Certified Homology Inference

Nicholas J. Cavanna* Kirk Gardner† Donald R. Sheehy‡

1 Introduction

The goal of homology inference is to compute a space's shape from a point cloud sampled near it. Given such a sample, one may want to know when we can reliably infer the homology of the space in question. Naturally, this requires making assumptions on the sample as well as the underlying space.

Niyogi, Smale, and Weinberger showed that one can infer the homology of a smooth manifold from finite points chosen uniformly at random near its surface [5]. Chazal and Lieutier relaxed this to include non-smooth bounded spaces in $\mathbb{R}^d$ via the so-called weak feature size [1]. Both assume that there is a sampled point within ε of every point in the space in their sample. In their work on sensor networks, de Silva and Ghrist give sampling conditions for checking coverage of a shrunken version of a space assuming one can compute the distance from points to the boundary [3].

We show how these approaches can be combined in order to provide a computable inference of the homology of domain from a coordinate-free point sample. We do so on more general spaces in $\mathbb{R}^d$, only assuming a lower bound on the weak feature size of a compact, locally contractible domain, and that we can compute the distance to the boundary and between close pairs of sample points.

2 Background

Distance Functions. For a compact set $A \subset \mathbb{R}^d$, and metric $\mathrm{d}(\cdot,\cdot)$, define the distance function from $x \in \mathbb{R}^d$ to A as $\mathrm{d}(x, A) := \min_{y\in A} \mathrm{d}(x, y)$. The **$\varepsilon$-offsets** of a set A are defined as $A^\varepsilon := \{x \in X \mid \mathrm{d}(x, A) \leq \varepsilon\} = \cup_{x\in A}\mathrm{ball}_\varepsilon(x)$. Recall the set difference, or relative complement, of two sets A and B is defined as $A \setminus B := \{a \in A \mid a \notin B\}$. In this paper, the ambient space will be the one-point compactification of $\mathbb{R}^d$, $\mathbb{R}^d \cup \{\infty\}$, which is homeomorphic to the d-sphere, S^d, and the metric will be the Euclidean metric, $\|\cdot\|$.

Given a compact set $A \subset \mathbb{R}^d$, the critical value associated to a critical point x of the distance function is $\mathrm{d}(x, A)$. The **weak feature size** of A is defined to be the least positive critical value of $\mathrm{d}(\cdot, A)$ and is denoted $\mathrm{wfs}(A)$.

*University of Connecticut
†University of Connecticut
‡University of Connecticut

Homology. Homology is a tool from algebraic topology that gives a characterization of the shape of a space with regards to its k-dimensional holes. It is a topological invariant and as such it is preserved under homeomorphisms. We assume singular homology over a field, which implies that the resulting homology groups of a space X, written $H_*(X)$ when considered over all dimensions, are vector spaces. When referencing homology with respect to a specific dimension k, we write $H_k(X)$. If there exists a map between two spaces, e.g. $f : X \to Y$, then there is a map at the level of homology, $f_* : H_*(X) \to H_*(Y)$. We will denote such a map by $f_* := H_*(X \to Y)$.

Čech and Rips Complexes. When computing homology in practice, one often needs a finite simplicial complex to represent the space as their homology can be calculated via matrix multiplication.

Given a finite collection of points $P \in \mathbb{R}^d$, its **Čech complex** at scale ε is defined as $\mathcal{C}_\varepsilon(P) := \{\sigma \subseteq P \mid \exists x \in \mathbb{R}^d : \max_{p\in\sigma} \|x - p\| \leq \varepsilon\}$.

Of note is that the Čech complex of a point set P at scale ε is the nerve of the collection $\{\mathrm{ball}_\varepsilon(p)\}_{p\in P}$, whose union is P^ε. By the nerve theorem [4] , $\mathcal{C}_\varepsilon(P)$ is homotopic to P^ε, and thus $H_*(\mathcal{C}_\varepsilon(P)) \cong H_*(P^\varepsilon)$. This implies that by studying the homology of the ε-Čech complex, one knows the homology of the ε-offsets.

As we simply know the distance between close points in the sample, this is not enough to compute the Čech complex. Instead we compute the Rips complex, as this can be computed by checking pairs of points' distance. The **(Vietoris-)Rips complex** of P at scale ε is defined as $\mathcal{R}_\varepsilon(P) := \{\sigma \subseteq P \mid \{p, q\} \in \mathcal{C}_\varepsilon(P) \text{ for all } p, q \in \sigma\}$. The following inclusions help us in relating the knowledge about a Čech complex to a Rips complex.

For all $\varepsilon > 0$ and a finite point set $P \subseteq \mathbb{R}^d$, $\mathcal{C}_\varepsilon(P) \subseteq \mathcal{R}_\varepsilon(P) \subseteq \mathcal{C}_{\vartheta_d\varepsilon}(P)$, where $\vartheta_d = \sqrt{\frac{2d}{d+1}}$.

3 Results

Throughout we assume a compact, locally contractible domain $\mathcal{D} \subset \mathbb{R}^d$, with boundary $\mathcal{B}$, and a finite point sampling $P \subset \mathcal{D}$. Given a constant $\beta \geq 0$, define $U_\beta := P \setminus B^\beta = \{p \in P \mid \mathrm{d}(p, \mathcal{B}) > \beta\}$. These are the points of P at least a distance of β away from the boundary. This definition leads to the following lemma relating a subsampling to a shrunken domain, assuming we have coverage.

Lemma 1 *Given a domain $\mathcal{D}$ with boundary $\mathcal{B}$ and a finite set $P \subset \mathcal{D}$ with $\alpha > 0$ such that $\mathcal{D} \setminus \mathcal{B}^{2\alpha} \subseteq P^\alpha$, for all γ, β such that $\gamma, \beta \geq \alpha$, we have the following.*

$$U^\gamma_{\beta+\gamma} \subseteq \mathcal{D} \setminus \mathcal{B}^\beta \text{ and } \mathcal{D} \setminus \mathcal{B}^{\beta+\gamma} \subseteq U^\gamma_\beta.$$

The following lemma relates the homology of the inclusion between two sub-samplings of P to the homology of the domain $\mathcal{D}$.

Lemma 2 *Given a domain $\mathcal{D}$ with boundary $\mathcal{B}$ and a finite set $P \subset \mathcal{D}$ with $\alpha > 0$ such that $\mathcal{D} \setminus \mathcal{B}^{2\alpha} \subseteq P^\alpha$, let $\beta, \gamma, \varepsilon, \delta$ be constants such that $\delta \geq \varepsilon \geq \gamma \geq \alpha$ and $\beta \geq \varepsilon + \delta + \gamma$. If $\mathrm{wfs}(\mathcal{B}) > \beta + \gamma$, then for all $\lambda \in (0, \mathrm{wfs}(\mathcal{D}))$,*

$$rk(H_*(U^\gamma_\beta \hookrightarrow U^\varepsilon_\delta)) = dim(H_*(\mathcal{D}^\lambda)).$$

From this lemma we can prove our main theorem by switching to the Čech and Rips complexes. We assume that we know the distance between all $p, q \in P$ such that $\mathrm{d}(p,q) < \vartheta_d \alpha$, where α corresponds to the scale at which we have coverage. The constants used are a result of Lemma 2 and those required to achieve the Rips-Čech inclusions.

The theorem tells that we can compute the homology of a small offset of the domain by computing the Rips complexes at the two scales, and computing the image of the induced map between their homology groups. Furthermore, if D is homotopic to D^λ, the image tells the homology of the domain itself.

Theorem 3 *Given $\mathcal{D} \subset \mathbb{R}^d$, a compact, locally contractible domain with boundary $\mathcal{B}$ such that $\mathrm{wfs}(\mathcal{B}) > (2\vartheta_d^2 + 4\vartheta_d + 2)\alpha$, and finite point set $P \subset \mathcal{D}$ such that $\mathcal{D} \setminus \mathcal{B}^{2\alpha} \subseteq P^\alpha$, the following holds for all $\lambda \in (0, \mathrm{wfs}(\mathcal{D}))$.*

$$im(H_*(\mathcal{R}_\alpha(U_{(2\vartheta_d^2+4\vartheta_d+1)\alpha}) \hookrightarrow \mathcal{R}_{\vartheta_d\alpha}(U_{(2\vartheta_d^2+\vartheta_d)\alpha}))) \cong H_*(\mathcal{D}^\lambda).$$

Proof. Let $0 < \alpha \leq \beta \leq \gamma$ and $\delta_1 \geq \delta_2 \geq \delta_3 \geq \delta_4 \geq 0$. This leads to the following inclusions of sub-samplings.

$$U^\alpha_{\delta_1} \hookrightarrow U^\beta_{\delta_2} \hookrightarrow U^\beta_{\delta_3} \hookrightarrow U^\gamma_{\delta_4}$$

If $\beta \geq \vartheta_d \alpha$ and $\gamma \geq \vartheta_d \beta \geq \vartheta_d^2 \alpha$, then we have the following diagram with Čech and Rips complexes, with the diagonal maps due to the Rips-Čech interleaving.

$$\begin{array}{ccccccc} \mathcal{C}_\alpha(U_{\delta_1}) & \hookrightarrow & \mathcal{C}_\beta(U_{\delta_2}) & \hookrightarrow & \mathcal{C}_\beta(U_{\delta_3}) & \hookrightarrow & \mathcal{C}_\gamma(U_{\delta_4}) \\ & \searrow & \uparrow & & \downarrow & \nearrow & \\ & & \mathcal{R}_\alpha(U_{\delta_1}) & & \mathcal{R}_\beta(U_{\delta_3}) & & \end{array}$$

By applying the homology functor to the previous two diagrams, we get a commutative diagram of maps at the level of homology and vertical isomorphisms due to the Persistent Nerve Lemma [2].

$$\begin{array}{ccccccc} H_*(U^\alpha_{\delta_1}) & \longrightarrow & H_*(U^\beta_{\delta_2}) & \longrightarrow & H_*(U^\beta_{\delta_3}) & \longrightarrow & H_*(U^\gamma_{\delta_4}) \\ \downarrow\cong & & \downarrow\cong & & \downarrow\cong & & \downarrow\cong \\ H_*(\mathcal{C}_\alpha(U_{\delta_1})) & \to & H_*(\mathcal{C}_\beta(U_{\delta_2})) & \to & H_*(\mathcal{C}_\beta(U_{\delta_3})) & \to & H_*(\mathcal{C}_\gamma(U_{\delta_4})) \\ & \searrow & \uparrow & & \downarrow & \nearrow & \\ & & H_*(\mathcal{R}_\alpha(U_{\delta_1})) & & H_*(\mathcal{R}_\beta(U_{\delta_3})) & & \end{array}$$

If $\delta_1, \delta_4, \alpha, \gamma$ and $\delta_2, \delta_3, \beta$ are chosen such that they satisfy the assumptions of Lemma 2, then we have the following isomorphisms, as each vector space in question is finite-dimensional

$$\mathrm{im}(H_*(U^\alpha_{\delta_1} \hookrightarrow U^\gamma_{\delta_4})) \cong \mathrm{im}(H_*(U^\beta_{\delta_2} \hookrightarrow U^\beta_{\delta_3})) \cong H_*(\mathcal{D}^\lambda).$$

This also gives us the following isomorphisms at the level of the Čech complexes.

$$\mathrm{im}(H_*(\mathcal{C}_\alpha(U_{\delta_1}) \hookrightarrow \mathcal{C}_\gamma(U_{\delta_4})) \cong \mathrm{im}(H_*(\mathcal{C}_\beta(U_{\delta_2}) \hookrightarrow \mathcal{C}_\beta(U_{\delta_3})) \cong H_*(\mathcal{D}^\lambda).$$

The conditions for the Rips-Čech interleaving and Lemma 2 are satisfied by making the following substitutions. $\beta = \vartheta_d \alpha$, $\gamma = \vartheta_d^2 \alpha$, $\delta_1 = (2\vartheta_d^2 + 4\vartheta_d + 1)\alpha$, $\delta_2 = (2\vartheta_d^2 + 3\vartheta_d)\alpha$, $\delta_3 = (2\vartheta_d^2 + \vartheta_d)\alpha$, and $\delta_4 = \vartheta_d^2 \alpha$.

Applying Lemma 3.2 from Chazal and Oudut [2] to the following homological sequence using the above Čech isomorphisms completes the proof.

$$\begin{array}{l} H_*(\mathcal{C}_\alpha(U_{\delta_1})) \to H_*(\mathcal{R}_\alpha(U_{\delta_1})) \to H_*(\mathcal{C}_\beta(U_{\delta_2})) \\ \swarrow \\ H_*(\mathcal{C}_\beta(U_{\delta_3})) \to H_*(\mathcal{R}_\beta(U_{\delta_3})) \to H_*(\mathcal{C}_\gamma(U_{\delta_4})) \end{array}$$

□

References

[1] Frédéric Chazal and André Lieutier. Weak feature size and persistent homology: Computing homology of solids in R^n from noisy data samples. In *Proceedings of the 21st ACM Symposium on Computational Geometry*, pages 255–262, 2005.

[2] Frédéric Chazal and Steve Yann Oudot. Towards persistence-based reconstruction in euclidean spaces. In *Proceedings of the Twenty-fourth Annual Symposium on Computational Geometry*, SCG '08, pages 232–241, New York, NY, USA, 2008. ACM.

[3] Vin de Silva and Robert Ghrist. Homological sensor networks. *Notices Amer. Math. Soc.*, 54(1):10–17, 2007.

[4] Allen Hatcher. *Algebraic Topology.* Cambridge University Press, 2001.

[5] Partha Niyogi, Stephen Smale, and Shmuel Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete & Computational Geometry*, 39(1-3):419–441, 2008.