

## MEDIANS OF POPULATIONS OF PERSISTENCE DIAGRAMS

KATHARINE TURNER

(communicated by Ralph Cohen)

### *Abstract*

Persistence diagrams are common objects in the field of Topological Data Analysis. They are topological summaries that capture both topological and geometric structure within data. Recently there has been a surge of interest in developing tools to statistically analyze populations of persistence diagrams, a process hampered by the complicated geometry of the space of persistence diagrams. In this paper we study the median of a set of diagrams, defined as the minimizer of an appropriate cost function analogous to the sum of distances used for samples of real numbers. We then characterize the local minima of this cost function and in doing so characterize the median. We also do some comparative analysis of the properties of the median and the mean.

### 1. Introduction

Topological data analysis (TDA) is a rapidly growing field that uses multi-scale topological features to find and analyze structure in data. An important use of TDA is as a preprocessing tool, providing topological summaries that may be more tractable than the raw information, and perhaps highlight geometric and topological features that are of particular interest. Examples of applications include the analysis of the shape of human jaws [14], plant root systems [3], shapes of calcanei bones of various primates [29] and retrieval of trademark symbols [8].

Persistence diagrams are a common topological summary statistic. Each diagram is a discrete summary of how the homology evolves over a nested sequence of topological spaces parameterized by some tuning parameter. Homology computes features such as the number of connected components, loops, tunnels, and void. Although persistent homology is defined via homology, the tuning parameter captures geometric information.

As part of the growing field of object oriented data analysis we wish to consider the structure of the space of persistence diagrams when performing statistical analysis. This is complicated by the geometry of the space of persistence diagrams — it is infinite in dimension and with no upper bound on curvature. At no point does it locally

---

Received September 20, 2016, revised August 22, 2018; published on December 11, 2019.

2010 Mathematics Subject Classification: 62G99, 52C35, 51F99, 55N99.

Key words and phrases: persistence diagram, median, mean, topological data analysis, object oriented data analysis.

Article available at <http://dx.doi.org/10.4310/HHA.2020.v22.n1.a15>

Copyright © 2019, International Press. Permission to copy for private use granted.

look Euclidean. The papers [22, 28] study the geometric and analytic properties of the space of persistence diagrams equipped with the metric analogous to the 2-Wasserstein distances of probability measures and to the  $L^2$  distances of functions on a discrete set. We denote this space  $(\mathcal{D}, d_2)$ . In [22] it was shown that it is possible to define the mean and variance of “nice” distributions via the Fréchet function. In [28] it was shown that  $(\mathcal{D}, d_2)$  was a non-negatively curved Alexandrov space and used this structure to characterize the mean of a population of persistence diagrams and to provide an algorithm to compute it.

After the mean, the median is next most common statistic to describe the center of a distribution. Furthermore, the median is often more robust than the mean. The purpose of this paper is to provide an analogous analysis for the space of the persistence diagrams under the corresponding metric for  $p = 1$  (instead of  $p = 2$ ) and to characterize the median of a population of persistence diagrams. We also are interested in comparing the properties of the mean and the median of populations of persistence diagrams.

In section 2 we establish geometric properties of the space of persistence diagrams, such as curvature and connected components, for a natural family of metrics  $\{(\mathcal{D}, d_p)\}$  that are analogous to the  $p$ -Wasserstein distances on the space of probability measures and to the  $L^p$  distances on the space of functions on a discrete set. Just as spaces of functions equipped with an  $L^2$  metric has nicer properties to those with an  $L^p$  metric with  $p \neq 2$ , we see that although  $(\mathcal{D}, d_2)$  is a non-negatively curved Alexandrov space, this does not hold for  $(\mathcal{D}, d_p)$  when  $p \neq 2$ .

The median of a population is the location that minimizes the average distance to each of the members of the population. In section 3 we study the median of a population of persistence diagrams. Unfortunately the proofs in [28] for characterizing and computing the mean require the extra Alexandrov space structure available in the case when  $p = 2$  (rather than  $p = 1$  which is the scenario for the median). This implies that we needed to develop new methods in order to characterize the median of a population of persistence diagrams. In the appendix we summarize how these new methods provide a new proof for previous results about the mean. Another aspect where the analysis here differs from that in [28] is that we consider persistence diagrams containing points with infinite persistence.

We can define the mean as the location that minimizes the average distance to the sample set. Unlike the mean, the median of a population of even size is not unique. For  $N$  odd, the median of a set of real numbers  $a_1, a_2, \dots, a_N$ , written in non-decreasing order, is  $a_{(N+1)/2}$ . If  $N$  is even, then every number in the interval  $[a_{N/2}, a_{(N+2)/2}]$  minimizes the average distance to the sample set and hence would be a valid choice as a median. To overcome this lack of uniqueness the general convention is to declare the median to be the midpoint of  $[a_{N/2}, a_{(N+2)/2}]$ . Analogous conventions could be applied for medians of an even number of persistence diagrams. However, the statements of theory and the clarity of exposition quickly becomes less clear. For this reason we will be restricting our attention to the case when the size of the populations is odd.

In section 4 we compare various properties of the median to the mean. We bound the number of off-diagonal points in the median compared to the mean, show the median is more robust than the mean, and prove that the mean is generically unique while the median is not.

**1.1. Related work**

Statistical analysis of persistence diagrams is an example of object oriented data analysis. This is an emerging area of statistics where the goal is to develop and apply statistical methods to objects such as functions, images, graphs or trees (e.g., [19, 21, 20, 25, 27, 30]) while respecting the structure of these objects.

There has been a growing movement of developing statistical methods to understand populations of persistence diagrams. This includes a significant body of work (for example, see [4, 7, 9, 10, 13]) which studies statistical methodology using the *bottleneck distance*, which is effectively the  $L^\infty$  distance within the space of persistence diagrams.

There has also been a parallel movement of statistical analysis in TDA by converting persistence diagrams into other functional summaries. These new topological summaries lie in spaces where it is easier to modify traditional statistical methods such as computing means or performing t-tests. However, this can be at the expense of making it harder to provide topological interpretations of the results. Examples of such functional summaries includes persistence landscapes [6] and persistent homology rank functions [26].

Other related work investigates the homology or persistent homology of random topological objects (for example, [1, 5, 15, 16, 32, 31]) including limit theorems and analysis of simulations.

**2. Geometry of the space of persistence diagrams**

**2.1. Background theory on persistent homology and persistence diagrams**

Here we will provide a very brief introduction to persistent homology. A more complete coverage can be found in [12]. Although homology can be computed over any ring, to compute persistent homology we need a field. This is usually  $\mathbb{F}_2$  for computational purposes.

To define persistent homology, we start with a nested sequence of topological spaces,

$$X_0 \subseteq X_1 \subseteq X_2 \subseteq \dots \subseteq X_n = X. \tag{2.1}$$

Often this sequence arises from the sublevel sets of a function,  $f: X \rightarrow \mathbb{R}$ , with  $X_i = f^{-1}(-\infty, t_i]$  for a sequence  $-\infty \leq t_0 \leq t_1 \leq \dots \leq t_n \leq \infty$ . The sequence induces linear maps on homology for any dimension r:

$$H_r(X_0) \rightarrow H_r(X_1) \rightarrow \dots \rightarrow H_r(X_n).$$

We are interested in when homology classes appear and disappear in this sequence. Let  $\phi_i^j: H_r(X_i) \rightarrow H_r(X_j)$  be the linear map on homology induced by the inclusion  $X_i \rightarrow X_j$ . Observe that if  $i < j < k$  then  $\phi_j^k \circ \phi_i^j = \phi_i^k$ . The homology class  $\gamma \in H_r(X_i)$  is said to be born at  $X_i$  if it is not in the image of  $\phi_{i-1}^i$ . This same class is said to die at  $X_j$  if its image in  $H_r(X_{j-1})$  is not in the image of  $\phi_{i-1}^{j-1}$ , but its image in  $H_r(X_j)$  is in the image of  $\phi_{i-1}^j$ . In the case that the spaces arose from the level sets of a function  $f$  as defined above, we define say that  $\gamma$  is born at time  $t_i$ , dies at time  $t_j$ , and its lifetime is  $t_j - t_i$ .

A discrete description of the persistent homology of (2.1) is the multiset of points in the extended plane where we include  $(t_i, t_j)$  with multiplicity the dimension of the set of homology classes born at  $t_i$  and dying at  $t_j$ . This is the information recorded as the ( $r$ th dimensional) persistence diagram.

Although we have here restricted our definition of persistent homology to that of a finite nested sequence, it can be defined more generally for nested parameterized families of topological spaces  $\{X_t : t \in \mathbb{R}\}$  with the condition that  $X_s \subseteq X_t$  whenever  $s \leq t$  (alongside some technical finiteness conditions). We can think of this as a filtered space with filtration parameter  $t$ . For “nice” filtered spaces the persistence diagram can be defined analogously to the finite case. For details see [11].

It is worth observing that although we are computing homology we are often capturing geometric information through the filtration parameter. For example, if we are considering a filtration of  $\mathbb{R}^2$  by the distance from a circle of radius  $R$  then the 1st dimensional persistent homology will have exactly one class born at 0 (the time the circle first appears) and dying at  $R$  (the time when the circle is filled in).

Before giving a technical definition of a persistence diagram we will need to introduce some notation. Let  $\mathbb{R}^{2+} = \{(a, b) \in \mathbb{R}^2 : a < b\}$ . This is the part of the plane above the diagonal. Let  $\Delta$  denote an abstract element representing the diagonal  $\{(x, x) : x \in \mathbb{R}\}$ .

Since we wish to also want to consider persistent homology classes of infinite duration we will also want to include lines at infinity. Let  $\mathcal{L}_{-\infty} := \{(-\infty, b) : b \in \mathbb{R}\}$  and  $\mathcal{L}_{\infty} := \{(a, \infty) : a \in \mathbb{R}\}$ . Persistence diagrams will be multiset of points in  $\mathcal{L}_{\infty} \cup \mathcal{L}_{-\infty} \cup \mathbb{R}^{2+} \cup \Delta$ . For tractability we will impose some finiteness conditions, namely that only finitely many classes have infinite lifetimes and that the sum of all the finite lifetimes is finite. This restriction is not onerous in applications where generally we have finite sized data as input.

**Definition 2.1.** A persistence diagram  $X$  is a multiset of  $\mathcal{L}_{\infty} \cup \mathcal{L}_{-\infty} \cup \mathbb{R}^{2+} \cup \Delta$  such that

- the number of elements in  $X|_{\mathcal{L}_{\infty}}$  and  $X|_{\mathcal{L}_{-\infty}}$  are finite
- $\sum_{(x_i, y_i) \in X|_{\mathbb{R}^{2+}}} (y_i - x_i) < \infty$
- there are countably infinite copies of  $\Delta$

## 2.2. Metrics on the space of persistence diagrams

In this paper we differ slightly from the historical definition of a persistence diagram. Instead of including every point along the diagonal of  $\mathbb{R}^2$  with infinite multiplicity we include countably many copies of the diagonal. This alternative description still has the same distances between different persistence diagrams but simplifies statements, arguments and calculations. This is because we do not need to specify which point on the diagonal is being used.

Let  $\mathcal{D}$  denote the space of all persistence diagrams. We will consider a family of metrics which are analogous to the  $p$ -Wasserstein distances on the space of probability measures and to the  $L^p$  distances on the space of functions on a discrete set.  $\mathbb{R}^{2+}$  inherits natural  $L^p$  distances from  $\mathbb{R}^2$ . For  $p \in [1, \infty)$  we have  $\|(a_1, b_1) - (a_2, b_2)\|_p^p = |a_1 - a_2|^p + |b_1 - b_2|^p$  and  $\|(a_1, b_1) - (a_2, b_2)\|_{\infty} = \max\{|a_1 - a_2|, |b_1 - b_2|\}$ .

Recall that  $\Delta$  represents the diagonal in  $\mathbb{R}^2$ . With a slight abuse of notation we write  $\|(a, b) - \Delta\|_p$  to denote the shortest  $L^p$  distance from  $(a, b)$  in to a point in the diagonal set in  $\mathbb{R}^2$ . Thus

$$\|(a, b) - \Delta\|_p = \inf_{t \in \mathbb{R}} \|(a, b) - (t, t)\|_p = 2^{\frac{1}{p}-1} |b - a|$$

for  $p < \infty$ , and  $\|(a, b) - \Delta\|_\infty = \inf_{t \in \mathbb{R}} \|(a, b) - (t, t)\|_\infty = |y - x|/2$ . Both  $\mathcal{L}_{-\infty}$  and  $\mathcal{L}_\infty$  inherit natural  $L^p$  distances from the  $L^p$  metric on  $\mathbb{R}$ ;  $\|(-\infty, b_1) - (-\infty, b_2)\|_p = |b_1 - b_2|$  and  $\|(a_1, \infty) - (a_2, \infty)\|_p = |a_1 - a_2|$ . We should also think of  $\mathcal{L}_{-\infty}$ ,  $\mathcal{L}_\infty$  and  $\mathbb{R}^{2+} \cup \Delta$  as three separate disjoint parts of a larger space.

Given persistence diagrams  $X$  and  $Y$  we can consider all the bijections from the set of off-diagonal points and copies of  $\Delta$  in  $X$ , to the set of off-diagonal points and copies of  $\Delta$  in  $Y$ . This set is non-empty as it contains the bijection which matches everything to a copy of  $\Delta$  in the other diagram. Each bijection provides a transport plan from  $X$  to  $Y$ . Analogous to the definition of Wasserstein distances, we will define our family of metrics in terms of the cost of most efficient transport plan.

For each  $p \in [1, \infty)$  define

$$d_p(X, Y) = \left( \inf_{\text{bijections } \phi: X \rightarrow Y} \sum_{x \in X} \|x - \phi(x)\|_p^p \right)^{1/p},$$

and  $d_\infty(X, Y) = \inf_{\text{bijections } \phi: X \rightarrow Y} \sup_{x \in X} \|x - \phi(x)\|_\infty$ .

These distances may be infinite — for example if  $X$  and  $Y$  contain a different number of points in  $\mathcal{L}_\infty$  then  $d_p(X, Y) = \infty$  for all  $p$ .

We will call a bijection between diagrams *optimal for  $d_p$*  if it achieves the infimum in the definition of  $d_p$  and this distance is finite. Figure 1 illustrates an example of an optimal bijection.

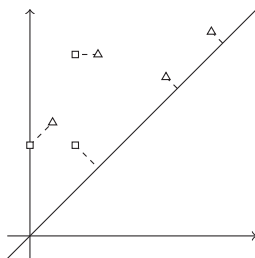


Figure 1: The dashed lines indicate an optimal bijection from the persistence diagram containing the square points (and copies of  $\Delta$ ) to the persistence diagram containing the triangle points (and copies of  $\Delta$ ).

Given the same pair of diagrams but different values of  $p$ , different bijections may be optimal. Furthermore, optimal bijections for a given  $p$  are not necessarily unique. For example, let  $X$  and  $Y$  to be diagrams containing pairs of opposite corners of a square located far from the diagonal. Because of symmetry, bijection the points vertically or horizontally involves the same cost. This example works for every  $p \in [1, \infty]$ . Other examples of non-uniqueness can involve  $\Delta$ ; it may be equally efficient to match two points to each other as it is to match both with a copy of  $\Delta$ .

In theory, for every pair  $p, q \in [1, \infty]$  one could construct a distance function of the form

$$\inf_{\phi: X \rightarrow Y} \left( \sum_{x \in X} \|x - \phi(x)\|_q^p \right)^{1/p},$$

with  $p$  and  $q$  different. Some of the computational topology literature uses a family of metrics  $d_{W_p}$  where  $p$  varies but  $q = \infty$  is fixed. The families  $\{d_p\}$  and  $\{d_{W_p}\}$  share many properties. The metrics  $d_p$  and  $d_{W_p}$  are bi-Lipschitz equivalent as for any  $x, y \in \mathbb{R}^2$  we have  $\|x - y\|_\infty \leq \|x - y\|_p \leq 2\|x - y\|_\infty$ , implying  $d_{W_p}(X, Y) \leq d_p(X, Y) \leq 2d_{W_p}(X, Y)$ . Any stability results for  $\{d_p\}$  or  $\{d_{W_p}\}$  would extend (with minor changes in constant) to stability results for the other.

We feel that the choice of setting  $q = p$  is cleaner in theory and in practice. The coordinates of the points within a persistence diagram have particular meanings; one is the birth time and one is the death time. They are often infinitesimally independent (even though not globally so). For example, if we have generated our persistence diagram from the distance function to a point cloud then each persistence class has its birth and death time (infinitesimally) determined by the location of two pairs of points which are often distinct. Whenever these pairs are distinct, moving any of these four points will change either the birth or the death but not both. The distinctness of the treatment of birth and death times as separate qualities may seem more philosophically pleasing to the reader in the setting of barcodes.

Another argument in favor of using  $p = q$  is computational power. The mean and the median are defined to be the minimizers of cost functions involving  $d_2$  and  $d_1$  respectively. Computationally the mean and median have far nicer characterizations, with easy algorithms to find them, when they are defined using the metrics where  $p = q$ . This is discussed later in section 3.

### 2.3. Geometry of the space of persistence diagrams

In this section we will describe the connected components of  $\mathcal{D}$  and show  $\mathcal{D}$  is a geodesic space, but first we need to introduce some notation. For each  $p < \infty$  we can define the distance between finite multisets  $A, B$  of points in  $\mathcal{L}_\infty$  as

$$d_p^{\mathcal{L}_\infty}(A, B)^p = \inf_{\text{bijections } \phi: A \rightarrow B} \sum_{a \in A} \|a - \phi(a)\|_p^p,$$

when  $|A| = |B|$  and infinity otherwise. We also set

$$d_\infty^{\mathcal{L}_\infty}(A, B) = \inf_{\text{bijections } \phi: A \rightarrow B} \sup_{a \in A} \|a - \phi(a)\|_\infty,$$

whenever  $|A| = |B|$  and infinity otherwise. We define  $d_p^{\mathcal{L}^{-\infty}}$  similarly. Let  $\mathcal{D}^{(k,l)}$  denote the space of persistence diagrams containing exactly  $k$  points in  $\mathcal{L}_{-\infty}$ , and exactly  $l$  points in  $\mathcal{L}_\infty$ .

**Lemma 2.2.** *Let  $p \in [1, \infty]$ . The connected components of  $(\mathcal{D}, d_p)$  are  $\mathcal{D}^{(k,l)}$ , with  $k, l$  non-negative integers. Let  $X, Y \in \mathcal{D}^{(k,l)}$ . For  $p < \infty$ ,*

$$\begin{aligned} d_p(X, Y)^p &= d_p(X|_{\mathbb{R}^2 + \cup \Delta}, Y|_{\mathbb{R}^2 + \cup \Delta})^p + d_p^{\mathcal{L}_\infty}(X|_{\mathcal{L}_\infty}, Y|_{\mathcal{L}_\infty})^p \\ &\quad + d_p^{\mathcal{L}^{-\infty}}(X|_{\mathcal{L}^{-\infty}}, Y|_{\mathcal{L}^{-\infty}})^p \end{aligned}$$

and

$$d_\infty(X, Y) = \max\{d_\infty(X|_{\mathbb{R}^{2+} \cup \Delta}, Y|_{\mathbb{R}^{2+} \cup \Delta}), d_\infty^{\mathcal{L}_\infty}(X|_{\mathcal{L}_\infty}, Y|_{\mathcal{L}_\infty}), d_\infty^{\mathcal{L}_{-\infty}}(X|_{\mathcal{L}_{-\infty}}, Y|_{\mathcal{L}_{-\infty}})\}.$$

The proof of this lemma follows from the observations that the connected components of  $\mathcal{L}_\infty \cup \mathcal{L}_{-\infty} \cup \mathbb{R}^{2+} \cup \Delta$  are  $\mathcal{L}_\infty$ ,  $\mathcal{L}_{-\infty}$  and  $\mathbb{R}^{2+} \cup \Delta$ , and that  $d_p(X|_{\mathbb{R}^{2+} \cup \Delta}, Y|_{\mathbb{R}^{2+} \cup \Delta})$  is always finite.

Let  $(\mathbb{X}, d)$  be a metric space. A curve  $\lambda: [0, 1] \rightarrow \mathbb{X}$  is called a *geodesic* if there exists constant  $C > 0$  such that  $d(\lambda(t_1), \lambda(t_2)) = C|t_1 - t_2|$  for all  $t_1, t_2 \in [0, 1]$  with  $|t_1 - t_2|$  sufficiently small.  $(\mathbb{X}, d)$  is called a *geodesic space* if for every  $x, y \in \mathbb{X}$  there exists a geodesic  $\lambda: [0, 1] \rightarrow \mathbb{X}$  such that  $\lambda(0) = x$  and  $\lambda(1) = y$ .

**Proposition 2.3.**  *$(\mathcal{D}, d_p)$  is a geodesic space for all  $p \in [1, \infty]$ .*

*Proof.* Fix  $p \in [1, \infty)$  and  $X, Y \in \mathcal{D}$  with  $d_p(X, Y) < \infty$ . We want to find a bijection  $\phi$  such that  $d_p(X, Y)^p = \sum_{x \in X} \|x - \phi(x)\|_p^p$  and from this bijection to construct a geodesic from  $X$  to  $Y$ .

Let  $\{\phi_i\}$  be a sequence of bijections such that  $\lim_{i \rightarrow \infty} \sum_{x \in X} \|x - \phi_i(x)\|_p^p = d_p(X, Y)^p$ . Fix some off-diagonal point  $\hat{x} \in X$ . The sequence  $\{\phi_i(\hat{x})\}$  must have a convergent subsequence  $\{\phi_{i_j}(\hat{x})\}$  which converges either to an off-diagonal point or to  $\Delta$ . This limit point must be in  $Y$  and we set  $\phi(\hat{x})$  to be this limit point. Observe our subsequence also satisfies  $\lim_{j \rightarrow \infty} \sum_{x \in X} \|x - \phi_{i_j}(x)\|_p^p = d_p(X, Y)^p$ .

We now replace our original sequence of bijections  $\{\phi_i\}$  with the subsequence  $\{\phi_{i_j}\}$ . In this manner we can determine a choice  $\phi(x)$  for each off-diagonal point  $x \in X$ . Similarly we can determine  $\phi^{-1}(y)$  for all the off-diagonal points  $y \in Y$ . Since we are always considering subsequences of previous subsequences we have consistency in our choices.

Since there are only countably many points off-the diagonal in the diagrams  $X$  and  $Y$  combined we can find an optimal bijection  $\phi: X \rightarrow Y$ . Let  $X_t$  be the diagram with off-diagonal points  $\{(1 - t)x + t\phi(x) : x \in X\}$  and set define the path  $\lambda: [0, 1] \rightarrow \mathcal{D}$  by  $\lambda(t) = X_t$ . By observation  $X_0 = X$ ,  $X_1 = Y$ , and  $\lambda$  is a geodesic.

The  $p = \infty$  case is similar. Let  $\{\phi_i\}$  be a sequence of bijections such that

$$\lim_{i \rightarrow \infty} \max_{x \in X} \|x - \phi_i(x)\|_\infty = d_\infty(X, Y)$$

and proceed as in the case  $p \in [1, \infty)$  to produce a bijection  $\phi$ , by assigning the values of  $\phi(x)$  and restricting to appropriate subsequences, such that  $d_\infty(X, Y) = \max_{x \in X} \|x - \phi(x)\|_\infty$ . This bijection will determine a geodesic by the same reasoning in the  $p \in [1, \infty)$  case. □

### 2.4. Curvature bounds on the space of persistence diagrams

In order to understand the space of persistence diagrams it is useful to analyze its curvature. Alexandrov spaces are geodesic spaces with curvature bounds. They come in two different forms; either their curvature is bounded from above (also known as *CAT* spaces) or their curvature is bounded from below. A bound on curvature in geodesic space  $(\mathbb{X}, d)$  is defined using comparison triangles. For each  $\kappa \in \mathbb{R}$  there is a model space  $M_\kappa$  with constant curvature  $\kappa$ . We compare triangles in  $\mathbb{X}$  to triangles in  $M_\kappa$ . Take three points  $x, y, z$ . If  $\kappa > 0$  we require  $d(x, y) + d(y, z) + d(z, x) \leq \sqrt{2\pi/\kappa}$ . These define a triangle  $\Delta(x, y, z)$  in  $\mathbb{X}$ . We can build a comparison triangle  $\Delta(\tilde{x}, \tilde{y}, \tilde{z})$

in the model space  $M_\kappa$  whose sides have the same length as the sides of  $\Delta(x, y, z)$ . The curvature of  $\mathbb{X}$  is bounded from below (above) by  $\kappa$  if, for every triangle  $\Delta(x, y, z)$  in  $\mathbb{X}$ , the distances between the points in  $\Delta(x, y, z)$  are less than or equal (respectively greater than or equal) the corresponding points in the comparison triangle  $\Delta(x', y', z')$  in  $M_\kappa$ . For more details see [18].

A  $CAT(k)$  space is a geodesic space whose curvature is bounded from above by  $k$ .  $CAT$ -spaces, in particular,  $CAT(0)$  spaces, have desirable properties. For example, the barycenter of any measure in a  $CAT(0)$  space is unique and in a  $CAT(k)$  space there is a length  $D_k$  such that balls of radius  $D_k$  are contractible. We first confirm that  $(\mathcal{D}, d_p)$  is not a  $CAT$ -space.

**Proposition 2.4.** *For all  $k > 0$  and  $p \in [1, \infty]$ ,  $(\mathcal{D}, d_p)$  is not in  $CAT(k)$ .*

*Proof.* If  $(\mathcal{D}, d_p)$  is a  $CAT(k)$  space then there is a constant  $K > 0$  such that for all pairs  $X, Y \in (\mathcal{D}, d_p)$  with  $d_p(X, Y)^2 < K$  there is a unique geodesic between them [18]. However, we can find pairs of diagrams  $X$  and  $Y$  which are arbitrarily close such that there are two distinct geodesics between them. One example is by taking  $X$  to be a diagram with two diagonally opposite corners of a square set far from the diagonal and  $Y$  the diagram with the other two corners. This is illustrated in Figure 2. The horizontal and vertical paths are equally optimal and we may choose the square to be as small as we wish.  $\square$

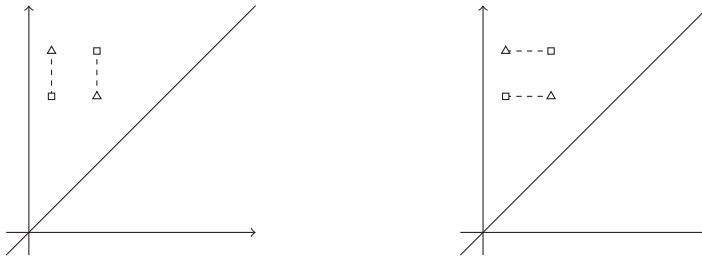


Figure 2: Two different optimal bijections between the triangle and the square diagrams.

Alexandrov spaces have nice properties. For example, for Alexandrov spaces we can define tangent cones (analogous to tangent planes), and exponential maps. From [24] we know that a geodesic space  $(\mathbb{X}, d)$  is an Alexandrov space with curvature bounded from below by zero if, and only if, for every geodesic  $\gamma: [0, 1] \rightarrow \mathbb{X}$  from  $X$  to  $Y$ , and every  $Z \in \mathbb{X}$  we have

$$d(Z, \gamma(t))^2 \geq td(Z, Y)^2 + (1 - t)d(Z, X)^2 - t(1 - t)d(X, Y)^2. \tag{2.2}$$

Using this characterization [28] showed that  $(\mathcal{D}, d_2)$  is an Alexandrov space with curvature bounded below by zero.

In contrast, using different counterexamples for  $p \in [1, 2)$  and for  $p \in (2, \infty]$  we can show that this curvature bound does not hold when  $p \neq 2$ . These counterexamples are illustrated in Figure 3.

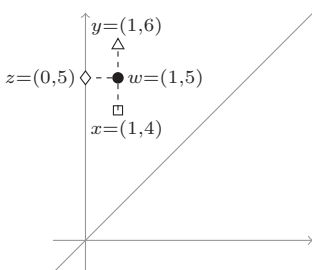
Let  $p \in [1, 2)$  and  $t = 1/2$ . Let  $X, Y$  and  $Z$  be a persistence diagram with only one off-diagonal point each in them at  $x = (1, 4), y = (1, 6)$  and  $z = (0, 5)$  respectively. The



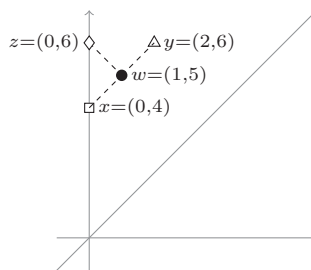
midway point between  $X$  and  $Y$  (playing the role of  $\gamma(1/2)$ ) is the diagram with the point  $w = (1, 5)$ . This set up is shown in Figure 3a. We can calculate  $d_p(Z, \gamma(1/2))^p = \|z - w\|_p^p = 1$ ,  $d_p(Z, X)^p = \|z - x\|_p^p = 2$ ,  $d_p(Z, Y)^p = \|z - y\|_p^p = 2$  and  $d_p(X, Y)^p = \|x - y\|_p^p = 2^p$ . Together they imply

$$\frac{1}{2}d_p(Z, Y)^2 + \frac{1}{2}d_p(Z, X)^2 - \frac{1}{4}d_p(X, Y)^2 = 2^{2/p} - 1.$$

But  $2^{2/p} - 1 > 1 = d_p(Z, \gamma(1/2))^2$  when  $1 \leq p < 2$ . This contradicts equation (2.2) and hence  $(\mathcal{D}, d_p)$  is not an Alexandrov space with curvature bounded below by zero.



(a) Counterexample for  $p \in [1, 2)$



(b) Counterexample for  $p \in (2, \infty]$

Figure 3: In both (a) and (b) the geodesic  $\gamma$  from  $X = \{x\}$  to  $Y = \{y\}$  has midpoint  $\gamma(1/2) = \{w\}$ . We consider the distances to  $Z = \{z\}$ . Here we are describing each persistence diagram by its list of off-diagonal points. In both examples  $d_p(Z, \gamma(1/2))^2 < \frac{1}{2}d_p(Z, Y)^2 + \frac{1}{2}d_p(Z, X)^2 - \frac{1}{4}d_p(X, Y)^2$  contradicting (2.2). Thus  $(\mathcal{D}, d_p)$  is not an Alexandrov space with curvature bounded below by zero when  $p \in [1, 2) \cup (2, \infty]$ .

Now let  $p \in (2, \infty)$  and  $t = 1/2$ . Let  $X, Y$  and  $Z$  be a persistence diagram with only one off-diagonal point each in them at  $x = (0, 4), y = (2, 6)$  and  $z = (0, 6)$  respectively. The midway point between  $X$  and  $Y$  (playing the role of  $\gamma(1/2)$ ) is the diagram with the point  $w = (1, 5)$ . This set up is shown in Figure 3b. Here  $d_p(Z, \gamma(1/2))^p = \|z - w\|_p^p = 2$ ,  $d_p(Z, X)^p = \|z - x\|_p^p = 2^p$ ,  $d_p(Z, Y)^p = \|z - y\|_p^p = 2^p$ , and  $d_p(X, Y)^p = \|x - y\|_p^p = 2^{p+1}$ .

Together they imply

$$\frac{1}{2}d_p(Z, Y)^2 + \frac{1}{2}d_p(Z, X)^2 - \frac{1}{4}d_p(X, Y)^2 = 2^2 - 2^{2/p}.$$

But  $2^2 - 2^{2/p} > 2^{2/p} = d_p(Z, \gamma(1/2))^2$  when  $p > 2$ . This contradicts equation (2.2) and hence  $(\mathcal{D}, d_p)$  is not an Alexandrov space with curvature bounded below by zero.

With this same set up we can calculate  $d_\infty(Z, \gamma(1/2)) = \|z - w\|_\infty = 1$ ,  $d_\infty(Z, X) = \|z - x\|_\infty = 2$ ,  $d_\infty(Z, Y) = \|z - y\|_\infty = 2$ , and  $d_\infty(X, Y) = \|x - y\|_\infty = 2$ . Hence

$$\frac{1}{2}d_\infty(Z, Y)^2 + \frac{1}{2}d_\infty(Z, X)^2 - \frac{1}{4}d_\infty(X, Y)^2 = 3 > 1 = d_\infty(Z, \gamma(1/2))^2.$$

This contradicts equation (2.2) and hence  $(\mathcal{D}, d_\infty)$  is not an Alexandrov space with curvature bounded below by zero.

### 3. The median of a population of persistence diagrams

Measures of central tendency (such as the mean and the median), or their corresponding measures of variability or dispersion (the variance and the average cost respectively) are common statistics used to describe distributions. Central tendencies are solutions for optimizing different cost functions which are based on  $p$ -Wasserstein metrics. The median corresponds to the case where  $p = 1$ .

The median of a set of real numbers  $a_1, a_2, \dots, a_N$ , written in non-decreasing order, is the number  $m$  which minimizes the mean absolute deviation function  $F_1^{\mathbb{R}}(x) = \frac{1}{N} \sum_{i=1}^N |a_i - x|$ . The average cost of moving a point in the sample data to the median is  $F_1(m)$ . For  $N$  odd the median is unique and equals  $a_{(N+1)/2}$ . If  $N$  even then every number in the interval  $[a_{N/2}, a_{(N+2)/2}]$  will minimize  $F_1^{\mathbb{R}}$  and would be a valid choice as a median. To overcome this lack of uniqueness the general convention is to declare the median to be the midpoint of  $[a_{N/2}, a_{(N+2)/2}]$ .

More generally the median of a population  $\{x_1, x_2, \dots, x_N\}$  within a connected metric space  $\mathbb{X}$  is the minimizer of the function  $F_1(y) = \frac{1}{N} \sum_{i=1}^N d(x_i, y)$  where  $d$  is an appropriate metric on  $\mathbb{X}$ .

To define the median of a population of persistence diagrams we need to fix a metric on the space of persistence diagrams. Unlike on the real line there are multiple reasonable options as explored in section 2.2. In this paper we argue that a well-motivated metric is  $d_1$ . This is for two main reasons. Firstly, the coordinates each have separate meanings and are infinitesimally independent. Thus for taking the median the birth should be the median of the relevant births, the deaths the median of the relevant deaths. Secondly, the computations become significantly easier. If we were to use  $d_2$  from section 2.2 then would need to compute, at some stage, the geometric median of a set of points in the plane. This is a problem as it was shown in [2] that in general there is no exact algorithm to find the geometric median of a set of  $k$  points in the plane.

The *median* is the persistence diagram  $m$  which minimizes the cost function  $F_1(Y) = \frac{1}{N} \sum_{i=1}^N d_1(X_i, Y)$ . The *total cost* is  $NF_1(m)$  and the *average cost* is  $F_1(m)$ .

It is worth observing that the median is only defined for populations  $\{X_1, \dots, X_N\}$  that lie within the same connected component of  $\mathcal{D}$  as otherwise for every  $Y$  there is some  $X_i$  such that  $d_1(X_i, Y) = \infty$ .

**Proposition 3.1.** *Fix  $k, l$  non-negative integers. Let  $X = \{X_1, X_2, \dots, X_N\}$  be a population of persistence diagrams in  $\mathcal{D}^{(k,l)}$ . For each  $i$  let  $X_i^{\mathbb{R}^{2+}}$  denote  $X_i|_{\mathbb{R}^{2+} \cup \Delta}$ ,  $X_i^\infty$  denote  $X_i|_{\mathcal{L}^\infty}$  and  $X_i^{-\infty}$  denote  $X_i|_{\mathcal{L}^{-\infty}}$ .*

*The diagram  $Y$  is a median of  $X$  if and only if  $Y = Y^{\mathbb{R}^{2+}} \cup Y^\infty \cup Y^{-\infty}$  where  $Y^{\mathbb{R}^{2+}}$  is a median of  $\{X_1^{\mathbb{R}^{2+}}, X_2^{\mathbb{R}^{2+}}, \dots, X_N^{\mathbb{R}^{2+}}\}$ ,  $Y^\infty$  is a median of  $\{X_1^\infty, X_2^\infty, \dots, X_N^\infty\}$  and  $Y^{-\infty}$  is a median of  $\{X_1^{-\infty}, X_2^{-\infty}, \dots, X_N^{-\infty}\}$ .*

*Proof.* Let  $Z \in \mathcal{D}^{(k,l)}$ . From Lemma 2.2 we know that

$$d_1(Z, X_i) = d_1(Z|_{\mathbb{R}^{2+}}, X_i^{\mathbb{R}^{2+}}) + d_1^{\mathcal{L}^\infty}(Z|_{\mathcal{L}^\infty}, X_i^\infty) + d_1^{\mathcal{L}^{-\infty}}(Z|_{\mathcal{L}^{-\infty}}, X_i^{-\infty})$$

and hence we can write  $F_1$  as a sum of three independent sums

$$F_1(Z) = \frac{1}{N} \sum_{i=1}^N d_1(Z|_{\mathbb{R}^{2+}}, X_i^{\mathbb{R}^{2+}}) + \frac{1}{N} \sum_{i=1}^N d_1^{\mathcal{L}^\infty}(Z|_{\mathcal{L}^\infty}, X_i^\infty) + \frac{1}{N} \sum_{i=1}^N d_1^{\mathcal{L}^{-\infty}}(Z|_{\mathcal{L}^{-\infty}}, X_i^{-\infty}).$$

If  $Y$  is a median then it must minimize each of these sums. □

In the remainder of this section we will characterize the median of a population of multisets in  $\mathcal{L}_\infty$  or in  $\mathcal{L}_{-\infty}$ . The next section will address the more complicated characterization of medians of populations in  $\mathcal{D}^{(0,0)}$ .

**Lemma 3.2.** *Fix  $N$  odd. Let  $A_1, A_2, \dots, A_N$  be each multisets of exactly  $k$  real numbers. Label the elements of each  $A_i$  so that  $A_i = \{a_{i,1}, a_{i,2}, \dots, a_{i,k}\}$  with  $a_{i,1} \leq a_{i,2} \leq \dots \leq a_{i,k}$ . Set  $B = \{b_1, b_2, \dots, b_k\}$  where  $b_j$  is the median of  $\{a_{1,j}, \dots, a_{N,j}\}$ . Then  $B$  is the unique multiset of  $k$  real numbers that minimizes*

$$f_1: Y \mapsto \sum_{i=1}^N \left( \inf_{\phi: A_i \rightarrow Y, \phi \text{ bijection}} \sum_{a \in A_i} |a - \phi(a)| \right).$$

*Proof.* The key to this proof is that for  $X = \{x_1, x_2, \dots, x_k\}$  and  $Y = \{y_1, y_2, \dots, y_k\}$  (each written in non-decreasing order) we have

$$\phi: X \rightarrow Y, \phi \text{ bijection} \sum_{j=1}^m |x_j - \phi(x_j)| = \sum_{j=1}^m |x_j - y_j|.$$

We are not claiming that  $\phi: x_i \mapsto y_i$  is the unique optimal transport from  $X$  to  $Y$  (this is not always true) but just that it achieves this optimality.

Suppose the multiset  $Y = \{y_1, y_2, \dots, y_k\}$  (written in non-decreasing order) minimizes  $f_1$ . The observation above implies

$$f_1(Y) = \sum_{i=1}^N \sum_{j=1}^k |a_{i,j} - y_j| = \sum_{j=1}^k \left( \sum_{i=1}^N |a_{i,j} - y_j| \right).$$

For each  $j$  let  $b_j$  be the median of  $\{a_{i,j}\}$ . Since  $N$  is odd,  $\sum_{i=1}^N |a_{i,j} - b_j| \leq \sum_{i=1}^N |a_{i,j} - y_j|$  with equality if and only if  $b_j = y_j$ . Since  $B$  minimizes  $f_1$  we conclude that  $y_j$  is the median of  $\{a_{i,j}\}$  □

If  $N$  is even, then the median is not unique, even if we use the midpoint convention for populations of real values. For example if  $A_1 = \{0, 2\}$  and  $A_2 = \{6, 12\}$ , the two medians would be  $\{\text{midpt}[0, 6] = 3, \text{midpt}[2, 12] = 8\}$  and  $\{\text{midpt}[0, 12] = 6, \text{midpt}[2, 6] = 4\}$ .

### 3.1. The mean and median of multisets of points in the plane and copies of the diagonal

We are splitting up our analysis into the different regions  $\mathbb{R}^{2+} \cup \Delta$ ,  $\mathcal{L}_\infty$  or  $\mathcal{L}_{-\infty}$ . This is because these are the disconnected components and all geodesics will keep the points in the persistence diagrams within these different regions separate. Thus in this section we will focus on the points in  $\mathbb{R}^{2+} \cup \Delta$ , that is those corresponding to persistent homology classes with finite lifetimes.

Before considering the problem of populations of multisets in  $\mathbb{R}^{2+} \cup \Delta$  we will first investigate the simpler problem of populations of singletons in  $\mathbb{R}^{2+} \cup \Delta$ . Here we are using the definition of median as the minimizer of the sum of  $d_1$  distances. Given a population  $S$  set  $f_S: \mathbb{R}^{2+} \rightarrow \mathbb{R}$  by  $f_S(z) = \frac{1}{N} \sum_{w \in S} \|z - w\|_1$ . Thus  $y$  is a median of  $S$  implies it is a minimizer of  $f_S$ .

Within the proposition the candidate minimizer is the point whose  $x$ - and  $y$ -coordinates are each the median of sets of numbers constructed from the  $x$ - and  $y$ -coordinates of the points in  $S$ , and from copies of the diagonal. The idea is that whenever a point  $(x, y)$  is matched with the diagonal it is effectively matching it with a point  $(t, t)$  with  $x \leq t \leq y$ . For calculating a median of the  $x$  coordinates with  $|S|$  larger than the number of the copies of the diagonal, a contribution of  $t \in [x, y]$  and a contribution of  $\infty$  will have the same effect. Similarly, for calculating the median of the  $y$ -coordinates, a contribution of  $t \in [x, y]$  will have the same effect as that of  $-\infty$ . This means that from the purposes of calculation, we can use  $\pm\infty$  in our lists of coordinates in the proposition below.

**Proposition 3.3.** *Fix  $N$  odd and suppose  $k > N/2$ . Let  $S = \{(a_1, b_1), (a_2, b_2), \dots, (a_k, b_k)\}$  and  $N - k$  copies of  $\Delta$  (where the  $(a_i, b_i) \in \mathbb{R}^{2+}$ ). Define  $f = f_S|_{\mathbb{R}^{2+}}$ , that is*

$$f((x, y)) = \sum_{i=1}^k \|(x, y) - (a_i, b_i)\|_1 + \sum_{i=k+1}^N \|(x, y) - \Delta\|_1.$$

Let  $(\tilde{x}, \tilde{y})$  be the point in  $\mathbb{R}^2$  where  $\tilde{x}$  is the median of  $\{a_1, a_2, \dots, a_k\}$  with  $N - k$  copies of  $\infty$  and  $\tilde{y}$  is the median of  $\{b_1, b_2, \dots, b_k\}$  with  $N - k$  copies of  $-\infty$ . If  $\tilde{x} < \tilde{y}$  then  $(\tilde{x}, \tilde{y})$  is the point in  $\mathbb{R}^{2+}$  which minimizes  $f$ . If  $\tilde{x} \geq \tilde{y}$  then  $f((x, y)) > \sum_{i=1}^k \|\Delta - (a_i, b_i)\|_1$  for all  $(x, y) \in \mathbb{R}^{2+}$ .

*Proof.* Since  $k > N/2$  we know that  $\tilde{x}$  and  $\tilde{y}$  are finite. Suppose that  $\tilde{x} < \tilde{y}$ . We want to show  $(\tilde{x}, \tilde{y})$  is the minimum of  $f$ . Since  $f$  is a convex function over  $\mathbb{R}^{2+}$  it is sufficient to show  $(\tilde{x}, \tilde{y})$  is a local minimum.

Consider pairs  $(u, v)$  such that  $|u| < \min_{a_i \neq \tilde{x}} |\tilde{x} - a_i|$ ,  $|v| < \min_{b_i \neq \tilde{y}} |\tilde{y} - b_i|$ , and  $|u| + |v| < \|(\tilde{x}, \tilde{y}) - \Delta\|_1$ . This is true for sufficiently small  $u$  and  $v$ . For such  $(u, v)$  we have

$$\begin{aligned} & \sum_{i=1}^k \|(\tilde{x} + u, \tilde{y} + v) - (a_i, b_i)\|_1 - \sum_{i=1}^k \|(\tilde{x}, \tilde{y}) - (a_i, b_i)\|_1 \\ &= |\{i : a_i < \tilde{x}\}|u + |\{i : a_i > \tilde{x}\}|(-u) + |\{i : a_i = \tilde{x}\}| |u| \\ & \quad + |\{i : b_i < \tilde{y}\}|v + |\{i : b_i > \tilde{y}\}|(-v) + |\{i : b_i = \tilde{y}\}| |v| \end{aligned}$$

and

$$\|(\tilde{x} + u, \tilde{y} + v) - \Delta\|_1 - \|(\tilde{x}, \tilde{y}) - \Delta\|_1 = ((\tilde{y} + v) - (\tilde{x} + u)) - (\tilde{y} - \tilde{x}) = v - u.$$

Together these imply that

$$\begin{aligned} & f((\tilde{x} + u, \tilde{y} + v)) - f((\tilde{x}, \tilde{y})) \\ &= |\{i : a_i < \tilde{x}\}|u + |\{i : a_i > \tilde{x}\}|(-u) + |\{i : a_i = \tilde{x}\}| |u| + (N - k)(-u) \\ & \quad + |\{i : b_i < \tilde{y}\}|v + |\{i : b_i > \tilde{y}\}|(-v) + |\{i : b_i = \tilde{y}\}| |v| + (N - k)v. \end{aligned}$$

Since  $\tilde{x}$  is the median of  $\{a_1, a_2, \dots, a_k\}$  with  $N - k$  copies of  $\infty$  we know that

$$\left| (|\{i : a_i > \tilde{x}\}| + (N - k)) - |\{i : a_i < \tilde{x}\}| \right| < |\{i : a_i = \tilde{x}\}|.$$

This implies that if  $u \neq 0$  then

$$|\{i : a_i < \tilde{x}\}|u + |\{i : a_i > \tilde{x}\}|(-u) + |\{i : a_i = \tilde{x}\}| |u| + (N - k)(-u) > 0.$$

Similarly if  $v \neq 0$  then

$$|\{i : b_i < \tilde{y}\}|v + |\{i : b_i > \tilde{y}\}|(-v) + |\{i : b_i = \tilde{y}\}||v| + (N - k)v > 0.$$

Thus  $f((\tilde{x} + u, \tilde{y} + v)) > f((\tilde{x}, \tilde{y}))$  for  $(\tilde{x} + u, \tilde{y} + v)$  sufficiently near, but not equal to,  $(\tilde{x}, \tilde{y})$ , and that  $(\tilde{x}, \tilde{y})$  is a local minimum. The convexity of  $f$  further implies that  $(\tilde{x}, \tilde{y})$  is the global minimum of  $f$  over the domain  $\mathbb{R}^{2+}$ . Remember we are not including the diagonal as a candidate of locations for the minimum here as  $f$  is a function over  $\mathbb{R}^{2+}$ .

Now suppose that  $(\tilde{x}, \tilde{y})$  lies on or below the diagonal. Let  $(x, y) \in \mathbb{R}^{2+}$ . Then either  $x < \tilde{x}$  or  $y > \tilde{y}$ . Suppose that  $x < \tilde{x}$ . Let  $x' \in (x, \tilde{x})$  with  $(x', y) \in \mathbb{R}^{2+}$ . Then

$$f((x, y)) - f((x', y)) = \sum_{i=1}^k (|x - a_i| - |x' - a_i|) + \sum_{i=k+1}^N (x' - x) > 0$$

as  $|x - a_i| - |x' - a_i| = x' - x$  whenever  $a_i \geq \tilde{x}$  and

$$|\{i : a_i \geq \tilde{x}\}| + (N - k) > |\{i : a_i < \tilde{x}\}|$$

as  $\tilde{x}$  is the median of the  $a_i$  and  $N - k$  copies of  $\infty$ .

A similar argument shows that if  $y > \tilde{y}$  and  $y' \in (\tilde{y}, y)$  with  $(x, y') \in \mathbb{R}^{2+}$  then  $f((x, y)) > f((x, y'))$ . Thus  $f$  decreases as  $(x, y)$  travels towards  $(\tilde{x}, \tilde{y})$  while staying within  $\mathbb{R}^{2+}$ . By considering limits there is some  $t$  such that

$$f((x, y)) > \sum_{i=1}^k \|(t, t) - (a_i, b_i)\|_1 + (N - k)\|(t, t) - \Delta\|_1.$$

The observation that  $\|(t, t) - (a_i, b_i)\|_1 \geq \|\Delta - (a_i, b_i)\|_1$  for all  $i, t$  completes the proof.  $\square$

**Lemma 3.4.** *If  $k < N/2$  then*

$$\sum_{i=1}^k \|(x, y) - (a_i, b_i)\|_1 + \sum_{i=k+1}^N \|(x, y) - \Delta\|_1 > \sum_{i=1}^k \|\Delta - (a_i, b_i)\|_1$$

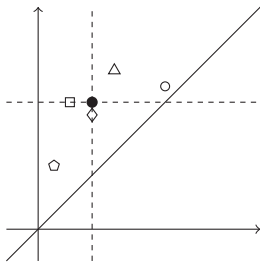
for every point  $(x, y) \in \mathbb{R}^{2+}$ .

*Proof.*  $\sum_{i=k+1}^N \|(x, y) - \Delta\|_1 > \sum_{i=1}^k \|(x, y) - \Delta\|_1$  as  $k < N/2$ . Then use the triangle inequality;  $\|(x, y) - (a_i, b_i)\|_1 + \|(x, y) - \Delta\|_1 \geq \|\Delta - (a_i, b_i)\|_1$ .  $\square$

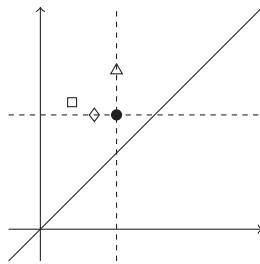
Using Proposition 3.3 and Lemma 3.4 we can characterize the median of a odd sized population of points in  $\mathbb{R}^{2+} \cup \Delta$ .

**Corollary 3.5.** *Let  $N$  be odd and  $S$  the population  $\{(a_1, b_1), \dots, (a_k, b_k), \Delta, \dots, \Delta\}$  containing  $N - k$  copies of  $\Delta$ . Let  $\tilde{x}$  be the median of  $\{a_1, a_2, \dots, a_k\}$  with  $N - k$  copies of  $\infty$  and let  $\tilde{y}$  be the median of  $\{b_1, b_2, \dots, b_k\}$  with  $N - k$  copies of  $-\infty$ . If  $(\tilde{x}, \tilde{y})$  lies above the diagonal then the median of  $S$  is either  $(\tilde{x}, \tilde{y})$  or  $\Delta$  (depending on whether  $f((\tilde{x}, \tilde{y}))$  or  $f(\Delta)$  is smaller). If  $(\tilde{x}, \tilde{y})$  lies on or below the diagonal (or is  $(\infty, -\infty)$  which philosophically lies below the diagonal) then  $\Delta$  is the median of  $S$ .*

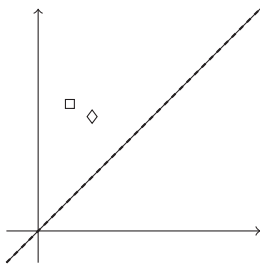
The median of  $S$  is at most two locations as the median of an odd number of (extended) real numbers is unique. In contrast, if  $N$  is even then we would have the same uniqueness issues as for sets of real numbers. For example, given the set



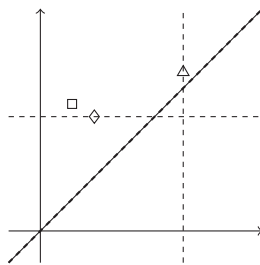
(a) The median of the square, triangle and diamond, circle and pentagon points. The order of the  $x$  coordinates are {pentagon, square, diamond, triangle, circle} and hence the median is that of the diamond. The order of the  $y$  coordinates are {pentagon, diamond, square, circle, triangle} and hence the  $y$  coordinates of the median is that of the square.



(b) The median of the square, triangle and diamond points alongside two copies of the diagonal. The order of the  $x$  coordinates are {square, diamond, triangle,  $\infty, \infty$ } and hence the median is that of the triangle. The order of the  $y$  coordinates are  $\{-\infty, -\infty, \text{diamond, square, triangle}\}$  and hence the  $y$  coordinates of the median is that of the diamond.



(c) The median of the square and diamond points alongside three copies of the diagonal. The order of the  $x$  coordinates are {square, diamond,  $\infty, \infty, \infty$ } and hence the median is " $\infty$ ". The order of the  $y$  coordinates are  $\{-\infty, -\infty, -\infty, \text{diamond, square}\}$  and hence the median is  $-\infty$ . This implies that the median is a copy of the diagonal.



(d) The median of the square and diamond points alongside three copies of the diagonal. The order of the  $x$  coordinates are {square, diamond, triangle,  $\infty, \infty$ } and hence the candidate median has  $x$  coordinate of the triangle. The order of the  $y$  coordinates are  $\{-\infty, -\infty, \text{diamond, square, triangle}\}$  and hence the candidate median has  $y$  coordinate of the diamond. However, this candidate lies below the diagonal and hence the median is a copy of the diagonal.

$X = \{2, 3, 4, 10\}$  every point in the interval  $[3, 4]$  will minimize the sums distances of points to  $X$ . However, by convention, we generally say that 3.5 is the median of  $X$ . When considering the median of an even number of points in  $\mathbb{R}^{2+} \cup \Delta$ , instead of a unique point in  $\mathbb{R}^{2+} \cup \Delta$  we have a rectangle of options, in which every point would minimize  $f_S$ . We could in theory adopt an analogous convention by using the barycenter of this rectangle. This may be application dependent. However, for clarity of exposition, and to make the statements of theorems much cleaner, we will be restricting our attention to  $N$  odd.

### 3.2. Characterizing the median(s) of sets of diagrams

In [28] there is a complete characterization of the local minima of  $F_2$  when the observations are finitely many persistence diagrams each with only finitely many off-diagonal points. However, the proof used the Alexandrov space structure of  $(\mathcal{D}, d_2)$  and hence we cannot adapt it to characterize the local minima of  $F_1$ . Here we use an alternative approach to prove analogous necessary and sufficient conditions for a persistence diagram to be a local minimum of  $F_1$ .

This characterization of local minima of  $F_2$  was rephrased in [23] in terms of selections, groupings and optimal pairings. Here we will characterize the local minima of  $F_1$  with this same terminology. Given a set of diagrams  $X_1, \dots, X_N$ , a *selection* is a choice of one point from each diagram (where that point can be  $\Delta$ ). A *grouping* is a set of selections so that every off-diagonal point of every diagram is part of exactly one selection. Our notation will be as follows. If  $S$  is a selection then let  $m_S$  be the median of that selection (chosen to be the off-diagonal point if not unique). A grouping  $G$  of  $X_1, \dots, X_N$  is the set of selections  $G = \{S_j\}$ . Let  $m(G)$  be the persistence diagram which contains  $\{m_{S_j} : S_j \in G\}$ . Each grouping  $G$  produces a candidate  $m(G)$  for the median. We will show that any median of  $X_1, \dots, X_N$  must be  $m(G)$  for some grouping  $G$  of  $X_1, \dots, X_N$ .

We will consider two groupings as equivalent if they only differ by selections containing only copies of the diagonal. Note that equivalent groupings produce the same persistence diagram as their median candidate. This implies that in theorems and algorithms we can restrict to groupings where each selection contains at least one off-diagonal point.

Since any minimum is also a local minimum, we will focus now on characterizing the local minima of  $F_1$ . We will first show that if  $Y$  is the median of  $\{X_1, \dots, X_N\}$  then  $Y = m(G)$  whenever  $G$  is an appropriate grouping constructed using optimal bijections  $\phi_i : Y \rightarrow X_i$ .

**Theorem 3.6.** *Let  $X_1, \dots, X_N$  be persistence diagrams in  $\mathcal{D}^{(0,0)}$ . Let  $Y = \{y_j\} \in \mathcal{D}^{(0,0)}$ . For each  $i$  let  $\phi_i : Y \rightarrow X_i$  be an optimal bijection between  $Y$  and  $X_i$  using the distance function  $d_1$ . For each  $y \in Y$  we have a selection  $\{\phi_i(y)\}$  (to make this well defined we think of the copies of  $\Delta$  when  $\phi_i^{-1}(x_j) = \Delta$  as each disjoint). Let  $G$  be the grouping  $\{\{\phi_i(y)\} : y \in Y\}$ .*

*If  $Y$  is a local minimum of  $F_1 : Z \mapsto \frac{1}{N} \sum_{i=1}^N d_1(X_i, Z)$  then  $Y = m(G)$ .*

*Proof.* Suppose that  $Y \neq m(G)$  and thus  $y \neq m_{\{\phi_i(y)\}}$  for some  $y \in Y$ . We need to split into cases depending on whether or not  $m_{\{\phi_i(y)\}}$  is the diagonal.

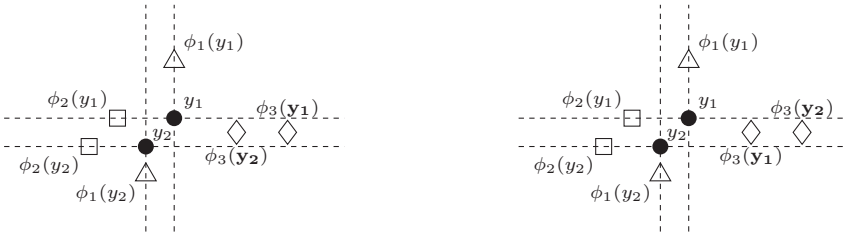
If  $y = \Delta$  then  $\{\phi_i(y)\}$  contains at most one off-diagonal point. By Lemma 3.4 we know that  $m_{\{\phi_i(y)\}} = \Delta$ .

Suppose now that  $y \neq \Delta$  and that more than half of  $\{\phi_i(y)\}$  are copies of the diagonal. As  $z$  moves from  $y$  to the closest point on the diagonal  $\sum_{\{i:\phi_i(y) \neq \Delta\}} \|z - \phi_i(y)\|_1$  increases less than  $\sum_{\{i:\phi_i(y) = \Delta\}} \|z - \Delta\|_1$  decreases and hence  $\sum_i \|z - \phi_i(y)\|_1$  must be decreasing. Let  $Z = \{z\} \cup Y \setminus y$ .  $F_1(Z)$  decreases as  $z$  moves from  $y$  towards the diagonal. Thus  $Y$  cannot be a local minimum.

Finally, suppose that  $y \neq \Delta$  and more than half the points of  $\{\phi_i(y)\}$  are off-the diagonal. Consider the point  $(\tilde{x}, \tilde{y}) \in \mathbb{R}^2$  introduced in Proposition 3.3. If  $(\tilde{x}, \tilde{y})$  lies above the diagonal then by Proposition 3.3 we know that  $\sum_i \|z - \phi_i(y)\|_1$  decreases as  $z$  travels along a straight line from  $y$  to  $m_{\{\phi_i(y)\}}$ .

If  $(\tilde{x}, \tilde{y})$  lies on or below the diagonal then the proof of Proposition 3.3 shows that  $\sum_i \|z - \phi_i(y)\|_1$  decreases as  $z$  moves from  $y$  to  $\Delta = m_{\{\phi_i(y)\}}$ . In both cases this implies that  $F_1$  would also be decreasing as  $z$  travels from  $y$  towards  $m_{\{\phi_i(y)\}}$ . Having now exhausted all the possibilities we conclude  $Y$  is not a local minimum.  $\square$

In the above theorem we made no assumption about the uniqueness of the optimal bijections  $\phi_i: Y \rightarrow X_i$ . Instead this necessary condition holds for *any* set of optimal bijections. This is slightly different to the scenario of the mean. In [28] it was shown that if  $Y$  was a local minimum of  $F_2$  the  $\phi_i$  were essentially unique, up to relabeling points in the persistent diagrams at the same location in  $\mathbb{R}^{2+}$ . However, this uniqueness does not hold for the local minima of  $F_1$ . This is because shifting an observation  $a_i$  within a population  $\{a_1, \dots, a_N\}$  of real numbers does not affect the median unless  $\text{sgn}(a_i - \text{median}\{a_1, \dots, a_N\})$  changes. Figure 5 provides an explicit example.



(a) A grouping where the  $y_2$  is the median of the selection containing the diamond on the *left*.

(b) A grouping where the  $y_2$  is the median of the selection containing the diamond on the *right*.

Figure 5:  $Y$  (black) is a local minimum of  $F_1$ . It is the unique median of the “triangle”, “square” and “diamond” diagrams. There are two optimal bijections  $\phi_3$  from  $Y$  to the “diamond” diagram, creating two groupings  $G_1$  and  $G_2$ . Here  $Y = m(G_1) = m(G_2)$ .

The following is a sufficient condition for a persistence diagram to be a local minimum of the function  $F_1: Z \mapsto \frac{1}{N} \sum_{i=1}^N d_1(X_i, Z)$  when we restrict to input diagrams  $X_i$  containing only finitely many off-diagonal points.

**Theorem 3.7.** *Let  $X_1, \dots, X_N \in \mathcal{D}^{(0,0)}$  be persistence diagrams with only finitely many off-diagonal points. Let  $Y = \{y_j\} \in \mathcal{D}$ . Suppose that whenever  $\phi_i: Y \rightarrow X_i$  are optimal bijections:*

1.  $y = m_{\{\phi_i(y)\}}$  whenever  $y \in Y$  is off-diagonal, and
2. for any selection  $S = \{x_1, x_2, \dots, x_N\}$  such that  $x_i \in X_i$  and  $\phi_i^{-1}(x_i) = \Delta$ , we have  $m_S = \Delta$ .

Then  $Y$  is a local minimum of  $F_1: Z \mapsto \frac{1}{N} \sum_{i=1}^N d_1(X_i, Z)$ .

*Proof.* Let  $\phi_i: Y \rightarrow X_i$  be optimal bijections. Assume that  $y = m_{\{\phi_i(y)\}}$  whenever  $y \in Y$  is off-diagonal. Since each of the  $X_i$  contain only finitely many off-diagonal points there can only be finitely many selections of the  $\{X_1, X_2, \dots, X_N\}$  containing some off-diagonal point. Thus there can only be finitely many off-diagonal points in  $Y$ .

Suppose that  $Y$  is not a local minimum. Then there exists a sequence  $Y_n$  that



converges to  $Y$  such that  $F_1(Y_n) < Y$  for all  $n$ . For each  $Y_n$  fix optimal bijections  $\psi_n: Y \rightarrow Y_n$ . Fix an off-diagonal point  $y \in Y$ . Since  $\|y - \Delta\| > 0$  and  $d_1(Y, Y_n) \rightarrow 0$  we know  $\psi_n(y) \neq \Delta$  for large enough  $n$ .

For each  $i$  choose optimal bijections  $\phi_i^n: Y_n \rightarrow X_i$ . Consider the sequence  $(\phi_i^n \circ \psi_n)(y) \in X_i$ . Since  $X_i$  has only finitely many off-diagonal points this sequence has a constant subsequence (here we think of the sequence containing only copies of the diagonal as constant). By taking subsequences of subsequences we can find a subsequence  $\hat{Y}_l$  of  $Y_n$  such that  $(\phi_i^l \circ \psi_l)(y)$  is constant for all off-diagonal  $y \in Y$  and all  $i$ .

Construct  $\beta_i: Y \rightarrow X_i$  by  $\beta_i(y) = \phi_i^l \circ \psi_l(y)$  and  $\beta_i(x) = \Delta$  for any remaining unmatched points  $x \in X_i$ . We will show these  $\beta_i: Y \rightarrow X_i$  are optimal bijections.

For each bijection  $\tau: A \rightarrow B$  let  $C(\tau) = \sum_{a \in A} \|a - \tau(a)\|_1$  denote the 1-Wasserstein transportation cost via the bijection  $\tau$ . Thus  $\tau: A \rightarrow B$  is an optimal bijection if and only if  $d_1(A, B) = C(\tau)$ .

Suppose that  $\beta_i: Y \rightarrow X_i$  is not optimal. This implies there is some bijection  $\alpha: Y \rightarrow X_i$  and  $\epsilon > 0$  with the  $C(\alpha) < C(\beta_i) - \epsilon$ .

Since  $\lim_{l \rightarrow \infty} \hat{Y}_l = Y$  there is some  $l$  and some bijection  $\psi_l: \hat{Y}_l \rightarrow Y$  such that  $C(\psi_l) < \epsilon/3$ . Let  $\hat{\beta}_i: \hat{Y}_l \rightarrow X_i$  be the transportation plan by first transporting  $\hat{Y}_l$  via  $\psi_l$  to  $Y$  and then transporting via  $\alpha$  to  $X_i$ . By construction  $C(\hat{\beta}_i) \leq C(\alpha) + C(\psi_l) < C(\beta_i) - 2\epsilon/3$ . But at the same time, by considering  $\beta_i$  as the composition of the transportation plans of  $\hat{\beta}_i$  and  $\psi_l$ , we know  $C(\beta_i) \leq C(\hat{\beta}_i) + C(\psi_l) < C(\hat{\beta}_i) + \epsilon/3$ .

Together these inequalities imply that  $C(\beta_i) < C(\beta_i) - \epsilon/3$  which is impossible. Thus the  $\beta_i = \phi_i^l \circ \psi_l: Y \rightarrow X_i$  are optimal bijections for all  $i$ .

Now

$$F_1(Y) = \frac{1}{N} \sum_{i=1}^N \left( \sum_{\{y \in Y: y \neq \Delta\}} \|y - \beta_i(y)\|_1 + \sum_{\{x \in X_i: \beta_i^{-1}(x) = \Delta\}} \|x - \Delta\|_1 \right)$$

and

$$\begin{aligned} F_1(Y_l) &= \frac{1}{N} \sum_{i=1}^N \left( \sum_{\{\hat{y} \in Y_l: \psi_l(\hat{y}) \neq \Delta\}} \|\psi_l(\hat{y}) - \phi_i^l(y)\|_1 + \sum_{\{\hat{y} \in Y_l: \psi_l(\hat{y}) = \Delta\}} \|\hat{y} - \phi_i^l(\hat{y})\|_1 \right) \\ &= \frac{1}{N} \sum_{i=1}^N \left( \sum_{\{y \in Y: y \neq \Delta\}} \|\psi_l^{-1}(y) - \beta_i(y)\|_1 + \sum_{\{\hat{y} \in Y_l: \psi_l(\hat{y}) = \Delta\}} \|\hat{y} - \phi_i^l(\hat{y})\|_1 \right). \end{aligned}$$

By assumption  $y = m_{\{\beta_i(y)\}}$  and hence

$$\sum_{i=1}^N \sum_{\{y \in Y: y \neq \Delta\}} \|y - \beta_i(y)\|_1 \leq \sum_{i=1}^N \sum_{\{y \in Y: y \neq \Delta\}} \|\psi_l^{-1}(y) - \beta_i(y)\|_1.$$

Thus for  $F_1(Y_l) < F_1(Y)$  to hold it must be true that

$$\sum_{i=1}^N \sum_{\{x \in X_i: \beta_i^{-1}(x) = \Delta\}} \|x - \Delta\|_1 > \sum_{i=1}^N \sum_{\{\hat{y} \in Y_l: \psi_l(\hat{y}) = \Delta\}} \|\hat{y} - \phi_i^l(\hat{y})\|_1.$$

Since  $\{\hat{y} \in Y_l : \psi_l(\hat{y}) = \Delta\} \subset \{\hat{y} \in Y_l : (\beta_i^{-1} \circ \phi_i^l)(\hat{y}) = \Delta\}$  we further know

$$\sum_{i=1}^N \sum_{\{x \in X_i : \beta_i^{-1}(x) = \Delta\}} \|x - \Delta\|_1 > \sum_{i=1}^N \sum_{\{x \in X_i : \beta_i^{-1}(x) = \Delta\}} \|x - (\phi_i^l)^{-1}(x)\|_1.$$

This implies that there is a selection  $S = \{x_1, x_2, \dots, x_N\}$  such that  $x_i \in X_i$ ,  $\beta_i^{-1}(x_i) = \Delta$  and  $m_S \neq \Delta$ , contradicting our second condition.  $\square$

Theorem 3.6 provides us with an (admittedly very slow) algorithm to find the median. We can consider the set of all groupings  $G$  up to equivalence, and their corresponding candidates  $m(G)$ . The median is one of these  $m(G)$  so we only need to compare the  $F_1(m(G))$  over all groupings  $G$ . We only have to use the necessary condition as it is always allowable to check (finitely many) extra options when looking for global minima as long as we know we have all the local minima in our list to check.

Alternatively we could use a gradient descent algorithm analogous to that used in [28]. The only modifications needed are replacing the optimal pairing using  $d_2$  with optimal pairings using  $d_1$  and replacing the means of the selections with the medians of the selections. This algorithm will terminate in finite time as at each iteration the cost function  $F_1$  decreases and uses a new grouping (of which there are only finitely many). This would not guarantee finding the global minimum but rather will terminate at a local minimum. Running multiple times from different initial locations can improve the estimate.

## 4. Comparing the median and the mean

### 4.1. Robustness of the median

For real numbers, the median is a robust measure of central tendency, while the mean is not. One measure of robustness is the breakdown point, which is of a bound on the proportion of incorrect observations (e.g. arbitrarily large observations) that an estimator can handle before giving an incorrect (e.g. arbitrarily large) result. The median has a breakdown point of 50%, while the mean has a breakdown point of 0% (a single large observation can throw it off). In this section we will prove that the breakdown points for medians and mean of persistence diagrams are the same as those of real numbers. For persistence diagrams we replace “arbitrarily large” with “arbitrarily far finite distance away.”

**Lemma 4.1.** *The breakdown point for the mean of a population of persistence diagrams lying in  $\mathcal{D}^{(0,0)}$  is 0%.*

*Proof.* Let  $X_1, X_2, \dots, X_N \in \mathcal{D}^{(0,0)}$  with mean  $Y$ . There is some  $M > 0$  such that every point in  $Y$  is at most distance  $M$  from the diagonal. Let  $\tilde{X}$  be a diagram with a single off-diagonal point  $p = (0, \sqrt{2}(KN + MN))$ . Observe that  $p$  is distance  $KN + MN$  from the diagonal.

Let  $Z$  be a mean of  $\{\tilde{X}, X_2, \dots, X_N\}$ . Using the characterization of the mean,  $Z$  must contain a point at least distance  $(KN + MN)/N = K + M$  from the diagonal. This implies that  $d_2(Z, Y) \geq K$ . By choosing  $K$  arbitrarily large we can ensure  $Y$  and  $Z$  are arbitrarily far away.  $\square$

The breakdown point for the median of a population of persistence diagrams lying in  $\mathcal{D}^{(0,0)}$  is 50%. Let  $\emptyset$  denote the persistence diagram only containing copies of the diagonal.

**Theorem 4.2.** *Let  $X_1, X_2, \dots, X_{n+1} \in \mathcal{D}^{(0,0)}$  each with only finitely many off-diagonal points. Then exists a constant  $M$  (dependent on  $X_1, X_2, \dots, X_n$ ) such that for any  $X_{n+2}, \dots, X_{2n+1} \in \mathcal{D}^{(0,0)}$ , any median of  $\{X_1, X_2, \dots, X_{n+1}, X_{n+2}, \dots, X_{2n+1}\}$  is of distance at most  $M$  from the persistence diagram only containing copies of the diagonal.*

*Proof.* Set our bound  $M$  as

$$\max_{\text{Groupings } G \text{ of } X_1, X_2, \dots, X_{n+1}} \sum_{\text{selections } s \in G} \max\{y\text{-coords in } s\} - \min\{x\text{-coords in } s\}$$

which only depends on the diagrams  $\{X_1, X_2, \dots, X_{n+1}\}$ .

Let  $Y = \{(a_j, b_j)\}$  be a median of  $\{X_1, X_2, \dots, X_{n+1}, X_{n+2}, \dots, X_{2n+1}\}$ . From optimal bijections  $\phi_i: Y \rightarrow X_i$  denote the coordinates of  $\phi_i((a_j, b_j))$  as  $(x_{i,j}, y_{i,j})$  (writing  $(\infty, -\infty)$  when  $\phi_i((a_j, b_j)) = \Delta$ ). Since  $a_j$  is the median of  $\{x_{1,j}, x_{2,j}, \dots, x_{2n+1,j}\}$  we know that  $a_j \geq \min\{x_{1,j}, x_{2,j}, \dots, x_{n+1,j}\}$ . Similarly  $b_j \leq \max\{y_{1,j}, y_{2,j}, \dots, y_{n+1,j}\}$ . This implies that for each  $j$ , we have the bound  $b_j - a_j \leq \max\{y_{1,j}, y_{2,j}, \dots, y_{n+1,j}\} - \min\{x_{1,j}, x_{2,j}, \dots, x_{n+1,j}\}$ . Thus

$$d_1(Y, \emptyset) = \sum_j b_j - a_j \leq \sum_j \max\{y_{1,j}, y_{2,j}, \dots, y_{n+1,j}\} - \min\{x_{1,j}, x_{2,j}, \dots, x_{n+1,j}\}.$$

From our characterization of the median of a set of persistence diagrams we know  $Y = m(\hat{G})$  for some grouping  $\hat{G}$  of  $\{X_1, X_2, \dots, X_{2n+1}\}$ . Let  $\hat{G}^{res}$  be the restriction of  $\hat{G}$  to the subset of diagrams  $\{X_1, X_2, \dots, X_{n+1}\}$ .

By construction

$$\begin{aligned} M &\geq \sum_{\text{selections } s \in \hat{G}^{res}} \max\{y\text{-coordinates in } s\} - \min\{x\text{-coordinates in } s\} \\ &= \sum_j \max\{y_{1,j}, y_{2,j}, \dots, y_{n+1,j}\} - \min\{x_{1,j}, x_{2,j}, \dots, x_{n+1,j}\} \\ &\geq d_1(Y, \emptyset) \end{aligned} \quad \square$$

**4.2. Number of points in the mean compared to the median**

One qualitative difference between the mean and the median is the presence or absence of points with small persistence. In some applications, such as when we have point cloud samples of an underlying shape of interest, these are heuristically the result of noise. The mean of a selection containing at least one point off the diagonal is a point off the diagonal. It is possible for the mean of  $N$  diagrams each with  $K$  points to contain  $NK$  off-diagonal points. In comparison the median of any selection with more than half copies of the diagonal will always be a copy of the diagonal. In the big picture this can add up to lots of extra points off the diagonal in the mean persistence diagram when compared to the median.

**Lemma 4.3.** *Let  $X_1, \dots, X_N$  be persistence diagrams such that the average number of off-diagonal points in the  $X_i$  is  $K$ . If  $Y$  is a median of the  $X_i$  then  $Y$  has less than  $2K$  points off the diagonal.*

*Proof.* Let  $y_1, y_2, \dots, y_n$  be the off-diagonal points in  $Y$ . Let  $\phi_i$  be optimal bijections between  $Y$  and the  $X_i$ . By Theorem 3.6 we know that  $y_j$  is the median of  $\{\phi_i(y_j)\}$  for each  $j$ . By Lemma 3.4 we know that for each  $j$  the sets  $\{\phi_i(y_j)\}$  must contain at least  $(N + 1)/2$  off-diagonal points. This implies that  $\cup_j \{\phi_i(y_j)\}$  must contain at least  $(N + 1)n/2$  points.

Since the combined total of all the off-diagonal points in the  $X_i$  is  $NK$  we can conclude that  $(N + 1)n/2 \leq NK$  and hence  $n < 2K$ .  $\square$

We illustrate this significant advantage of the median with a simulated geometric example. We generated point clouds of the unit circle by drawing 25 points from the uniform measure on the unit circle convoluted with Gaussian noise with variance  $\sigma^2$ . We then build the  $H_1$  persistence diagrams from the corresponding Rips filtration of this point cloud (described in the appendix). For the 5 persistence diagrams thus produced we then computed the mean and the median. These are illustrated in Figure 6.

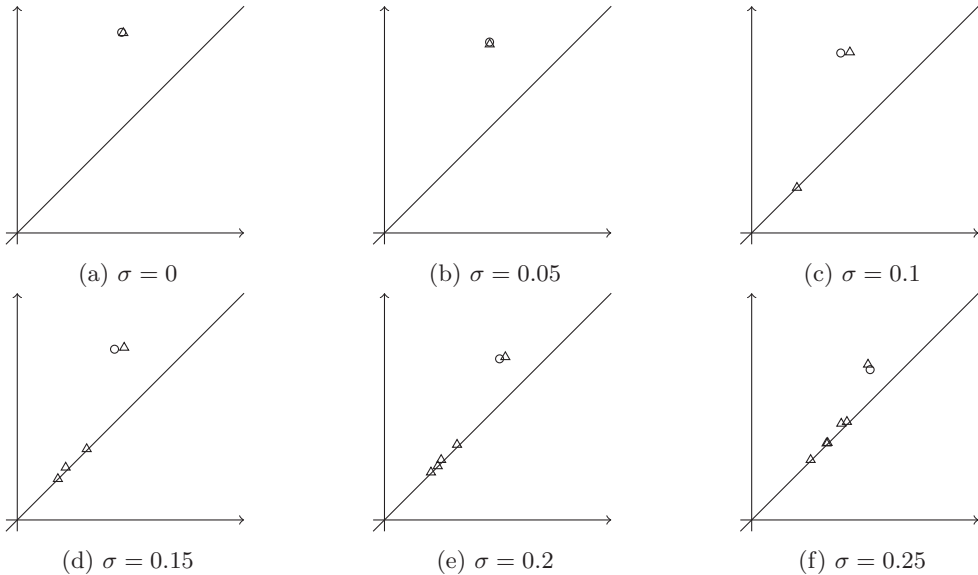


Figure 6: For each standard deviation  $\sigma$  we randomly generated five noisy point clouds of the circle each 25 points drawn i.i.d. from the convolution of the uniform measure on the unit circle convoluted with Gaussian noise with standard deviation  $\sigma$ . From these five point clouds we constructed five Rips filtrations and their  $H_1$  persistence diagrams. The corresponding median diagram is then depicted using circles and the mean diagram using triangles.

Since the underlying shape of interest is a circle there should be one point in each persistence diagram far from the diagonal corresponding to the  $H_1$  class of the circle, alongside extra “noisy” points near the diagonal (with more as the noise parameter in the sampling process increases). In the simulated data the mean and median each have one point far from the diagonal (and these are close to each other) but with

larger noise the mean diagram has more extra points near the diagonal than the median diagram.

### 4.3. Discontinuities and non-uniqueness of the mean and the median

Unfortunately both the mean and median are neither continuous nor always unique. There can be discontinuities when the grouping  $G$  which provides us with the optimal candidate for the mean or the median switches. This is illustrated in the Figures 7 and 8.

In these examples we have three diagrams, one consists only of copies of the diagonal, one containing off-diagonal points denoted by squares, and the other denoted by triangles. In this example as  $z$  increases in the squares diagram travels across the optimal grouping changes from  $\{\{x_1, (1, z), \Delta\}, \{x_2, \Delta, \Delta\}\}$  to  $\{\{x_1, \Delta, \Delta\}, \{x_2, (1, z), \Delta\}\}$  leading to a discontinuity to both the mean and the median (note that the value of  $z$  where the switch occurs is different for the mean and median). At the time it switches both groupings are equally optimal and hence we have non-uniqueness.

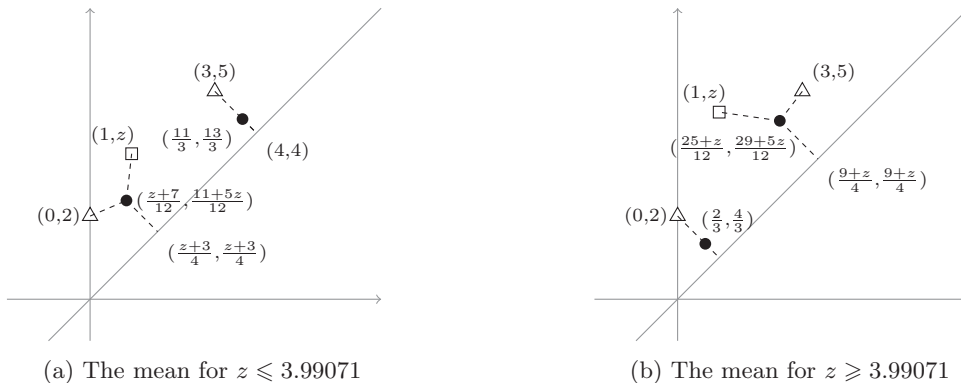


Figure 7: We have three diagrams, one consists only of copies of the diagonal, one containing off-diagonal points denoted by squares, and the other containing off-diagonal points denoted by triangles. In (a)  $F_2(\text{circles}) = \frac{8639 - 3995z + 1268z^2}{6534}$  and in (b)  $F_2(\text{circles}) = \frac{191 - 58z + 7z^2}{36}$ . When  $z < 3.99070$  the optimal grouping is  $\{(0, 2), (1, z), \Delta\}$  and  $\{(3, 5), \Delta, \Delta\}$  (used in (a)). When  $z > 3.99072$  then the optimal grouping is  $\{(0, 2), \Delta, \Delta\}$  and  $\{(3, 5), (1, z), \Delta\}$  (used in (b)). Both groupings are optimal when  $z \simeq 3.99071$  and as a result we do not have a unique mean.

The mean is generically unique but the median is not. To show this rigorously we shall restrict ourselves to the case where we have  $N$  diagrams each with only finitely many off-diagonal points. Let  $k_1, k_2, \dots, k_N$  be non-negative integers. Let  $U(k_1, k_2, \dots, k_N)$  denote the space of sets of diagrams  $X = \{X_1, X_2, \dots, X_N\}$  such that  $X_i$  has  $k_i$  off-diagonal points.  $U(k_1, k_2, \dots, k_N)$  is the quotient of  $(\mathbb{R}^{2+})^{k_1+k_2+\dots+k_N}$  by a finite group of symmetries  $\Gamma$ . There is a quotient map

$$q: (\mathbb{R}^{2+})^{k_1+k_2+\dots+k_N} \rightarrow U(k_1, k_2, \dots, k_N) = (\mathbb{R}^{2+})^{k_1+k_2+\dots+k_N} / \Gamma.$$

Let  $\lambda$  be Lebesgue measure on  $(\mathbb{R}^{2+})^{k_1+k_2+\dots+k_N}$  and let  $\rho = q_*(\lambda)$  be the push forward of Lebesgue measure onto  $U(k_1, k_2, \dots, k_N)$ .

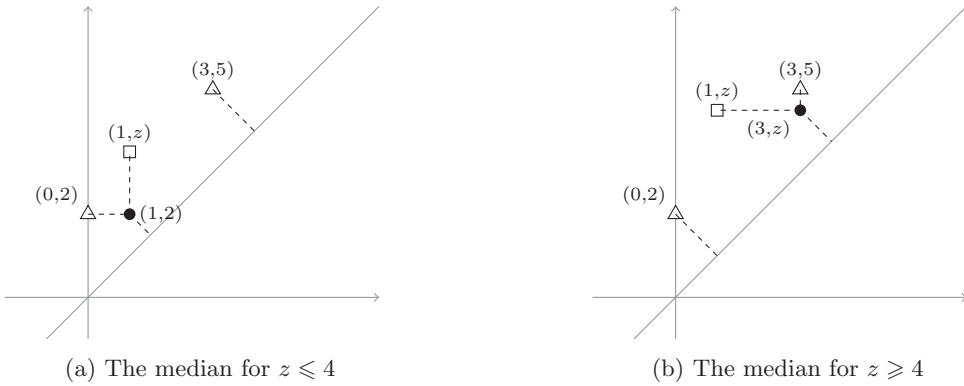


Figure 8: We have three diagrams, one consists only of copies of the diagonal, one containing off-diagonal points denoted by squares, and the other containing off-diagonal points denoted by triangles. When  $z < 4$  the optimal grouping is  $\{(0, 2), (1, z), \Delta\}$  and  $\{(3, 5), \Delta, \Delta\}$  (the grouping used in (a)). When  $z > 4$  then the optimal grouping is  $\{(0, 2), \Delta, \Delta\}$  and  $\{(3, 5), (1, z), \Delta\}$  (the grouping used in (b)). Both are optimal when  $z = 4$  and as a result we do not have a unique median.

**Proposition 4.4.** *The sets of diagrams in  $U(k_1, k_2, \dots, k_N)$  which do not have a unique mean has measure zero.*

*Proof.* Let  $\tilde{A}$  be the set of sets of diagrams in  $U(k_1, k_2, \dots, k_N)$  which do not have a unique mean. Then  $A = q^{-1}(\tilde{A})$  is the set of vectors of labeled diagrams (in  $(\mathbb{R}^{2+})^{k_1+k_2+\dots+k_N}$ ) which do not have a unique mean. Since  $\rho(\tilde{A}) = \lambda(q^{-1}(\tilde{A}))$  it is sufficient to show  $\lambda(A) = 0$ .

Let  $S$  be a selection containing the points  $\{(a_1, b_1), (a_2, b_2), \dots, (a_k, b_k)\}$  with  $N - k$  copies of the diagonal. In the appendix we define the mean of the selection  $S$  as the minimizer of  $f_S(y) = \sum_{x \in S} \|x - y\|_2^2$ . which occurs at

$$\mu_S = \left( \frac{1}{N} \left( k\hat{x} + (N - k)\frac{\hat{x} + \hat{y}}{2} \right), \frac{1}{N} \left( k\hat{y} + (N - k)\frac{\hat{x} + \hat{y}}{2} \right) \right)$$

where  $\hat{x}$  and  $\hat{y}$  are the means of  $a_1, a_2, \dots, a_k$  and  $b_1, b_2, \dots, b_k$  respectively.

For each pair of distinct groupings,  $G_1$  and  $G_2$ , of the labeled diagrams let  $A(G_1, G_2) = \{X = (X_1, X_2, \dots, X_N) : \sum_{S \in G_1} f_S(\mu_S) = \sum_{S \in G_2} f_S(\mu_S)\}$ . Any  $X \in A(G_1, G_2)$  must satisfy a quadratic equation so either  $A(G_1, G_2) = (\mathbb{R}^{2+})^{k_1+k_2+\dots+k_N}$  or  $\lambda(A(G_1, G_2)) = 0$ . It is clear that there exists a vector of labeled persistence diagrams  $X = (X_1, X_2, \dots, X_N) \in (\mathbb{R}^{2+})^{k_1+k_2+\dots+k_N}$  such that  $X \notin A(G_1, G_2)$  we conclude that  $\lambda(A(G_1, G_2)) = 0$ .

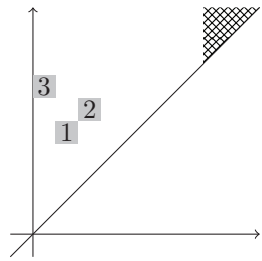
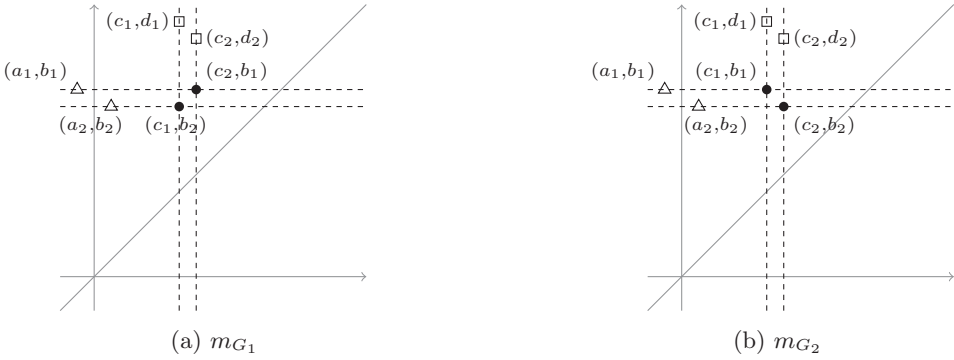
If  $X$  has more than one mean then by Proposition B.2 there must be groupings  $G_1, G_2$  such that  $\mu_{G_1} \neq \mu_{G_2}$  but  $\sum_{S \in G_1} f_S(\mu_S) = F_2(\mu_{G_1}) = F_2(\mu_{G_2}) = \sum_{S \in G_2} f_S(\mu_S)$ . This implies  $A \subseteq \bigcup_{G_1 \neq G_2 \text{ groupings}} A(G_1, G_2)$ . There are only finitely many groupings so  $\lambda(A) = 0$ . □

This proof of generic uniqueness contrasts sharply to the case of the median which is not generically unique.

**Proposition 4.5.** *Let  $N \geq 3$  be an odd number. Let  $k_1, k_2, \dots, k_{(N+1)/2} \geq 2$ . The sets of diagrams in  $U(k_1, k_2, \dots, k_N)$  which do not have a unique median has positive measure.*

*Proof.* We will first illustrate this with the case  $U(2, 2, 0)$  which shows the idea of the general case. Suppose  $X_1$  and  $X_2$  each contain two off-diagonal points  $\{(a_1, a_2), (a_2, b_2)\}$ , and  $\{(c_1, d_1), (c_2, d_2)\}$  respectively, and  $X_3$  has no off-diagonal points. Further suppose that  $a_1, a_2 < c_1, c_2 \leq b_1, b_2 < d_1, d_2$ .

First consider the grouping  $G_1 = \{S_{(1,2)}, S_{(2,1)}\}$  where  $S_{(1,2)} := \{(a_1, b_1), (c_2, d_2), \Delta\}$  and  $S_{(2,1)} := \{(a_2, b_2), (c_1, d_1), \Delta\}$ . The median of the selection  $S_{(1,2)}$  is  $(c_2, b_1)$  and the median of the selection  $S_{(2,1)}$  is  $(c_1, b_2)$ . This implies that  $m_{G_1}$  has off-diagonal points  $\{(c_2, b_1), (c_1, b_2)\}$ . Also consider the grouping  $G_2 = \{S_{(1,2)}, S_{(2,1)}\}$  where  $S_{(1,2)} := \{(a_1, b_1), (c_2, d_2), \Delta\}$  and  $S_{(2,1)} := \{(a_2, b_2), (c_1, d_1), \Delta\}$ . Analogous calculations show the off-diagonal points of  $m_{G_2}$  are  $\{(c_1, b_1), (c_2, b_2)\}$ . These groupings are illustrated in Figures 9a and 9b.



(c) The different regions of  $\mathbb{R}^{2+}$  for constructing sets of populations of persistence diagrams with non unique medians.

Figure 9

Now  $F_1(m_{G_1}) = -a_1 + d_1 - a_2 + d_2 = F_1(m_{G_2})$  and that  $F_1(m_G) \geq -a_1 + d_1 - a_2 + d_2$  for all other groupings  $G$ . This implies that  $m_{G_1}$  and  $m_{G_2}$  are both medians of  $X$ . If  $b_1 \neq b_2$  and  $c_1 \neq c_2$  these medians are distinct and thus we do not have a unique median. The measure of such sets of diagrams  $\{X_1, X_2, X_3\}$  has non-zero measure in  $U(2, 2, 0)$ .

The extension of this example to when  $k_1, k_2, \dots, k_{(N+1)/2} > 2$  is illustrated in Figure 9c. We need to find an example of a non-zero measure set of  $(X_1, X_2, \dots, X_N) \in (\mathbb{R}^{2+})^{k_1+k_2+\dots+k_N}$  with  $k_1, k_2, \dots, k_{(N+1)/2} > 2$  with non unique medians.

We will require that:

- $X_1$  has two points  $(a_1, b_1)$  and  $(a_2, b_2)$  in region 1,
- $X_2$  has two points  $(c_1, d_1)$  and  $(c_2, d_2)$  in region 2,
- $X_3, X_4, \dots, X_{(N-3)/2}$  each contains two points in region 3, and
- every other off-diagonal point in the  $X_i$  lies in the region patterned by cross-hatch.

Note that this set of populations of persistence diagrams is of non-zero measure in  $U(k_1, k_2, \dots, k_N)$ .

Every median of  $\{X_i\}$  can be written as  $m_G$  where each selection in  $G$  contains either points in the cross-hatch region (and potentially copies of the diagonal) or they contain one point each from regions 1 and 2,  $(N-3)/2$  points from region 3, and  $(N-1)/2$  copies of the diagonal. There is a median  $m$  with off-diagonal points  $\{(c_1, b_1), (c_2, b_2)\}$  alongside other points determined by the points in the cross-hatched region. Another median  $\tilde{m}$  is the same as  $m$  but switching  $\{(c_1, b_1), (c_2, b_2)\}$  for  $\{(c_2, b_1), (c_1, b_2)\}$ .  $\square$

## 5. Discussion and further directions

There are many parallels between the mean and median of populations of persistence diagrams. This suggests some future directions could involve extending work that has been done on the mean to the corresponding results for the median. For example, in [23] they explore an alternative probabilistic definition of the mean which combines the tradition mean used with the notion of a shaking hand equilibrium in game theory. This alternate definition is unique and continuous. We believe a similar idea would work to create a probabilistic definition of the median.

Another future direction is to combine the median with sampling theorems to find conditions to infer the correct homology with high probability. For example, the homology of a set can be inferred from the persistence diagram corresponding to a point cloud with small Hausdorff distance to the original set. Under certain sampling conditions we can ensure that this Hausdorff distance is small with high probability. Perhaps with higher probability the median of independently obtained persistence diagrams under such samplings conditions will provide the correct homology.

There is scope for further developments in algorithms, both in design and implementation. In this paper we have discussed very referred to some of the computational aspects, including mentioning a brute force algorithm and a gradient descent approach. Perhaps there could be significant improvements by using geometry analogous to the work in [17] where they show that by exploiting the inherit geometry of the points in persistence diagrams lying in a plane we can approximate the Wasserstein distances much faster.



### Appendix A. Rips filtration

The Rips filtration  $\mathcal{R}$  is the filtration of the flag complex on 25 vertices where at time  $t$ ,  $\mathcal{R}_t$  contains all the vertices  $[v]$ , the edges  $[v_0, v_1]$  whenever  $\|v_0 - v_1\| < t$ , the 2-simplicies  $[v_0, v_1, v_2]$  whenever  $[v_0, v_1]$ ,  $[v_1, v_2]$  and  $[v_2, v_0]$  are all included in  $\mathcal{R}_t$ , and so on including higher dimensional simplicies whenever all their boundary faces are in  $\mathcal{R}_t$ .

In other words  $\mathcal{R}_t$  is the flag complex (also known as the clique complex) on the graph containing all edges of length at most  $t$ .

### Appendix B. Mean diagram

The methods here provide a proof for the necessary condition for a persistence diagram to be a local minimum of the Fréchet function which is far simpler than that in [28]. We also extend the results to persistence diagrams containing points in  $\mathcal{L}_\infty$  and  $\mathcal{L}_{-\infty}$ . Due to the similarities to the earlier material we omit many of the details.

For the mean we also split our analysis into the restrictions to  $\mathbb{R}^{2+} \cup \Delta$ ,  $\mathcal{L}_\infty$  and  $\mathcal{L}_{-\infty}$ . If  $X_1, X_2, \dots, X_N \in \mathcal{D}^{(k,l)}$  then  $Y$  is a mean of the  $X_i$  if and only if  $Y|_{\mathbb{R}^{2+} \cup \Delta}$ ,  $Y|_{\mathcal{L}_\infty}$ , and  $Y|_{\mathcal{L}_{-\infty}}$  are means of the  $X_i$  each restricted to the appropriate domain.

As  $\mathcal{L}_\infty$  and  $\mathcal{L}_{-\infty}$  are effectively copies of  $\mathbb{R}$  we can easily characterize the means of populations of multisets in them. Suppose  $A_1, A_2, \dots, A_N$  are each multisets of exactly  $k$  real numbers and we label the elements of each  $A_i$  so that  $A_i = \{a_{i,1}, a_{i,2}, \dots, a_{i,k}\}$  with  $a_{i,1} \leq a_{i,2} \leq \dots \leq a_{i,k}$ . Set  $B = \{b_1, b_2, \dots, b_k\}$  where  $b_j$  is the mean of  $\{a_{1,j}, \dots, a_{N,j}\}$ . Then  $B$  is the unique multiset of  $k$  real numbers that minimizes

$$f_2: Y \mapsto \sum_{i=1}^N \left( \inf_{\phi: A_i \rightarrow Y, \phi \text{ bijection}} \sum_{a \in A_i} |a - \phi(a)|^2 \right)$$

and hence  $B$  is the mean.

Characterizing the means of populations of persistence diagrams in  $\mathcal{D}^{(0,0)}$  is analogously achieved through the means of selections.

**Lemma B.1.** *Let  $(a_1, b_1), (a_2, b_2), \dots, (a_k, b_k)$  be points in the plane. Let  $\hat{x}$  be the mean of  $a_1, a_2, \dots, a_k$  and  $\hat{y}$  be the mean of  $b_1, b_2, \dots, b_k$ . Then*

$$(\tilde{x}, \tilde{y}): = \left( \frac{1}{N} \left( k\hat{x} + (N - k) \frac{\hat{x} + \hat{y}}{2} \right), \frac{1}{N} \left( k\hat{y} + (N - k) \frac{\hat{x} + \hat{y}}{2} \right) \right)$$

is the unique point in  $\mathbb{R}^{2+}$  which minimizes

$$f(x, y) = \sum_{i=1}^k \|(x, y) - (a_i, b_i)\|_2^2 + \sum_{i=k+1}^N \|(x, y) - \Delta\|_2^2.$$

Let  $S$  be a multiset in  $\mathbb{R}^{2+} \cup \Delta$  containing  $(a_1, b_1), (a_2, b_2), \dots, (a_k, b_k)$  and  $N - k$  copies of the diagonal and let  $(x, y)$  be the point in  $\mathbb{R}^{2+}$  found in Lemma B.1. We call this  $(x, y)$  the *mean* of  $S$  and denote it by  $\mu_S$ . An example is illustrated in Figure 10.

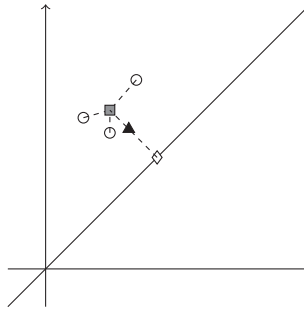


Figure 10: We want the mean of the three points marked by circles alongside two copies of the diagonal. The gray square is the arithmetic mean of the three points marked by circles. The diamond is the point on the diagonal closest to the square. The triangle is the mean of the circles and 2 copies of the diamond. It is the weighted average of the square and the diamond.

Let  $\mu(G)$  be the persistence diagram which contains  $\{\mu_{S_j} : S_j \in G\}$ . Each grouping  $G$  produces a candidate  $\mu(G)$  for the mean. We will show that any mean must be  $\mu(G)$  for some grouping  $G$ .

**Theorem B.2.** *Let  $X_1, \dots, X_N, Y \in \mathcal{D}^{(0,0)}$  be persistence diagrams each with finitely many off-diagonal points. Let  $F_2: Z \mapsto \frac{1}{N} \sum_{i=1}^N d_2(X_i, Z)^2$ . For each  $i$  fix an optimal bijection  $\phi_i: Y \rightarrow X_i$  (for  $d_2$ ). For each  $y \in Y$  we have a selection  $\{\phi_i(y)\}$ . Let  $G_Y$  be the grouping  $\{\{\phi_i(y)\}: y \in Y\}$ . If  $Y$  is a local minimum of  $F_2$  then  $Y = \mu_{G_Y}$ .*

*Proof.* Suppose that  $Y \neq \mu_{G_Y}$  and thus  $y_0 \neq \mu_{\{\phi_i(y_0)\}}$  for some  $y_0 \in Y$ . Set  $Y_t$  to be the diagram which agrees with  $Y$  except the point  $y_0$  is replaced with  $(1-t)y_0 + t\mu_{\{\phi_i(y_0)\}}$ ,

$$\begin{aligned} F_2(Y_t) &= \frac{1}{N} \sum_{i=1}^N \inf_{\phi_i^t: Y_t \rightarrow X_i} \sum_{y_t \in Y_t} \|y_t - \phi_i^t(y_t)\|^2 \\ &\leq \frac{1}{N} \sum_{i=1}^N \sum_{y \in Y, y \neq y_0} \|y - \phi_i(y)\|^2 + (((1-t)y_0 + t\mu_{\{\phi_i(y_0)\}}) - \phi_i(y_0))^2. \end{aligned}$$

We thus can conclude that for all  $t \in (0, 1)$

$$F_2(Y_t) - F_2(Y_0) \leq \frac{1}{N} \sum_{i=1}^N (y_0 - \phi_i(y_0))^2 - (((1-t)y_0 + t\mu_{\{\phi_i(y_0)\}}) - \phi_i(y_0))^2,$$

which we know is negative from the proof of Lemma B.1. This implies that  $Y_0 = Y$  can not be a local minimum.  $\square$

### References

[1] R.J. Adler, O. Bobrowski, and S. Weinberger. Crackle: The homology of noise. *Discrete & Computational Geometry*, 52(4):680–704, 2014.

- [2] Ch. Bajaj. The algebraic degree of geometric optimization problems. *Discrete & Computational Geometry*, 3(2):177–191, 1988.
- [3] P. Bendich, H. Edelsbrunner, and M. Kerber. Computing robustness and persistence for images. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1251–1260, 2010.
- [4] A.J. Blumberg, I. Gal, M.A. Mandell, and M. Pancia. Robust statistics, hypothesis testing, and confidence intervals for persistent homology on metric measure spaces. *Foundations of Computational Mathematics*, 14(4):745–789, 2014.
- [5] O. Bobrowski and R.J. Adler. Distance functions, critical points, and the topology of random čech complexes. *Homology, Homotopy and Applications*, 16(2):311–344, 2014.
- [6] P. Bubenik. Statistical topological data analysis using persistence landscapes. *The Journal of Machine Learning Research*, 16(1):77–102, 2015.
- [7] P. Bubenik, G. Carlsson, P.T. Kim, and Z.-M. Luo. Statistical topology via Morse theory persistence and nonparametric estimation. *Algebraic methods in statistics and probability II*, 516:75–92, 2010.
- [8] A. Cerri, M. Ferri, and D. Giorgi. Retrieval of trademark images by means of size functions. *Graphical Models*, 68(5):451–471, 2006.
- [9] F. Chazal, B.T. Fasy, F. Lecci, A. Rinaldo, A. Singh, and L. Wasserman. On the bootstrap for persistence diagrams and landscapes. *Modeling and Analysis of Information Systems*, 6(20):96–105, 2013.
- [10] F. Chazal, M. Glisse, C. Labruère, and B. Michel. Convergence rates for persistence diagram estimation in topological data analysis. *The Journal of Machine Learning Research*, 16(1):3603–3635, 2015.
- [11] W. Crawley-Boevey. Decomposition of pointwise finite-dimensional persistence modules. *Journal of Algebra and Its Applications*, 14(05):1550066, 2015.
- [12] H. Edelsbrunner and J.L. Harer. *Computational Topology: An Introduction*. Amer. Math. Soc., 2010.
- [13] B.T. Fasy, F. Lecci, A. Rinaldo, L. Wasserman, S. Balakrishnan, A. Singh, et al. Confidence sets for persistence diagrams. *The Annals of Statistics*, 42(6):2301–2339, 2014.
- [14] J. Gamble and G. Heo. Exploring uses of persistent homology for statistical analysis of landmark-based shape data. *Journal of Multivariate Analysis*, 101(9):2184–2199, 2010.
- [15] M. Kahle. Random geometric complexes. *Discrete & Computational Geometry*, 45(3):553–573, 2011.
- [16] M. Kahle, E. Meckes, et al. Limit theorems for Betti numbers of random simplicial complexes. *Homology, Homotopy and Applications*, 15(1):343–374, 2013.
- [17] M. Kerber, D. Morozov, and A. Nigmatov. Geometry helps to compare persistence diagrams. *Journal of Experimental Algorithmics (JEA)*, 22:1–4, 2017.
- [18] W.A. Kirk. Geodesic geometry and fixed point theory. In *Seminar of Mathematical Analysis (Malaga/Seville, 2002/2003)*, volume 64, pages 195–225, 2003.

- [19] X. Lu, J.S. Marron, and P. Haaland. Object-oriented data analysis of cell images. *Journal of the American Statistical Association*, 109(506):548–559, 2014.
- [20] J.S. Marron and A.M. Alonso. Overview of object oriented data analysis. *Biometrical Journal*, 56(5):732–753, 2014.
- [21] J.S. Marron, J.O. Ramsay, L.M. Sangalli, A. Srivastava, et al. Functional data analysis of amplitude and phase variation. *Statistical Science*, 30(4):468–484, 2015.
- [22] Y. Mileyko, S. Mukherjee, and J. Harer. Probability measures on the space of persistence diagrams. *Inverse Problems*, 27(12):124007, 2011.
- [23] E. Munch, K. Turner, P. Bendich, S. Mukherjee, J. Mattingly, J. Harer, et al. Probabilistic fréchet means for time varying persistence diagrams. *Electronic Journal of Statistics*, 9(1):1173–1204, 2015.
- [24] S.-i. Ohta. Barycenters in Alexandrov spaces of curvature bounded below. *Adv. Geom.*, 12:571–587, 2012.
- [25] J.O. Ramsay. *Functional data analysis*. Wiley Online Library, 2006.
- [26] V. Robins and K. Turner. Principal component analysis of persistent homology rank functions with case studies of spatial point patterns, sphere packing and colloids. *Physica D: Nonlinear Phenomena*, 334:99–117, 2016.
- [27] L.M. Sangalli, P. Secchi, and S. Vantini. Object oriented data analysis: a few methodological challenges. *Biometrical Journal*, 56(5):774–777, 2014.
- [28] K. Turner, Y. Mileyko, S. Mukherjee, and J. Harer. Fréchet means for distributions of persistence diagrams. *Discrete & Computational Geometry*, 52(1):44–70, July 2014.
- [29] K. Turner, S. Mukherjee, and D.M. Boyer. Persistent homology transform for modeling shapes and surfaces. *Information and Inference*, 3(4):310–344, 2014.
- [30] H. Wang, J.S. Marron, et al. Object oriented data analysis: Sets of trees. *The Annals of Statistics*, 35(5):1849–1873, 2007.
- [31] D. Yogeshwaran, R.J. Adler, et al. On the topology of random complexes built over stationary point processes. *The Annals of Applied Probability*, 25(6):3338–3380, 2015.
- [32] D. Yogeshwaran, E. Subag, and R.J. Adler. Random geometric complexes in the thermodynamic regime. *Probability Theory and Related Fields*, pages 1–36, 2014.

Katharine Turner [katharine.turner@anu.edu.au](mailto:katharine.turner@anu.edu.au)

Mathematical Sciences Institute, Building 145, Science Road, ANU, Acton, Australia