways in which persistence diagrams are defined in the literature. However, the main variances are normally in the number of points that one allows (e.g finite, bounded, countable) or some restriction on the domain of the points. One may assume the above definition going forward unless otherwise stated.

**Definition 1.2.7.** A **partial matching** between two persistence diagrams $D_1$ and $D_2$ is a triple $(B_1, B_2, f)$ where $B_1 \subset D_1$, $B_2 \subset D_2$ and $f\colon B_1 \to B_2$ is a bijection.

**Definition 1.2.8.** For $1 \leq p < \infty$ The $p$**-cost** of a partial matching $(B_1, B_2, f)$ between diagrams $D_1$ and $D_2$ is defined as,

$$\text{cost}_p(f) = \left( \sum_{(b,d)\in B_1} d((b,d), f((b,d)))^p + \sum_{(b,d)\in D_1 \setminus B_1} d((b,d), \Delta_{\mathbb{R}})^p + \sum_{(b,d)\in D_2 \setminus B_2} d((b,d), \Delta_{\mathbb{R}})^p \right)^{\frac{1}{p}}$$

and for $p = \infty$ is defined as,

$$\text{cost}_\infty(f) = \max\{ \sup_{(b,d)\in B_1)} d((b,d), f((b,d))), \sup_{(b,d)\in D_1 \setminus B_1} d((b,d), \Delta_{\mathbb{R}}), \sup_{(b,d)\in D_2 \setminus B_2} d((b,d), \Delta_{\mathbb{R}})\}$$

where $\Delta_{\mathbb{R}} = \{(x,x) \in \mathbb{R}^2\}$ is the diagonal.

Figure 1.3. Two examples of partial matchings between two persistence diagrams (red,blue). Note how the unmatched points in both cases are visualized as being matched with the diagonal.

**Definition 1.2.9.** Let $1 \leq p \leq \infty$ and $D_1$, $D_2$ be persistence diagrams. Define,

$$W_p(D_1, D_2) = \inf_f \{\text{cost}_p(f)\}.$$

$W_p$ is a metric called the **Wasserstein metric** on persistence diagrams. Let $(\mathcal{D}, W_p)$ denote the space of persistence diagrams under this metric.

In applications, one often wishes to use the information stored in a persistence diagram by plugging it into some machine learning algorithm. In order to do this one must seek a map $f\colon \mathcal{D} \to \mathbb{R}^n$ for some $n$, or more generally to a normed vector space such as a (real) Hilbert space. Such a map is often called a vectorization. Some choices of vectorizations of persistence diagrams are persistence landscapes, persistence images, and persistence curves [1, 6, 12]. We will provide a brief description of the development of the persistence curve from [12]. Note that $[\cdot]$ denotes a multiset.

**Lemma 1.2.10** (Fundamental Lemma of Persistence Homology [15])**.** *Let $D$ be the $k$-dimensional diagram with respect to a filtration $\{F_i\}_i$. Then,*

$$\beta_k(F_t) = \#\left[(b,d) \in D \mid b \leq t < d\right] = \#(B_t \cap D) = \#D_t.$$

*Here $B_t = [(x,y) \mid x \leq t < y \in \mathbb{R}$ multiplicity $(x,y) = \infty]$ and $D_t = F_t \cap D$.*

So, the fundamental lemma states the the $k$-th persistent Betti number can be computed from the persistence diagram associated to the filtration by simply counting the number of points in the so called fundamental box, $B_t$, with multiplicity. In fact one can already see that this will yield a vectorization of the persistence diagram where the output at $t$ is the $k$-th persistent Betti number. This is often refered to as the Betti curve. Chung and Lawson generalize this as follows: Let $\mathcal{F}$ be the set of all function $\psi\colon \mathcal{D} \times \mathbb{R}^3 \to \mathbb{R}$ and let $\mathcal{T}$ represent the set of all summary statistics on multi sets, and $\mathcal{R}$ represent the set of functions on $\mathbb{R}$.

**Definition 1.2.11.** Define a map $P\colon \mathcal{D} \times \mathcal{F} \times \mathcal{T} \to \mathcal{R}$ by

$$P(D, \psi, T)(t) = T([\psi(D, b, d, t) \mid (b, d) \in D_t]), t \in \mathbb{R}$$

The function $P(D, \psi, T)$ is called the **persistence curve** of $D$ with respect to $\psi$ and $T$.

In other words the persistence curve framework provides a general class of vectorizations. One simply makes choices of summary statistic, applies this statistic to points in the fundamental box, and accumulates the information.

## 1.3 Optimal Transport

In this section we will give a brief overview of the field of optimal transport which, simply put, seeks to find the most efficient way to move one mass to another. We will follow the development and notation in [21] unless otherwise stated. We begin with the introduction of some simple definitions and notation. Let

$$\Sigma_n = \left\{ a \in \mathbb{R}^n \mid \sum_{i=1}^{n} a_i = 1 \text{ and } a_i \geq 0 \text{ for all } i = 1, \ldots, n \right\}.$$

$\Sigma_n$ is called the **probability simplex** (**standard simplex**) and an element $a \in \Sigma_n$ is referred to as a **probability vector**. Note that $\Sigma_n$ is a simplex (Definition 1.2.1) with the vertices in question being the standard unit vectors in $\mathbb{R}^n$.

**Definition 1.3.1.** A **discrete measure** with weights $a \in \mathbb{R}^n$ and locations $x_1, \ldots, x_n$ in a space $\mathcal{X}$ is denoted,

$$\alpha = \sum_{i=1}^{n} a_i \delta_{x_i}.$$

If $a$ is in fact an element of the probability simplex $\Sigma_n$ then the measure is called a **probability measure**.

We will begin with a brief example of a type of optimization problem called the assignment problem.

**Example 1.3.2** (Assignment Problem)**.** Let $A$ and $B$ be two sets such that $|A| = |B|$ and let $C\colon A \times B \to \mathbb{R}$ denote a function (typically thought of as representing a cost of movement). Then the assignment problem is an optimization problem of the form:

$$\min_{f\colon A \to B,\ f \text{ a bijection}} \frac{1}{n} \sum_{a \in A} C(a, f(a)).$$

One natural analogy for such problems is to consider the sets $A$ and $B$ as representing the locations of warehouses and stores respectively. With this in mind the cost function represents the cost of travel between an individual warehouse and a store and the problem seek a transportation plan which seeks to provide each store with a necessary good while minimizing cost. Of course one should note that the requirement that $f$ is a bijection here is already a strong requirement given this analogy. It would be quite reasonable to allow goods from multiple warehouse to go to a single store and yet this would not be allowed in the assignment problem.

With this in mind we consider a relaxation of this bijection condition as considered by Monge in 1781.

**Example 1.3.3** (Monge Assignment Problem Between Discrete Measures)**.** Let

$$\alpha = \sum_{i=1}^{n} a_i \delta_{x_i} \text{ and } \beta = \sum_{j=1}^{m} b_j \delta_{y_j}$$

be two discrete measure with weights $a \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ and locations $x_1, \ldots, x_n \in \mathcal{X}$ and $y_1, \ldots, y_m \in \mathcal{Y}$. Let $C \colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ denote a cost function. One wishes to consider maps $T \colon \{x_1, \ldots, x_n\} \to \{y_1, \ldots, y_m\}$ that satisfy:

$$\forall\, j \in [m], \quad b_j = \sum_{i \colon T(x_i) = y_j} a_i. \tag{1.1}$$

One might recognize that Equation (1.1) requires that the pushforward of the measure $\alpha$ is equal to $\beta$. In light of this the notation $T_{\#}\alpha = \beta$ is used to represent this criterion. Finally, Monge's assignment problem can be stated as:

$$\min_{T} \left\{ \sum_{i} C(x_i, T(x_i)) \mid T_{\#}\alpha = \beta \right\}.$$

Note that Monge's problem can in fact be seen as a generalization of the assignment problem. In the case when the measures are uniform probability measures of weight $\frac{1}{n}$ the constraint $T_{\#}\alpha = \beta$ implies that $T$ is a bijection. Thus Monge's assignment problem is the assignment problem with the same cost function.

As is often the case neither the assignment problem or Monge's assignment problem necessarily have unique solutions (see Figure 1.5).

Again we should note that even with the relaxation provided in the Monge problem there are still some issues in light of our analogue as a transport plan. In this case, while it is possible to send the goods from different warehouses to a single store, it is still impossible to send goods from one warehouse to distinct stores. Such a condition in regards to measures is often called "mass splitting".

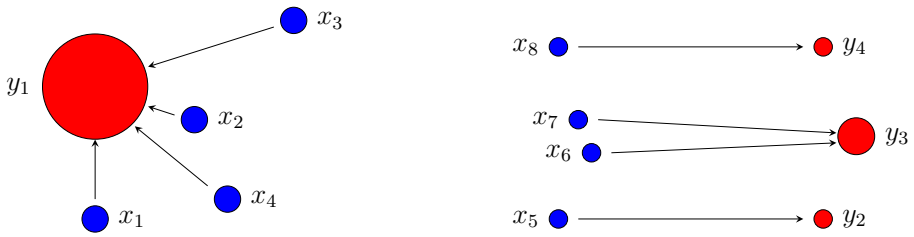



Figure 1.4. One possible matching for two discrete measures (blue and red) in the Monge assignment problem. Note that it is possible for the blue measure to have more supported locations than the red measure but not vice versa.

Figure 1.5. Two transport plans which are solutions to Monge's assignment problem. Here the measures (blue and red) are uniform measures of weight $\frac{1}{2}$ and the matching are denoted by the blue arrows.

**Example 1.3.4** (Kantorovich's relaxation)**.** Let

$$U(a,b) = \left\{P \in \mathbb{R}_+^{n\times m} \mid P\mathbb{1}_m = a \text{ and } P^T\mathbb{1}_n = b\right\}.$$

Then, for a given cost matrix $C$, Kantorovich's relaxation is the optimization problem:

$$L_C(a,b) = \min_{P\in U(a,b)} \sum_{i,j} C_{i,j}P_{i,j}.$$

Technically, this is not formulated as a problem between discrete measures. However, it is easy to make the translation. If $\alpha$ and $\beta$ are discrete measures with weights $a$ and $b$ respectively and $C$ denotes a cost matrix defined on their supports then one has Kantorovich's problem between discrete measures:

$$\mathcal{L}_C(\alpha,\beta) = L_C(a,b).$$

Note that so far we have restricted our attention to the discrete versions of these problems. There are of course analogues that can be stated for arbitrary measures. We will be particularly focused on restricting to the case of discrete measures in Chapter 2 though, so we have chosen this path for some consistency. Nevertheless, we will conclude this section with the necessary definition to introduce the formal statement of Wasserstein space.

**Definition 1.3.5.** Let $X$ be a set. A **$\sigma$-algebra** on $X$ is a collection $\mathcal{B}$ of subsets of $X$ that satisfy the following:

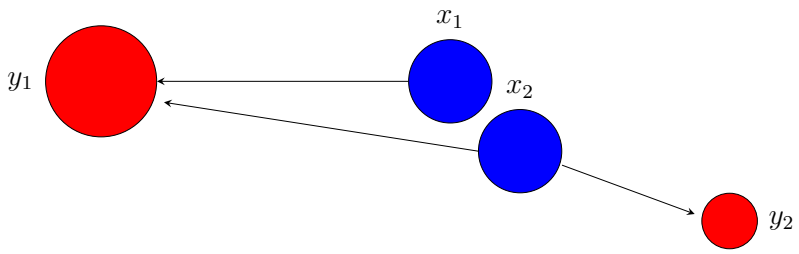



Figure 1.6. An example of an admissible coupling between the two measures (blue and red). Note how the mass located at $x_2$ goes to two distinct location in the red measures.

1. $\varnothing \in \mathcal{B}$.
2. If $E \in \mathcal{B}$, then $E^C \in \mathcal{B}$.
3. If $E_1, E_2, \cdots \in \mathcal{B}$, then $\cup_{n=1}^{\infty} E_n \in \mathcal{B}$.

**Definition 1.3.6.** Let $X$ be a metric space. The **Borel-$\sigma$-algebra** of $X$, denoted $\mathcal{B}[X]$, is defined to be the $\sigma$-algebra generated by the open subsets of $X$.

Note here that the $\sigma$-algebra generated by a family of sets can equivalently be thought of as the coarsest $\sigma$-algebra that contains the family or the intersection of all $\sigma$-algebras that contain the family.

**Definition 1.3.7.** Let $(X, \mathcal{B})$ be a measurable space. A **measure** $\mu$ on $\mathcal{B}$ is a map $\mu \colon \mathcal{B} \to [0, \infty]$ that satisfies,

1. $\mu(\varnothing) = 0$.
2. Whenever $E_1, E_2, \cdots \in \mathcal{B}$ are a countable sequence of disjoint measurable sets, then

$$\mu(\cup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} \mu(E_n).$$

A triplet $(X, \mathcal{B}, \mu)$, where $(X, \mathcal{B})$ is a measurable space and $\mu \colon \mathcal{B} \to [0, \infty]$ is a measure is known as a **measure space**. Moreover, if the total measure of the space is one, e.g. $\mu(X) = 1$, then the space is said to be a **probability space** and the measure a **probability measure**.

With this we can define Wasserstein space. We say that a probability distribution $\alpha$ has **finite $p$-th moment** if for every $x_0$

$$\int d(x_0, x)^p d\alpha(x) < \infty.$$

**Definition 1.3.8.** Let $(X, d_x)$ be a complete separable metric space and for $p \geq 1$ let $\mathcal{P}_p(x)$ denote the space of all Borel probability measures on $X$ with finite $p$-th moments. Further, let $\mathcal{U}(a, b)$ denote the set of Borel probability measures $\gamma$ on $X^2$ with marginals $\alpha$ and $\beta$ (i.e. the pushforwards of $\gamma$ along the projections are $\alpha$ and $\beta$ respectively). Then the Wasserstein $p$ distance between $\alpha, \beta \in \mathcal{P}_p(X)$ is

$$W_p(\alpha, \beta)^p = \inf_{\gamma \in \mathcal{U}(\alpha,\beta)} \int d(x, y)^p d\gamma(x, y)$$

The metric space $(\mathcal{P}_p(X), W_p)$ is called the **Wasserstein $p$-space over** $X$.

We call an element of $\mathcal{U}(\alpha, \beta)$ a **coupling from** $\alpha$ **to** $\beta$. The connection to see here is that the distance plays the role of the cost function and the coupling $\gamma$ plays the role of the transportation map.

## 1.4 Coarse Geometry

In coarse geometry one seeks to study the large scale structure of metric spaces. More specifically one seeks to study properties of spaces which are invariant under coarse equivalences. Historically this has developed out of work of Gromov in geometric group theory, [17], and Roe in index theory, [23]. We begin by recalling the notion of coarse equivalence before mentioning two coarse invariants.

**Definition 1.4.1.** Let $X$ and $Y$ be metric spaces and let $f\colon X \to Y$ be a map. $f$ is said to be

1. **bornologous** if for all $R > 0$ there exists an $S > 0$ such that $d(x, x') < R$ implies $d(f(x), f(x')) < S$.
2. **proper** if for each $x \in X$ and each $R > 0$ there exists $S > 0$ such that $f^{-1}(B_R(f(x))) \subseteq B_S(x)$.

Moreover if $f$ is both bornologous and proper $f$ is said to be **coarse**.

**Definition 1.4.2.** Two metric spaces $X$ and $Y$ are said to be **coarsely equivalent** if there exist two coarse maps $f\colon X \to Y$ and $g\colon Y \to X$ such that $gf$ and $fg$ are each close to the identity map. Here two maps $h, i$ are said to be **close** if the set $\{d(h(x), i(x)) \mid x \in X\}$ is bounded.

Note in this formulation a **coarse embedding**, $f\colon X \to Y$ is a coarse equivalence of $X$ and $f(X)$.

**Definition 1.4.3** ([28])**.** A (discrete) metric space $X$ is said to have **property A** if for all $R > 0$ and all $\epsilon > 0$, there exists a family $\{A_x\}_{x \in X}$ of finite, non-empty subsets of $X \times \mathbb{Z}_{\geq 1}$ such that

1. for all $x, y \in X$ with $d(x, y) \leq R$, we have $\frac{|(A_x \Delta A_y)|}{|(A_x \cap A_y)|} \leq \epsilon$, and

2. there exists a $B > 0$ such that for every $x \in X$, if $(y, n) \in A_x$, then $d(x, y) \leq B$.

One should also recall the following results of Yu which connect spaces with property A to the coarse Baum-Connes conjecture.

**Theorem 1.4.4** ([28])**.** *If a discrete metric space* $\Gamma$ *has property A, then* $\Gamma$ *admits a coarse embedding into Hilbert space.*

**Theorem 1.4.5** ([28])**.** *Let* $\Gamma$ *be a discrete metric space with bounded geometry. If* $\Gamma$ *admits a coarse embedding into Hilbert space, then the coarse Baum-Connes conjecture holds for* $\Gamma$*.*

Of course in light of this it is perhaps interesting to ask if the space of persistence diagrams has Yu's property A. This question was answered in the negative in joint work of the author and Bell et. al., [4].

# Chapter 2: Coarse Embeddability of Wasserstein Space and the Space of Persistence Diagrams

## 2.1 Introduction

In this chapter, we consider vectorizations of two kinds of non-linear data: persistence diagrams, and probability distributions. A persistence diagram is an unordered set of points in the plane which arises as a summary of the topological information in a dataset (e.g. a point cloud or a grayscale image). Persistence diagrams have proven to capture important information in applications using image data [10], geospatial data [16], time series data [24], and more. The set of persistence diagrams is endowed with a family of natural metrics called Wasserstein distances (see Section 2.2). The analysis of persistence diagrams is hampered by the fact that the set of persistence diagrams is not readily identifiable with a subset of Euclidean space. Hence, many vectorizations (i.e. embeddings from the set of persistence diagrams to a Hilbert space) have been introduced in recent years. Examples of embeddings for persistence diagrams include persistence landscapes [6], persistence images [1] and persistence curves [12]. Ideally one would like such embeddings to be isometric; however, not only are these embeddings not so, it can be shown that no isometric embedding of persistence diagrams into Hilbert space exists. Even worse, even if one relaxes the isometric condition (say, to that of a coarse embedding), such an embedding is still theoretically proven not to exist in most cases (see Section 2.3 for a survey of such results).

This chapter also studies probability distributions as objects. The space of all (Borel) probability distributions on $\mathbb{R}^n$ is equipped with a family of metrics, also called Wasserstein distances. Indeed, the persistence diagram distance metrics inherited the name from these distances, which come from the field of optimal transport, due to the similarity between their definitions. Optimal transport and Wasserstein distances have been applied in a variety of areas including economics, machine learning, computer graphics and fluid dynamics. Probability distributions, with Wasserstein metrics, are also difficult to embed in Euclidean space (see Section 2.3).

Despite the large number of negative results regarding embeddings, some important cases remain open. In the case of persistence diagrams, it is not known whether the set of persistence diagrams with the $p$-Wasserstein metric coarsely embeds into Hilbert space for $1 \leq p \leq 2$. In the case of probability distributions, it is not known whether the set of probability distributions on $\mathbb{R}^2$ with the $p$-Wasserstein metric coarsely embeds into Hilbert space for the same range of $p$ values. These spaces are somewhat similar in that persistence diagrams can be thought of as discrete distributions, and one might expect that the answers to these two open questions should be related. In this paper, we confirm this by leveraging a result of Nowak on coarse embeddings of finite subsets [19].

In particular, we show that all finite sets of distributions with the $p$-Wasserstein metric uniformly coarsely embed into the space of persistence diagrams with the $p$-Wasserstein metric (Proposition 2.4.2). If $p > 1$, we obtain the other direction: that finite sets of persistence diagrams embed into Wasserstein space (Proposition 2.4.5). As a corollary, we obtain that if the space of persistence diagrams with the $p$-Wasserstein metric embeds into Hilbert space, for $1 \leq p < \infty$, then so does the space of distributions (with finite $p^{th}$ moment) in $\mathbb{R}^2$ with the $p$-Wasserstein metric; the converse holds if $p > 1$ (Theorem 2.4.8).

## 2.2 Preliminaries

### 2.2.1 Wasserstein Space

We recall the following basic notions in optimal transport from [21].

**Definition 2.2.1.** Let $(X, d_X)$ be a complete separable metric space and let $\mathcal{P}_p(X)$ denote the space of all Borel probability measures on $X$ with finite $p$-th moments, for $1 \leq p < \infty$. The $p$-Wasserstein distance between $\alpha, \beta \in \mathcal{P}_p(X)$ is given by,

$$(W_p(\alpha, \beta))^p = \inf_{\gamma \in \mathcal{U}(\alpha,\beta)} \int d(x, y)^p d\gamma(x, y),$$

Where $\mathcal{U}(\alpha, \beta)$ is the set of Borel probability measures on $X^2$ with marginals $\alpha$ and $\beta$. The metric space $(\mathcal{P}_p(X), W_p)$ is called the **Wasserstein** $p$ **space** over $(X, d_X)$.

While the $p$-Wassertein distance is defined over all measures, in this paper we will be particularly focused on discrete measures. In this case there is an equivalent formulation of the distance which will be useful. Let $\alpha = \sum_{j=1}^{m} a_j \delta_{x_j}$ and $\beta = \sum_{j=1}^{n} b_j \delta_{y_j}$ be discrete measures then,

$$(W_p(\alpha, \beta))^p = \min_{P \in U(a,b)} \sum_{i,j} d(x_i, x_j)^p P_{ij}.$$

Here $U(a, b)$ is the set of $n \times m$ matrices such that $P\mathbb{1}_m = a$ and $P^T\mathbb{1}_n = b$. Moreover, if one additionally assumes that $\alpha$ and $\beta$ have rational coefficients then by rewriting $\alpha =$

$\sum_{j=1}^{N} \frac{1}{N}\delta_{x'_j}$ and $\beta = \sum_{j=1}^{N} \frac{1}{N}\delta_{y'_j}$ it follows from the Birkhoff von Neumann theorem that the distance can be re-expressed as,

$$(W_p(\alpha, \beta))^p = \min_{\sigma \in P(N)} \frac{1}{N} \sum_{j=1}^{N} d(x'_j, y'_{\sigma(j)})^p \tag{2.1}$$

where $P(N)$ is the set of permutations on $\{1, \ldots, N\}$ [21].

### 2.2.2 The Space of Persistence Diagrams

Persistence diagrams typically appear in topological data analysis as a way to store topological information from a sequence of complexes.

**Definition 2.2.2.** Denote by $\mathbb{R}^2_<$ the set $\{(b, d) \in \mathbb{R}^2 \mid b < d\}$. A **persistence diagram** is finite multiset $D \subseteq \mathbb{R}^2_<$.

**Definition 2.2.3.** A **partial matching** between two persistence diagrams $D_1$ and $D_2$ is a triple $(B_1, B_2, f)$ where $B_1 \subseteq D_1$, $B_2 \subseteq D_2$ and $f \colon B_1 \to B_2$ is a bijection.

**Definition 2.2.4.** For $1 \leq p < \infty$ The $p$**-cost** of a partial matching $(B_1, B_2, f)$ between diagrams $D_1$ and $D_2$ is defined as,

$$\text{cost}_p(f) = \left( \sum_{(b,d) \in B_1} d((b, d), f((b, d)))^p \sum_{(b,d) \in D_1 \setminus B_1} d((b, d), \Delta_{\mathbb{R}})^p + \sum_{(b,d) \in D_2 \setminus B_2} d((b, d), \Delta_{\mathbb{R}})^p \right)^{\frac{1}{p}}$$

where $\Delta_{\mathbb{R}} = \{(x, x) \in \mathbb{R}^2$ is the diagonal.

The distance $d$ in the above definition is a distance between points in the plane. Common choices are an $\ell^q$ distance, for $q \geq 1$, or $\ell^\infty$ distance. Since all $\ell^q$ distances on $\mathbb{R}^2$ are bi-Lipschitz equivalent, our results do not depend on the particular choice of $d$; when necessary we assume the $\ell^\infty$ distance.

Now, for two persistence diagrams $D_1, D_2$ we define the possibly infinite distance function

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

Note that $W_p$ is a metric on persistence diagrams, which we call the $p$-**Wasserstein metric** on persistence diagrams.

**Definition 2.2.5.** For $1 \leq p < \infty$, let $\mathcal{D}$ denote the collection of persistence diagrams $D$ that satisfy $W_p(D, \varnothing) < \infty$ modulo the relation $D_1 \sim D_2$ if $W_p(D_1, D_2) = 0$. Here $\varnothing$ represents the empty diagram. The metric space $(\mathcal{D}, W_p)$ is called the **space of persistence diagrams** in the $p$-Wasserstein distance.

We use $W_p$ to denote both the Wasserstein distance between probability distributions and the Wasserstein distance between persistence diagrams; the meaning will always be clear from the arguments. Note that the Wasserstein distance between persistence diagrams is not the same as the Wasserstein distance between the corresponding empirical distributions supported on the points in each diagram as a result of the partial matchings and the use of the diagonal as a universal point to match to. In addition we will assume that the ground norms for both the Wasserstein metric on diagrams and distributions is the infinity norm unless otherwise stated. While other options are available the equivalence of norms makes the choice somewhat superfluous for our arguments.

### 2.2.3 Embeddings

A metric embedding is a map between metric spaces which preserves distances in some manner. We will consider a number of different types of embeddings.

**Definition 2.2.6.** Let $(X, d_X)$ and $(Y, d_Y)$ be metric spaces and let $f\colon X \to Y$ be a map. We say that $f$ is

- an **isometric embedding** if $d(x, y) = d(f(x), f(y)) \forall x, y \in X$;
- a **bi-Lipschitz embedding** if there exists a constant $A \geq 1$ such that $A^{-1} d(x, y) \leq d(f(x), f(y)) \leq A d(x, y)$;
- **$\epsilon$-quasi-isometric embedding** if there exists an $\epsilon > 0$ such that $d(x, y) - \epsilon \leq d(f(x), f(y)) \leq d(x, y) + \epsilon$;
- a **coarse embedding** if there exist non-decreasing functions $\rho_1, \rho_2\colon [0, \infty) \to [0, \infty)$ satisfying $\rho_1(d(x, y)) \leq d(f(x), f(y)) \leq \rho_2(d(x, y)) \quad \forall x, y \in X$ and $\lim_{t \to \infty} \rho_1(t) = +\infty$.

Suppose now that $D \in [1, \infty)$, then $f$ is said to have **distortion at most** $D$ if there exists an $s \in (0, \infty)$ such that

$$\forall x, y \in X \quad s d(x, y) \leq d(f(x), f(y)) \leq D s d(x, y).$$

Moreover, the distortion of $f$, denoted **dist**$(f)$ is the infimum over all $D$ such that the inequality above holds. If a map $f\colon X \to Y$ with distortion $D$ exists, then $X$ is said to embed into $Y$ with distortion $D$. We introduce the notation

$$c_Y(X) = \inf_{f\colon X \to Y} \text{dist}(f).$$

If $\theta \in (0, 1]$, then the **$\theta$-snowflake** of a metric space $(X, d)$ is the metric space $(X, d^\theta)$, where the metric is the obtained by raising $d$ to the $\theta$ power. Following [2], a metric space $X$ is said to be **$\theta$-snowflake universal** if for every finite metric space $(Y, d_Y)$, $c_X(Y, d_Y^\theta) = 1$.

## 2.3 Past results on embeddings

In this section we will give an overview of some embedding results for $p$-Wasserstein space and the space of persistence diagrams. In general, most of these spaces do not embed into Hilbert or Euclidean spaces except under severe restrictions.

**Theorem 2.3.1** (Turner et al., 2014 [25]). *$(\mathcal{D}, W_p)$ does not admit an isometric embedding into Hilbert space for any $1 \leq p \leq \infty$*

**Theorem 2.3.2** (Carrière and Bauer, 2018 [8]). *Let $n \in \mathbb{N}$. Then for any $N \in \mathbb{N}$ and $L > 0$ $(\mathcal{D}_N^L, W_p)$ does not bi-Lipschitz embed into $\mathbb{R}^n$ for $p \in \mathbb{N} \cup \infty$*

Note here the $\mathcal{D}_N^L$ denotes a restricted space of persistence diagrams; namely the space of all diagrams which have fewer than $N$ points and whose points all lie in the region $[-L, L]^2$. To obtain a positive result, we need not only a cardinality restriction, but a relaxation of the embedding type.

**Theorem 2.3.3** (Mitra and Virk, 2018 [18]). *$(\mathcal{D}_N, W_p)$ coarsely embeds into Hilbert space for $1 \leq p \leq \infty$.*

Without the cardinality restriction, even coarse embeddability is not possible in many natural cases. A first result in this direction showed that the space of all persistence diagrams fails to have Yu's Property A [28], a sufficient but not necessary condition for embeddability in Hilbert space originally introduced in Yu's work on the coarse Baum-Connes and Novikov Conjectures.

**Theorem 2.3.4** (Bell et al., 2019 [4]). *$(\mathcal{D}, W_p)$ does not have property $A$ for $1 \leq p < \infty$.*

Later it was shown for all $p > 2$ that the space of persistence diagrams fails to coarsely embed into Hilbert space.

**Theorem 2.3.5** (Bubenik and Wagner, 2020 [7]). *$(\mathcal{D}, W_\infty)$ does not coarsely embed into Hilbert space.*

**Theorem 2.3.6** (Wagner, 2021 [27]). *$(\mathcal{D}, W_p)$ does not coarsely embed into Hilbert space for $2 < p < \infty$.*

Turning to Wasserstein space, Andoni, Naor, and Neiman show that $p$-Wasserstein space on $\mathbb{R}^3$ is $\frac{1}{p}$-snowflake universal. The proof of this result relies on an explicit embedding of any snowflake of a finite metric space into $p$-Wasserstein space on $\mathbb{R}^3$ as a uniform measure.

**Theorem 2.3.7** (Andoni-Naor-Neiman, 2018 [2]). *If $p \in (1, \infty)$ then for every finite metric space $(X, d_X)$ we have*

$$c_{\mathcal{P}_p(\mathbb{R}^3)}\left(X, d_X^{\frac{1}{p}}\right) = 1.$$

As a corollary, Andoni-Naor-Neiman prove that $p$-Wasserstein space on $\mathbb{R}^3$ fails to coarsely embed into Hilbert space for $p > 1$. The case for $\mathbb{R}^2$ and $p = 1$ remains open.

**Theorem 2.3.8** (Andoni-Naor-Neiman, 2018 [2]). *If $p > 1$ then $\mathcal{P}_p(\mathbb{R}^3)$ does not admit a coarse embedding into any Banach space of nontrivial type. In particular, for $p > 1$, $\mathcal{P}_p(\mathbb{R}^3)$ does not admit a coarse embedding into Hilbert space.*

To summarize, coarse embeddability in Hilbert space remains an open question for persistence diagrams when $1 \leq p \leq 2$ and for Wasserstein space when the underlying space is the plane (the most closely related context to that of persistence diagrams). Our main results prove that these two questions are equivalent.

## 2.4 Main Results

For convenience, we will use the notation $d(\cdot, \cdot)$ for distances in $(\mathcal{D}, W_p)$ and $\mathcal{P}_p(\mathbb{R}^2)$.

The following characterization, by Nowak, of coarse embedabilty into Hilbert space by way of finite subsets will prove quite useful.

**Theorem 2.4.1** (Nowak, 2005 [19]). *A metric space $X$ admits a coarse embedding into a Hilbert space if and only if there exists non-decreasing functions $\rho_-, \rho_+ \colon [0, \infty) \to [0, \infty)$ such that $\lim_{t\to\infty} \rho_-(t) = \infty$ and for every finite subset $A \subset X$ there exists a map $f_A \colon A \to \ell_2$ satisfying,*

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

*for every $x, y \in A$.*

Nowak's characterization of coarse embeddings into Hilbert space allows one to restrict one's attention to maps on finite subsets. In light of this the proofs of Theorems 2.4.8 and 2.4.9 rely on finding low distortion maps between finite subsets in these spaces. We now proceed to show the existence of such a map from any finite subset of measures in $\mathcal{P}_p(\mathbb{R}^2)$.

**Proposition 2.4.2.** *Suppose $A = \{\alpha_1, \cdots, \alpha_n\} \subseteq \mathcal{P}_p(\mathbb{R}^2)$. Then for all $\epsilon > 0$ there exists an $\epsilon$-quasi-isometric embedding $f \colon A \to (\mathcal{D}, W_p)$.*

*Proof.* Let $A = \{\alpha_1, \cdots, \alpha_n\} \subseteq \mathcal{P}_p(\mathbb{R}^2)$ and $\epsilon > 0$ be given. It follows [26] that there exists a collection $B = \{\beta_1, \cdots, \beta_n\} \subseteq \mathcal{P}_p(\mathbb{R}^2)$ of rational discrete measures (i.e. discrete measures with rational weights) such that $d(\alpha_i, \beta_i) < \frac{\epsilon}{2}$. By Lemma (2.4.3) below there exists an isometry $f\colon B \to \mathcal{D}$. Define $\bar{f}\colon A \to \mathcal{D}$ by $\alpha_i \mapsto f(\beta_i)$. We check that this is an $\epsilon$-quasi-isometry. Indeed:

$$\begin{aligned} d(\bar{f}(\alpha_i), \bar{f}(\alpha_j)) &= d(f(\beta_i), f(\beta_j)) \\ &= d(\beta_i, \beta_j) \\ &\leq d(\beta_i, \alpha_i) + d(\alpha_i, \alpha_j) + d(\alpha_j, \beta_j) \\ &< d(\alpha_i, \alpha_j) + \epsilon \end{aligned}$$

and

$$\begin{aligned} d(\alpha_i, \alpha_j) &\leq d(\alpha_i, \beta_i) + d(\beta_i, \beta_j) + d(\beta_j, \alpha_j) \\ &< \epsilon + d(\bar{f}(\alpha_i), \bar{f}(\alpha_j)). \end{aligned}$$

□

The proof of Proposistion 2.4.2 relied on the existence of an isometry from finite subsets of rational measures in $\mathcal{P}_p(\mathbb{R}^2)$ to $\mathcal{D}_p$. The desired map is simply a translation of points, but the details are included in Lemma 2.4.3 for completeness.

**Lemma 2.4.3.** *Suppose $A = \{\alpha_1, \cdots, \alpha_n\}$ is a finite subset of discrete rational measures in $\mathcal{P}_p(\mathbb{R}^2)$. Then there exists an isometry $f\colon A \to \mathcal{D}$*

*Proof.* Let $N$ denote the common denominator of all coefficients in the measures $\alpha_1, \cdots, \alpha_n$. We write each $\alpha_i$ as a sum of uniformly weighted Dirac measures (possibly with duplication):

$$\alpha_i = \sum_{j=1}^{N} \frac{1}{N} \delta_{x_j^i}.$$

Let $D$ denote the diameter of the set $\{\frac{1}{N} x_j^i\}_{i,j}$ in $\mathbb{R}^2$. Note that there exists an $x \in \mathbb{R}^2$ such that

$$\{\frac{1}{N} x_j^i + x\} \subset \{(a, b) \in \mathbb{R}^2 \mid a < b\} \text{and}$$
$$d_\infty(\{\frac{1}{N} x_j^i + x\}, \Delta_{\mathbb{R}^2}) > 2D.$$

Define $f\colon A \to \mathcal{D}$ by $\alpha_i \mapsto \{\frac{1}{N} x_j^i + x\}$. From (2.1) we have that

$$(W_p(\alpha, \beta))^p = \min_{\sigma \in P(N)} \frac{1}{N} \sum_{j=1}^{N} \|x'_j - y'_{\sigma(j)}\|_\infty^p.$$

Now, since the diagrams are sufficiently far from the diagonal, we must have that the distance between $f(\alpha)$ and $f(\beta)$ is achieved by a perfect matching, which proves the result. $\square$

We now consider the other direction, namely embedding persistence diagrams into Wasserstein space. The main obstacle here is the existence of matchings to the diagonal.

**Lemma 2.4.4.** *Suppose $D = \{D_1, D_2, \cdots, D_n\} \subset (\mathcal{D}, W_p)$ is a finite subset of diagrams whose points all have multiplicity one. Let $p > 1$, then for all $\epsilon > 0$ there exists an $N \in \mathbb{N}$ and map $f\colon D \to \mathcal{P}^p(\mathbb{R}^2)$ satisfying the following for sufficiently large $s$:*

$$\frac{1}{N+s+1} d(D_i, D_j) \leq d(f(D_i), f(D_j)) \leq \frac{1}{N+s+1} d(D_i, D_j) + \frac{1}{N+s} \epsilon \tag{2.2}$$

*Proof.* Let $D$ be as above and let $\epsilon > 0$ be given. We want to define $f$ to be the map that sends each diagram to a uniform measure on the same number of points. To this end we will first map each diagram to subsets of the plane with the same cardinalities. We begin by fixing some notation. Let $\pi_{ij}$ denote the optimal partial matching between $D_i$ and $D_j$, $\mathcal{U}_{ij}$ denote those points in $D_i$ that are unmatched under this partial matching, $N_i = |D_i|$, $N = \max_i N_i$. Furthermore, define the following:

$$\begin{aligned}
M &= \max_{(x,y) \in \cup_i D_i} \frac{x+y}{2} \\
m &= \min_{(x,y) \in \cup_i D_i} \frac{x+y}{2} \\
\rho_x &= \operatorname*{argmin}_x \{d(x,y) \mid y \in \Delta_{\mathbb{R}}\} \\
I &= \{(m + \frac{t(M-m)}{s}, m + \frac{t(M-m)}{s})\}_{t=0}^{s} \\
I_i &= \{(m + \frac{t(M-m)}{s}, m + \frac{t(M-m)}{s})\}_{t=s+1}^{s+N-N_i}.
\end{aligned}$$

Here $s > N$ is taken so large so that $\frac{M-m}{s} < \frac{\epsilon^p}{3N}$ and $(s+1+N-N_j-|\mathcal{U}_{ij}|)(\frac{N(M-m)}{s})^p < \frac{\epsilon^p}{3}$. We note that $|I| = s+1$ and $|I_i| = N - N_i$. Define $\bar{D}_i = D_i \cup I \cup I_i$ and note $|\bar{D}_i| = N_i + N - N_i + s + 1 = N + s + 1$ for all $i$. Finally, define $f\colon D \to \mathcal{P}^p$ by sending each $D_i$ to the uniform measure on $\bar{D}_i$. We first check that $f$ has the desired upper bound. Let $\sigma_{ij} : \mathcal{U}_{ij} \to \sigma_{ij}(\mathcal{U}_{ij})$ be an optimal matching from $\mathcal{U}_{ij}$ to a subset of $I$ and $\tau_{ji}$ denote an optimal matching from $\mathcal{U}_{ji}$ to a (possibly different) subset of $I$. Now, note that,

$$\begin{aligned}
|I \cup I_i \setminus \tau_{ji}(\mathcal{U}_{ji})| &= s + 1 + N - N_i - |\mathcal{U}_{ji}| \\
&= s + 1 + N - N_j - |\mathcal{U}_{ij}| \\
&= |I \cup I_j \setminus \sigma_{ij}(\mathcal{U}_{ij})|.
\end{aligned}$$

Assume that these sets have been ordered as $\{x_i\}$ and $\{y_i\}$ with the order induced from the standard order on $\mathbb{R}$ and let $\omega$ denote the matching which sends $x_i \in I \cup I_i \setminus \tau_{ji}(\mathcal{U}_{ji})$ to

$y_i \in I \cup I_j \setminus \sigma_{ij}(\mathcal{U}_{ij})$. We define a coupling,$\pi$, between the uniform measures $f(D_i)$ and $f(D_j)$ by,

$$\pi = \frac{1}{N+s+1}\left[\sum_{x\in D_i\setminus\mathcal{U}_{ij}} \delta_{(x,\pi_{ij}(x))} + \sum_{x\in\mathcal{U}_{ij}} \delta_{(x,\sigma_{ij}(x))} + \sum_{x\in\tau_{ji}(\mathcal{U}_{ji})} \delta_{(x,\tau_{ji}^{-1}(x))} + \sum_{x\in I\cup I_i\setminus\mathcal{U}_{ij}} \delta_{(x,\omega(x))}\right].$$

With $x = \left(m + \frac{t_1(M-m)}{s}, m + \frac{t_1(M-m)}{s}\right)$ and $\omega(x) = \left(m + \frac{t_2(M-m)}{s}, m + \frac{t_2(M-m)}{s}\right)$ we note that,

$$\begin{aligned}\|x-\omega(x)\|_\infty &= \frac{|t_1-t_2|(M-m)}{s}\\ &\leq \frac{N(M-m)}{s}.\end{aligned}$$

Thus,

$$\sum_{x\in I\cup I_i\setminus\mathcal{U}_{ij}} \|x-\omega(x)\|_\infty^p \leq (s+1+N-N_j-|\mathcal{U}_{ij}|)(\frac{N(M-m)}{s})^p < \frac{\epsilon^p}{3}.$$

Putting it together,

$$
\begin{aligned}
& d(f(D_i), f(D_j)) \\
\leq & \Big(\int_{\mathbb{R}^2\times\mathbb{R}^2} \|x-y\|_\infty^p d\pi(x,y)\Big)^{\frac{1}{p}} \\
= & \frac{1}{N+s+1}\Bigg[\sum_{x\in D_i\setminus\mathcal{U}_{ij}} \|x-\pi_{ij}(x)\|_\infty^p + \sum_{x\in\mathcal{U}_{ij}} \|x-\sigma_{ij}(x)\|_\infty^p + \\
& \sum_{x\in\tau_{ji}(\mathcal{U}_{ji})} \|x-\tau_{ji}^{-1}(x)\|_\infty^p + \sum_{x\in I\cup I_i\setminus\mathcal{U}_{ij}} \|x-\omega(x)\|_\infty^p\Bigg]^{\frac{1}{p}} \\
\leq & \frac{1}{N+s+1}\Bigg[\sum_{x\in D_i\setminus\mathcal{U}_{ij}} \|x-\pi_{ij}(x)\|_\infty^p + \sum_{x\in\mathcal{U}_{ij}} \left[\|x-\rho_x\|_\infty + \|\rho_x-\sigma_{ij}(x)\|_\infty\right]^p + \\
& \sum_{x\in\tau_{ji}(\mathcal{U}_{ji})} \left[\|x-\rho_x\|_\infty + \|\rho_x-\tau_{ji}^{-1}(x)\|_\infty\right]^p + \sum_{x\in I\cup I_i\setminus\mathcal{U}_{ij}} \|x-\omega(x)\|_\infty^p\Bigg]^{\frac{1}{p}} \\
\leq & \frac{1}{N+s+1}\Bigg[\sum_{x\in D_i\setminus\mathcal{U}_{ij}} \|x-\pi_{ij}(x)\|_\infty^p + \sum_{x\in\mathcal{U}_{ij}} \|x-\rho_x\|_\infty^p + \sum_{x\in\tau_{ji}(\mathcal{U}_{ji})} \|x-\rho_x\|_\infty^p\Bigg]^{\frac{1}{p}} + \\
& \frac{1}{N+s+1}\Bigg[\sum_{x\in\mathcal{U}_{ij}} \|\rho_x-\sigma_{ij}(x)\|_\infty^p + \sum_{x\in\tau_{ji}(\mathcal{U}_{ji})} \|\rho_x-\tau_{ji}^{-1}(x)\|_\infty^p + \sum_{x\in I\cup I_i\setminus\mathcal{U}_{ij}} \|x-\omega(x)\|_\infty^p\Bigg]^{\frac{1}{p}} \\
\leq & \frac{1}{N+s+1}\left[d(D_i,D_j)+\epsilon\right].
\end{aligned}
$$

This completes the proof of the upper bound. For the lower bound, note that any bijective coupling of uniform measures $f(D_i)$ and $f(D_j)$ induces a partial matching on the diagrams $D_i$ and $D_j$. Let $\pi$ denote the optimal coupling and $\pi_{ij}$ the induced partial matching between diagrams $D_i$ and $D_j$, further let $\mathcal{U}_{ij}$ denote those points unmatched under $\pi_{ij}$. Then we have,

$$
\begin{aligned}
d(D_i,D_j)^p &\leq \sum_{x\in D_i\setminus\mathcal{U}_{ij}} \|x-\pi_{ij}(x)\|_\infty^p + \sum_{x\in\mathcal{U}_{ij}} \|x-\rho_x\|_\infty^p + \sum_{x\in\mathcal{U}_{ji}} \|x-\rho_x\|_\infty^p \\
&\leq \sum_{x\in D_i\setminus\mathcal{U}_{ij}} \|x-\pi(x)\|_\infty^p + \sum_{x\in\mathcal{U}_{ij}} \|x-\pi(x)\|_\infty^p \\
&\quad + \sum_{x\in\mathcal{U}_{ji}} \|x-\pi^{-1}(x)\|_\infty^p + \sum_{x\in I\cup I_i\setminus\mathcal{U}_{ij}} \|x-\pi(x)\|_\infty^p \\
&= (N+s+1)d(f(D_i),f(D_j))^p
\end{aligned}
$$

□

For a discrete measure $\alpha = \sum a_i \delta_{x_i}$ and a real number $r$ denote by $\lambda_r(\alpha)$ the dilated measure $\lambda_r(\alpha) = \sum a_i \delta_{rx_i}$. Note that by defining $\bar{f} = \lambda_{N+s+1}(f)$ we may obtain a map on finite subsets of diagrams where the coefficients in the inequality in Lemma 2.4.4 become 1. Moreover, for any diagram there exists a diagram whose points have multiplicity one and which is within $\epsilon$ of the original diagram. Thus, one may relax the restriction that each diagram has multiplicity one simply at the cost of a $-\epsilon$ on the lower bound after scaling. This results in the following form of Lemma 2.4.4 which will be more useful.

**Proposition 2.4.5.** *Suppose $D = \{D_1, D_2, \cdots, D_n\} \subset (\mathcal{D}, W_p)$ is a finite subset of diagrams. Let $p > 1$, then for all $\epsilon > 0$ there exists an $\epsilon$-quasi-isometric embedding $f\colon D \to \mathcal{P}_p(\mathbb{R}^2)$.*

**Proposition 2.4.6.** *Let $X$ and $Y$ be metric spaces and suppose for all $\epsilon > 0$ and all finite subsets $A \subset X$ there exists a map $f\colon A \to Y$ satisfying*

$$d(x_i, x_j) - \epsilon \leq d(f(x_i), f(x_j)) \leq d(x_i, x_j) + \epsilon \tag{2.3}$$

*for all $x_i, x_j \in A$. If $Y$ coarsely embeds into a Hilbert space, then $X$ coarsely embeds into a Hilbert space.*

*Proof.* We proceed by applying Theorem (2.4.1). Let $A$ be a finite subset of $X$. Then by assumption and (2.4.1) there exists $\rho_-, \rho_+\colon [0,\infty) \to [0,\infty)$ with $\lim_{t\to\infty} \rho_-(t) = \infty$ and $g\colon f(A) \to \ell_2$ satisfying

$$\rho_-(d(f(x_i), f(x_j))) \leq \|g(f(x_i)) - g(f(x_j))\| \leq \rho_+(d(f(x_i), f(x_j))).$$

Define $\bar{g} = g \circ f$, $\bar{\rho}_+(t) = \rho_+(t+\epsilon)$ and

$$\bar{\rho}_-(t) = \begin{cases} \rho_-(t-\epsilon) & t > \epsilon \\ 0 & \text{otherwise} \end{cases}$$

Then we have,

$$\begin{aligned} \bar{\rho}_-(d(x_i, x_j)) \leq \rho_-(d(f(x_i), f(x_j))) \leq \|g(f(x_i)) - g(f(x_j))\| &\leq \rho_+(d(f(x_i), f(x_j))) \\ &\leq \bar{\rho}_+(d(x_i, x_j)) \end{aligned}$$

□

**Lemma 2.4.7.** *Let $X$ and $Y$ be metric spaces and suppose for all $\epsilon > 0$ and all finite subsets $A \subset X$ there exists a map $f\colon A \to Y$ satisfying,*

$$d(x_i, x_j) - \epsilon \leq d(f(x_i), f(x_j)) \leq d(x_i, x_j) + \epsilon \tag{2.4}$$

*for all $x_i, x_j \in A$. Then $X$ being $\frac{1}{p}$-snowflake universal implies $Y$ is $\frac{1}{p}$-snowflake universal.*

*Proof.* Let $\theta \in (0,1)$ and $(W, d_W^\theta)$ be the $\theta$-snowflake of a finite metric space $(W, d_W)$. Let $\epsilon > 0$ be given and take $\delta < 1$ so small so that $\frac{1+\delta}{1-\delta} < 1 + \epsilon$. Take $\epsilon_1 < \frac{\delta}{2}$. Then by assumption there exists a map $g\colon W \to X$ and a constant $k_1$ such that,

$$k_1 d(x_i, x_j) \leq d(g(x_i), g(x_j)) \leq k_1(1+\epsilon_1) d(x_i, x_j).$$

Let $M = \min_{x_i \neq x_j} d(x_i, x_j)$ and take $\epsilon_2 < \frac{\delta k_1 M}{2}$. Let $f\colon g(W) \to Y$ be a map satisfying (2.4) with respect to $\epsilon_2$. We claim that $f \circ g$ is the desired map. Indeed,

$$\begin{aligned} d(f(g(x_i)), f(g(x_j))) &\leq d(g(x_i), g(x_j)) + \epsilon_2 \\ &\leq k_1(1+\epsilon_1) d(x_i, x_j) + \epsilon_2 \\ &\leq k_1(1+\delta) d(x_i, x_j) \\ &\leq k_1(1-\delta)(1+\epsilon) d(x_i, x_j). \end{aligned}$$

Further,

$$\begin{aligned} d(f(g(x_i)), f(g(x_j))) &\geq d(g(x_i), g(x_j)) - \epsilon_2 \\ &\geq k_1 d(x_i, x_j) - \epsilon_2 \\ &= k_1 d(x_i.x_j) \left[1 - \frac{\epsilon_2}{k_1 d(x_i, x_j)}\right] \\ &\geq k_1(1-\delta) d(x_i, x_j) \end{aligned}$$

□

Together, Proposition 2.4.2, Theorem 2.4.5 and Proposition 2.4.6 complete the proof our main result, which connects the embeddability questions for persistence diagrams and Wasserstein space.

**Theorem 2.4.8** (A)**.** *If $(\mathcal{D}, W_p)$ coarsely embeds into Hilbert space, for $1 \leq p < \infty$, then so does $\mathcal{P}_p(\mathbb{R}^2)$. The converse holds if $p > 1$.*

Since snowflake universality depends only on finite subsets, Lemma 2.4.7 and Propositions 2.4.2, 2.4.5 yield the following.

**Theorem 2.4.9** (B)**.** *If $\mathcal{P}_p(\mathbb{R}^2)$ is $\frac{1}{p}$-snowflake universal, for $1 \leq p < \infty$, then so is $(\mathcal{D}, W_p)$. The converse holds if $p > 1$.*

We conclude this section with a corollary which follows from Theorems 2.3.5, 2.3.6, and 2.4.8. We note that a direct proof is also possible adapting techniques from Wagner [27].

**Corollary 2.4.10.** *The space $\mathcal{P}_p(\mathbb{R}^2)$ does not admit a coarse embedding into a Hilbert space if $p > 2$.*

## 2.5 Concluding remarks

Wasserstein space and the space of persistence diagrams have many similarities, especially when viewing persistence diagrams as discrete distributions; this similarity has been explored elsewhere using partial optimal transport [13]. The main difference between the spaces is the presence of the diagonal as a sink for unmatched points. Our results suggest that, for $p > 1$, this difference does not affect the coarse embeddability of these spaces. Obstructions to embeddability developed in either case will therefore work just as well for the other. On the other hand, for $p = 1$, our construction degenerates in a similar way to that in [2] and so we do not obtain an equivalence. The $p = 1$ case is important since it appears in many stability results for vectorizations [1, 12]. Note that stability is one half of coarse (or bi-Lipschitz) embeddability, as it bounds the distortion in Hilbert space in terms of the distance between diagrams. Our hope is that our results motivate the use of techniques for Wasserstein space to be used to resolve the question of embeddability of persistence diagrams for $1 \leq p \leq 2$.

# Chapter 3: Gaussian persistence curves

*This chapter is adapted from the preprint "Gaussian Persistence Curves" by Yu-Min Chung, Michael Hull, Austin Lawson and Neil Pritchard [11].*

## 3.1 Introduction

One of the main tools of topological data analysis (TDA) is persistent homology, which measures how certain topological features of a data set appear and disappear at different scales. This information can be stored and visualized in a concise format called a persistence diagram.

Functional summaries play an important role in topological data analysis, as they allow one to apply machine and deep learning techniques to analyze topological information contained in persistence diagrams. In [10] a new class of one-dimensional smooth functional summaries was introduced called Gaussian persistence curves (GPC's). These functional summaries were built by combining (a slight variation of) the persistence curve framework from [12] with the persistence surfaces construction from [1], and they were used to study the texture classification of grey-scale images [10].

In this paper, we investigate the injectivity of both persistence surfaces and GPC's. Loosely speaking the injectivity of a summary implies that the summary can distinguish between distinct diagrams. We show that under mild conditions unweighted GPC's are almost injective (Theorem 3.3.4). Furthermore we show that unweighted persistence surfaces are injective (Theorem 3.3.3).

Other summaries in topological data analysis include persistence landscapes [6], the persistent entropy summary function [3], persistence silhouettes [9], persistence surfaces and persistence images [1]. We refer to [5] for a review of the properties and applications of these summaries. All of these other summaries are known to be stable, but among them only the persistence landscapes are known to be injective. Note that persistence landscapes can be viewed as a sequences of one-dimensional functions and for any $n \geq 1$ injectivity will fail if only the first $n$ terms of the sequences are considered.

The outline of this chapter is as follows. In Section 3.2 we introduce Gaussian persistence curves and derive some basic properties and useful formulas. In Section 3.3 we investigate

the extent to which the functional summaries produced by persistence surfaces and GPC's are injective.

## 3.2 Gaussian Persistence Curves

Here we describe the construction of Gaussian persistence curves which can be viewed as a combination of the persistence surfaces introduced in [1] and the persistence curve framework from [12]. We refer to [14] for background on persistent homology and persistence diagrams.

The input to our construction is a persistence diagram, by which we mean a finite multiset $D$ of points in the plane that lie above the main diagonal $y = x$. Let $\mathbf{\Sigma}$ be a symmetric, positive semi-definite $2 \times 2$ matrix. For a point $\boldsymbol{\mu} \in \mathbb{R}^2$, Let $g_{\boldsymbol{\mu},\mathbf{\Sigma}}$ be the probability density function (PDF) of a bivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$. That is,

$$g_{\boldsymbol{\mu},\Sigma}(\mathbf{x}) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)}{2\pi|\mathbf{\Sigma}|^{1/2}}.$$

Let $\kappa\colon \mathbb{R}^2 \to \mathbb{R}$ be a function with $\kappa(b,b) = 0$ for all $b \in \mathbb{R}$. We refer to any such $\kappa$ as a weighting function.

**Definition 3.2.1.** [1] The persistence surface associated to the diagram $D$ with weight $\kappa$ is the function

$$\rho_{D,\kappa}(x,y) = \sum_{(b,d)\in D} \kappa(b,d) g_{(b,d),\mathbf{\Sigma}}(x,y).$$

We will assume that $\sigma$ is fixed in advance and so do not include it in the notation for the persistence surface or curve.

In [1] the authors choose a grid on the plane, integrate the persistence surface over each box in the grid, and then use these values to produce a vector summary of the original diagram. Instead, we look at this surface from the perspective of the persistence curve framework from [12]. This framework produces a function $G\colon \mathbb{R} \to \mathbb{R}$ such that the value of $G(t)$ depends on measuring some property of the diagram inside the *fundamental box* $F_t = \{(x,y) \mid x < t, y > t\}$.

**Definition 3.2.2.** Let $\rho_{D,\kappa}$ be a persistence surface. The corresponding **Gaussian persistence curve** is the function

$$G_{D,\kappa}(t) = \int_{F_t} \rho_{D,\kappa}(x,y)dxdy.$$

If $\kappa(x,y) = 1$ for all $(x,y)$, then we drop $\kappa$ from the notation and denote the corresponding surface and curve by $\rho_D(t)$ and $G_D(t)$ respectively. We refer to this curve as the unweighted Gaussian persistence curve on $D$.

We will always consider $\Sigma$ to be fixed ahead of time and do not include it in the notation for the curve $G_{D,\kappa}(t)$. While the definition makes sense for more general $\Sigma$, in this chapter we fix $\Sigma$ to be a multiple of the identity matrix by a scalar $\sigma^2$. This allows us to split $g_{\boldsymbol{\mu},\boldsymbol{\Sigma}}$ as

$$g_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(x,y) = \phi(\frac{y-d}{\sigma})\phi(\frac{x-b}{\sigma})$$

where $\phi$ is the PDF of the standard normal distribution. This assumption allows one to easily preform the integration over the fundamental box and obtain the **CDF realization** of $G_{D,\kappa}(t)$ as

$$G_{D,\kappa}(t) = \sum_{(b,d)\in D} \kappa_D(b,d)\Phi(\frac{t-b}{\sigma})\Phi(\frac{d-t}{\sigma}),$$

where $\Phi$ denotes the CDF of the standard normal distribution.

Gaussian persistence curves fit into (a slight modification of) the persistence curve framework from [12] in the following way. Let $\mathcal{D}$ be the set of all persistence diagrams, $\Psi$ be the set of all functions $\psi : \mathcal{D} \times \mathbb{R}^3 \to \mathbb{R}$ with $\psi(D;x,x,t) = 0$ for all $(x,x) \in \mathbb{R}^2$ and $D \in \mathcal{D}$. Let $\mathcal{R}$ represent the set of functions from $\mathbb{R}$ to $\mathbb{R}$. Let $\mathcal{T}$ be a set of operators $T(S,f)$ that read in a multi-set $S$ and real-valued function $f$ and returns a scalar. Given $D \in \mathcal{D}$, $\psi \in \Psi$, and $T \in \mathcal{T}$, the corresponding persistence curve is the function

$$P_{D,\psi,T}(t) := T(F_t, \psi(D;x,y,t)),\ t \in \mathbb{R}.$$

The function $P_{D,\psi,T}(t)$ is called a **persistence curve** on $D$ with respect to $\psi$ and $T$. In this notation, choosing $\psi(D;x,y,t) = \rho_{D,\kappa}(x,y)$ and $T(f,S) = \int_S f(x,y)dxdy$, we obtain$P_{D,\psi,T}(t) = G_{D,\kappa}(t)$.

We start with a few examples of Gaussian persistence curves, which are smooth versions of persistence curves appearing in [12].

**Example 3.2.3.** When $\kappa_D(b,d) = 1$ for all $(b,d)$, the resulting unweighted Gaussian persistence curve can be viewed as a smooth version of the Betti curve from [12]. For this reason we also refer to the unweighted Gaussian persistence curve $G_D(t)$ as the *Gaussian Betti Curve*.

**Example 3.2.4.** Let $\ell_D(x,y) := (y-x)\cdot\chi_D(x,y)$, where $\chi$ denotes the characteristic function, and let $\ell_{sum} = \sum_{(b,d)\in D}\ell(b,d)$. Define $\kappa_D(x,y) = \frac{\ell(x,y)}{\ell_{sum}}$. The corresponding Gaussian persistence curve is a smooth version of the life curve from [12] which we call the *Gaussian Life Curve*.

**Example 3.2.5.** Let $m_{sum} = \sum_{(b,d)\in D}(b+d)$ and define $\kappa_D(x,y) = \frac{x+y}{m_{sum}}$. The corresponding Gaussian persistence curve is a smooth version of the midlife curve from [12] which we call the *Gaussian midlife Curve*.

With this set up, generating new Gaussian persistence curves is only a matter of selecting a covariance matrix $\Sigma$, which controls the smoothness of the curve and a function $\kappa$, which is a weighting function. For example, by using weight functions such as the entropy function ($-\frac{d-b}{\sum_{(b,d)\in D} d-b} \log \frac{d-b}{\sum_{(b,d)\in D} d-b}$) and multiplicative life function ($\frac{d}{b}$), we can obtain Gaussian versions of the life entropy and multiplicative life persistence curves. In general, we can produce a Gaussian version of any function in the persistence curve framework.

The next two lemmas will be used to compute the $L^1$–norm of a Gaussian persistence curve. Their proofs are elementary exercises in Calculus. We include their elementary proofs for completeness.

**Lemma 3.2.6.** *Given $b > 0 \in \mathbb{R}$,*

$$\int_{-\infty}^{\infty} \Phi(\frac{b-t}{\sigma})\Phi(\frac{t-b}{\sigma})\mathrm{d}t = \frac{\sigma}{\sqrt{\pi}}. \tag{3.1}$$

*Proof.* We will first prove the result for $b = 0$ and $\sigma = 1$. By integration by parts (by letting $u = \Phi(-t)$ and $\mathrm{d}v = \Phi(t)\mathrm{d}t$), we obtain

$$\int_{-\infty}^{\infty} \Phi(-t)\Phi(t)\mathrm{d}t = \underbrace{[t\Phi(-t)\Phi(t) + \Phi(-t)\phi(t)]_{-\infty}^{\infty}}_{I} + \underbrace{\int_{-\infty}^{\infty} t\phi(t)\Phi(t) + \phi^2(t)\mathrm{d}t}_{II}.$$

For $I$, by the elementary facts $\Phi(t) \to 1$ as $t \to \infty$ and L'Hopital's rule, one can evaluate that $I = 0$.

For the $II$, we consider each integral separately. By [20],
we have that $\int_{\infty}^{\infty} t\Phi(a+bt)\phi(t)\mathrm{d}t = \frac{b}{\sqrt{1+b^2}}\phi(\frac{a}{\sqrt{1+b^2}})$. Since in our case $a = 0$ and $b = 1$,

$$\int_{-\infty}^{\infty} t\Phi(t)\phi(t)\mathrm{d}t = \frac{1}{\sqrt{2}}\phi(0) = \frac{1}{2\sqrt{\pi}}.$$

By [20] again, we know that

$$\int_{-\infty}^{\infty} \phi^2(t)\mathrm{d}t = \frac{1}{2\sqrt{\pi}}\Phi(\sqrt{2}t)\Big|_{-\infty}^{\infty} = \frac{1}{2\sqrt{\pi}}.$$

Thus, sum over them to obtain the desired result. To obtain the final result simply apply the substitution, $s = \frac{t-b}{\sigma}$ then $ds = \frac{1}{\sigma}dt$.

$$\int_{-\infty}^{\infty} \Phi(\frac{b-t}{\sigma})\Phi(\frac{t-b}{\sigma})\mathrm{d}t = \int_{-\infty}^{\infty} \Phi(-s)\Phi(s)\sigma\mathrm{d}s = \frac{\sigma}{\sqrt{\pi}}.$$

□

**Lemma 3.2.7.**

$$\int_{-\infty}^{\infty} \Phi(at+b_1)(\Phi(at+b_2)-\Phi(at+b_3))\mathrm{d}t \tag{3.2}$$

$$= \frac{-\sqrt{2}}{a}\left[\frac{b_1-b_2}{\sqrt{2}}\Phi(\frac{b_1-b_2}{\sqrt{2}})+\phi(\frac{b_1-b_2}{\sqrt{2}})-\frac{b_1-b_3}{\sqrt{2}}\Phi(\frac{b_1-b_3}{\sqrt{2}})-\phi(\frac{b_1-b_3}{\sqrt{2}})\right] \tag{3.3}$$

*Proof.* Let $s = u + (at + b_3)$. Then $\mathrm{d}s = \mathrm{d}u$.

$$\begin{aligned}
&\int_{-\infty}^{\infty} \Phi(at+b_1)(\Phi(at+b_2)-\Phi(at+b_3))\mathrm{d}t \\
&= \int_{-\infty}^{\infty} \Phi(at+b_1)\int_{at+b_3}^{at+b_2} \phi(s)\,\mathrm{d}s\mathrm{d}t \\
&= \int_{-\infty}^{\infty}\int_{at+b_3}^{at+b_2} \Phi(at+b_1)\phi(s)\,\mathrm{d}s\mathrm{d}t \\
&= \int_{-\infty}^{\infty}\int_{0}^{b_2-b_3} \Phi(at+b_1)\phi(at+b_3+u)\,\mathrm{d}u\mathrm{d}t \\
&= \int_{0}^{b_2-b_3}\int_{-\infty}^{\infty} \Phi(at+b_1)\phi(at+b_3+u)\,\mathrm{d}t\mathrm{d}u.
\end{aligned}$$

To evaluate the integral, we recall that $\int_{-\infty}^{\infty}\Phi(x+\epsilon)\phi(x)\,\mathrm{d}x = \Phi(\frac{\epsilon}{\sqrt{2}})$. Consider another substitution: $x = at + b_3 + u$, so $\mathrm{d}x = a\mathrm{d}t$.

$$\begin{aligned}
&\int_{-\infty}^{\infty} \Phi(at+b_1)\phi(at+b_3+u)\,\mathrm{d}t \\
&= \int_{-\infty}^{\infty} \Phi(x-u-b_3+b_1)\phi(x)\frac{1}{a}\,\mathrm{d}x = \frac{1}{a}\Phi(\frac{b_1-b_3-u}{\sqrt{2}}).
\end{aligned}$$

Let $v = \frac{b_1-b_3-u}{\sqrt{2}}$. Then $\mathrm{d}v = \frac{-1}{\sqrt{2}}\mathrm{d}u$. Finally, we obtain

$$\begin{aligned}
&\int_{0}^{b_2-b_3}\int_{-\infty}^{\infty} \Phi(at+b_1)\phi(at+b_3+u)\,\mathrm{d}t\mathrm{d}u \\
&= \int_{0}^{b_2-b_3} \frac{1}{a}\Phi(\frac{b_1-b_3-u}{\sqrt{2}})\,\mathrm{d}u. \\
&= \int_{\frac{b_1-b_3}{\sqrt{2}}}^{\frac{b_1-b_2}{\sqrt{2}}} \frac{-1}{a}\sqrt{2}\Phi(v)\mathrm{d}v = \frac{-\sqrt{2}}{a}\left[v\Phi(v)+\phi(v)\right]_{\frac{b_1-b_3}{\sqrt{2}}}^{\frac{b_1-b_2}{\sqrt{2}}} \\
&= \frac{-\sqrt{2}}{a}\left[\frac{b_1-b_2}{\sqrt{2}}\Phi(\frac{b_1-b_2}{\sqrt{2}})+\phi(\frac{b_1-b_2}{\sqrt{2}})-\frac{b_1-b_3}{\sqrt{2}}\Phi(\frac{b_1-b_3}{\sqrt{2}})-\phi(\frac{b_1-b_3}{\sqrt{2}})\right].
\end{aligned}$$

□

**Proposition 3.2.8.** *Let $G_D(t)$ be an unweighted Gaussian persistence curve on a diagram $D$. Then*

$$\|G_D(t)\|_1 = \sum_{(b,d)\in D} \left[ (d-b)\Phi(\frac{d-b}{\sqrt{2}\sigma}) + \sqrt{2}\sigma\phi(\frac{d-b}{\sqrt{2}\sigma}) \right]. \tag{3.4}$$

*Proof.*

$$\|G_D\|_1 = \int_{-\infty}^{\infty} \left| \sum_{(b,d)\in D} \Phi(\frac{t-b}{\sigma})\Phi(\frac{d-t}{\sigma}) \right| \mathrm{d}t \tag{3.5}$$

$$= \int_{-\infty}^{\infty} \sum_{(b,d)\in D} \Phi(\frac{t-b}{\sigma})\Phi(\frac{d-t}{\sigma})\mathrm{d}t \tag{3.6}$$

$$= \sum_{(b,d)\in D} \int_{-\infty}^{\infty} \Phi(\frac{t-b}{\sigma})\Phi(\frac{d-t}{\sigma})\mathrm{d}t. \tag{3.7}$$

By adding and subtracting $\Phi(\frac{t-b}{\sigma}) + \Phi^2(\frac{t-b}{\sigma})$ and using the CDF property $\Phi(-t) = 1 - \Phi(t)$ we obtain,

$$= \sum_{(b,d)\in D} \int_{-\infty}^{\infty} \Phi(\frac{t-b}{\sigma})\Phi(\frac{b-t}{\sigma}) + \Phi(\frac{t-b}{\sigma}) \left( \Phi(\frac{t-b}{\sigma}) - \Phi(\frac{t-d}{\sigma}) \right) \mathrm{d}t \tag{3.8}$$

$$= \sum_{(b,d)\in D} \frac{\sigma}{\sqrt{\pi}} + (d-b)\Phi(\frac{d-b}{\sqrt{2}\sigma}) + \sqrt{2}\sigma\phi(\frac{d-b}{\sqrt{2}\sigma}) - \frac{\sigma}{\sqrt{\pi}} \tag{3.9}$$

$$= \sum_{(b,d)\in D} \left[ (d-b)\Phi(\frac{d-b}{\sqrt{2}\sigma}) + \sqrt{2}\sigma\phi(\frac{d-b}{\sqrt{2}\sigma}) \right]. \tag{3.10}$$

where (9) follows from Lemmas 3.2.6 and 3.2.7. □

Let $L_D = \sum_{\{(b,d)\in D\}} d - b$. We refer to $L_D$ as the **total lifespan** of $D$. We also define $\delta_D = \min_{(b,d)\in D} d - b$, that is $\delta_D$ is the **minimum lifespan** of a point in $D$. We note that by convention, $\min \emptyset = \infty$ and $\frac{1}{\min \emptyset} = 0$.

**Corollary 3.2.9.** *For any persistence diagram $D$,*

$$\|G_D(t)\|_1 \leq \sum_{(b,d)\in D} \left[ (d-b) + \frac{\sigma}{\sqrt{\pi}} \right] \leq (1 + \frac{\sigma}{\sqrt{\pi}\delta_D})L_D.$$

Apply the same argument as above gives a bound for the weighted case as well. Let $M_{D,\kappa} = \max_{(b,d)\in D} |\kappa(b,d)|$.

**Corollary 3.2.10.**

$$\|G_{D,\kappa}(t)\|_1 = \sum_{(b,d)\in D} |\kappa(b,d)| \left[ (d-b)\Phi(\frac{d-b}{\sqrt{2}\sigma}) + \sqrt{2}\sigma\phi(\frac{d-b}{\sqrt{2}\sigma}) \right].$$
$$\leq \sum_{(b,d)\in D} |\kappa(b,d)| \left[ (d-b) + \frac{\sigma}{\sqrt{\pi}} \right] \leq (1 + \frac{\sigma}{\sqrt{\pi}\delta_D}) M_{D,\kappa} L_D$$

As long as the diagram $D$ is finite, then the Gaussian persistence curve given by Definition 3.2.2 will be a Lipschitz function with respect to the input $t \in \mathbb{R}$ (see [10]). Together with some mild assumptions on the weight functions $\kappa$, this implies that whenever there is is a process for randomly sampling persistence diagrams the associated Gaussian persistence curves will satisfy a version of the central limit theorem. See [10] for details, or [5] for more general results about statistical properties of functional summaries.

## 3.3 Injectivity

In this section we study the injectivity of the transformation from a persistence diagram to either the persistence surface or the corresponding Gaussian persistence curve. By injectivity here we mean that distinct diagrams produce distinct persistence surfaces or curves. In general, this notion depends on the choice of weight functions, see example 3.3.5. However, we show the injectivity of unweighted persistence surfaces and (in most cases) unweighted Gaussian persistence curves. We also conjecture that any weight functions defined independently of the diagrams will produce injective persistence surfaces and curves.

We first show injectivity for unweighted persistence surfaces. Given two diagrams $C$ and $D$, this amounts to showing that the multi-set of means in the corresponding surfaces is equal. We achieve this by setting up an infinite system of equations, which can only be solved when the multi-sets of means are exactly equal. We now proceed to some techinical lemmas, but, first recall that for a positive integer $n$, $n!! = \Pi_{k=0}^{\lceil \frac{n}{2} \rceil - 1}(n-2k)$.

**Lemma 3.3.1.** *Let $A, B \subset \mathbb{R}$ be finite sets of equal cardinality. Suppose*

$$\sum_{a\in A} \phi\left(\frac{x-a}{\sigma}\right) = \sum_{b\in B} \phi\left(\frac{x-b}{\sigma}\right).$$

*Then for every $n \in \mathbb{N}$,*

$$\sum_{a\in A} a^n = \sum_{b\in B} b^n$$

*Proof.* Let $n \in \mathbb{N}$ and suppose

$$\sum_{a\in A} \phi\left(\frac{x-a}{\sigma}\right) = \sum_{b\in B} \phi\left(\frac{x-b}{\sigma}\right).$$

Multiplying by $\frac{x^n}{\sigma}$ and integrating over $\mathbb{R}$ with respect to $x$ gives

$$\sum_{a\in A}\int_{\mathbb{R}} x^n \frac{1}{\sigma}\phi(\frac{x-a}{\sigma})dx = \sum_{b\in B}\int_{\mathbb{R}} x^n \frac{1}{\sigma}\phi(\frac{x-b}{\sigma})dx \tag{3.11}$$

$$\sum_{a\in A}\sum_{k \text{ even}}^{n}\binom{n}{k}a^{n-k}\sigma^k(k-1)!! = \sum_{b\in B}\sum_{k \text{ even}}^{n}\binom{n}{k}b^{n-k}\sigma^k(k-1)!!. \tag{3.12}$$

Here (3.12) follows from computing the $n^{th}$ moment of the normal distribution. We will now prove that $\sum_{a\in A} a^n = \sum_{b\in B} b^n$ for all $n \in \mathbb{N}$ by induction. When $n = 1$ this follows immediately from equation (3.12) above. Now assume that $\sum_{a\in A} a^m = \sum_{b\in B} b^m$ for all $m < n$. Then again by equation (3.12) we have,

$$\begin{aligned}
\sum_{a\in A}\sum_{k \text{ even}}^{n}\binom{n}{k}a^{n-k}\sigma^k(k-1)!! &= \sum_{b\in B}\sum_{k \text{ even}}^{n}\binom{n}{k}b^{n-k}\sigma^k(k-1)!! \\
\sum_{k \text{ even}}^{n}\binom{n}{k}\sigma^k(k-1)!!\sum_{a\in A}a^{n-k} &= \sum_{k \text{ even}}^{n}\binom{n}{k}\sigma^k(k-1)!!\sum_{b\in B}b^{n-k} \\
\sum_{a\in A}a^n &= \sum_{b\in B}b^n,
\end{aligned}$$

where the last step follows by our inductive hypothesis. □

**Lemma 3.3.2.** *Let $A, B \subset \mathbb{R}^2$ be finite sets of equal cardinality. Suppose*

$$\sum_{(a,b)\in A}\phi\left(\frac{x-a}{\sigma}\right)\phi\left(\frac{y-b}{\sigma}\right) = \sum_{(\alpha,\beta)\in B}\phi\left(\frac{x-\alpha}{\sigma}\right)\phi\left(\frac{y-\beta}{\sigma}\right).$$

*Then for any $m_1, m_2 \in \mathbb{N}$,*

$$\sum_{(a,b)\in A} a^{m_1}b^{m_2} = \sum_{(\alpha,\beta)\in B}\alpha^{m_1}\beta^{m_2}.$$

*Proof.* Let $m_1, m_2 \in \mathbb{N}$ and suppose

$$\sum_{(a,b)\in A}\phi\left(\frac{x-a}{\sigma}\right)\phi\left(\frac{y-b}{\sigma}\right) = \sum_{(\alpha,\beta)\in B}\phi\left(\frac{x-\alpha}{\sigma}\right)\phi\left(\frac{y-\beta}{\sigma}\right).$$

Multiplying by $x^{m_1}y^{m_2}$ and integrating over $\mathbb{R}$ with respect to $x$ and $y$ yields,

$$\begin{aligned}
&\sum_{(a,b)\in A}\int_{\mathbb{R}} x^{m_1}\phi\left(\frac{x-a}{\sigma}\right)dx\int_{\mathbb{R}} y^{m_2}\phi\left(\frac{y-b}{\sigma}\right)dy \\
&=\sum_{(\alpha,\beta)\in B}\int_{\mathbb{R}} x^{m_1}\phi\left(\frac{x-\alpha}{\sigma}\right)dx\int_{\mathbb{R}} y^{m_2}\phi\left(\frac{y-\beta}{\sigma}\right)dy \qquad (3.13)\\
&\sum_{(a,b)\in A}\sum_{k\text{ even}}^{m_1}\binom{m_1}{k}a^{m_1-k}\sigma^k(k-1)!!\sum_{l\text{ even}}^{m_2}\binom{m_2}{l}b^{m_2-l}\sigma^l(l-1)!! \\
&=\sum_{(\alpha,\beta)\in B}\sum_{k\text{ even}}^{m_1}\binom{m_1}{k}\alpha^{m_1-k}\sigma^k(k-1)!!\sum_{l\text{ even}}^{m_2}\binom{m_2}{l}\beta^{m_2-l}\sigma^l(l-1)!!. \qquad (3.14)
\end{aligned}$$

Now when $m_1 = m_2 = 1$, equation 3.14 yields

$$\sum_{(a,b)\in A} ab = \sum_{(\alpha,\beta)\in B}\alpha\beta.$$

We proceed by proving that the claim is true for all $m_1 + m_2 = n \in \mathbb{N}$ by strong induction on $n$. The base case has been proven above so assume that the statement is true for all $m_1 + m_2 = j < n \in \mathbb{N}$. Note that $m_1 - k + m_2 - l = j$ where $j \in \mathbb{N}$ and $j < n$ for all $k, l \in 2\mathbb{N}$ with $k \le m_1, l \le m_2$. Then the inductive hypothesis and equation 3.14 yields

$$\sum_{(a,b)\in A} a^{m_1}b^{m_2} = \sum_{(\alpha,\beta)\in B}\alpha^{m_1}\beta^{m_2}$$

as desired. □

**Theorem 3.3.3.** *Let $C$ and $D$ be persistence diagrams and let $\rho_C$ and $\rho_D$ be the corresponding unweighted persistence surfaces. If $\rho_C \equiv \rho_D$, then $C = D$.*

*Proof.* We will denote the points of $C$ by $(b_i^C, d_i^C)$ and analogously the points of $D$ will be denoted $(b_i^D, d_i^D)$. Assume that $\rho_C = \rho_D$. We first note that $\int_{\mathbb{R}^2}\rho_C = |C|$. Thus, we must have $|C| = |D|$. We will assign $N = |C|$. Thus, we have

$$\sum_{i=1}^{N}\phi\left(\frac{x-b_i^C}{\sigma}\right)\phi\left(\frac{y-d_i^C}{\sigma}\right) = \sum_{i=1}^{N}\phi\left(\frac{x-b_i^D}{\sigma}\right)\phi\left(\frac{y-d_i^D}{\sigma}\right).$$

Integrating over $\mathbb{R}$ with respect to $y$ and dividing by $\sigma$ yields

$$\sum_{i=1}^{N}\phi\left(\frac{x-b_i^C}{\sigma}\right) = \sum_{i=1}^{N}\phi\left(\frac{x-b_i^D}{\sigma}\right).$$

An application of Lemma 3.3.1 yields $\sum_{i=1}^{N}(b_i^C)^n = \sum_{i=1}^{N}(b_i^D)^n$ for all $n \in \mathbb{N}$. A similar method yields $\sum_{i=1}^{N}(d_i^C)^n = \sum_{i=1}^{N}(d_i^D)^n$. Next, an application of Lemma 3.3.2 yields that for each $m_1, m_2 \in \mathbb{N}$ we have $\sum_{i=1}^{N}(b_i^C)^{m_1}(d_i^C)^{m_2} = \sum_{i=1}^{N}(b_i^D)^{m_1}(d_i^D)^{m_2}$.

Before proceeding we recall the following basic fact: suppose $a_1, ..., a_k$ are non-negative real numbers and $c_1, ..., c_k$ are positive real numbers such that $a_k > a_i$ for all $1 \leq i < k$. Then there exists $m \in \mathbb{N}$ such that

$$c_k a_k^m > \sum_{i=1}^{k-1} c_i a_i^m.$$

Now, label the points in $C$ such that $b_1^C \leq b_2^C ... \leq b_N^C$ and whenever $b_i^C = b_{i+1}^C$, $d_i^C \leq d_{i+1}^C$. Label the points in $D$ in the same way.

Suppose that some $b_k^C \neq b_k^D$. Choosing $k$ to be the largest such index, it follows from above that for all $n \geq 1$.

$$\sum_{i=1}^{k}(b_i^C)^n = \sum_{i=1}^{k}(b_i^D)^n.$$

Without loss of generality, we assume that $b_k^C > b_k^D$. Since $b_k^D \geq b_{k-1}^D \geq ... \geq b_1^D$, we can find an $m$ such that

$$\sum_{i=1}^{k}(b_i^C)^m \geq (b_k^C)^m > \sum_{i=1}^{k}(b_i^D)^m$$

which is a contradiction. Hence we have that $b_i^C = b_i^D$ for all $1 \leq i \leq N$.

Now suppose that for some $k$, $d_k^C \neq d_k^D$. Again we choose $k$ to be largest such index and assume without loss of generality that $d_k^C > d_k^D$. Now we choose $m_1$ such that $(b_k^C)^{m_1} d_k^C > (b_i^D)^{m_1} d_i^D$ for all $1 \leq i < k$. $m_1$ exists since for all $1 \leq i \leq k$, either $b_k^C = b_k^D > b_i^D$, or $b_k^C = b_k^D = b_i^D$ and $d_k^C > d_k^D \geq d_i^D$.

Given this $m_1$, we can apply the above fact again to find $m_2$ such that

$$\sum_{i=1}^{k}((b_i^C)^{m_1} d_i^C)^{m_2} \geq ((b_k^C)^{m_1} d_k^C)^{m_2} > \sum_{i=1}^{k}((b_i^D)^{m_1} d_i^D)^{m_2}$$

which is another contradiction. Hence, we must have that $d_i^C = d_i^D$ for all $1 \leq i \leq N$ which means we have shown that $C = D$.

□

We prove a partial result for the unweighted Gaussian Persistence Curve.

**Theorem 3.3.4.** *Let $C$ and $D$ be two persistence diagrams with maximum death values $d_{max}^C, d_{max}^D$ and minimum birth values $b_{min}^C, b_{min}^D$ respectively. Suppose that either $d_{max}^C \neq d_{max}^D$ or $b_{min}^C \neq b_{min}^D$. Then $G_C \neq G_D$.*

*Proof.* Let $(b_1, d_1)$ be a point of $C$ and $(b_2, d_2)$ a point of $D$. We first show that if $d_1 > d_2$ then,

$$\lim_{t\to\infty} \frac{\Phi(\frac{t-b_2}{\sigma})\Phi(\frac{d_2-t}{\sigma})}{\Phi(\frac{t-b_1}{\sigma})\Phi(\frac{d_1-t}{\sigma})} = 0. \tag{3.15}$$

We will proceed by cases. If $b_1 = b_2$ then with L'Hopital's rule we have,

$$\lim_{t\to\infty} \frac{\Phi(\frac{t-b_2}{\sigma})\Phi(\frac{d_2-t}{\sigma})}{\Phi(\frac{t-b_1}{\sigma})\Phi(\frac{d_1-t}{\sigma})} = \lim_{t\to\infty} \frac{\phi(\frac{d_2-t}{\sigma})}{\phi(\frac{d_1-t}{\sigma})} = \lim_{t\to\infty} e^{(\frac{d_2+d_1-2t}{\sigma})(\frac{d_1-d_2}{\sigma})} = 0. \tag{3.16}$$

Now if $b_1 < b_2$ we note that $\Phi(\frac{t-b_1}{\sigma}) > \Phi(\frac{t-b_2}{\sigma})$ for any sufficiently large value of $t$. Applying this inequality to the limit we reduce back to the first case. Finally if $b_1 > b_2$ we have,

$$\lim_{t\to\infty} \frac{\Phi(\frac{t-b_2}{\sigma})\Phi(\frac{d_2-t}{\sigma})}{\Phi(\frac{t-b_1}{\sigma})\Phi(\frac{d_1-t}{\sigma})} = \lim_{t\to\infty} \frac{\Phi(\frac{t-b_2}{\sigma})}{\Phi(\frac{t-b_1}{\sigma})} \lim_{t\to\infty} \frac{\Phi(\frac{d_2-t}{\sigma})}{\Phi(\frac{d_1-t}{\sigma})} = 0. \tag{3.17}$$

Assume that $d^C_{max} > d^D_{max}$, and hence $d^C_{max}$ is larger then the death value of any point of $D$. Let $(b^C, d^C_{max})$ be a point of $C$. Then

$$\begin{aligned}\lim_{t\to\infty} \frac{G_D(t)}{G_C(t)} &= \lim_{t\to\infty} \frac{\sum_{(b,d)\in D} \Phi(\frac{t-b}{\sigma})\Phi(\frac{d-t}{\sigma})}{\sum_{(b,d)\in C} \Phi(\frac{t-b}{\sigma})\Phi(\frac{d-t}{\sigma})} \leq \lim_{t\to\infty} \frac{\sum_{(b,d)\in D} \Phi(\frac{t-b}{\sigma})\Phi(\frac{d-t}{\sigma})}{\Phi(\frac{t-b^C}{\sigma})\Phi(\frac{d^C_{max}-t}{\sigma})} \\ &= \sum_{(b,d)\in D} \lim_{t\to\infty} \frac{\Phi(\frac{t-b}{\sigma})\Phi(\frac{d-t}{\sigma})}{\Phi(\frac{t-b^C}{\sigma})\Phi(\frac{d^C_{max}-t}{\sigma})} = 0\end{aligned}$$

Hence, for any sufficiently large value of $t$, $G_C(t) > G_D(t)$. A similar argument applied when assuming distinct minimum birth values, one simply needs to look at the limit as $t$ approaches negative infinity instead.

□

Obtaining a result for general surfaces or curves may prove to be challenging. The next example shows that the injectivity of persistence surfaces (hence Gaussian persistence curves) cannot be generalized to arbitrary weight functions.

**Example 3.3.5.** Let $\kappa_D(b, d) = \frac{d-b}{L_D}$ where $L_d = \sum_{(b,d)\in D} d - b$ is the total lifespan of $D$. Take $C = \{(b, d)\}$ and $D = \{(b, d), (b, d)\}$. That is, $C$ is a diagram with a single point and

$D$ is a diagram with two points both at the same place as $C$. Then $D \neq C$ but

$$\begin{aligned}\rho_{C,\kappa_C}(x,y) &= \phi\left(\frac{x-b}{\sigma}\right)\phi\left(\frac{y-d}{\sigma}\right) \\ &= \frac{1}{2}\phi\left(\frac{x-b}{\sigma}\right)\phi\left(\frac{y-d}{\sigma}\right) + \frac{1}{2}\phi\left(\frac{x-b}{\sigma}\right)\phi\left(\frac{y-d}{\sigma}\right) = \rho_{D,\kappa_D}(x,y).\end{aligned}$$

Note that the weight function in the above example is a natural one to consider in practice as it produced a strong stability result in the original persistence curve setting and also performed well in computer experiments [12]. However, in this example at least the failure of injectivity of the persistence surface is clearly tied to the fact that different diagrams produce different weighting functions. We conjecture that this is the only way for injectivity to fail.

**Conjecture.** Let $\kappa\colon \mathbb{R}^2 \to \mathbb{R}^+$ be a weighting function and let $C$ and $D$ be persistence diagrams. Suppose that $\rho_{C,\kappa} = \rho_{D,\kappa}$. Then $C = D$.

Since $\rho_{C,\kappa_C} = \rho_{D,\kappa_D}$ implies that $G_{C,\kappa_C} = G_{D,\kappa_D}$, example 3.3.5 also shows that it is possible for distinct diagrams to produce the same Gaussian persistence curve. However, as with persistence surfaces we conjecture that this cannot happen for weighting functions which are independent of the diagrams.

**Conjecture.** Let $\kappa\colon \mathbb{R}^2 \to \mathbb{R}^+$ be a weighting function and let $C$ and $D$ be persistence diagrams. Suppose that $G_{C,\kappa} = G_{D,\kappa}$. Then $C = D$.

# References


[1] 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. *The Journal of Machine Learning Research*, 18(1):218–252, 2017.

[2] Alexandr Andoni, Assaf Naor, and Ofer Neiman. Snowflake universality of Wasserstein spaces. In *Annales Scientifiques de l'Ecole Normale Superieure*, volume 51, pages 657–700. Societe Mathematique de France, 2018.

[3] Nieves Atienza, Rocío González-Díaz, and M. Soriano-Trigueros. On the stability of persistent entropy and new summary functions for TDA. *CoRR*, abs/1803.08304, 2018.

[4] Greg Bell, Austin Lawson, Neil Pritchard, and Dan Yasaki. The space of persistence diagrams fails to have Yu's Property A. In *Topology Proceedings*, volume 58, pages 279–288, 2021.

[5] Eric Berry, Yen-Chi Chen, Jessi Cisewski-Kehe, and Brittany Terese Fasy. Functional summaries of persistence diagrams. *Journal of Applied and Computational Topology*, 4(2):211–262, 2020.

[6] Peter Bubenik. Statistical topological data analysis using persistence landscapes. *The Journal of Machine Learning Research*, 16(1):77–102, 2015.

[7] Peter Bubenik and Alexander Wagner. Embeddings of persistence diagrams into Hilbert spaces. *Journal of Applied and Computational Topology*, 4(3):339–351, 2020.

[8] Mathieu Carrière and Ulrich Bauer. On the metric distortion of embedding persistence diagrams into separable Hilbert spaces. *arXiv preprint arXiv:1806.06924*, 2018.

[9] Frédéric Chazal, Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, Aarti Singh, and Larry Wasserman. On the bootstrap for persistence diagrams and landscapes. *arXiv preprint arXiv:1311.0376*, 2013.

[10] Yu-Min Chung, Michael Hull, and Austin Lawson. Smooth summaries of persistence diagrams and texture classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 840–841, 2020.

[11] Yu-Min Chung, Michael Hull, Austin Lawson, and Neil Pritchard. Gaussian persistence curves. *arXiv preprint arXiv:2205.11353*, 2022.

[12] Yu-Min Chung and Austin Lawson. Persistence curves: A canonical framework for summarizing persistence diagrams. *Advances in Computational Mathematics*, 48(1):6, 2022.

[13] Vincent Divol and Théo Lacombe. Understanding the topology and the geometry of the space of persistence diagrams via optimal partial transport. *Journal of Applied and Computational Topology*, 5:1–53, 2021.

[14] Herbert Edelsbrunner, John Harer, et al. Persistent homology-a survey. *Contemporary mathematics*, 453:257–282, 2008.

[15] Herbert Edelsbrunner and John L Harer. *Computational topology: an introduction*. American Mathematical Society, 2022.

[16] Michelle Feng and Mason A Porter. Persistent homology of geospatial data: A case study with voting. *SIAM Review*, 63(1):67–99, 2021.

[17] Mikhael Gromov. Asymptotic invariants of infinite groups. Technical report, P00001028, 1992.

[18] Atish Mitra and Žiga Virk. The space of persistence diagrams on n points coarsely embeds into Hilbert space. *Proceedings of the American Mathematical Society*, 149(6):2693–2703, 2021.

[19] Piotr Nowak. Coarse embeddings of metric spaces into Banach spaces. *Proceedings of the American Mathematical Society*, 133(9):2589–2596, 2005.

[20] Jagdish K Patel and Campbell B Read. *Handbook of the normal distribution*, volume 150. CRC Press, 1996.

[21] Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

[22] John Roe. *Lectures on coarse geometry*. Number 31. American Mathematical Soc., 2003.

[23] John Roe et al. *Index theory, coarse geometry, and topology of manifolds*, volume 90. American Mathematical Soc., 1996.

[24] Lee M Seversky, Shelby Davis, and Matthew Berger. On time-series topological data analysis: New data and opportunities. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 59–67, 2016.

[25] Katharine Turner, Yuriy Mileyko, Sayan Mukherjee, and John Harer. Fréchet means for distributions of persistence diagrams. *Discrete & Computational Geometry*, 52(1):44–70, 2014.

[26] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer, 2009.

[27] Alexander Wagner. Nonembeddability of persistence diagrams with $p > 2$ wasserstein metric. *Proceedings of the American Mathematical Society*, 149(6):2673–2677, 2021.

[28] Guoliang Yu. The coarse Baum–Connes conjecture for spaces which admit a uniform embedding into Hilbert space. *Inventiones mathematicae*, 139(1):201–240, 2000.