Topological Analysis of Nerves, Reeb Spaces, Mappers, and Multiscale Mappers*†

Tamal K. Dey[1], Facundo Mémoli[2], and Yusu Wang[3]

1 **Department of Computer Science and Engineering, The Ohio State University, Columbus, OH, USA**
`tamaldey@cse.ohio-state.edu`

2 **Department of Mathematics and Department of Computer Science and Engineering, The Ohio State University, Columbus, OH, USA**
`memoli@math.osu.edu`

3 **Department of Computer Science and Engineering, The Ohio State University, Columbus, OH, USA**
`yusu@cse.ohio-state.edu`

Abstract

Data analysis often concerns not only the space where data come from, but also various types of maps attached to data. In recent years, several related structures have been used to study maps on data, including Reeb spaces, mappers and multiscale mappers. The construction of these structures also relies on the so-called *nerve* of a cover of the domain.

In this paper, we aim to analyze the topological information encoded in these structures in order to provide better understanding of these structures and facilitate their practical usage.

More specifically, we show that the one-dimensional homology of the nerve complex $N(\mathcal{U})$ of a path-connected cover $\mathcal{U}$ of a domain X cannot be richer than that of the domain X itself. Intuitively, this result means that no new H_1-homology class can be "created" under a natural map from X to the nerve complex $N(\mathcal{U})$. Equipping X with a pseudometric d, we further refine this result and characterize the classes of $H_1(X)$ that may survive in the nerve complex using the notion of *size* of the covering elements in $\mathcal{U}$. These fundamental results about nerve complexes then lead to an analysis of the H_1-homology of Reeb spaces, mappers and multiscale mappers.

The analysis of H_1-homology groups unfortunately does not extend to higher dimensions. Nevertheless, by using a map-induced metric, establishing a Gromov-Hausdorff convergence result between mappers and the domain, and interleaving relevant modules, we can still analyze the persistent homology groups of (multiscale) mappers to establish a connection to Reeb spaces.

1998 ACM Subject Classification F.2.2 Nonnumerical Algorithms and Problems

Keywords and phrases Topology, nerves, mapper, multiscale mapper, Reeb spaces

Digital Object Identifier 10.4230/LIPIcs.SoCG.2017.36

1 Introduction

Data analysis often concerns not only the space where data come from, but also various types of information attached to data. For example, each node in a road network can contain information about the average traffic flow passing this point, a node in protein-protein interaction network can be associated with biochemical properties of the proteins involved.

* A full version of the paper is available at `https://arxiv.org/abs/1703.07387`.

† This work was partially supported by National Science Foundation under grant CCF-1526513.

33rd International Symposium on Computational Geometry (SoCG 2017).
Editors: Boris Aronov and Matthew J. Katz; Article No. 36; pp. 36:1–36:16
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Such information attached to data can be modeled as maps defined on the domain of interest; note that the maps are not necessarily $\mathbb{R}^d$-valued, e.g, the co-domain can be $\mathbb{S}^1$. Hence understanding data benefits from analyzing maps relating two spaces rather than a single space with no map on it.

In recent years, several related structures have been used to study general maps on data, including Reeb spaces [9, 11, 13, 18], mappers (and variants) [4, 8, 21] and multiscale mappers [10]. More specifically, given a map $f : X \to Z$ defined on a topological space X, the Reeb space R_f w.r.t. f (first studied for piecewise-linear maps in [13]), is a generalization of the so-called Reeb graph for a scalar function which has been used in various applications [2]. It is the quotient space of X w.r.t. an equivalence relation that asserts two points of X to be equivalent if they have the same function value and are connected to each other via points of the same function value. All equivalent points are collapsed into a single point in the Reeb space. Hence R_f provides a way to view X from the perspective of f.

The Mapper structure, originally introduced in [21], can be considered as a further generalization of the Reeb space. Given a map $f : X \to Z$, it also considers a cover $\mathcal{U}$ of the co-domain Z that enables viewing the structure of f at a coarser level. Intuitively, the equivalence relation between points in X is now defined by whether points are within the same connected component of the pre-image of a cover element $U \in \mathcal{U}$. Instead of a quotient space, the mapper takes the nerve complex of the cover of X formed by the connected components of the pre-images of all elements in $\mathcal{U}$ (i.e, the cover formed by those equivalent points). Hence the mapper structure provides a view of X from the perspective of both f and a cover of the co-domain Z.

Finally, both the Reeb space and the mapper structures provide a fixed snapshot of the input map f. As we vary the cover $\mathcal{U}$ of the co-domain Z, we obtain a family of snapshots at different granularities. The *multiscale mapper* [10] describes the sequence of the mapper structures as one varies the granularity of the cover of Z through a sequence of covers of Z connected via cover maps.

New work. While these structures are meaningful in that they summarize the information contained in data, there has not been any qualitative analysis of the precise information encoded by them with the only exception of [4] and [14] [1]. In this paper, we aim to analyze the *topological information* encoded by these structures, so as to provide better understanding of these structures and facilitate their practical usage [12, 17]. In particular, the construction of the mapper and multiscale mapper use the so-called *nerve* of a cover of the domain. To understand the mappers and multiscale mappers, we first provide a quantitative analysis of the topological information encoded in the nerve of a reasonably well-behaved cover for a domain. Given the generality and importance of the nerve complex in topological studies, this result is of independent interest.

More specifically, in Section 3, we first obtain a general result that relates the one dimensional homology H_1 of the nerve complex $N(\mathcal{U})$ of a path-connected cover $\mathcal{U}$ (where each open set contained is path-connected) of a domain X to that of the domain X itself. Intuitively, this result says that no new H_1-homology classes can be "created" under a natural map from X to the nerve complex $N(\mathcal{U})$. Equipping X with a pseudometric d, we further

[1] Carrière and Oudot [4] analyzed certain persistence diagram of mappers induced by a real-valued function, and provided a characterization for it in terms of the persistence diagram of the corresponding Reeb graph. Gasparovic et al [14] provides full description of the persistence homology information encoded in the *intrinsic Čech complex* (a special type of nerve complex) of a metric graph.

refine this result and quantify the classes of $H_1(X)$ that may survive in the nerve complex (Theorem 21, Section 4). This demarcation is obtained via a notion of *size* of covering elements in $\mathcal{U}$. These fundamental results about nerve complexes then lead to an analysis of the H_1-homology classes in Reeb spaces (Theorem 27), mappers and multiscale mappers (Theorem 29). The analysis of H_1-homology groups unfortunately does not extend to higher dimensions. Nevertheless, we can still provide an interesting analysis of the persistent homology groups for these structures (Theorem 36, Section 5). During this course, by using a map-induced metric, we establish a Gromov-Hausdorff convergence between the mapper structure and the domain. This offers an alternative to [18] for defining the convergence between mappers and the Reeb space, which may be of independent interest.

All missing proofs in what follows are deferred to the full version of this paper on arXiv.

2 Topological background and motivation

Space, paths, covers. Let X denote a path connected topological space. Since X is path connected, there exists a path $\gamma : [0,1] \to X$ connecting every pair of points $\{x, x'\} \in X \times X$ where $\gamma(0) = x$ and $\gamma(1) = x'$. Let $\Gamma_X(x, x')$ denote the set of all such paths connecting x and x'. These paths play an important role in our definitions and arguments.

By a cover of X we mean a collection $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$ of open sets such that $\bigcup_{\alpha \in A} U_\alpha = X$. A cover $\mathcal{U}$ is *path connected* if each U_α is path connected. In this paper, we consider only path connected covers.

Later to define maps between X and its nerve complexes, we need X to be *paracompact*, that is, every cover $\mathcal{U}$ of X has a subcover $\mathcal{U}' \subseteq \mathcal{U}$ so that each point $x \in X$ has an open neighborhood contained in *finitely many* elements of $\mathcal{U}'$. Such a cover $\mathcal{U}'$ is called *locally finite*. From now on, we assume X *to be compact* which implies that it is paracompact too.

▶ **Definition 1** (Simplicial complex and maps)**.** A simplicial complex K with a vertex set V is a collection of subsets of V with the condition that if $\sigma \in 2^V$ is in K, then all subsets of σ are in K. We denote the geometric realization of K by $|K|$. Let K and L be two simplicial complexes. A map $\phi : K \to L$ is *simplicial* if for every simplex $\sigma = \{v_1, v_2, \ldots, v_p\}$ in K, the simplex $\phi(\sigma) = \{\phi(v_1), \phi(v_2), \ldots, \phi(v_p)\}$ is in L.

▶ **Definition 2** (Nerve of a cover)**.** Given a cover $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$ of X, we define the *nerve* of the cover $\mathcal{U}$ to be the simplicial complex $N(\mathcal{U})$ whose vertex set is the index set A, and where a subset $\{\alpha_0, \alpha_1, \ldots, \alpha_k\} \subseteq A$ spans a k-simplex in $N(\mathcal{U})$ if and only if $U_{\alpha_0} \cap U_{\alpha_1} \cap \ldots \cap U_{\alpha_k} \neq \emptyset$.

Maps between covers. Given two covers $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$ and $\mathcal{V} = \{V_\beta\}_{\beta \in B}$ of X, a *map of covers* from $\mathcal{U}$ to $\mathcal{V}$ is a set map $\xi : A \to B$ so that $U_\alpha \subseteq V_{\xi(\alpha)}$ for all $\alpha \in A$. *By a slight abuse of notation we also use* ξ *to indicate the map* $\mathcal{U} \to \mathcal{V}$. Given such a map of covers, there is an induced simplicial map $N(\xi) : N(\mathcal{U}) \to N(\mathcal{V})$, given on vertices by the map ξ. Furthermore, if $\mathcal{U} \xrightarrow{\xi} \mathcal{V} \xrightarrow{\zeta} \mathcal{W}$ are three covers of X with the intervening maps of covers between them, then $N(\zeta \circ \xi) = N(\zeta) \circ N(\xi)$ as well. The following simple result is useful.

▶ **Proposition 3** (Maps of covers induce contiguous simplicial maps [10])**.** *Let* $\zeta, \xi : \mathcal{U} \to \mathcal{V}$ *be any two maps of covers. Then, the simplicial maps* $N(\zeta)$ *and* $N(\xi)$ *are contiguous.*

Recall that two simplicial maps $h_1, h_2 : K \to L$ are *contiguous* if for all $\sigma \in K$ it holds that $h_1(\sigma) \cup h_2(\sigma) \in L$. In particular, contiguous maps induce identical maps at the homology level [19]. Let $H_k(\cdot)$ denote the k-dimensional homology of the space in its argument. This homology is *singular* or *simplicial* depending on if the argument is a topological space or a

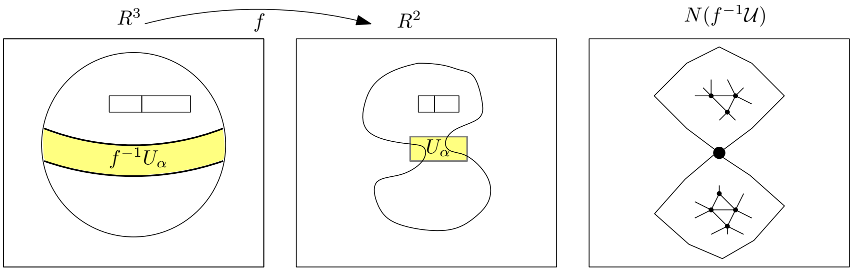


Figure 1 The map $f : \mathbb{S}^2 \subset \mathbb{R}^3 \to \mathbb{R}^2$ takes the sphere to $\mathbb{R}^2$. The pullback of the cover element U_α makes a band surrounding the equator which causes the nerve $N(f^{-1}\mathcal{U})$ to pinch in the middle creating two 2-cycles. This shows that the map $\phi_* : X \to N(*)$ may not induce a surjection in H_2.

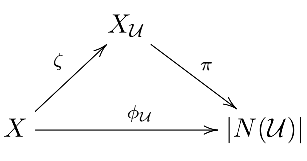
simplicial complex respectively. All homology groups in this paper are defined over the field $\mathbb{Z}_2$. Proposition 3 implies that the map $H_k(N(\mathcal{U})) \to H_k(N(\mathcal{V}))$ arising out of a cover map can be deemed canonical.

3 Surjectivity in H_1-persistence

In this section we first establish a map $\phi_\mathcal{U}$ between X and the geometric realization $|N(\mathcal{U})|$ of a nerve complex $N(\mathcal{U})$. This helps us to define a map $\phi_\mathcal{U}*$ from the singular homology groups of X to the simplicial homology groups of $N(\mathcal{U})$ via the singular homology of $|N(\mathcal{U})|$. The famous nerve theorem [3, 16] says that if the elements of $\mathcal{U}$ intersect only in contractible spaces, then $\phi_\mathcal{U}$ is a homotopy equivalence and hence $\phi_\mathcal{U}*$ leads to an isomorphism between $H_*(X)$ and $H_*(N(\mathcal{U}))$. The contractibility condition can be weakened to a *homology ball* condition to retain the isomorphism between the two homology groups [16]. In absence of such conditions of the cover, simple examples exist to show that $\phi_\mathcal{U}*$ is neither a monophorphism (injection) nor an epimorphism (surjection). Figure 1 gives an example where $\phi_{\mathcal{U}*}$ is not sujective in H_2. However, for one dimensional homology we show that, for any path connected cover $\mathcal{U}$, the map $\phi_{\mathcal{U}*}$ is necessarily a surjection. One implication of this is that the simplicial maps arising out of cover maps induce a surjection among the one dimensional homology groups of two nerve complexes.

3.1 Nerves

The proof of the nerve theorem [15] uses a construction that connects the two spaces X and $|N(\mathcal{U})|$ via a third space $X_\mathcal{U}$ that is a product space of $\mathcal{U}$ and the geometric realization $|N(\mathcal{U})|$.

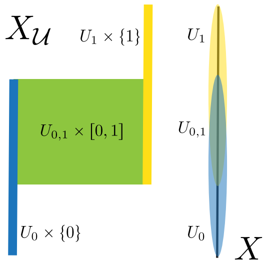


In our case $\mathcal{U}$ may not satisfy the contractibility condition. Nevertheless, we use the same construction to define three maps, $\zeta : X \to X_\mathcal{U}$, $\pi : X_\mathcal{U} \to |N(\mathcal{U})|$, and $\phi_\mathcal{U} : X \to |N(\mathcal{U})|$ where $\phi_\mathcal{U} = \pi \circ \zeta$ is referred to as the *nerve map*. Details about the construction of these maps follow.

Denote the elements of the cover $\mathcal{U}$ as U_α for α taken from some indexing set A. The vertices of $N(\mathcal{U})$ are denoted by $\{u_\alpha, \alpha \in A\}$, where each u_α corresponds to the cover element U_α. For each finite non-empty intersection $U_{\alpha_0,\ldots,\alpha_n} := \bigcap_{i=0}^{n} U_{\alpha_i}$ consider the product $U_{\alpha_0,\ldots,\alpha_n} \times \Delta^n_{\alpha_0,\ldots,\alpha_n}$, where $\Delta^n_{\alpha_0,\ldots,\alpha_n}$ denotes the n-dimensional simplex with vertices $u_{\alpha_0}, \ldots, u_{\alpha_n}$. Consider now the disjoint union

$$M := \bigsqcup_{\alpha_0,\ldots,\alpha_n \in A:\, U_{\alpha_0,\ldots,\alpha_n} \neq \emptyset} U_{\alpha_0,\ldots,\alpha_n} \times \Delta^n_{\alpha_0,\ldots,\alpha_n}$$

together with the following identification: each point $(x, y) \in M$, with $x \in U_{\alpha_0,\ldots,\alpha_n}$ and $y \in [\alpha_0, \ldots, \widehat{\alpha}_i, \ldots, \alpha_n] \subset \Delta^n_{\alpha_0,\ldots,\alpha_n}$ is identified with the corresponding point in the product $U_{\alpha_0,\ldots,\widehat{\alpha}_i,\ldots,\alpha_n} \times \Delta_{\alpha_0,\ldots,\widehat{\alpha}_i,\ldots,\alpha_n}$ via the inclusion $U_{\alpha_0,\ldots,\alpha_n} \subset U_{\alpha_0,\ldots,\widehat{\alpha}_i,\ldots,\alpha_n}$. Here $[\alpha_0, \ldots, \widehat{\alpha}_i, \ldots, \alpha_n]$ denotes the i-th face of the simplex $\Delta^n_{\alpha_0,\ldots,\alpha_n}$. Denote by $\sim$ this identification and now define the space $X_{\mathcal{U}} := M \,/ \sim$. An example for the case when X is a line segment and $\mathcal{U}$ consists of only two open sets is shown below.



▶ **Definition 4.** A collection of real valued continuous functions $\{\varphi_\alpha :\to [0,1], \alpha \in A\}$ is called a *partition of unity* if (i) $\sum_{\alpha \in A} \varphi_\alpha(x) = 1$ for all $x \in X$, (ii) For every $x \in X$, there are only finitely many $\alpha \in A$ such that $\varphi_\alpha(x) > 0$.

If $\mathcal{U} = \{U_\alpha,\, \alpha \in A\}$ is any open cover of X, then a partition of unity $\{\varphi_\alpha,\, \alpha \in A\}$ is *subordinate* to $\mathcal{U}$ if $\mathrm{supp}(\varphi_\alpha)$ is contained in U_α for each $\alpha \in A$.

Since X is paracompact, for any open cover $\mathcal{U} = \{U_\alpha,\, \alpha \in A\}$ of X, there exists a partition of unity $\{\varphi_\alpha, \alpha \in A\}$ subordinate to $\mathcal{U}$ [20]. For each $x \in X$ such that $x \in U_\alpha$, denote by x_α the corresponding copy of x residing in $X_{\mathcal{U}}$. Then, the map $\zeta : X \to X_{\mathcal{U}}$ is defined as follows: for any $x \in X$,

$$\zeta(x) := \sum_{\alpha \in A} \varphi_\alpha(x)\, x_\alpha.$$

The map $\pi : X_{\mathcal{U}} \to |N(\mathcal{U})|$ is induced by the individual projection maps

$$U_{\alpha_0,\ldots,\alpha_n} \times \Delta^n_{\alpha_0,\ldots,\alpha_n} \to \Delta^n_{\alpha_0,\ldots,\alpha_n}.$$

Then, it follows that $\phi_{\mathcal{U}} = \pi \circ \zeta : X \to |N(\mathcal{U})|$ satisfies, for $x \in X$,

$$\phi_{\mathcal{U}}(x) = \sum_{\alpha \in A} \varphi_\alpha(x)\, u_\alpha. \tag{1}$$

We have the following fact [20, pp. 108]:

▶ **Fact 5.** *ζ is a homotopy equivalence.*

3.2 From space to nerves

Now, we show that the nerve maps at the homology level are surjective for one dimensional homology. Interestingly, the result is not true beyond one dimensional homology (see Figure 1) which is probably why this simple but important fact has not been observed before. First, we make a simple observation that connects the classes in singular homology of $|N(\mathcal{U})|$ to those in the simplicial homology of $N(\mathcal{U})$. The result follows immediately from the isomorphism between singular and simplicial homology induced by the geometric realization; see [19, Theorem 34.3]. In what follows let $[c]$ denote the class of a cycle c.

▶ **Proposition 6.** *Every 1-cycle ξ in $|N(\mathcal{U})|$ has a 1-cycle γ in $N(\mathcal{U})$ so that $[\xi] = [|\gamma|]$.*

▶ **Proposition 7.** *If $\mathcal{U}$ is path connected, $\phi_{\mathcal{U}*} : H_1(X) \to H_1(|N(\mathcal{U})|)$ is a surjection.*

Proof. Let $[\gamma]$ be any class in $H_1(|N(\mathcal{U})|)$. Because of Proposition 6, we can assume that $\gamma = |\gamma'|$, where γ' is a 1-cycle in the 1-skeleton of $N(\mathcal{U})$. We construct a 1-cycle $\gamma_{\mathcal{U}}$ in $X_{\mathcal{U}}$ so that $\pi(\gamma_{\mathcal{U}}) = \gamma$. Recall the map $\zeta : X \to X_{\mathcal{U}}$ in the construction of the nerve map $\phi_{\mathcal{U}}$ where $\phi_{\mathcal{U}} = \pi \circ \zeta$. There exists a class $[\gamma_X]$ in $H_1(X)$ so that $\zeta_*([\gamma_X]) = [\gamma_{\mathcal{U}}]$ because ζ_* is an isomorphism by Fact 5. Then, $\phi_{\mathcal{U}*}([\gamma_X]) = \pi_*(\zeta_*([\gamma_X]))$ because $\phi_{\mathcal{U}*} = \pi_* \circ \zeta_*$. It follows $\phi_{\mathcal{U}*}([\gamma_X]) = \pi_*([\gamma_{\mathcal{U}}]) = [\gamma]$ showing that $\phi_{\mathcal{U}*}$ is surjective.

Therefore, it remains only to show that a 1-cycle $\gamma_{\mathcal{U}}$ can be constructed given γ in $|N(\mathcal{U})|$ so that $\pi(\gamma_{\mathcal{U}}) = \gamma$. See the full version for this construction. ◀

Since we are eventually interested in the simplicial homology groups of the nerves rather than the singular homology groups of their geometric realizations, we make one more transition using the known isomorphism between the two homology groups. Specifically, if $\iota_{\mathcal{U}} : H_k(|N(\mathcal{U})|) \to H_k(N(\mathcal{U}))$ denotes this isomorphism, we let $\bar{\phi}_{\mathcal{U}*}$ denote the composition $\iota_{\mathcal{U}} \circ \phi_{\mathcal{U}*}$. As a corollary to Proposition 7, we obtain:

▶ **Theorem 8.** *If $\mathcal{U}$ is path connected, $\bar{\phi}_{\mathcal{U}*} : H_1(X) \to H_1(N(\mathcal{U}))$ is a surjection.*

3.3 From nerves to nerves

In this section we extend the result in Theorem 8 to simplicial maps between two nerves induced by cover maps. The following proposition is key to establishing the result.

▶ **Proposition 9** (Coherent partitions of unity). *Suppose $\{U_\alpha\}_{\alpha\in A} = \mathcal{U} \xrightarrow{\theta} \mathcal{V} = \{V_\beta\}_{\beta\in B}$ are open covers of the paracompact topological space X and $\theta : A \to B$ is a map of covers. Then there exists a partition of unity $\{\varphi_\alpha\}_{\alpha\in A}$ subordinate to the cover $\mathcal{U}$ such that if for each $\beta \in B$ we define*

$$\psi_\beta := \begin{cases} \sum_{\alpha\in\theta^{-1}(\beta)} \varphi_\alpha & \text{if } \beta \in \mathrm{im}(\theta); \\ 0 & \text{otherwise.} \end{cases}$$

then the set of functions $\{\psi_\beta\}_{\beta\in B}$ is a partition of unity subordinate to the cover $\mathcal{V}$.

Proof is deferred to the full version.

Let $\{U_\alpha\}_{\alpha\in A} = \mathcal{U} \xrightarrow{\theta} \mathcal{V} = \{V_\beta\}_{\beta\in B}$ be two open covers of X connected by a map of covers. Apply Proposition 9 to obtain coherent partitions of unity $\{\varphi_\alpha\}_{\alpha\in A}$ and $\{\psi_\beta\}_{\beta\in B}$ subordinate to $\mathcal{U}$ and $\mathcal{V}$, respectively. Let the nerve maps $\phi_{\mathcal{U}} : X \to |N(\mathcal{U})|$ and $\phi_{\mathcal{V}} : X \to |N(\mathcal{V})|$ be defined as in (1) above. Let $N(\mathcal{U}) \xrightarrow{\tau} N(\mathcal{V})$ be the simplicial map induced by the cover map θ. Then, τ can be extended to a continuous map $\hat{\tau}$ on the image of $\phi_{\mathcal{U}}$ as follows: for $x \in X$, $\hat{\tau}(\phi_{\mathcal{U}}(x)) = \Sigma_{\alpha\in A}\varphi_\alpha(x)\, v_{\theta(\alpha)}$.

▶ **Proposition 10.** *Let $\mathcal{U}$ and $\mathcal{V}$ be two covers of X connected by a cover map $\mathcal{U} \xrightarrow{\theta} \mathcal{V}$. Then, the nerve maps $\phi_{\mathcal{U}}$ and $\phi_{\mathcal{V}}$ satisfy $\phi_{\mathcal{V}} = \hat{\tau} \circ \phi_{\mathcal{U}}$ where $\tau : N(\mathcal{U}) \to N(\mathcal{V})$ is the simplicial map induced by the cover map θ.*

Proof. For any point $p \in \mathrm{im}(\phi_{\mathcal{U}})$, there is $x \in X$ where $p = \phi_{\mathcal{U}}(x) = \Sigma_{\alpha \in A} \varphi_\alpha(x) u_\alpha$. Then,

$$\hat{\tau} \circ \phi_{\mathcal{U}}(x) = \hat{\tau}\left(\sum_{\alpha \in A} \varphi_\alpha(x) u_\alpha\right) = \sum_{\alpha \in A} \varphi_\alpha(x) \tau(u_\alpha) = \sum_{\alpha \in A} \varphi_\alpha(x)\, v_{\theta(\alpha)}$$
$$= \sum_{\beta \in B} \sum_{\alpha \in \theta^{-1}(\beta)} \varphi_\alpha(x)\, v_{\theta(\alpha)} = \sum_{\beta \in B} \psi_\beta(x) v_\beta = \phi_{\mathcal{V}}(x)\,. \quad \blacktriangleleft$$

An immediate corollary of the above Proposition is:

▶ **Corollary 11.** *The induced maps of* $\phi_{\mathcal{U}*} : H_k(X) \to H_k(|N(\mathcal{U})|)$, $\phi_{\mathcal{V}*} : H_k(X) \to H_k(|N(\mathcal{V})|)$, *and* $\hat{\tau}_* : H_k(|N(\mathcal{U})|) \to H_k(|N(\mathcal{V})|)$ *at the homology levels commute, that is,* $\phi_{\mathcal{V}*} = \hat{\tau}_* \circ \phi_{\mathcal{U}*}$.

With transition from singular to simplicial homology, Corollary 11 implies that:

▶ **Proposition 12.** $\bar{\phi}_{\mathcal{V}*} = \tau_* \circ \bar{\phi}_{\mathcal{U}*}$ *where* $\bar{\phi}_{\mathcal{V}*} : H_k(X) \to H_k(N(\mathcal{V}))$, $\bar{\phi}_{\mathcal{U}*} : H_k(X) \to H_k(N(\mathcal{U}))$ *and* $\tau : N(\mathcal{U}) \to N(\mathcal{V})$ *is the simplicial map induced by a cover map* $\mathcal{U} \to \mathcal{V}$.

Proposition 12 extends Theorem 8 to the simplicial maps between two nerves.

▶ **Theorem 13.** *Let* $\tau : N(\mathcal{U}) \to N(\mathcal{V})$ *be a simplicial map induced by a cover map* $\mathcal{U} \to \mathcal{V}$ *where both* $\mathcal{U}$ *and* $\mathcal{V}$ *are path connected. Then,* $\tau_* : H_1(N(\mathcal{U})) \to H_1(N(\mathcal{V}))$ *is a surjection.*

Proof. Consider the maps

$$H_1(X) \stackrel{\bar{\phi}_{\mathcal{U}*}}{\to} H_1(N(\mathcal{U})) \stackrel{\tau_*}{\to} H_1(N(\mathcal{V})), \text{ and } H_1(X) \stackrel{\bar{\phi}_{\mathcal{V}*}}{\to} H_1(N(\mathcal{V})).$$

By Proposition 12, $\tau_* \circ \bar{\phi}_{\mathcal{U}*} = \bar{\phi}_{\mathcal{V}*}$. By Theorem 8, the map $\bar{\phi}_{\mathcal{V}*}$ is a surjection. It follows that τ_* is a surjection. ◀

3.4 Mapper and multiscale mapper

In this section we extend the previous results to the structures called mapper and multiscale mapper. Recall that X is assumed to be compact. Consider a cover of X obtained indirectly as a pullback of a cover of another space Z. This gives rise to the so called *Mapper* and *Multiscale Mapper*. Let $f : X \to Z$ be a continuous map where Z is equipped with an open cover $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$ for some index set A. Since f is continuous, the sets $\{f^{-1}(U_\alpha),\, \alpha \in A\}$ form an open cover of X. For each α, we can now consider the decomposition of $f^{-1}(U_\alpha)$ into its path connected components, so we write $f^{-1}(U_\alpha) = \bigcup_{i=1}^{j_\alpha} V_{\alpha,i}$, where j_α is the number of path connected components $V_{\alpha,i}$'s in $f^{-1}(U_\alpha)$. We write $f^*\mathcal{U}$ for the cover of X obtained this way from the cover $\mathcal{U}$ of Z and refer to it as the *pullback* cover of X induced by $\mathcal{U}$ via f. Note that by its construction, this pullback cover $f^*\mathcal{U}$ is path-connected.

Notice that there are pathological examples of f where $f^{-1}(U_\alpha)$ may shatter into infinitely many path components. This motivates us to consider *well-behaved* functions f: we require that for every path connected open set $U \subseteq Z$, the preimage $f^{-1}(U)$ has *finitely* many open path connected components. Henceforth, all such functions are assumed to be well-behaved.

▶ **Definition 14** (Mapper [21])**.** Let $f : X \to Z$ be a continuous map. Let $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$ be an open cover of Z. The *mapper* arising from these data is defined to be the nerve simplicial complex of the pullback cover: $\mathrm{M}(\mathcal{U}, f) := N(f^*\mathcal{U})$.

When we consider a continuous map $f : X \to Z$ and we are given a map of covers $\xi : \mathcal{U} \to \mathcal{V}$ between covers of Z, we observed in [10] that there is a corresponding map of covers between the respective pullback covers of X: $f^*(\xi) : f^*\mathcal{U} \longrightarrow f^*\mathcal{V}$. Furthermore, if $\mathcal{U} \xrightarrow{\xi} \mathcal{V} \xrightarrow{\theta} \mathcal{W}$ are three different covers of a topological space with the intervening maps of covers between them, then $f^*(\theta \circ \xi) = f^*(\theta) \circ f^*(\xi)$.

In the definition below, *objects* can be covers, simplicial complexes, or vector spaces.

▶ **Definition 15** (Tower). A *tower* $\mathfrak{W}$ with resolution $r \in \mathbb{R}$ is any collection $\mathfrak{W} = \{\mathcal{W}_\varepsilon\}_{\varepsilon \geq r}$ of objects $\mathcal{W}_\varepsilon$ indexed in $\mathbb{R}$ together with maps $w_{\varepsilon,\varepsilon'} : \mathcal{W}_\varepsilon \to \mathcal{W}_{\varepsilon'}$ so that $w_{\varepsilon,\varepsilon} = \mathrm{id}$ and $w_{\varepsilon',\varepsilon''} \circ w_{\varepsilon,\varepsilon'} = w_{\varepsilon,\varepsilon''}$ for all $r \leq \varepsilon \leq \varepsilon' \leq \varepsilon''$. Sometimes we write $\mathfrak{W} = \{\mathcal{W}_\varepsilon \xrightarrow{w_{\varepsilon,\varepsilon'}} \mathcal{W}_{\varepsilon'}\}_{r \leq \varepsilon \leq \varepsilon'}$ to denote the collection with the maps. Given such a tower $\mathfrak{W}$, $\mathrm{res}(\mathfrak{W})$ refers to its resolution.

When $\mathfrak{W}$ is a collection of covers equipped with maps of covers between them, we call it *a tower of covers*. When $\mathfrak{W}$ is a collection of simplicial complexes equipped with simplicial maps between them, we call it *a tower of simplicial complexes*.

The pullback properties described at the end of section 2 make it possible to take the pullback of a given tower of covers of a space via a given continuous function into another space, so that we obtain the following.

▶ **Proposition 16** ([10]). *Let $\mathfrak{U} = \{\mathcal{U}_\varepsilon\}$ be a tower of covers of Z and $f : X \to Z$ be a continuous function. Then, $f^*\mathfrak{U} = \{f^*\mathcal{U}_\varepsilon\}$ is a tower of (path-connected) covers of X.*

In general, given a tower of covers $\mathfrak{W}$ of a space X, the nerve of each cover in $\mathfrak{W}$ together with each map of $\mathfrak{W}$ provides a tower of simplicial complexes which we denote by $N(\mathfrak{W})$.

▶ **Definition 17** (Multiscale Mapper [10]). Let $f : X \to Z$ be a continuous map. Let $\mathfrak{U}$ be a tower of covers of Z. Then, the *multiscale mapper* is defined to be the tower of the nerve simplicial complexes of the pullback: $\mathrm{MM}(\mathfrak{U}, f) := N(f^*\mathfrak{U})$.

As we indicated earlier, in general, no surjection between X and its nerve may exist at the homology level. It follows that the same is true for the mapper $N(f^*\mathcal{U})$. But, for H_1, we can apply the results contained in previous section to claim the following.

▶ **Theorem 18.** *Consider the following multiscale mapper arising out of a tower of path connected covers:*

$$N(f^*\mathcal{U}_0) \to N(f^*\mathcal{U}_1) \to \cdots \to N(f^*\mathcal{U}_n).$$

- *There is a surjection from $H_1(X)$ to $H_1(N(f^*\mathcal{U}_i))$ for each $i \in [0, n]$.*
- *Consider a H_1-persistence module of a multiscale mapper as shown below.*

$$\mathrm{H}_1\big(N(f^*\mathcal{U}_0)\big) \to \mathrm{H}_1\big(N(f^*\mathcal{U}_1)\big) \to \cdots \to \mathrm{H}_1\big(N(f^*\mathcal{U}_n)\big). \tag{2}$$

  All connecting maps in the above module are surjections.

The above result implies that, as we proceed forward through the multiscale mapper, no new homology classes are born. They can only die. Consequently, all bar codes in the persistence diagram of the H_1-persistence module induced by it have the left endpoint at 0.

4 Analysis of persistent H_1-classes

Using the language of persistent homology, the results in the previous section imply that one dimensional homology classes can die in the nerves, but they cannot be born. In this section, we analyze further to identify the classes that survive. The distinction among the classes is made via a notion of 'size'. Intuitively, we show that the classes with 'size' much larger than the 'size' of the cover survive. The 'size' is defined with the pseudometric that the space X is assumed to be equipped with. Precise statements are made in the subsections.

4.1 H_1-classes of nerves of pseudometric spaces

Let (X, d) be a pseudometric space, that is, d satisfies the axioms of a metric except that $d(x, x') = 0$ may not necessarily imply $x = x'$. Assume X to be compact as before. We define a 'size' for a homology class that reflects how big the smallest generator in the class is in the metric d.

▶ **Definition 19.** The size $s(X')$ of a subset X' of the pseudometric space (X, d) is defined to be its diameter, that is, $s(X') = \sup_{x,x' \in X' \times X'} d(x, x')$. The size of a class $c \in H_k(X)$ is defined as $s(c) = \inf_{z \in c} s(z)$.

▶ **Definition 20.** A set of k-cycles $z_1, z_2, \ldots, z_n$ of $H_k(X)$ is called a generator basis if the classes $[z_1], [z_2], \ldots, [z_n]$ together form a basis of $H_k(X)$. It is called a minimal generator basis if $\Sigma_{i=1}^{n} s(z_i)$ is minimal among all generator bases.

Lebesgue number of a cover. Our goal is to characterize the classes in the nerve of $\mathcal{U}$ with respect to the sizes of their preimages in X via the map $\phi_{\mathcal{U}}$. The Lebesgue number of a cover $\mathcal{U}$ becomes useful in this characterization. It is the largest number $\lambda(\mathcal{U})$ so that any subset of X with size at most $\lambda(\mathcal{U})$ is contained in at least one element of $\mathcal{U}$. Formally,

$$\lambda(\mathcal{U}) = \sup\{\delta \,|\, \forall X' \subseteq X \text{ with } s(X') \leq \delta, \exists U_\alpha \in \mathcal{U} \text{ where } U_\alpha \supseteq X'\}\,.$$

We observe that a homology class of size no more than $\lambda(\mathcal{U})$ cannot survive in the nerve. Further, the homology classes whose sizes are significantly larger than the maximum size of a cover do necessarily survive where we define the maximum size of a cover as $s_{max}(\mathcal{U}) := \max_{U \in \mathcal{U}}\{s(U)\}$.

Let $z_1, z_2, \ldots, z_g$ be a non-decreasing sequence of the generators with respect to their sizes in a minimal generator basis of $H_1(X)$. Consider the map $\phi_{\mathcal{U}} : X \to |N(\mathcal{U})|$ as introduced in Section 3. We have the following result.

▶ **Theorem 21.** *Let $\mathcal{U}$ be a path-connected cover of X.*

(i) *Let $\ell = g + 1$ if $\lambda(\mathcal{U}) > s(z_g)$. Otherwise, let $\ell \in [1, g]$ be the smallest integer so that $s(z_\ell) > \lambda(\mathcal{U})$. If $\ell \neq 1$, the class $\bar{\phi}_{\mathcal{U}*}[z_j] = 0$ for $j = 1, \ldots, \ell - 1$. Moreover, if $\ell \neq g + 1$, the classes $\{\bar{\phi}_{\mathcal{U}*}[z_j]\}_{j=\ell,\ldots,g}$ generate $H_1(N(\mathcal{U}))$.*

(ii) *The classes $\{\bar{\phi}_{\mathcal{U}*}[z_j]\}_{j=\ell',\ldots,g}$ are linearly independent where $s(z_{\ell'}) > 4s_{max}(\mathcal{U})$.*

The result above says that only the classes of $H_1(X)$ generated by generators of large enough size survive in the nerve. To prove this result, we use a map ρ that sends each 1-cycle in $N(\mathcal{U})$ to a 1-cycle in X. We define a chain map $\rho : \mathcal{C}_1(N(\mathcal{U})) \to \mathcal{C}_1(X)$ among one dimensional chain groups as follows [2]. It is sufficient to exhibit the map for an elementary chain of an edge, say $e = \{u_\alpha, u_{\alpha'}\} \in \mathcal{C}_1(N(\mathcal{U}))$. Since e is an edge in $N(\mathcal{U})$, the two cover elements U_α and $U_{\alpha'}$ in X have a common intersection. Let $a \in U_\alpha$ and $b \in U_{\alpha'}$ be two points that are arbitrary but fixed for U_α and $U_{\alpha'}$ respectively. Pick a path $\xi(a, b)$ (viewed

[2] We note that the high level framework of defining such a chain map and analyzing what it does to homologous cycles is similar to the work by Gasparovic et al. [14]. The technical details are different.

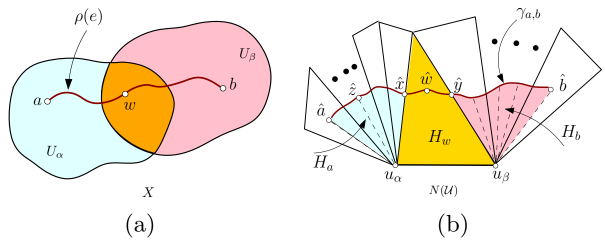


Figure 2 Illustration for proof of Proposition 22.

as a singular chain) in the union of U_α and $U_{\alpha'}$ which is path connected as both U_α and $U_{\alpha'}$ are. Then, define $\rho(e) = \xi(a,b)$. The following properties of $\phi_\mathcal{U}$ and ρ turn out to be useful.

▶ **Proposition 22.** *Let γ be any 1-cycle in $N(\mathcal{U})$. Then, $[\phi_\mathcal{U}(\rho(\gamma))] = [|\gamma|]$.*

Proof. Let $e = (u_\alpha, u_\beta)$ be an edge in γ with u_α and u_β corresponding to U_α and U_β respectively. Let a and b be the corresponding fixed points for set U_α and U_β respectively. Consider the path $\rho(e) = \xi(a,b)$ in X as constructed above, and set $\gamma_{a,b} = \phi_\mathcal{U}(\xi(a,b))$ to be the image of $\rho(e)$ in $|N(\mathcal{U})|$. See Figure 2 for an illustration. Given an oriented path ℓ and two points $x, y \in \ell$, we use $\ell[x,y]$ to denote the subpath of ℓ from x to y. For a point $x \in X$, for simplicity we set $\hat{x} = \phi_\mathcal{U}(x)$ to be its image in $|N(\mathcal{U})|$.

Now, let $w \in \rho(e)$ be a point in $U_\alpha \cap U_\beta$, and $\hat{w} = \phi_\mathcal{U}(w)$ be its image in $\gamma_{a,b}$. Furthermore, let $\sigma_w \in N(\mathcal{U})$ be the lowest-dimensional simplex containing $\hat{w}$. While u_α and u_β may not be vertices of σ_w, we can show that $Vert(\sigma_w) \cup \{u_\alpha, u_\beta\}$ must span a simplex $\bar{\sigma}_w$, in $N(\mathcal{U})$ (see full version). Let $\gamma_{a,b}[\hat{x}, \hat{y}]$ be the maximal subpath of $\gamma_{a,b}$ containing $\hat{w}$ that is contained within $|\bar{\sigma}_w|$. One can construct a homotopy H_a that takes $\gamma_{a,b}[\hat{a}, \hat{x}]$ to u_α under which any point $\hat{z} \in \gamma_{a,b}[\hat{a}, \hat{x}]$ moves monotonically along the segment $\hat{z}u_\alpha$ within the geometric realization of the simplex containing both $\hat{z}$ and u_α. See the details in the full version.

Similarly, there is a homotopy H_b that takes $\gamma_{a,b}[\hat{y}, \hat{b}]$ to u_β under which any point $\hat{z} \in \gamma_{a,b}[\hat{y}, \hat{b}]$ moves monotonically along the segment $\hat{z}u_\beta$. Finally, for the middle subpath $\gamma_{a,b}[\hat{x}, \hat{y}]$, since it is within simplex $\bar{\sigma}_w$ with $e = (u_\alpha, u_\beta)$ being an edge of it, we can construct a homotopy H_w that takes $\gamma_{a,b}[\hat{x}, \hat{y}]$ to $|u_\alpha u_\beta|$ under which $\hat{x}$ and $\hat{y}$ move monotonically along the segments $\hat{x}u_\alpha$ and $\hat{y}u_\beta$ within the geometric realization of simplex $\bar{\sigma}_w$, respectively. Concatenating H_a, H_w and H_b, we obtain a homotopy $H_{\alpha,\beta}$ taking $\gamma_{a,b}$ to $|e|$. A concatenation of these homotopies $H_{\alpha,\beta}$ considered over all edges in γ, brings $\phi_\mathcal{U}(\rho(\gamma))$ to $|\gamma|$ with a homotopy in $|N(\mathcal{U})|$. Hence, their homology classes are the same. ◀

▶ **Proposition 23.** *Let z be a 1-cycle in $\mathcal{C}_1(X)$. Then, $[\phi_\mathcal{U}(z)] = 0$ if $\lambda(\mathcal{U}) > s(z)$.*

Proof of Theorem 21.

Proof of (i): By Proposition 23, we have $\phi_{\mathcal{U}*}[z] = [\phi_\mathcal{U}(z)] = 0$ if $\lambda(\mathcal{U}) > s(z)$. This establishes the first part of the assertion because $\bar{\phi}_{\mathcal{U}*} = \iota \circ \phi_{\mathcal{U}*}$ where ι is an isomorphism between the singular homology of $|N(\mathcal{U})|$ and the simplicial homology of $N(U)$. To see the second part, notice that $\bar{\phi}_{\mathcal{U}*}$ is a surjection by Theorem 8. Therefore, the classes $\bar{\phi}_{\mathcal{U}*}(z)$ where $\lambda(\mathcal{U}) \not> s(z)$ contain a basis for $H_1(N(\mathcal{U}))$. Hence they generate it.

Proof of (ii): Suppose on the contrary, there is a subsequence $\{\ell_1, \ldots, \ell_t\} \subset \{\ell', \ldots, g\}$ such that $\Sigma_{j=1}^{t} [\phi_\mathcal{U}(z_{\ell_j})] = 0$. Let $z = \Sigma_{j=1}^{t} \phi_\mathcal{U}(z_{\ell_j})$. Let γ be a 1-cycle in $N(\mathcal{U})$ so that $[z] = [|\gamma|]$ whose existence is guaranteed by Proposition 6. It must be the case that there is a 2-chain D in $N(\mathcal{U})$ so that $\partial D = \gamma$. Consider a triangle $t = \{u_{\alpha_1}, u_{\alpha_2}, u_{\alpha_3}\}$ contributing to D. Let $a_i' = \phi_\mathcal{U}^{-1}(u_{\alpha_i})$. Since t appears in $N(\mathcal{U})$, the covers $U_{\alpha_1}, U_{\alpha_2}, U_{\alpha_3}$

containing a'_1, a'_2, and a'_3 respectively have a common intersection in X. This also means that each of the paths $a'_1 \rightsquigarrow a'_2$, $a'_2 \rightsquigarrow a'_3$, $a'_3 \rightsquigarrow a'_1$ has size at most $2s_{max}(\mathcal{U})$. Then, $\rho(\partial t)$ is mapped to a 1-cycle in X of size at most $4s_{max}(\mathcal{U})$. It follows that $\rho(\partial D)$ can be written as a linear combination of cycles of size at most $4s_{max}(\mathcal{U})$. Each of the 1-cycles of size at most $4s_{max}(\mathcal{U})$ is generated by basis elements $z_1, \ldots, z_k$ where $s(z_k) \leq 4s_{max}(\mathcal{U})$. Therefore, the class of $z' = \phi_{\mathcal{U}}(\rho(\gamma))$ is generated by a linear combination of the basis elements whose preimages have size at most $4s_{max}(\mathcal{U})$. The class $[z']$ is same as the class $[|\gamma|]$ by Proposition 22. But, by assumption $[|\gamma|] = [z]$ is generated by a linear combination of the basis elements whose sizes are larger than $4s_{max}(\mathcal{U})$ reaching a contradiction. ◀

4.2 H_1-classes in Reeb space

In this section we prove an analogue of Theorem 21 for Reeb spaces, which to our knowledge is new. The Reeb space of a function $f : X \to Z$, denoted R_f, is the quotient of X under the equivalence relation $x \sim_f x'$ if and only if $f(x) = f(x')$ and there exists a continuous path $\gamma \in \Gamma_X(x, x')$ such that $f \circ \gamma$ is constant. The induced quotient map is denoted $q : X \to R_f$ which is of course surjective. We show that q_* at the homology level is also surjective for H_1 when the codomain Z of f is a metric space. In fact, we prove a stronger statement: only 'vertical' homology classes (classes with strictly positive size) survive in a Reeb space which extends the result of Dey and Wang [11] for Reeb graphs.

Let $\mathcal{V}$ be a path-connected cover of R_f. This induces a pullback cover denoted $\mathcal{U} = \{U_\alpha\}_{\alpha \in A} = \{q^{-1}(V_\alpha)\}_{\alpha \in A}$ on X. Let $N(\mathcal{U})$ and $N(\mathcal{V})$ denote the corresponding nerve complexes of $\mathcal{U}$ and $\mathcal{V}$ respectively. It is easy to see that $N(\mathcal{U}) \cong N(\mathcal{V})$ because $U_\alpha \cap U_{\alpha'} \neq \emptyset$ if and only if $V_\alpha \cap V_{\alpha'} \neq \emptyset$. There are nerve maps $\phi_{\mathcal{V}} : R_f \to |N(\mathcal{V})|$ and $\phi_{\mathcal{U}} : X \to |N(\mathcal{U})|$ so that the following holds:

▶ **Proposition 24.** *Consider the sequence* $X \xrightarrow{q} R_f(X) \xrightarrow{\phi_{\mathcal{V}}} |N(\mathcal{V})| = |N(\mathcal{U})|$. *Then,* $\phi_{\mathcal{U}} = \phi_{\mathcal{V}} \circ q$.

Let the codomain of the function $f : X \to Z$ be a *metric space* (Z, d_Z). We first impose a pseudometric on X induced by f; the one-dimensional version of this pseudometric is similar to the one used in [1] for Reeb graphs. Recall that given two points $x, x' \in X$ we denote by $\Gamma_X(x, x')$ the set of all continuous paths $\gamma : [0, 1] \to X$ such that $\gamma(0) = x$ and $\gamma(1) = x'$.

▶ **Definition 25.** We define a pseudometric d_f on X as follows: for $x, x' \in X$,

$$d_f(x, x') := \inf_{\gamma \in \Gamma_X(x, x')} \mathrm{diam}_Z(f \circ \gamma).$$

▶ **Proposition 26.** $d_f : X \times X \to \mathbb{R}_+$ *is a pseudometric.*

Similar to X, we endow R_f with a distance $\tilde{d}_f$ that descends via the map q: for any equivalence classes $r, r' \in R_f$, pick $x, x' \in X$ with $r = q(x)$ and $r' = q(x')$, then define

$$\tilde{d}_f(r, r') := d_f(x, x').$$

The definition does not depend on the representatives x and x' chosen. In this manner we obtain the pseudometric space $(R_f, \tilde{d}_f)$. Let $z_1, \ldots, z_g$ be a minimal generator basis of $H_1(X)$ defined with respect to the pseudometric d_f and $q : X \to R_f$ be the quotient map.

▶ **Theorem 27.** *Let* $\ell \in [1, g]$ *be the smallest integer so that* $s(z_\ell) \neq 0$. *If no such* ℓ *exists,* $H_1(R_f)$ *is trivial, otherwise,* $\{[q(z_i)]\}_{i=\ell,\ldots g}$ *is a basis for* $H_1(R_f)$.

Proof. Consider the sequence $X \xrightarrow{q} R_f \xrightarrow{\phi_{\mathcal{V}}} |N(\mathcal{V})|$ where $\mathcal{V}$ is a cover of R_f. It is shown in the full version that q_* is a surjection for H_1-homology. Then, $\{[q(z_i)]\}_{i=1,\ldots,g}$ generate $H_1(R_f)$. First, assume that ℓ as stated in the theorem exists. Let the cover $\mathcal{V}$ be fine enough so that $0 < s_{\max}(\mathcal{U}) \leq \delta$ where $\delta = \frac{1}{4}\min\{s(z_i)\,|\,s(z_i) \neq 0\}$. Then, by applying Theorem 21(ii), we obtain that $[\phi_{\mathcal{U}}(z_i)]_{i=\ell,\ldots,g}$ are linearly independent in $H_1(|N(\mathcal{U})|) = H_1(|N(\mathcal{V})|$. Since $[\phi_{\mathcal{U}}(z_i)] = [\phi_{\mathcal{V}} \circ q(z_i)]$ by Proposition 24, $\{[q(z_i)]\}_{i=\ell,\ldots,g}$ are linearly independent in $H_1(R_f)$. But, $[q(z_i)] = 0$ for $s(z_i) = 0$ and $\{[q(z_i)]\}_{i=1,\ldots,g}$ generate $H_1(R_f)$. Therefore, $\{[q(z_i)]\}_{i=\ell,\ldots,g}$ is a basis. In the case when ℓ does not exist, we have $s(z_i) = 0$ for every $i \in [1, g]$. Then, $[q(z_i)] = 0$ for every i rendering $H_1(R_f)$ trivial. ◀

4.3 Persistence of H_1-classes in mapper and multiscale mapper

To apply the results for nerves in section 4.1 to mappers and multiscale mappers, the Lebesgue number of the pullback covers of X becomes important. The following observation in this respect is useful. Remember that the size of a subset in X and hence the cover elements are measured with respect to the pseudometric d_f.

▶ **Proposition 28.** *Let $\mathcal{U}$ be a cover for the codomain Z. Then, the pullback cover $f^*\mathcal{U}$ has Lebesgue number $\lambda(\mathcal{U})$.*

Notice that the smallest size $s_{min}(f^*\mathcal{U})$ of an element of the pullback cover can be arbitrarily small even if $s_{min}(\mathcal{U})$ is not. However, the Lebesgue number of $\mathcal{U}$ can be leveraged for the mapper due to the above proposition.

Given a cover $\mathcal{U}$ of Z, consider the mapper $N(f^*\mathcal{U})$. Let $z_1, \ldots, z_g$ be a set of minimal generator basis for $H_1(X)$ where the metric in question is d_f. Then, as a consequence of Theorem 21 we have:

▶ **Theorem 29.**

(i) *Let $\ell = g + 1$ if $\lambda(\mathcal{U}) > s(z_g)$. Otherwise, let $\ell \in [1, g]$ be the smallest integer so that $s(z_\ell) > \lambda(\mathcal{U})$. If $\ell \neq 1$, the class $\phi_{\mathcal{U}*}[z_j] = 0$ for $j = 1, \ldots, \ell - 1$. Moreover, if $\ell \neq g + 1$, the classes $\{\phi_{\mathcal{U}*}[z_j]\}_{j=\ell,\ldots,g}$ generate $H_1(N(f^*\mathcal{U}))$.*

(ii) *The classes $\{\phi_{\mathcal{U}*}[z_j]\}_{j=\ell',\ldots,g}$ are linearly independent where $s(z_{\ell'}) > 4s_{max}(\mathcal{U})$.*

(iii) *Consider a H_1-persistence module of a multiscale mapper induced by a tower of path connected covers:*

$$\mathrm{H}_1\big(N(f^*\mathcal{U}_{\varepsilon_0})\big) \xrightarrow{s_{1*}} \mathrm{H}_1\big(N(f^*\mathcal{U}_{\varepsilon_1})\big) \xrightarrow{s_{2*}} \cdots \xrightarrow{s_{n*}} \mathrm{H}_1\big(N(f^*\mathcal{U}_{\varepsilon_n})\big)\,. \tag{3}$$

Let $\hat{s}_{i*} = s_{i*} \circ s_{(i-1)*} \circ \cdots \circ \bar{\phi}_{\mathcal{U}_{\varepsilon_0}*}$. Then, the assertions in (i) and (ii) hold for $H_1(N(f^*\mathcal{U}_{\varepsilon_i}))$ with the map $\hat{s}_{i*} : X \to N(f^*\mathcal{U}_{\varepsilon_i})$.

▶ Remark (Persistence diagram approximation.). The persistence diagram of the H_1-persistence module considered in Theorem 29(iii) contains points whose birth coordinates are exactly zero. This is because all connecting maps are surjective by (i) and thus every class is born only at the beginning. The death coordinate of a point that corresponds to a minimal basis generator of size s is in between the index ε_i and ε_j where $s \geq 4s_{max}(\mathcal{U}_{\varepsilon_i})$ and $s \leq \lambda(\mathcal{U}_{\varepsilon_j})$ because of the assertions (i) and (ii) in Theorem 29. Assuming covers whose λ and s_{max} values are within a constant factor of each other (such as the ones described in next subsection), we can conclude that a generator of size s dies at some point cs for some constant c. Therefore, by computing a minimal generator basis of $N(\mathcal{U}_{\varepsilon_0})$ and computing their sizes provide a 4-approximation to the persistence diagram of the multiscale mapper in the log scale.

4.4 Two special covers and intrinsic Čech complex

We discuss two special covers, one can be effectively computed and the other one is relevant in the context of the intrinsic Čech complex of a metric space. We say a cover $\mathcal{U}$ of a metric space (Y, d) is (α, β)-cover if $\alpha \leq \lambda(\mathcal{U})$ and $\beta \geq s_{max}(\mathcal{U})$.

A $(\delta, 4\delta)$-cover: Consider a δ-sample P of Y, that is, every metric ball $B(y, \delta)$, $y \in Y$, contains a point in P. Observe that the cover $\mathcal{U} = \{B(p, 2\delta)\}_{p \in P}$ is a $(\delta, 4\delta)$-cover for Z. Clearly, $s_{max}(\mathcal{U}) \leq 4\delta$. To determine $\lambda(\mathcal{U})$, consider any subset $Y' \subseteq Y$ with $s(Y') \leq \delta$. There is a $p \in P$ so that $d_Y(p, Y') \leq \delta$. Let y' be the furthest point in Y' from p. Then, $d_Y(p, y') \leq d_Y(p, Y) + \mathrm{diam}(Y') \leq 2\delta$ establishing that $\lambda(\mathcal{U}) \geq \delta$.

A $(\delta, 2\delta)$-cover: Consider the infinite cover $\mathcal{U}$ of Y where $\mathcal{U} = \{B(y, \delta)\}_{y \in Y}$. These are the set of all metric balls of radius δ. Clearly, $s_{max}(\mathcal{U}) \leq 2\delta$. Any subset $Y' \subseteq Y$ with $s(Y') \leq \delta$ is contained in a ball $B(y, \delta)$ where y is any point in Y'. This shows that $\lambda(\mathcal{U}) \geq \delta$. A consequence of this observation and Theorem 21 is that the intrinsic Čech complexes satisfy some interesting property.

▶ **Definition 30.** Given a metric space (Y, d_Y), its intrinsic Čech complex $C^\delta(Y)$ at scale δ is defined to be the nerve complex of the set of intrinsic δ-balls $\{B(y, \delta)\}_{y \in Y}$.

▶ **Observation 31.** *Let $C^\delta(Y)$ denote the intrinsic Čech complex of a metric space Y at scale δ. Let $\mathcal{U}$ denote the corresponding possibly infinite cover of Y. Let $z_1, \ldots, z_g$ be a minimal generator basis for $H_1(Y)$. Then, $\{\bar{\phi}_{\mathcal{U}*}(z_i)\}_{i=\ell,\ldots,g}$ generate $H_1(C^\delta(Y))$ if ℓ is the smallest integer with $s(z_\ell) > \delta$. Furthermore, $\{\bar{\phi}_{\mathcal{U}*}(z_i)\}_{i=\ell',\ldots,g}$ are linearly independent if $s(z'_\ell) > 8\delta$.*

5 Higher dimensional homology groups

We have already observed that the surjectivity of the map $\phi_{\mathcal{U}*} : H_1(X) \to H_1(|N(\mathcal{U})|)$ in one dimensional homology does not extend to higher dimensional homology groups. This means that we cannot hope for analogues to Theorem 21(i) and Theorem 29 to hold for higher dimensional homology groups. However, under the assumption that $f : X \to Z$ is a continuous map from a compact space to a metric space, we can provide some characterization of the persistent diagrams of the mapper and the multiscale mapper as follows:

- We define a metric d_δ on the vertex set P_δ of $N(\mathcal{U})$ where $s_{\max}(\mathcal{U}) \leq \delta$ and then show that the Gromov-Hausdorff distance between the metric spaces (P_δ, d_δ) and $(R_f, \tilde{d}_f)$ is at most 5δ. The same proof also applies if we replace $(R_f, \tilde{d}_f)$ with the pseudometric space (X, d_f). See the full version.
- Previous result implies that the persistence diagrams of the intrinsic Čech complex of the metric space (X, d_f) and that of the metric space (P_δ, d_δ) have a bottleneck distance of $O(\delta)$. This further implies that the persistence diagram of the mapper structure $N(\mathcal{U})$ (approximated as the metric structure (P_δ, d_δ)) is close to that of the intrinsic Čech complex of the pseudometric space (X, d_f).
- We show that the intrinsic Čech complexes of (X, d_f) interleave with $\mathrm{MM}(\mathfrak{U}, f)$ thus connecting their persistence diagrams. See Section 5.1.
- It follows that the persistence diagrams of the multiscale mapper $\mathrm{MM}(\mathfrak{U}, f)$ and (P_δ, d_δ) are close, both being close to that of (X, d_f). This shows that the multiscale mapper encodes similar information as the mapper under an appropriate map-induced metric.

▶ **Definition 32** (Intrinsic Čech filtration). The *intrinsic Čech filtration of the metric space* (Y, d_Y) is

$$\mathfrak{C}(Y) = \{C^r(Y) \subseteq C^{r'}(Y)\}_{0<r<r'}.$$

The *intrinsic Čech filtration at resolution s* is defined as $\mathfrak{C}_s(Y) = \{C^r(Y) \subseteq C^{r'}(Y)\}_{s\leq r<r'}$.

Whenever (Y, d_Y) is totally bounded, the persistence modules induced by taking homology of this intrinsic Čech filtration become q-tame [7]. This implies that one may define its persistence diagram Dg $\mathfrak{C}(Y)$ which provides one way to summarize the topological information of the space Y through the lens of its metric structure d_Y.

We argue that the pseudometric space (X, d_f) is totally bounded. This requires us to show that for any $\varepsilon > 0$ there is a finite subset of $P \subseteq X$ so that open balls centered at points in P with radii ε cover X. Recall that we have assumed that X is a compact topological space, that (Z, d_Z) is a metric space, and that $f : X \to Z$ is a continuous map. Consider a cover $\mathcal{U}$ of Z where each cover element is a ball of radius most $\varepsilon/2$ around a point in Z. Then, the pullback cover $f^*\mathcal{U}$ of X has all elements with diameter at most ε in the metric d_f. Since X is compact, a finite sub-cover of $f^*\mathcal{U}$ still covers X. A finite set P consisting of one arbitrary point in each element of this finite sub-cover is such that the union of d_f-balls of radius ϵ around points in P covers X. Since $\varepsilon > 0$ was arbitrary, (X, d_f) is totally bounded.

Consider the mapper $N(f^*\mathcal{U})$ w.r.t a cover $\mathcal{U}$ of the codomain Z. We can equip its vertex set, denoted by P_δ, with a metric structure (P_δ, d_δ), where δ is an upper bound on the diameter of each element in $\mathcal{U}$. Hence we can view the persistence diagram Dg $\mathfrak{C}(P_\delta)$ w.r.t. the metric d_δ as a summary of the mapper $N(f^*\mathcal{U})$. Using the Gromov-Hausdorff distance between the metric spaces (P_δ, d_δ) and (X, d_f), we relate this persistent summary to the persistence diagram Dg $\mathfrak{C}(X)$ induced by the intrinsic Čech filtration of (X, d_f). Specifically, we show that $d_{GH}((P_\delta, d_\delta), (X, d_f)) \leq 5\delta$. Theorem 32 in the full version extends to this result. With (X, d_f) being totally bounded, by results of [7], it follows that the bottleneck-distance between the two resulting persistence diagrams satisfies:

$$d_B(\text{Dg } \mathfrak{C}(P_\delta), \text{Dg } \mathfrak{C}(X)) \leq 2 * 5\delta = 10\delta. \tag{4}$$

5.1 $\mathrm{MM}(\mathfrak{W}, f)$ for a tower of covers $\mathfrak{W}$

Above we discussed the information encoded in a certain persistence diagram summary of a single Mapper structure. We now consider the persistent homology of multiscale mappers. Given any tower of covers (TOC) $\mathfrak{W}$ of the co-domain Z, by applying the homology functor to its multiscale mapper $\mathrm{MM}(\mathfrak{W}, f)$, we obtain a persistent module, and we can thus discuss the persistent homology induced by a tower of covers $\mathfrak{W}$. However, as discussed in [10], this persistent module is not necessarily stable under perturbations (of e.g the map f) for general TOCs. To address this issue, Dey et al. introduced a special family of the so-called (c,s)-good TOC in [10], which is natural and still general. Below we provide an equivalent definition of the (c,s)-good TOC based on the Lebesgue number of covers.

▶ **Definition 33** ((c, s)-good TOC). Give a tower of covers $\mathfrak{U} = \{\mathcal{U}_\varepsilon\}_{\varepsilon\geq s}$, we say that it is *(c,s)-good TOC* if for any $\varepsilon \geq s$, we have that (i) $s_{max}(\mathcal{U}_\varepsilon) \leq \varepsilon$ and (ii) $\lambda(\mathcal{U}_{c\varepsilon}) \geq \varepsilon$.

As an example, the TOC $\mathfrak{U} = \{\mathcal{U}_\varepsilon\}_{\varepsilon\geq s}$ with $\mathcal{U}_\varepsilon := \{B_{\varepsilon/2}(z) \mid z \in Z\}$ is an (2,s)-good TOC of the co-domain Z.

We now characterize the persistent homology of multiscale mappers induced by (c,s)-good TOCs. Connecting these persistence modules is achieved via the interleaving of towers of simplicial complexes originally introduced in [5]. Below we include the slightly generalized version of the definition from [10].

▶ **Definition 34** (Interleaving of simplicial towers, [10])**.** Let $\mathfrak{S} = \left\{\mathcal{S}_\varepsilon \xrightarrow{s_{\varepsilon,\varepsilon'}} \mathcal{S}_{\varepsilon'}\right\}_{r\le\varepsilon\le\varepsilon'}$ and $\mathfrak{T} = \left\{\mathcal{T}_\varepsilon \xrightarrow{t_{\varepsilon,\varepsilon'}} \mathcal{T}_{\varepsilon'}\right\}_{r\le\varepsilon\le\varepsilon'}$ be two towers of simplicial complexes where $\mathrm{res}(\mathfrak{S}) = \mathrm{res}(\mathfrak{T}) = r$. For some $c \ge 0$, we say that they are *c-interleaved* if for each $\varepsilon \ge r$ one can find simplicial maps $\varphi_\varepsilon : \mathcal{S}_\varepsilon \to \mathcal{T}_{\varepsilon+c}$ and $\psi_\varepsilon : \mathcal{T}_\varepsilon \to \mathcal{S}_{\varepsilon+c}$ so that:

(i) for all $\varepsilon \ge r$, $\psi_{\varepsilon+c} \circ \varphi_\varepsilon$ and $s_{\varepsilon,\varepsilon+2c}$ are contiguous,
(ii) for all $\varepsilon \ge r$, $\varphi_{\varepsilon+\eta} \circ \psi_\varepsilon$ and $t_{\varepsilon,\varepsilon+2c}$ are contiguous,
(iii) for all $\varepsilon' \ge \varepsilon \ge r$, $\varphi_{\varepsilon'} \circ s_{\varepsilon,\varepsilon'}$ and $t_{\varepsilon+c,\varepsilon'+c} \circ \varphi_\varepsilon$ are contiguous,
(iv) for all $\varepsilon' \ge \varepsilon \ge r$, $s_{\varepsilon+c,\varepsilon'+c} \circ \psi_\varepsilon$ and $\psi_{\varepsilon'} \circ t_{\varepsilon,\varepsilon'}$ are contiguous.

Analogously, if we replace the operator '+' by the multiplication '$\cdot$' in the above definition, then we say that $\mathfrak{S}$ and $\mathfrak{T}$ are *c-multiplicatively interleaved.*

Our main results of this section are the following whose proofs are deferred to the full version. First, Theorem 35 states that the multiscale-mappers induced by any two (c, s)-good towers of covers interleave with each other, implying that their respective persistence diagrams are also close under the bottleneck distance. From this point of view, the persistence diagrams induced by any two (c,s)-good TOCs contain roughly the same information. Next in Theorem 36, we show that the multiscale mapper induced by any (c, s)-good TOC interleaves (at the homology level) with the intrinsic Čech filtration of (X, d_f), thereby implying that the persistence diagram of the multiscale mapper w.r.t. any (c, s)-good TOC is close to that of the intrinsic Čech filtration of (X, d_f) under the bottleneck distance.

▶ **Theorem 35.** *Given a map $f : X \to Z$, let $\mathfrak{V} = \{\mathcal{V}_\varepsilon \xrightarrow{v_{\varepsilon,\varepsilon'}} \mathcal{V}_{\varepsilon'}\}_{\varepsilon\le\varepsilon'}$ and $\mathfrak{W} = \{\mathcal{W}_\varepsilon \xrightarrow{w_{\varepsilon,\varepsilon'}} \mathcal{W}_{\varepsilon'}\}_{\varepsilon\le\varepsilon'}$ be two (c, s)-good tower of covers of Z. Then the corresponding multiscale mappers $\mathrm{MM}(\mathfrak{V}, f)$ and $\mathrm{MM}(\mathfrak{W}, f)$ are c-multiplicatively interleaved.*

▶ **Theorem 36.** *Let $\mathfrak{C}_s(X)$ be the intrinsic Čech filtration of (X, d_f) starting with resolution s. Let $\mathfrak{U} = \{\mathcal{U}_\varepsilon \xrightarrow{u_{\varepsilon,\varepsilon'}} \mathcal{U}_{\varepsilon'}\}_{s\le\varepsilon\le\varepsilon'}$ be a (c, s)-good TOC of the compact connected metric space Z. Then the multiscale mapper $\mathrm{MM}(\mathfrak{U}, f)$ and $\mathfrak{C}_s(X)$ are $2c$-multiplicatively interleaved.*

Finally, given a persistence diagram Dg, we denote its *log-scaled version* $\mathrm{Dg}_{\log}$ to be the diagram consisting of the set of points $\{(\log x, \log y) \mid (x, y) \in \mathrm{Dg}\}$. Since interleaving towers of simplicial complexes induce interleaving persistent modules, using results of [5, 6], we have the following corollary.

▶ **Corollary 37.** *Given a continuous map $f : X \to Z$ and a (c, s)-good TOC $\mathfrak{U}$ of Z, let $\mathrm{Dg}_{\log}\mathrm{MM}(\mathfrak{U}, f)$ and $\mathrm{Dg}_{\log}\mathfrak{C}_s$ denote the log-scaled persistence diagram of the persistence modules induced by $\mathrm{MM}(\mathfrak{U}, f)$ and by the intrinsic Čech filtration $\mathfrak{C}_s$ of (X, d_f) respectively. We have that*

$$d_B(\mathrm{Dg}_{\log}\mathrm{MM}(\mathfrak{U}, f), \mathrm{Dg}_{\log}\mathfrak{C}_s) \le 2c.$$

Acknowledgments. We thank the reviewers for helpful comments.

References

1 U. Bauer, X. Ge and Y. Wang. Measuring distance between Reeb graphs. *Proc. 30th Annual Symp. Comput. Geom., SoCG* (2014), 464–473.

2 S. Biasotti, D. Giorgi, M. Spagnuolo, and B. Falcidieno. Reeb graphs for shape analysis and applications. *Theor. Comput. Sci.*, 392(1-3):5–22, 2008.

3 K. Borsuk. On the imbedding of systems of compacta in simplicial complexes. *Fund. Math.* 35 (1948), 217–234.

4 M. Carrière and S. Y. Oudot. Structure and Stability of the 1-Dimensional Mapper. *Proc. 32nd Int'l Symp. Comput. Geom., SoCG* (2016), 25:1–25:16.

5 F. Chazal, D. Cohen-Steiner, M. Glisse, L. Guibas, and S. Oudot. Proximity of persistence modules and their diagrams. *Proc. 25th Annual Symp. Comput. Geom., SoCG* (2009), 237–246.

6 F. Chazal, V. de Silva, M. Glisse, and S. Oudot. The structure and stability of persistence modules. *SpringerBriefs in Mathematics,* eBook ISBN 978-3-319-42545-0, Springer, 2016.

7 F. Chazal, V. de Silva, and S. Oudot. Persistence stability for geometric complexes. *Geometric Dedicata*, 173(1):193–214, 2014.

8 F. Chazal and J. Sun. Gromov-Hausdorff approximation of filament structure using Reeb-type graph. *Proc. 30th Annual Symp. Comput. Geom., SoCG* (2014), 491–500.

9 V. de Silva, E. Munch, and A. Patel. Categorified Reeb graphs. ArXiv preprint arXiv:1501.04147, (2015).

10 T. K. Dey, F. Mémoli, and Y. Wang. Multiscale mapper: Topological summarization via codomain covers. *ACM-SIAM Symp. Discrete Alg., SODA* (2016), 997–1013.

11 T. K. Dey and Y. Wang. Reeb graphs: Approximation and persistence. *Discrete Comput. Geom.* 49 (2013), 46–73.

12 H. Edelsbrunner and J. Harer. *Computational Topology: An Introduction.* Amer. Math. Soc., Providence, Rhode Island, 2009.

13 H. Edelsbrunner, J. Harer, and A. K. Patel. Reeb spaces of piecewise linear mappings. In *Proc. 24th Annual Symp. Comput. Geom., SoCG* (2008), 242–250.

14 E. Gasparovic, M. Gommel, E. Purvine, R. Sazdanovic, B. Wang, Y. Wang and L. Ziegelmeier. A complete characterization of the one-dimensional intrinsic Čech persistence diagrams for metric graphs. *Manuscript, an earlier version appeared as a report for IMA Workshop for Women in Computational Topology (WinCompTop)*, 2016.

15 A. Hatcher. Algebraic Topology. Cambridge U. Press, New York, 2002.

16 J. Leray. L'anneau spectral et l'anneau filtré d'homologie d'un espace localement compact et d'une application continue. *J. Math. Pures Appl.* 29 (1950), 1–139.

17 P. Y. Lum, G. Singh, A. Lehman, T. Ishkhanikov, M. Vejdemo-Johansson, M. Alagappan, J. Carlsson, and G. Carlsson. Extracting insights from the shape of complex data using topology. *Scientific reports* 3 (2013).

18 E. Munch and B. Wang. Convergence between categorical representations of Reeb space and mapper. *32nd Int'l Symp. Comput. Geom., SoCG* (2016), 53:1–53:16.

19 J. R. Munkres, *Topology*, Prentice-Hall, Inc., New Jersey, 2000.

20 V. Prasolov. *Elements of combinatorial and differential topology.* American Mathematical Soc., Vol. 74, 2006.

21 G. Singh, F. Mémoli, and G. Carlsson. Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition. *Symp. Point Based Graphics*, 2007.