Measuring Similarities in Data using Reeb Graphs and the Interleaving Distance

Brian Bollen
Professor Josh Levine

Topological Data Analysis

The study of data using techniques from algebraic topology. These techniques are often characterized by being able to extract the "shape" of the data and ignore potential noise.

The Canonical Structures:

Persistence Diagrams and Reeb Graphs

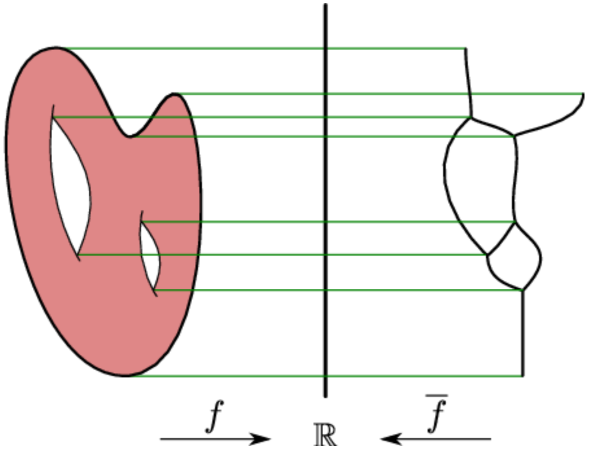
Persistent Homology is the study of how the homology groups of a simplicial complex $\mathcal{K}$ changes as a parameter ρ is changed overtime. Persistence diagrams track this evolution.

The Reeb Graph tracks the evolution of scalar field as we sweep upwards, observing how the connected components of level sets change as we advance.

Scalar Fields and Reeb Graphs

Let $\mathbb{X}$ be a topological space and $f : \mathbb{X} \to \mathbb{R}$ be a scalar function. We call the pair $(\mathbb{X}, f)$ a scalar field. Define an equivalence relation on $\mathbb{X}$ by stating that $x \sim y$ if and only if $f(x) = f(y) = a$ and x, y are in the same path connected component of $f^{-1}(a)$. We call $f^{-1}(a)$ the **level-set** of a or the a-**fiber**.

We then define a quotient space on $\mathbb{X}$ using this equivalence relation and denote it as $\bar{\mathbb{X}}$. **The Reeb Graph** is the scalar field $(\bar{\mathbb{X}}, \bar{f}\,)$ where $\bar{f}$ is the function inherited from f.



Categorifying Reeb Graphs

Background

A **category** $\mathbf{C}$ is a collection of objects. Maps between objects in a category are called **morphisms**. Often, morphisms are defined in such a way as to "preserve structure" such as a homomorphism for groups or isomorphisms for vector spaces.

A **functor** is a mapping $\mathsf{F} : \mathbf{B} \to \mathbf{C}$ is a mapping between two categories such that for each object $B \in \mathbf{B}$ there is a corresponding object $\mathsf{F}(B) \in \mathbf{C}$ and for every morphism α in $\mathbf{B}$, where is a corresponding morphism $\mathsf{F}[\alpha]$ in $\mathbf{C}$. Furthermore, a functor must respect identities and composition.

Category of Scalar Fields and Reeb Graphs

We consider can consider the the space of scalar fields to be a category, called $\mathbb{R} - \mathbf{Top}$ where morphisms between the scalar fields are function preserving, continuous maps.

We can then think of the category of Reeb Graphs denoted $\mathbf{Reeb}$ as being the subcategory of $\mathbb{R}$-$\mathbf{Top}$. There exists a functor $\mathcal{R}$ that maps from the category of scalar fields to the reeb graphs by taking the reeb graph of each scalar field.

Pre-Cosheafs

A **pre-cosheaf** is a functor $\mathsf{F}\colon \mathbf{Int} \to \mathbf{Set}$ where $\mathbf{Int}$ is the category of intervals of $\mathbb{R}$ and $\mathbf{Set}$ is the category of sets. The morphisms in both these categories are inclusion.

We can think of a pre-cosheaf as a way to identify each interval with a specific set. We relate this to Reeb Graphs by stating that $\mathsf{F}(I) := \pi_0(f^{-1}(I))$, where π_0 represents taking the path connected components of the space.

ε-compatible maps

Let $I = (a, b)$ and then denote $I^\varepsilon = (a - \varepsilon, b + \varepsilon)$. We say that **F**,**G** are epsilon compatible if there exsists families of maps φ_ε : **F**(I) $\to$**G**(I^ε) and ψ_ε : **G**(I) $\to$ **F**(I^ε) such that the following diagrams commute:

$$
\begin{tikzcd}
\mathsf{F}(I) \arrow[rr, "\mathsf{F}\left[I \subseteq I^{2\varepsilon}\right]"] \arrow[dr, "\varphi^{\varepsilon}"'] & & \mathsf{F}\left(I^{2\varepsilon}\right) \\
& \mathsf{G}\left(I^{\varepsilon}\right) \arrow[ur, "\psi^{2\varepsilon}"'] &
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathsf{G}(I) \arrow[rr, "\mathsf{G}\left[I \subseteq I^{2\varepsilon}\right]"] \arrow[dr, "\psi^{\varepsilon}"'] & & \mathsf{G}\left(I^{2\varepsilon}\right) \\
& \mathsf{F}\left(I^{\varepsilon}\right) \arrow[ur, "\varphi^{2\varepsilon}"'] &
\end{tikzcd}
$$

If $\varepsilon = 0$, this is precisley an isomorphism between the two functors. If there exists an ε such that the above statements are true, we say that the two pre-cosheafs are ε-**interleaved**.

Interleaving Distance

The interleaving distance between two pre-cosheafs **F**,**G** is the minimum ε such that the two pre-cosheafs are ε-interleaved. That is,

$$d_I(\mathbf{F},\mathbf{G}) = \inf\{\varepsilon \mid \text{there exists an } \varepsilon\text{-compatible map between } \mathbf{F} \text{ and } \mathbf{G}\}$$

How much do we have to squint until there is an isomorphism between them?

Interleaving of Reeb Graphs

The Thickening Functor

For $\varepsilon \geq 0$, can define the **thickening functor** $\mathcal{T}_\varepsilon : \mathbb{R}\text{-}\mathbf{Top} \to \mathbb{R}\text{-}\mathbf{Top}$ as follows:

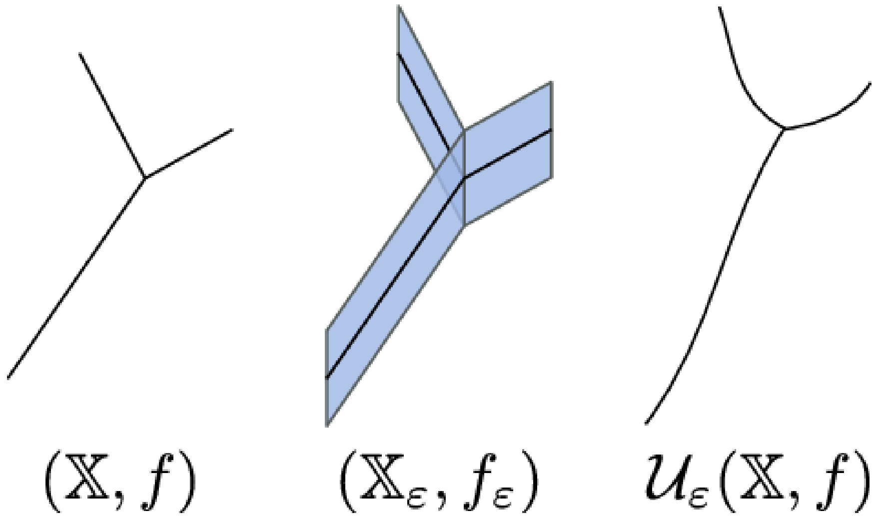
- $\mathcal{T}_\varepsilon(\mathbb{X}, f) = (\mathbb{X}_\varepsilon, f_\varepsilon)$ where $\mathbb{X}_\varepsilon = \mathbb{X} \times [-\varepsilon, \varepsilon]$ and $f_\varepsilon(x, t) = f(x) + t$
- For a morphism $\alpha : (\mathbb{X}, f) \to (\mathbb{Y}, g)$ we have $\mathcal{T}_\varepsilon[\alpha] : (\mathbb{X}_\varepsilon, f_\varepsilon) \to (\mathbb{Y}_\varepsilon, g_\varepsilon)$ where $(x, t) \mapsto (\alpha(x), t)$

It turns out that thickening a Reeb Graph by ε and then looking at its pre-cosheafs is equivalent to looking at ε-smoothed" pre-cosheafs.

Interleaving of Reeb Graphs

Geometric Realization

Since thickening a Reeb Graph and viewing the cosheafs is the same as smoothing the cosheaf itself, we can picture what finding ε-compatible maps looks like by thickening a Reeb Graph and then looking at the Reeb Graph of the newly created scalar field



Computation

Usefulness in Mapper Graphs (?)

Future Work