PyTDA

Kirk P. Gardner[*]   Donald R. Sheehy[†]

1 Sensor Networks and Simplicial Complexes

Suppose we would like to learn about the structure of some unknown domain. Computational techniques are naturally suited to discrete domains, such as a collection of functions on a grid. One such example is image data in which each image is simply a function on a grid, where each function value is a pixel in the image. More often the domain in question is not discrete but continuous. In this case we need some method of discretizing the space in a way that faithfully captures its structure. This is often achieved by drawing a collection of sample points from the space and studying the relationships between them. If the sample is representative of the domain and their relationships are induced by a measure on the domain itself we may model the structure of the domain as a graph consisting of vertices for each sample point and edges reflecting their pairwise relationships.

Given an unknown space $\mathcal{D}$ imbued with a metric $\mathrm{d} : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ and a set of sample points $P \subset \mathcal{D}$ we may construct a graph with edges between each pair of points within some distance $\varepsilon > 0$. This particular construction is the traditional model for a sensor network in which each node can communicate only with nodes within distance ε. A graph of this network can provide information regarding the connectivity of the nodes, however more information is required if we are interested in the domain itself.

Suppose we have a collection of points $P \subset \mathcal{D}$ with **coverage radius** $\varepsilon > 0$. We define the **coverage region** of a node $p \in P$ as the set of points in $\mathcal{D}$ within distance ε of p, formally

$$\mathrm{ball}_\varepsilon(p) := \{x \in \mathcal{D} \mid \mathrm{d}(p, x) \leq \varepsilon\},$$

and define the ε-offset of P, denoted P^ε, as

$$P^\varepsilon := \bigcup_{p \in P} \mathrm{ball}_\varepsilon(p).$$

A susbet $D \subseteq \mathcal{D}$ is **covered** by P if $D \subset P^\varepsilon$ - each point $x \in D$ is within distance ε of at least one point in P. While a graph is not sufficient for determining if a

[*]University of Connecticut.
[†]University of Connecticut.

domain is covered a more robust structure known as a simplicial complex is. A **simplicial complex** K is a collection of subsets, called **simplices**, of a vertex set V that is closed under taking subsets. That is, for all $\sigma \in K$ and $\tau \subset \sigma$ it must follow that $\tau \in K$. The **dimension** of a simplex $\sigma \in K$ is defined as $\dim(\sigma) := |\sigma| - 1$ where $|\cdot|$ denotes set cardinality. The dimension of a simplicial complex K is the maximum dimension of any simplex in K.

There is a specific simplicial complex defined for a collection of sets that precisely captures the coverage information we require. The **Čech complex** of a finite collection of points P at scale $\varepsilon > 0$ is defined

$$\check{\mathcal{C}}^{\varepsilon}(P) := \left\{ \sigma \subseteq P \mid \bigcap_{p \in \sigma} \mathrm{ball}_{\varepsilon}(p) \neq \emptyset \right\}.$$

The Čech complex is a special case of a more general construction known as a the **nerve** $\mathbb{N}(\mathcal{U})$ of a collection of sets $\mathcal{U} = \{U_i\}_{i \in I}$, where I is any indexing set. The nerve of $\mathcal{U}$ is defined as the simplicial complex with vertex set I such that $\sigma \subseteq I$ is a simplex if and only if

$$\bigcap_{i \in \sigma} U_i \neq \emptyset.$$

The collection $\mathcal{U}$ is a **good cover** if for each $\sigma \subset I$ the set $\bigcap_{i \in \sigma} U_i$ is contractible if it is nonempty. The **nerve lemma** states that if $\mathcal{U}$ is a good cover then its nerve $\mathbb{N}(\mathcal{U})$ is homotopy equivalent to $\bigcup_{i \in I} U_i$. That is, for a set of nodes $P \subset \mathcal{D}$ such that $\mathcal{U} = \{\mathrm{ball}_{\varepsilon}(p) \mid p \in P\}$ is a good cover the nerve $\mathbb{N}(\mathcal{U})$ is homotopy equivalent to $P^{\varepsilon} = \bigcup_{p \in P} \mathrm{ball}_{\varepsilon}(p)$. It follows that the Čech complex $\check{\mathcal{C}}_{\varepsilon}(P)$ of P at scale ε is a suitable representation of the coverage region P^{ε}.

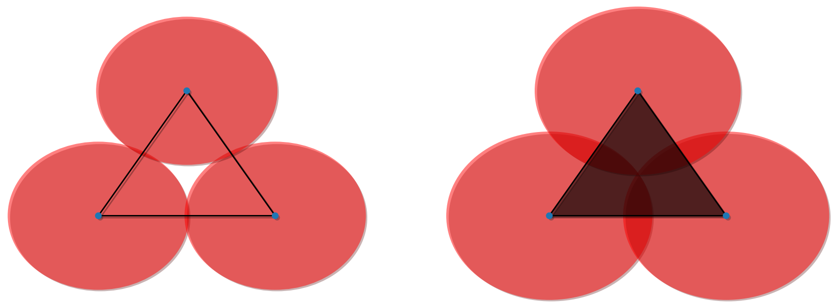

Figure 1: Čech complexes of three points at scales $\varepsilon, \varepsilon'$.

While the Čech complex captures the topology of the coverage region in question exactly it can only be computed when the precise pairwise distances between nodes is known. If instead we are only provided pairwise *proximity* information indicating when nodes are within some fixed distance we may use the **(Vietoris-)Rips complex**, defined for a set P at scale $\varepsilon > 0$ as

$$\mathcal{R}^\varepsilon(P) := \{\sigma \subseteq P \mid \forall p, q \in \sigma,\ \mathrm{d}(p,q) \leq \varepsilon\}\,.$$

An important result about the relationship of Čech and Rips complexes follows from Jung's Theorem [2] relating the diameter of a point set P and the radius of the minimum enclosing ball:

$$\check{\mathcal{C}}^\varepsilon(P) \subseteq \mathcal{R}^\varepsilon(P) \subseteq \check{\mathcal{C}}^{\vartheta_d \varepsilon}(P), \tag{1}$$

where the constant $\vartheta_d = \sqrt{\frac{2d}{d+1}}$ (see [1]).

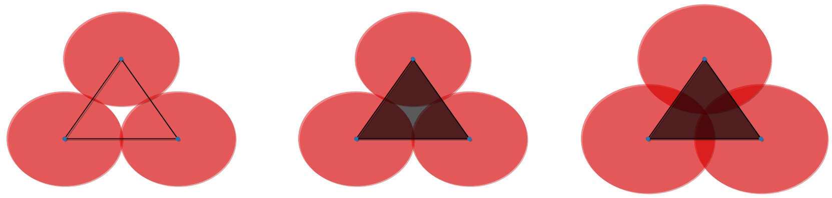

Figure 2: The Rips-Čech interleaving $\check{\mathcal{C}}^\varepsilon(P) \subseteq \mathcal{R}^\varepsilon(P) \subseteq \check{\mathcal{C}}^{\vartheta_d \varepsilon}(P)$

As we will see the Rips complex may be used to verify coverage of a domain satisfying some minimal assumptions when we allow the sensors in our network to communicate at two radii α and β. In short, if the the inclusion of Rips complexes at scales $\alpha < \beta$ resembles the structure of a subset of the domain then the Čech complex at scale α, and therefore P^α, does as well.

2 Functions on Complexes

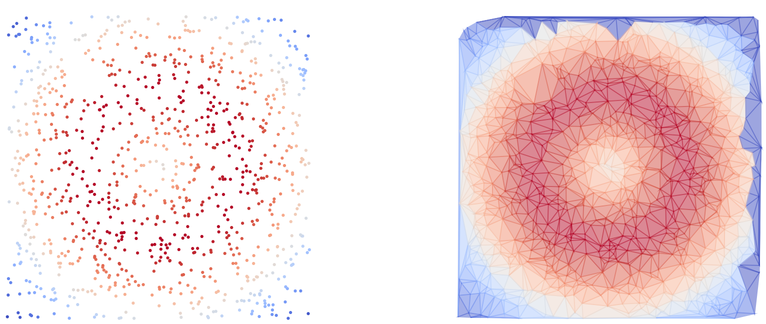

Figure 3: (Middle) Function values on random sample. (Right) Function values on the 2-simplices of a simplicial complex on the random sample.

3 Homology and Cohomology

Simplicial complexes not only provide a comprehensive discretization of continuous domains, but are the primary tool for concrete calculations in algebratic topology. In particular, the study of simplicial homology groups and its dual cohomology groups rely on simplicial complexes in order to study important topological invariants of a discretized space.

3.1 Homology

The following vector spaces may be defined over any field $\mathbb{F}$, however we will assume the field $\mathbb{F}_2$ in order to avoid orienting the simplices in K. Let $C_k(K)$ denote the vector space over some field $\mathbb{F}$ consisting of linear combinations of k-simplices in K, which form a basis for $C_k(K)$, known as **k-chains**. These vector spaces are connected by **boundary maps** $\partial_k : C_k(K) \to C_{k+1}(K)$ which are linear transformations taking basis elements of $C_k(K)$ to the abstract sum of basis $(k+1)$-simplex faces. The collection of chains and boundary maps forms a sequence of vector spaces known as the **chain complex** $\mathcal{C} = (C_*, \partial_*)$

$$\ldots \xrightarrow{\partial_{k+1}} C_k(K) \xrightarrow{\partial_k} C_{k-1}(K) \xrightarrow{\partial_{k-1}} \ldots \xrightarrow{\partial_2} C_1(K) \xrightarrow{\partial_1} C_0(K) \xrightarrow{\partial_0} 0.$$

An important property of the boundary maps ∂_k is that the composition of subsequent boundary maps is null. That is, for all k

$$\partial_k \circ \partial_{k-1} = 0.$$

As a result the image of ∂_{k+1}, denoted $\text{im}\,\partial_{k+1} = \{\partial_{k+1}c \mid c \in C_{k+1}(K)$ is a subspace of the kernel, $\ker \partial_k = \{c \in C_k(K) \mid \partial_k c = 0\}$, of ∂_k. A **k-cycle** of $\mathcal{C}$ is a k-chain with empty boundary - an element of $\ker \partial_k$. Two cycles in $\ker \partial_k$ are said to be **homologous** if they differ by an element of $\text{im}\,\partial_{k+1}$. This leads us to the definition of the **homology groups** of K as quotient vector spaces $H_k(K)$ over $\mathbb{F}$, defined for $k \in \mathbb{N}$ as

$$H_k(K) := \ker \partial_k / \text{im}\,\partial_{k+1}.$$

3.2 Cohomology

A **cochain complex** is a sequence $\mathcal{C} = (C^*, \delta_*)$ of $\mathbb{R}$-modules C^k consisting of **cochains** and module homomorphisms known as **coboundary maps** $\delta_k : C^k \to C^{k+1}$. As in homology we have the property that $\delta_{k+1} \circ \delta_k = 0$ for all k, leading to a familiar definition of the **cohomology** of $\mathcal{C}$

$$H^k(\mathcal{C}) = \ker \delta_k / \text{im}\,\delta_{k-1}.$$

The equivalence classes of $H^k(\mathcal{C})$ consist of **k-cocycles**: elements of $\ker \delta_k$ that differ by a **k-coboundary** in $\text{im}\,\partial_{k-1}$. Such cocycles are said to be **cohomologous** if they belong to the same equivalence class in $H^k(\mathcal{C})$.

The simplest construction of a cochain complex is to dualize a chain complex. For a simplicial complex K with chain complex (C_*, ∂_*) define $C^k(K)$ to be the module of homomorphisms $\psi : C_k \to \mathbb{R}$. The coboundary maps δ_k are defined for cochains $\psi : C_k \to \mathbb{R}$ and k-simplices $\sigma \in K$ as

$$\delta_k \psi(\sigma) = \psi(\partial_k \sigma).$$

3.3 Relative Homology and Cohomology

3.4 Representative Cycles and Cocycles

We can find a basis for each homology group $H_p(K)$ and cohomology group $H^p(K)$ consisting of p-cycles and p-cocycles, respectively. In general, p-cycles represent p-dimensional holes in the simplicial complex K, where p-cocycles can be understood as "blocking chains."

3.5 Persistent Homology and Cohomology

References

[1] Mickaël Buchet, Frédéric Chazal, Steve Y. Oudot, and Donald R. Sheehy. Efficient and robust persistent homology for measures. In *ACM-SIAM Symposium on Discrete Algorithms*, pages 168–180, 2015.

[2] Heinrich Jung. Über die kleinste kugel, die eine räumliche figur einschließt. *J. Reine Angew. Math*, 123:214–257, 1901.