Reeb graph metrics: a survey

immediate

January 19, 2021

Abstract

We survey the available options for reeb graph metrics

Contents

1 Preliminary Definitions

1.1 Scalar Fields and Reeb Graphs

Assumptions in this section is that we will be dealing with morse functions defined on two-manifolds

Definition 1.1. *A **scalar field** (equivalently an $\mathbb{R}$**-space**) is a pair $(\mathbb{X}, f)$ where $\mathbb{X}$ is topological space and $f : \mathbb{X} \to \mathbb{R}$ is a continuous real-valued function.*

Definition 1.2. *We define an equivalence relation $\sim_f$ on $\mathbb{X}$ by stating that $x \sim_f y$ if $f(x) = f(y) = a$ and x and y both lie in the same connected component of the levelset $f^{-1}(a)$. We define $\mathbb{X}_f$ to be the quotient space $\mathbb{X} \ \sim_f$ and define $\tilde{f} : \mathbb{X}_f \to \mathbb{R}$ to be the restriction of f to the domain $\mathbb{X}_f$. The pair $R(f) := (X_f, \tilde{f})$ is called the **Reeb Graph** of $(\mathbb{X}, f)$.*

Here, we only consider f to be a Morse function and $\mathbb{X}$ to be a two-manifold (surface). In the Reeb graph of a surface, we have three different types of nodes corresponding to three different indices of critical points. A degree 1 node with one edge exiting the node corresponds to an index 0 or minima of the scalar field, a degree 1 node with one edge entering the node corresponds to an index 2 or maxima of the scalar field, and a degree 3 node corresponds to a index 1 or saddle point of the scalar field. We call saddle points with two edges entering the node a **down-fork** and saddle points with two edges exiting the node an **up-fork**.

Provide example of a scalar field with its associated Reeb graph

1.2 Stability

Definition 1.3. *Let $f, g : \mathbb{X} \to \mathbb{R}$ be two real-valued continuous functions defined on the same domain. The L^∞ **distance** between f and g is defined as*

$$||f - g||_\infty := \max_{x \in \mathbb{X}} |f(x) - g(x)|$$

*. If d is a metric between the two Reeb graphs $R(f), R(g)$, we say that d is **stable** if*

$$d(R(f), R(g)) \leq ||f - g||_\infty$$

Stability helps us ensure that these metrics do not compute arbitrarily large values for two Reeb graphs defined on the same domain. However, each of these metrics can be defined on two Reeb graphs that differ in both function *and* the topological space on which they are defined. A metric that is stable does not state any conclusion about these cases.

1.3 (Extended) Persistence

Persistence intuitively captures the length of time that features of a scalar field (or other data sets **CITE**) take to disappear once they have been introduced. The notion of persistence has been widely used as a tool in topological data analysis (**CITE**). The data provided by persistence can be compactly encoded into a **barcode** –where each feature is provided is associated with a horizontal bar corresponding to the length of time that the feature exists in the data – or in a **persistence diagram** – where each feature is a coordinate pair (a, b) with a representing the birth time and b representing the death time.

While a Reeb graph provides a finer grained detail of the topology of a scalar field, the persistence diagram has been proven useful for its computability and the slew of metrics that can be defined on these persistence diagrams which are also computationally tractable (**CITE**).

Let $(\mathbb{X}, f)$ be a scalar field with critical points $\{v_1, \dots, v_k\}$. We define $\mathbb{X}_a := f^{-1}(-\infty, a]$ to be **sublevel set** of $\mathbb{X}$ at a. Now, let $\{b_0, \dots, b_k\}$ be a set of real numbers such that

$$b_0 < f(v_1) < b_1 < f(v_2) < \dots < b_{k-1} < f(v_k) < b_k.$$

This induces a sequence of nested subspaces

$$\emptyset = \mathbb{X}_{b_0} \subset \mathbb{X}_{b_1} \subset \dots \subset \mathbb{X}_{b_{k-1}} \subset \mathbb{X}_{b_k} = \mathbb{X},$$

called a **filtration** of the scalar field (X, f). We can then associate each $\mathbb{X}_{b_i}$ with a corresponding homology group $H_n(\mathbb{X}_{b_i})$ for a fixed dimension n to obtain the sequence

$$\emptyset = H_n(\mathbb{X}_{b_0}) \to H_n(\mathbb{X}_{b_1}) \to \dots \to H_n(\mathbb{X}_{b_{k-1}}) \to H_n(\mathbb{X}_{b_k}),$$

where each arrow between homology group represents the homomorphism $h_n^{i,j} : H_n(\mathbb{X}_{b_i}) \to H_n(\mathbb{X}_{b_j})$ induced by the inclusion $\mathbb{X}_{b_i} \subset \mathbb{X}_{b_j}$. Note that we have chosen these $\{b_0, \dots, b_k\}$ to be specifically interleaved between the critical values of f so that the homology of the sequence changes at every iteration [1].

Definition 1.4. *The* ***n^{th}-persistent homology groups*** *are the images of the hmology group homomorphisms,* $H_n^{i,j} := \mathsf{Im}(h_n^{i,j})$ *and the* ***n^{th}-peristent Betti numbers*** *are their corresponding ranks,* $\beta_n^{i,j} = \mathsf{Rank}(H_n^{i,j})$.

The classes of homology groups intuitively represent various n-dimensional holes in the surface. Classes of H_0 represent connected components and classes of H_1 represent closed loops. The Betti numbers corresponding to a particular topological space then simply count the number of holes. Thus, the -0^{th} Betti numbers tell us the number of connected components, while the 1^{st} Betti numbers tell us the number of loops, and so on.

Definition 1.5. *We say that a class α is* ***born*** *at i if $\alpha \in H_n(\mathbb{X}^i) - \mathsf{Im}(h_n^{i-1,i})$. We say that α* ***dies*** *at j if $h_n^{i,j-1}(\alpha) \notin \mathsf{Im}(h_n^{i-1,j-1}(\alpha))$ but $h_n^{i,j}(\alpha) \in \mathsf{Im}(h_n^{i-1,j}(\alpha))$. The* ***persistence*** *of α is $f(v_j) - f(v_i)$. If α never dies, then we say that the persistence of α is $+\infty$.*

Low persistence features are often attributed to *noise* or *unimportant* features of the data being studied, while high persistence features are often associated with *important* features.

Definition 1.6. *The* ***n^{th}-persistence diagram*** *of f, denoted as $Dgm_n(f)$, is a visualization of its persistent features that records each feature α of the n^{th}-persistent homology groups as a coordinate pair $(f(v_i), f(v_j))$, where α is born in $\mathbb{X}_i$ and dies entering $\mathbb{X}_j$, in the extended plane $\bar{\mathbb{R}}^2 := \mathbb{R} \cup \{+\infty\} \times \mathbb{R} \cup \{+\infty\}$.*

The ordered pairs of a persistence diagram correspond directly to pairs of critical points in a scalar field, which in turn correspond to nodes in the Reeb graph. Index 0 critical points create connected components

[1] The homology groups do not necessarily change, but the only possible places that these sublevelsets have different topologies is when we pass over critical points. Choosing values that are not surrounding the critical points would cause our sequence to have multiple homology groups that are gauranteed to be repeated

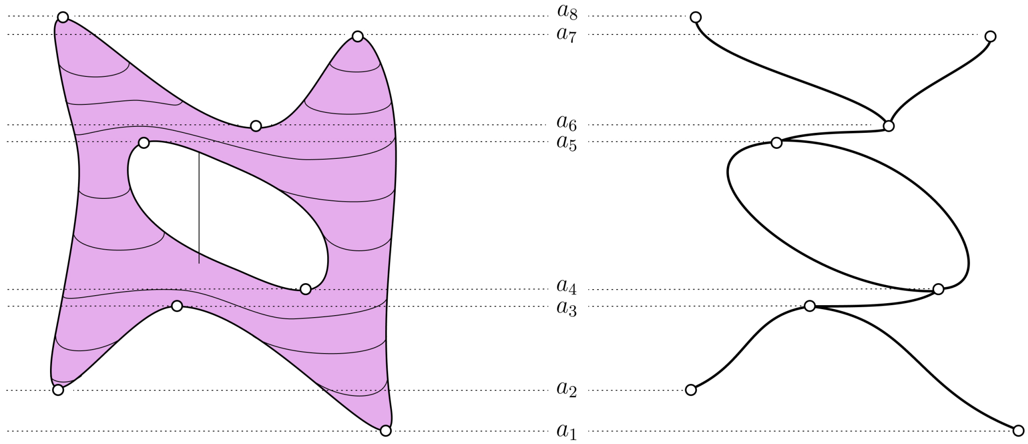

Figure 1: Here is a caption

(0-dimensional homology classes) which are then destroyed by down-forks while up forks create closed loops (1-dimensional homology classes) which are destroyed by index 2 critical points.

However, there are homology classes that are born at some point in the filtration but never die, called **essential** homology classes. Otherwise, the classes are called **inessential**. Critical points which create essential homology classes are present in the Reeb graph, yet they do not correspond to any persistence pair by definition. To alleviate this, we can leverage Poincare and Lefshetz duality to create a new sequence of homology groups where we begin and end with the trivial group [4]. This guarantees that each homology class that is born will also die, meaning each critical point in the Reeb graph will be matched with another critical point at least once.

Let $X^a := f^{-1}[a, \infty)$ be the **superlevel set** of $\mathbb{X}$ at a and let $H_n(\mathbb{X}, \mathbb{X}^a)$ denote the relative homology group. From this, we can create a new sequence of homology groups

$$\begin{aligned} 0 =& H_n(\mathbb{X}_{b_0}) \to H_n(\mathbb{X}_{b_1}) \to \ldots \to H_n(\mathbb{X}_{b_{k-1}}) \to H_n(\mathbb{X}_{b_k}) \\ =& H_n(\mathbb{X}, \mathbb{X}^{b_k}) \to H_n(\mathbb{X}, \mathbb{X}^{b_{k-1}}) \to \ldots \to H_n(\mathbb{X}, \mathbb{X}^{b_1}) \to H_n(\mathbb{X}, \mathbb{X}^{b_0}) = 0. \end{aligned}$$

We can now construct an **extended persistence** diagram from the new pairs in the same fashion as before. To differ between the two types persistence, we will often use the term **ordinary persistence** for the former process. However, we should note that extended persistence is strictly more powerful than ordinary persistence because it captures all inessential homology classes as well as essential homology classes, where as ordinary persistence will only capture inessential classes. We will soon define metrics on these persistence diagrams which will then compare with the metrics defined directly on Reeb graphs. It is natural for us to consider only the extended persistence diagrams rather than ordinary persistence since this most closely relates to the Reeb graph.

Perliminary results in [1] showed a pairing between all critical points of a 2-manifold while [4] extended this to general manifolds later. We provide a standard example of extended persistence and the process behind it to give the reader some intuition.

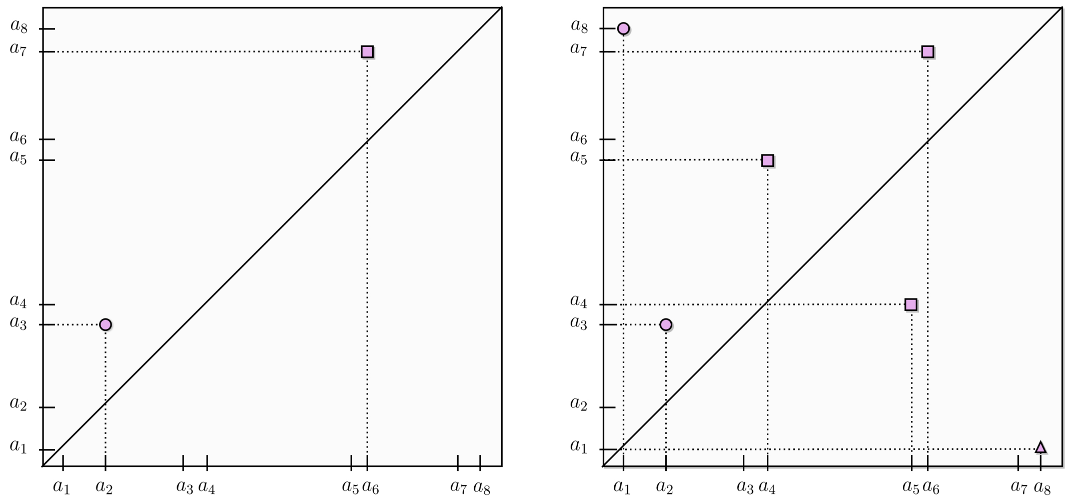

Figure 2: The ordinary persistence and extended persistence diagrams for the surface in Figure 1. Circles denote H_0 classes, squares denote H_1 classes, and triangles denote H_2 classes.

Example 1.7. *Figure 1 shows a genus-1 surface embedded in $\mathbb{R}^3$ along with its Reeb graph. Both a_1 and a_2 create classes of H_0, a_4, a_5 and a_6 create classes in H_1, and a_8 creates a class in H_2. We pair a_2 with a_3 since a_3 is a down-fork which destroys that connected component, and we pair a_6 with a_7 since a_7 is a maximum which destroys the closed loop that a_6 created. However, we have yet to pair a_1,a_4,a_5, and a_8. Continuing in the sequence, we now have begin with a superlevel set which will eventually grow to the entire shape. We first encounter a_8 which destroys the H_0 class created by a_1, a_7 creates an H_1 class which is destroyed by a_6, a_5 destroys the H_1 class created by a_4 going upwards, a_4 destroys the H_1 class that was created by a_5 going upwards, a_3 gives birth to a class in H_2 that is destroyed by a_2, and finally a_1 destroys the class in H_2 that was created by a_8. Figure 2 shows the ordinary persistence and extended persistence diagrams for this surface.*

2 Interleaving Distance

2.1 Definition

A **pre-cosheaf** is a functor F from the category of intervals on the real line **Int** to the category of sets **Set** [2]. Intuitively, it is a way to assign data to the open intervals of the real line in a way that respects inclusion of the intervals. We can construct a pre-cosheaf F from a given Reeb graph $(\mathbb{X}, f)$ by the formulas

$$\mathsf{F}(I) = \pi_0(f^{-1}(I)), \qquad \mathsf{F}[I \subseteq J] = \pi_0[f^{-1}(I) \subseteq f^{-1}(J)],$$

where $\pi_0(U)$ is the set of path connected components of the set U. Stating that two Reeb graphs $(\mathbb{X}, f), (\mathbb{Y}, g)$ are isomorphic is equivalent to stating that their associated pre-cosheafs F, G are isomorphic. Recall that an isomoprhism between two functors is a pair of natural transformations $\varphi : \mathsf{F} \Rightarrow \mathsf{G}$, $\psi : \mathsf{G} \Rightarrow \mathsf{F}$ such that

[2] In general, we can define pre-cosheafs where the domain category is the open sets of any topological space and the range category is unrestricted.

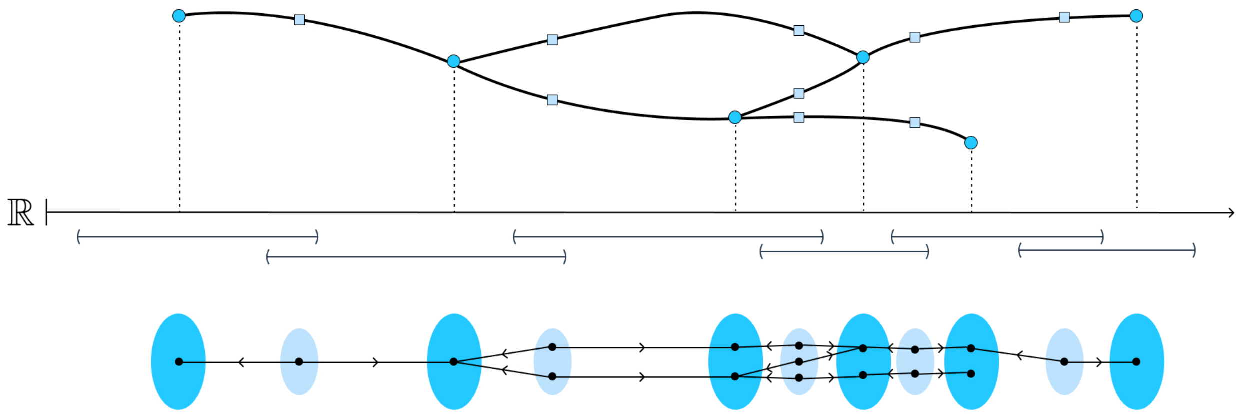

Figure 3: Caption

$\psi_I \circ \varphi_I = \mathbf{Id}_{\mathsf{F}(I)}$ and $\varphi_I \circ \psi_I = \mathbf{Id}_{\mathsf{G}(I)}$, for all $I \in \mathbf{Int}$. If we do not have a true isomorphism between these pre-cosheafs, we can approximate the isomorphisms to form an ε-**interleaving**.

Definition 2.1. *Let $I = (a, b) \subseteq \mathbb{R}$ and $I^\varepsilon = (a - \varepsilon, b + \varepsilon)$. The ε-**expansion functor**, Ω_ε, where $\varepsilon > 0$, is defined as $\Omega_\varepsilon(I) = I^\varepsilon$.*

Explain smoothing here and even refer to the figure maybe

Definition 2.2. *Let $\sigma_{\mathsf{F}}^{\varepsilon}$ be the natural transformation $\mathsf{F} \Rightarrow \mathsf{F}\Omega_\varepsilon$ created by noting that $I \subseteq I^\varepsilon$ implies $\mathsf{F}(I) \to \mathsf{F}(I^\varepsilon) = \mathsf{F}\Omega_\varepsilon(I)$. We say that two pre-cosheafs F, G are ε-**interleaved** if there exists a pair of natural transformations $\varphi : \mathsf{F} \Rightarrow \mathsf{G}\Omega_\varepsilon$ and $\psi : \mathsf{G} \Rightarrow \mathsf{F}\Omega_\varepsilon$ such that the following diagrams commute:*

$$\begin{tikzcd}
\mathsf{F} \arrow[dd, Rightarrow, "\sigma_{\mathsf{F}}^{2\varepsilon}"'] \arrow[dr, Rightarrow, "\varphi"] & \\
& \mathsf{G}\Omega_\varepsilon \arrow[dl, Rightarrow, "\psi\Omega_\varepsilon"] \\
\mathsf{F}\Omega_{2\varepsilon} &
\end{tikzcd}
\qquad\qquad
\begin{tikzcd}
& \mathsf{G} \arrow[dd, Rightarrow, "\sigma_{\mathsf{G}}^{2\varepsilon}"] \arrow[dl, Rightarrow, "\psi"'] \\
\mathsf{F}\Omega_\varepsilon \arrow[dr, Rightarrow, "\varphi\Omega_\varepsilon"'] & \\
& \mathsf{G}\Omega_{2\varepsilon}
\end{tikzcd}$$

If $\varepsilon = 0$, then this is exactly the definition of an isomorphism between F and G. When two pre-cosheafs are ε-interleaved, we say that there exists an ε-**interleaving** between them.

The pre-cosheafs $\mathsf{F}\Omega_\varepsilon, \mathsf{G}\Omega_\varepsilon$ are reffered to as the ε-**smoothed** pre-cosheafs of F, G. The term comes from the fact that increasing the minimum size of the intervals for which we assign data to means that the data is seen from a more coarse view. More specifically in the Reeb graph case, the larger the interval is, the

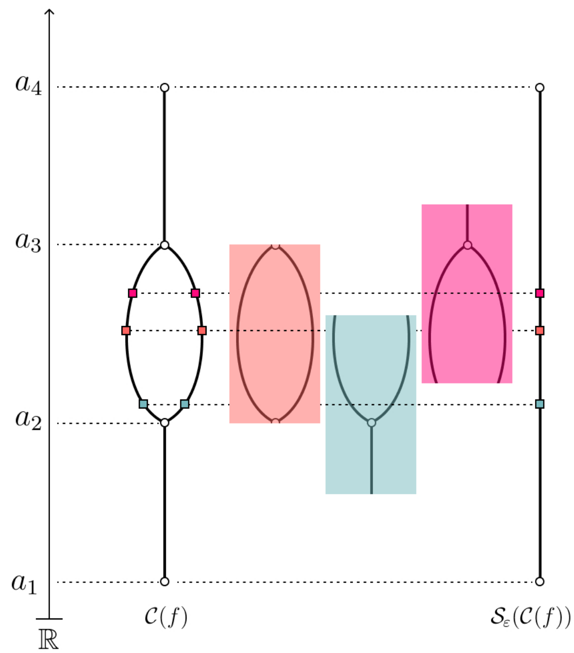

Figure 4: A Reeb graph of a torus along with its ε-smoothed version. To cover the hole completely, ε has to be large enough so that every interval I, the expanded interval I^ε will only have one path connected component. Setting $\varepsilon \geq \frac{a_3-a_2}{}$ will guarantee this.

larger the pre-image of f, g are. Increasing this pre-image will only allow for more chances of the set being path-connected, essentially reducing the number of path-connected components on this interval.

To smooth a Reeb graph $f = (\mathbb{X}, f)$ by ε, we can start at the bottom of the Reeb graph and track how the path connected components change as we look at intervals of the form $(a-\varepsilon, a+\varepsilon)$, where we vary a across all values in the range of f. Since we are considering Reeb graphs with a finite number of nodes and edges, there are only a finite amount of points where the connected components can change.

2.2 Properties and Examples

Suppose we have two Reeb graphs $f = (\mathbb{X}, f)$ and $g = (\mathbb{X}, g)$. Finding the interleaving distance between them can be boiled down into certain situations that we need to avoid. Consider three sets $A = \{a_1, a_2\}$, $B = \{b_1\}$, and $C = \{c_1, c_2\}$. Now, suppose there exsists a map $\gamma : A \to C$ such that $a_1 \mapsto c_1$ and $a_2 \mapsto c_2$. In this case, there does *not* exist maps $\alpha : A \to B$ and $\beta : B \to C$ such that $\beta \circ \alpha = \gamma$ since either $\beta(b_1) = c_1$ or $\beta(b_1) = c_2$, but not both.

Let $\sigma^{2e}_{\mathsf{F},a}$ be the map from $\pi_0(f^{-1}(a))$ to $\pi_0(f^{-1}(a-2\varepsilon, a+2\varepsilon))$, for some $a \in \mathbb{R}$. From the statement above, if $|\mathrm{Im}(\sigma^{2e}_{\mathsf{F},a})| > |\pi_0(g^{-1}(a-\varepsilon, a+\varepsilon)|$, then f, g are cannot be ε-interleaved.

- All loops of size $\varepsilon/2$ will be removed in the ε smoothing
- If $\beta_0(\mathbb{X}) \neq \beta_0(\mathbb{Y})$, then the interleaving distance between any Reeb graphs defined on these spaces is ∞.

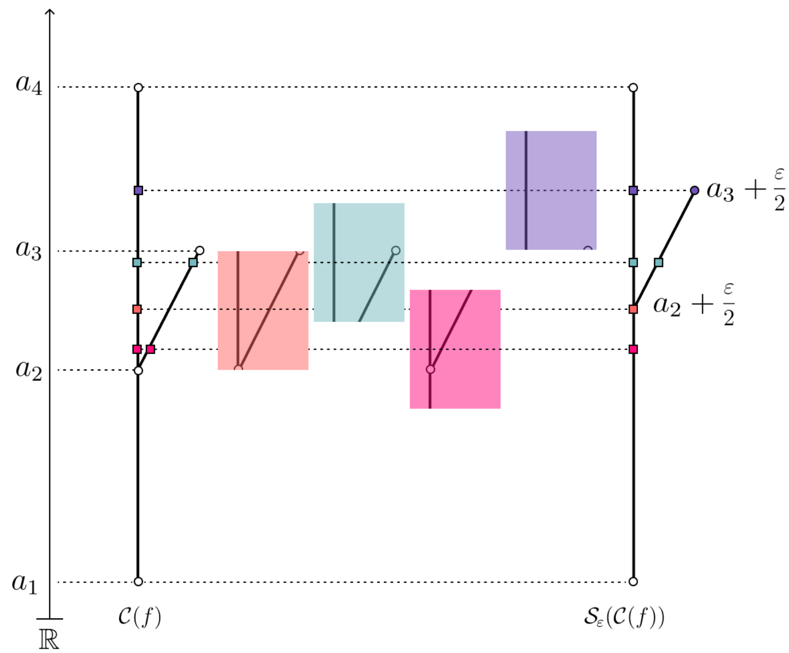

Figure 5: Leaves of a Reeb graph will be shifted by $\frac{\varepsilon}{2}$. As the center of the interval passes $\frac{a_3+a_2}{2}$, the number of components changes from one to two, essentially creating a leaf in the smoothed version that is shifted upwards. Note that the last component (purple) maps to only *one* component in the smoothed Reeb graph.

- Show interleaving example for leaves. Although shifting doesn't remove the leaf, until its shifted past the highest value, shifting it completely past the original spot of the leaf (so stem of new leaf higher than tip of original), will be able to make it map to a single component.
- Perhaps provide upper bound stuff?

We can list the various properties of interleaving distance in this section instead and then provide examples for each of them. Furthermore, we can list the stability results here.

The example should show the various properties of interleaving distance such as robustness to reflections, removing of holes needs certain values (dont remember the fraction), and removal of leaves

General example of interleaving distance (simply connected and non-simply connected domain?)

Example of two scalar fields with high interleaving distance but only due to one outlier

3 Functional Distortion Distance

3.1 History

3.2 Definition

3.3 Properties

4 Edit Distance

4.1 History

4.2 Definition

4.3 Properties

5 Comparison of Distances

Here, we only consider regular and extended persistence diagrams. I assume that we will try and make it so that all the relationships are in terms of extended persistence (which might not be difficult).

Dependencies:

1. Definitions of Edit, Functional Distortion, and Interleaving Distance.
2. Bottleneck distance for 0-dimensional persistence diagrams
3. Bottleneck distance for 1-Dimensional extended persistence Diagrams. Extended persistence diagrams were first introduced in [4]. The motivation behind these diagrams is that there are examples of morse functions where some critical points are never matched. A simple example is a torus with a height function (with orientation so that its actually Morse). The first critical point is the minima of the torus and the last critical point is its maxima. When we normally pair critical points in for persistence, the index of the critical point always differs by 1. By this definition, the minima and maximum would never get paired with each other, and the class born at the minima is considered an *essential* homology class. Extended persistence finds a way to pair this minima and maxima together.

Facts:

1. In [5], we are shown that the Reeb graph edit distance is greater than the functional distortion distance.
2. In [5], we are shown that the Reeb graph Edit distance is greater than the bottleneck distance.
3. In [3], we are shown that the Interleaving Distance and Functional Distortion Distance are strongly equivalent:
$$d_I(f,g) \leq d_{FD}(f,g) \leq 3d_I(f,g)$$

4. In [3] we are shown the relationship between interleaving distance on Reeb graphs compared to the bottleneck distance on persistence diagrams:

$$d_B(\mathrm{Dg}_0(f), \mathrm{Dg}_0(g)) \leq 3d_I(f, g)$$

$$d_B(\mathrm{ExDg}_1(f), \mathrm{ExDg}_1(g)) \leq 9d_I(f, g)$$

5. In [2] we are shown the relationship between functional distortion distance on Reeb graphs and bottleneck distance on persistence Diagrams:

$$d_B(\mathrm{Dg}_0(f), \mathrm{Dg}_0(g)) \leq d_{FD}(f, g)$$

$$d_B(\mathrm{ExDg}_1(f), \mathrm{ExDg}_1(g)) \leq 3d_{FD}(f, g)$$

6 Computation

References

[1] Pankaj Agarwal, Herbert Edelsbrunner, John Harer, and Yusu Wang. "Extreme elevation on a 2-manifold". In: Jan. 2004, pp. 357–365. DOI: 10.1145/997817.997871.

[2] Ulrich Bauer, Xiaoyin Ge, and Yusu Wang. "Measuring Distance between Reeb Graphs". In: *Annual Symposium on Computational Geometry - SOCG'14*. ACM Press, 2014. DOI: 10.1145/2582112.2582169.

[3] Ulrich Bauer, Elizabeth Munch, and Yusu Wang. "Strong Equivalence of the Interleaving and Functional Distortion Metrics for Reeb Graphs". In: *31st International Symposium on Computational Geometry (SoCG 2015)*. Ed. by Lars Arge and János Pach. Vol. 34. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2015, pp. 461–475. ISBN: 978-3-939897-83-5. DOI: http://dx.doi.org/10.4230/LIPIcs.SOCG.2015.461.

[4] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. "Extending Persistence Using Poincaré and Lefschetz Duality". In: *Foundations of Computational Mathematics* 9 (Feb. 2009), pp. 79–103. DOI: 10.1007/s10208-008-9027-z.

[5] Barbara Di Fabio and Claudia Landi. "The Edit Distance for Reeb Graphs of Surfaces". In: *Discrete & Computational Geometry* 55.2 (Jan. 2016), pp. 423–461. DOI: 10.1007/s00454-016-9758-6.