Journal of Machine Learning Research 16 (2015) 77-102 Submitted 7/14; Published 1/15

Statistical Topological Data Analysis using Persistence Landscapes

Peter Bubenik PETER.BUBENIK@GMAIL.COM
Department of Mathematics
Cleveland State University
Cleveland, OH 44115-2214, USA

Editor: David Dunson

Abstract

We define a new topological summary for data that we call the persistence landscape. Since this summary lies in a vector space, it is easy to combine with tools from statistics and machine learning, in contrast to the standard topological summaries. Viewed as a random variable with values in a Banach space, this summary obeys a strong law of large numbers and a central limit theorem. We show how a number of standard statistical tests can be used for statistical inference using this summary. We also prove that this summary is stable and that it can be used to provide lower bounds for the bottleneck and Wasserstein distances.

Keywords: topological data analysis, statistical topology, persistent homology, topological summary, persistence landscape

1. Introduction

Topological data analysis (TDA) consists of a growing set of methods that provide insight to the "shape" of data (see the surveys Ghrist, 2008; Carlsson, 2009). These tools may be of particular use in understanding global features of high dimensional data that are not readily accessible using other techniques. The use of TDA has been limited by the difficulty of combining the main tool of the subject, the *barcode* or *persistence diagram* with statistics and machine learning. Here we present an alternative approach, using a new summary that we call the *persistence landscape*. The main technical advantage of this descriptor is that it is a function and so we can use the vector space structure of its underlying function space. In fact, this function space is a separable Banach space and we apply the theory of random variables with values in such spaces. Furthermore, since the persistence landscapes are sequences of piecewise-linear functions, calculations with them are much faster than the corresponding calculations with barcodes or persistence diagrams, removing a second serious obstruction to the wider use of topological methods in data analysis.

Notable successes of TDA include the discovery of a subgroup of breast cancers by Nicolau et al. (2011), an understanding of the topology of the space of natural images by Carlsson et al. (2008) and the topology of orthodontic data by Heo et al. (2012), and the detection of genes with a periodic profile by Dequéant et al. (2008). De Silva and Ghrist (2007b,a) used topology to prove coverage in sensor networks.

In the standard paradigm for TDA, one starts with data that one encodes as a finite set of points in $\mathbb{R}^n$ or more generally in some metric space. Then one applies some geometric construction to which one applies tools from algebraic topology. The end result is a topological summary of the data. The standard topological descriptors are the barcode and the persistence diagram (Edelsbrunner et al., 2002; Zomorodian and Carlsson, 2005; Cohen-Steiner et al., 2007), which give a multiscale representation of the *homology* (Hatcher, 2002) of the geometric construction. Roughly, homology in degree 0 describes the connectedness of the data; homology in degree 1 detects holes or tunnels; homology in degree 2 captures voids; and so on. Of particular interest are the homological features that persist as the resolution changes. We will give precise definitions and an illustrative example of this method, called *persistent homology* or *topological persistence*, in Section 2.

Now let us take a statistical view of this paradigm. We consider the data to be sampled from some underlying abstract probability space. Composing the constructions above, we consider our topological summary to be a random variable with values in some summary space $\mathcal{S}$. In detail, the probability space $(\Omega, \mathcal{F}, \mathcal{P})$ consists of a sample space Ω, a σ-algebra $\mathcal{F}$ of events, and a probability measure $\mathcal{P}$. Composing our constructions gives a function $X : (\Omega, \mathcal{F}, \mathcal{P}) \to (\mathcal{S}, \mathcal{A}, \mathcal{P}_*)$, where $\mathcal{S}$ is the summary space, which we assume has some metric, $\mathcal{A}$ is the corresponding Borel σ-algebra, and $\mathcal{P}_*$ is the probability measure on $\mathcal{S}$ obtained by pushing forward $\mathcal{P}$ along X. We assume that X is measurable and thus X is a random variable with values in $\mathcal{S}$.

Here is a list of what we would like to be able to do with our topological summary. Let $X_1, \ldots, X_n$ be a sample of independent random variables with the same distribution as X. We would like to have a good notion of the mean μ of X and the mean $\overline{X}_n$ of the sample; know that $\overline{X}_n$ converges to μ; and be able to calculate $\overline{X}_n(\omega)$, for $\omega \in \Omega$, efficiently. We would like to have information the difference $\overline{X}_n - \mu$, and be able to calculate approximate confidence intervals related to μ. Given two such samples for random variables X and Y with values in our summary space, we would like to be able to test the hypothesis that $\mu_X = \mu_Y$. In order to answer these questions we also need an efficient algorithm for calculating distances between elements of our summary space. In this article, we construct a topological summary that we call the persistence landscape which meets these requirements.

Our basic idea is to convert the barcode into a function in a somewhat additive manner. The are many possible variations of this construction that may result in more suitable summary statistics for certain applications. Hopefully, the theory presented here will also be helpful in those situations.

We remark that while the persistence landscape has a corresponding barcode and persistence diagram, the mean persistence landscape does not. This is analogous to the situation in which an integer-valued random variable having a Poisson distribution has a summary statistic, the rate parameter, that is not an integer.

We also remark that the reader may restrict our Banach space results results to the perhaps more familiar Hilbert space setting. However we will need this generality to prove stability of the persistence landscape for, say, functions on the n-dimensional sphere where $n > 2$.

There has been progress towards combining the persistence diagram and statistics (Mileyko et al., 2011; Turner et al., 2014; Munch et al., 2013; Chazal et al., 2013; Fasy et al., 2014). Blumberg et al. (2014) give a related statistical approach to TDA. Kovacev-Nikolic

et al. (2014) use the persistence landscape defined here to study the maltose binding complex and Chazal et al. (2014) apply the bootstrap to the persistence landscape. The persistence landscape is related to the well group defined by Edelsbrunner et al. (2011).

In Section 2 we provide the necessary background and define the persistence landscape and give some of its properties. In Section 3 we introduce the statistical theory of persistence landscapes, which we apply to a few examples in Section 4. In Section 5 we prove that the persistence landscape is stable and that it provides lower bounds for the previously defined bottleneck and Wasserstein distances.

2. Topological Summaries

The two standard topological summaries of data are the *barcode* and the *persistence diagram*. We will define a new closely-related summary, the *persistence landscape*, and then compare it to these two previous summaries. All of these summaries are derived from the *persistence module*, which we now define.

2.1 Persistence Modules

The main algebraic object of study in topological data analysis is the persistence module. A *persistence module* M consists of a vector space M_a for all $a \in \mathbb{R}$ and linear maps $M(a \leq b) : M_a \to M_b$ for all $a \leq b$ such that $M(a \leq a)$ is the identity map and for all $a \leq b \leq c$, $M(b \leq c) \circ M(a \leq b) = M(a \leq c)$.

There are many ways of constructing a persistence module. One example starts with a set of points $X = \{x_1, \ldots, x_n\}$ in the plane $M = \mathbb{R}^2$ as shown in the top left of Figure 1. To help understand this configuration, we "thicken" each point, by replacing each point, x, with $B_x(r) = \{y \in M \mid d(x, y) \leq r\}$, a disk of fixed radius, r, centered at x. The resulting union, $X_r = \bigcup_{i=1}^n B_r(x_i)$, is shown in Figure 1 for various values of r. For each r, we can calculate $H(X_r)$, the homology of the resulting union of disks. To be precise, $H(-)$ denotes $H_k(-, \mathbb{F})$, the singular homology functor in degree k with coefficients in a field $\mathbb{F}$. So $H(X_r)$ is a vector space that is the quotient of the k-cycles modulo those that are boundaries. As r increases, the union of disks grows, and the resulting inclusions induce maps between the corresponding homology groups. More precisely, if $r \leq s$, the inclusion $\iota_r^s : X_r \hookrightarrow X_s$ induces a map $H(\iota_r^s) : H(X_r) \to H(X_s)$. The images of these maps are the *persistent homology* groups. The collection of vector spaces $H(X_r)$ and linear maps $H(\iota_r^s)$ is a persistence module. Note that this construction works for any set of points in $\mathbb{R}^n$ or more generally in a metric space.

The union of balls X_r has a nice combinatorial description. The *Čech complex*, $\check{C}_r(X)$, of the set of balls $\{B_{x_i}(r)\}$ is the simplicial complex whose vertices are the points $\{x_i\}$ and whose k-simplices correspond to $k+1$ balls with nonempty intersection (see Figure 1). This is also called the *nerve*. It is a basic result that if the ambient space is $\mathbb{R}^n$, X_r is homotopy equivalent to its Čech complex (Borsuk, 1948). So to obtain the singular homology of the union of balls, one can calculate the simplicial homology of the corresponding Čech complex. The Čech complexes $\{\check{C}_r(X)\}$ together with the inclusions $\check{C}_r(X) \subseteq \check{C}_s(X)$ for $r \leq s$ form a filtered simplicial complex. Applying simplicial homology we obtain a persistence module. There exist efficient algorithms for calculating the persistent homology of filtered simplicial complexes (Edelsbrunner et al., 2002; Milosavljević et al., 2011; Chen and Kerber, 2013).

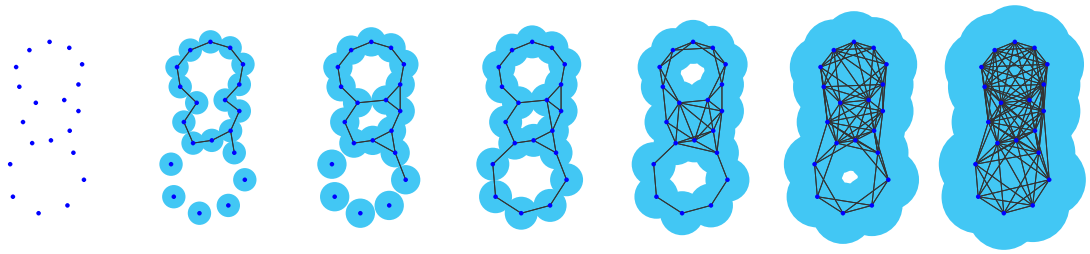

Figure 1: A growing union of balls and the 1-skeleton of the corresponding Čech complex. As the radius grows, features—such as connected components and holes—appear and disappear. Here, the complexes illustrate the births and deaths of three holes, homology classes in degree one. The corresponding birth-death pairs are plotted as part of the top left of Figure 2.

The Čech complex is often computationally expensive, so many variants have been used in computational topology. A larger, but simpler complex called the Rips complex has as vertices the points x_i and has k-simplices corresponding to $k+1$ balls with all pairwise intersections nonempty. Other possibilities include the witness complexes of de Silva and Carlsson (2004), graph induced complexes by Dey et al. (2013) and complexes built using kernel density estimators and triangulations of the ambient space (Bubenik et al., 2010). Some of these are used in the examples in Section 4.

Given any real-valued function $f : S \to \mathbb{R}$ on a topological space S, we can define the associated persistence module, $M(f)$, where $M(f)(a) = H(f^{-1}((\infty, a]))$ and $M(f)(a \leq b)$ is induced by inclusion. Taking f to be the the minimum distance to a finite set of points, X, we obtain the first example.

2.2 Persistence Landscapes

In this section we define a number of functions derived from a persistence module. Examples of each of these are given in Figure 2.

Let M be a persistence module. For $a \leq b$, the corresponding *Betti number* of M, is given by the dimension of the image of the corresponding linear map. That is,

$$\beta^{a,b} = \dim(\mathrm{im}(M(a \leq b))). \tag{1}$$

Lemma 1 *If $a \leq b \leq c \leq d$ then $\beta^{b,c} \geq \beta^{a,d}$.*

Proof Since $M(a \leq d) = M(c \leq d) \circ M(b \leq c) \circ M(a \leq b)$, this follows from (1). ■

Our simplest function, which we call the *rank function* is the function $\lambda : \mathbb{R}^2 \to \mathbb{R}$ given by

$$\lambda(b, d) = \begin{cases} \beta^{b,d} & \text{if } b \leq d \\ 0 & \text{otherwise.} \end{cases}$$

Now let us change coordinates so that the resulting function is supported on the upper half plane. Let

$$m = \frac{b+d}{2}, \quad \text{and} \quad h = \frac{d-b}{2}. \tag{2}$$

The *rescaled rank function* is the function $\lambda : \mathbb{R}^2 \to \mathbb{R}$ given by

$$\lambda(m,h) = \begin{cases} \beta^{m-h,m+h} & \text{if } h \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Much of our theory will apply to these simple functions. However, the following version, which we will call the *persistence landscape*, will have some advantages.

First let us observe that for a fixed $t \in \mathbb{R}$, $\beta^{t-\bullet,t+\bullet}$ is a decreasing function. That is,

Lemma 2 *For* $0 \leq h_1 \leq h_2$,

$$\beta^{t-h_1,t+h_1} \geq \beta^{t-h_2,t+h_2}.$$

Proof Since $t - h_2 \leq t - h_1 \leq t + h_1 \leq t + h_2$, by Lemma 1, $\beta^{t-h_2,t+h_2} \leq \beta^{t-h_1,t+h_1}$. ■

Definition 3 *The* persistence landscape *is a function* $\lambda : \mathbb{N} \times \mathbb{R} \to \overline{\mathbb{R}}$, *where* $\overline{\mathbb{R}}$ *denotes the extended real numbers,* $[-\infty, \infty]$. *Alternatively, it may be thought of as a sequence of functions* $\lambda_k : \mathbb{R} \to \overline{\mathbb{R}}$, *where* $\lambda_k(t) = \lambda(k,t)$. *Define*

$$\lambda_k(t) = \sup(m \geq 0 \mid \beta^{t-m,t+m} \geq k).$$

The persistence landscape has the following properties.

Lemma 4 *1.* $\lambda_k(t) \geq 0$,

2. $\lambda_k(t) \geq \lambda_{k+1}(t)$, *and*

3. λ_k *is 1-Lipschitz.*

The first two properties follow directly from the definition. We prove the third in the appendix.

To help visualize the graph of $\lambda : \mathbb{N} \times \mathbb{R} \to \overline{\mathbb{R}}$, we can extend it to a function $\overline{\lambda} : \mathbb{R}^2 \to \overline{\mathbb{R}}$ by setting

$$\overline{\lambda}(x,t) = \begin{cases} \lambda(\lceil x \rceil, t), & \text{if } x > 0, \\ 0, & \text{if } x \leq 0. \end{cases} \tag{3}$$

We remark that the non-persistent Betti numbers, $\{\dim(M(t))\}$, of a persistence module M can be read off from the diagonal of the rank function, the m-axis of the rescaled rank function, and from the support of the persistence landscape.

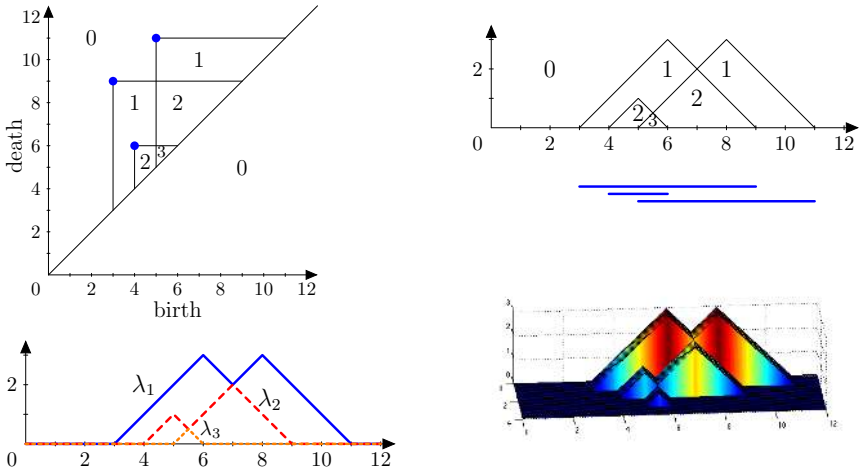

Figure 2: Persistence landscapes for the homology in degree 1 of the example in Figure 1. For the rank function (top left) and rescaled rank function (top right) the values of the functions on the corresponding region are given. The top left graph also contains the three points of the corresponding persistence diagram. Below the top right graph is the corresponding barcode. We also have the corresponding persistence landscape (bottom left) and its 3d-version (bottom right). Notice that λ_1 gives a measure of the dominant homological feature at each point of the filtration.

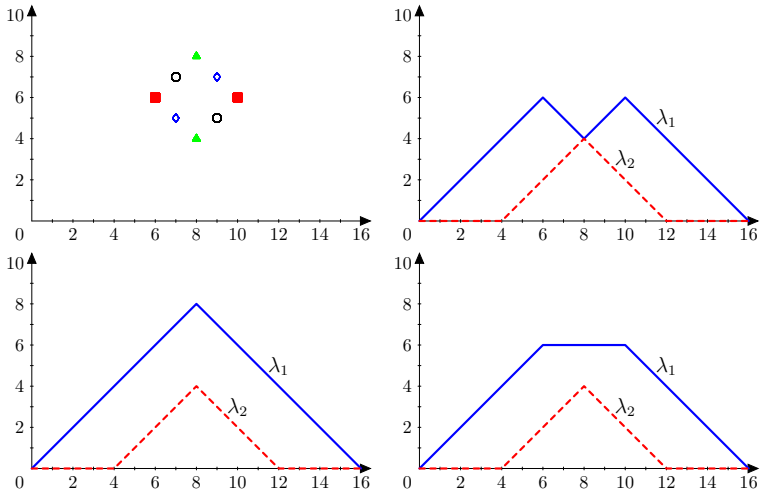

Figure 3: Means of persistence diagrams and persistence landscapes. Top left: the rescaled persistence diagrams $\{(6,6),(10,6)\}$ and $\{(8,4),(8,8)\}$ have two (Fréchet) means: $\{(7,5),(9,7)\}$ and $\{(7,7),(9,5)\}$. In contrast their corresponding persistence landscapes (top right and bottom left) have a unique mean (bottom right).

2.3 Barcodes and Persistence Diagrams

All of the information in a (tame) persistence module is completely contained in a multiset of intervals called a *barcode* (Zomorodian and Carlsson, 2005; Crawley-Boevey, 2012; Chazal et al., 2012). Mapping each interval to its endpoints we obtain the *persistence diagram*.

There exist maps in both directions between these topological summaries and our functions. For an example of corresponding persistence diagrams, barcodes and persistence landscapes, see Figure 2. Informally, the persistence diagram consists of the "upper-left corners" in our rank function. In the other direction, $\lambda(b,d)$ counts the number of points in the persistence diagram in the upper left quadrant of (b,d). Informally, the barcode consists of the "bases of the triangles" in the rescaled rank function, and the other direction is obtained by "stacking isosceles triangles" whose bases are the intervals in the barcode. We invite the reader to make the mappings precise. For example, given a persistence diagram $\{(b_i,d_i)\}_{i=1}^n$,

$$\lambda_k(t) = k\text{th largest value of } \min(t-b_i, d_i-t)_+,$$

where c_+ denotes $\max(c,0)$. The fact that barcodes are a complete invariant of persistence modules is central to these equivalences.

The geometry of the space of persistence diagrams makes it hard to work with. For example, sets of persistence diagrams need not have a unique (Fréchet) mean (Mileyko et al., 2011). In contrast, the space of persistence landscapes is very nice. So a set of persistence landscapes has a unique mean (4). See Figure 3.

Compared to the persistence diagram, the barcode has extra information on whether or not the endpoints of the intervals are included. This finer information is seen in the rank

function and rescaled rank function, but not in the persistence landscape. However when we pass to the corresponding L^p space in Section 2.4, this information disappears.

2.4 Norms for Persistence Landscapes

Recall that for a measure space $(\mathcal{S}, \mathcal{A}, \mu)$, and a function $f : \mathcal{S} \to \mathbb{R}$ defined μ-almost everywhere, for $1 \leq p < \infty$, $\|f\|_p = \left[\int |f|^p d\mu\right]^{\frac{1}{p}}$, and $\|f\|_\infty = \operatorname{ess\,sup} f = \inf\{a \mid \mu\{s \in \mathcal{S} \mid f(s) > a\} = 0\}$. For $1 \leq p \leq \infty$, $\mathcal{L}^p(\mathcal{S}) = \{f : \mathcal{S} \to \mathbb{R} \mid \|f\|_p < \infty\}$ and define $L^p(\mathcal{S}) = \mathcal{L}^p(\mathcal{S})/\sim$, where $f \sim g$ if $\|f - g\|_p = 0$.

On $\mathbb{R}$ and $\mathbb{R}^2$ we will use the Lebesgue measure. On $\mathbb{N} \times \mathbb{R}$, we use the product of the counting measure on $\mathbb{N}$ and the Lebesgue measure on $\mathbb{R}$. For $1 \leq p < \infty$ and $\lambda : \mathbb{N} \times \mathbb{R} \to \overline{\mathbb{R}}$,

$$\|\lambda\|_p^p = \sum_{k=1}^{\infty} \|\lambda_k\|_p^p,$$

where $\lambda_k(t) = \lambda(k, t)$. By Lemma 4(2), $\|\lambda\|_\infty = \|\lambda_1\|_\infty$. If we extend f to $\overline{\lambda} : \mathbb{R}^2 \to \overline{\mathbb{R}}$, as in (3), we have $\|\lambda\|_p = \|\overline{\lambda}\|_p$, for $1 \leq p \leq \infty$.

If λ is any of our functions corresponding to a barcode that is a finite collection of finite intervals, then $\lambda \in \mathcal{L}^p(\mathcal{S})$ for $1 \leq p \leq \infty$, where $\mathcal{S}$ equals $\mathbb{N} \times \mathbb{R}$ or $\mathbb{R}^2$.

Let λ_{bd} and λ_{mh} denote the rank function and the rescaled rank function corresponding to a persistence landscape λ, and let D be the corresponding persistence diagram. Let $\operatorname{pers}_2(D)$ denote the sum of the squares of the lengths of the intervals in the corresponding barcode, and let $\operatorname{pers}_\infty(D)$ be the length of the longest interval.

Proposition 5 *1.* $\|\lambda\|_1 = \|\lambda_{mh}\|_1 = \frac{1}{2}\|\lambda_{bd}\|_1 = \frac{1}{4}\operatorname{pers}_2(D)$, *and*

2. $\|\lambda\|_\infty = \|\lambda_1\|_\infty = \frac{1}{2}\operatorname{pers}_\infty(D)$.

Proof

1. To see that $\|\lambda\|_1 = \|\lambda_{mh}\|$ we remark that both are the volume of the same solid. The change of coordinates implies that $\|\lambda_{mh}\|_1 = \frac{1}{2}\|\lambda_{bd}\|_1$. If $D = \{(b_i, d_i)\}$, then each point (b_i, d_i) contributes h_i^2 to the volume $\|\lambda_{mh}\|_1$, where $h_i = \frac{d_i - b_i}{2}$. So $\|\lambda_{mh}\|_1 = \sum_i h_i^2$. Finally, $\operatorname{pers}_2(D) = \sum_i (2h_i)^2 = 4\sum_i h_i^2$.
2. Lemma 4(2) implies that $\|\lambda\|_\infty = \|\lambda_1\|_\infty$. If $D = \{(b_i, d_i)\}$, then $\|\lambda\|_\infty = \sup_i \frac{d_i - b_i}{2}$.

■

We remark that the quantities in 1 and 2 also equal $W_2(D, \emptyset)^2$ and $W_\infty(D, \emptyset)$ respectively (see Section 5 for the corresponding definitions).

3. Statistics with Landscapes

Now let us take a probabilistic viewpoint. First, we assume that our persistence landscapes lie in $L^p(\mathcal{S})$ for some $1 \leq p < \infty$, where $\mathcal{S}$ equals $\mathbb{N} \times \mathbb{R}$ or $\mathbb{R}^2$. In this case, $L^p(\mathcal{S})$ is a separable Banach space. When $p = 2$ we have a Hilbert space; however, we will not use this structure. In some examples, the persistence landscapes will only be stable for some $p > 2$ (see Theorem 16).

3.1 Landscapes as Banach Space Valued Random Variables

Let X be a random variable on some underlying probability space $(\Omega, \mathcal{F}, P)$, with corresponding persistence landscape Λ, a Borel random variable with values in the separable Banach space $L^p(\mathcal{S})$. That is, for $\omega \in \Omega$, $X(\omega)$ is the data and $\Lambda(\omega) = \lambda(X(\omega)) =: \lambda$ is the corresponding topological summary statistic.

Now let $X_1, \ldots, X_n$ be independent and identically distributed copies of X, and let $\Lambda^1, \ldots, \Lambda^n$ be the corresponding persistence landscapes. Using the vector space structure of $L^p(\mathcal{S})$, the *mean landscape* $\overline{\Lambda}^n$ is given by the pointwise mean. That is, $\overline{\Lambda}^n(\omega) = \overline{\lambda}^n$, where

$$\overline{\lambda}^n(k,t) = \frac{1}{n}\sum_{i=1}^{n} \lambda^i(k,t). \tag{4}$$

Let us interpret the mean landscape. If $B_1, \ldots, B_n$ are the barcodes corresponding to the persistence landscapes $\lambda^1, \ldots, \lambda^n$, then for $k \in \mathbb{N}$ and $t \in \mathbb{R}$, $\overline{\lambda}^n(k,t)$ is the average value of the largest radius interval centered at t that is contained in k intervals in the barcodes $B_1, \ldots, B_n$.

For those used to working with persistence diagrams, it is tempting to try to find a persistence diagram whose persistence landscape is closest to a given mean landscape. While this is an interesting mathematical question, we would like to suggest that the more important practical issue is using the mean landscape to understand the data.

We would like to be able to say that the mean landscape converges to the expected persistence landscape. To say this precisely we need some notions from probability in Banach spaces.

3.2 Probability in Banach Spaces

Here we present some results from probability in Banach spaces. For a more detailed exposition we refer the reader to Ledoux and Talagrand (2011).

Let $\mathcal{B}$ be a real separable Banach space with norm $\|\cdot\|$. Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $V : (\Omega, \mathcal{F}, P) \to \mathcal{B}$ be a Borel random variable with values in $\mathcal{B}$. The composite $\|V\| : \Omega \xrightarrow{V} \mathcal{B} \xrightarrow{\|\cdot\|} \mathbb{R}$ is a real-valued random variable. Let $\mathcal{B}^*$ denote the topological dual space of continuous linear real-valued functions on $\mathcal{B}$. For $f \in \mathcal{B}^*$, the composite $f(V) : \Omega \xrightarrow{V} \mathcal{B} \xrightarrow{f} \mathbb{R}$ is a real-valued random variable.

For a real-valued random variable $Y : (\Omega, \mathcal{F}, P) \to \mathbb{R}$, the *mean* or *expected value*, is given by $E(Y) = \int Y \, dP = \int_\Omega Y(\omega) \, dP(\omega)$. We call an element $E(V) \in \mathcal{B}$ the *Pettis integral* of V if $E(f(V)) = f(E(V))$ for all $f \in \mathcal{B}^*$.

Proposition 6 *If $E\|V\| < \infty$, then V has a Pettis integral and $\|E(V)\| \leq E\|V\|$.*

Now let $(V_n)_{n\in\mathbb{N}}$ be a sequence of independent copies of V. For each $n \geq 1$, let $S_n = V_1 + \cdots + V_n$. For a sequence (Y_n) of $\mathcal{B}$-valued random variables, we say that (Y_n) *converges almost surely* to a $\mathcal{B}$-valued random variable Y, if $P(\lim_{n\to\infty} Y_n = Y) = 1$.

Theorem 7 (Strong Law of Large Numbers) *$(\frac{1}{n}S_n) \to E(V)$ almost surely if and only if $E\|V\| < \infty$.*

For a sequence (Y_n) of $\mathcal{B}$-valued random variables, we say that (Y_n) *converges weakly* to a $\mathcal{B}$-valued random variable Y, if $\lim_{n\to\infty} E(\varphi(Y_n)) = E(\varphi(Y))$ for all bounded continuous functions $\varphi : \mathcal{B} \to \mathbb{R}$. A random variable G with values in $\mathcal{B}$ is said to be *Gaussian* if for each $f \in \mathcal{B}^*$, $f(G)$ is a real valued Gaussian random variable with mean zero. The *covariance structure* of a $\mathcal{B}$-valued random variable, V, is given by the expectations $E[(f(V) - E(f(V)))(g(V) - E(g(V)))]$, where $f, g \in B^*$. A Gaussian random variable is determined by its covariance structure. From Hoffmann-Jørgensen and Pisier (1976) we have the following.

Theorem 8 (Central Limit Theorem) *Assume that $\mathcal{B}$ has type 2. (For example $\mathcal{B} = L^p(\mathcal{S})$, with $2 \leq p < \infty$.) If $E(V) = 0$ and $E(\|V\|^2) < \infty$ then $\frac{1}{\sqrt{n}} S_n$ converges weakly to a Gaussian random variable $G(V)$ with the same covariance structure as V.*

3.3 Convergence of Persistence Landscapes

Now we will apply the results of the previous section to persistence landscapes.

Theorem 7 directly implies the following.

Theorem 9 (Strong Law of Large Numbers for persistence landscapes) *$\overline{\Lambda}^n \to E(\Lambda)$ almost surely if and only if $E\|\Lambda\| < \infty$.*

Theorem 10 (Central Limit Theorem for peristence landscapes) *Assume $p \geq 2$. If $E\|\Lambda\| < \infty$ and $E(\|\Lambda\|^2) < \infty$ then $\sqrt{n}[\overline{\Lambda}^n - E(\Lambda)]$ converges weakly to a Gaussian random variable with the same covariance structure as Λ.*

Proof Apply Theorem 8 to $V = \lambda(X) - E(\lambda(X))$. ■

Next we apply a functional to the persistence landscapes to obtain a real-valued random variable that satisfies the usual central limit theorem.

Corollary 11 *Assume $p \geq 2$, $E\|\Lambda\| < \infty$ and $E(\|\Lambda\|^2) < \infty$. For any $f \in L^q(\mathcal{S})$ with $\frac{1}{p} + \frac{1}{q} = 1$, let*

$$Y = \int_{\mathcal{S}} f\Lambda = \|f\Lambda\|_1. \tag{5}$$

Then

$$\sqrt{n}[\overline{Y}_n - E(Y)] \xrightarrow{d} N(0, \mathrm{Var}(Y)). \tag{6}$$

where d denotes convergence in distribution and $N(\mu, \sigma^2)$ is the normal distribution with mean μ and variance σ^2.

Proof Since $V = \Lambda - E(\Lambda)$ satisfies the central limit theorem in $L^p(\mathcal{S})$, for any $g \in L^p(\mathcal{S})^*$, the real random variable $g(V)$ satisfies the central limit theorem in $\mathbb{R}$ with limiting Gaussian law with mean 0 and variance $E(g(V)^2)$. If we take $g(h) = \int_{\mathcal{S}} fh$, where $f \in L^q(\mathcal{S})$, with $\frac{1}{p} + \frac{1}{q} = 1$, then $g(V) = Y - E(Y)$ and $E(g(V)^2) = \mathrm{Var}(Y)$. ■

3.4 Confidence Intervals

The results of Section 3.3 allow us to obtain approximate confidence intervals for the expected values of functionals on persistence landscapes.

Assume that $\lambda(X)$ satisfies the conditions of Corollary 11 and that Y is a corresponding real random variable as defined in (5). By Corollary 11 and Slutsky's theorem we may use the normal distribution to obtain the approximate $(1-\alpha)$ confidence interval for $E(Y)$ using

$$\overline{Y}_n \pm z^* \frac{S_n}{\sqrt{n}},$$

where $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (Y_i - \overline{Y}_n)^2$, and z^* is the upper $\frac{\alpha}{2}$ critical value for the normal distribution.

3.5 Statistical Inference using Landscapes I

Here we apply the results of Section 3.3 to hypothesis testing using persistence landscapes.

Let $X_1, \ldots, X_n$ be an iid copies of the random variable X and let $X'_1, \ldots, X'_{n'}$ be an iid copies of the random variable X'. Assume that the corresponding persistence landscapes Λ, Λ' lie in $L^p(\mathcal{S})$, where $p \geq 2$. Let $f \in L^q(\mathcal{S})$, where $\frac{1}{p} + \frac{1}{q} = 1$. Let Y and Y' be defined as in (5). Let $\mu = E(Y)$ and $\mu' = E(Y')$. We will test the null hypothesis that $\mu = \mu'$. First we recall that the sample mean $\overline{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$ is an unbiased estimator of μ and the sample variance $s_Y^2 = \frac{1}{n-1}\sum_{i=1}^n (Y_i - \overline{Y})^2$ is an unbiased estimator of $\mathrm{Var}(Y)$ and similarly for $\overline{Y'}$ and $s_{Y'}^2$. By Corollary 11, Y and Y' are asymptotically normal.

We use the two-sample z-test. Let

$$z = \frac{\overline{Y} - \overline{Y'}}{\sqrt{\frac{S_Y^2}{n} + \frac{S_{Y'}^2}{n'}}},$$

where the denominator is the standard error for the difference. From this standard score a p-value may be obtained from the normal distribution.

3.6 Choosing a Functional

To apply the above results, one needs to choose a functional, $f \in L^q(\mathcal{S})$. This choice will need to be made with an understanding of the data at hand. Here we present a couple of options.

If each $\lambda = \Lambda(\omega)$ is supported by $\{1, \ldots, K\} \times [-B, B]$, take

$$f(k,t) = \begin{cases} 1 & \text{if } t \in [-B,B] \text{ and } k \leq K \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Then $\|f\Lambda\|_1 = \|\Lambda\|_1$.

If the parameter values for which the persistence landscape is nonzero are bounded by $\pm B$, then we have a nice choice of functional for the persistence landscape that is unavailable for the (rescaled) rank function. We can choose a functional that is sensitive of the first K dominant homological features. That is, using f in (7), $\|f\lambda\|_1 = \sum_{k=1}^K \|\Lambda_k\|_1$.

Under this weaker assumption we can also take $f_k(t) = \frac{1}{k^r}\chi_{[-B,B]}$, where $r > 1$. Then $\|f\Lambda\|_1 = \sum_{k=1}^{\infty} \frac{1}{k^r}\|\Lambda_k(t)\|_1$.

The condition that λ is supported by $\mathbb{N}\times[-B,B]$ can often be enforced by using reduced homology or by applying extended persistence (Cohen-Steiner et al., 2009; Bubenik and Scott, 2014) or by simply truncating the intervals in the corresponding barcode at some fixed values. We remark that certain experimental data may have bounds on the number of intervals. For example, in the protein data considered using the ideas presented here in Kovacev-Nikolic et al. (2014), the simplicial complexes have a fixed number of vertices.

3.7 Statistical Inference using Landscapes II

The functionals suggested in Section 3.6 in the hypothesis test given in Section 3.5 may not have enough power to discriminate between two groups with different persistence in some examples.

To increase the power, one can apply a vector of functionals and then apply Hotelling's T^2 test. For example, consider $Y = (\int(\Lambda_1 - \Lambda_1'), \ldots, \int(\Lambda_K - \Lambda_K'))$, where $K \ll n_1 + n_2 - 2$.

This alternative will not be sufficient if the persistence landscapes are translates of each other, (see Figure 7). An additional approach is to compute the distance between the mean landscapes of the two groups and obtain a p-value using a permutation test. This is done in the Section 4.3. This test has been applied to persistence diagrams and barcodes (Chung et al., 2009; Robinson and Turner, 2013).

4. Examples

The persistent homologies in this section were calculated using javaPlex (Tausz et al., 2011) and Perseus by Nanda (2013). Another publicly available alternative is Dionysus by Morozov (2012). In Section 4.2 we use Matlab code courtesy of Eliran Subag that implements an algorithm from Wood and Chan (1994).

4.1 Linked Annuli

We start with a simple example to illustrate the techniques. Following Munch et al. (2013), we sample 200 points from the uniform distribution on the union of two annuli. We then calculate the corresponding persistence landscape in degree one using the Vietoris-Rips complex. We repeat this 100 times and calculate the mean persistence landscape. See Figure 4.

Note that in the degree one barcode of this example, it is very likely that there will be one large interval, one smaller interval born at around the same time, and all other intervals are smaller and die around the time the larger two intervals are born.

4.2 Gaussian Random Fields

The topology of Gaussian random fields is of interest in statistics. The Euler characteristic of superlevel sets of a Gaussian random field may be calculated using the Gaussian Kinematic Formula of Adler and Taylor (2007). The persistent homology of Gaussian random fields

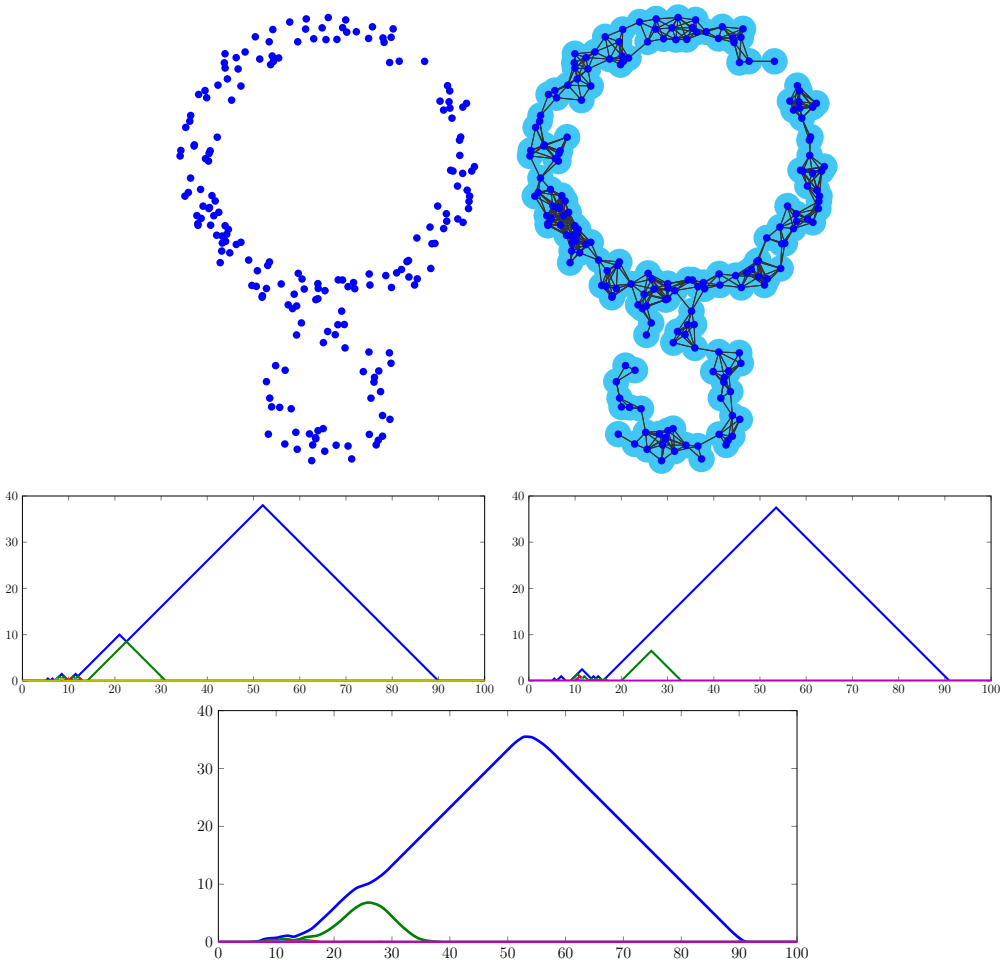

Figure 4: 200 points were sampled from a pair of linked annuli. Here we show the points and a corresponding union of balls and 1-skeleton of the Čech complex. This was repeated 100 times. Next we show two of the degree one persistence landscapes and the mean degree one persistence landscape.

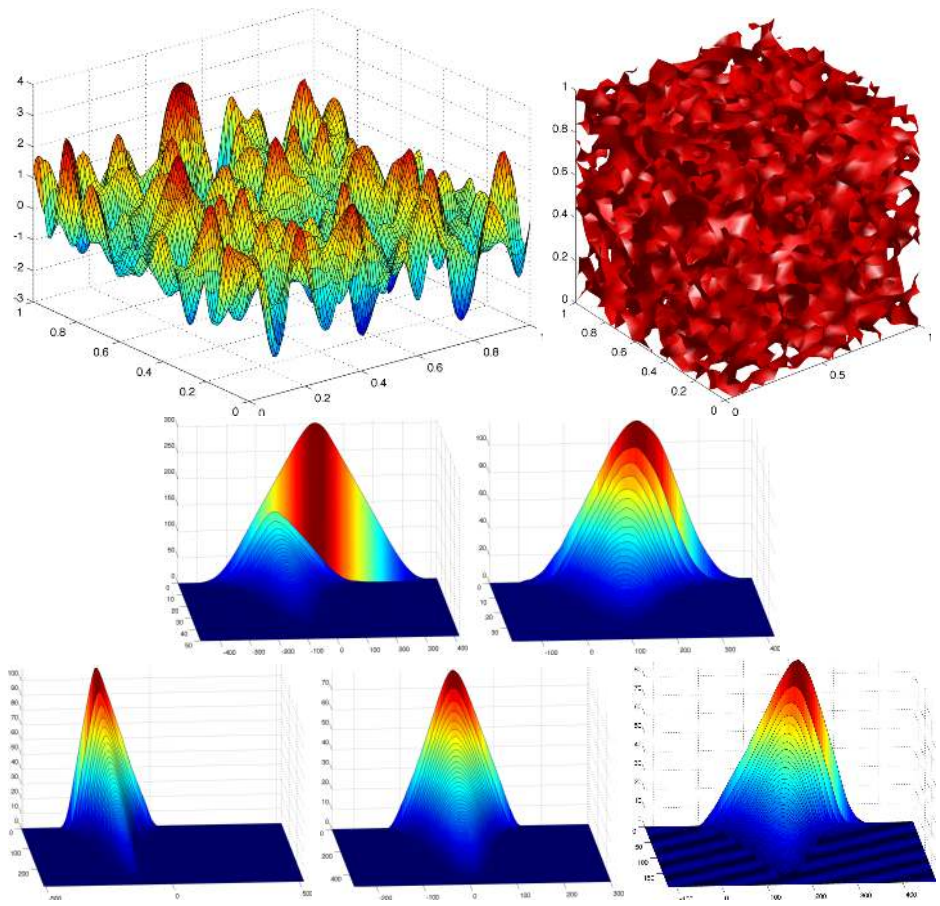

Figure 5: Mean landscapes of Gaussian random fields. The graph of a Gaussian random field on $[0,1]^2$ (top left) and its corresponding mean landscapes (middle row) in degrees 0 and 1. The 0-isosurface of a Gaussian random field on $[0,1]^3$ (top right) and the corresponding mean landscapes in degrees 0, 1 and 2 (bottom row).

has been considered by Adler et al. (2010) and its expected Euler characteristic has been obtained by Bobrowski and Borman (2012).

Here we consider a stationary Gaussian random field on $[0,1]^2$ with autocovariance function $\gamma(x,y) = e^{-400(x^2+y^2)}$. See Figure 5. We sample this field on a 100 by 100 grid, and calculate the persistence landscape of the sublevel set. For homology in degree 0, we truncate the infinite interval at the maximum value of the field. We calculate the mean persistence landscapes in degrees 0 and 1 from 100 samples (see Figure 5, where we have rescaled the filtration by a factor of 100).

In the Gaussian random field literature, it is more common to consider superlevel sets. However, by symmetry, the expected persistence landscape in this case is the same except for a change in the sign of the filtration.

We repeat this calculation for a similar Gaussian random field on $[0,1]^3$, this time using reduced homology. See Figure 5. This time we sample on a $25 \times 25 \times 25$ grid.

4.3 Torus and Sphere

Here we combine persistence landscapes and statistical inference to discriminate between iid samples of 1000 points from a torus and a sphere in $\mathbb{R}^3$ with the same surface area, using the uniform surface area measure as described by Diaconis et al. (2012) (see Figure 6). To be precise, we use the torus given by $(r-2)^2 + z^2 = 1$ in cylindrical coordinates, and the sphere given by $r^2 = 2\pi$ in spherical coordinates.

For these points, we construct a filtered simplicial complex as follows. First we triangulate the underlying space using the Coxeter–Freudenthal–Kuhn triangulation, starting with a cubical grid with sides of length $\frac{1}{2}$. Next we smooth our data using a triangular kernel with bandwidth 0.9. We evaluate this kernel density estimator at the vertices of our simplicial complex. Finally, we filter our simplicial complex as follows. For filtration level $-r$, we include a simplex in our triangulation if and only if the kernel density estimator has values greater than or equal to r at all of its vertices. Three stages in the filtration for one of the samples are shown in (see Figure 6). We then calculate the persistence landscape of this filtered simplicial complex for 100 samples and plot the mean landscapes (see Figure 6). We observe that the large peaks correspond to the Betti numbers of the torus and sphere.

Since the support of the persistence landscapes is bounded, we can use the integral of the landscapes to obtain a real valued random variable that satisfies (6). We use a two-sample z-test to test the null hypothesis that these random variables have equal mean. For the landscapes in dimensions 0 and 2 we cannot reject the null hypothesis. In dimension 1 we do reject the null hypothesis with a p-value of 3×10^{-6}.

We can also choose a functional that only integrates the persistence landscape $\lambda(k,t)$ for certain ranges of k. In dimension 1, with $k = 1$ or $k = 2$ there is a statistically significant difference (p-values of 10^{-8} and 3×10^{-6}), but not for $k > 2$. In dimension 2, there is not a significant difference for $k = 1$, but there is a significant difference for $k > 1$ (p-value $< 10^{-4}$).

Now we increase the difficulty by adding a fair amount of Gaussian noise to the point samples (see Figure 7) and using only 10 samples for each surface. This time we calculate the L^2 distances between the mean landscapes. We use the permutation test with 10,000 repetitions to determine if this distance is statistically significant. There is a significant difference in dimension 0, with a p value of 0.0111. This is surprising, since the mean landscapes look very similar. However, on closer inspection, they are shifted slightly (see Figure 7). Note that we are detecting a geometric difference, not a topological one. This shows that this statistic is quite powerful. There is also a significant difference in dimensions 1 and 2, with p values of 0.0000 and 0.0000, respectively.

5. Landscape Distance and Stability

In this section we define the landscape distance and use it to show that the persistence landscape is a stable summary statistic. We also show that the landscape distance gives lower bounds for the bottleneck and Wasserstein distances. We defer the proofs of the results of this section to the appendix.

Let M and M' be persistence modules as defined in Section 2.1 and let λ and λ' be their corresponding persistence landscapes as defined in Section 2.2. For $1 \leq p \leq \infty$, define the

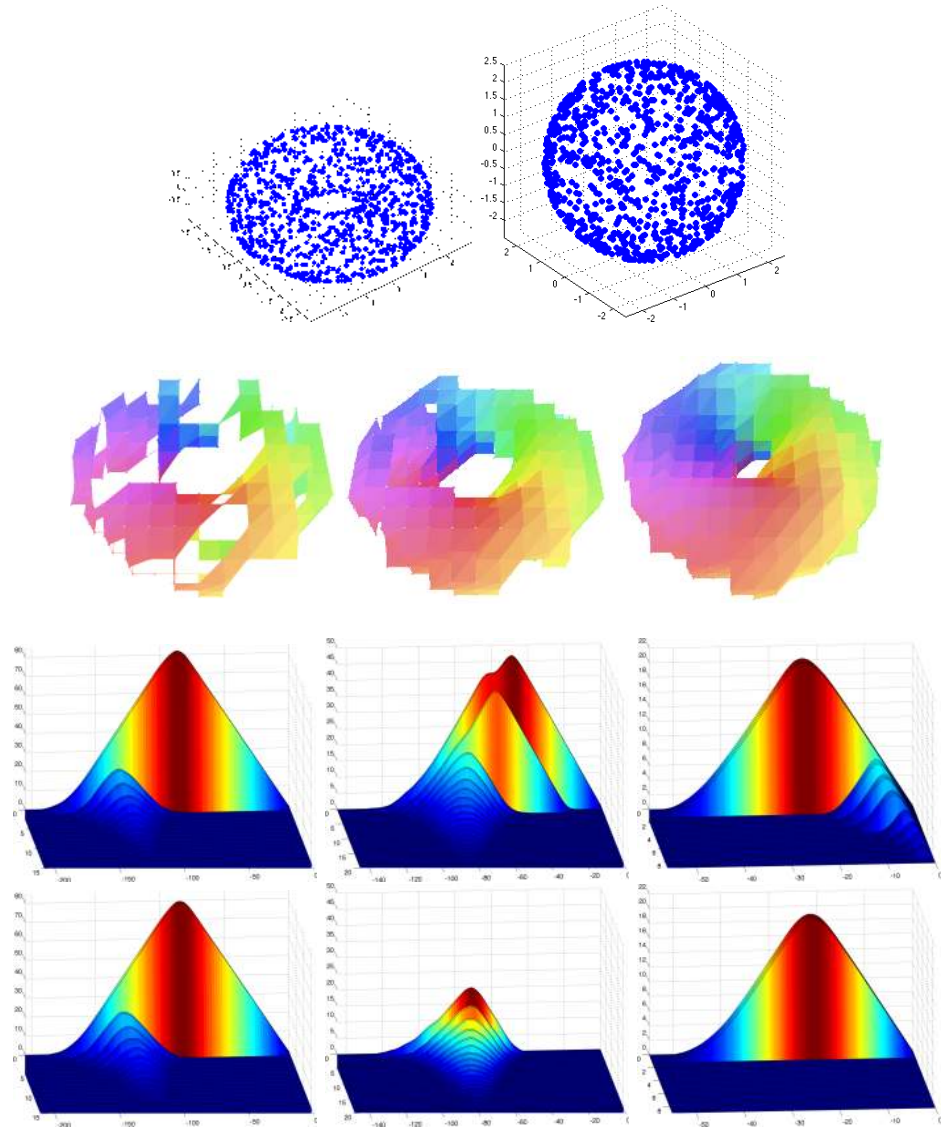

Figure 6: We sample 1000 points for a torus and sphere, 100 times each, construct the corresponding filtered simplicial complexes and calculate persistent homology. In columns 1, 2 and 3, we have the mean persistence landscape in dimension 0, 1 and 2 of the torus in row 3 and the sphere in row 4.

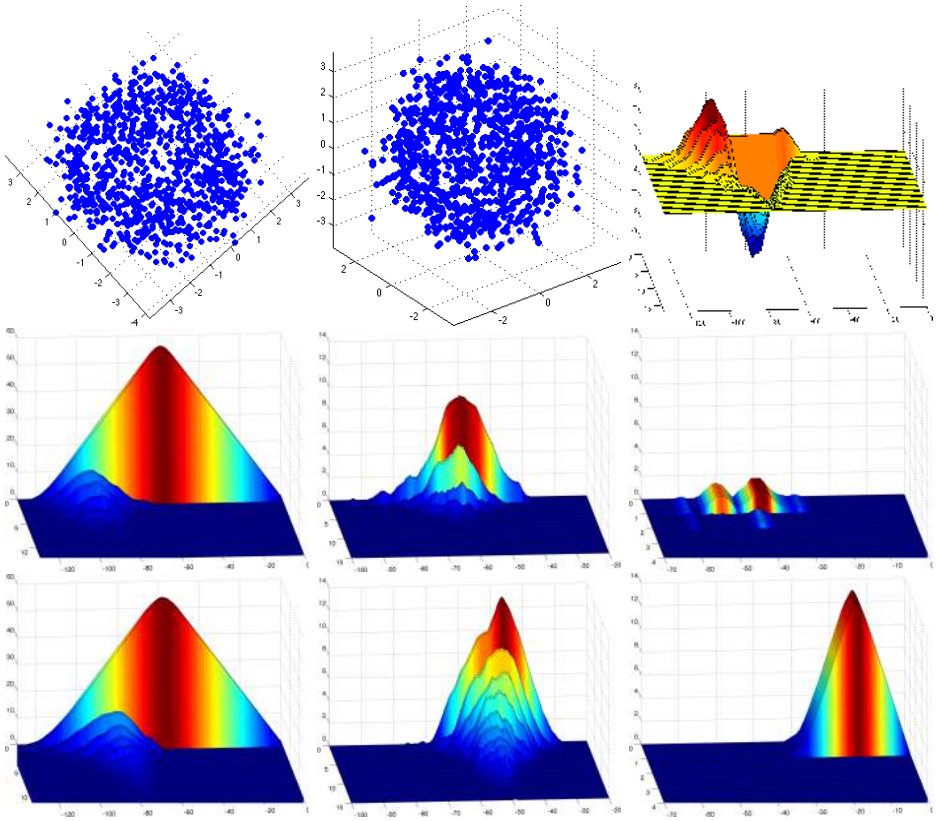

Figure 7: We again sample 1000 points sampled from a torus (top left) and sphere (top middle), this time with Gaussian noise. We show the torus from the perspective that makes it easiest to see the hole in the middle. We calculate persistent homology from 10 samples. In columns 1, 2 and 3, we have the mean persistence landscape in dimension 0, 1 and 2, respectively, with the torus in row 2 and the sphere in row 3. The top right is a graph of the difference between the mean landscapes in dimension 0.

p-landscape distance between M and M' by

$$\Lambda_p(M, M') = \|\lambda - \lambda'\|_p.$$

Similarly, if λ and λ' are the persistence landscapes corresponding to persistence diagrams D and D' (Section 2.3), then we define

$$\Lambda_p(D, D') = \|\lambda - \lambda\|_p.$$

Given a real valued function $f : X \to \mathbb{R}$ on a topological space X, let $M(f)$ denote be the corresponding persistence module defined at the end of Section 2.1.

Theorem 12 (∞-Landscape Stability Theorem) *Let $f, g : X \to \mathbb{R}$. Then*

$$\Lambda_\infty(M(f), M(g)) \leq \|f - g\|_\infty.$$

Thus the persistence landscape is stable with respect to the supremum norm. We remark that there are no assumptions on f and g, not even the q-tame condition of Chazal et al. (2012).

Let D be a persistence diagram. For $x = (b, d) \in D$, let $\ell = d - b$ denote the *persistence* of x. If $D = \{x_j\}$, let $\mathrm{Pers}_k(D) = \sum_j \ell_j^k$ denote the *degree-k total persistence* of D.

Now let us consider a persistence diagram to be an equivalence class of multisets of pairs (b, d) with $b \leq d$, where $D \sim D \amalg \{(t, t)\}$ for any $t \in \mathbb{R}$. That is, to any persistence diagram, we can freely adjoin points on the diagonal. This is reasonable, since points on the diagonal have zero persistence. Each persistence diagram has a unique representative $\hat{D}$ without any points on the diagonal. We set $|D| = |\hat{D}|$. We also remark that $\mathrm{Pers}_k(D)$ is well defined.

By allowing ourselves to add as many points on the diagonal as necessary, there exists bijections between any two persistence diagrams. Any bijection $\varphi : D \xrightarrow{\cong} D'$ can be represented by $\varphi : x_j \mapsto x'_j$, where $j \in J$ with $|J| = |D| + |D'|$. For a given φ, let $x_j = (b_j, d_j)$, $x'_j = (b'_j, d'_j)$ and $\varepsilon_j = \|x_j - x'_j\|_\infty = \max(|b_j - b'_j|, |d_j - d'_j|)$.

The *bottleneck distance* (Cohen-Steiner et al., 2007) between persistence diagrams D and D' is given by

$$W_\infty(D, D') = \inf_{\varphi : D \xrightarrow{\cong} D'} \sup_j \varepsilon_j,$$

where the infimum is taken over all bijections from D to D'. It follows that for the empty persistence diagram $\emptyset$, $W_\infty(D, \emptyset) = \frac{1}{2} \sup_j \ell_j$.

The ∞-landscape distance is bounded by the bottleneck distance.

Theorem 13 *For persistence diagrams D and D',*

$$\Lambda_\infty(D, D') \leq W_\infty(D, D').$$

For $p \geq 1$, the *p-Wasserstein distance* (Cohen-Steiner et al., 2010) between D and D' is given by

$$W_p(D, D') = \inf_{\varphi : D \xrightarrow{\cong} D'} \left[\sum_j \varepsilon_j^p \right]^{\frac{1}{p}}.$$

We remark that the Wasserstein distance gives equal weighting to the ε_j while the landscape distance gives a stronger weighting to ε_j if x_j has larger persistence. The landscape distance is most closely related to a weighted version of the Wasserstein distance that we now define. The *persistence weighted p-Wasserstein distance* between D and D' is given by

$$\overline{W}_p(D, D') = \inf_{\varphi: D \xrightarrow{\cong} D'} \left[\sum_j \ell_j \varepsilon_j^p \right]^{\frac{1}{p}}.$$

Note that it is asymmetric.

For the remainder of the section we assume that D and D' are finite. The following result bounds the p-landscape distance. Recall that ℓ_j is the persistence of $x_j \in D$ and when $\varphi : x_j \mapsto x'_j$, $\varepsilon_j = \|x_j - x'_j\|_\infty$

Theorem 14 *If $n = |D| + |D|$ then*

$$\Lambda_p(D, D')^p \leq \min_{\varphi: D \xrightarrow{\cong} D'} \left[\sum_{j=1}^{n} \ell_j \varepsilon_j^p + \frac{2}{p+1} \sum_{j=1}^{n} \varepsilon_j^{p+1} \right].$$

From this we can obtain a lower bound on the p-Wasserstein distance.

Corollary 15 $W_p(D, D')^p \geq \min\left(1, \frac{1}{2}\left[W_\infty(D, \emptyset) + \frac{1}{p+1}\right]^{-1} \Lambda_p(D, D')^p\right)$.

For our final stability theorem, we use ideas from Cohen-Steiner et al. (2010). Let $f : X \to \mathbb{R}$ be a function on a topological space. We say that f is *tame* if for all but finitely many $a \in \mathbb{R}$, the associated persistence module $M(f)$ is constant and finite dimensional on some open interval containing a. For such an f, let $D(f)$ denote the corresponding persistence diagram. If X is a metric space we say that f is *Lipschitz* if there is some constant c such that $|f(x) - f(y)| \leq c\, d(x, y)$ for all $x, y \in X$. We let $\mathrm{Lip}(f)$ denote the infimum of all such c. We say that a metric space X *implies bounded degree-k total persistence* if there is a constant $C_{X,k}$ such that $\mathrm{Pers}_k(D(f)) \leq C_{X,k}$ for all tame Lipschitz functions $f : X \to \mathbb{R}$ such that $\mathrm{Lip}(f) \leq 1$. For example, as observed by Cohen-Steiner et al. (2010), if X is the n-dimensional sphere, then $X = S^n$ has bounded k-persistence for $k = n + \delta$ for any $\delta > 0$, but does not have bounded k-persistence for $k < n$.

Theorem 16 (p-Landscape stability theorem) *Let X be a triangulable, compact metric space that implies bounded degree-k total persistence for some real number $k \geq 1$, and let f and g be two tame Lipschitz functions. Then*

$$\Lambda_p(D(f), D(g))^p \leq C \|f - g\|_\infty^{p-k},$$

for all $p \geq k$, where $C = C_{X,k}\|f\|_\infty(\mathrm{Lip}(f)^k + \mathrm{Lip}(g)^k) + C_{X,k+1}\frac{1}{p+1}(\mathrm{Lip}(f)^{k+1} + \mathrm{Lip}(g)^{k+1})$.

Thus the persistence diagram is stable with respect to the p-landscape distance if $p > k$, where X has bounded degree-k total persistence. This is the same condition as for the stability of the p-Wasserstein distance in Cohen-Steiner et al. (2010). Equivalently, the

persistence landscape is stable with respect to the p-norm if $p > k$, where X has bounded degree-k total persistence.

Acknowledgments

The author would like to thank Robert Adler, Frederic Chazal, Herbert Edelsbrunner, Giseon Heo, Sayan Mukherjee and Stephen Rush for helpful discussions. Thanks to Junyong Park for suggesting Hotelling's T^2 test. Also thanks to the anonymous referees who made a number of helpful comments to improve the exposition. In addition, the author gratefully acknowledges the support of the Air Force Office of Scientific Research (AFOSR grant FA9550-13-1-0115).

Appendix A. Proofs

Proof [Proof of Lemma 4(3)] We will prove that λ_k is 1-Lipschitz. That is, $|\lambda_k(t)-\lambda_k(s)| \leq |t-s|$, for all $s, t \in \mathbb{R}$.

Let $s, t \in \mathbb{R}$. Without loss of generality, assume that $\lambda_k(t) \geq \lambda_k(s) \geq 0$. If $\lambda_k(t) \leq |t-s|$, then $\lambda_k(t) - \lambda_k(s) \leq \lambda_k(t) \leq |t-s|$ and we are done. So assume that $\lambda_k(t) > |t-s|$.

Let $0 < h < \lambda_k(t) - |t-s|$. Then $t - \lambda_k(t) < s - h < s + h < t + \lambda_k(t)$. Thus, by Lemma 1 and Definition 3, $\beta^{s-h,s+h} \geq k$. It follows that $\lambda_k(s) \geq \lambda_k(t) - |t-s|$. Thus $\lambda_k(t) - \lambda_k(s) \leq |t-s|$. ■

Theorems 12 and 13 follow from the next result which is of independent interest. Following Chazal et al. (2009), we say that two persistence modules M and M' are *ε-interleaved* if for all $a \in \mathbb{R}$ there exist linear maps $\varphi_a : M_a \to M'_{a+\varepsilon}$ and $\psi : M'_a \to M_{a+\varepsilon}$ such that for all $a \in \mathbb{R}$, $\psi_{a+\varepsilon} \circ \varphi_a = M(a \leq a + 2\varepsilon)$ and $\varphi_{a+\varepsilon} \circ \psi_a = M'(a \leq a + 2\varepsilon)$ and for all $a \leq b$ $M'(a + \varepsilon \leq b + \varepsilon) \circ \varphi_a = \varphi_b \circ M(a \leq b)$ and $M(a + \varepsilon \leq b + \varepsilon) \circ \psi_a = \psi_b \circ M'(a \leq b)$. For persistence modules M and M' define the *interleaving distance* between M and M' by

$$d_I(M, M) = \inf(\varepsilon \mid M \text{ and } M' \text{ are } \varepsilon\text{-interleaved}).$$

Theorem 17 $\Lambda_\infty(M, M') \leq d_I(M, M')$.

Proof Assume that M and M' are ε-interleaved. Then for $t \in \mathbb{R}$ and $m \geq \varepsilon$, the map $M(t - m \leq t + m)$ factors through the map $M'(t - m + \varepsilon \leq t + m - \varepsilon)$. So by Lemma 1, $\beta^{t-m+\varepsilon,t+m-\varepsilon}(M') \geq \beta^{t-m,t+m}(M)$. Thus by Definition 3, $\lambda'(k, t) \geq \lambda(k, t) - \varepsilon$ for all $k \geq 1$. It follows that $\|\lambda - \lambda'\|_\infty \leq \varepsilon$. ■

Proof [Proof of Theorem 12] Combining Theorem 17 with the stability theorem of Bubenik and Scott (2014), we have $\Lambda_\infty(M(f), M(g)) \leq d_I(M(f), M(g)) \leq \|f - g\|_\infty$. ■

Proof [Proof of Theorem 13] For a persistence diagram D, consider the persistence module given by the corresponding sum of interval modules (Chazal et al., 2012), $M(D) = \oplus_{(a,b)\in\hat{D}} \mathbb{I}(a, b)$. Combining Theorem 17 with Theorem 4.9 of Chazal et al. (2012) we have

$\Lambda_\infty(M(D), M(D')) \leq d_I(M(D), M(D')) \leq W_\infty(D, D')$. ■

Proof [Proof of Theorem 14] Let $\varphi : D \xrightarrow{\cong} D'$ with $\varphi(x_j) = x'_j$. Let $\lambda = \lambda(D)$ and $\lambda' = \lambda(D')$. So $\Lambda_p(D, D')^p = \|\lambda - \lambda'\|_p^p$.

$$\begin{aligned}\|\lambda - \lambda'\|_p^p &= \int |\lambda(k,t) - \lambda'(k,t)|^p \\ &= \sum_{k=1}^n \int |\lambda_k(t) - \lambda'_k(t)|^p \, dt \\ &= \int \sum_{k=1}^n |\lambda_k(t) - \lambda'_k(t)|^p \, dt\end{aligned}$$

Fix t. Let $u_j(t) = \lambda(\{x_j\})(1,t)$ and $v_j(t) = \lambda(\{x'_j\})(1,t)$. For each t, let $u_{(1)}(t) \leq \cdots \leq u_{(n)}(t)$ denote an ordering of $u_1(t), \ldots, u_n(t)$ and define $v_{(k)}(t)$ for $1 \leq k \leq n$ similarly. Then $u_{(k)}(t) = \lambda_k(t)$ and $v_{(k)}(t) = \lambda'_k(t)$ (see Figure 2). We obtain the result from the following where the two inequalities are proven in Lemmata 18 and 19.

$$\begin{aligned}\|\lambda - \lambda'\|_p^p &= \int \sum_{k=1}^n |u_{(k)}(t) - v_{(k)}(t)|^p \, dt \\ &\leq \int \sum_{k=1}^n |u_k(t) - v_k(t)|^p \, dt \\ &= \sum_{j=1}^n \int |u_j(t) - v_j(t)|^p \, dt \\ &\leq \sum_{j=1}^n \ell_j \varepsilon_j^p + \frac{2}{p+1} \sum_{j=1}^n \varepsilon_j^{p+1}.\end{aligned}$$

■

Lemma 18 *Let $u_1, \ldots, u_n \in \mathbb{R}$ and $v_1, \ldots, v_n \in \mathbb{R}$. Order them $u_{(1)} \leq \cdots \leq u_{(n)}$ and $v_{(1)} \leq \cdots \leq v_{(n)}$. Then*

$$\sum_{j=1}^n |u_{(j)} - v_{(j)}|^p \leq \sum_{j=1}^n |u_j - v_j|^p.$$

Proof Assume $u_1 < \cdots < u_n$, $v_1 < \cdots < v_n$, and $p \geq 1$. Let u and v denote $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_n)$. Let Σ_n denote the symmetric group on n letters and let $f_n : \Sigma_n \to \mathbb{R}$ be given by $f_n(\sigma) = \sum_{j=1}^n |u_j - v_{\sigma(j)}|^p$. We will prove by induction that if $f_n(\sigma)$ is minimal then σ is the identity, which we denote by 1.

For $n = 1$ this is trivial. For $n = 2$ assume without loss of generality that $u_1 = 0$, $u_2 = 1$ and $0 \leq v_1 < v_2$. Let 1 and τ denote the elements of Σ_2. Then $f(1) = v_1^p + |1 - v_2|^p$ and $f(\tau) = v_2^p + |1 - v_1|^p$. Notice that $f(1) < f(\tau)$ if and only if $v_1^p - |1 - v_1|^p < v_2^p - |1 - v_2|^p$. The result follows from checking that $g(x) = x^p - |1 - x|^p$ is an increasing function for $x \geq 0$.

Now assume that the statement is true for some $n \geq 2$. Assume that $f_{n+1}(\sigma^*)$ is minimal. Fix $1 \leq i \leq n+1$. Let $u' = (u_1, \ldots, \hat{u}_i, \ldots, u_{n+1})$ and $v' = (v_1, \ldots, \hat{v}_{\sigma^*(i)}, \ldots, v_{n+1})$, where $\hat{\cdot}$ denotes omission. Since $f_{n+1}(\sigma^*)$ is minimal for u and v, it follows that $\sum_{j=1, j\neq i}^{n} |u_j - v_{\sigma^*(j)}|$ is minimal for u' and v'. By the induction hypothesis, for $1 \leq j < k \leq n+1$ and $j, k \neq i$, $\sigma^*(j) < \sigma^*(k)$. Therefore $\sigma^* = 1$. Thus, by induction, the statement is true for all n.

Hence $\sum_{j=1}^{n} |u_{(j)} - v_{(j)}|^p \leq \sum_{j=1}^{n} |u_j - v_j|^p$ if $u_{(1)} < \cdots < u_{(n)}$ and $v_{(1)} < \cdots < v_{(n)}$. The statement in the lemma follows by continuity. ■

Lemma 19 *Let $x = (b, d)$ and $x' = (b', d')$ where $b \leq d$ and $b' \leq d'$. Let $\ell = d - b$ and $\varepsilon = \|x - x'\|_\infty$. Then $\|\lambda(\{x\}) - \lambda(\{x'\})\|_p^p \leq \ell\varepsilon^p + \frac{2}{p+1}\varepsilon^{p+1}$.*

Proof Let $\lambda = \lambda(\{x\})$ and $\lambda' = \lambda(\{x'\})$. First $\lambda_k = \lambda'_k = 0$ for $k > 1$; so $\|\lambda - \lambda'\|_p = \|\lambda_1 - \lambda'_1\|_p$. Second $\lambda_1(t) = (h - |t - m|)_+$, where $h = \frac{d-b}{2}$, $m = \frac{b+d}{2}$, and $y_+ = \max(y, 0)$, and similarly for λ'_1 (see Figure 2).

Fix x and ε. As x' moves along the square $\|x - x'\|_\infty = \varepsilon$, $\|\lambda_1 - \lambda'_1\|_p^p$ has a maximum if $x' = (a - \varepsilon, b + \varepsilon)$. In this case $\|\lambda_1 - \lambda'_1\|_p^p = 2\int_0^h \varepsilon^p \, dt + 2\int_0^\varepsilon t^p \, dt = \ell\varepsilon^p + \frac{2}{p+1}\varepsilon^{p+1}$. ■

Proof [Proof of Corollary 15] Let $\varphi : D \xrightarrow{\cong} D'$ be a minimizer for $W_p(D, D')$, with corresponding $\{\varepsilon_j\}$. Assume that $W_p(D, D') \leq 1$. Then $W_p(D, D')^p = \sum_{j=1}^{n} \varepsilon_j^p \leq 1$. So for $1 \leq j \leq n$, $\varepsilon_j \leq 1$. Combining this with Theorem 14, we have that

$$\Lambda_p(D, D')^p \leq \sum_{j=1}^{n} \left(\ell_j + \frac{2}{p+1} \right) \varepsilon_j^p. \tag{8}$$

Since $W_\infty(D, \emptyset) = \max \frac{1}{2}\ell_j$, $\ell_j \leq 2\, W_\infty(D, \emptyset)$. Hence

$$\Lambda_p(D, D')^p \leq 2 \left(W_\infty(D, \emptyset) + \frac{1}{p+1} \right) W_p(D, D')^p. \tag{9}$$

Therefore $W_p(D, D')^p \geq 1$ or $W_p(D, D')^p \geq \frac{1}{2}\left[W_\infty(D, \emptyset) + \frac{1}{p+1} \right]^{-1} \Lambda_p(D, D')^p$. The statement of the corollary follows. ■

Theorem 16 follows from the following corollary to Theorem 14 which is of independent interest.

Corollary 20 *Let $p \geq k \geq 1$. Then*

$$\Lambda_p(D, D')^p \leq W_\infty(D, D')^{p-k} \Big[W_\infty(D, \emptyset)(\mathrm{Pers}_k(D) + \mathrm{Pers}_k(D')) + \frac{1}{p+1}(\mathrm{Pers}_{k+1}(D) + \mathrm{Pers}_{k+1}(D')) \Big]$$

Proof Let φ be a minimizer for $W_\infty(D, D')$ with corresponding $\{\varepsilon_j\}$. If $\varepsilon_j > \frac{\ell_j}{2} + \frac{\ell'_j}{2}$ then modify φ to pair $x_j = (b_j, d_j)$ with $\bar{x}_j = (\frac{b_j+d_j}{2}, \frac{b_j+d_j}{2})$ and similarly for x'_j. Note that $\|x_j - \bar{x}_j\|_\infty = \frac{\ell_j}{2}$ and $\|x'_j - \bar{x'}_j\|_\infty = \frac{\ell'_j}{2}$, so φ is still a minimizer for $W_\infty(D, D')$.

Recall that for all j, $\ell_j \leq 2\, W_\infty(D, \emptyset)$. Since φ is a minimizer for $W_\infty(D, D')$, for all j, $\varepsilon_j \leq W_\infty(D, D')$. So applying our choice of φ to Theorem 14 we have,

$$\Lambda_p(D, D')^p \leq W_\infty(D, D')^{p-k} \left[2\, W_\infty(D, \emptyset) \sum_{j=1}^{n} \varepsilon_j^k + \frac{2}{p+1} \sum_{j=1}^{n} \varepsilon_j^{k+1} \right].$$

Now $\varepsilon_j^q \leq \left(\frac{\ell_j}{2} + \frac{\ell'_j}{2}\right)^q \leq \frac{1}{2}\left((\ell_j)^q + (\ell'_j)^q\right)$ for $q \geq 1$, where the right hand side follows by the convexity of $\alpha(x) = x^q$ for $q \geq 1$. Thus $\sum_{j=1}^n \varepsilon_j^q \leq \frac{1}{2}(\mathrm{Pers}_q(D) + \mathrm{Pers}_q(D'))$ for $q \geq 1$. The result follows. ■

Proof [Proof of Theorem 16] Theorem 16 follows from Corollary 20 by the following two observations. First, by the stability theorem of Cohen-Steiner et al. (2007), $W_\infty(D(f), D(g)) \leq \|f - g\|_\infty$ and $W_\infty(D(f), \emptyset) \leq \|f\|_\infty$. Second, if $\mathrm{Pers}_q(D(f)) \leq C_{X,q}$ for all tame Lipschitz functions $f : X \to \mathbb{R}$ with $\mathrm{Lip}(f) \leq 1$, then for general tame Lipschitz functions, $\mathrm{Pers}_q(D(f)) \leq C_{X,q} \,\mathrm{Lip}(f)^q$. ■

References

Robert J. Adler and Jonathan E. Taylor. *Random Fields and Geometry*. Springer Monographs in Mathematics. Springer, New York, 2007. ISBN 978-0-387-48112-8.

Robert J. Adler, Omer Bobrowski, Matthew S. Borman, Eliran Subag, and Shmuel Weinberger. Persistent homology for random fields and complexes. In *Borrowing Strength: Theory Powering Applications—a Festschrift for Lawrence D. Brown*, volume 6 of *Inst. Math. Stat. Collect.*, pages 124–143. Inst. Math. Statist., Beachwood, OH, 2010.

Andrew J. Blumberg, Itamar Gal, Michael A. Mandell, and Matthew Pancia. Robust statistics, hypothesis testing, and confidence intervals for persistent homology on metric measure spaces. *Found. Comput. Math.*, 14(4):745–789, 2014. ISSN 1615-3375. doi: 10.1007/s10208-014-9201-4. URL `http://dx.doi.org/10.1007/s10208-014-9201-4`.

Omer Bobrowski and Matthew Strom Borman. Euler integration of Gaussian random fields and persistent homology. *J. Topol. Anal.*, 4(1):49–70, 2012. ISSN 1793-5253.

Karol Borsuk. On the imbedding of systems of compacta in simplicial complexes. *Fund. Math.*, 35:217–234, 1948. ISSN 0016-2736.

Peter Bubenik and Jonathan A. Scott. Categorification of persistent homology. *Discrete Comput. Geom.*, 51(3):600–627, 2014. ISSN 0179-5376.

Peter Bubenik, Gunnar Carlsson, Peter T. Kim, and Zhi-Ming Luo. Statistical topology via Morse theory persistence and nonparametric estimation. In *Algebraic Methods in Statistics and Probability II*, volume 516 of *Contemp. Math.*, pages 75–92. Amer. Math. Soc., Providence, RI, 2010.

Gunnar Carlsson. Topology and data. *Bull. Amer. Math. Soc. (N.S.)*, 46(2):255–308, 2009. ISSN 0273-0979.

Gunnar Carlsson, Tigran Ishkhanov, Vin de Silva, and Afra Zomorodian. On the local behavior of spaces of natural images. *Int. J. Comput. Vision*, 76(1):1–12, 2008. ISSN 0920-5691.

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. ISBN 978-1-60558-501-7.

Frederic Chazal, Vin de Silva, Marc Glisse, and Steve Oudot. The structure and stability of persistence modules. arXiv:1207.3674 [math.AT], 2012.

Frédéric Chazal, Marc Glisse, Catherine Labruère, and Bertrand Michel. Optimal rates of convergence for persistence diagrams in topological data analysis. 2013. arXiv:1305.6239 [math.ST].

Frédéric Chazal, Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, and Larry Wasserman. Stochastic convergence of persistence landscapes and silhouettes. *Symposium on Computational Geometry (SoCG)*, 2014.

Chao Chen and Michael Kerber. An output-sensitive algorithm for persistent homology. *Comput. Geom.*, 46(4):435–447, 2013. ISSN 0925-7721.

Moo K. Chung, Peter Bubenik, and Peter T. Kim. Persistence diagrams in cortical surface data. In *Information Processing in Medical Imaging (IPMI) 2009*, volume 5636 of *Lecture Notes in Computer Science*, pages 386–397, 2009.

David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete Comput. Geom.*, 37(1):103–120, 2007. ISSN 0179-5376.

David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Extending persistence using Poincaré and Lefschetz duality. *Found. Comput. Math.*, 9(1):79–103, 2009. ISSN 1615-3375.

David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Yuriy Mileyko. Lipschitz functions have L_p-stable persistence. *Found. Comput. Math.*, 10(2):127–139, 2010. ISSN 1615-3375.

William Crawley-Boevey. Decomposition of pointwise finite-dimensional persistence modules. arXiv:1210.0819 [math.RT], 2012.

Vin de Silva and Gunnar Carlsson. Topological estimation using witness complexes. *Eurographics Symposium on Point-Based Graphics*, 2004.

Vin De Silva and Robert Ghrist. Coverage in sensor networks via persistent homology. *Algebr. Geom. Topol.*, 7:339–358, 2007a.

Vin De Silva and Robert Ghrist. Homological sensor networks. *Notic. Amer. Math. Soc.*, 54(1):10–17, 2007b.

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, 08 2008.

Tamal Krishna Dey, Fengtao Fan, and Yusu Wang. Graph induced complex on point data. In *Proceedings of the Twenty-ninth Annual Symposium on Computational Geometry*, SoCG '13, pages 107–116, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2031-3.

Persi Diaconis, Susan Holmes, and Mehrdad Shahshahani. Sampling from a manifold. arXiv:1206.6913 [math.ST], 2012.

Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological persistence and simplification. *Discrete Comput. Geom.*, 28(4):511–533, 2002. ISSN 0179-5376. Discrete and computational geometry and graph drawing (Columbia, SC, 2001).

Herbert Edelsbrunner, Dmitriy Morozov, and Amit Patel. Quantifying transversality by measuring the robustness of intersections. *Found. Comput. Math.*, 11(3):345–361, 2011. ISSN 1615-3375.

Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, Larry Wasserman, Sivaraman Balakrishnan, and Aarti Singh. Confidence sets for persistence diagrams. *Ann. Statist.*, 42(6):2301–2339, 2014. ISSN 0090-5364. doi: 10.1214/14-AOS1252. URL `http://dx.doi.org/10.1214/14-AOS1252`.

Robert Ghrist. Barcodes: the persistent topology of data. *Bull. Amer. Math. Soc. (N.S.)*, 45(1):61–75, 2008. ISSN 0273-0979.

Allen Hatcher. *Algebraic Topology*. Cambridge University Press, Cambridge, 2002. ISBN 0-521-79160-X; 0-521-79540-0.

Giseon Heo, Jennifer Gamble, and Peter T. Kim. Topological analysis of variance and the maxillary complex. *J. Amer. Statist. Assoc.*, 107(498):477–492, 2012. ISSN 0162-1459.

J. Hoffmann-Jørgensen and G. Pisier. The law of large numbers and the central limit theorem in Banach spaces. *Ann. Probability*, 4(4):587–599, 1976.

Violeta Kovacev-Nikolic, Giseon Heo, Dragan Nikolić, and Peter Bubenik. Using cycles in high dimensional data to analyze protein binding. 2014. arXiv:1412.1394 [stat.ME].

Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces*. Classics in Mathematics. Springer-Verlag, Berlin, 2011. ISBN 978-3-642-20211-7. Isoperimetry and processes, Reprint of the 1991 edition.

Yuriy Mileyko, Sayan Mukherjee, and John Harer. Probability measures on the space of persistence diagrams. *Inverse Problems*, 27(12):124007, 22, 2011. ISSN 0266-5611.

Nikola Milosavljević, Dmitriy Morozov, and Primož Škraba. Zigzag persistent homology in matrix multiplication time. In *Computational Geometry (SCG'11)*, pages 216–225. ACM, New York, 2011.

Dimitriy Morozov. Dionysus: a C++ library with various algorithms for computing persistent homology. Software available at `http://www.mrzv.org/software/dionysus/`, 2012.

Elizabeth Munch, Paul Bendich, Katharine Turner, Sayan Mukherjee, Jonathan Mattingly, and John Harer. Probabilistic fréchet means and statistics on vineyards. 2013. arXiv:1307.6530 [math.PR].

Vidit Nanda. Perseus: the persistent homology software. Software available at `http://www.math.rutgers.edu/~vidit/perseus/index.html`, 2013.

Monica Nicolau, Arnold J. Levine, and Gunnar Carlsson. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proc. Nat. Acad. Sci.*, 108(17):7265–7270, 2011.

Andrew Robinson and Katharine Turner. Hypothesis testing for topological data analysis. 2013. arXiv:1310.7467 [stat.AP].

Andrew Tausz, Mikael Vejdemo-Johansson, and Henry Adams. Javaplex: a research software package for persistent (co)homology. Software available at `http://code.google.com/javaplex`, 2011.

Katharine Turner, Yuriy Mileyko, Sayan Mukherjee, and John Harer. Fréchet means for distributions of persistence diagrams. *Discrete Comput. Geom.*, 52(1):44–70, 2014.

Andrew T. A. Wood and Grace Chan. Simulation of stationary Gaussian processes in $[0, 1]^d$. *J. Comput. Graph. Statist.*, 3(4):409–432, 1994. ISSN 1061-8600.

Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. *Discrete Comput. Geom.*, 33(2):249–274, 2005. ISSN 0179-5376.