

Probabilistic Fréchet Means and Statistics on Vineyards

Elizabeth Munch*, Paul Bendich†, Katharine Turner‡,
Sayan Mukherjee§, Jonathan Mattingly¶ and John Harer||

July 24, 2013

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 (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. As an unfortunate consequence, one sees that the means of a continuously-varying set of diagrams do not themselves 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 the (Fréchet mean) persistence diagram of a perturbation of the input diagrams. We show that this new definition defines a (Hölder) continuous map, for each k , from $(D_p)^k \rightarrow P(D_p)$, and we present several examples to show how it may become a useful statistic on 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 χ in some possibly high-dimensional metric space, the persistence diagram $d(\chi)$ 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 Figure 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. To address this there has been a recent effort to study the mean and variance of a set of persistence diagrams [3, 5, 21, 27], as well a very recent paper that gives 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 χ , there is a computational

*Department of Mathematics, Duke University and Institute for Mathematics and its Applications, University of Minnesota

†Department of Mathematics, Duke University

‡Department of Mathematics, University of Chicago

§Departments of Statistical Science, Computer Science, Mathematics, and Institute for Genome Sciences & Policy, Duke University

¶Department of Mathematics, Duke University

||Departments of Mathematics, Computer Science, and Electrical and Computer Engineering, Duke University

and statistical advantage to subsampling the data to produce smaller point clouds χ_1, \dots, χ_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, 27] make careful study of the geometric and analytic properties of the space (D_p, W_p) of persistence diagrams equipped with the Wasserstein metric, which enables a definition of mean and variance in the former paper, and an algorithm for their computation in the latter. There is, however, an unfortunate problem with the definition in [21]: namely, the mean of a set of diagrams can itself be a set of diagrams, rather than a single unique diagram. This results in means that do not continuously vary as the input set of diagrams vary.

In this paper, we provide an alternative definition for the mean of a set of diagrams. The key idea is that our mean is not itself a diagram, but is rather a probabilistic mixture of diagrams, thus an element of $\mathcal{P}(D_p)$, 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 (Theorem 13) continuity of the map $(D_p)^k \rightarrow \mathcal{P}(D_p)$ taking a set of k persistence diagrams to its mean; in fact, we show that this map is Hölder with exponent $\frac{1}{2}$. 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 (D_p, W_p) . The contributions of [21] and [27] are reviewed more fully in Section 3, and the non-uniqueness issue is also discussed in that section. We give our new definition in Section 4, and prove its desirable theoretical properties in Section 5, although some technical details are confined to the Appendix. Examples, implementation details, and a discussion of visualization, all come in Section 6, and the paper concludes with some discussion in Section 7.

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 (D_p, W_p) . For more details on persistence, see [15]. We assume the reader is familiar with homology; [24] is a good reference. For the expert, we note that all homology groups in this paper need to be computed with field coefficients.

2.1 Persistent Homology

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

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

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

$$d_\chi(x) = \inf_{v \in \chi} \|x - v\|$$

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

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

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

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

Let $\varphi_i^j : H_r(\mathbb{X}_i) \longrightarrow H_r(\mathbb{X}_j)$ be the composition of necessary maps from Equation (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 φ_{i-1}^i . This same class is said to die

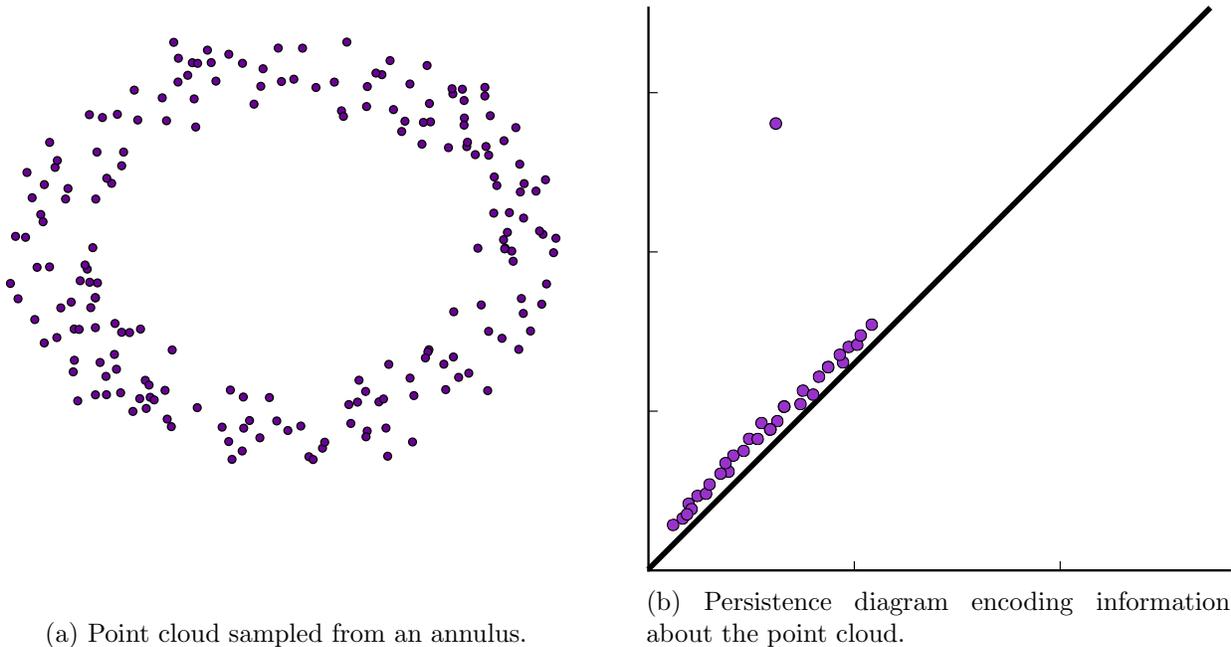


Figure 1: A point cloud, shown at left, is sampled from an annulus. In order to summarize the topological data, we look at the level sets of the distance function from the set of points, then construct the persistence diagram, shown at right. The points near the diagonal are considered noise, while the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

at \mathbb{X}_j if its image in $H_r(\mathbb{X}_{j-1})$ is not in the image of φ_{i-1}^{j-1} , but its image in $H_r(\mathbb{X}_j)$ is in the image of φ_{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 γ 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,

$$V_1 \longrightarrow V_2 \longrightarrow \cdots \longrightarrow V_n,$$

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 many copies of the points on the diagonal $\Delta = \{(x, x) \in \mathbb{R}^2\}$. For each class γ 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 Figure 1b for an example.

2.2 The Space (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 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 Δ having countably infinite multiplicity.*

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

Definition 2. The p^{th} Wasserstein distance between two persistence diagrams X and Y is given by

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

where $1 \leq p \leq \infty$, σ is a metric on the plane, and φ 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 remainder of this paper, we will be using $W_p[L_2]$, which we refer to as W_p for brevity.

The persistence $\text{pers}(u)$ of a point $u = (x, y)$ in a diagram is defined to be $y - x$, and the p^{th} total persistence of a diagram d is defined to be the sum of the p^{th} -powers of the persistences of the off-diagonal points in d .

Definition 3. The space of persistence diagrams is

$$D_p = \{d \mid W_p(d, d_\emptyset) < \infty\} = \{d \mid \text{Pers}_p(d) < \infty\}$$

along with the p^{th} -Wasserstein metric, $W_p = W_p[L_2]$, from Definition 2.

The authors in [21] show that (D_p, W_p) is a Polish (complete and separable) space, and they also describe all of the compact sets in this space. In [27], it is shown that D_p is a geodesic space, and thus that every pair of diagrams has a minimal geodesic between them. This geodesic can be defined using a bijection between the diagrams which minimizes Wasserstein distance.

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, $D(\mathbb{X}(t))$, 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 Figure 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 4. The space of abstract vineyards is

$$V = \{v : [0, 1] \rightarrow D_p\}$$

the space of maps from the unit interval to D_p where v is continuous with respect to W_p .

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 [27]. 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.

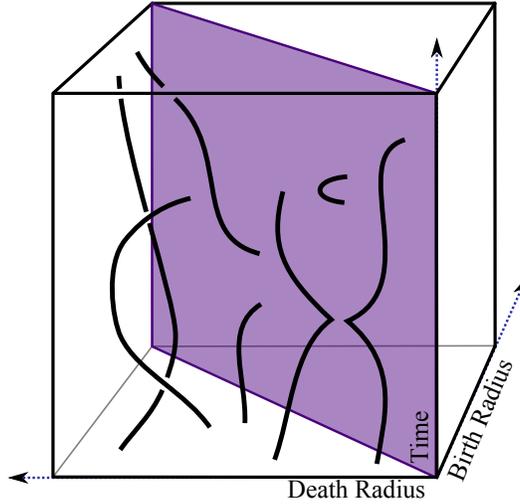


Figure 2: An example of a vineyard. For each time, given on the z -axis, there is a persistence diagram. Since vineyards arising from continuous point clouds are continuous, each point in the diagram traces out a path called a vine. These vines can have endpoints on the starting or ending times, or on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

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 $(D_p, \mathcal{B}(D_p), \mathcal{P})$, we can define the Fréchet mean as follows.

Definition 5. *Given a probability space $(D_p, \mathcal{B}(D_p), \mathcal{P})$, the quantity*

$$Var_{\mathcal{P}} = \inf_{X \in D_p} \left[F_{\mathcal{P}}(X) = \int_{D_p} W_p(X, Y)^2 d\mathcal{P}(Y) < \infty \right]$$

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 $(D_p, W_p[L_{\infty}])$ is non-empty for certain types of well-behaved probability measures on D_p .

Theorem 6. *Let \mathcal{P} be a probability measure on $(D_p, \mathcal{B}(D_p))$ 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 Algorithm

The focus of [21] was to develop the probability theory required for statistical procedures on persistence diagrams, including defining a mean. In [27] an algorithm to compute an estimate of the Fréchet mean of a set of diagrams was given. It is also 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, Definition 5. In order to present the algorithm for computing the Fréchet

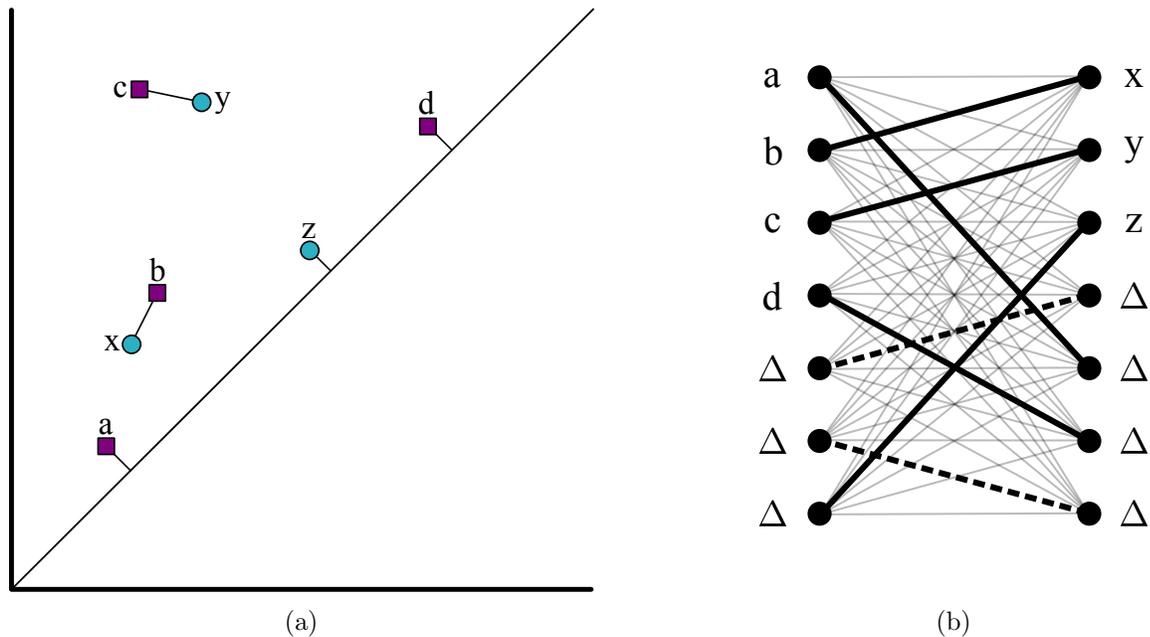


Figure 3: Computation of the Wasserstein distance between d_{\blacksquare} and d_{\bullet} in Figure (a). The problem is turned into the problem of computing a minimum cost coupling on the weighted graph in Figure (b). The coupling chosen, shown in the bold edges in (b), is used to determine the bijection for the diagrams in (a). Dashed edges in (b) correspond to $\Delta - \Delta$ pairings, which contribute nothing to the total distance.

mean, we must first describe the algorithm for computation of Wasserstein distance. The representation of a diagram in this algorithm is a list of its off-diagonal points, $X = [x_1, \dots, x_k]$. In order to compute the Wasserstein distance between two diagrams, we will reduce the problem to computing a minimum cost coupling 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 bipartite graph with vertex set $U \cup V$. U has a vertex for each x_i , as well as m vertices representing the abstract diagonal Δ . Similarly, V has a vertex for each y_i as well as k vertices representing Δ . Take all edges from U to V so that this is a complete bipartite graph. The edge between points x_i and y_j is given weight $\|x_i - y_j\|^p$. Each edge (x_i, Δ) and (Δ, 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 Δ are given weight 0. The minimum cost coupling algorithm typically used is the Hungarian algorithm of Munkres [23].

A minimum cost coupling in the bipartite graph immediately gives a bijection $\varphi : U \rightarrow V$ 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 couplings for a bipartite graph which minimize the cost, there could be multiple couplings which minimize the Wasserstein distance. To compute the mean diagram, we will actually be more interested in the bijection returned in this algorithm than in the distance itself. Figure 3 displays an example of a pair of diagrams and their corresponding bipartite graph.

Definition 7. 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 Δ . The trivial selection for a particular off-diagonal point $x \in X_i$ is the selection s_x which chooses x for X_i and Δ for every other diagram.

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

Given a set of diagrams, a coupling is the analog of a bijection for a pair of diagrams. A coupling for N diagrams which has k selections can be stored as a $k \times N$ matrix G where entry $G[j, X_i] = x$ means that the j^{th} selection has point $x \in X_i$. Note that we consider couplings to be equivalent up to

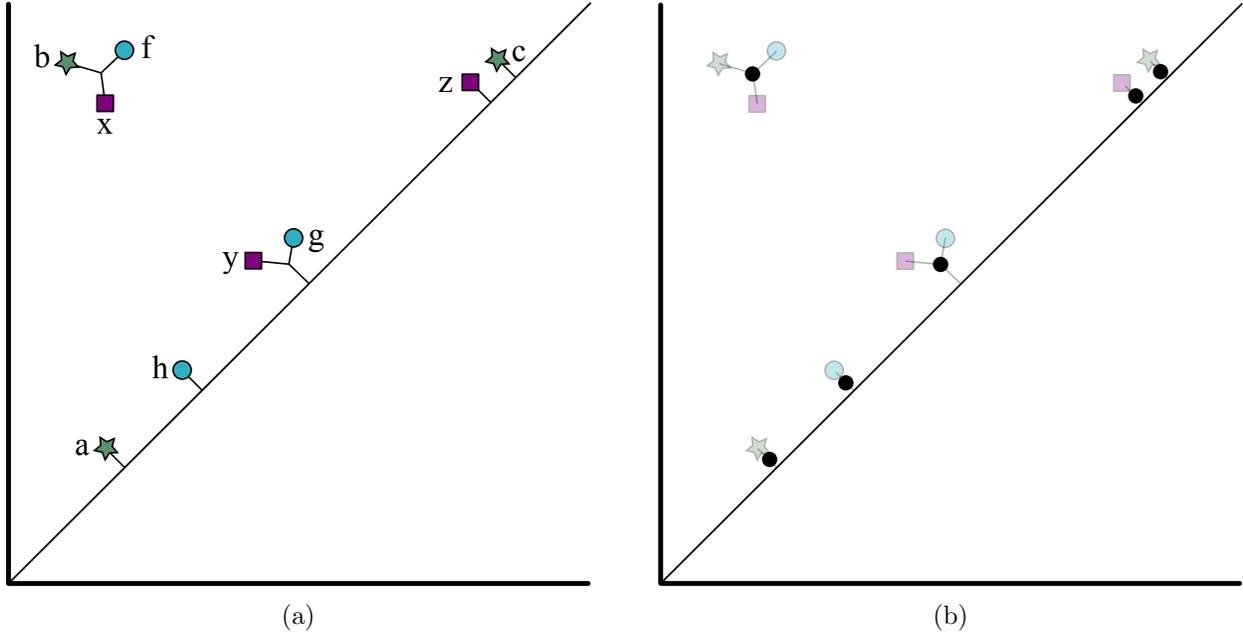


Figure 4: An example of a coupling for three overlaid persistence diagrams, d_{\blacksquare} , d_{\star} , and d_{\bullet} is given in (a). In this example, the coupling has four selections and the corresponding coupling matrix is given in Equation 3. The dark circles in diagram (b) give the mean diagram associated to this particular coupling.

reordering of the selections. See Figure 4 for an example. In this case, the coupling shown is given by the matrix

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

where Δ represents the diagonal.

The *mean of a selection* is the point which minimizes the sum of the square distances to the elements of the selection. 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 the rest copies of the diagonal Δ . A quick computation gives this point as

$$\text{mean}_X(s) = \frac{1}{2Nk} \left((N+k) \sum_i x_i + (N-k) \sum_i y_i, \right. \tag{4} \\
 \left. (N-k) \sum_i x_i + (N+k) \sum_i y_i \right).$$

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, Equation 4 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. If the single off-diagonal point is at $x = (a, b)$ and there are a total of N diagrams,

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

so the point is placed at distance $\frac{1}{N}\|x - \Delta\|$ from the diagonal.

The *mean of a coupling*, $\text{mean}(G)$, is a diagram in D_p 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 coupling yields a diagram.

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 bijection for $W_2[L_2](Y, X_i)$. We denote this as $\text{WassersteinPairing}(Y, X_i)$. From this, we have a coupling G where $G[y_j, X_i]$ 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. [27] uses the structure of $(D_2, W_2[L_2])$ to prove that this algorithm terminates at a local minimum of the Fréchet function. See Algorithm 1 for the pseudocode.

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

▷ $G[y_j, X_i]$ = 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 $G[y_j, X_i] = x_k$

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.

$y'_j = \text{mean}\{G[y_j, X_1], \dots, G[y_j, X_N]\}$

 ▷ $Y' = \text{mean}_X(G)$

 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

3.3 Issues with extensions to Vineyards

Consider the example of Figure 5. Here we have 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 pairing 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 elongated configuration of Figure 6a and move along the dotted lines to the configuration in Figure 6b. At the bend of the dotted line, the points are at the corners of a square, so as in the

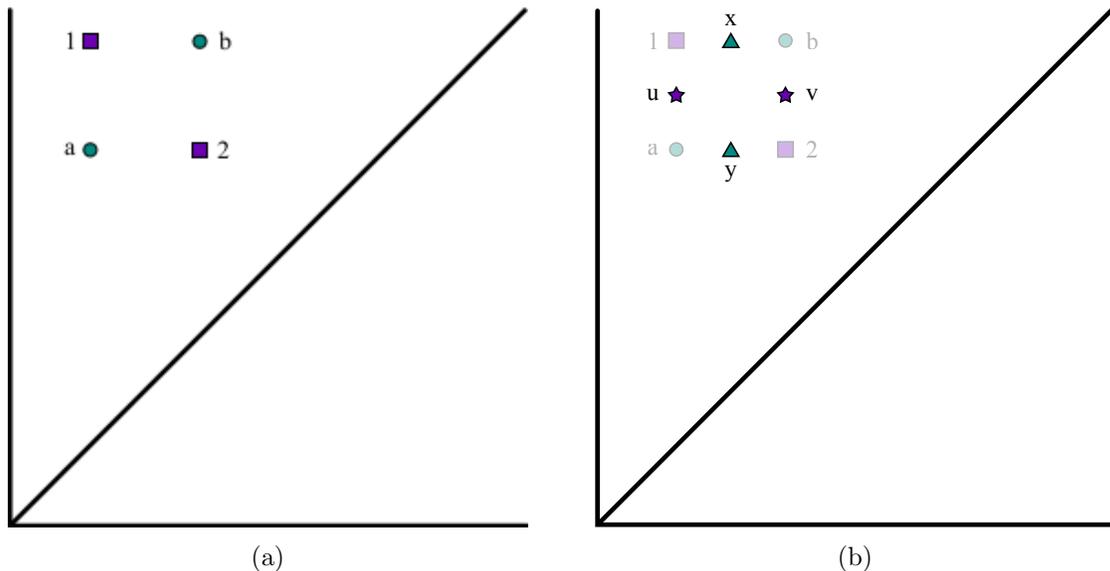


Figure 5: A counterexample to uniqueness of the Fréchet mean in D_p . Figure (a) shows two diagrams overlaid: d_{\blacksquare} has points 1 and 2, d_{\bullet} has points a and b. Since the coupling given by the Wasserstein distance is not unique, neither is the Fréchet mean. The two possible means are given in Figure (b): one has points x and y , the other has points u and v .

example of Figure 5, 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.

4 The Mean as a Distribution

To overcome the lack of continuity of the mean vineyard and the non-uniqueness illustrated in Figure 5 we will define a mean diagram that is a mixture of diagrams or a distribution over persistence diagrams.

In order to prove continuity, the diagrams will be limited to $S_{M,K}$, the set of diagrams in D_2 with at most K off-diagonal points, and all points $x = (x_1, x_2)$ satisfy $0 \leq x_1, x_2 \leq M$.

Consider the space $\mathcal{P}(S_{M,K})$, the space of probability measures with finite second moment on $S_{M,K} \subset D_2$. This is of course a metric space with the standard probability Wasserstein distance as defined below.

Definition 8. *The p^{th} -Wasserstein distance between two probability distributions, ν and η , on metric space $(\mathbb{X}, d_{\mathbb{X}})$ is*

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

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

Thus, we can use $W_2[W_2[L_2]]$ as the distance function on $\mathcal{P}(S_{M,K})$, where the outside W_2 is the Wasserstein distance of Definition 8 and the inside W_2 is the deterministic Wasserstein distance of Definition 2. Note that the map $Y \rightarrow \delta_Y$, where δ_Y is the delta measure concentrated on the diagram Y , gives an isometric embedding of S into $\mathcal{P}(S)$.

4.1 Intuition

We first give an intuitive description of the main ideas behind our new definition. The basic ideas we use to achieve continuity of a mean diagram is to think of diagrams as probabilistic objects and couplings between diagrams as probabilistic objects. These two ideas imply that the mean of a set of diagrams is not a diagram but a distribution over diagrams.

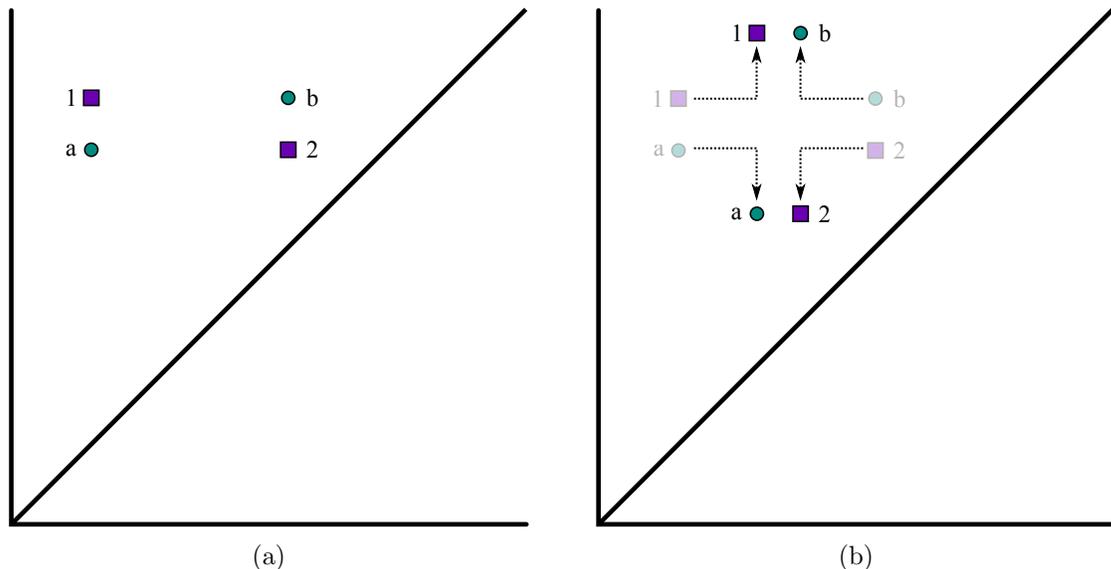


Figure 6: Two vineyards whose pointwise mean is not continuous. The two vineyards start at the configuration in Figure (a) and end in the configuration of Figure (b). The mean is continuous until the points get to the turn of the dotted line, where they form a square, and the mean jumps discontinuously.

A useful perspective on Wasserstein distance metrics is considering them as couplings. The general definition of a coupling follows [28].

Definition 9. Let (\mathcal{X}, ν) and (\mathcal{Y}, η) be two probability spaces. Coupling ν and η means constructing two random variables X and Y on some probability space (Ω, \mathcal{P}) , in such a way that law $(X) = \nu$, law $(Y) = \eta$. The couple (X, Y) is called a coupling of ν, η .

A natural interpretation of the above definition is that a coupling consists of constructing a joint probability space with marginals ν and η . Deterministic couplings follow from the above definition as the case where the law γ on (X, Y) is concentrated on the graph of a measurable function $G : \mathcal{X} \rightarrow \mathcal{Y}$. There are two problems with deterministic couplings: (1) given two diagrams there may be multiple optimal couplings (see Figure 5), (2) given two diagrams slight perturbations of the point configurations can result in radically different optimal couplings, consider slight perturbations of Figure 5. These two problems led to the problem shown in Figure 6, where a jump between optimal couplings led to a discontinuity in the mean diagram. The first problem is addressed by considering probabilistic couplings, so that γ is a probability distribution. The second problem is addressed by considering each diagram as a distribution of point configurations.

The distribution over couplings we construct is a discrete probability distribution over deterministic couplings. The idea behind probabilistic couplings is that given two diagrams we consider *all* possible (optimal) couplings between the diagrams, and assign a probability to each coupling. For now consider each coupling to be equiprobable. For example, the mean of d_{\blacksquare} and d_{\bullet} in Figure 5 will be

$$\frac{1}{2} \cdot \delta_{d_{\blacksquare}} + \frac{1}{2} \delta_{d_{\bullet}},$$

since there are two deterministic couplings that are equally probable.

This addresses the non-uniqueness issue but the second problem of instability of the coupling remains. To see this, consider a slight perturbation of the point configuration in Figure 5, this will result in a mean of either d_{\blacktriangle} or d_{\star} since the perturbation will either result in either one of the two configurations to be optimal couplings. To address this problem we consider diagrams as probabilistic objects, given a diagram with off diagonal points $\{p_1, \dots, p_\ell\}$ we consider the collection as a probability density function on point configurations in \mathbb{R}^2 with

$$f(x) = \frac{1}{\ell} \sum_{i=1}^{\ell} k(x, p_i; h),$$

where $k(x, p_i; h)$ is a density function centered at p_i and localized to a region h in \mathbb{R}^2 . In this case, slightly perturbing the points in Figure 5 will result in a mean diagram of

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

where $p_{\blacktriangle} = 1 - p_{\blackstar} \approx .5$ since the probability of the two optimal matches of the localized probability distributions around the points in Figure 5 is about equal, whether the points are slightly perturbed or not.

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 10.

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

Here the sum is taken over all possible couplings G on the set of diagrams, and $\text{mean}_X(G) \in S$ is the mean diagram for the specific coupling G . Note that the mean $\text{mean}_X(G)$ of a set of N diagrams, each with at most K points, can itself have at most NK points, so that μ_X , as defined, is indeed an element of $\mathcal{P}(S_{M,NK})$. The weights $\mathbb{P}(\mathcal{H} = G)$ are derived from a random variable \mathcal{H} which can be thought of either as a probabilistic coupling 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 matching. 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 $S_{M,K}$. Sometimes, where appropriate, we will also use $X = \bigcup_i X_i$ to represent the set of off-diagonal points in all the X_i . We now define \mathcal{H} , a coupling valued random variable.

First, we fix an $\alpha > 0$ and a probability distribution η on \mathbb{R}^2 which is absolutely continuous with respect to Lebesgue measure λ . This means that η has a Radon-Nikodym derivative; there exists a measurable non-negative function f on \mathbb{R}^2 such that the equation $\eta(A) = \int_A f d\lambda$ holds for every measurable set A . We further require that f is bounded with support of a ball of radius α , $B(0, \alpha)$, and is radially symmetric. Examples of distributions satisfying these requirements are the uniform distribution $\mathcal{U}(B_\alpha(0))$ or the truncated normal distribution $\mathcal{N}(0, \sigma^2, \alpha)$ which is just the portion of the standard normal distribution $\mathcal{N}(0, \sigma^2)$ contained inside of $B_\alpha(0)$ and normalized to have total mass 1. For each $x \in \mathbb{R}^2$, let η_x be the translation of η centered around x .

We will often be considering the measure of a ball of a radius smaller than α centered around x , so we use the following notation

$$\eta(r) = \eta_x(B_r(x)).$$

By construction we know that $\eta(r) \leq \pi r^2 f(0)$ where f is the probability density function corresponding to the density η .

We now define, for each input diagram X_i , a persistence-diagram-valued random variable X'_i , as follows. For each off diagonal point $x \in X_i$ which is more than distance α from the diagonal, draw a point from the distribution η_x and add it to the diagram X'_i . For each off-diagonal point $x \in X_i$ which is less than distance α from the diagonal, draw a point x' from η_x and add it to the diagram X'_i only if it is contained in the ball of radius $\beta = \|x - \Delta\|$ centered at x . In this way, the probability that a point gets added to the diagram from one of these points close to the diagonal decreases as the distance to the diagonal decreases. See Figure 7.

For each particular draw of the random variables X'_1, \dots, X'_N , we can compute a mean diagram via Algorithm 1. In particular, we need the coupling used by the algorithm, and can discard the computed mean diagram. Since each point in a draw of X'_i is associated to a point in X_i , we consider the coupling used to compute the mean of a drawn $\{X'_i\}$ to be the same as the coupling using the corresponding points of X_i . However, since some points in the X_i did not get corresponding points in the draw X'_i , we can extend this to a full coupling by adding the trivial selection for these points. That is, if a point

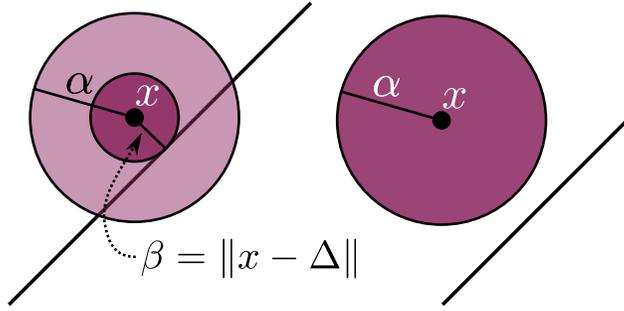


Figure 7: The method for drawing points. For a point $x \in X_i$ which is more than distance α from the diagonal as at right, a point is drawn from the distribution η_x , which has mean α and is contained in the ball of radius α centered at x . This point is then added to the diagram X'_i . For a point $x \in X_i$ which is less than distance α from the diagonal as at left, a point is still drawn from the distribution η_x , however the point is only added to X'_i if it is inside the ball of radius $\beta = \|x - \Delta\|$ centered at x .

$x \in X_i$ did not lead to a drawn point in X'_i , we add the selection which chooses x for X_i and Δ for every other diagram. This completes the definition of the random variable \mathcal{H} : as described above, each draw of X'_1, \dots, X'_N leads to the draw $\mathcal{H} = G$, where G is the coupling on X described above.

4.3 Example

Here is an example to make the discussion above a little more clear. Consider the three overlaid diagrams in Figure 8a. Points are drawn in the ball of radius α centered at each point. Since a, c, h, g, y , and z are near the diagonal there is a chance that no point is drawn for them. In this particular draw, given in Figure 8b, no point is drawn for a or c .

When computing the mean of the diagrams in Figure 8b, the coupling used is

$$\begin{array}{c}
 d'_\star \quad d'_\square \quad d'_\circ \\
 1 \begin{pmatrix} b' & x' & f' \\
 2 \begin{pmatrix} \Delta & y' & g' \\
 3 \begin{pmatrix} \Delta & \Delta & h' \\
 4 \begin{pmatrix} \Delta & z' & \Delta \end{pmatrix}. \tag{5}
 \end{array}$$

So, to find the corresponding coupling 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:

$$\begin{array}{c}
 d_\star \quad d_\square \quad d_\circ \\
 1 \begin{pmatrix} b & x & f \\
 2 \begin{pmatrix} \Delta & y & g \\
 3 \begin{pmatrix} \Delta & \Delta & h \\
 4 \begin{pmatrix} \Delta & z & \Delta \\
 5 \begin{pmatrix} a & \Delta & \Delta \\
 6 \begin{pmatrix} c & \Delta & \Delta \end{pmatrix}. \tag{6}
 \end{array}$$

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 start by developing some more detailed theory about the measures η_x , as well as fixing some notation which will be needed in the proof.

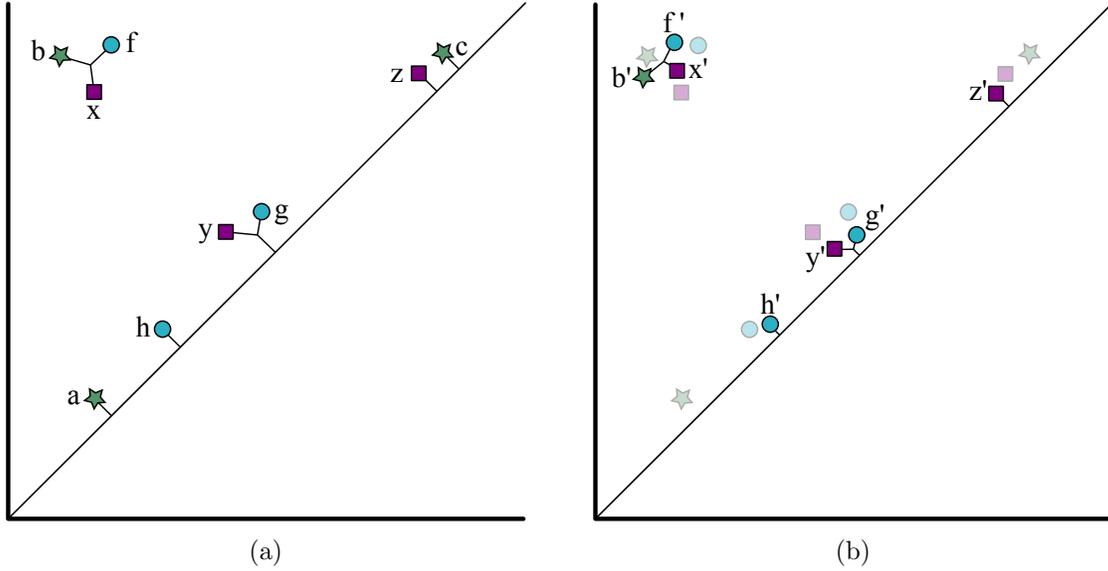


Figure 8: An example of corresponding couplings for a given draw. The original diagrams are d_{\star} , d_{\square} and d_{\circ} in Figure (a). A point is drawn near each point away from the diagonal, and points are drawn for some points near the diagonal to construct d'_{\star} , d'_{\square} and d'_{\circ} in Figure (b). The coupling for the mean of these three diagrams is computed using Algorithm 1 and the associated coupling is given in Equation 5. Then the coupling is converted to a coupling for d_{\star} , d_{\square} and d_{\circ} in Equation 6 and drawn in Figure (a).

5.1 Product Measures

For each point $x \in \mathbb{R}^2$, recall that we have the probability distribution η_x , with support contained in the ball $B_{\alpha}(x)$. We use this to define a new distribution η'_x , given by the formula:

$$\eta'_x = \frac{1}{\eta_x(B_{\|x-\Delta\|}(x))} \eta_x \Big|_{B_{\|x-\Delta\|}(x)} \quad (7)$$

Note that $\eta'_x = \eta_x$ iff x is more than α away from the diagonal.

As before, we start with a set $X = \{X_i\}$ of input diagrams, and we let $X = \bigcup_i X_i$ be the set of all off-diagonal points in all the X_i . Recall that, for each draw of the set of random variables $\{X'_i\}$, some of the points in each X_i created a drawn point in X'_i and some did not. To formalize this, we define the subset-valued random variable $\Omega \subset X$ to be the set of points in X which do indeed lead to a drawn point; note that each draw of Ω will always contain the points in X which are more than α away from the diagonal, but it may contain more points as well.

For each subset $\omega \subset X$, consider the space $(\mathbb{R}^2)^{|\omega|}$, where we have associated a copy of \mathbb{R}^2 to each point in ω . For each coupling G on X , we define $\mathcal{U}_{\omega, G} \subset (\mathbb{R}^2)^{|\omega|}$ to be the set of points which lead to this coupling. Notice that this set is completely independent of the locations of the points from the original diagrams; it relies only on the combinatorial structure of G .

Let $\eta'_{\omega} = \prod_{x \in \omega} \eta'_x$ be the product measure on $(\mathbb{R}^2)^{|\omega|}$. Then the probability that G is the coupling used assuming ω is chosen is

$$\mathbb{P}(\mathcal{H} = G \mid \Omega = \omega) = \eta'_{\omega}(\mathcal{U}_{\omega, G}).$$

For example, say X consists of two diagrams, X_1 and X_2 , where X_1 has a single off-diagonal point x_1 , and X_2 has a single off diagonal point x_2 . Then there are exactly two possible couplings:

$$G_1 = \frac{1}{2} \begin{pmatrix} X_1 & X_2 \\ x_1 & \Delta \\ \Delta & x_2 \end{pmatrix}, \quad G_2 = \frac{1}{2} \begin{pmatrix} X_1 & X_2 \\ x_1 & x_2 \end{pmatrix}.$$

If $\omega = \{x_1, x_2\}$, then the space we are interested in is $(\mathbb{R}^2)^2$. So $\mathcal{U}_{\omega, G_1}$ is the set of points which would rather be paired with the diagonal than with each other, thus

$$\mathcal{U}_{\omega, G_1} = \{(a, b) \in (\mathbb{R}^2)^2 \mid \|a - \Delta\|^2 + \|b - \Delta\|^2 \leq \|a - b\|^2\}.$$

Likewise, $\mathcal{U}_{\omega, G_2}$ is the set of points where being paired to each other is better, so

$$\mathcal{U}_{\omega, G_2} = \{(a, b) \in (\mathbb{R}^2)^2 \mid \|a - b\|^2 \leq \|a - \Delta\|^2 + \|b - \Delta\|^2\}.$$

For the purposes of determining the probability of G_1 , we must take the locations of x_1 and x_2 into account. Now we are only interested in the $(a, b) \in \mathcal{U}_{\omega, G_1}$ for which a is in the support of η'_{x_1} and b is in the support of η'_{x_2} . So, the probability that G_1 is chosen assuming both points are picked is

$$\mathbb{P}(\mathcal{H} = G_1 \mid \Omega = \omega) = \eta'_{\omega}(\mathcal{U}_{\omega, G_1}) = (\eta'_{x_1} \times \eta'_{x_2})(\mathcal{U}_{\omega, G_1}).$$

Likewise, the probability that G_2 is chosen assuming both points are picked is

$$\mathbb{P}(\mathcal{H} = G_2 \mid \Omega = \omega) = \eta'_{\omega}(\mathcal{U}_{\omega, G_2}) = (\eta'_{x_1} \times \eta'_{x_2})(\mathcal{U}_{\omega, G_2}).$$

This construction is particularly useful since we can consider the same coupling for two sets of diagrams which are close. Since $\mathcal{U}_{\omega, G}$ is independent of the actual diagrams, the only thing varying is the measure.

5.2 The Proof

We will prove Hölder continuity for the map that takes a set of persistence diagrams to its mean distribution. First we give a lemma which bounds the distance between the measures associated to nearby points in the plane. The proof, which is purely analytical, can be found in the Appendix.

Lemma 11. *Let f_x and f_y be the Radon-Nikodym derivatives of $\eta_x|_{B(x, \|x-\Delta\|)}$ and $\eta_y|_{B(y, \|y-\Delta\|)}$ respectively. Then*

$$\int |f_x - f_y| d\lambda \leq 4\pi\alpha \|x - y\| f(0)$$

where f is the Radon-Nikodym derivative of η .

Next we prove a result which relates the mean of a set of diagrams to the mean of a new set which is obtained by simply moving one of the points in one of the diagrams to a new location.

Proposition 12. *Let $Z = \{Z_1, Z_2, \dots, Z_N\} \in (S_{M,K})^N$ with mean $\mu \in \mathcal{P}(S_{M,NK})$. Let \hat{Z}_1 be the same diagram as Z_1 except that one of the off-diagonal points, x , has been moved to the new location y (while still maintaining $\hat{Z}_1 \in S_{M,K}$). Let $\hat{\mu}$ be the mean of $\hat{Z} = \{\hat{Z}_1, Z_2, Z_3, \dots, Z_N\}$. Then*

$$W_2(\mu, \hat{\mu})^2 \leq 8\pi\alpha f(0) \|x - y\| \overline{M}^2 + \|x - y\|^2.$$

Here \overline{M} is the maximal distance between diagrams in $S_{M,NK}$ and f is the Radon-Nikodym derivative of η .

Proof. Let \mathcal{G} be the set of couplings on the set of diagrams $Z = \{Z_1, Z_2, \dots, Z_N\}$ and $\hat{\mathcal{G}}$ the set of couplings on the set of diagrams $\hat{Z} = \{\hat{Z}_1, Z_2, \dots, Z_N\}$. The bijection $\varphi : Z_1 \rightarrow \hat{Z}_1$, which sends x to y and fixes every other point, induces a bijection between \mathcal{G} and $\hat{\mathcal{G}}$ which we will also denote by φ .

In order to understand the bound we will put on $W_2(\mu, \hat{\mu})$, it is best to use the earth mover analogy for Wasserstein distance. Each distribution is thought of as dirt piles, and the cost of moving a dirt pile is the amount of dirt times the distance it must be moved.

In this analogy, since $\mu = \sum \mathbb{P}(\mathcal{H} = G) \delta_{\text{mean}(G)}$, $\text{mean}(G)$ is the location of the dirt piles and $\mathbb{P}(\mathcal{H} = G)$ is the quantity of dirt in that location. Then we will pair the couplings in \mathcal{G} with couplings in $\hat{\mathcal{G}}$ via φ . This pairs up the piles in a way that the amount of dirt in the matched piles are approximately

the same. Then we can argue that we move most of the dirt a short distance, and the amount of leftover dirt is small enough to move it anywhere without too much penalty.

Note that $\min\{\mathbb{P}(\mathcal{H}_Z = G), \mathbb{P}(\mathcal{H}_{\hat{Z}} = \varphi(G))\}$ for $G \in \mathcal{G}$ is the maximum amount of dirt that can be moved from $\text{mean}_Z(G)$ to $\text{mean}_{\hat{Z}}(\varphi(G))$. This fact, combined with the argument in the paragraph above, gives the inequality

$$\begin{aligned} W_2(\mu, \hat{\mu})^2 &\leq \sum_{G \in \mathcal{G}} \min\{\mathbb{P}(\mathcal{H}_Z = G), \mathbb{P}(\mathcal{H}_{\hat{Z}} = \varphi(G))\} \cdot W_2(\text{mean}_Z(G), \text{mean}_{\hat{Z}}(\varphi(G)))^2 \\ &\quad + \sum_{G \in \mathcal{G}} |\mathbb{P}(\mathcal{H}_Z = G) - \mathbb{P}(\mathcal{H}_{\hat{Z}} = \varphi(G))| \cdot \overline{M}^2 \end{aligned} \quad (8)$$

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

First, we will bound $W_2(\text{mean}_Z(G), \text{mean}_{\hat{Z}}(\varphi(G)))$. For selections m in Z which do not contain x the corresponding selection, $\varphi(m)$, in \hat{Z} is the same set of points and hence $\text{mean}_Z(m) = \text{mean}_{\hat{Z}}(\varphi(m))$.

If m is a selection of Z with points x, z_2, \dots, z_N then the corresponding selection $\varphi(m)$ in \hat{Z} is $\varphi(x), z_2, \dots, z_N$. Consider the means of the selections m and $\varphi(m)$ in rotated coordinates with axes $(1/\sqrt{2}, 1/\sqrt{2})$ and $(-1/\sqrt{2}, 1/\sqrt{2})$. Without loss of generality, let x, z_2, \dots, z_k be the off-diagonal points, which leaves $N - k$ copies of the diagonal. Writing $z_i = (u_i, v_i)$, $x = (a, b)$ and $\varphi(x) = (\hat{a}, \hat{b})$ in these coordinates, we have

$$\begin{aligned} \text{mean}_Z(m) &= \left(\frac{1}{k} \left(a + \sum_{i=2}^k u_i \right), \frac{1}{N} \left(b + \sum_{i=2}^k v_i \right) \right), \\ \text{mean}_{\hat{Z}}(\varphi(m)) &= \left(\frac{1}{k} \left(\hat{a} + \sum_{i=2}^k u_i \right), \frac{1}{N} \left(\hat{b} + \sum_{i=2}^k v_i \right) \right), \end{aligned}$$

Thus $\|\text{mean}_Z(m) - \text{mean}_{\hat{Z}}(\varphi(m))\|^2 = \left(\frac{a-\hat{a}}{k}\right)^2 + \left(\frac{b-\hat{b}}{N}\right)^2 \leq (a - \hat{a})^2 + (b - \hat{b})^2 = \|x - \varphi(x)\|^2$. Given a coupling G exactly one selection will involve x and hence for every coupling $G \in \mathcal{G}$ we have $W_2(\text{mean}_Z(G), \text{mean}_{\hat{Z}}(\varphi(G)))^2 \leq \|x - \varphi(x)\|^2$ and hence

$$\sum_{G \in \mathcal{G}} \min\{\mathbb{P}(\mathcal{H}_Z = G), \mathbb{P}(\mathcal{H}_{\hat{Z}} = \varphi(G))\} \cdot W_2(\text{mean}_Z(G), \text{mean}_{\hat{Z}}(\varphi(G)))^2 \leq \|x - \varphi(x)\|^2. \quad (9)$$

Let Ω and $\hat{\Omega}$ be random variables giving the subsets of Z and \hat{Z} , respectively, which do indeed lead to a drawn point.

We now wish to bound $\sum_{G \in \mathcal{G}} |\mathbb{P}(\mathcal{H}_Z = G) - \mathbb{P}(\mathcal{H}_{\hat{Z}} = \varphi(G))|$. Let us fix a subset $\omega \subset Z$ such that $x \in \omega$. For $z \in \mathbb{R}^2$ let f_z be the Radon-Nikodym derivative for the measure $\eta_z|_{B(z, \|z-\Delta\|)}$. Let F_ω and $F_{\varphi(\omega)}$ be functions over $z = (z_1, z_2, \dots, z_{|\omega|}) \in (\mathbb{R}^2)^{|\omega|}$ of the form $F_\omega(z) = f_x(z_1) \prod_{z_i \in \omega, z_i \neq x} f_{z_i}(z_i)$ and $F_{\varphi(\omega)}(z) = f_y(z_1) \prod_{x_i \in \omega, x_i \neq x} f_{x_i}(z_i)$.

Since G and $\varphi(G)$ have the same combinatorial structure, $\mathcal{U}_{\omega, G} = \mathcal{U}_{\varphi(\omega), \varphi(G)}$. Also observe that the set of points for which multiple couplings minimize the Wasserstein distance has Lebesgue measure zero and thus the distinct $\mathcal{U}_{G, \omega}$ are disjoint except for a set of measure zero. Therefore for each ω containing x :

$$\begin{aligned} &\sum_{G \in \mathcal{G}} |\mathbb{P}(\mathcal{H}_Z = G \text{ and } \Omega = \omega) - \mathbb{P}(\mathcal{H}_{\hat{Z}} = \varphi(G) \text{ and } \hat{\Omega} = \varphi(\omega))| \\ &\leq \sum_{G \in \mathcal{G}} \int_{\mathcal{U}_{G, \omega}} |F_\omega - F_{\varphi(\omega)}| d\Lambda \\ &\leq \int_{(\mathbb{R}^2)^{|\omega|}} |F_\omega - F_{\varphi(\omega)}| d\lambda \\ &\leq \frac{\mathbb{P}(\Omega = \omega)}{\mathbb{P}(x \in \Omega)} \int_{\mathbb{R}^2} |f_x - f_{\varphi(x)}| d\lambda \end{aligned}$$

From Lemma 14 we have $\int_{\mathbb{R}^2} |f_x - f_{\varphi(x)}| d\lambda \leq 4\pi\alpha f(0)\|x - \varphi(x)\|$.

We also know that $\mathbb{P}(x \in \Omega) = \sum_{\{\omega: x \in \omega\}} \mathbb{P}(\Omega = \omega)$ because if x is chosen in Ω then exactly one of the events $\Omega = \omega$ with $\omega \in \{\omega : x \in \omega\}$ occurs and the events $\Omega = \omega$ with $\omega \in \{\omega : x \in \omega\}$ are disjoint. Therefore,

$$\begin{aligned} & \sum_{\{\omega: x \in \omega\}} \sum_{G \in \mathcal{G}_X} |\mathbb{P}(\mathcal{H}_Z = G \text{ and } \Omega = \omega) - \mathbb{P}(\mathcal{H}_{\hat{Z}} = \varphi(G) \text{ and } \hat{\Omega} = \varphi(\omega))| \\ & \leq \sum_{\{\omega: x \in \omega\}} \frac{\mathbb{P}(\Omega = \omega)}{\mathbb{P}(x \in \Omega)} 4\pi\alpha f(0)\|x - \varphi(x)\| \\ & = 4\pi\alpha f(0)\|x - \varphi(x)\| \end{aligned} \quad (10)$$

Now consider the subsets ω such that $x \notin \omega$. Here

$$\frac{\mathbb{P}(\mathcal{H}_Z = G \text{ and } \Omega = \omega)}{\mathbb{P}(x \notin \Omega)} = \frac{\mathbb{P}(\mathcal{H}_{\hat{Z}} = \varphi(G) \text{ and } \Omega = \varphi(\omega))}{\mathbb{P}(\varphi(x) \notin \hat{\Omega})}$$

and hence

$$|\mathbb{P}(\mathcal{H}_Z = G \text{ and } \Omega = \omega) - \mathbb{P}(\mathcal{H}_{\hat{Z}} = \varphi(G) \text{ and } \hat{\Omega} = \varphi(\omega))| = \mathbb{P}(\mathcal{H}_Z = G \text{ and } \Omega = \omega) \left| 1 - \frac{\mathbb{P}(\varphi(x) \notin \hat{\Omega})}{\mathbb{P}(x \notin \Omega)} \right|.$$

Taking the sum over ω not containing x and all $G \in \mathcal{G}$ we have

$$\begin{aligned} & \sum_{\{\omega: x \notin \omega\}} \sum_{G \in \mathcal{G}} |\mathbb{P}(\mathcal{H}_Z = G \text{ and } \Omega = \omega) - \mathbb{P}(\mathcal{H}_{\hat{Z}} = \varphi(G) \text{ and } \hat{\Omega} = \varphi(\omega))| \\ & = \mathbb{P}(x \notin \Omega) \left| 1 - \frac{\mathbb{P}(\varphi(x) \notin \hat{\Omega})}{\mathbb{P}(x \notin \Omega)} \right| \\ & = |\mathbb{P}(x \notin \Omega) - \mathbb{P}(\varphi(x) \notin \hat{\Omega})| \\ & = |(1 - \mathbb{P}(x \in \Omega)) - (1 - \mathbb{P}(\varphi(x) \in \hat{\Omega}))| \\ & = |\mathbb{P}(x \in \Omega) - \mathbb{P}(\varphi(x) \in \hat{\Omega})| \end{aligned}$$

This is bounded by

$$|\mathbb{P}(x \in \Omega) - \mathbb{P}(\varphi(x) \in \hat{\Omega})| \leq \left| \int f_x d\lambda - \int f_{\varphi(x)} d\lambda \right| \leq \int |f_x - f_{\varphi(x)}| d\lambda \leq 4\pi\alpha f(0)\|x - \varphi(x)\|,$$

which implies

$$|\mathbb{P}(\mathcal{H}_Z = G \text{ and } \Omega = \omega) - \mathbb{P}(\mathcal{H}_{\hat{Z}} = \varphi(G) \text{ and } \hat{\Omega} = \varphi(\omega))| \leq 4\pi\alpha f(0)\|x - \varphi(x)\|. \quad (11)$$

Thus considering both (10) and (11), we have

$$\begin{aligned} & \sum_{G \in \mathcal{G}} |\mathbb{P}(\mathcal{H}_Z = G) - \mathbb{P}(\mathcal{H}_{\hat{Z}} = \varphi(G))| \\ & \leq \sum_{\omega} \sum_{G \in \mathcal{G}} |\mathbb{P}(\mathcal{H}_Z = G \text{ and } \Omega = \omega) - \mathbb{P}(\mathcal{H}_{\hat{Z}} = \varphi(G) \text{ and } \hat{\Omega} = \varphi(\omega))| \\ & = \sum_{\{\omega: x \in \omega\}} \sum_{G \in \mathcal{G}} |\mathbb{P}(\mathcal{H}_Z = G \text{ and } \Omega = \omega) - \mathbb{P}(\mathcal{H}_{\hat{Z}} = \varphi(G) \text{ and } \hat{\Omega} = \varphi(\omega))| \\ & \quad + \sum_{\{\omega: x \notin \omega\}} \sum_{G \in \mathcal{G}} |\mathbb{P}(\mathcal{H}_Z = G \text{ and } \Omega = \omega) - \mathbb{P}(\mathcal{H}_{\hat{Z}} = \varphi(G) \text{ and } \hat{\Omega} = \varphi(\omega))| \\ & \leq 8\pi\alpha f(0)\|x - \varphi(x)\| \end{aligned} \quad (12)$$

Combining inequalities (9) and (12) provides the final bound

$$W_2(\mu, \hat{\mu})^2 \leq \|x - \varphi(x)\|^2 + 8\pi\alpha f(0)\|x - \varphi(x)\|\overline{M}^2.$$

□

With these last two steps in hand, we can finally prove our main theorem.

Theorem 13. *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 mean as defined above. Then Φ is Hölder continuous with exponent $1/2$. That is, there exists a constant C such that

$$W_2(\mu_X, \mu_Y) \leq C\sqrt{W_2(X, Y)}$$

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

The main idea of the proof is, given two close sets of diagrams (X_1, \dots, X_N) and (Y_1, \dots, Y_N) , to create a bijection from most of the couplings of the X_i to most of the couplings of the Y_i in such a way that the probability of getting the corresponding couplings is similar and the associated mean diagrams are close. This allows for a transportation plan which moves most of the mass a short way, and we can then argue that although the rest of the mass could be moved a long distance, its weight is negligible. Equations 14 and 8 each encode this information for different transportation plans.

Proof. Let $X = (X_1, \dots, X_N)$ and $Y = (Y_1, \dots, Y_N)$ denote sets of diagrams in $(S_{M,K})^N$ with corresponding means μ_X and μ_Y . We wish to find a constant C such that $W_2(\mu_X, \mu_Y) \leq C\sqrt{W_2(X, Y)}$.

For the moment assume that $W_2(X, Y) \leq 1$.

For each i , let φ_i be an optimal bijection given by $W_2(X_i, Y_i)$. Let \tilde{X}_i be the diagram consisting of points x in X_i such that $\varphi_i(x) \neq \Delta$. Likewise, let \tilde{Y}_i be the diagram consisting of points $y \in Y_i$ such that $\varphi_i^{-1}(y) \neq \Delta$.

Let \mathcal{G}_X be the set of all couplings for the diagrams X_1, \dots, X_N , and let $\mathcal{G}_{\tilde{X}}$ be the set of all couplings for the diagrams $\tilde{X}_1, \dots, \tilde{X}_N$. \mathcal{G}_Y and $\mathcal{G}_{\tilde{Y}}$ are defined similarly. Note that since all diagrams are in $S_{M,K}$, they have finitely many off-diagonal points, and therefore finitely many couplings.

There is an injection $i_{\tilde{X}} : \mathcal{G}_{\tilde{X}} \hookrightarrow \mathcal{G}_X$ which takes a coupling $G \in \mathcal{G}_{\tilde{X}}$ and maps it to the coupling $G' \in \mathcal{G}_X$ which has all the same selections along with the trivial selection for each unused point $x \in X_i \setminus \tilde{X}_i$. Likewise, there is an injection $i_{\tilde{Y}} : \mathcal{G}_{\tilde{Y}} \hookrightarrow \mathcal{G}_Y$.

We will bound $W_2(\mu_X, \mu_Y)$ using the triangle inequality:

$$W_2(\mu_X, \mu_Y) \leq W_2(\mu_X, \mu_{\tilde{X}}) + W_2(\mu_{\tilde{X}}, \mu_{\tilde{Y}}) + W_2(\mu_{\tilde{Y}}, \mu_Y). \quad (13)$$

This will be done in two parts; first we will bound $W_2(\mu_X, \mu_{\tilde{X}})$ and $W_2(\mu_Y, \mu_{\tilde{Y}})$, then we will bound $W_2(\mu_{\tilde{X}}, \mu_{\tilde{Y}})$.

Part 1: In order to understand the bound we will put on $W_2(\mu_X, \mu_{\tilde{X}})$, it is best to use the earth mover analogy for Wasserstein distance. Each distribution is thought of as dirt piles, and the cost of moving a dirt pile is the amount of dirt times the distance it must be moved.

In this analogy, since $\mu_X = \sum \mathbb{P}(\mathcal{H} = G)\delta_{\text{mean}(G)}$, $\text{mean}(G)$ is the location of the dirt piles and $\mathbb{P}(\mathcal{H} = G)$ is the quantity of dirt in that location. Then we will pair up most of the couplings in \mathcal{G}_X with couplings in $\mathcal{G}_{\tilde{X}}$ in such a way that the dirt piles are approximately the same. Then we can argue that we move most of the dirt a short distance, and the amount of leftover dirt is small enough to move it anywhere without too much penalty.

The pairing comes from the map $i_{\tilde{X}} : \mathcal{G}_{\tilde{X}} \rightarrow \mathcal{G}_X$. Note that $\min\{\mathbb{P}(\mathcal{H}_X = i_{\tilde{X}}(G), \mathbb{P}(\mathcal{H}_{\tilde{X}} = G)\}$ for $G \in \mathcal{G}_{\tilde{X}}$ is the maximum amount of dirt that can be moved from $\text{mean}_X(i_{\tilde{X}}(G))$ to $\text{mean}_{\tilde{X}}(G)$. The

above argument gives the inequality

$$\begin{aligned}
W_2(\mu_X, \mu_{\tilde{X}})^2 &\leq \sum_{G \in \mathcal{G}_{\tilde{X}}} \min\{\mathbb{P}(\mathcal{H}_X = i_{\tilde{X}}(G), \mathbb{P}(\mathcal{H}_{\tilde{X}} = G)\} \cdot W_2(\text{mean}_X(i_{\tilde{X}}(G)), \text{mean}_{\tilde{X}}(G))^2 \\
&\quad + \sum_{G \in \mathcal{G}_{\tilde{X}}} |\mathbb{P}(\mathcal{H}_X = i_{\tilde{X}}(G)) - \mathbb{P}(\mathcal{H}_{\tilde{X}} = G)| \cdot \overline{M}^2 \\
&\quad + \sum_{\substack{G \in \mathcal{G}_X \\ G \notin \text{Im}(i_{\tilde{X}})}} |\mathbb{P}(\mathcal{H}_X = G)| \cdot \overline{M}^2
\end{aligned} \tag{14}$$

where \overline{M} is the maximum distance between any two diagrams in $S_{M,NK}$. In this equation, the first term corresponds to moving as much dirt as possible from $\text{mean}_X(i_{\tilde{X}}(G))$ to $\text{mean}_{\tilde{X}}(G)$, the second term accounts for the leftover dirt from this motion, and the last term is the amount of dirt from couplings which have no pair.

In order to bound $W_2(\text{mean}_{\tilde{X}}(G), \text{mean}_X(i_{\tilde{X}}(G)))^2$, 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. Thus, using the bijection sending each of these additional points to the diagonal,

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

Therefore,

$$\sum_{G \in \mathcal{G}_{\tilde{X}}} \min\{\mathbb{P}(\mathcal{H}_X = i_{\tilde{X}}(G)), \mathbb{P}(\mathcal{H}_{\tilde{X}} = G)\} \cdot W_2(\text{mean}_X(i_{\tilde{X}}(G)), \text{mean}_{\tilde{X}}(G))^2 \leq \sum_{x \in X \setminus \tilde{X}} \frac{\|x - \Delta\|^2}{N^2} \tag{15}$$

Let Ω and $\tilde{\Omega}$ be random variables defined on subsets of X and \tilde{X} respectively which gives the set of points which do indeed lead to a drawn point. Note that

$$\mathbb{P}(\mathcal{H}_X = i_{\tilde{X}}(G)) = \mathbb{P}(\mathcal{H}_X = i_{\tilde{X}}(G) \text{ and } \Omega \subset \tilde{X}) + \mathbb{P}(\mathcal{H}_X = i_{\tilde{X}}(G) \text{ and } \Omega \not\subset \tilde{X})$$

and

$$\mathbb{P}(\mathcal{H}_{\tilde{X}} = G) = \mathbb{P}(\mathcal{H}_{\tilde{X}} = G \text{ and } \tilde{\Omega} \subset \tilde{X}).$$

Hence,

$$\begin{aligned}
&\sum_{G \in \mathcal{G}_{\tilde{X}}} |\mathbb{P}(\mathcal{H}_X = i_{\tilde{X}}(G)) - \mathbb{P}(\mathcal{H}_{\tilde{X}} = G)| \\
&\leq \sum_{G \in \mathcal{G}_{\tilde{X}}} |\mathbb{P}(\mathcal{H}_X = i_{\tilde{X}}(G) \text{ and } \Omega \subset \tilde{X}) - \mathbb{P}(\mathcal{H}_{\tilde{X}} = G \text{ and } \tilde{\Omega} \subset \tilde{X})| \\
&\quad + \sum_{G \in \mathcal{G}_{\tilde{X}}} \mathbb{P}(\mathcal{H}_X = i_{\tilde{X}}(G) \text{ and } \Omega \not\subset \tilde{X}).
\end{aligned} \tag{16}$$

Note that for any subset $\omega \subset \tilde{X}$ and coupling $G \in \mathcal{G}_{\tilde{X}}$,

$$\begin{aligned}
\mathbb{P}(\mathcal{H}_X = i_{\tilde{X}}(G) \text{ and } \Omega = \omega) &= \mathbb{P}(\mathcal{H}_X = i_{\tilde{X}}(G) \mid \Omega = \omega) \cdot \mathbb{P}(\Omega = \omega) \\
&= \eta'_\omega(\mathcal{U}_{i_{\tilde{X}}(G), \omega}) \prod_{x \in \omega} \mathbb{P}(x \in \Omega) \prod_{\substack{x \in X \\ x \notin \omega}} \mathbb{P}(x \notin \Omega)
\end{aligned}$$

and likewise

$$\begin{aligned}
\mathbb{P}(\mathcal{H}_{\tilde{X}} = G \text{ and } \Omega = \omega) &= \mathbb{P}(\mathcal{H}_{\tilde{X}} = G \mid \Omega = \omega) \cdot \mathbb{P}(\tilde{\Omega} = \omega) \\
&= \eta'_\omega(\mathcal{U}_{G, \omega}) \prod_{x \in \omega} \mathbb{P}(x \in \Omega) \prod_{\substack{x \in \tilde{X} \\ x \notin \omega}} \mathbb{P}(x \notin \tilde{\Omega}).
\end{aligned}$$

Comparing these two equations gives

$$\mathbb{P}(\mathcal{H}_X = i_{\tilde{X}}(G) \text{ and } \Omega = \omega) = \mathbb{P}(\mathcal{H}_{\tilde{X}} = G \text{ and } \Omega = \omega) \cdot \prod_{x \in X \setminus \tilde{X}} \mathbb{P}(x \notin \Omega).$$

This equation with the triangle inequality gives

$$\begin{aligned} & \sum_{G \in \mathcal{G}_{\tilde{X}}} |\mathbb{P}(\mathcal{H}_X = i_{\tilde{X}}(G) \text{ and } \Omega \subset \tilde{X}) - \mathbb{P}(\mathcal{H}_{\tilde{X}} = G \text{ and } \Omega \subset \tilde{X})| \\ & \leq \sum_{\omega \subset \tilde{X}} \sum_{G \in \mathcal{G}_{\tilde{X}}} |\mathbb{P}(\mathcal{H}_{\tilde{X}} = G \text{ and } \Omega = \omega) - \mathbb{P}(\mathcal{H}_{\tilde{X}} = G \text{ and } \Omega = \omega)| \\ & \leq \sum_{\omega \subset \tilde{X}} \sum_{G \in \mathcal{G}_{\tilde{X}}} \mathbb{P}(\mathcal{H}_{\tilde{X}} = G \text{ and } \Omega = \omega) \cdot \left(1 - \prod_{x \in X \setminus \tilde{X}} \mathbb{P}(x \notin \Omega)\right) \\ & \leq \sum_{\omega \subset \tilde{X}} \sum_{G \in \mathcal{G}_{\tilde{X}}} \mathbb{P}(\mathcal{H}_{\tilde{X}} = G \text{ and } \Omega = \omega) \cdot \sum_{x \in X \setminus \tilde{X}} \mathbb{P}(x \in \Omega), \end{aligned} \tag{17}$$

where the final inequality follows from the union bound. From our method of drawing points $\mathbb{P}(x \in \Omega) = \eta(\|x - \Delta\|) \leq f(0)\|x - \Delta\|^2$. Thus we can further continue equation (17) as follows.

$$\begin{aligned} & \sum_{G \in \mathcal{G}_{\tilde{X}}} |\mathbb{P}(\mathcal{H}_X = i_{\tilde{X}}(G) \text{ and } \Omega \subset \tilde{X}) - \mathbb{P}(\mathcal{H}_{\tilde{X}} = G \text{ and } \Omega \subset \tilde{X})| \\ & \leq \mathbb{P}(\Omega \subset \tilde{X}) \sum_{x \in X \setminus \tilde{X}} f(0)\|x - \Delta\|^2 \\ & \leq \sum_{x \in X \setminus \tilde{X}} f(0)\pi\|x - \Delta\|^2 \end{aligned} \tag{18}$$

This bounds the first piece of Equation 16, and thus all that remains to be bounded from Equation 14 is the leftover from Equation 16 and the leftover from Equation 14,

$$\sum_{G \in \mathcal{G}_{\tilde{X}}} \mathbb{P}(\mathcal{H}_X = i_{\tilde{X}}(G) \text{ and } \tilde{\Omega} \not\subset \tilde{X}) + \sum_{\substack{G \in \mathcal{G}_X \\ G \notin \text{Im}(i_{\tilde{X}})}} |\mathbb{P}(\mathcal{H}_X = G)|.$$

Note that for any $G \in \mathcal{G}_X \setminus \text{Im}(i_{\tilde{X}})$, it is impossible to form the coupling G without at least one point outside of \tilde{X} , so

$$\mathcal{P}(H_X = G) = \mathbb{P}(\mathcal{H}_X = G \text{ and } \Omega \not\subset \tilde{X}).$$

Therefore,

$$\begin{aligned} & \sum_{G \in \mathcal{G}_{\tilde{X}}} \mathbb{P}(\mathcal{H}_X = i_{\tilde{X}}(G) \text{ and } \Omega \not\subset \tilde{X}) + \sum_{\substack{G \in \mathcal{G}_X \\ G \notin \text{Im}(i_{\tilde{X}})}} |\mathbb{P}(\mathcal{H}_X = G)| \\ & = \sum_{G \in \mathcal{G}_X} \mathbb{P}(\mathcal{H}_X = G \text{ and } \Omega \not\subset \tilde{X}) \\ & = \mathbb{P}(\Omega \not\subset \tilde{X}) \\ & \leq \sum_{x \in X \setminus \tilde{X}} \eta(\|x - \Delta\|) \\ & \leq \sum_{x \in X \setminus \tilde{X}} f(0)\pi\|x - \Delta\|^2 \end{aligned} \tag{19}$$

So combining Equation 14 with Equations 15, 16, 18, and 19, we have

$$W_2(\mu_X, \mu_{\tilde{X}})^2 \leq (2\pi f(0)\overline{M}^2 + 1/N^2) \sum_{x \in X \setminus \tilde{X}} \|x - \Delta\|^2 \leq (2\pi f(0)\overline{M}^2 + 1/N^2) W_2(X, Y)^2.$$

Note that the argument for Y is similar, so we also have

$$W_2(\mu_Y, \mu_{\tilde{Y}})^2 \leq (2\pi f(0)\overline{M}^2 + 1/N^2)W_2(X, Y)^2.$$

Together these imply that

$$W_2(\mu_X, \mu_{\tilde{X}}) + W_2(\mu_Y, \mu_{\tilde{Y}}) \leq 2\sqrt{2\pi f(0)\overline{M}^2 + 1/N^2}W_2(X, Y) \quad (20)$$

Part 2: We now wish to bound $W_2(\mu_{\tilde{X}}, \mu_{\tilde{Y}})$. To do this we will use the triangle inequality alongside Proposition 12.

Consider an arbitrary ordering of the points in \tilde{X} and order the points in \tilde{Y} accordingly via φ . For each j , we create a set of diagrams X^j by taking the first j points to be from \tilde{Y} and the remaining points from \tilde{X} . Note that $X^0 = \tilde{X}$ and $X^m = \tilde{Y}$, where m is the total number of off-diagonal points contained in the diagrams in \tilde{X} , and therefore the triangle inequality gives:

$$W_2(\mu_{\tilde{X}}, \mu_{\tilde{Y}}) \leq \sum_{j=0}^{m-1} W_2(\mu_{X^j}, \mu_{X^{j+1}}).$$

If x is the point in X^j which is moved to $\varphi(x)$ in X^{j+1} then Proposition 12 shows that $W_2(\mu_{X^j}, \mu_{X^{j+1}})^2 \leq 8\pi f(0)\alpha\|x - \varphi(x)\|\overline{M}^2 + \|x - \varphi(x)\|^2$. Using the stipulation that $W_2(X, Y) \leq 1$ we know that $\|x - \varphi(x)\| \leq 1$ for every $x \in \tilde{X}$ and hence

$$W_2(\mu_{X^j}, \mu_{X^{j+1}}) \leq \sqrt{(8\pi f(0)\alpha\overline{M}^2 + 1)\|x - \varphi(x)\|}.$$

This implies that

$$W_2(\mu_{\tilde{X}}, \mu_{\tilde{Y}}) \leq \sqrt{8\pi f(0)\alpha\overline{M}^2 + 1} \sum_{x \in \tilde{X}} \sqrt{\|x - \varphi(x)\|}.$$

Now $\|x - \varphi(x)\| \leq W_2(X, Y)$ for all $x \in \tilde{X}$. Since there are at most NK points in \tilde{X} we know

$$W_2(\mu_{\tilde{X}}, \mu_{\tilde{Y}}) \leq NK\sqrt{W_2(X, Y)}\sqrt{8\pi f(0)\alpha\overline{M}^2 + 1}. \quad (21)$$

We finally can say, assuming $W_2(X, Y) \leq 1$, that

$$\begin{aligned} W_2(\mu_X, \mu_Y) &\leq W_2(\mu_X, \mu_{\tilde{X}}) + W_2(\mu_{\tilde{X}}, \mu_{\tilde{Y}}) + W_2(\mu_{\tilde{Y}}, \mu_Y) \\ &\leq 2W_2(X, Y)\sqrt{2\pi f(0)\overline{M}^2 + 1/N^2} + NK\sqrt{W_2(X, Y)}\sqrt{8\pi f(0)\alpha\overline{M}^2 + 1} \\ &\leq (2\sqrt{2\pi f(0)\overline{M}^2 + 1/N^2} + NK\sqrt{8\pi f(0)\alpha\overline{M}^2 + 1})\sqrt{W_2(X, Y)}. \end{aligned} \quad (22)$$

Set

$$C := \max\{(2\sqrt{2\pi f(0)\overline{M}^2 + 1/N^2} + NK\sqrt{8\pi f(0)\alpha\overline{M}^2 + 1}), \overline{M}\}.$$

If $W_2(X, Y) \leq 1$ then equation (22) implies that $W_2(\mu_X, \mu_Y) \leq C\sqrt{W_2(X, Y)}$. If $W_2(X, Y) > 1$ we also know $W_2(\mu_X, \mu_Y) \leq C\sqrt{W_2(X, Y)}$ as \overline{M} satisfies $W_2(\mu_X, \mu_Y) \leq \overline{M} \leq C$. \square

6 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 matchings among the diagrams, with the weights given by the probability that a perturbation of the diagrams would produce that matching.

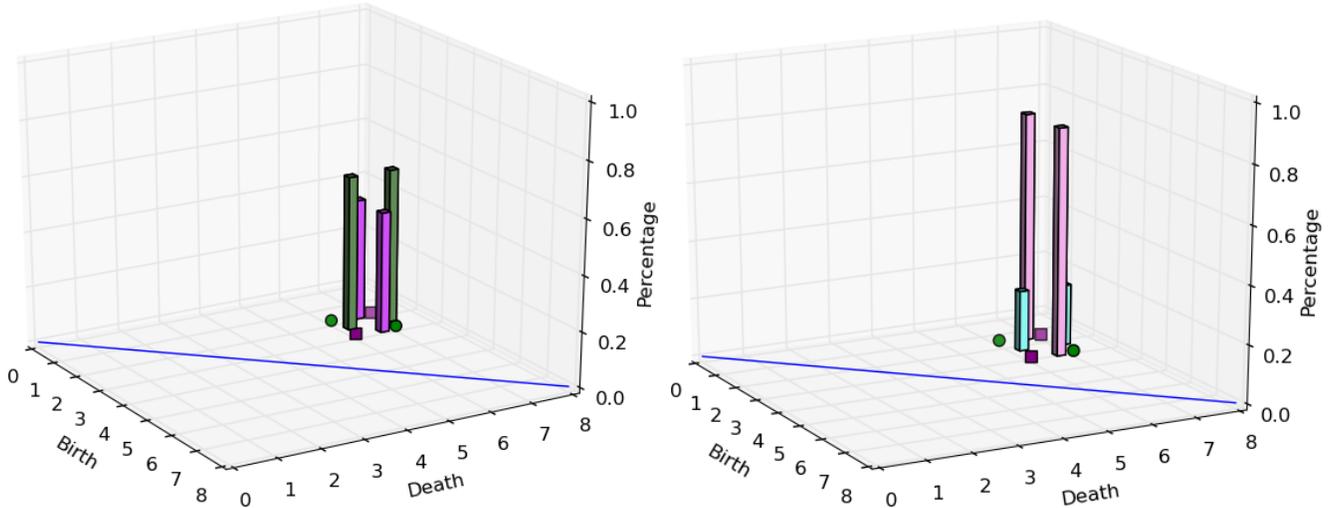


Figure 9: Change in mean distribution as a set of two diagrams moves through the problematic configuration of Figure 5. On the left we see the mean of two diagrams which form a rectangle that is slightly longer in the death-axis direction. On the right is the result for a rectangle that is quite a bit longer in the birth-axis direction.

In Figure 9, we show a resolution to the discontinuity issue raised in Figure 5, although the figure needs some explanation. The flat colored dots on the left side of this 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 η_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 Figure 10. Then we computed the one-dimensional 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.

Finally, we computed the mean distribution of these thirty diagrams, using the same approximation scheme as above. On the left of Figure 11, 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.

7 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,

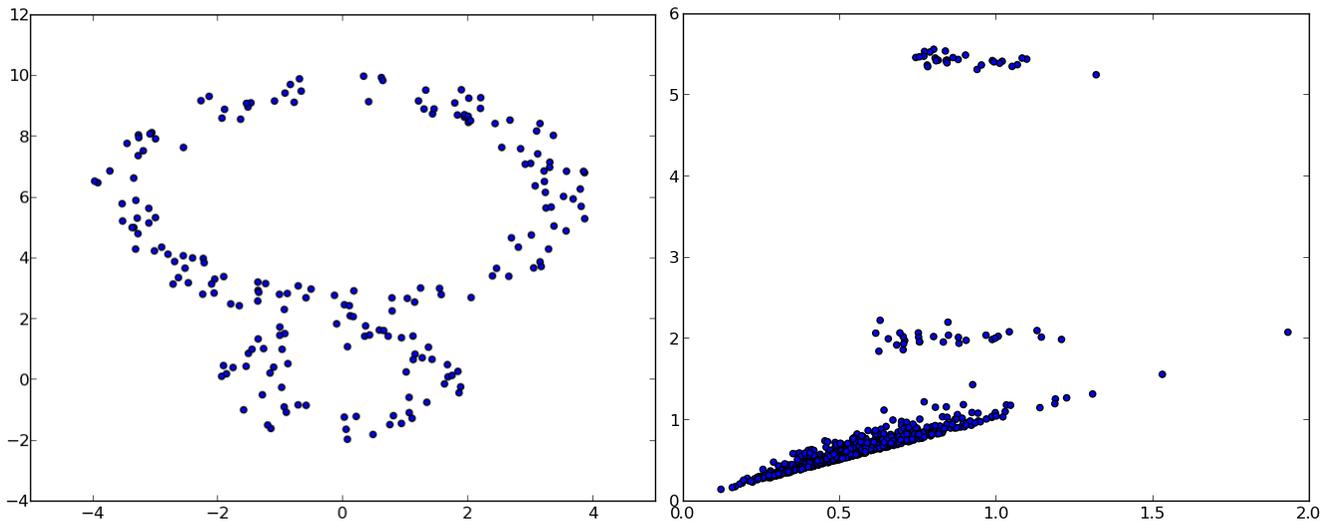


Figure 10: Thirty diagrams were created from thirty point clouds drawn from a double annulus. One such point cloud is shown on the left. All thirty diagrams are overlaid on the right.

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.

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 [26].

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.
- [5] Peter Bubenik. Statistical topology using persistence landscapes. *arXiv:1207.6437*, July, 2012.

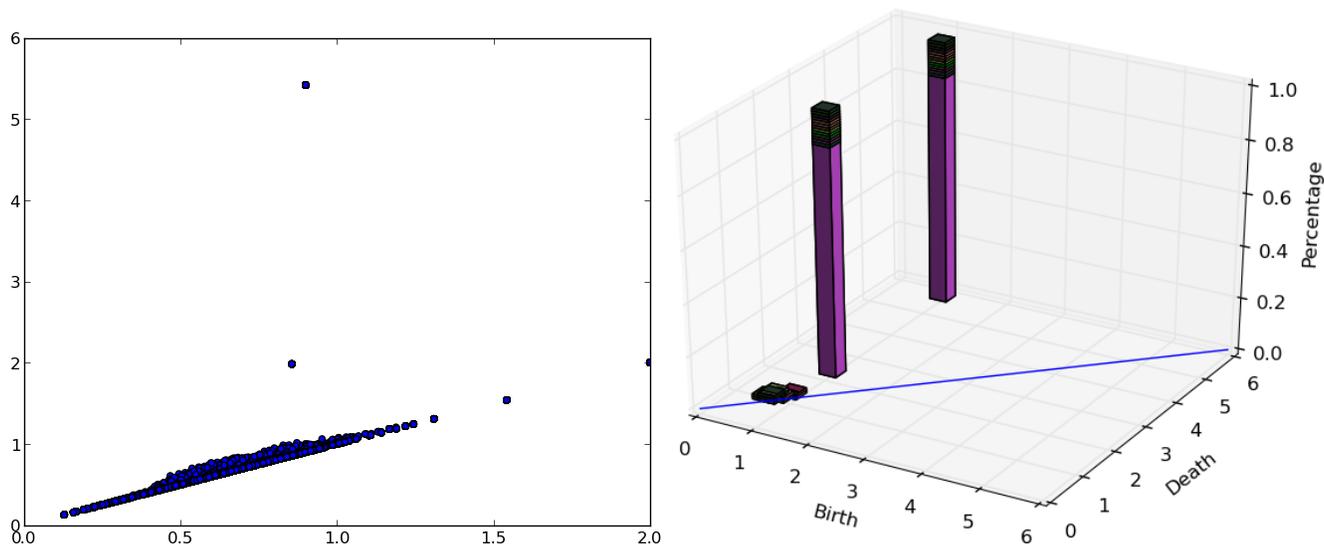


Figure 11: The mean distribution for a set of thirty diagrams sampled from a double annulus. The left side shows all positive-weight diagrams in the mean overlaid in one figure, while the right side indicates the weights in a three-dimensional plot.

- [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] Bradley Efron and Robert J. Tibshirani. *An Introduction to the Bootstrap*. Chapman and Hall/CRC, 1994.
- [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] Katharine Turner. Medians and means for sets of persistence diagrams. *In preparation*, 2013.
- [27] Katharine Turner, Yuriy Mileyko, Sayan Mukherjee, and John Harer. Fréchet means for distributions of persistence diagrams. *ArXiv*, page arXiv:1206.2790, 2011.
- [28] Cédric Villani. *Optimal Transport: Old and New*. Springer, 2009.

A Appendix

Lemma 14. *Let f_x and f_y be the Radon-Nikodym derivatives of $\eta_x|_{B(x, \|x-\Delta\|)}$ and $\eta_y|_{B(y, \|y-\Delta\|)}$ respectively. Then*

$$\int |f_x - f_y| d\lambda \leq 4\pi\alpha \|x - y\| f(0)$$

where f is the Radon-Nikodym derivative of η .

Proof. Without loss of generality we may assume that $\|x - \Delta\| \geq \|y - \Delta\|$.

The Radon-Nikodym derivative f is the limit of functions of the form $\sum_i a_i 1_{B(0, r_i)}$ where 1_S is the indicator function over the set S and the a_i are positive. By taking limits it is sufficient to prove the result for functions f of the form $\sum_i a_i 1_{B(0, r_i)}$. If $f = \sum_i a_i 1_{B(0, r_i)}$ then

$$f_y = \sum_{\{i: r_i \leq \|y-\Delta\|\}} a_i 1_{B(y, r_i)} + \sum_{\{i: r_i > \|y-\Delta\|\}} a_i 1_{B(y, \|y-\Delta\|)}$$

and

$$\begin{aligned}
f_x &= \sum_{\{i:r_i \leq \|x-\Delta\|\}} a_i 1_{B(x,r_i)} + \sum_{\{i:r_i > \|x-\Delta\|\}} a_i 1_{B(x,\|x-\Delta\|)} \\
&= \sum_{\{i:r_i \leq \|y-\Delta\|\}} a_i 1_{B(x,r_i)} + \sum_{\{i:\|y-\Delta\| < r_i \leq \|x-\Delta\|\}} a_i (1_{B(x,\|y-\Delta\|)} + 1_{A(x, [\|y-\Delta\|, r_i])}) \\
&\quad + \sum_{\{i:r_i > \|x-\Delta\|\}} a_i (1_{B(x,\|y-\Delta\|)} + 1_{A(x, [\|y-\Delta\|, \|x-\Delta\|])}).
\end{aligned}$$

where $A(x, [a, b])$ denotes the annulus centered at x with inner radius a and outer radius b . From these last two equations, we compute:

$$\begin{aligned}
\int |f_x - f_y| d\lambda &\leq \sum_{\{i:r_i \leq \|y-\Delta\|\}} a_i \int |1_{B(x,r_i)} - 1_{B(y,r_i)}| d\lambda + \sum_{\{i:r_i > \|y-\Delta\|\}} a_i \int |1_{B(x,\|y-\Delta\|)} - 1_{B(y,\|y-\Delta\|)}| d\lambda \\
&\quad + \sum_{\{i:\|y-\Delta\| < r_i \leq \|x-\Delta\|\}} a_i \int |1_{A(x, [\|y-\Delta\|, r_i])}| d\lambda + \sum_{\{i:r_i > \|x-\Delta\|\}} a_i \int |1_{A(x, [\|y-\Delta\|, \|x-\Delta\|])}| d\lambda.
\end{aligned}$$

For each i , if $\|x - y\| < r_i$, then a ball of diameter $2r_i - \|x - y\|$ fits in the intersection of $B(x, r_i)$ and $B(y, r_i)$. Thus, recalling that the absolute value of the difference between the indicator functions of two sets is just the indicator function of their symmetric difference, we compute:

$$\int |1_{B(x,r_i)} - 1_{B(y,r_i)}| d\lambda \leq 2\pi r_i^2 - 2\pi(r_i - \|x - y\|/2)^2 \leq 2\pi r_i \|x - y\| \leq 2\pi\alpha \|x - y\|. \quad (23)$$

The bound in (23) also holds if $\|x - y\| \geq r_i$ as then the integral is $2\pi r_i^2$.

For each r with $\alpha \geq \|x - \Delta\| \geq r \geq \|y - \Delta\|$

$$\begin{aligned}
\int |1_{A(x, [\|y-\Delta\|, r])}| d\lambda &= \pi r^2 - \pi \|y - \Delta\|^2 \\
&\leq \pi \|x - \Delta\|^2 - \pi \|y - \Delta\|^2 \\
&\leq \pi \|x - \Delta\|^2 - \pi (\|x - \Delta\| - \|x - y\|)^2 \\
&\leq 2\pi \|x - \Delta\| \|x - y\| \\
&\leq 2\alpha\pi \|x - y\|
\end{aligned}$$

Similarly if $\|x - \Delta\| > \alpha \geq r \geq \|y - \Delta\|$ then $\int |1_{A(x, [\|y-\Delta\|, r])}| d\lambda \leq 2\alpha\pi \|x - y\|$ also.

Together these imply that

$$\int |f_x - f_y| d\lambda \leq 4\alpha\pi \|x - y\| \sum_i a_i$$

Observing that $\sum_i a_i$ is the value of f at 0 completes the proof. □