

# PROBABILISTIC FRÉCHET MEANS FOR TIME VARYING PERSISTENCE DIAGRAMS

ELIZABETH MUNCH<sup>1</sup>, KATHARINE TURNER<sup>2</sup>, PAUL BENDICH<sup>2</sup>, SAYAN MUKHERJEE<sup>4</sup>,  
JONATHAN MATTINGLY<sup>5</sup>, AND JOHN HARER<sup>6</sup>

ABSTRACT. In order to use persistence diagrams as a true statistical tool, it would be very useful to have a good notion of mean and variance for a set of diagrams. In [21], Mileyko and his collaborators made the first study of the properties of the Fréchet mean in  $(\mathcal{D}_p, W_p)$ , the space of persistence diagrams equipped with the  $p$ -th Wasserstein metric. In particular, they showed that the Fréchet mean of a finite set of diagrams always exists, but is not necessarily unique. The means of a continuously-varying set of diagrams do not themselves (necessarily) vary continuously, which presents obvious problems when trying to extend the Fréchet mean definition to the realm of vineyards.

We fix this problem by altering the original definition of Fréchet mean so that it now becomes a probability measure on the set of persistence diagrams; in a nutshell, the mean of a set of diagrams will be a weighted sum of atomic measures, where each atom is itself a persistence diagram determined using a perturbation of the input diagrams. This definition gives for each  $N$  a map  $(\mathcal{D}_p)^N \rightarrow \mathbb{P}(\mathcal{D}_p)$ . We show that this map is Hölder continuous on finite diagrams and thus can be used to build a useful statistic on time-varying persistence diagrams, better known as vineyards.

## 1. INTRODUCTION

The field of topological data analysis (TDA) was first introduced [14] in 2000, and has rapidly been applied to many different areas: for example, in the study of protein structure [1, 2, 20], plant root structure [18], speech patterns [4], image compression and segmentation [6, 16], neuroscience [11], orthodontia [19], gene expression [13], and signal analysis [25].

A key tool in TDA is the *persistence diagram* [7, 14]. Given a set of points  $S$  in some possibly high-dimensional metric space, the persistence diagram  $D(S)$  is a computable summary of the data which provides a compact two-dimensional representation of the multi-scale topological information carried by the point cloud; see Fig. 1 for an example of such a diagram and Section 2 for a more rigorous description. If the point cloud varies continuously over time (or some other parameter) then the persistence diagrams vary continuously over time [9]; the diagrams stacked on top of each other then form what is called a vineyard [10].

A key part of data analysis is to model variation in data. In particular, there is an interest in object oriented data analysis where the data of study is a more complicated object than just points in Euclidean space. Thus, there has been a recent effort to study the mean and variance of a set of persistence diagrams [3, 5, 21, 28], as well as nice convergence rates for persistence diagrams of larger and larger point clouds sampled from a compactly-supported measure [8]. There are a variety of reasons to want to characterize statistical properties of diagrams. For example, given a massive point cloud  $S$ , there is a computational and statistical advantage to subsampling the data to produce smaller point clouds  $S_1, \dots, S_n$ , and computing the mean and variance of the set of persistence diagrams obtained from the  $n$  subsampled data sets. In statistical terminology, this example consists of computing a bootstrap estimate [17] of persistence diagram of the data. This procedure requires a good definition for the mean (and variance) of a set of persistence diagrams.

The papers [21, 28] make careful study of the geometric and analytic properties of the space  $(\mathcal{D}_p, W_p)$  of persistence diagrams equipped with the Wasserstein metric. This enables defining the mean and variance via the Fréchet function [21], and an algorithm for their computation [28]. There are, however, unfortunate problems with using the Fréchet mean: the mean of a set of diagrams is not necessarily unique nor continuous.

---

*Date:* November 19, 2014.

AMS SUBJECT CLASSIFICATION: 55, 60

In this paper, we provide an alternative definition for the mean of a set of diagrams which we call the Probabilistic Fréchet Mean (PFM). By combining the notions of the Fréchet mean and the trembling hand equilibrium in game theory [26], we construct a mean that is not itself a diagram, but is rather a probabilistic mixture of diagrams and thus an element of  $\mathcal{P}(\mathcal{D}_p)$ , the space of probability distributions over persistence diagrams. Uniqueness of this new mean will be obvious from the definition we propose. More crucially, we prove the following corollary to the main technical result, Thm. 5.3.

**Corollary 5.4.** *Let*

$$\begin{aligned} \Phi : \quad (S_{M,K})^N &\longrightarrow \mathcal{P}(S_{M,NK}) \\ (X_1, \dots, X_N) &\longmapsto \mu_X \end{aligned}$$

*be the map which sends a set of diagrams to its PFM. Then  $\Phi$  is Hölder continuous with exponent  $1/2$ . That is, there is a constant  $C'$  such that*

$$\mathcal{W}_2(\mu_X, \mu_Y) \leq C' \sqrt{\vec{d}_2(X, Y)}$$

*for all  $X, Y \in (S_{M,K})^N$ . Here  $\mathcal{W}_2$  is the Wasserstein metric on the space of probability distributions over the space of persistence diagrams, and  $\vec{d}_2$  is the metric on  $(S_{M,K})^N$  induced from using the Wasserstein analogous metric on the space of persistence diagrams coordinate-wise.*

Thus, if we compute the new mean on each step of a path in  $\mathcal{D}_p$ , the resulting object gives a path in  $\mathcal{P}(\mathcal{D}_p)$ , thus making the construction amenable to analyzing distributions of vineyards. This is stated specifically in the following corollary.

**Corollary 5.5.** *Let  $\gamma_1, \dots, \gamma_N : [0, 1] \rightarrow \mathcal{D}_2$  be vineyards in  $\mathcal{V}_2$ . Then*

$$\begin{aligned} \mu_\gamma : \quad [0, 1] &\longrightarrow \mathcal{P}(S_{M,K}) \\ t &\longmapsto \mu_{\gamma_1(t), \dots, \gamma_N(t)} \end{aligned}$$

*is continuous.*

Finally, we give examples of this mean computed on diagrams drawn from samples of various point clouds, and introduce a useful way to visualize them.

**Outline.** Section 2 contains definitions for persistence diagrams and vineyards, as well as a discussion of the space  $(\mathcal{D}_p, W_p)$ . The contributions of [21] and [28] are reviewed more fully in Section 3, and the non-uniqueness issue is also discussed in that section. We give our new definition, the probabilistic Fréchet mean (PFM), in Section 4, and prove its desirable theoretical properties in Section 5. Examples, implementation details, and a discussion of visualization are in Section 7, and the paper concludes with some discussion in Section 8.

## 2. DIAGRAMS AND VINEYARDS

Here we give the basic definitions for persistence diagrams and vineyards, and then move on to a description of the metric space  $(\mathcal{D}_p, W_p)$ . For more details on persistence, see [15]. We assume the reader is familiar with homology; [24] is a good reference. We note that all homology groups in this paper are computed with field coefficients.

**2.1. Persistent Homology.** To define persistent homology, we start with a nested sequence of topological spaces,

$$(1) \quad \emptyset = \mathbb{X}_0 \subseteq \mathbb{X}_1 \subseteq \mathbb{X}_2 \subseteq \dots \subseteq \mathbb{X}_n = \mathbb{X}.$$

Often this sequence arises from the sublevel sets of a continuous function,  $f : \mathbb{X} \rightarrow \mathbb{R}$ , where  $\mathbb{X}_i = f^{-1}((-\infty, a_i])$  with  $a_0 \leq a_1 \leq \dots \leq a_n$ . For many applications, this function is the distance function

$$d_S(x) = \inf_{v \in S} \|x - v\|$$

from a point cloud  $S$  such as in the example of Fig. 1. In this case, a sublevel set can be visualized as a union of balls around the points in  $S$ .

The sequence of inclusion maps from Eqn. (1) induces maps on homology for any dimension  $r$ ,

$$(2) \quad 0 \longrightarrow H_r(\mathbb{X}_1) \longrightarrow H_r(\mathbb{X}_2) \longrightarrow \dots \longrightarrow H_r(\mathbb{X}_n).$$

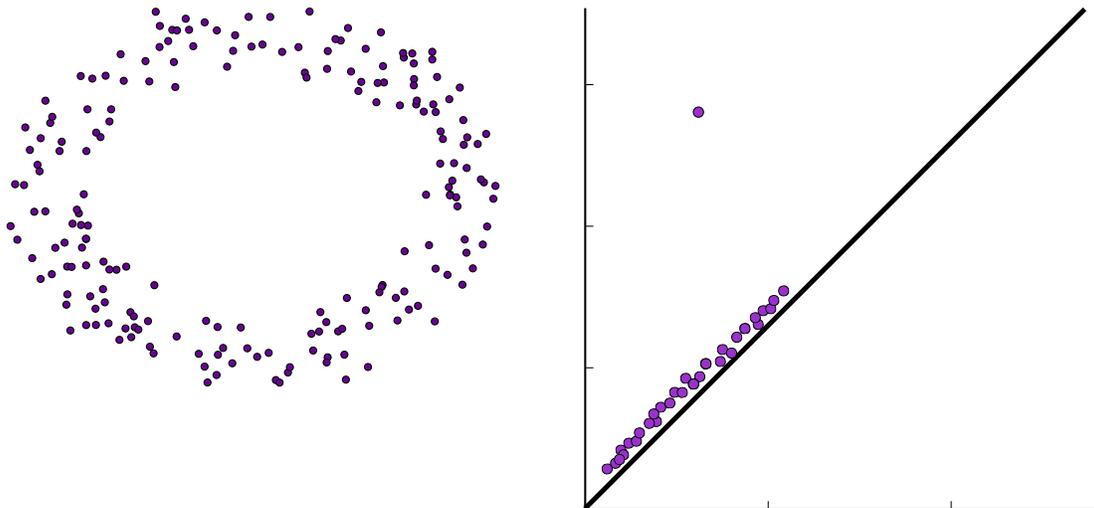


FIGURE 1. A point cloud, shown at left, is sampled from an annulus. In order to summarize the topological data, we look at the sublevel sets of the distance function from the set of points, then construct the persistence diagram, shown at right. The points near the diagonal are considered noise, while the single point far from the diagonal gives information about the hole in the annulus.

In order to understand the changing space, we look at where homology classes appear and disappear in this sequence.

Let  $\varphi_i^j : H_r(\mathbb{X}_i) \rightarrow H_r(\mathbb{X}_j)$  be the composition of the appropriate maps from Eqn. (2). The homology class  $\gamma \in H_r(\mathbb{X}_i)$  is said to be born at  $\mathbb{X}_i$  if it is not in the image of  $\varphi_{i-1}^i$ . This same class is said to die at  $\mathbb{X}_j$  if its image in  $H_r(\mathbb{X}_{j-1})$  is not in the image of  $\varphi_{i-1}^{j-1}$ , but its image in  $H_r(\mathbb{X}_j)$  is in the image of  $\varphi_{i-1}^j$ . In the case that the spaces arose from the level sets of a function  $f$  as defined above, we define the persistence of a class  $\gamma$  which is born at  $\mathbb{X}_i = f^{-1}((-\infty, a_i])$  and dies at  $\mathbb{X}_j = f^{-1}((-\infty, a_j])$  to be  $\text{pers}(\gamma) = a_j - a_i$ .

Notice that this equivalence can also be seen from working with persistence modules [7], an abstraction of the definition presented here where persistence is defined at the algebraic level. In fact, given any set of maps between vector spaces,

$$V_1 \longrightarrow V_2 \longrightarrow \dots \longrightarrow V_n,$$

we can analogously define the birth and death of classes in the vector spaces.

In order to visualize the changing homology, we draw a persistence diagram  $d_r$  for each dimension  $r$ . A persistence diagram is a set of points with multiplicity in the upper half plane  $\{(b, d) \in \mathbb{R}^2 \mid d \geq b\}$  along with countably infinite copies of the points on the diagonal  $\Delta = \{(x, x) \in \mathbb{R}^2\}$ . For each class  $\gamma$  which is born at  $\mathbb{X}_i$  and dies at  $\mathbb{X}_j$ , we draw a point at  $(a_i, a_j)$ . A point in the persistence diagram which is close to the diagonal represents a class which was born and died very quickly. A point which is far from the diagonal had a longer life. Depending on the context, this may mean the class is more important, or more telling of the inherent topology of the space. See Fig. 1 for an example.

**2.2. The Space  $(\mathcal{D}_p, W_p)$ .** In order to define a framework for statistics, we will ignore the connection to topological spaces or maps between vector spaces and instead focus on the space of persistence diagrams abstractly.

**Definition 2.1.** *An abstract persistence diagram is a countable multiset of points along with the diagonal,  $\Delta = \{(x, x) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$ , with points in  $\Delta$  having countably infinite multiplicity.*

The distance between these abstract diagrams is the  $p^{\text{th}}$  Wasserstein distance.

**Definition 2.2.** *The  $p^{\text{th}}$  Wasserstein distance between two persistence diagrams  $X$  and  $Y$  is given by*

$$W_p[\sigma](X, Y) := \inf_{\varphi: X \rightarrow Y} \left[ \sum_{x \in X} \sigma(x, \varphi(x))^p \right]^{1/p}$$

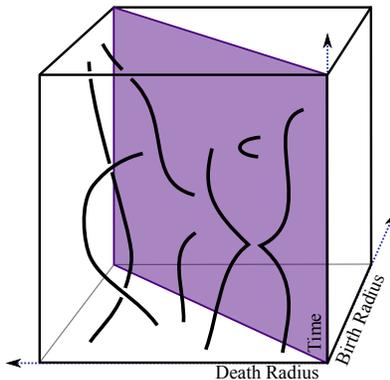


FIGURE 2. An example of a vineyard. For each time, given on the  $z$ -axis, there is a persistence diagram. Since vineyards arising from continuous point clouds are continuous, each point in the diagram traces out a path called a vine. These vines can have endpoints on the starting or ending times, or on the plane which projects to the diagonal.

where  $1 \leq p \leq \infty$ ,  $\sigma$  is a metric on the plane, and  $\varphi$  ranges over bijections between  $X$  and  $Y$ .

We often use  $\sigma = L_q$ . Notice that for  $p = \infty$ ,

$$W_\infty[L_q](X, Y) := \inf_{\varphi: X \rightarrow Y} \sup_{x \in X} \|x - \varphi(x)\|_q.$$

$W_\infty[L_\infty]$  is often referred to as the bottleneck distance. For the majority of this paper, we will be using  $W_2[L_2]$ , which we refer to as  $W_2$  for brevity. We also assume that  $\|\cdot\|$  implies  $L_2$  distance.

**Definition 2.3.** *The space of persistence diagrams  $\mathcal{D}_p$  consists of abstract persistence diagrams with finite distance to the empty diagram  $D_\emptyset$ , which is the diagram which consists of only the points on the diagonal. That is,*

$$\mathcal{D}_p = \{X \mid W_p(X, D_\emptyset) < \infty\}$$

along with the  $p^{\text{th}}$ -Wasserstein metric,  $W_p = W_p[\sigma]$ , from Definition 2.2.

The authors in [21] show that  $(\mathcal{D}_p, W_p[L_\infty])$  is a Polish (complete and separable) space. They also give a description of all of the compact sets in this space. In [28], it is shown that these results hold if we work instead with  $(\mathcal{D}_2, W_2[L_2]) = (\mathcal{D}_2, W_2)$ , with the added benefit that it is a non-negatively curved Alexandrov space (a geodesic space with a lower bound on curvature). Thus, every pair of diagrams has a minimal geodesic between them and this geodesic can be defined using a matching between the diagrams which minimizes Wasserstein distance. So, for the remainder of the paper, we will focus on the space  $(\mathcal{D}_2, W_2)$ .

**2.3. Vineyards.** The first definitions of vineyards [10, 22] were used in the well-behaved case of a homotopy between two functions. In this case, each off-diagonal point of a diagram varies continuously in time and is called a vine. Vines can start and end at off diagonal points at times 0 or 1, or have starting or ending points on the diagonal for any  $t$ , see Fig. 2.

As we do with persistence diagrams, let us consider the space of abstract vineyards to be the space of paths in persistence diagram space.

**Definition 2.4.** *The space of abstract vineyards is*

$$\mathcal{V}_2 = \{v : [0, 1] \rightarrow \mathcal{D}_2 \mid v \text{ is continuous}\},$$

the space of continuous maps from the unit interval to  $\mathcal{D}_2$  where  $v$  is continuous with respect to  $W_2$ .

### 3. FRÉCHET MEANS OF DIAGRAMS

This section reviews previous definitions of the mean of a set of diagrams [21] and an algorithm to compute the mean [28]. We will define the mean of a diagram as the Fréchet mean, give the algorithm for the computation of this mean, and finally present the non-uniqueness problem.

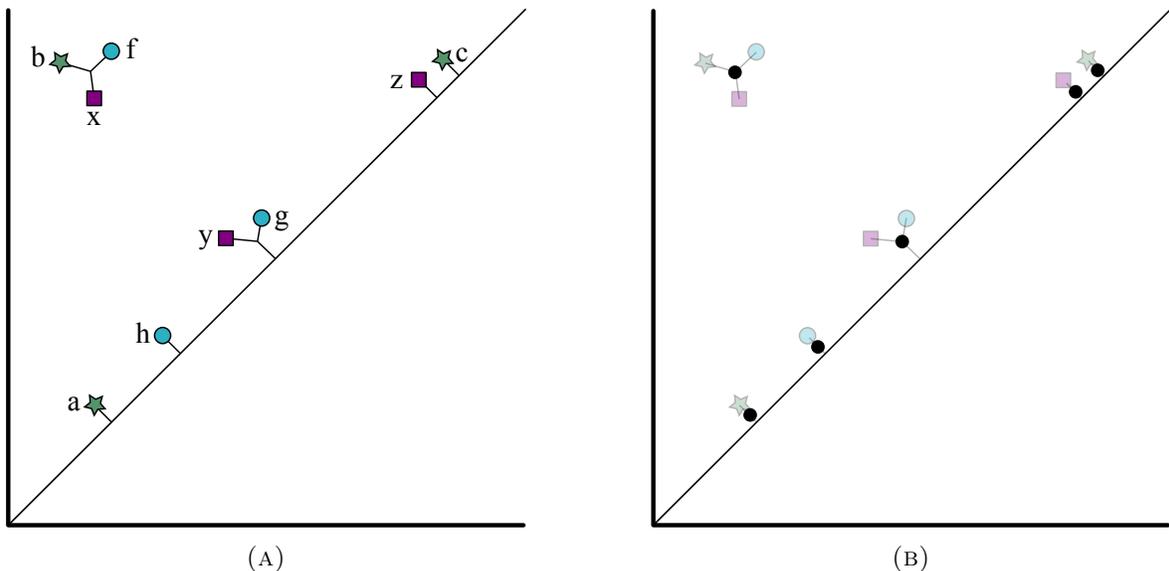


FIGURE 3. An example of a grouping for three overlaid persistence diagrams,  $D_{\blacksquare}$ ,  $D_{\star}$ , and  $D_{\bullet}$  is given in (A). In this example, the grouping has four selections and the corresponding grouping matrix is given in Eqn. 3. The dark circles in diagram (B) give the mean diagram associated to this particular grouping.

**3.1. Fréchet Means.** The Fréchet mean generalizes the mean of a set of points or a distribution in Euclidean space to any metric space. It can be thought of as a generalization of the arithmetic mean in that it minimizes the sum of the square distances to points in the distribution. Given a probability space  $(\mathcal{D}_2, \mathcal{B}(\mathcal{D}_2), \mathcal{P})$  where  $\mathcal{B}(\mathcal{D}_2)$  consists of Borel sets of  $\mathcal{D}_2$ , we can define the Fréchet mean as follows.

**Definition 3.1.** Given a probability space  $(\mathcal{D}_2, \mathcal{B}(\mathcal{D}_2), \mathcal{P})$ ,

$$F_{\mathcal{P}} : \mathcal{D}_2 \longrightarrow \mathbb{R}$$

$$X \longmapsto \int_{\mathcal{D}_2} W_2(X, Y)^2 d\mathcal{P}(Y)$$

is the Fréchet function. The quantity

$$Var_{\mathcal{P}} = \inf_{X \in \mathcal{D}_2} [F_{\mathcal{P}}(X)]$$

is the Fréchet variance of  $\mathcal{P}$  and the set at which the value is obtained

$$\mathbb{E}(\mathcal{P}) = \{X \mid F_{\mathcal{P}}(X) = Var_{\mathcal{P}}\}$$

is the Fréchet expectation, also called Fréchet mean.

The mean in this case need not be a single diagram, but may be a set of diagrams. In fact, there is no guarantee that  $\mathbb{E}(\mathcal{P})$  is even non-empty. However, it was proved in [21] that the Fréchet mean for  $(\mathcal{D}_p, W_p[L_{\infty}])$  is non-empty for certain types of well-behaved probability measures on  $\mathcal{D}_p$ , and this result can be immediately extended to  $(\mathcal{D}_2, W_2[L_2])$  as in [28].

**Theorem 3.2.** Let  $\mathcal{P}$  be a probability measure on  $(\mathcal{D}_2, \mathcal{B}(\mathcal{D}_2))$  with a finite second moment. If  $\mathcal{P}$  has compact support, then  $\mathbb{E}(\mathcal{P}) \neq \emptyset$ .

A similar result holds when the tail probabilities of the distribution  $\mathcal{P}$  decay fast enough, see [21] for details.

**3.2. Matchings, Selections, and Groupings.** The focus of [21] was to develop the probability theory required for statistical procedures on persistence diagrams, including defining a mean. In [28] an algorithm to compute an estimate of the Fréchet mean of a set of diagrams was given. This algorithm centered around understanding an analogue to the Wasserstein distance matching in order to work with more than two diagrams.

The representation of a diagram for the purposes of these definitions is a list of its off-diagonal points,  $X = [x_1, \dots, x_k]$ . We implicitly assume that every diagram has infinitely many copies of the diagonal.

Since our main theorem is stated with regards to diagrams with finitely many off-diagonal points, we will also implicitly assume that this list is finite.

**Definition 3.3.** Let  $X = [x_1, \dots, x_k]$  and  $Y = [y_1, \dots, y_m]$  be diagrams. A matching between  $X$  and  $Y$  is a bijection  $\varphi : X \rightarrow Y$ . An optimal matching is one which attains the Wasserstein distance in Def. 2.2.

We now need to understand how to define matchings when we have  $N$  diagrams instead of just two. For this, we define selections and groupings which restrict to matchings when  $N = 2$ .

**Definition 3.4.** Given a set of diagrams  $X_1, \dots, X_N$ , a selection is a choice of one point from each diagram, where that point could be  $\Delta$ . The trivial selection for a particular off-diagonal point  $x \in X_i$  is the selection  $m_x$  which chooses  $x$  for  $X_i$  and  $\Delta$  for every other diagram.

A grouping is a set of selections so that every off-diagonal point of every diagram is part of exactly one selection.

A grouping for  $N$  diagrams which has  $k$  selections can be stored as a  $k \times N$  matrix  $G$  where entry  $G[j, i] = x$  means that the  $j^{\text{th}}$  selection has point  $x \in X_i$ . See Fig. 3 for an example; in this case, the grouping shown is given by the matrix

$$(3) \quad \begin{matrix} & D_{\star} & D_{\blacksquare} & D_{\bullet} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} & \begin{pmatrix} b & x & f \\ a & \Delta & \Delta \\ \Delta & y & g \\ \Delta & z & \Delta \\ \Delta & \Delta & h \\ c & \Delta & \Delta \end{pmatrix} \end{matrix}$$

where  $\Delta$  represents the diagonal. Note that we consider grouping to be equivalent up to reordering of the selections, and we can add or remove as many  $(\Delta, \Delta, \dots, \Delta)$  rows as we want.

The mean of a selection  $s$  is the point denoted  $\text{mean}(s)$  which minimizes the sum of the square distances to the elements of the selection. When necessary, the notation  $\text{mean}_X(s)$  is used to emphasize the diagram set of interest. The computation of this point will be discussed in Sect. 6.1.

The mean of a grouping,  $\text{mean}(G)$ , is a diagram in  $\mathcal{D}_2$  with a point at the mean of each selection. When it is unclear as to the set of diagrams from which this mean arose, we will denote it as  $\text{mean}_X(G)$ . Note that the mean of the selection yields a point while the mean of a grouping yields a diagram.

It should be noted that there are close ties between these groupings and the Fréchet mean. The diagrams in the Fréchet mean (and all other local minima of the Fréchet function) are  $\text{mean}(G)$  for some grouping  $G$  [28]. Thus, we define an optimal grouping  $G$  to be one such that  $\text{mean}_X(G)$  is in the Fréchet mean.

**3.3. Issues with extensions to Vineyards.** The algorithm given in [28] utilizes the tight relationship between groupings and their means in order to find a local minimum of the Fréchet function. Assume we are working with a set of diagrams  $X_1, \dots, X_N$ . The idea of the algorithm is to find a candidate diagram  $Y$  for the mean, compute a minimal matching for each pair  $(Y, X_i)$ , and build up these matchings into a grouping for the whole set  $X_1, \dots, X_N$ . Then the candidate diagram  $Y$  is replaced with the mean of the computed grouping and the process is repeated until it terminates. See Appendix A for details.

However, it is important to note that the non-uniqueness of optimal groupings leads to non-uniqueness of the Fréchet mean. Consider the example of two diagrams in Fig. 4A. Their mean diagrams are given in 4B. In Fig. 4A, there are two persistence diagrams overlaid:  $D_{\blacksquare}$  has square points 1 and 2,  $D_{\bullet}$  has circle points  $a$  and  $b$ . Since the four points lie exactly on a square, the grouping (which can also be called matching in this instance) to give the Wasserstein distance could either be  $\{(a, 1), (b, 2)\}$  or  $\{(a, 2), (b, 1)\}$ . Thus there are two diagrams which give a minimum of the Fréchet function: the diagram with  $u$  and  $v$ , or the diagram with  $x$  and  $y$ .

If two vineyards pass through this configuration, the mean of the vineyards constructed by finding the mean at each time will not be continuous. Consider for example two vineyards of two points each who start in the grayed configuration of Fig. 4C and move along the dotted line to the darkened configuration. At the bend of the dotted line, the points are at the corners of a square, so as in the

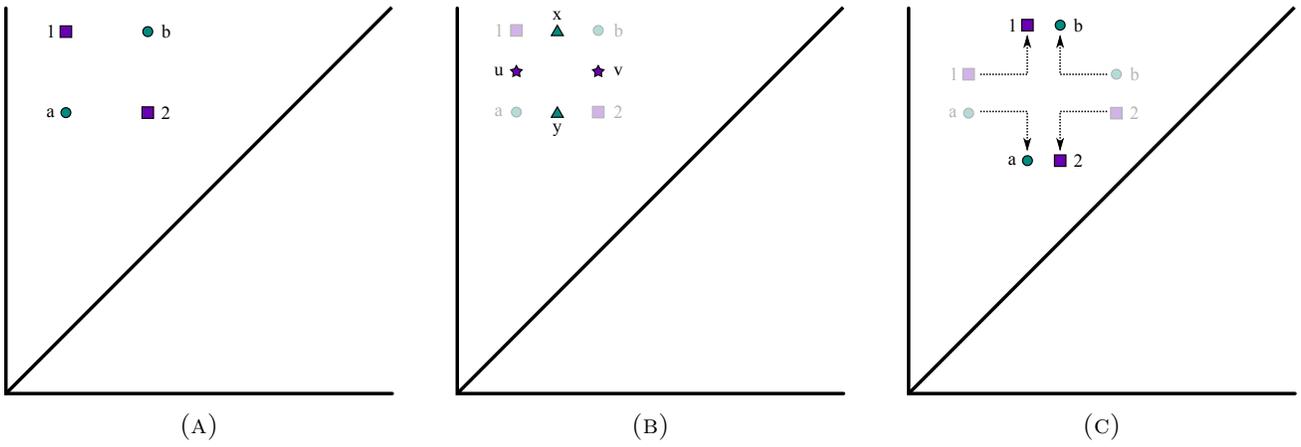


FIGURE 4. A counterexample to uniqueness of the Fréchet mean in  $D_2$  gives an example of issue with means of vineyards. Fig. (A) shows two diagrams overlaid:  $D_{\blacksquare}$  has points 1 and 2,  $D_{\bullet}$  has points  $a$  and  $b$ . Since the grouping given by the Wasserstein distance is not unique, neither is the Fréchet mean. The two possible means are given in Fig. (B): one has points  $x$  and  $y$ , the other has points  $u$  and  $v$ . In Fig. (C), we have two vineyards which pass through the configuration of Fig. (B). The mean is continuous until the points get to the turn of the dotted line, where they form a square, and the mean jumps discontinuously.

example of Fig. 4B, there are two possible choices for the mean. One is close to the means from the previous times, and one is close to the means from the following times.

This means that it is not beneficial to define the mean of a set of vineyards  $\{V_i : [0, 1] \rightarrow \mathcal{D}_2\}$  by using the pointwise Fréchet mean on the set of diagrams  $\{V_i(t)\}$  for a fixed  $t$  as there is no notion of continuity. Thus we must be more creative with our definition.

#### 4. THE MEAN AS A DISTRIBUTION

To overcome the lack of continuity of the mean vineyard and the non-uniqueness illustrated in Fig. 4, we will define the mean of a set of diagrams to be a distribution over persistence diagrams. In order to prove continuity, the diagrams will be limited to  $S_{M,K}$ , the set of diagrams in  $\mathcal{D}_2$  with at most  $K$  off-diagonal points, and all points  $x = (x_1, x_2)$  satisfy  $0 \leq x_1, x_2 \leq M$ .

One important property of  $S_{M,K}$  which we will utilize is that its diameter is finite. For a coarse bound, we note that for a diagram  $X \in S_{M,K}$ , any point is at most distance  $\frac{\sqrt{2}}{2}M$  from the diagonal. Thus,  $W_2(X, D_\emptyset) \leq \frac{\sqrt{2}}{2}KM$ , and so the diameter is bounded by  $\sqrt{2}KM$ . Obviously, however, this is a massive overestimate.

Consider the space  $\mathcal{P}(S_{M,K})$ , the space of probability measures with finite second moment on  $S \subset \mathcal{D}_2$ . This is of course a metric space with the standard probability Wasserstein distance as defined below.

**Definition 4.1.** *The  $p^{\text{th}}$ -Wasserstein distance between two probability distributions,  $\nu$  and  $\eta$ , on metric space  $(\mathbb{X}, d_{\mathbb{X}})$  is*

$$\mathcal{W}_p[d_{\mathbb{X}}](\nu, \eta) = \left[ \inf_{\gamma \in \Gamma(\nu, \eta)} \int_{\mathbb{X} \times \mathbb{X}} d_{\mathbb{X}}(x, y)^p d\gamma(x, y) \right]^{1/p}$$

where  $\Gamma(\nu, \eta)$  is the space of distributions on  $\mathbb{X} \times \mathbb{X}$  with marginals  $\nu$  and  $\eta$  respectively. When  $d_{\mathbb{X}}$  is obvious from context, we will instead write  $\mathcal{W}_p(\nu, \eta)$ .

Thus, we can use  $\mathcal{W}_2[W_2[L_2]]$  as the distance function on  $\mathcal{P}(S_{M,K})$ , where the outside  $\mathcal{W}_2$  is the Wasserstein distance of Def. 4.1 and the inside  $W_2$  is the deterministic Wasserstein distance of Def. 2.2. Note that the map  $Y \rightarrow \delta_Y$ , where  $\delta_Y$  is the delta measure concentrated on the diagram  $Y$ , gives an isometric embedding of  $S_{M,K}$  into  $\mathcal{P}(S_{M,K})$ .

This section is organized as follows. First, we give the intuition for the PFM in Section 4.1, then we discuss the precise definition in Section 4.3, and finally work out an example in Section 4.3.

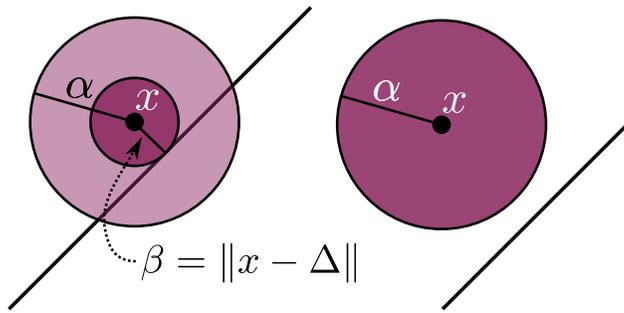


FIGURE 5. The method for drawing points. For a point  $x \in X_i$  where  $\|x - \Delta\| \geq \alpha$  as at right, a point is drawn from the uniform distribution on the ball of radius  $\alpha$  centered at  $x$ . This point is then added to the diagram  $X'_i$ . For a point  $x \in X_i$  where  $\|x - \Delta\| < \alpha$  as at left, a point is still drawn from ball of radius  $\alpha$ , however the point is only added to  $X'_i$  if it is inside the ball of radius  $\beta = \|x - \Delta\|$  centered at  $x$ .

**4.1. Intuition.** We first give an intuitive description of the main ideas behind the new definition. The basic ideas we use to achieve continuity of a mean diagram is to think of diagrams as probabilistic objects and groupings between diagrams as probabilistic objects. We track the probabilities of different groupings being optimal for of a set of diagrams drawn by perturbing the points in the original input diagrams. The mean of a set of diagrams is not a diagram but a distribution over diagrams which are each the minimizer for some grouping. The weight on a diagram is the probability that the corresponding grouping is optimal.

The Fréchet mean is generically unique. More precisely, the measure of sets of diagrams in  $(S_{M,K})^N$  with non-unique Fréchet means is of measure zero [28]. However, the Fréchet mean is not stable. To see this, consider a slight perturbation of the point configuration in Fig. 4A. The result is a mean which contains exactly one of  $D_\blacktriangle$  and  $D_\blackstar$  since the perturbation will result in exactly one groupings to be optimal.

To address this problem, we consider diagrams as probabilistic objects. So, given a diagram with labeled off diagonal points  $\{p_1, \dots, p_\ell\}$  we consider the collection as a probability density function on sets of labeled points  $\{x_1, x_2, \dots, x_\ell\} \subset \mathbb{R}^2 \cup \Delta$  where each  $x_i$  is a perturbation of  $p_i$  and can be either off the diagonal or a copy of the diagonal. This is the "trembling hands" of the points considered as players in game theory; they sometimes play a slightly different strategy than the one they intended to play.

We will use the density function for an individual perturbation to be a linear combination of a uniform distribution over a small ball centered at  $p_i$  and the Dirac function over the diagonal. This is for ease of calculations and there is no theoretical restriction.

For example, consider Fig. 4A. After slightly perturbing the points in the diagrams, the probability of the two optimal matches is about equal. This will result in a mean diagram of

$$p_\blacktriangle \cdot \delta_{D_\blacktriangle} + p_\blackstar \delta_{D_\blackstar},$$

where  $p_\blacktriangle = 1 - p_\blackstar \approx .5$ .

In general, if  $X = \{X_1, \dots, X_N\}$  is a set of diagrams from  $S_{M,K}$ , we define its mean to be the following distribution on  $S_{M,NK}$ :

**Definition 4.2.** *The probabilistic Fréchet mean (PFM) for a set of diagrams  $\{X_1, \dots, X_n\}$  is the distribution given by*

$$\mu_X = \sum_G \mathbb{P}(\mathcal{H}_X = G) \cdot \delta_{\text{mean}_X(G)}.$$

Here the sum is taken over all possible groupings  $G$  on the set of diagrams, and  $\text{mean}_X(G)$  is the mean diagram for the specific grouping  $G$ . The weights  $\mathbb{P}(\mathcal{H} = G)$  are derived from a random variable  $\mathcal{H}$  which can be thought of either as a probabilistic grouping where each point in the input diagram is replaced with a localized distribution centered on the point or as the probability that a stochastic perturbation of the input diagrams would lead to  $G$  being the optimal grouping. We now explain this in more detail.

**4.2. The Definition of  $\mathcal{H}$ .** We are given a set  $X = \{X_1, \dots, X_N\}$  of diagrams from  $S_{M,K}$ . We now define  $\mathcal{H}$ , a grouping valued random variable. As with other notation, we will write  $\mathcal{H}_X$  when it is necessary to distinguish the diagrams of interest.

First, choose parameter  $\alpha > 0$  which determines how much we perturb the points. Label all the points within the  $X_i$ . When we perturb the points in the  $X_i$  remember the labels. We need to define the distributions  $\eta_x$  for each  $x \in X_i$  which gives the method for perturbing the off-diagonal points of the persistence diagrams. This is based on the uniform distribution over the ball of radius  $\alpha$ , but must be modified if the initial point is within  $\alpha$  of the diagonal. This can be thought of as instead adding a point on the diagonal if the drawn  $x'$  is outside of  $B(x, \|x - \Delta\|)$ ; however, since we only keep track of off-diagonal points for storing a persistence diagram, this point can safely be ignored. Note that if  $x$  is more than  $\alpha$  away from the diagonal, an off-diagonal point is always added to the diagram. However, the probability that an off-diagonal point gets added decreases as the distance to the diagonal decreases. See Fig. 5.

Let  $x \in X_i$  and set  $r = \min\{\alpha, \|x - \Delta\|\}$ . Define the density function on  $\mathbb{R}^2 \cup \Delta$  for  $x$  as

$$\eta_x = \frac{1}{\pi\alpha^2} \mathbf{1}_{B(x,r)} + \frac{\alpha^2 - r^2}{\alpha^2} \delta_\Delta.$$

If we wished, we could work with a more general distribution such as a normal distribution restricted to a disc  $B(x, r)$ ; however, for clarity of the proof we will use the uniform distribution.

We now draw perturbations  $X'_1, \dots, X'_N$  of original input diagrams  $X_1, X_2, \dots, X_N$  by drawing the  $x'$  from the corresponding density functions  $\eta_x$ . We are interested in the optimal groupings for these diagrams where we have kept track of the labeling. Since each point in a draw of  $X'_i$  is associated to a point in  $X_i$ , we associate a grouping of the  $\{X'_i\}$  with the grouping using the corresponding points of  $X_i$ . To make sure this is well-defined, whenever a copy of the diagonal is used in a selection within a grouping for  $X'_1, X'_2, \dots, X'_N$  we assume that it came from an unlabeled copy of the diagonal (not from an off-diagonal point in some  $X_i$ ). Exactly the points in the  $X_i$  which are sent to an off-diagonal point in  $X'_i$  are contained in some selection. Some points in the  $X_i$  did not get corresponding off-diagonal points in the draw  $X'_i$  and therefore are not represented in the selections. We can extend this set of selections to a full grouping by adding the trivial selection for these points. That is, if a point  $x \in X_i$  did not lead to a off-diagonal point in  $X'_i$ , we add the selection which chooses  $x$  for  $X_i$  and  $\Delta$  for every other diagram.

Note that the diagram  $\text{mean}_X(G)$ , where  $X \in (S_{M,K})^N$  can have at most  $NK$  points and those points will be contained in a box of size  $M$  (as their coordinates are an affine combination of values in  $[0, M]$ ). Thus,  $\mu_X$  is an element of  $\mathcal{P}(S_{M,NK})$ .

With probability one there is a unique optimal grouping for  $X'_1, \dots, X'_N$  [27]. This implies that we have constructed a grouping valued random element which we call  $\mathcal{H}$ .

**4.3. Example.** Here is an example to make the discussion above a little more clear. Consider the three overlaid diagrams in Fig. 6A. Points are drawn in the ball of radius  $\alpha$  centered at each point. Since  $a, c, h, g, y$ , and  $z$  are near the diagonal there is a chance that the diagonal is drawn for them. In this particular draw, given in Fig. 6B, the diagonal is drawn for  $a$  and  $c$ .

For the diagrams in Fig. 6B, the optimal grouping shown is

$$(4) \quad \begin{array}{c} D'_\star \quad D'_\square \quad D'_\circ \\ 1 \left( \begin{array}{ccc} b' & x' & f' \\ \Delta & y' & g' \\ \Delta & \Delta & h' \\ \Delta & z' & \Delta \end{array} \right) \\ 2 \\ 3 \\ 4 \end{array}.$$

9

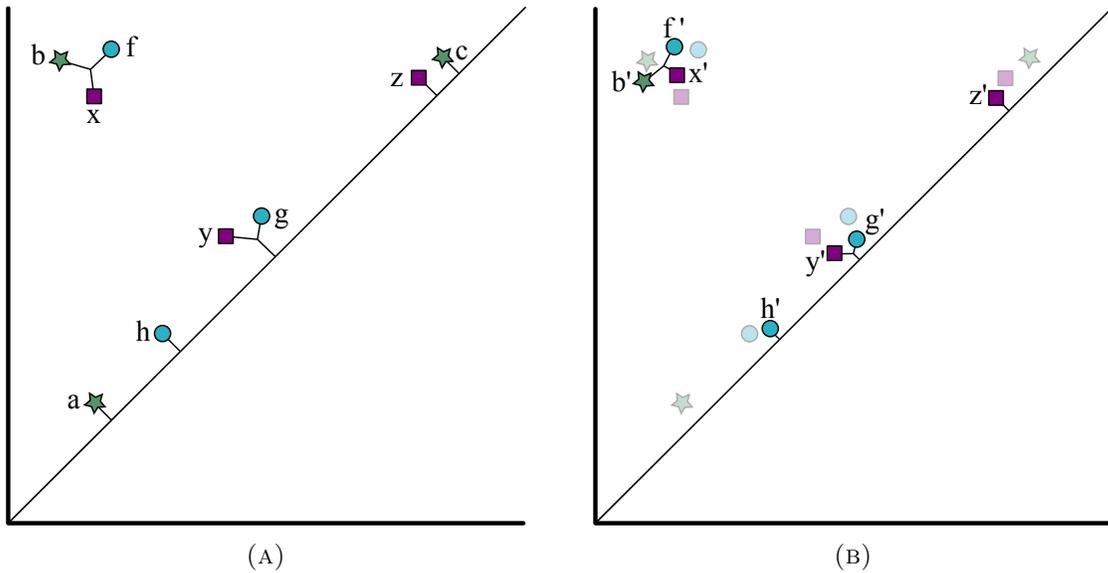


FIGURE 6. An example of corresponding groupings for a given draw. The original diagrams are  $D_{\star}$ ,  $D_{\square}$  and  $D_{\circ}$  in Fig. (A). A point is drawn near each point away from the diagonal, and points are drawn for some points near the diagonal to construct  $D'_{\star}$ ,  $D'_{\square}$  and  $D'_{\circ}$  in Fig. (B). The grouping for the mean of these three diagrams is computed using Algorithm 1 and the associated grouping is given in Eqn. 4. Then the grouping is converted to a grouping for  $D_{\star}$ ,  $D_{\square}$  and  $D_{\circ}$  in Eqn. 5 and drawn in Fig. (A).

So, to find the corresponding grouping for the original diagrams, we replace each point with its corresponding point, and add in the trivial selection for the points that were not chosen:

$$(5) \quad \begin{matrix} & D_{\star} & D_{\square} & D_{\circ} \\ 1 & \left( \begin{array}{ccc} b & x & f \\ \Delta & y & g \\ \Delta & \Delta & h \\ \Delta & z & \Delta \\ a & \Delta & \Delta \\ c & \Delta & \Delta \end{array} \right) \end{matrix}.$$

## 5. CONTINUITY

In this section, we prove our main theorem: that the mean distribution varies continuously when faced with a continuously varying set of input diagrams. We first must give  $(S_{M,K})^N$  and  $\mathcal{P}(S_{M,NK})$  metrics.

**Definition 5.1.** Let  $X = \{X_1, \dots, X_N\}$  and  $Y = \{Y_1, \dots, Y_N\}$  be elements of  $(S_{M,K})^N$ . Then the space  $(S_{M,K})^N$  is given the metric  $\vec{d}_2(X, Y) = \left( \sum_{i=1}^N W_2(X_i, Y_i)^2 \right)^{1/2}$ .

**Definition 5.2.** The space  $\mathcal{P}(S_{M,NK})$  consists of distributions with the Wasserstein metric  $\mathcal{W}_2$  of Def. 4.1.

With these structures, we can state our main theorem.

**Theorem 5.3.** Let  $X = (X_1, \dots, X_N)$  and  $Y = (Y_1, \dots, Y_N)$  denote sets of diagrams in  $(S_{M,K})^N$  with PFM's  $\mu_X$  and  $\mu_Y$  respectively and let  $\varphi_i : X_i \rightarrow Y_i$  for  $i \in \{1, \dots, N\}$  be any set of optimal matchings between the pairs. Let  $\tilde{X}_i$  be the diagram consisting of points  $x \in X_i$  such that  $\varphi_i(x) = \Delta$ . Likewise, let  $\tilde{Y}_i$  be the diagram consisting of points  $y \in Y_i$  such that  $\varphi^{-1}(y) = \Delta$ .

Then

$$\mathcal{W}_2(\mu_X, \mu_Y) \leq C \left( \sum_{x \in \tilde{X}} \|x - \varphi(x)\| + \vec{d}_2(X, Y)^2 \right)^{\frac{1}{2}}$$

$$\text{where } C = \sqrt{\left(2 \left(\frac{1}{N^2} + \frac{\bar{M}^2}{\alpha^2}\right) + \frac{4\bar{M}^2}{\alpha} + 1\right)}.$$

Before delving into the proof, we note that there are two major results that are now corollaries to this theorem. First, we have that the map which takes a set of diagrams to its PFM is Hölder continuous.

**Corollary 5.4.** *Let*

$$\begin{aligned} \Phi : (S_{M,K})^N &\longrightarrow \mathcal{P}(S_{M,NK}) \\ (X_1, \dots, X_N) &\longmapsto \mu_X \end{aligned}$$

be the map which sends a set of diagrams to its PFM. Then  $\Phi$  is Hölder continuous with exponent  $1/2$  and constant  $C' = \max\{C\sqrt{NK+1}, \bar{M}\}$ . That is,

$$\mathcal{W}_2(\mu_X, \mu_Y) \leq C' \sqrt{\vec{d}_2(X, Y)}$$

for all  $X, Y \in (S_{M,K})^N$ .

*Proof.* First, notice that for any  $x \in \tilde{X}_i$ ,  $\|x - \varphi(x)\| \leq W_2(X_i, Y_i) \leq \vec{d}_2(X, Y)$  and the number of off-diagonal points in  $\tilde{X}$  is at most  $NK$ . Thus, if  $\vec{d}_2(X, Y) < 1$ , the theorem implies that

$$\begin{aligned} \mathcal{W}_2(\mu_X, \mu_Y) &\leq C \left( NK \vec{d}_2(X, Y) + \vec{d}_2(X, Y)^2 \right)^{1/2} \\ &\leq C(NK + 1)^{1/2} \sqrt{\vec{d}_2(X, Y)}. \end{aligned}$$

If  $\vec{d}_2(X, Y) \geq 1$ , then we note that since  $\mathcal{W}_2(\mu_X, \mu_Y) \leq \bar{M}$ ,  $\mathcal{W}_2(\mu_X, \mu_Y) \leq \bar{M} \sqrt{\vec{d}_2(X, Y)}$  trivially.  $\square$

Secondly, if we define a mean pointwise for a set of vineyards, we can stitch the results together into a continuous path in  $\mathcal{P}(S_{M,NK})$ . This result is immediate given the previous corollary.

**Corollary 5.5.** *Let  $\gamma_1, \dots, \gamma_N : [0, 1] \rightarrow S_{M,K}$  be vineyards in  $\mathcal{V}_2$  restricted to  $S_{M,K}$ . Then*

$$\begin{aligned} \mu_\gamma : [0, 1] &\longrightarrow \mathcal{P}(S_{M,NK}) \\ t &\longmapsto \mu_{\gamma_1(t), \dots, \gamma_N(t)} \end{aligned}$$

is continuous.

## 6. PROOF OF THEOREM 5.3

The proof of Thm. 5.3 is organized as follows. We start by discussing means of selections in Section 6.1. Then, in order to prove the theorem, we need to match groupings for  $X$  to groupings for  $Y$ . We do this in Section 6.2 by first working with those off-diagonal points in  $X$  which are close to off-diagonal points in  $Y$  where there is an obvious association between groupings in  $X$  and groupings in  $Y$ . Then, the issue of points matched to the diagonal is discussed in Section 6.3, and finally, the full proof is given in Section 6.4.

**6.1. Means of selections.** Consider the mean of the selection  $s$  consisting of  $N$  points:  $\{p_1, \dots, p_k\}$  with  $p_i = (x_i, y_i)$  off-diagonal, and  $N - k$  copies of the diagonal  $\Delta$ . A quick computation gives this point as

$$(6) \quad \text{mean}_X(s) = \frac{1}{2Nk} \left( (N+k) \sum_i x_i + (N-k) \sum_i y_i, \right. \\ \left. (N-k) \sum_i x_i + (N+k) \sum_i y_i \right).$$

Sometimes it may be simpler to consider the mean of two selections in rotated coordinates with axes  $(1/\sqrt{2}, 1/\sqrt{2})$  and  $(-1/\sqrt{2}, 1/\sqrt{2})$ . Writing  $p_i = (a_i, b_i)$  in these coordinates, Eqn. 6 becomes

$$\text{mean}_X(s) = \left( \frac{1}{k} \sum_{i=1}^k a_i, \frac{1}{N} \sum_{i=1}^k b_i \right).$$

In these coordinates, it is easier to see what happens for the mean of a trivial selection.

**Lemma 6.1.** *For any off-diagonal point  $x$ ,  $\|\text{mean}(s_x) - \Delta\| = \frac{1}{N}\|x - \Delta\|$ .*

*Proof.* If the single off-diagonal point is at  $x = (a, b)$  in the rotated coordinates and there are a total of  $N$  diagrams,

$$\text{mean}_X(s_x) = \left( a, \frac{1}{N}b \right)$$

so the distance to the diagonal is minimized at the point  $(a, 0)$ , again in rotated coordinates. Thus  $\|x - \Delta\| = b$  and hence

$$\|\text{mean}(s_x) - \Delta\|^2 = \left(\frac{1}{N}b\right)^2 = \frac{1}{N^2}b^2 = \frac{1}{N^2}\|x - \Delta\|^2.$$

□

To conclude the section, we show that the distance between points which are the means of selections is bounded by the distance between the points which build the selections.

**Lemma 6.2.** *Let  $s$  be a selection of points  $z_1, z_2, \dots, z_N$  and  $\hat{s}$  a selection of points  $\hat{z}_1, \hat{z}_2, \dots, \hat{z}_N$ . Further, assume  $z_1, \dots, z_k$  and  $\hat{z}_1, \dots, \hat{z}_k$  are all off-diagonal for  $1 \leq k \leq N$  and  $z_{k+1} = \dots = z_N = \hat{z}_{k+1} = \dots = \hat{z}_N = \Delta$ . Then*

$$\|\text{mean}(s) - \text{mean}(\hat{s})\|^2 \leq \sum_{i=1}^k \|z_i - \hat{z}_i\|^2.$$

*Proof.* Consider the means of the selections  $s$  and  $\hat{s}$  in rotated coordinates with axes  $(1/\sqrt{2}, 1/\sqrt{2})$  and  $(-1/\sqrt{2}, 1/\sqrt{2})$ . Writing  $z_i = (u_i, v_i)$  and  $\hat{z}_i = (\hat{u}_i, \hat{v}_i)$  in these coordinates, we have

$$\begin{aligned} \text{mean}_Z(s) &= \left( \frac{1}{k} \left( \sum_{i=1}^k u_i \right), \frac{1}{N} \left( \sum_{i=1}^k v_i \right) \right), \\ \text{mean}_{\hat{Z}}(\hat{s}) &= \left( \frac{1}{k} \left( \sum_{i=1}^k \hat{u}_i \right), \frac{1}{N} \left( \sum_{i=1}^k \hat{v}_i \right) \right), \end{aligned}$$

hence, using Cauchy-Schwartz,

$$\begin{aligned} \|\text{mean}_Z(s) - \text{mean}_{\hat{Z}}(\hat{s})\|^2 &= \left( \frac{1}{k} \sum_{i=1}^k (u_i - \hat{u}_i) \right)^2 + \left( \frac{1}{N} \sum_{i=1}^k (v_i - \hat{v}_i) \right)^2 \\ &\leq \frac{1}{k} \sum_{i=1}^k (u_i - \hat{u}_i)^2 + \frac{1}{N} \sum_{i=1}^k (v_i - \hat{v}_i)^2 \\ &\leq \sum_{i=1}^k (u_i - \hat{u}_i)^2 + \sum_{i=1}^k (v_i - \hat{v}_i)^2 \\ &= \sum_{i=1}^k \|z_i - \hat{z}_i\|^2. \end{aligned}$$

□

**6.2. Proof for off-diagonal points.** In this section, our goal is Prop. 6.5 where we bound the distance between the PFM of diagrams where we do not have to worry about the issue of points close to the diagonal. Let  $X = \{X_1, X_2, \dots, X_N\}$  and  $Y = \{Y_1, Y_2, Y_3, \dots, Y_N\}$  be in  $(S_{M,K})^N$  where  $\mathcal{G}(X)$  and  $\mathcal{G}(Y)$  are the sets of possible groupings for  $X$  and  $Y$  respectively. For the sake of notation, let  $\bar{X} = \bigcup_i X_i$  and  $\bar{Y} = \bigcup_i Y_i$  be the set of all off-diagonal points in the input sets

Let  $\varphi_i : X_i \rightarrow Y_i$  be optimal matching and consider the case where  $X_i$  and  $Y_i$  have the same number of off-diagonal points and that  $\varphi_i$  maps off-diagonal points to off-diagonal points. By remembering the labeling and using the fact that  $\varphi$  matches off-diagonal points to off-diagonal points,  $\varphi$  induces a bijection  $\varphi : \mathcal{G}(X) \rightarrow \mathcal{G}(Y)$  which we can use to associate the probability masses in order to construct a transportation plan for Prop. 6.5.

First, we bound the difference in probability for associated groupings under  $\varphi$ .

**Lemma 6.3.** *Let  $X = \{X_1, X_2, \dots, X_N\}$  and  $Y = \{Y_1, Y_2, Y_3, \dots, Y_N\}$  be in  $(S_{M,K})^N$ . Let  $\varphi_i : X_i \rightarrow Y_i$  be optimal matchings, which can be thought of as  $\varphi : \bar{X} \rightarrow \bar{Y}$ . Let us assume that  $X_i$  and  $Y_i$  have the same number of off-diagonal points and that  $\varphi_i$  maps off-diagonal points to off-diagonal points. Then*

$$\sum_{G \in \mathcal{G}(X)} \max \{ \mathbb{P}(\mathcal{H}_X = G) - \mathbb{P}(\mathcal{H}_Y = \varphi(G)), 0 \} \leq \frac{4}{\alpha} \sum_{x \in \bar{X}} \|x - \varphi(x)\|.$$

*Proof.* The proof is immediate whenever  $\|x - \varphi(x)\| \geq \alpha$  for some  $x \in X$  as

$$\sum_{G \in \mathcal{G}(X)} \max \{ \mathbb{P}(\mathcal{H}_X = G) - \mathbb{P}(\mathcal{H}_Y = \varphi(G)), 0 \} \leq \sum_{G \in \mathcal{G}(X)} \mathbb{P}(\mathcal{H}_X = G) \leq 1.$$

Thus, assume that  $\|x - \varphi(x)\| < \alpha$  for all  $x \in X$ .

For inputs  $X = \{X_1, X_2, \dots, X_N\}$  and  $Y = \{Y_1, Y_2, \dots, Y_N\}$ , we have probability density functions  $f_X$  and  $f_Y$  for the population of  $N$  diagrams drawn by suitably perturbing the points within them. The random grouping valued element is determined by the populations of diagrams drawn by  $f_X$  and  $f_Y$ . Thus, to bound  $\sum_{G \in \mathcal{G}(X)} \max \{ \mathbb{P}(\mathcal{H}_X = G) - \mathbb{P}(\mathcal{H}_Y = \varphi(G)), 0 \}$  it is enough to bound

$$\int_{S_{M+\alpha, NK}^N} (f_X - f_Y)_+ d\rho.$$

Let  $m$  be the total number of points in the diagrams  $X_i$ , so  $m = \#\{\bar{X}\}$ . Construct a sequence of populations of diagrams  $X = Z^0, Z^1, \dots, Z^m = Y$  where at each stage, a point  $x$  is moved to  $\varphi(x)$  while fixing the rest of the points. Let  $f_k$  denote the probability density function over  $S_{M+\alpha, NK}^N$  corresponding to  $Z^k$ . Since

$$(7) \quad \int_{S_{M+\alpha, NK}^N} (f_X - f_Y)_+ d\rho \leq \int_{S_{M+\alpha, NK}^N} |f_X - f_Y| d\rho \leq \sum_{k=1}^m \int_{S_{M+\alpha, NK}^N} |f_{k-1} - f_k| d\rho,$$

we wish to bound  $\int |f_{k-1} - f_k| d\rho$  in terms of  $\|x - \varphi(x)\|$  where  $x$  is the point that was moved between  $Z^{k-1}$  and  $Z^k$ .

Let  $r_x = \min\{\alpha, \|x - \Delta\|\}$ . As the perturbations of the points in the diagrams are independent we can integrate out the effects of the other points in the diagrams. Therefore,

$$\begin{aligned} \int_{S_{M+\alpha, NK}^N} |f_{k-1} - f_k| d\rho &= \int_{\mathbb{R}^2 \cup \Delta} |\eta_x - \eta_{\varphi(x)}| d\rho \\ &\leq \int_{\mathbb{R}^2} \left| \frac{1}{\pi\alpha^2} \mathbf{1}_{B(x, r_x)} - \frac{1}{\pi\alpha^2} \mathbf{1}_{B(x_{\varphi(x)}, r_{\varphi(x)})} \right| d\rho + \left| \frac{\alpha^2 - r_x^2}{\alpha^2} - \frac{\alpha^2 - r_{\varphi(x)}^2}{\alpha^2} \right| \\ &\leq \frac{1}{\pi\alpha^2} \int_{\mathbb{R}^2} \mathbf{1}_{B(x, r_x) \Delta B(x_{\varphi(x)}, r_{\varphi(x)})} d\rho + \frac{|r_x^2 - r_{\varphi(x)}^2|}{\alpha^2} \end{aligned}$$

where  $U \Delta V$  denotes the symmetric difference of  $U$  and  $V$ .

Now,  $B(x, r_x) \cap B(x_{\varphi(x)}, r_{\varphi(x)})$  contains a ball with diameter  $r_x + r_{\varphi(x)} - \|x - \varphi(x)\|$ . Note also that  $r_x, r_{\varphi(x)} \leq \alpha$ , and  $|r_x - r_{\varphi(x)}| \leq \|x - \varphi(x)\|$ . Then

$$\begin{aligned} \int_{\mathbb{R}^2} \mathbf{1}_{B(x, r_x) \Delta B(x_{\varphi(x)}, r_{\varphi(x)})} d\rho &\leq \pi r_x^2 + \pi r_{\varphi(x)}^2 - 2\pi \left( \frac{1}{2}(r_x + r_{\varphi(x)} - \|x - \varphi(x)\|) \right)^2 \\ &= \frac{\pi}{2} \left( (r_x - r_{\varphi(x)})^2 + \|x - \varphi(x)\| (2r_x + 2r_{\varphi(x)} - \|x - \varphi(x)\|) \right) \\ &\leq \frac{\pi}{2} \left( \|x - \varphi(x)\|^2 + \|x - \varphi(x)\| (4\alpha - \|x - \varphi(x)\|) \right) \\ &\leq 2\pi\alpha \|x - \varphi(x)\| \end{aligned}$$

and therefore

$$\begin{aligned} \int_{S_{M+\alpha, NK}^N} |f_{k-1} - f_k| d\rho &\leq \frac{\pi\alpha \|x - \varphi(x)\|}{\pi\alpha^2} + \frac{|r_x^2 - r_{\varphi(x)}^2|}{\alpha^2} \\ &\leq \frac{2\|x - \varphi(x)\|}{\alpha} + \frac{2\alpha\|x - \varphi(x)\|}{\alpha^2} \\ &\leq \frac{4\|x - \varphi(x)\|}{\alpha}. \end{aligned}$$

Together with Eqn. (7) we can conclude that

$$\int_{S_{M+\alpha, NK}^N} |f_X - f_Y| d\rho \leq \sum_{x \in X} \frac{4\|x - \varphi(x)\|}{\alpha}.$$

□

Next, we use the previous lemma to bound the distance between the mean diagrams for associated groupings.

**Lemma 6.4.** *Let  $X = \{X_1, X_2, \dots, X_N\}$  and  $Y = \{Y_1, Y_2, Y_3, \dots, Y_N\}$  be in  $(S_{M,K})^N$ . Let  $\varphi_i : X_i \rightarrow Y_i$  be an optimal matching. Assume that  $X_i$  and  $Y_i$  have the same number of off-diagonal points and that  $\varphi_i$  maps off-diagonal points to off-diagonal points. Then*

$$W_2(\text{mean}_X G, \text{mean}_Y \varphi(G)) \leq \vec{d}_2(X, Y)$$

for all  $G \in \mathcal{G}$ .

*Proof.* Let  $m_1, \dots, m_\ell$  be the selections in  $G$ . Thus  $\varphi(m_1), \dots, \varphi(m_\ell)$  are the selections of  $\varphi(G)$ ,  $\text{mean}(m_1), \dots, \text{mean}(m_\ell)$  are the off-diagonal points of  $\text{mean}_X(G)$  and  $\text{mean}(\varphi(m_1)), \dots, \text{mean}(\varphi(m_\ell))$  are the off-diagonal points of  $\text{mean}_Y(\varphi(G))$ . Define a bijection  $\psi : \text{mean}_X G \rightarrow \text{mean}_Y \varphi(G)$  by sending  $\text{mean}(m_i)$  to  $\text{mean}(\varphi(m_i))$ . Thus,

$$\begin{aligned} W_2(\text{mean}_X G, \text{mean}_Y \varphi(G))^2 &= \sum_{i=1}^{\ell} \|\text{mean}(m_i) - \text{mean}(\varphi(m_i))\|^2 \\ &\leq \sum_{i=1}^{\ell} \sum_{x \in m_i} \|x - \varphi(x)\|^2 \\ &= \sum_{i=1}^N \sum_{x \in X_i} \|x - \varphi(x)\|^2 \\ &= \sum_{i=1}^N W_2(X_i, Y_i)^2 \\ &= \vec{d}_2(X, Y)^2. \end{aligned}$$

□

Finally, we can bound the distance between the PFMs for the sets of diagrams.

**Proposition 6.5.** Let  $X = \{X_1, X_2, \dots, X_N\}, Y = \{Y_1, Y_2, Y_3, \dots, Y_N\} \in (S_{M,K})^N$  with PFM's  $\mu_X, \mu_Y \in \mathcal{P}(S_{M,NK})$  respectively. Let  $\varphi_i : X_i \rightarrow Y_i$  be optimal matchings. Further, assume that  $X_i$  and  $Y_i$  have the same number of off-diagonal points and that  $\varphi_i$  maps off-diagonal points to off-diagonal points. Then

$$\mathcal{W}_2(\mu_X, \mu_Y) \leq \left( \frac{4\bar{M}^2}{\alpha} \sum_{x \in X} \|x - \varphi(x)\| \right)^{1/2} + \vec{d}_2(X, Y)$$

where  $\bar{M}$  is the maximal distance between diagrams in  $S_{M,NK}$ .

*Proof.* Let  $\mathcal{G}(X)$  be the set of groupings on the set of diagrams  $X = \{X_1, X_2, \dots, X_N\}$  and  $\mathcal{G}(Y)$  the set of diagrams  $Y = \{Y_1, Y_2, \dots, Y_N\}$ . Recall that

$$\mu_X = \sum_{G \in \mathcal{G}(X)} \mathbb{P}(\mathcal{H}_X = G) \delta_{\text{mean}_X G}.$$

Using the bijection  $\varphi : \mathcal{G}(X) \rightarrow \mathcal{G}(Y)$ , we can write

$$\mu_Y = \sum_{G \in \mathcal{G}(X)} \mathbb{P}(\mathcal{H}_Y = \varphi(G)) \delta_{\text{mean}_Y \varphi(G)}.$$

Thus

$$\begin{aligned} \mathcal{W}_2(\mu_X, \mu_Y) &= \mathcal{W}_2 \left( \sum_{G \in \mathcal{G}(X)} \mathbb{P}(\mathcal{H}_X = G) \delta_{\text{mean}_X G}, \sum_{G \in \mathcal{G}(X)} \mathbb{P}(\mathcal{H}_Y = \varphi(G)) \delta_{\text{mean}_Y \varphi(G)} \right) \\ (8) \quad &\leq \mathcal{W}_2 \left( \sum_{G \in \mathcal{G}(X)} \mathbb{P}(\mathcal{H}_X = G) \delta_{\text{mean}_X G}, \sum_{G \in \mathcal{G}(X)} \mathbb{P}(\mathcal{H}_Y = \varphi(G)) \delta_{\text{mean}_X G} \right) \\ &\quad + \mathcal{W}_2 \left( \sum_{G \in \mathcal{G}(X)} \mathbb{P}(\mathcal{H}_Y = \varphi(G)) \delta_{\text{mean}_X G}, \sum_{G \in \mathcal{G}(X)} \mathbb{P}(\mathcal{H}_Y = \varphi(G)) \delta_{\text{mean}_Y \varphi(G)} \right) \end{aligned}$$

by the triangle inequality.

We want to bound the first term,  $\mathcal{W}_2 \left( \sum_{G \in \mathcal{G}(X)} \mathbb{P}(\mathcal{H}_X = G) \delta_{\text{mean}_X G}, \sum_{G \in \mathcal{G}(X)} \mathbb{P}(\mathcal{H}_Y = \varphi(G)) \delta_{\text{mean}_X G} \right)$ , by constructing a transportation plan which keeps most of the mass stationary. Consider the following plan:

- If  $\mathbb{P}(\mathcal{H}_X = G) \leq \mathbb{P}(\mathcal{H}_Y = \varphi(G))$  then keep all the mass at  $\delta_{\text{mean}_X G}$  at the same spot.
- If  $\mathbb{P}(\mathcal{H}_X = G) > \mathbb{P}(\mathcal{H}_Y = \varphi(G))$  then keep  $\mathbb{P}(\mathcal{H}_Y = \varphi(G))$  worth of mass at  $\delta_{\text{mean}_X G}$  and redistribute the rest as needed.

The amount of mass that moves is then  $\sum_{G \in \mathcal{G}} \max\{\mathbb{P}(\mathcal{H}_X = G) - \mathbb{P}(\mathcal{H}_Y = \varphi(G)), 0\}$ . By Lemma 6.3,

$$\sum_{G \in \mathcal{G}(X)} \max\{\mathbb{P}(\mathcal{H}_X = G) - \mathbb{P}(\mathcal{H}_Y = \varphi(G)), 0\} \leq \frac{4}{\alpha} \sum_{x \in X} \|x - \varphi(x)\|.$$

The distance between diagrams is bounded by  $\bar{M}$ . Therefore,

$$\mathcal{W}_2 \left( \sum_{G \in \mathcal{G}(X)} \mathbb{P}(\mathcal{H}_X = G) \delta_{\text{mean}_X G}, \sum_{G \in \mathcal{G}(X)} \mathbb{P}(\mathcal{H}_Y = \varphi(G)) \delta_{\text{mean}_X G} \right)^2 \leq \frac{4\bar{M}^2}{\alpha} \sum_{x \in X} \|x - \varphi(x)\|.$$

Focusing on the second term of Eqn. 8,

$$\begin{aligned} \mathcal{W}_2 \left( \sum_{G \in \mathcal{G}(X)} \mathbb{P}(\mathcal{H}_Y = \varphi(G)) \delta_{\text{mean}_X G}, \sum_{G \in \mathcal{G}(X)} \mathbb{P}(\mathcal{H}_Y = \varphi(G)) \delta_{\text{mean}_Y \varphi(G)} \right) &\leq \max_{G \in \mathcal{G}} \mathcal{W}_2 \left( \text{mean}_X G, \text{mean}_Y \varphi(G) \right) \\ &\leq \vec{d}_2(X, Y) \end{aligned}$$

by Lemma 6.4. Together, this implies

$$\begin{aligned} \mathcal{W}_2(\mu_X, \mu_Y) &= \mathcal{W}_2 \left( \sum_{G \in \mathcal{G}(X)} \mathbb{P}(\mathcal{H}_X = G) \delta_{\text{mean}_X G}, \sum_{G \in \mathcal{G}(X)} \mathbb{P}(\mathcal{H}_Y = \varphi(G)) \delta_{\text{mean}_Y \varphi(G)} \right) \\ &\leq \left( \frac{4\bar{M}^2}{\alpha} \sum_{x \in X} \|x - \varphi(x)\| \right)^{1/2} + \vec{d}_2(X, Y). \end{aligned}$$

□

**6.3. Proof for points close to the diagonal.** In Section 6.2, we were able to use the fact that a set of optimal matchings  $\varphi_i : X_i \rightarrow Y_i$  which associate off-diagonal points together induces a bijection  $\varphi : \mathcal{G}(X) \rightarrow \mathcal{G}(Y)$ . However, if we have a point  $x \in X_i$  such that  $\varphi_i(x) = \Delta$ ,  $\varphi(G)$  is no longer injective as different selections can map to the same selection. For example given two diagrams, the grouping  $G_1 = \{(x, \Delta), (y, \Delta)\}$  and  $G_2 = \{(x, y)\}$  have the same image. Thus, in the following proposition, we bound the distance between PFMs for diagrams which only differ by points that are matched to the diagonal.

**Proposition 6.6.** *Let  $X = (X_1, \dots, X_N)$  and  $\tilde{X} = (\tilde{X}_1, \dots, \tilde{X}_N)$  denote sets of diagrams in  $(S_{M,K})^N$  where the off-diagonal points in each  $\tilde{X}_i$  is a subset of those in  $X_i$ . Then*

$$\mathcal{W}_2(\mu_X, \mu_{\tilde{X}})^2 \leq \left( \frac{1}{N^2} + \frac{\bar{M}^2}{\alpha^2} \right) \sum_{x \in X \setminus \tilde{X}} \|x - \Delta\|^2$$

where  $\bar{M}$  is the maximum distance between any two diagrams in  $S_{M,NK}$ .

*Proof.* If  $\|x - \Delta\| \geq \alpha$  for some  $x \in X \setminus \tilde{X}$  then  $\sum_{x \in X \setminus \tilde{X}} \|x - \Delta\|^2 > \alpha^2$  and the theorem automatically holds as  $\mathcal{W}_2(\mu_X, \mu_{\tilde{X}})^2 \leq \bar{M}^2$  by the definition of  $\bar{M}$ . From now on assume that  $\|x - \Delta\| < \alpha$  for all  $x \in X \setminus \tilde{X}$ .

Let  $\mathcal{G}(\tilde{X})$  and  $\mathcal{G}(X)$  be the sets of groupings for  $\tilde{X}$  and  $X$  respectively. There is an injection  $i_{\tilde{X}} : \mathcal{G}(\tilde{X}) \hookrightarrow \mathcal{G}(X)$  which maps a grouping  $G \in \mathcal{G}(\tilde{X})$  and to the grouping in  $\mathcal{G}(X)$  which has all the same selections as  $G$  along with the trivial selection for each unused point  $x \in X_i \setminus \tilde{X}_i$ . In order to bound  $\mathcal{W}_2(\mu_X, \mu_{\tilde{X}})$ , construct a transportation plan from  $\mu_{\tilde{X}}$  to  $\mu_X$  as follows:

- If  $\mathbb{P}(\mathcal{H}_{\tilde{X}} = G) \leq \mathbb{P}(\mathcal{H}_X = i_{\tilde{X}}(G))$  then move all the mass at  $\delta_{\text{mean}_{\tilde{X}} G}$  to  $\delta_{\text{mean}_X i_{\tilde{X}}(G)}$
- If  $\mathbb{P}(\mathcal{H}_{\tilde{X}} = G) > \mathbb{P}(\mathcal{H}_X = i_{\tilde{X}}(G))$  then move  $\mathbb{P}(\mathcal{H}_X = i_{\tilde{X}}(G))$  worth of mass from  $\delta_{\text{mean}_{\tilde{X}} G}$  to  $\delta_{\text{mean}_X i_{\tilde{X}}(G)}$  and redistribute the rest as needed.

First note that for any  $G \in \mathcal{G}(\tilde{X})$ , the amount of mass moved from  $\delta_{\text{mean}_{\tilde{X}} G}$  to its corresponding  $\delta_{\text{mean}_X i_{\tilde{X}}(G)}$  is bounded from above by  $\mathbb{P}(\mathcal{H}_{\tilde{X}} = G)$ . Secondly, the amount of mass not moved from  $\delta_{\text{mean}_{\tilde{X}} G}$  to its corresponding  $\delta_{\text{mean}_X i_{\tilde{X}}(G)}$ , is

$$\sum_{G \in \mathcal{G}_{\tilde{X}}} \max \left\{ \mathbb{P}(\mathcal{H}_{\tilde{X}} = G) - \mathbb{P}(\mathcal{H}_X = i_{\tilde{X}}(G)), 0 \right\}.$$

Therefore using this transport plan,

$$\begin{aligned} \mathcal{W}_2(\mu_X, \mu_{\tilde{X}})^2 &\leq \sum_{G \in \mathcal{G}_{\tilde{X}}} \mathbb{P}(\mathcal{H}_{\tilde{X}} = G) \mathcal{W}_2(\text{mean}_{\tilde{X}} G, \text{mean}_X i_{\tilde{X}}(G))^2 \\ (9) \quad &+ \bar{M}^2 \sum_{G \in \mathcal{G}_{\tilde{X}}} \max \{ \mathbb{P}(\mathcal{H}_{\tilde{X}} = G) - \mathbb{P}(\mathcal{H}_X = i_{\tilde{X}}(G)), 0 \} \end{aligned}$$

where  $\bar{M}$  is the maximum distance between any two diagrams in  $S_{M,NK}$ .

In order to bound  $\mathcal{W}_2(\text{mean}_{\tilde{X}}(G), \text{mean}_X(i_{\tilde{X}}(G)))^2$ , observe that every off diagonal point that appears in  $\text{mean}_{\tilde{X}}(G)$  also appears in  $\text{mean}_X(i_{\tilde{X}}(G))$  and that the additional points in  $\text{mean}_X(i_{\tilde{X}}(G))$  correspond to the trivial selections  $m_x$  for all  $x \in X \setminus \tilde{X}$ . Each of these additional points are at distance  $\|x - \Delta\|/N$

to the diagonal by Lemma 6.1. Thus, using the matching sending each of these additional points to the diagonal,

$$W_2(\text{mean}_{\tilde{X}}(G), \text{mean}_X(i_{\tilde{X}}(G)))^2 \leq \sum_{x \in X \setminus \tilde{X}} \frac{\|x - \Delta\|^2}{N^2}.$$

for all  $G \in \mathcal{G}_{\tilde{X}}$ . Since  $\sum_{G \in \mathcal{G}_{\tilde{X}}} \mathbb{P}(\mathcal{H}_{\tilde{X}} = G) = 1$ ,

$$\sum_{G \in \mathcal{G}_{\tilde{X}}} \mathbb{P}(\mathcal{H}_{\tilde{X}} = G) W_2(\text{mean}_{\tilde{X}} G, \text{mean}_X i_{\tilde{X}}(G))^2 \leq \sum_{x \in X \setminus \tilde{X}} \frac{\|x - \Delta\|^2}{N^2}.$$

Now we can consider the second half of Eqn. 9. Let  $E$  be the event that all the points  $x \in X \setminus \tilde{X}$  are perturbed to the diagonal.

$$\mathbb{P}(E) = \prod_{x \in X \setminus \tilde{X}} \left(1 - \frac{\|x - \Delta\|^2}{\alpha^2}\right).$$

When  $E$  is conditioned to be true, the randomly drawn diagrams for  $X$  have the same distribution as that of  $\tilde{X}$ . This implies that  $\mathbb{P}(\mathcal{H}_X = i_{\tilde{X}}(G) | E) = \mathbb{P}(\mathcal{H}_{\tilde{X}} = G)$ .

Thus, to bound  $\sum_{G \in \mathcal{G}_{\tilde{X}}} \max\{\mathbb{P}(\mathcal{H}_{\tilde{X}} = G) - \mathbb{P}(\mathcal{H}_X = i_{\tilde{X}}(G)), 0\}$  observe that

$$\begin{aligned} \mathbb{P}(\mathcal{H}_X = i_{\tilde{X}}(G)) &> \mathbb{P}(\mathcal{H}_X = i_{\tilde{X}}(G) \text{ and } E) \\ &= \mathbb{P}(\mathcal{H}_X = i_{\tilde{X}}(G) | E) \mathbb{P}(E) \\ &= \mathbb{P}(\mathcal{H}_{\tilde{X}} = G) \prod_{x \in X \setminus \tilde{X}} \left(1 - \frac{\|x - \Delta\|^2}{\alpha^2}\right). \end{aligned}$$

This implies that

$$\begin{aligned} \max\{\mathbb{P}(\mathcal{H}_{\tilde{X}} = G) - \mathbb{P}(\mathcal{H}_X = i_{\tilde{X}}(G)), 0\} &\leq \mathbb{P}(\mathcal{H}_{\tilde{X}} = G) \left(1 - \prod_{x \in X \setminus \tilde{X}} \left(1 - \frac{\|x - \Delta\|^2}{\alpha^2}\right)\right) \\ &\leq \mathbb{P}(\mathcal{H}_{\tilde{X}} = G) \sum_{x \in X \setminus \tilde{X}} \frac{\|x - \Delta\|^2}{\alpha^2} \end{aligned}$$

and hence

$$\begin{aligned} \sum_{G \in \mathcal{G}_{\tilde{X}}} \max\{\mathbb{P}(\mathcal{H}_{\tilde{X}} = G) - \mathbb{P}(\mathcal{H}_X = i_{\tilde{X}}(G)), 0\} &\leq \sum_{G \in \mathcal{G}_{\tilde{X}}} \mathbb{P}(\mathcal{H}_{\tilde{X}} = G) \left(\sum_{x \in X \setminus \tilde{X}} \frac{\|x - \Delta\|^2}{\alpha^2}\right) \\ &\leq \sum_{x \in X \setminus \tilde{X}} \frac{\|x - \Delta\|^2}{\alpha^2}. \end{aligned}$$

Substituting into (9)

$$\begin{aligned} \mathcal{W}_2(\mu_X, \mu_{\tilde{X}})^2 &\leq \sum_{x \in X \setminus \tilde{X}} \frac{\|x - \Delta\|^2}{N^2} + \overline{M}^2 \sum_{x \in X \setminus \tilde{X}} \frac{\|x - \Delta\|^2}{\alpha^2} \\ &\leq \left(\frac{1}{N^2} + \frac{\overline{M}^2}{\alpha^2}\right) \sum_{x \in X \setminus \tilde{X}} \|x - \Delta\|^2 \end{aligned}$$

where  $\overline{M}$  is the maximum distance between any two diagrams in  $S_{M,NK}$ . □

6.4. **Proof of Theorem 5.3.** With these results in hand, particularly Props. 6.6 and 6.5, we can prove the main theorem.

*Proof of Thm. 5.3.* Let  $X = (X_1, \dots, X_N)$  and  $Y = (Y_1, \dots, Y_N)$  denote sets of diagrams in  $(S_{M,K})^N$  with PFMs  $\mu_X$  and  $\mu_Y$  respectively. We wish to find a constant  $C$  such that  $\mathcal{W}_2(\mu_X, \mu_Y) \leq C \vec{d}_2(X, Y)$ .

For the moment assume that  $\vec{d}_2(X, Y) \leq 1$ . For each  $i$ , let  $\varphi_i : X_i \rightarrow Y_i$  be an optimal matching. Let  $\tilde{X}_i$  be the diagram consisting of points  $x \in X_i$  such that  $\varphi_i(x) = \Delta$ . Likewise, let  $\tilde{Y}_i$  be the diagram consisting of points  $y \in Y_i$  such that  $\varphi_i^{-1}(y) = \Delta$ . We will bound  $\mathcal{W}_2(\mu_X, \mu_Y)$  using the triangle inequality,

$$\mathcal{W}_2(\mu_X, \mu_Y) \leq \mathcal{W}_2(\mu_X, \mu_{\tilde{X}}) + \mathcal{W}_2(\mu_{\tilde{X}}, \mu_{\tilde{Y}}) + \mathcal{W}_2(\mu_{\tilde{Y}}, \mu_Y).$$

Using Prop. 6.6 for the first and third portions and Prop. 6.5 for the second, we have

$$\begin{aligned} \mathcal{W}_2(\mu_X, \mu_Y) &\leq \left( \left( \frac{1}{N^2} + \frac{\bar{M}^2}{\alpha^2} \right) \sum_{x \in X \setminus \tilde{X}} \|x - \Delta\|^2 \right)^{1/2} + \left( \left( \frac{4\bar{M}^2}{\alpha} \sum_{x \in \tilde{X}} \|x - \varphi(x)\| \right)^{1/2} + \vec{d}_2(\tilde{X}, \tilde{Y}) \right) \\ &\quad + \left( \left( \frac{1}{N^2} + \frac{\bar{M}^2}{\alpha^2} \right) \sum_{y \in Y \setminus \tilde{Y}} \|y - \Delta\|^2 \right)^{1/2}. \end{aligned}$$

Let  $U = \frac{1}{N^2} + \frac{\bar{M}^2}{\alpha^2}$  and  $V = \frac{4\bar{M}^2}{\alpha}$ . Then via Cauchy-Schwartz,

$$\begin{aligned} &\mathcal{W}_2(\mu_X, \mu_Y)^2 \\ &\leq \left( U^{1/2} \left( \sum_{x \in X \setminus \tilde{X}} \|x - \Delta\|^2 \right)^{1/2} + V^{1/2} \left( \sum_{x \in \tilde{X}} \|x - \varphi(x)\| \right)^{1/2} + \vec{d}_2(\tilde{X}, \tilde{Y}) + U^{1/2} \left( \sum_{y \in Y \setminus \tilde{Y}} \|y - \Delta\|^2 \right)^{1/2} \right)^2 \\ &\leq (2U + V + 1) \left( \sum_{x \in \tilde{X}} \|x - \varphi(x)\| + \sum_{y \in Y \setminus \tilde{Y}} \|y - \Delta\|^2 + \sum_{x \in X \setminus \tilde{X}} \|x - \Delta\|^2 + \vec{d}_2(\tilde{X}, \tilde{Y})^2 \right). \end{aligned}$$

Finally, we combine this with the fact that

$$\sum_{y \in Y \setminus \tilde{Y}} \|y - \Delta\|^2 + \sum_{x \in X \setminus \tilde{X}} \|x - \Delta\|^2 + \vec{d}_2(\tilde{X}, \tilde{Y})^2 = \vec{d}_2(X, Y)^2$$

to get the bound in the theorem. □

## 7. EXAMPLES

We now give some examples of the probabilistic Fréchet mean of a set of diagrams, introducing a useful way to visualize them along the way. Recall that the mean distribution of a set of diagrams is a weighted sum of delta-measures, each one concentrated on the mean of one of the possible groupings among the diagrams, with the weights given by the probability that a perturbation of the diagrams would produce that grouping.

In Fig. 7, we show a resolution to the discontinuity issue raised in Fig. 4, although this figure needs some explanation. The flat colored dots on the left side of the figure represent a pair of diagrams which form a rectangle that is slightly longer in the death-axis direction. To approximate the probability of each possible matching between the pair, we perturbed the diagrams 100 times with  $\alpha = 0.3$  and  $\eta_0$  equal to the uniform distribution, and simply counted the number of times each matching occurred. The results are shown on the left side of the figure, where the height of a colored stack represents the weight of the diagram which contains the point at the bottom of the stack; note that the green stacks are slightly taller than the purple ones. On the other hand, the right side of the same figure shows the mean distribution for a pair of diagrams which forms a rectangle that is quite a bit longer in the birth-axis direction.

For a more complicated example, we drew thirty different point clouds from a pair of linked annuli of different radii; one such point cloud is shown on the left of Fig. 8. Then we computed the one-dimensional

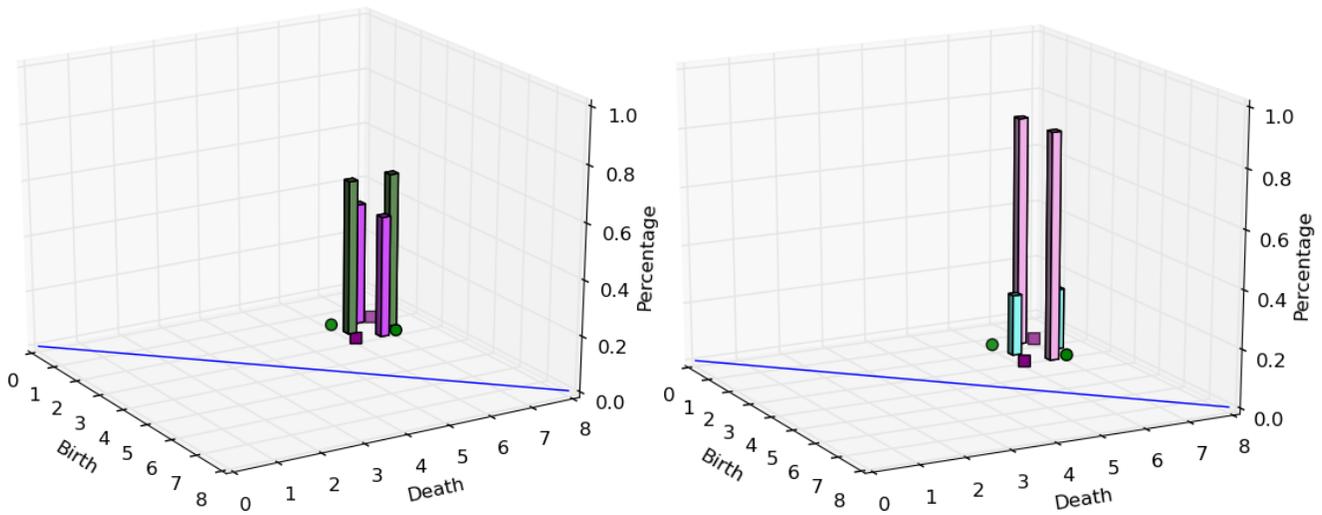


FIGURE 7. Change in mean distribution as a set of two diagrams moves through the problematic configuration of Fig. 4. On the left we see the mean of two diagrams which form a rectangle that is slightly longer in the death-axis direction. On the right is the result for a rectangle that is quite a bit longer in the birth-axis direction.

persistence diagram for each point cloud, using the recently-developed  $M_{12}$  software package [12]. As one might expect, each diagram contained a point for the big annulus, and point for the small annulus, and a good bit of noise along the diagonal. However, the birth times of the non-noisy points varied quite widely. The set of thirty diagrams, overlaid in one picture, is shown on the right of the same figure.

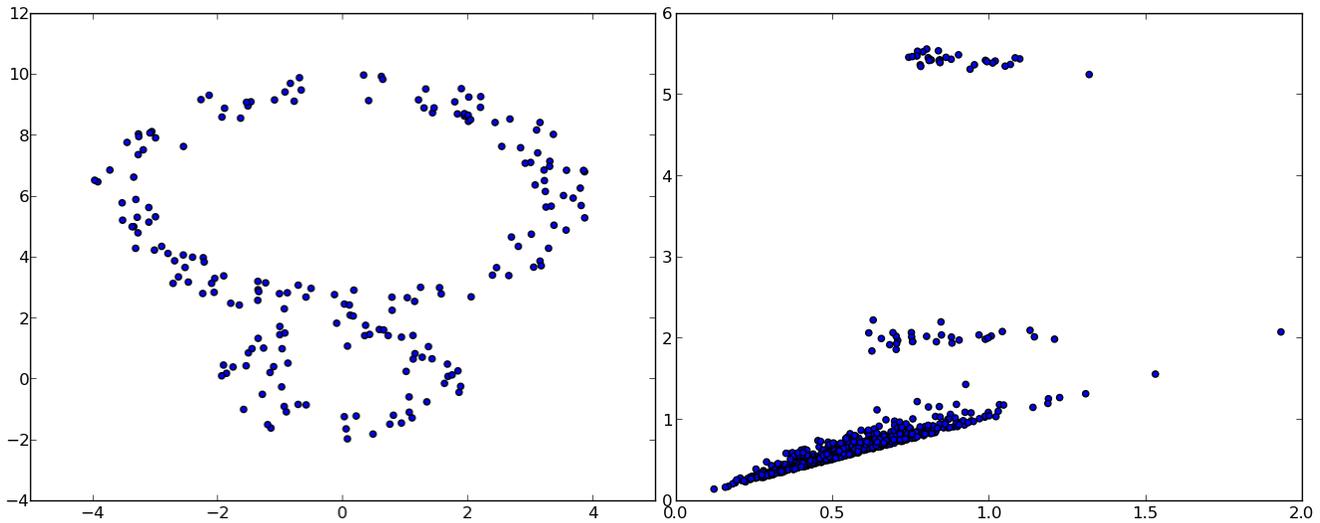


FIGURE 8. Thirty diagrams were created from thirty point clouds drawn from a double annulus. One such point cloud is shown on the left. All thirty diagrams are overlaid on the right.

Finally, we computed the mean distribution of these thirty diagrams, using the same approximation scheme as above. On the left of Fig. 9, we see an overlay of the set of all diagrams which receive positive weight in the mean distribution, while the right side of the same figure displays the mean distribution using the same colored-stack scheme as in the example above. Notice that the two very large stacks are actually at height one, which indicates that every single diagram in the mean contains the two non-noisy dots from the left side of the figure.

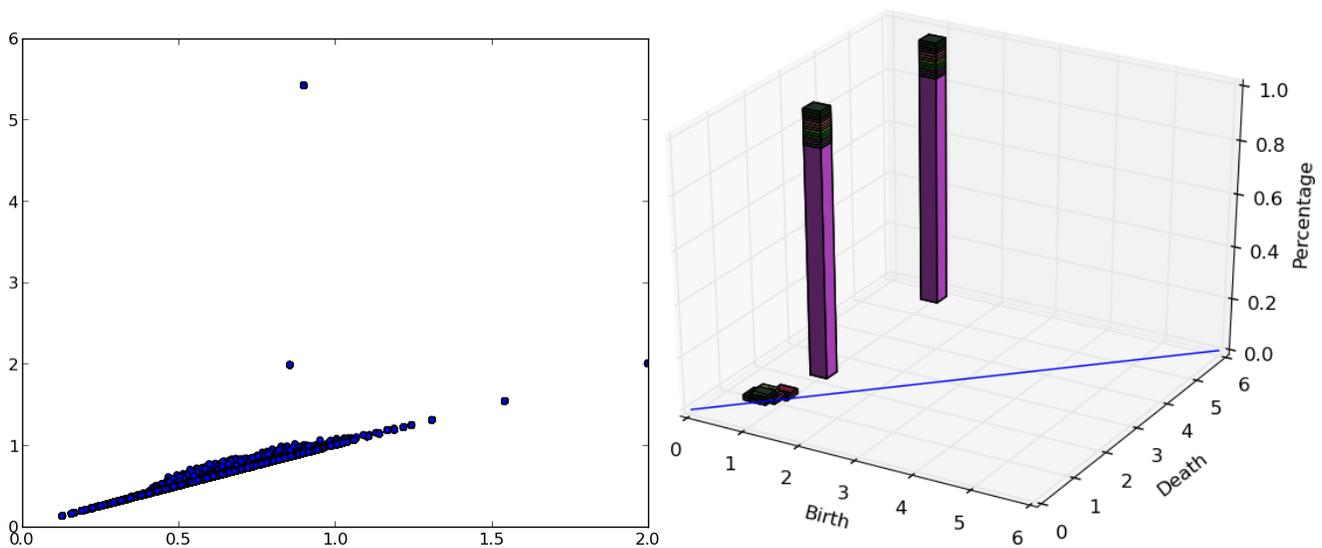


FIGURE 9. The mean distribution for a set of thirty diagrams sampled from a double annulus. The left side shows all positive-weight diagrams in the mean overlaid in one figure, while the right side indicates the weights in a three-dimensional plot.

## 8. CONCLUSIONS AND FUTURE WORK

In this paper, we have defined a new mean which, unlike its predecessor, is continuous for continuously varying diagrams. This mean is, in fact, a distribution on diagram space which is one feature of the distribution of diagrams from which it arose. We hope that this new definition will provide a useful statistical tool for topological data analysis. We also believe that this is an important step in the overall project of establishing persistent homology as an important shape statistic. Several questions remain, however, and there are obviously many directions for future research. We list some of them here.

The most pressing need, of course, is to study how far we can take this new definition into the realm of traditional statistics. In particular, can we prove laws of large numbers, central limit theorems, and the like? Will this mean actually provide a useful tool towards the bootstrapping idea discussed in the introduction? Can we use this new mean, and the associated variance function, to provide more insight into the convergence rate theorems of [8]?

On a more technical level, can we improve our continuity theorem to remove the reliance on the subspaces  $S_{M,K}$ ? At the moment, we can not find counterexamples to a more general statement, but nor can we prove the theorem without making finiteness assumptions. We also conjecture that the constant can be improved. In particular, we are making a vast over-estimate by using  $\overline{M}$ . It would also be interesting to understand exactly how large a role  $\alpha$  plays. Of course, if  $\alpha$  goes to 0, the PFM should converge to the regular Fréchet mean, so can we make a good choice of  $\alpha$  based on the diagrams of study?

Note, too, that we have only addressed means and variances in this paper. Another interesting statistical summary of data is the median; this will be addressed in an upcoming paper [27]. Perhaps the most important project is to understand under what conditions persistence diagrams provide sufficient statistics for an object, a data cloud, etc. The work in this paper will be a critical part of this effort.

## REFERENCES

- [1] Pankaj K. Agarwal, Herbert Edelsbrunner, John Harer, and Yusu Wang. Extreme elevation on a 2-manifold. *Discrete & Computational Geometry*, 36:553–572, 2006.
- [2] Yih-En Andrew Ban, Herbert Edelsbrunner, and Johannes Rudolph. Interface surfaces for protein-protein complexes. In *Proceedings of the eighth annual international conference on Research in computational molecular biology*, RECOMB ’04, pages 205–212, New York, NY, USA, 2004. ACM.
- [3] Andrew J. Blumberg, Itamar Gal, Michael A. Mandell, and Matthew Pancia. Persistent homology for metric measure spaces, and robust statistics for hypothesis testing and confidence intervals. *arXiv:1206.4581*, 2012.
- [4] Kenneth A. Brown and Kevin P. Knudson. Nonlinear statistics of human speech data. *International Journal of Bifurcation and Chaos*, 19(07):2307–2319, 2009.

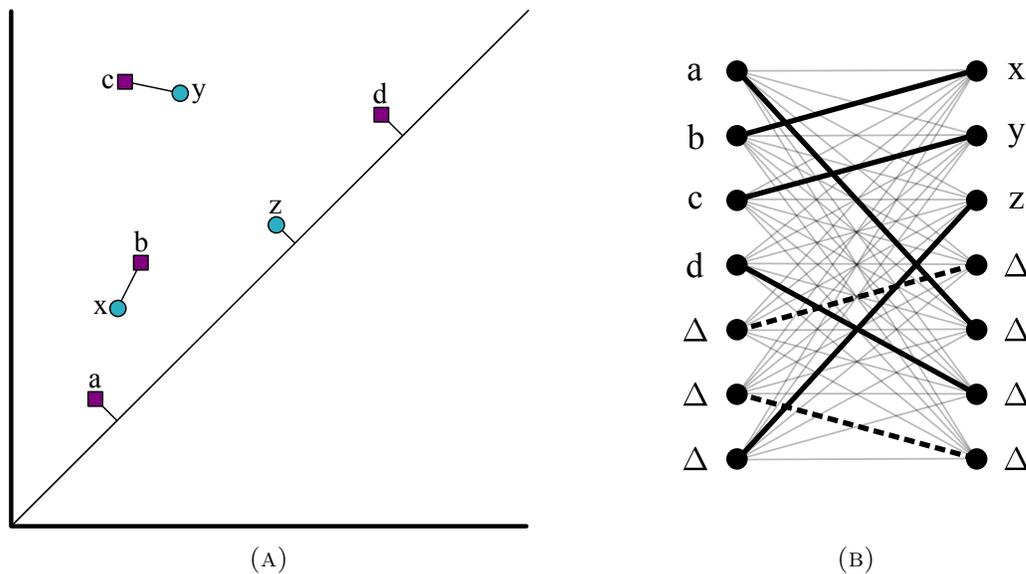


FIGURE 10. Computation of the Wasserstein distance between  $d_{\blacksquare}$  and  $d_{\bullet}$  in Fig. (a). The problem is turned into the problem of computing a minimum cost grouping on the weighted graph in Fig. (b). The grouping chosen, shown in the bold edges in (b), is used to determine the matching for the diagrams in (a). Dashed edges in (b) correspond to  $\Delta - \Delta$  pairings, which contribute nothing to the total distance.

- [5] Peter Bubenik. Statistical topology using persistence landscapes. *arXiv:1207.6437*, July, 2012.
- [6] Gunnar Carlsson, Tigran Ishkhanov, Vin de Silva, and Afra Zomorodian. On the local behavior of spaces of natural images. *International Journal of Computer Vision*, 76:1–12, 2008. 10.1007/s11263-007-0056-x.
- [7] Frédéric Chazal, David Cohen-Steiner, Marc Glisse, Leonidas J. Guibas, and Steve Y. Oudot. Proximity of persistence modules and their diagrams. In *Proceedings of the 25th annual symposium on Computational geometry*, SCG '09, pages 237–246, New York, NY, USA, 2009. ACM.
- [8] Frederic Chazal, Marc Glisse, Catharine Labruere, and Bertrand Michel. Optimal rates of convergence for persistence diagrams in topological data analysis. *arXiv:1305.6239*, 2013.
- [9] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete Comput. Geom.*, 37(1):103–120, January 2007.
- [10] David Cohen-Steiner, Herbert Edelsbrunner, and Dmitriy Morozov. Vines and vineyards by updating persistence in linear time. *Proceedings of the twenty-second annual symposium on Computational geometry - SCG '06*, page 119, 2006.
- [11] Y Dabaghian, F Mémoli, L Frank, and G Carlsson. A topological paradigm for hippocampal spatial map formation using persistent homology. *PLoS Comput Biol*, 8(8):e1002581, 08 2012.
- [12] Anastasia Deckard, Jose' A. Perea, John Harer, and Steve Haase. Sw1pers: Sliding windows and 1-persistence scoring; discovering periodicity in gene expression time series data. *To appear*, 2013.
- [13] Mary-Lee Dequéant, Sebastian Ahnert, Herbert Edelsbrunner, Thomas M. A. Fink, Earl F. Glynn, Gaye Hattem, Andrzej Kudlicki, Yuriy Mileyko, Jason Morton, Arcady R. Mushegian, Lior Pachter, Maga Rowicka, Anne Shiu, Bernd Sturmfels, and Olivier Pourquié. Comparison of pattern detection methods in microarray time series of the segmentation clock. *PLoS ONE*, 3(8):e2856, 2008.
- [14] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. In *Foundations of Computer Science, 2000. Proceedings. 41st Annual Symposium on*, pages 454–463, 2000.
- [15] Herbert Edelsbrunner and John Harer. *Persistent homology — a survey*, chapter Surveys on Discrete and Computational Geometry: Twenty Years Later, pages 257–282. American Mathematical Society, Providence, RI, 2008.
- [16] Herbert Edelsbrunner, John Harer, Vijay Natarajan, and Valerio Pascucci. Morse-smale complexes for piecewise linear 3-manifolds. In *Proceedings of the nineteenth annual symposium on Computational geometry*, SCG '03, pages 361–370, New York, NY, USA, 2003. ACM.
- [17] B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, New York, 1993.
- [18] T. Galkovskiy, Y. Mileyko, A. Bucksch, B. Moore, O. Symonova, C. A. Price, C. N. Topp, A. S. Iyer-Pascuzzi, P. R. Zurek, S. Fang, J. Harer, P. N. Benfey, and J. S. Weitz. Gia roots: Software for high throughput analysis of plant root system architecture. *BMC Plant Biology*, 12(116), 2012.
- [19] Jennifer Gamble and Giseon Heo. Exploring uses of persistent homology for statistical analysis of landmark-based shape data. *Journal of Multivariate Analysis*, 101(9):2184 – 2199, 2010.
- [20] Jeffrey J. Headd, Y. E. Andrew Ban, Paul Brown, Herbert Edelsbrunner, Madhuwanti Vaidya, and Johannes Rudolph. Protein-protein interfaces: properties, preferences, and projections. *Journal of Proteome Research*, 6(7):2576–2586, 2007. PMID: 17542628.

- [21] Yuriy Mileyko, Sayan Mukherjee, and John Harer. Probability measures on the space of persistence diagrams. *Inverse Problems*, 27(12):124007, 2011.
- [22] Dmitriy Morozov. *Homological Illusions of Persistence and Stability*. PhD thesis, Duke University, 2008.
- [23] James Munkres. Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial and Applied Mathematics*, 5(1):32–38, 1957.
- [24] James R. Munkres. *Elements of Algebraic Topology*. Addison Wesley, 1993.
- [25] José A. Perea and John Harer. Sliding windows and persistence: An application of topological methods to signal analysis. *To appear*, 2013.
- [26] Reinhard Selten. Reexamination of the perfectness concept for equilibrium points in extensive games. *International journal of game theory*, 4(1):25–55, 1975.
- [27] Katharine Turner. Medians and means for sets of persistence diagrams. *In preparation*, 2013.
- [28] Katharine Turner, Yuriy Mileyko, Sayan Mukherjee, and John Harer. Fréchet means for distributions of persistence diagrams. *ArXiv*, page arXiv:1206.2790, 2011.

## APPENDIX A. ALGORITHMS

---

**Algorithm 1** Algorithm for computing the Fréchet Mean of a finite set of diagrams

---

**Input:** Persistence diagrams  $X_1, \dots, X_N$

**Output:**  $Y$ , a persistence diagram giving a local min of the Fréchet function

Choose one of the  $X_i$  randomly, set  $Y = X_i$

Initialize matching  $G$

▷  $G[j, i] =$  the  $x_k \in X_i$  matched  
 ▷ with the point  $y_j \in Y$

stop = False

**while** stop == False **do**

**for** each diagram  $X_i$  **do**

  ▷ Determine the best  $G$

$P = \text{WassersteinPairing}(Y, X_i)$

**for** each pair  $(y_j, x_k) \in P$  **do**

      Set  $G[j, i] = x_k$

**end for**

**end for**

Initialize empty diagram  $Y'$

▷ Move each point to the

**for** each point  $y_j \in Y$  **do**

▷ barycenter of its selection.

$y'_j = \text{mean}\{G[j, 1], \dots, G[j, N]\}$

▷  $Y' = \text{mean}_X(G)$

  Add  $y'_j$  to  $Y'$

**end for**

**if**  $\text{WassersteinPairing}(Y, X_i) = \text{WassersteinPairing}(Y', X_i) \forall i$  **then**

  stop = True

**end if**

$Y = Y'$

**end while**

**return**  $Y$

---

Here, we discuss the algorithm to compute an estimate of the Fréchet mean of a set of diagrams as given in [28] using the vocabulary developed in this paper. It is shown there that the Fréchet function is semiconcave for distributions with bounded support, so we can make use of a gradient descent algorithm to find local minima of the Fréchet function, Def. 3.1. In order to present the algorithm for computing the Fréchet mean, we must first describe the algorithm for computation of Wasserstein distance. In order to compute the Wasserstein distance between two diagrams, we will reduce the problem to computing a minimum cost grouping of a complete, weighted bipartite graph.

Let  $X = [x_1, \dots, x_k]$  and  $Y = [y_1, \dots, y_m]$  be diagrams. In order to compute  $W_2[L_2](X, Y)$ , we construct a complete bipartite graph with vertex set  $U \cup V$ . There is a vertex in  $U$  for each  $x_i$ , as well as  $m$  vertices representing the abstract diagonal  $\Delta$ ; similarly,  $V$  has a vertex for each  $y_i$  as well as  $k$  vertices representing  $\Delta$ . The edge between points  $x_i$  and  $y_j$  is given weight  $\|x_i - y_j\|^p$ . Each edge  $(x_i, \Delta)$

and  $(\Delta, y_j)$  has weight  $\|x_i - \Delta\|^p$  and  $\|y_j - \Delta\|^p$  respectively where  $\|a - \Delta\| = \min_{z \in \Delta} \|a - z\|$ . Finally, edges between two vertices representing  $\Delta$  are given weight 0. The minimum cost grouping algorithm typically used is the Hungarian algorithm of Munkres [23].

A minimum cost grouping in the bipartite graph immediately gives a matching  $\varphi : U \rightarrow V$  and the Wasserstein distance is given by the square root of the sum of the squares of the weights of the edges. Notice that since there could be multiple groupings for a bipartite graph which minimize the cost, there could be multiple groupings which minimize the Wasserstein distance. To compute the mean diagram, we will actually be more interested in the matching returned in this algorithm than in the distance itself. Fig. 10 displays an example of a pair of diagrams and their corresponding bipartite graph.

Now we are ready to give the algorithm for the Fréchet mean of a set of diagrams. Given a finite set of diagrams  $\{X_1, \dots, X_N\}$ , start with a candidate for the mean,  $Y$ , and compute the matching for  $W_2(Y, X_i)$ . We denote this as  $\text{WassersteinPairing}(Y, X_i)$ . From this, we have a grouping  $G$  where  $G[j, i]$  gives the point in  $X_i$  which was paired to point  $y_j \in Y$ . Set  $Y' = \text{mean}(G)$ . This new diagram is now the candidate for the mean and the process is repeated. The algorithm terminates when the Wasserstein pairing does not change. In [28], the structure of  $(D_2, W_2)$  is used to prove that this algorithm terminates at a local minimum of the Fréchet function. See Algorithm 1 for the pseudocode.

<sup>1</sup> *Corresponding author*, DEPT OF MATHEMATICS & STATISTICS, UNIVERSITY AT ALBANY – SUNY, *emunch@albany.edu*.

<sup>2</sup> DEPT OF MATHEMATICS, UNIVERSITY OF CHICAGO, *kate@math.uchicago.edu*.

<sup>3</sup> DEPT OF MATHEMATICS, DUKE UNIVERSITY, *bendich@math.duke.edu*.

<sup>4</sup> DEPTS OF STATISTICAL SCIENCE, COMPUTER SCIENCE, MATHEMATICS, AND INSTITUTE FOR GENOME SCIENCES & POLICY, DUKE UNIVERSITY, *sayan@stat.duke.edu*.

<sup>5</sup> DEPT OF MATHEMATICS, DUKE UNIVERSITY, *jonm@math.duke.edu*.

<sup>6</sup> DEPTS OF MATHEMATICS, COMPUTER SCIENCE, AND ELECTRICAL AND COMPUTER ENGINEERING, DUKE UNIVERSITY, *john.harer@duke.edu*.