Derek Sorensen

Stability of Persistent Homology

1 Introduction

One important use of persistent homology in computational algebraic topology is to identify the general shape or form of a given dataset, or to identify the general shape or manifold on which points in a dataset might lie. This is useful for many reasons, for example facial and shape recognition [CCSGMO09].

Given a dataset, hereafter referred to as a point cloud, the goal is to fill in the empty space between the points and make a resulting topological space in a reasonable way which will give us information about our dataset. Our process of using the tools of persistent homology essentially consists of assigning to any dataset a filtered simplicial complex, called its *Rips filtration*, and observing how the homology of these spaces evolves using *barcodes* or *persistence modules*. This gives us a directed system of homology groups. However, in order for it to be useful we ought to have some notion of stability. That is, we ought to have some notion of what it means for two datasets to be 'similar,' or 'close together,' what it means for the corresponding persistence modules to be 'similar' or 'close together.'

This is precisely where the notion of stability comes in: The goal of our paper is to show that this process of taking the persistent homology of the Rips filtration is stable in the sense that a 'small change' in point clouds or Rips filtrations will yield a 'small change' in the corresponding persistence modules. If we slightly weaken the result, we are saying that the map which sends a space to its corresponding persistence module is continuous. This is important in any applied science; a method of measurement or analysis is hardly useful if minuscule changes in the data yield wildly different results.

We will treat two distinct stability theorems, Theorems 4.1 and 3.1, which will bound the distance of either the barcodes or the persistence modules (which we will show to be essentially equivalent notions) obtained through different filtrations by the original distance of the two corresponding filtrations.

Although the notions of persistent homology may be used most widely for analyzing point clouds, the theory developed here can be generalized; in the case of Theorem 3.1 to finite metric spaces and the persistent homology of their corresponding Rips filtrations, and in the case of Theorem 4.1 to any pair of filtrations of an arbitrary topological space X which are derived from real-valued maps $f, g : X \to \mathbb{R}$.

In both cases, however, we construct two separate filtrations of complexes, indexed over the category $(\mathbb{R}, \leq)$, and measure the distance of their corresponding persistent homologies. As it turns out, under certain circumstances we can take any pair of Rips filtrations of a point cloud and construct a topological space X with maps $f, g : X \to \mathbb{R}$ corresponding, respectively, to the two Rips filtrations. This allows us to show stability of the persistent homology of the point cloud (in the sense of Theorem 3.1) via stability in the sense of Theorem 4.1. This will not quite show that

Theorem 3.1 is a special case of Theorem 4.1, but it does support the argument that the latter theorem is considerably more general than the former.

Throughout this paper we will assume that the reader is familiar with the material from an introductory course to algebraic topology and category theory, including homology with coefficients from an abelian group, functors, universal properties, universal objects, abstract simplicial complexes, etc. We also assume the reader is familiar with tools used in persistent homology, although we will supply the definition of anything we need aside from the foundational material.

Finally, in this paper unless otherwise stated we will be considering the persistent homology with coefficients in a field $\mathbb{F}$. The motivation for this is that the k^{th} homology group of a topological space X with coefficients in a field is always a vector space over $\mathbb{F}$; if it is finite-dimensional, then we have that $H_k(X, \mathbb{F}) = \bigoplus_{i=1}^{n} \mathbb{F}$ for some n. This allows us to very concisely consider the persistent homology in terms of $\mathbb{F}$-vector spaces. Since persistent homology will generally apply to point clouds and their Rips complexes, we will assume that every homology group which we consider will be a finite-dimensional $\mathbb{F}$ vector space. We will denote the category of finite-dimensional vector spaces over $\mathbb{F}$ as $\mathrm{Vec}_{\mathbb{F}}$ or, abusing notation, simply as Vec.

2 Definitions and Metrics

Let $\mathbb{F}$ be a field. Recall that, for a category $\mathcal{J}$, a diagram of shape $\mathcal{J}$ in a category $\mathcal{C}$ is a functor $\mathcal{J} \to \mathcal{C}$. Also, recall that for categories $\mathcal{C}$ and $\mathcal{D}$, $\mathcal{C}^{\mathcal{D}}$ is the category of diagrams $F : \mathcal{C} \to \mathcal{D}$. Finally, we denote by $[n]$ the full subcategory of the category $(\mathbb{Z}, \leq)$ which consists of objects $1, ..., n \in \mathbb{Z}$ and morphisms $i \leq j$ for all $1 \leq i \leq j \leq n$.

Definition 2.1 (Persistence Module). *A persistence module is any diagram in* $Vec^{(\mathbb{R},\leq)}$, $Vec^{[n]}$ *or* $Vec^{(\mathbb{Z}_+,\leq)}$.

One can think of a persistence module as a special kind of directed system of finite-dimensional $\mathbb{F}$ vector spaces. Recall from the introduction that a point cloud is a collection of discrete data points, often in $\mathbb{R}^d$. In particular, point clouds often form finite metric spaces. It is for these spaces that we will define a filtration of topological spaces in $\mathrm{Top}^{(\mathbb{R},\leq)}$ as follows:

Definition 2.2 (Rips Filtration of a Discrete Metric Space). *Let* (X, d_X) *be a discrete metric space endowed with a real-valued labeling* $f : X \to \mathbb{R}$ *(note that any choice of* f *is continuous because* X *is discrete). Let* $\varepsilon \in \mathbb{R}$*. Then the Rips complex* $\mathcal{R}_\varepsilon(X, \mathrm{d}_X)$ *is the abstract simplicial complex whose simplices are of the form*

$$\sigma \subseteq X \textit{ such that } \exists \textit{ some } \varepsilon\textit{-ball } U \subseteq X \textit{ such that } \sigma \subseteq U.$$

The Rips filtration of (X, d_X) *with regard to the map* f *and denoted* $\mathcal{R}(X, \mathrm{d}_X, f)$ *is the filtration of Rips complexes given by*

$$\{\mathcal{R}_\varepsilon(X_\varepsilon)\}_{\varepsilon\in\mathbb{R}} \textit{ where } X_\varepsilon := f^{-1}((-\infty, \varepsilon]) \subseteq X$$

In order to properly continue into our two stability theorems, or even think about stability at all, we will need to define a few different metrics so that we can sensibly talk about what it means for two barcodes or maps in $\mathrm{Sets}(X, \mathbb{R})$ to be "close together." We will define metrics on $\mathrm{Sets}(X, \mathbb{R})$, on compact metric spaces and on persistence modules. By $\mathrm{Sets}(X, \mathbb{R})$ we mean the set of maps $X \to \mathbb{R}$ where X is the underlying set of an object in Top and $\mathbb{R}$ is considered as a set, and not a topological space.

The metric on $\mathrm{Sets}(X, \mathbb{R})$ is quite standard and straightforward, and given by

$$\mathrm{d}_\infty(f, g) = \|f - g\|_\infty := \sup_{x \in X} |f(x) - g(x)|. \tag{1}$$

Note that if X is compact and f is continuous, $\|f - g\|_\infty$ is actually $\max_{x \in X} |f(x) - g(x)|$. It is a standard result that this results in a metric on $\mathrm{Sets}(X, \mathbb{R})$.

2.1 Gromov-Hausdorff Metric

Our next task is to define a metric which will be suitable for the Rips complexes of finite metric spaces. We would also like to somehow incorporate into this metric arbitrary real-valued labelings of X, or continuous maps $f : X \to \mathbb{R}$. As it turns out, the most natural metric for this will readily generalize to a metric on compact metric spaces, so for the sake of generality we will weaken our hypotheses and let X simply be a compact metric space.

Following the conventions of [CCSGMO09], we define a category $\mathcal{X}_1$ with objects of the form (X, d_X, f), where (X, d_X) is a compact metric space and $f : X \to \mathbb{R}$ is continuous. A morphism between (X, d_X, f) and (Y, d_Y, g) is a morphism $\varphi : X \to Y$ of compact metric spaces which is compatible with the maps f and g, i.e. $f = g \circ \varphi$. Thus the identity map is simply $id : X \to X$, composition of maps is well-defined and an isomorphism is simply a morphism which is also an isometry of metric spaces.

By defining a metric on $\mathcal{X}_1$, we define a metric on any subset of the class of objects of $\mathcal{X}_1$. In order to do so, we will first have to give a brief definition.

Definition 2.3. *A correspondence between two sets X and Y is a subset $C \subseteq X \times Y$ such that the canonical projection maps $\pi_X : C \to X$ and $\pi_Y : C \to Y$ are surjective. This yields a category $\mathcal{C}$ with objects as sets and morphisms as correspondences. Thus the set of all correspondences between X and Y is denoted $\mathcal{C}(X, Y)$.*

Definition 2.4 (Gromov-Hausdorff Metric on $\mathcal{X}_1$)**.** *Let $(X, \mathrm{d}_X), (Y, \mathrm{d}_Y)$ be compact metric spaces endowed with maps $f : X \to \mathbb{R}$ and $g : Y \to \mathbb{R}$. Then we define the Gromov-Hausdorff metric on $\mathcal{X}_1$ as follows:*

$$d^1_{GH}((X, \mathrm{d}_X, f), (Y, \mathrm{d}_Y, g)) := \tag{2}$$

$$\inf_{C \in \mathcal{C}(X,Y)} \max\{\frac{1}{2} \sup_{(x,y),(x',y') \in C} |d_X(x, x') - d_Y(y, y')|; \sup_{(x,y) \in C} |f(x) - g(y)|\}$$

This metric looks a little bit perplexing, but we hope to motivate its definition. Notice that if we consider the diagonal correspondence, $C = \{(x, x) \,|\, x \in X\}$, then for all $(x, x), (x', x') \in C$,

$$\sup_{(x,x),(x',x')\in C} |d_X(x, x') - d_Y(x, x')| = 0$$

and so the right-hand side of (2) becomes

$$\sup_{(x,x)\in C} |f(x) - g(x)|,$$

which is precisely the equation for $\|f - g\|_\infty$.

Now in the case that $f = g \equiv 0$,

$$\mathrm{d}^1_{\mathrm{GH}}((X, \mathrm{d}_X, f), (Y, \mathrm{d}_Y, g)) = \inf_{C\in\mathcal{C}(X,Y)} \frac{1}{2} \sup_{(x,y),(x',y')\in C} |d_X(x, x') - d_Y(y, y')|. \quad (3)$$

It turns out that the right-hand side of (3) is a metric on compact metric spaces, which we will denote as simply d_{GH}. As given in (3), the metric gives essentially no intuition on what it might be measuring. There is an equivalent definition (which is in practice much more difficult to use) that will shed some light on what d_{GH} is actually doing.

This metric will measure, using the Hausdorff distance, how similarly X and Y can be embedded into some larger metric space (Z, d_Z). Recall that the Hausdorff distance of two subsets X and Y of (Z, d_Z) is given by

$$\mathrm{d}^Z_H(X, Y) = \max\{\sup_{x\in X} \inf_{y\in Y} \mathrm{d}_Z(x, y); \sup_{y\in Y} \inf_{x\in X} \mathrm{d}_Z(x, y)\} \quad (4)$$

Definition 2.5 (Gromov-Hausdorff Metric)**.** *The Gromov-Hausdorff distance between compact metric spaces (X, d_x) and (Y, d_Y) is give by:*

$$\mathrm{d}_{GH}((X, \mathrm{d}_x), (Y, \mathrm{d}_Y)) := \inf_{(Z,\mathrm{d}_Z),\gamma_X,\gamma_Y} \mathrm{d}^Z_H(\gamma_X(X), \gamma_Y(Y)) \quad (5)$$

where (Z, d_Z) ranges over all metric spaces and γ_X, γ_Y range over isometric embeddings of (X, d_X) and (Y, d_Y) into (Z, d_Z).

The definition in (5), intuitively, shows us that d_{GH} is measuring how far apart (X, d_X) and (Y, d_Y) are from being isometric. These two definitions of d_{GH} are equivalent by Theorem 7.3.25 of [BBI01]. Furthermore, d_{GH} defines a metric on isometry classes of compact metric spaces by Theorem 7.3.30 of [BBI01] and $\mathrm{d}^1_{\mathrm{GH}}$ defines a metric on isomorphism classes of the objects of $\mathcal{X}_1$ by Theorem 2.5 in [CCSGMO09].

2.2 Metrics on Persistence Modules

The final metrics which we will define will be on persistence modules of filtered topological spaces in $\mathrm{Top}^{(\mathbb{R},\leq)}$. The first will be the more general inteleaving distance, which measures distance on diagrams of the form $\mathcal{D}^{(\mathbb{R},\leq)}$, for any category $\mathcal{D}$. Using the interleaving metric we will define the bottleneck distance d_B, which gives a metric on barcodes. These two metrics will turn out to be equivalent in the case that our barcodes are finite. Since our main focus of this paper is to analyze the persistent homology of finite metric spaces (point clouds), we are guaranteed that the barcode corresponding to the persistent homology will be finite in the cases which we're interested in. For that reason, after we prove the equivalence of metrics on diagrams of finite type at the end of this section, we will work entirely with persistence modules and the interleaving metric, noting that all of the theorems can be equivalently stated simply using the bottleneck distance and barcodes. This will give us a more general approach to our subject, and in particular will make comparing our two main theorems much simpler.

Recall that a diagram $F \in \mathrm{Vec}^{(\mathbb{R},\leq)}$ is *of finite type* if $F = \bigoplus_{k=1}^{n} \chi_{I_k}$, where I_k is an interval in $\mathbb{R}$ and $\chi_{I_k} \in \mathrm{Vec}^{(\mathbb{R},\leq)}$ is a persistence module defined by $\chi_{I_k}(a) = 0$ if $a \notin I_k$, and $\mathbb{F}$ if $a \in I_k$, and $\chi_{I_k}(a \leq b) = 0$ if a or $b \notin I_k$ and is an iso otherwise.

We will quickly define two maps before giving the definition of the interleaving metric. Given $\varepsilon > 0$, let $T_\varepsilon : (\mathbb{R},\leq) \to (\mathbb{R},\leq)$ be the functor given by $a \mapsto a+\varepsilon$ and let $\eta_\varepsilon : \mathrm{Id}_{(\mathbb{R},\leq)} \Rightarrow T_\varepsilon$ be the natural transformation given by component maps $\eta_\varepsilon(a) : a \leq a+\varepsilon$. We define the interleaving metric, d, as follows:

Definition 2.6 (Interleaving Metric)**.** *Let $\mathcal{D}$ be any category, and let $F, G \in \mathcal{D}^{(\mathbb{R},\leq)}$. An $\varepsilon-$interleaving of F and G consists of natural transformation $\varphi : F \Rightarrow GT_\varepsilon$ and $\psi : G \Rightarrow FT_\varepsilon$ such that the following diagrams commute:*

$$\begin{array}{ccc} F(a) & \xrightarrow{\qquad\qquad} & F(a+2\varepsilon) \\ & {\scriptstyle\varphi(a)}\searrow \quad \nearrow{\scriptstyle\psi(a+\varepsilon)} & \\ & G(a+\varepsilon) & \end{array} \qquad \begin{array}{ccc} & F(a+\varepsilon) & \\ & {\scriptstyle\psi(a)}\nearrow \quad \searrow{\scriptstyle\varphi(a+\varepsilon)} & \\ G(a) & \xrightarrow{\qquad\qquad} & G(a+2\varepsilon) \end{array}$$

i.e. such that $(\psi T_\varepsilon)\varphi = F\eta_{2\varepsilon}$ and $(\varphi T_\varepsilon)\psi = G\eta_{2\varepsilon}$. If there is an $\varepsilon-$interleaving between F and G, we say that F and G are $\varepsilon-$interleaved.

Then define the interleaving metric as $\mathrm{d}(F,G) := \inf\{\varepsilon \geq 0 \mid F \text{ and } G \text{ are } \varepsilon\text{−interleaved}\}$. *We write* $\mathrm{d}(F,G) = \infty$ *if there is no $\varepsilon \geq 0$ for which F and G are $\varepsilon-$interleaved.*

We saw in our course that d defined above is a pseudometric on any subset of $\mathcal{D}^{(\mathbb{R},\leq)}$, and if we identify the diagrams F and G which satisfy $\mathrm{d}(F,G) = 0$, we have a metric on the resulting equivalence classes. Furthermore, in the homework exercises we proved that for any functor $H : \mathcal{D} \to \mathcal{C}$,

$$\mathrm{d}(HF, HG) \leq \mathrm{d}(F,G). \tag{6}$$

We can use the interleaving metric to define the bottleneck distance, which is a metric on barcodes. More generally, it's a metric on multisets. In the end, we will care only about finite barcodes (which correspond to diagrams of finite type), but in the meantime we can define d_B on any pair of multisets. In the following definition, we will consider any persistence module in $F \in \mathrm{Vec}^{(\mathbb{R},\leq)}$ equivalently as a barcode, where the barcode of F is the multiset $\{I_\alpha\}_{\alpha\in P}$, where $F = \bigoplus_{\alpha\in P} \chi_{I_\alpha}$. We also note here that we are following the conventions and definitions in [BS14].

Definition 2.7 (Bottleneck Distance)**.** *If A and B are multisets, then define A_B to be the disjoint union of A and the multiset containing the empty interval $\emptyset$ with cardinality $|B|$. A partial matching between two multisets A and B is a bijection $f : A_B \to B_A$ and is written $f : A \rightleftharpoons B$.*

Then the bottleneck distance d_B between A and B is given by

$$\mathrm{d}_B(A, B) := \inf_{f:A\rightleftharpoons B} \sup_{I_A\in A} \mathrm{d}(\chi_I, \chi_{f(I)})$$

where $\mathrm{d}(\chi_I, \chi_{f(I)})$ is the interleaving distance.

As stated before, we can identify these metrics in the case that our barcodes are finite (and thus the corresponding diagrams are of finite type). For all intents and purposes of this analysis, this suffices and so we will do so. After the proof of the following theorem, instead of considering barcodes with the bottleneck distance, we will only consider diagrams of finite type with the interleaving distance. This will make our comparison in Section 5 much more straightforward.

Theorem 2.1 (Theorem 4.16 in [BS14])**.** *Let $\mathcal{B}$ be the set of finite barcodes, d_B be the bottleneck distance and d the interleaving distance. Then $\chi : \mathcal{B} \to Vec^{(\mathbb{R},\leq)}$ given by $\chi(\{I_k\}_{k=1}^n) = \oplus_{k=1}^n \chi_{I_k}$ is an isometric embedding of metric spaces*

$$\chi : (\mathcal{B}, \mathrm{d}_B) \hookrightarrow (Vec^{(\mathbb{R},\leq)}, \mathrm{d}). \tag{7}$$

Proof. We know already from Theorem 4.4 of [CCSGMO09] that

$$\mathrm{d}_B(B, B') \leq \mathrm{d}(\chi(B), \chi(B')).$$

So to show the isometric embedding we only need to show that

$$\mathrm{d}(\chi(B), \chi(B')) \leq \mathrm{d}_B(B, B'). \tag{8}$$

Note that the inequality in (8) is trivial if $\mathrm{d}_B(B, B') = \infty$. So we assume that $\mathrm{d}_B(B, B') < \infty$. Thus there is some map $f : B \rightleftharpoons B'$ such that $\sup_{I\in B_{B'}} \mathrm{d}(\chi_I, \chi_{f(I)}) = \gamma < \infty$ (recall the notation of $B_{B'}$ and B'_B, where f is a bijection $B_{B'} \xrightarrow{\sim} B'_B$). Note that for all $I \in B_{B'}$, by definition $\mathrm{d}(\chi_I, \chi_{f(I)}) \leq \gamma$. Thus if we take some $\varepsilon > \gamma$, χ_I and $\chi_{f(I)}$ are $\varepsilon-$interleaved. Since for any $F \in \mathrm{Vect}^{(\mathbb{R},\leq)}$, $F \cong F \oplus \chi_\emptyset$, it follows that $\chi(B) \cong \bigoplus_{I\in B_{B'}} \chi_I$ and $\chi(B') \cong \bigoplus_{I\in B'_B} \chi_I$. Corollary 7.11 in [BS14] gives us that

$$\mathrm{d}(\bigoplus_{I\in B_{B'}} \chi_I, \bigoplus_{I\in B'_B} \chi_I) \leq \sup_{I\in B_{B'}} \mathrm{d}(\chi_I, \chi_{f(I)}),$$

and so since $\mathrm{d}(\chi_I, \chi_{f(I)}) \leq \varepsilon$, we have that $\mathrm{d}(\bigoplus_{I\in B_{B'}} \chi_I, \bigoplus_{I\in B'_B} \chi_I) = \mathrm{d}(\chi(B), \chi(B')) \leq \varepsilon$.

We can run this process over all $f : B \rightleftharpoons B'$, and by definition of d_B we have that $\mathrm{d}(\chi(B), \chi(B')) \leq \varepsilon$ for all $\varepsilon > \mathrm{d}_B(B, B')$. This gives straight away that (8) holds by definition of $\mathrm{d}(\chi(B), \chi(B'))$ as an infemum, and we have our result. ∎

3 Stability Theorem for Point Clouds

Now that we have the definitions and some basic results about our metrics, we are finally ready to look at our stability theorems. We will first treat the less general case—that of the stability theorem for point clouds—and then move onto the more general result.

For intuition, the way that we are going to present the first result on stability is by defining the function

$$\Psi_k : (\mathcal{X}_1, \mathrm{d}^1_{\mathrm{GH}}) \to (\mathrm{Vec}^{(\mathbb{R},\leq)}, \mathrm{d}) \tag{9}$$

which maps a finite metric space (X, d_X) to the k^{th} persistence diagram of $\mathcal{R}(X, \mathrm{d}_X, f)$. Note that this is not a functor, but rather strictly a map between collections of objects. Since we have the metrics $\mathrm{d}^1_{\mathrm{GH}}$ and d, we can consider whether or not Ψ_k is continuous—i.e. whether a small change in (X, d_X, f) will result in a small change in its corresponding persistent module. Our main result will will give an bound on the distance between persistent modules in terms of the Gromov-Hausdorff distance of the original spaces, which implies immediately that Ψ_k is continuous for all $k \in \mathbb{N}$.

Theorem 3.1. *Let $(X, \mathrm{d}_X, f), (Y, \mathrm{d}_Y, g) \in \mathcal{X}_1$, where X and Y are finite metric spaces (such as a point cloud in $\mathbb{R}^n$). Then*

$$\mathrm{d}(\mathcal{R}(X, \mathrm{d}_X, f), \mathcal{R}(Y, \mathrm{d}_Y, g)) \leq d^1_{GH}((X, \mathrm{d}_X, f), (Y, \mathrm{d}_Y, g)). \tag{10}$$

Corollary 3.1.1. *By Equation 6, for any category $\mathcal{D}$ and functor $H : Top \to \mathcal{D}$,*

$$\mathrm{d}(H(\mathcal{R}(X, \mathrm{d}_X, f)), H(\mathcal{R}(Y, \mathrm{d}_Y, g))) \leq d^1_{GH}((X, \mathrm{d}_X, f), (Y, \mathrm{d}_Y, g))$$

In particular, if we choose H to be the k^{th} homology functor H_k, then Theorem 3.1 gives us stability in taking the persistence modules of our filtered complexes $\mathcal{R}(X, \mathrm{d}_X, f)$ and $\mathcal{R}(Y, \mathrm{d}_Y, g)$.

Corollary 3.1.2. *Ψ_k is continuous for all $k \in \mathbb{N}$.*

As long as our metrics are sensible, the notion that Ψ_k is continuous means that small changes in the data cloud will result in small changes in the corresponding persistent homology of our space. That is, by varying (X, d_X, f) or (Y, d_Y, g) around in an $\varepsilon-$ball, we can be sure that their corresponding barcodes will also vary in an $\varepsilon-$ball in the space of finite barcodes. The inequality is thus quite a bit stronger than simply stating that Ψ_k is continuous for all k. Perhaps the best thing that comes out of this is that it indicates that this method of analyzing a point cloud is nontrivial, and could even be useful.

Before proceeding with the proof of Theorem 3.1, we include here a brief sketch of its main points. We first let $\varepsilon = \mathrm{d}^1_{\mathrm{GH}}((X, \mathrm{d}_X, f), (Y, \mathrm{d}_Y, g))$, with the intent to show that $\mathrm{d}(\mathcal{R}(X, \mathrm{d}_X, f), \mathcal{R}(Y, \mathrm{d}_Y, g)) \leq \varepsilon$. The first step is to find an isometric embedding of both (X, d_X) and (Y, d_Y) into $\mathbb{R}$ with the l^∞ metric. We do this in order to be able to compare the filtrations $\{\gamma \circ \gamma_X(X_\alpha)^\alpha\}_{\alpha>0}$ and $\{\gamma \circ \gamma_Y(Y_\alpha)^\alpha\}_{\alpha>0}$ which we define in the proof, and show that they are $\varepsilon-$interleaved. The final step is to show, through a string of theorems and lemmas, that the filtrations $\{\gamma \circ \gamma_X(X_\alpha)^\alpha\}_{\alpha>0}$ and $\{\gamma \circ \gamma_Y(Y_\alpha)^\alpha\}_{\alpha>0}$ being $\varepsilon-$interleaved implies that the Rips filtrations $\mathcal{R}(X, \mathrm{d}_X, f)$ and $\mathcal{R}(Y, \mathrm{d}_Y, g)$ are $\varepsilon-$interleaved as well. We proceed now with the proof.

Proof of Theorem 3.1. Let $\varepsilon = \mathrm{d}^1_{\mathrm{GH}}((X, \mathrm{d}_X, f), (Y, \mathrm{d}_Y, g))$ and recall the notation $X_\alpha := f^{-1}(-\infty, \alpha]$ and $Y_\alpha := g^{-1}(-\infty, \alpha]$ for any $\alpha \in \mathbb{R}$. Since both X and Y are finite, the infemum given in the definition of $\mathrm{d}^1_{\mathrm{GH}}$ is actually a minimum, achieved by some $C \in \mathcal{C}(X, Y)$.

We note that in the proof of Theorem 7.3.25 of [BBI01] it shows that for any $C \in \mathcal{C}(X, Y)$, there exists a metric d_Z on the disjoint union $Z = X \sqcup Y$ such that the canonical inclusions $\gamma_X : X \hookrightarrow Z$ and $\gamma_Y : Y \hookrightarrow Z$ are isometric embeddings, and such that the following inequalities hold for all $(x, y) \in C$:

$$\mathrm{d}_Z(\gamma_X(x), \gamma_Y(y)) \leq \frac{1}{2} \sup_{(x,y),(x',y') \in C} |\mathrm{d}_X(x, x') - \mathrm{d}_Y(y, y')| \tag{11}$$

$$|f(x) - g(y)| \leq \|f - g\|_{\ell^\infty(C)} := \sup_{(x,y) \in C} |f(x) - g(y)|. \tag{12}$$

By definition of $\mathrm{d}^1_{\mathrm{GH}}$, (11) and (12) give us that $\mathrm{d}_Z(\gamma_X(x), \gamma_Y(y)), |f(x) - g(y)| \leq \varepsilon$.

Now note that $(\gamma_X(X) \cup \gamma_Y(Y), d_Z)$ is a finite metric space. We will cite Lemma 2.8 of [CCSGMO09] here, which says that any finite metric space of cardinality n can be isometrically embedded into $\mathbb{R}^n$ endowed with the l^∞ metric. Thus if $n = \#X + \#Y$, we have an isometric embedding $\gamma : (\gamma_X(X) \cup \gamma_Y(Y), d_Z) \hookrightarrow (\mathbb{R}^n, l^\infty)$.

This gives us that $\gamma \circ \gamma_X$ and $\gamma \circ \gamma_Y$ are isometric embeddings of (X, d_X) and (Y, d_Y) respectively into $(\mathbb{R}, l^\infty)$. Furthermore, because $\gamma \circ \gamma_X$ and $\gamma \circ \gamma_Y$ are isometries, for all $(x, y) \in C$, we have that

$$\|\gamma \circ \gamma_X(x) - \gamma \circ \gamma_Y(y)\|_\infty \leq \varepsilon.$$

We will denote by $\gamma \circ \gamma_X(X)^\alpha$ the union of the open l^∞ balls of radius α centered at each point of $\gamma \circ \gamma_X(X)$. Our goal is to show that $\{\gamma \circ \gamma_X(X_\alpha)^\alpha\}_{\alpha>0}$ and $\{\gamma \circ \gamma_Y(Y_\alpha)^\alpha\}_{\alpha>0}$ are $\varepsilon-$interleaved.

To do so, suppose $p \in \gamma \circ \gamma_X(X_\alpha)^\alpha$, and note that there must exist some $x \in X$ such that $\|p - \gamma \circ \gamma_X(x)\|_\infty \leq \alpha$ by definition. Then take $y \in Y$ such that $(x, y) \in C$. By our embedding, we have that $\|\gamma \circ \gamma_X(x) - \gamma \circ \gamma_Y(y)\|_\infty \leq \varepsilon$, as well as that $g(y) \leq f(x) + \varepsilon \leq \alpha + \varepsilon$. Thus $y \in Y_{\alpha+\varepsilon}$, and the triangle inequality gives us that

$$\|p - \gamma \circ \gamma_Y(y)\|_\infty = \|p - \gamma \circ \gamma_Y(y) + \gamma \circ \gamma_X(x) - \gamma \circ \gamma_X(x)\|_\infty \leq$$

$$\|\gamma \circ \gamma_X(x) - \gamma \circ \gamma_Y(y)\|_\infty + \|p - \gamma \circ \gamma_X(x)\|_\infty = \alpha + \varepsilon.$$

Thus $p \in \gamma \circ \gamma_Y(Y_{\alpha+\varepsilon})^{\alpha+\varepsilon}$, and we have that $\gamma \circ \gamma_X(X_\alpha)^\alpha \subseteq \gamma \circ \gamma_Y(Y_{\alpha+\varepsilon})^{\alpha+\varepsilon}$. By a symmetric argument, $\gamma \circ \gamma_Y(Y_{\alpha+\varepsilon})^{\alpha+\varepsilon} \subseteq \gamma \circ \gamma_X(X_{\alpha+2\varepsilon})^{\alpha+2\varepsilon}$. Thus we get the inclusions

$$\gamma \circ \gamma_X(X_\alpha)^\alpha \subseteq \gamma \circ \gamma_Y(Y_{\alpha+\varepsilon})^{\alpha+\varepsilon} \subseteq \gamma \circ \gamma_X(X_{\alpha+2\varepsilon})^{\alpha+2\varepsilon} \subseteq \gamma \circ \gamma_Y(Y_{\alpha+3\varepsilon})^{\alpha+3\varepsilon}. \tag{13}$$

With the inclusion maps as our natural transformations ϕ and ψ, (13) shows that they satisfy the two diagrams for an $\varepsilon-$interleaving, as desired, and we have that $\{\gamma \circ \gamma_X(X_\alpha)^\alpha\}_{\alpha>0}$ and $\{\gamma \circ \gamma_Y(Y_\alpha)^\alpha\}_{\alpha>0}$ are $\varepsilon-$interleaved.

As it turns out, by a string of theorems and lemmas, this is sufficient to show that the Rips filtrations $\mathcal{R}(X, \mathrm{d}_X, f)$ and $\mathcal{R}(Y, \mathrm{d}_Y, g)$ are $\varepsilon-$close, as desired. We get by Theorem 4.4 of [CCSGGO09] that, under the bottleneck distance, the persistence diagrams of $\{\gamma \circ \gamma_X(X_\alpha)^\alpha\}_{\alpha>0}$ and $\{\gamma \circ \gamma_Y(Y_\alpha)^\alpha\}_{\alpha>0}$ are $\varepsilon-$close. Lemma 2.10 of [CCSGMO09] gives us the same result for the Čech filtrations $\{C_\alpha(\gamma \circ \gamma_X(X_\alpha), \mathbb{R}^n, l^\infty)\}_{\alpha>0}$ and $\{C_\alpha(\gamma \circ \gamma_Y(Y_\alpha), \mathbb{R}^n, l^\infty)\}_{\alpha>0}$. And Lemma 2.9 of [CCSGMO09] gives the same result for the persistence diagrams of the Rips filtrations $\{\mathcal{R}_\alpha(\gamma \circ \gamma_X(X_\alpha), l^\infty)\}_{\alpha>0}$ and $\{R_\alpha(\gamma \circ \gamma_Y(Y_\alpha), l^\infty)\}_{\alpha>0}$. Since $\gamma \circ \gamma_X$ and $\gamma \circ \gamma_Y$ are isometric embeddings of (X, d_X) and (Y, d_Y) respectively, these are precisely the Rips filtrations $\mathcal{R}(X, \mathrm{d}_X, f)$ and $\mathcal{R}(Y, \mathrm{d}_Y, g)$, and we have our result. □

4 Stability Theorem for Level Sets

We will now step into quite a bit more generality as we consider the stability theorem for level sets. This theorem has quite a different feel to it. It is true that we are comparing the persistence module of different filtrations of a topological space. However, our filtrations are defined differently in this case. Instead of taking the Rips filtration, we will filter a topological space X with a (not necessarily continuous) map $f : X \to \mathbb{R}$ via the filter $X_\varepsilon = f^{-1}(-\infty, \varepsilon]$ for $\varepsilon \in \mathbb{R}$.

This gives us a diagram $F \in \mathrm{Top}^{(\mathbb{R},\leq)}$ of the form $F(a) = X_a$ and $F(a \leq b) = (X_a \hookrightarrow X_b)$, the inclusion map. Just like before, we can calculate the distance between the persistence modules of diagrams F induced by $f : X \to \mathbb{R}$ and G induced by $g : X \to \mathbb{R}$ and see whether or not this process of assigning to a map in $\mathrm{Sets}(X, \mathbb{R})$ to a diagram in $\mathrm{Vec}^{(\mathbb{R},\leq)}$ is continuous. We actually get an equally strong result as we did in Section 3 in the sense of bounding the distance persistence diagrams by the distance of the original maps.

But this case is much better; before we worked strictly in the case that X was a finite metric space and we wanted to measure persistence modules on the Rips filtration of X. This time we can define essentially any filtration we please on a topological space, measure its distance, and then instead of being restricted to only persistence modules, we can bound the distance of the image of our filtrations under *any functor* $\mathrm{Top} \to \mathcal{D}$, for an arbitrary category $\mathcal{D}$. Thus we can apply our stability result to the singular homology functor $H : \mathrm{Top} \to grAb$ with coefficients in any abelian group, homotopy groups, etc. It is probably the most computationally reasonable to use the singular homology functor with coefficients in $\mathbb{F}_2$, but this theorem is in full generality in the best sense. We can truly consider *any* functor H with domain $\mathrm{Top}^{(\mathbb{R},\leq)}$. Regardless of whether or not this generality provides more practical uses of the theory, this is a beautiful result for the theory.

Theorem 4.1. *Consider any $X \in Top$ and let f and g be (not necessarily continuous) maps $X \to \mathbb{R}$. Let $H : Top \to \mathcal{D}$ be any functor, and define $F \in Top^{(\mathbb{R},\leq)}$ to be $F(a) = f^{-1}(-\infty, a]$ and G likewise using g. Then*

$$\mathrm{d}(HF, HG) \leq \|f - g\|_\infty$$

Proof. Let $\varepsilon = \|f - g\|_\infty$. Then because f and g are $\varepsilon-$close,

$$F(a) = f^{-1}(-\infty, a] \subseteq g^{-1}(-\infty, a + \varepsilon] = G(a + \varepsilon)$$

and, likewise, $G(a) \subseteq F(a + \varepsilon)$. Thus F and G are $\varepsilon-$interleaved, giving that HF and GF are as well, and thus giving that $\mathrm{d}(HF, HG) \leq \|f - g\|_\infty$. □

Just like in Section 3 we can define a map for $X \in \mathrm{Top}$ and functor $H : \mathrm{Top} \to \mathcal{D}$:

$$\Phi_X : \mathrm{Sets}(X, \mathbb{R}) \to \mathcal{D}^{(\mathbb{R}, \leq)} \tag{14}$$

which is continuous.

5 Revisiting Stability of Point Clouds

We would like to show in this section how Theorem 3.1 can be thought of as a special case of Theorem 4.1. We will begin by considering a compact metric space X and a continuous map $f : X \to \mathbb{R}$. Note that since X is compact and f continuous, f is bounded. So for α small enough, we have that $\mathcal{R}_\alpha(X_\alpha) = \emptyset$. For α large enough, we have that $\mathcal{R}_\alpha(X_\alpha)$ is a constant topological space X' for all $\beta \geq \alpha$.

Thus our filtration $\mathcal{R}(X, \mathrm{d}_X, f)$ stabilizes as α goes to infinity. Now take X' and define a map $f' : X' \to \mathbb{R}$ which sends an element $x \in X'$ to

$$\inf\{\alpha \in \mathbb{R} \,|\, x \in R_\alpha(X_\alpha)\}.$$

Note that $\{\alpha \in \mathbb{R} \,|\, x \in R_\alpha(X_\alpha)\}$ is bounded below, so this map (which is probably not continuous) is well-defined, and thus defines a filtration on X' given by the level sets $X'_\alpha = (f')^{-1}(-\infty, \alpha]$. Note further that when restricted to $X \subseteq X'$, $f'|_X = f$.

Using the notation from Section 4, we have a diagram $F' \in \mathrm{Top}^{(\mathbb{R}, \leq)}$, induced by f', and given by $F'(a) = (f')^{-1}(-\infty, a]$. Furthermore, $\mathcal{R}(X, \mathrm{d}_X, f) \in \mathrm{Top}^{(\mathbb{R}, \leq)}$ and so we can compare the two filtrations. It is possible to that they might differ at some points because of the definition, but it is easy to see that

$$\mathrm{d}(\mathcal{R}(X, \mathrm{d}_X, f), F') = 0.$$

Now consider $g : X \to \mathbb{R}$. Note that, $\mathcal{R}(X, \mathrm{d}_X, g)$ stabilizes to the same X' as $\mathcal{R}(X, \mathrm{d}_X, f)$ just by the definition of the Rips filtration and so, in the same way as we did before, we can construct $g' : X' \to \mathbb{R}$ and obtain a diagram $G' \in \mathrm{Top}^{(\mathbb{R}, \leq)}$. Then we have that

$$\mathrm{d}(\mathcal{R}(X, \mathrm{d}_X, f), \mathcal{R}(X, \mathrm{d}_X, g)) \leq \mathrm{d}(\mathcal{R}(X, \mathrm{d}_X, f), G') + \mathrm{d}(G', \mathcal{R}(X, \mathrm{d}_X, g))$$

$$= \mathrm{d}(\mathcal{R}(X, \mathrm{d}_X, f), G') \leq \mathrm{d}(\mathcal{R}(X, \mathrm{d}_X, f), F') + \mathrm{d}(F', G') = \mathrm{d}(F', G')$$

and thus by Theorem 4.1,

$$\mathrm{d}(\mathcal{R}(X, \mathrm{d}_X, f), \mathcal{R}(X, \mathrm{d}_X, g)) \leq \mathrm{d}(F', G') \leq \|f' - g'\|_\infty. \tag{15}$$

We would like to say that because (15) gives stability, Theorem 3.1 is a special case of our more general Theorem 4.1. And in a sense this is true for the case we have considered of comparing $\mathcal{R}(X, \mathrm{d}_X, f)$ and $\mathcal{R}(Y, \mathrm{d}_Y, g)$, where $(X, \mathrm{d}_X) = (Y, \mathrm{d}_Y)$. In this case, the essence of Theorem 3.1 is captured by Theorem 4.1.

However, we run into a slight problem here. Since $f'|_X = f$ and $g'|_X = g$, we have that

$$\|f - g\|_\infty \leq \|f' - g'\|_\infty,$$

simply because f' and g' are defined on a larger domain. Recall that on the diagonal correspondence C,

$$\|f - g\|_\infty = \max\{\frac{1}{2} \sup_{(x,x),(y,y)\in C} |d_x(x, x) - d_y(y, y)|; \sup_{(x,x)\in C} |f(x) - g(x)|\}. \tag{16}$$

If we recall that $\mathrm{d}^1_{\mathrm{GH}}$ is defined to be the infemum of such values given in (16), ranging over all correspondences in $\mathcal{C}(X, X)$, we have that

$$\mathrm{d}^1_{\mathrm{GH}}(\mathcal{R}(X, \mathrm{d}_X, f), \mathcal{R}(X, \mathrm{d}_X, g)) \leq \|f - g\|_\infty \leq \|f' - g'\|_\infty.$$

Thus we might compare Theorems 3.1 and 4.1 by saying that Theorem 3.1 puts a lower bound than Theorem 4.1 does on $\mathrm{d}(\mathcal{R}(X, \mathrm{d}_X, f), \mathcal{R}(Y, \mathrm{d}_Y, g))$ in the special case to which Theorem 3.1 applies (i.e. when $(X, \mathrm{d}_X) = (Y, \mathrm{d}_Y)$).

We can even compare our theorems on a broader spectrum, where we compare $\mathcal{R}(X, \mathrm{d}_X, f)$ and $\mathcal{R}(Y, \mathrm{d}_Y, g)$ with $(X, \mathrm{d}_X) \neq (Y, \mathrm{d}_Y)$. We know by the proof of Theorem 3.1 that we can embed both X and Y into $(\mathbb{R}^n, l^\infty)$, where $n = \#X + \#Y$, and we get a point cloud Z in $\mathbb{R}^m$ for some $m \leq n$. Since both X and Y are finite metric spaces, both f and g are bounded above by some N. Choose $M > N$ and then extend f and g to Z by letting both f and g take the value of M on any point of Z which does not lie in the embedding of, respectively, X or Y. This gives us a filtered complex like in the case above, and again we get the stability given by Theorem 3.1 using Theorem 4.1—however, just as before, the bound given in Theorem 3.1 is better than the bound given in Theorem 4.1.

6 Conclusion

In this document, we have introduced notions for distance on diagrams of the form $\mathcal{C}^{(\mathbb{R},\leq)}$, on finite metric spaces endowed with continuous maps, and on maps in $\mathrm{Sets}(X, \mathbb{R})$ for a metric space X. This has allowed us to define notions of stability, and then to introduce two different theorems on stability. The first, Theorem 3.1, shows that for $(X, \mathrm{d}_X, f), (Y, \mathrm{d}_Y, g) \in \mathcal{X}_1$,

$$\mathrm{d}(\mathcal{R}(X, \mathrm{d}_X, f), \mathcal{R}(Y, \mathrm{d}_Y, g)) \leq \mathrm{d}^1_{\mathrm{GH}}((X, \mathrm{d}_X, f), (Y, \mathrm{d}_Y, g)).$$

This in particular shows that the interleaving distance between the k^{th} persistence homology diagrams $H_k(\mathcal{R}(X, \mathrm{d}_X, f), \mathbb{F})$ and $H_k(\mathcal{R}(Y, \mathrm{d}_Y, g), \mathbb{F})$ is bounded above by $\mathrm{d}^1_{\mathrm{GH}}((X, \mathrm{d}_X, f), (Y, \mathrm{d}_Y, g))$. This gives us that taking persistent homology of (X, d_X, f) can be thought of as a continuous operation, meaning that small changes in (X, d_X, f) with regard to the metric $\mathrm{d}^1_{\mathrm{GH}}$ will result in possibly smaller changes in $H_k(\mathcal{R}(X, \mathrm{d}_X, f), \mathbb{F})$ with regard to the interleaving distance on diagrams $\mathrm{Top}^{(\mathbb{R},\leq)}$.

We then showed a more general result, Theorem 4.1, which defines filtrations F (resp. G) on any topological space X endowed with a (not necessarily continuous) map $f : X \to \mathbb{R}$ (resp. $g : X \to \mathbb{R}$) and shows that, for any category $\mathcal{D}$ and functor $H : \mathrm{Top} \to \mathcal{D}$:

$$\mathrm{d}(HF, HG) \leq \|f - g\|_\infty.$$

Again we can consider H as the k^{th} homology functor, H_k, and we get the same notions of continuity that our previous theorem gave us.

These theorems are slightly different, but can be compared as we showed in Section 5. It turns out that Theorem 3.1 is not quite a special case of Theorem 4.1, but instead puts a lower bound on $\mathrm{d}(\mathcal{R}(X, \mathrm{d}_X, f), \mathcal{R}(Y, \mathrm{d}_Y, g))$ the special case that we consider the filtered complexes $\mathcal{R}(X, \mathrm{d}_X, f)$ and $\mathcal{R}(Y, \mathrm{d}_Y, g)$.

References

[BBI01] Burago D., Burago Y., Ivanov S.: *A Course in Metric Geometry,* vol. 33 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, 2001.

[BS14] Peter Bubenik and Jonathan A. Scott. Categorification of Persistent Homology. arXiv:1205.3669v3 [math.AT], 2014.

[CCSGGO09] Chazal F., Cohen-Steiner D., Guibas L. J., Glisse M., Oudot S. Y.: Proximity of persistence modules and their diagrams. In *Proc. 25th ACM Symp. Comp. Geom.* (2009).

[CCSGMO09] Frédéric Chazal, David Cohen-Snider, Leonidas J. Guibas, Facundo Mémoli, and Steve Y. Oudut. Gromov-Hausdorff Stable Signatures for Shapes using Persistence. *Eurographics Symposium on Geometry Processing 2009* 28 (2009), 5.