Computational Topology for Data Analysis: Notes from Book by

Tamal Krishna Dey Department of Computer Science Purdue University West Lafayette, Indiana, USA 46907

Yusu Wang Halıcıoğlu Data Science Institute University of California, San Diego La Jolla, California, USA 92093

Topic 13: Nerve, Cover, Mapper

Data can be complex both in terms of the domain where they come from and in terms of properties/observations associated with them which are often modeled as functions/maps. For example, we can have a set of patients, where each patient is associated with multiple biological markers, giving rise to a multivariate function from the space of patients to an image domain that may or may not be the Euclidean space. To this end, we need to analyze not only real-valued scalar fields as we did in so far in the book, but also more complex maps defined on a given domain, such as multivariate, circle valued, sphere valued maps, etc.



Figure 13.1: The function values on a hand model are binned into intervals as indicated by different colors. The mapper corresponding to these intervals (cover) is shown with the graph below; picture is taken from [15].

One way to analyze complex maps is to use the *Mapper* methodology introduced by Singh et al. in [15]. In particular, given a map $f : X \to Z$, the mapper M(f, U) creates a topological metaphor for the structure behind f by pulling back a cover U of the space Z to a cover on X through f. This mapper methodology can work with any (reasonably tame) continuous maps between two topological spaces. It converts complex maps and covers of the target space into simplicial complexes, which are much easier to process computationally. One can view the map f and a finite cover of the space Z as the lens through which the input data X is examined. It is in some sense related to Reeb graphs which also summarizes f but without any particular attention to a cover of the codomain. Figure 13.1 shows a mapper construction where the reader can see its similarity to the Reeb graph. The choice of different maps and covers allows the user to capture different aspects of the input data. The mapper methodology has been successfully applied to analyzing various types of data, we have shown an example in the Prelude of the book, for others see e.g. [10, 13].

To understand the Mapper and its multiscale version *Multiscale Mapper* better, we study first some properties of nerves as they are at the core of these constructions. We already know Nerve Theorem which states that if every intersection of cover elements in a cover \mathcal{U} is contractible, then the nerve $N(\mathcal{U})$ is homotopy equivalent to the space $X = \bigcup \mathcal{U}$. However, we cannot hope for such a good cover all the time and need to investigate what happens if the cover is not good. Sections 13.1 and 13.2 are devoted to this study. Specifically, we show that if every cover element satisfies a weaker property that it is only path connected, then the nerve may not preserve homotopy, but satisfies a surjectivity property in one-dimensional homology.

One limitation of the mapper is that it is defined with respect to a fixed cover of the target space. Naturally, the behavior of the mapper under a change of cover is of interest because it has the potential to reveal the property of the map at different scales. Keeping this in mind, we study a multiscale version of mapper, which we refer to as *multiscale mapper*. It is capable of producing a multiscale summary in the form of a persistence diagram using a cover of the codomain at different scales. In Section 13.4, we discuss the stability of the multiscale mapper under changes in the input map and/or in the tower \mathfrak{U} of covers. An efficient algorithm for computing mapper and multiscale mapper for a real valued PL-function is presented in Sections 13.5. In Section 13.6, we consider the more general case of a map $f : X \to Z$ where X is a simplicial complex but Z is not necessarily Euclidean. We show that we can use an even simpler combinatorial version of the multiscale mapper, which only acts on *vertex sets* of X with connectivity given by the 1-skeleton graph of X. The cost we pay here is that the resulting persistence diagram *approximates* (instead of computing exactly) the persistence diagram of the standard multiscale mapper if the tower of covers of Z is "good" in certain sense.

13.1 Covers and nerves

In this section we present several facts about covers of a topological space and their nerves. Specifically, we focus on maps between covers and the maps they induce between nerves and their homology groups.

Let X denote a path connected topological space. Recall that by this we mean that there exists a continuous function called path $\gamma : [0, 1] \rightarrow X$ connecting every pair of points $\{x, x'\} \in X \times X$ where $\gamma(0) = x$ and $\gamma(1) = x'$. Also recall that for a topological space X, a collection $\mathcal{U} = \{U_{\alpha}\}_{\alpha \in A}$ of open sets such that $\bigcup_{\alpha \in A} U_{\alpha} = X$ is called an open cover of X. Although it is not required in general, we will always assume that each open cover U_{α} is path connected.

Maps between covers. If we have two covers $\mathcal{U} = \{U_{\alpha}\}_{\alpha \in A}$ and $\mathcal{V} = \{V_{\beta}\}_{\beta \in B}$ of a space 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 every $\alpha \in A$. We abuse the notation ξ to also indicate the map $\mathcal{U} \to \mathcal{V}$. The following proposition connects a map between covers to a simplicial map between their nerves.



Figure 13.2: Cover maps ξ and ζ indicated by solid arrows induce simplicial maps $N(\xi)$ and $N(\zeta)$ whose corresponding vertex maps are indicated by dashed arrows.

Proposition 1. Given a map of covers $\xi : \mathcal{U} \to \mathcal{V}$, there is an induced simplicial map $N(\xi) : N(\mathcal{U}) \to N(\mathcal{V})$ given on vertices by the map ξ .

PROOF. Write $\mathcal{U} = \{U_{\alpha}\}_{\alpha \in A}$ and $\mathcal{V} = \{V_{\beta}\}_{\beta \in B}$. Then, for all $\alpha \in A$ we have $U_{\alpha} \subseteq V_{\xi(\alpha)}$. Now take any $\sigma \in N(\mathcal{U})$. We need to prove that $\xi(\sigma) \in N(\mathcal{V})$. For this observe that

$$\bigcap_{\beta \in \xi(\sigma)} V_{\beta} = \bigcap_{\alpha \in \sigma} V_{\xi(\alpha)} \supseteq \bigcap_{\alpha \in \sigma} U_{\alpha} \neq \emptyset,$$

where the last step follows because $\sigma \in N(\mathcal{U})$.

An example is given in Figure 13.2, where both maps $N(\xi)$ and $N(\zeta)$ are simplicial. Furthermore, if $\mathcal{U} \xrightarrow{\xi} \mathcal{V} \xrightarrow{\zeta} \mathcal{W}$ are three different covers of a topological space with the intervening maps of covers between them, then $N(\zeta \circ \xi) = N(\zeta) \circ N(\xi)$ as well.

The following fact will be very useful later for defining multiscale mappers.

Proposition 2 (Induced maps are contiguous). 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.

PROOF. Write $\mathcal{U} = \{U_{\alpha}\}_{\alpha \in A}$ and $\mathcal{V} = \{V_{\beta}\}_{\beta \in B}$. Then, for all $\alpha \in A$ we have both

$$U_{\alpha} \subseteq V_{\zeta(\alpha)}$$
 and $U_{\alpha} \subseteq V_{\xi(\alpha)}$; $\Rightarrow U_{\alpha} \subseteq V_{\zeta(\alpha)} \cap V_{\xi(\alpha)}$

Now take any $\sigma \in N(\mathcal{U})$. We need to prove that $\zeta(\sigma) \cup \xi(\sigma) \in N(\mathcal{V})$. For this write

$$\bigcap_{\beta \in \zeta(\sigma) \cup \xi(\sigma)} V_{\beta} = \left(\bigcap_{\alpha \in \sigma} V_{\zeta(\alpha)}\right) \cap \left(\bigcap_{\alpha \in \sigma} V_{\xi(\alpha)}\right)$$
$$= \bigcap_{\alpha \in \sigma} \left(V_{\zeta(\alpha)} \cap V_{\xi(\alpha)}\right) \supseteq \bigcap_{\alpha \in \sigma} U_{\alpha} \neq \emptyset,$$

where the last step follows from assuming that $\sigma \in N(\mathcal{U})$. It implies that the vertices in $\zeta(\sigma) \cup \xi(\sigma)$ span a simplex in $N(\mathcal{V})$.

In Figure 13.2, the two maps $N(\xi)$ and $N(\zeta)$ can be verified to be contiguous. Furthermore, contiguous maps induce identical maps at the homology level. Proposition 2 implies that the map $H_*(N(\mathcal{U})) \rightarrow H_*(N(\mathcal{V}))$ thus induced can be deemed canonical.

Maps at homology level. Now we focus on establishing various maps at the homology levels for covers and their nerves. 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 $N(\mathcal{U})$ (through the singular homology of $|N(\mathcal{U})|$). The nerve theorem says that if the elements of \mathcal{U} intersect only in contractible spaces, then $\phi_{\mathcal{U}}$ is a homotopy equivalence and hence $\phi_{\mathcal{U}*}$ is 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 [8]. In absence of such conditions of the cover, simple examples exist to show that $\phi_{\mathcal{U}*}$ could be neither a monophorphism (injection) nor an epimorphism

(surjection). Figure 13.3 gives an example where ϕ_{U_*} is not surjective in H₂. However, for one dimensional homology groups, the map ϕ_{U_*} is necessarily a surjection when each element in the cover U is path connected. We call such a cover U *path connected*. The simplicial maps arising out of cover maps between path connected covers induce a surjection between the 1-st homology groups of two nerve complexes.



Figure 13.3: 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 $\phi_{\mathcal{U}} : X \to N(\mathcal{U})$ may not induce a surjection in H₂.

Blow up space. The proof of the nerve theorem given by Hatcher in [7] uses a construction that connects the two spaces X and $|N(\mathcal{U})|$ via a blow-up 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 as in that proof. Nevertheless, we use a similar 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*; see Figure 13.4(left). Details about the construction of these maps follow.



Figure 13.4: (left) Various maps used for blow up space; (right) example of a blow up space.

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 $U_{\alpha_0,...,\alpha_n} := \bigcap_{i=0}^n U_{\alpha_i}$ consider the product $U_{\alpha_0,...,\alpha_n} \times \Delta_{\alpha_0,...,\alpha_n}^n$, where $\Delta_{\alpha_0,...,\alpha_n}^n$ denotes the *n*-dimensional simplex with vertices $u_{\alpha_0}, \ldots, u_{\alpha_n}$. Consider now the disjoint union

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

together with the following identification: each point $(x, y) \in M$, with $x \in U_{\alpha_0,...,\alpha_n}$ and $y \in [\alpha_0, ..., \widehat{\alpha}_i, ..., \alpha_n] \subset \Delta^n_{\alpha_0,...,\alpha_n}$ is identified with the corresponding point in the product $U_{\alpha_0,...,\widehat{\alpha}_i,...,\alpha_n} \times \Delta_{\alpha_0,...,\widehat{\alpha}_i,...,\alpha_n}$ via the inclusion $U_{\alpha_0,...,\widehat{\alpha}_i,...,\alpha_n} \subset U_{\alpha_0,...,\widehat{\alpha}_i,...,\alpha_n}$. Here $[\alpha_0, ..., \widehat{\alpha}_i, ..., \alpha_n]$ denotes the *i*-th face of the simplex $\Delta^n_{\alpha_0,...,\alpha_n}$. Denote by ~ 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 in Figure 13.4(right).

In what follows we assume that the space X is compact. The main motivation behind restricting X to such spaces is that they admit a condition called partition of unity which we use to establish further results.

Definition 1 (Locally finite). An open cover $\{\mathcal{U}_{\alpha}, \alpha \in A\}$ of *X* is called a refinement of another open cover $\{\mathcal{V}_{\beta}, \beta \in B\}$ of *X* if every element $U_{\alpha} \in \mathcal{U}$ is contained in an element $V_{\beta} \in \mathcal{V}$. Furthermore, \mathcal{U} is called locally finite if every point $x \in X$ has a neighborhood contained in finitely many elements of \mathcal{U} .

Definition 2 (Partition of unity). A collection of real valued continuous functions { $\varphi_{\alpha} : X \rightarrow [0, 1], \alpha \in A$ } 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 the support¹ supp (φ_{α}) of φ_{α} is contained in U_{α} for each $\alpha \in A$.

Fact 1 ([14]). For any open cover $\mathcal{U} = \{U_{\alpha}, \alpha \in A\}$ of a compact space *X*, there exists a partition of unity $\{\varphi_{\alpha}, \alpha \in A\}$ subordinate to \mathcal{U} .

We assume that X is compact and hence for an open cover $\mathcal{U} = \{U_{\alpha}\}_{\alpha}$ of X, we can choose any partition of unity $\{\varphi_{\alpha}, \alpha \in A\}$ subordinate to \mathcal{U} according to Fact 1. For each $x \in X$ such that $x \in U_{\alpha}$, denote by x_{α} the corresponding copy of x residing in $X_{\mathcal{U}}$. For our choice of $\{\varphi_{\alpha}, \alpha \in A\}$, define the map $\zeta : X \to X_{\mathcal{U}}$ as:

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 $x \in X$,

$$\phi_{\mathcal{U}}(x) = \sum_{\alpha \in A} \varphi_{\alpha}(x) \, u_{\alpha}. \tag{13.1}$$

We have the following fact [14, pp. 108]:

Fact 2. ζ is a homotopy equivalence.

¹The support of a real-valued function is the subset of the domain whose image is non-zero.

13.1.1 Special case of H_1

Now, we show that the nerve maps at the homology level are surjective for one dimensional homology groups, namely all homology classes in $N(\mathcal{U})$ arise from those in $X = \bigcup \mathcal{U}$. Furthermore, if we assume that X is equipped with a pseudo-metric, we can define a size for cycles with this pseudo-metric and show that all homology classes with representative cycles having a large enough size survive in the nerve $N(\mathcal{U})$. Note that the result is not true beyond one dimensional homology (recall Figure 13.3).

To prove this result for H₁, first, we make a simple observation that connects the classes in singular homology of |N(U)| to those in the simplicial homology of N(U). The result follows immediately from the isomorphism between singular and simplicial homology induced by the geometric realization; see [12]. Recall that [c] denotes the class of a cycle c. If c is simplicial, |c| denotes its underlying space.

Proposition 3. Every 1-cycle γ in $|N(\mathcal{U})|$ has a 1-cycle γ' in $N(\mathcal{U})$ so that $[\gamma] = [|\gamma'|]$.

Proposition 4. If \mathcal{U} is path connected, $\phi_{\mathcal{U}*} : H_1(X) \to H_1(|N(\mathcal{U})|)$ is a surjection, where $\phi_{\mathcal{U}*}$ is the homomorphism induced by the nerve map defined in Eqn. (13.1).

PROOF. Let $[\gamma]$ be any class in $H_1(|N(\mathcal{U})|)$. Because of Proposition 3, we can assume that $\gamma = |\gamma'|$, where γ' is a 1-cycle in the 1-skeleton of $N(\mathcal{U})$. We will construct a 1-cycle $\gamma_{\mathcal{U}}$ in $X_{\mathcal{U}}$ so that $\pi(\gamma_{\mathcal{U}}) = \gamma$. Assume first that such a $\gamma_{\mathcal{U}}$ can be constructed. Then, consider 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 2. 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 $\pi(\gamma_{\mathcal{U}}) = \gamma = |\gamma'|$. Let $e_0, e_1, \ldots, e_{r-1}, e_r = e_0$ be an ordered sequence of edges on γ' . Recall the construction of the space $X_{\mathcal{U}}$. In that terminology, let $e_i = \Delta_{\alpha_i \alpha_{(i+1)} \mod r}^n$. Let $v_i = e_{(i-1) \mod r} \cap e_i$ for $i \in [0, r-1]$. The vertex $v_i = v_{\alpha_i}$ corresponds to the cover element U_{α_i} where $U_{\alpha_i} \cap U_{\alpha_{(i+1) \mod r}} \neq \emptyset$ for every $i \in [0, r-1]$. Choose a point x_i in the common intersection $U_{\alpha_i} \cap U_{\alpha_{(i+1) \mod r}}$ for every $i \in [0, r-1]$. Then, the *edge* path $\tilde{e}_i = e_i \times x_i$ is in $X_{\mathcal{U}}$ by construction. Also, letting x_{α_i} to be the lift of x_i in the lifted U_{α_i} , we can choose a *vertex* path $x_{\alpha_i} \rightsquigarrow x_{\alpha_{(i+1) \mod r}}$ residing in the lifted U_{α_i} and hence in $X_{\mathcal{U}}$ because U_{α_i} is path connected. Consider the following cycle obtained by concatenating the edge and vertex paths

$$\gamma_{\mathfrak{U}} = \tilde{e}_0 x_{\alpha_0} \rightsquigarrow x_{\alpha_1} \tilde{e}_1 \cdots \tilde{e}_{r-1} x_{\alpha_{r-1}} \rightsquigarrow x_{\alpha_0}$$

By projection, we have $\pi(\tilde{e}_i) = e_i$ for every $i \in [0, r-1]$ and $\pi(x_{\alpha_i} \rightsquigarrow x_{\alpha_{(i+1) \mod r}}) = v_{\alpha_i}$ and thus $\pi(\gamma_{\mathcal{U}}) = \gamma$ as required.

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 $\iota_{\mathfrak{U}} : H_p(|N(\mathfrak{U})|) \rightarrow$ $H_p(N(\mathfrak{U}))$ denotes this isomorphism, we let

$$\bar{\phi}_{\mathcal{U}*} : \mathbf{H}_1(X) \to \mathbf{H}_1(N(\mathcal{U}))$$
 denote the composition $\iota_{\mathcal{U}} \circ \phi_{\mathcal{U}*}$. (13.2)

As a corollary to Proposition 4, we obtain:

Theorem 5. If \mathcal{U} is path connected, $\bar{\phi}_{\mathcal{U}*} : H_1(X) \to H_1(N(\mathcal{U}))$ is a surjection.



Figure 13.5: Sequence of cover maps induce a simplicial tower and hence a persistence module: classes in H_1 can only die.

From nerves to nerves. We now extend the result in Theorem 5 to simplicial maps between two nerves induced by cover maps. Figure 13.5 illustrates this fact. The following proposition is key to establishing the result.

Proposition 6 (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 a compact 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

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

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

PROOF. The proof closely follows that of [14, Corollary pp. 97]. Since X is compact, there exists a partition of unity $\{\varphi_{\alpha}\}_{\alpha \in A}$ subordinate to \mathcal{U} . The fact that the sum in the expression of ψ_{β} is well defined and continuous follows from the fact that the family $\{\supp(\varphi_{\alpha})\}_{\alpha}$ is locally finite. Let $C_{\beta} := \bigcup_{\alpha \in \theta^{-1}(\beta)} \supp(\varphi_{\alpha})$. The set C_{β} is closed, $C_{\beta} \subset U_{\beta}$, and $\psi_{\beta}(x) = 0$ for $x \notin C_{\beta}$ so that $\supp(\psi_{\beta}) \subset C_{\beta} \subset V_{\beta}$. Now, to check that the family $\{C_{\beta}\}_{\beta \in B}$ is locally finite pick any point $x \in X$. Since $\{\supp(\varphi_{\alpha})\}_{\alpha}$ is locally finite there is an open set O containing x such that O intersects only finitely many elements in \mathcal{U} . Denote these cover elements by $U_{\alpha_1}, \ldots, U_{\alpha_\ell}$. Now, notice if $\beta \in B$ and $\beta \notin \{\theta(\alpha_i), i = 1, \ldots, \ell\}$, then O does not intersect C_{β} . Then, the family $\{\supp(\psi_{\beta})\}_{\beta \in B}$ is locally finite. It then follows that for $x \in X$ one has

$$\sum_{\beta \in B} \psi_{\beta}(x) = \sum_{\beta \in B} \sum_{\alpha \in \theta^{-1}(\beta)} \varphi_{\alpha}(x) = \sum_{\alpha \in A} \varphi_{\alpha}(x) = 1.$$

We have obtained that $\{\psi_{\beta}\}_{\beta \in B}$ is a partition of unity subordinate to \mathcal{V} as needed by the proposition.

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 $\theta : A \to B$. Apply Proposition 6 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 Eqn. (13.1) using these coherent partitions of unity. Let $N(\mathcal{U}) \xrightarrow{\tau} N(\mathcal{V})$ be the simplicial map induced by the cover map θ . The map τ can be extended to a (linear) continuous map $\hat{\tau} : |N(\mathcal{U})| \to |N(\mathcal{V})|$ by assigning $y \in |N(\mathcal{U})|$ to $\hat{\tau}(y) \in |N(\mathcal{V})|$ where

$$y = \sum_{\alpha} t_{\alpha} u_{\alpha} \implies \hat{\tau}(y) = \sum_{\alpha} t_{\alpha} \hat{\tau}(u_{\alpha}), \text{ with } \sum_{\alpha} t_{\alpha} = 1.$$

Claim 1. The map $\hat{\tau}$ satisfies the property that, for $x \in X$, $\hat{\tau}(\phi_{U}(x)) = \phi_{V}(x)$.

PROOF. For any point $x \in X$, one has $\phi_{\mathcal{U}}(x) = \sum_{\alpha \in A} \varphi_{\alpha}(x) u_{\alpha}$ where u_{α} is the vertex corresponding to $U_{\alpha} \in \mathcal{U}$ in $|N(\mathcal{U})|$. Then,

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

An immediate corollary of the above claim is:

Corollary 7. The induced maps of $\phi_{\mathcal{U}*} : H_p(X) \to H_p(|N(\mathcal{U})|), \phi_{\mathcal{V}*} : H_p(X) \to H_p(|N(\mathcal{V})|), and <math>\hat{\tau}_* : H_p(|N(\mathcal{U})|) \to H_p(|N(\mathcal{V})|)$ commute, that is, $\phi_{\mathcal{V}*} = \hat{\tau}_* \circ \phi_{\mathcal{U}*}$.



Figure 13.6: Maps relevant for Proposition 8; $\bar{\phi}_{\mathcal{V}*} = \iota_{\mathcal{V}} \circ \phi_{\mathcal{V}*}$ and $\bar{\phi}_{\mathcal{U}*} = \iota_{\mathcal{U}} \circ \phi_{\mathcal{U}*}$. The triangular 'roof' and the square 'room' commute, so does the entire 'house'.

With the fact that isomorphism between singular and simplicial homology commutes with simplicial maps and their linear continuous extensions, Corollary 7 implies that:

Proposition 8. $\bar{\phi}_{\mathcal{V}*} = \tau_* \circ \bar{\phi}_{\mathcal{U}*}$ where $\bar{\phi}_{\mathcal{V}*} : \mathsf{H}_p(X) \to \mathsf{H}_p(N(\mathcal{V})), \ \bar{\phi}_{\mathcal{U}*} : \mathsf{H}_p(X) \to \mathsf{H}_p(N(\mathcal{U}))$ and $\tau : N(\mathcal{U}) \to N(\mathcal{V})$ is the simplicial map induced by a cover map $\mathcal{U} \to \mathcal{V}$.

PROOF. Consider the diagram in Figure 13.6. The upper triangle commutes by Corollary 7. The bottom square commutes by the property of simplicial maps, see Theorem 34.4 in [12]. The claim in the proposition follows by combining these two commutating subdiagrams. □ Proposition 8 extends Theorem 5 to the simplicial maps between two nerves.

Theorem 9. Let $\tau : N(\mathfrak{U}) \to N(\mathfrak{V})$ be a simplicial map induced by a cover map $\mathfrak{U} \to \mathfrak{V}$ where both \mathfrak{U} and \mathfrak{V} are path connected. Then, $\tau_* : H_1(N(\mathfrak{U})) \to H_1(N(\mathfrak{V}))$ is a surjection.

PROOF. Consider the maps

$$\mathsf{H}_{1}(X) \xrightarrow{\bar{\phi}_{\mathcal{U}^{*}}} \mathsf{H}_{1}(N(\mathcal{U})) \xrightarrow{\tau_{*}} \mathsf{H}_{1}(N(\mathcal{V})), \text{ and } \mathsf{H}_{1}(X) \xrightarrow{\bar{\phi}_{\mathcal{V}^{*}}} \mathsf{H}_{1}(N(\mathcal{V})).$$

By Proposition 8, $\tau_* \circ \bar{\phi}_{U*} = \bar{\phi}_{\mathcal{V}*}$. By Theorem 5, the map $\bar{\phi}_{\mathcal{V}*}$ is a surjection. It follows that τ_* is a surjection.

13.2 Analysis of persistent H₁-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 further characterize 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 using a pseudometric that the space X is assumed to be equipped with. Precise statements are made in the subsections. Let (X, d) be a pseudometric space, meaning that d satisfies the axioms of a metric except the first axiom, that is, d(x, x') = 0 may not necessarily imply x = x'. Assume X is compact. We define a 'size' for a homology class that reflects how big the smallest cycle in the class is w.r.t. the metric d.

Definition 3. 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 $c \in H_p(X)$ is defined as $s(c) = \inf_{z \in c} s(z)$.

Recall that a set of *p*-cycles $z_1, z_2, ..., z_n$ of $H_p(X)$ is called a cycle basis if the classes $[z_1], [z_2], ..., [z_n]$ together form a basis of $H_p(X)$. It is called an optimal cycle basis if $\sum_{i=1}^n s(z_i)$ is minimal among all cycle 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 real 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, the Lebesgue number $\lambda(\mathcal{U})$ of \mathcal{U} is defined as:

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

As we will see below, a homology class of size no more than $\lambda(\mathcal{U})$ cannot survive in the nerve (Proposition 12). 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

$$s_{max}(\mathcal{U}) := \max_{U \in \mathcal{U}} \{ s(U) \}.$$

Theorem 10 summarizes these observations.

Let z_1, z_2, \ldots, z_g be a non-decreasing sequence of the cycles with respect to their sizes in an optimal cycle basis of $H_1(X)$. Consider the map $\phi_{\mathcal{U}} : X \to |N(\mathcal{U})|$ as introduced in Eqn. (13.1), and the map $\bar{\phi}_{\mathcal{U}*}$ as defined by Eqn. (13.2). We have the following result.

Theorem 10. Let U be a path connected cover of X and z_1, z_2, \ldots, z_g be a sequence of an optimal cycle basis of $H_1(X)$ as stated above.

- *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 $\ell \neq 1$, then we have that the class $\bar{\phi}_{\mathcal{U}*}[z_j] = 0$ for $j = 1, \dots, \ell 1$. Moreover, if $\ell \neq g + 1$, then the classes $\{\bar{\phi}_{\mathcal{U}*}[z_j]\}_{j=\ell,\dots,g}$ generate $H_1(N(\mathcal{U}))$.
- ii. The classes $\{\bar{\phi}_{\mathcal{U}*}[z_j]\}_{j=\ell',\dots,g}$ 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 cycles 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 $\rho : C_1(N(\mathcal{U})) \to C_1(X)$ among one dimensional chain groups as follows. It is sufficient to exhibit the map for an elementary chain of an edge, say $e = \{u_{\alpha}, u_{\alpha'}\} \in C_1(N(\mathcal{U}))$. 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 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 $\rho(e) = \xi(a, b)$. A cycle γ when pushed back by ρ and then pushed forward by $\phi_{\mathcal{U}}$ remains in the same class. The following proposition states this fact whose proof appears in [6].

Proposition 11. Let γ be any 1-cycle in $N(\mathcal{U})$. Then, $[\phi_{\mathcal{U}}(\rho(\gamma))] = [|\gamma|]$.

The following proposition provides a sufficient characterization of the cycles whose classes become trivial after the push forward.

Proposition 12. Let z be a 1-cycle in $C_1(X)$. Then, $[\phi_U(z)] = 0$ if $\lambda(U) > s(z)$.

PROOF. It follows from the definition of the Lebesgue number that there exists a cover element $U_{\alpha} \in \mathcal{U}$ such that $z \subseteq U_{\alpha}$ because $s(z) < \lambda(\mathcal{U})$. We claim that there is a homotopy equivalence that sends $\phi_{\mathcal{U}}(z)$ to a vertex in $N(\mathcal{U})$ and hence $[\phi_{\mathcal{U}}(z)]$ is trivial.

Let x be any point in z. Recall that $\phi_{\mathcal{U}}(x) = \sum_i \varphi_i(x) u_{\alpha_i}$. Since U_{α} has a common intersection with each U_{α_i} so that $\varphi_{\alpha_i}(x) \neq 0$, we can conclude that $\phi_{\mathcal{U}}(x)$ is contained in a simplex with the vertex u_{α} . Continuing this argument with all points of z, we observe that $\phi_{\mathcal{U}}(z)$ is contained in simplices that share the vertex u_{α} . It follows that there is a homotopy that sends $\phi_{\mathcal{U}}(z)$ to u_{α} , a vertex of $N(\mathcal{U})$.

PROOF. [Proof of Theorem 10]

Proof of (i): By Proposition 12, 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 $N(\mathcal{U})$. To see the second part, notice that $\bar{\phi}_{\mathcal{U}*}$ is a surjection by Theorem 5. Therefore, the classes $\bar{\phi}_{\mathcal{U}*}(z)$ where $s(z) \ge \lambda(\mathcal{U})$ contain a basis for

 $H_1(N(\mathcal{U}))$. Hence they generate it.

Proof of (ii): For a contradiction, assume that there is a subsequence $\{\ell_1, \ldots, \ell_t\} \subset \{\ell', \ldots, g\}$ so that $\sum_{j=1}^t [\phi_{\mathfrak{U}}(z_{\ell_j})] = 0$. Let $z = \sum_{j=1}^t \phi_{\mathfrak{U}}(z_{\ell_j})$. Let γ be a 1-cycle in $N(\mathfrak{U})$ so that $[z] = [|\gamma|]$ whose existence is guaranteed by Proposition 3. As $\sum_{j=1}^t [\phi_{\mathfrak{U}}(z_{\ell_j})] = 0$, it must be that there is a 2-chain D in $N(\mathfrak{U})$ so that $\partial D = \gamma$. Consider a triangle $t = \{u_{\alpha_1}, u_{\alpha_2}, u_{\alpha_3}\}$ contributing to D. Let $a'_i = \phi_{\mathfrak{U}}^{-1}(u_{\alpha_i})$. Since t appears in $N(\mathfrak{U})$, 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 $2s_{max}(\mathfrak{U})$. Then, $\rho(\partial t)$ is mapped to a 1-cycle in X of size at most $4s_{max}(\mathfrak{U})$. It follows that $\rho(\partial D)$ can be written as a linear combination of cycles of size at most $4s_{max}(\mathfrak{U})$. Since z_1, \ldots, z_g form an optimal cycle basis of $H_1(X)$, each of the 1-cycles of size at most $4s_{max}(\mathfrak{U})$ is generated by basis elements z_1, \ldots, z_k where $s(z_k) \leq 4s_{max}(\mathfrak{U})$. Therefore, the class of $z' = \phi_{\mathfrak{U}}(\rho(\gamma))$ is generated by a linear combination of the basis elements whose preimages have size at most $4s_{max}(\mathfrak{U})$. The class [z'] is same as the class $[|\gamma|]$ by Proposition 11. But, by assumption $[|\gamma|] = [z]$ is generated by a linear combination of the basis elements whose sizes are larger than $4s_{max}(\mathfrak{U})$ reaching a contradiction. Hence the assumption cannot hold and (ii) is true. \Box

13.3 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*. More precisely, 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, and 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. By construction, every element in 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. Consequently, all nerves of pullbacks of finite covers become finite.

Definition 4 (Mapper). Let *X* and *Z* be topological spaces and let $f : X \to Z$ be a well-behaved and continuous map. Let $\mathcal{U} = \{U_{\alpha}\}_{\alpha \in A}$ be a finite open cover of *Z*. The *mapper* arising from these data is defined to be the nerve of the pullback cover $f^*(\mathcal{U})$ of *X*; that is, $M(\mathcal{U}, f) := N(f^*(\mathcal{U}))$. See an illustration in Figure 13.7.

Notice that we define the mapper using finite covers which allow us to extend definitions of persistence modules and persistence diagrams from previous chapters to the case of mappers. However, in the next Remark and later we allow infinite covers for simplicity. The definition of mapper remains valid with infinite covers.



Figure 13.7: Mapper construction: (left) a map $f : X \to Z$ from a circle to a subset $Z \subset \mathbb{R}$, (middle) the inverse map f^{-1} induces a cover of circle from a cover \mathcal{U} of Z, (right) the nerves of the two covers of X and Z: the nerve on the left (quadrangle shaped) is the mapper induced by f and \mathcal{U} .

Remark 1. The construction of mapper is quite general if we allow the cover \mathcal{U} to be infinite. For example, it can encompass both the Reeb graph and merge trees: consider X a topological space and $f : X \to \mathbb{R}$. Then, consider the following two options for $\mathcal{U} = \{U_{\alpha}\}_{\alpha \in A}$, the other ingredient of the construction:

- $U_{\alpha} = (-\infty, \alpha)$ for $\alpha \in A = \mathbb{R}$. This corresponds to sublevel sets which in turn lead to merge trees. See, for example, the construction in Figure 13.8(b).
- $U_{\alpha} = (\alpha \varepsilon, \alpha + \varepsilon)$ for $\alpha \in A = \mathbb{R}$, for some fixed $\varepsilon > 0$. This corresponds to (ε -thick) level sets, which induce a relaxed notion of Reeb graphs. See the description in "Mapper for PCD" below and Figure 13.8(a).

In these two examples, for simplicity of presentation, the set A is allowed to have infinite cardinality. Also, note one can take any open cover of \mathbb{R} in this definition. This may give rise to other constructions beyond merge trees or Reeb graphs. For instance, using the infinite setting for simplicity again, one may choose any point $r \in \mathbb{R}$ and let $U_{\alpha} = (r - \alpha, r + \alpha)$ for each $\alpha \in A = \mathbb{R}$ or other constructions.

Mapper for PCD: Consider a finite metric space (P, d_P) , that is, a point set P with distances between every pair of points. For a real $r \ge 0$, one can construct a graph $G^r(P)$ with every point in P as a vertex where an edge (p, p') is in $G^r(P)$ if and only if the distance between pand p' is at most r. Let $f : P \to \mathbb{R}$ be a real-valued function on the point set P. For a set of intervals \mathcal{U} covering \mathbb{R} , we can construct the mapper as follows. For every interval $(a, b) \in \mathcal{U}$, let $P_{(a,b)} = f^{-1}((a,b))$ be the set of points with function values in the range (a, b). Each such set consists of a partition $P_{(a,b)} = \{P_{(a,b)}^i\}$ determined by the graph connectivity of $G^r(P)$. Each set $P_{(a,b)}^i$ consists of the vertices of a connected component of the subgraph of $G^r(P)$ spanned by the vertices in $P_{(a,b)}$. The vertex sets $\bigcup_{(a,b)\in\mathcal{U}}\{P_{(a,b)}^i\}$ thus obtained over all intervals constitute a cover $f^{-1}(\mathcal{U})$ of P. The nerve of this cover is the mapper M(P, f). Here the intersection between cover elements is determined by the intersection of discrete sets.

Observe that, in the above construction, if one takes the intervals of $\mathcal{U} = \{U_i\}_{i \in \mathbb{Z}}$ where $U_i = (i - \varepsilon, i + \varepsilon)$ for some $\varepsilon \in (0, 1)$ causing only two consecutive intervals overlap partially, then we get a discretized approximation of the Reeb graphs of the function that f approximates on the discretized sample P. Figure 13.8 illustrates this observation. In the limit that each interval

degenerates to a point, the discretized Reeb converges to the original Reeb graph as shown in [6, 11].



Figure 13.8: Mapper construction for point cloud, a map $f : P \to Z$ from a PCD P to a subset $Z \subset \mathbb{R}$; the graph G^r is not shown: (a) covers are intervals; points are colored with the interval colors, gray points have values in two overlapping intervals, the mapper is a discretized Reeb graph; (b) the covers are sublevel sets, points are colored with the smallest levelset they belong to, discretized Reeb graph does not have the central loop any more.

13.3.1 Multiscale Mapper

A mapper $M(\mathcal{U}, f)$ is a simplicial complex encoding the structure of f through the lens of Z. However, the simplicial complex $M(\mathcal{U}, f)$ provides only *one* snapshot of X at a *fixed* scale determined by the scale of the cover \mathcal{U} . Using the idea of persistent homology, we study the evolution of the mapper $M(f, \mathcal{U}_a)$ for a *tower of covers* $\mathfrak{U} = {\mathcal{U}_a}_{a \in A}$. The tower by definition coarsens the cover with increasing indices and hence provides mappers at multiple scales.

As an intuitive example, consider a real-valued function $f : X \to \mathbb{R}$, and a cover $\mathcal{U}_{\varepsilon}$ of \mathbb{R} consisting of all possible intervals of length ε . Intuitively, as ε tends to 0, the corresponding Mapper $M(f, \mathcal{U}_{\varepsilon})$ approaches the Reeb graph of f. As ε increases, we look at the Reeb graph at coarser and coarser resolution. The multiscale mapper in this case roughly encodes this simplification process.

The idea of multiscale mapper requires a sequence of covers of the target space connected by cover maps. Through pullbacks, it generates a sequence of covers on the domain. In particular, first we have:

Proposition 13. Let $f : X \to Z$, and \mathcal{U} and \mathcal{V} be two covers of Z with a map of covers $\xi : \mathcal{U} \to \mathcal{V}$. Then, there is a corresponding map of covers between the respective pullback covers of X: $f^*(\xi) : f^*(\mathcal{U}) \longrightarrow f^*(\mathcal{V})$. PROOF. Indeed, we only need to note that if $U \subseteq V$, then $f^{-1}(U) \subseteq f^{-1}(V)$, and therefore it is clear that each path connected component of $f^{-1}(U)$ is included in exactly one path connected component of $f^{-1}(V)$. More precisely, let $\mathcal{U} = \{U_{\alpha}\}_{\alpha \in A}$, $\mathcal{V} = \{V_{\beta}\}_{\beta \in B}$, with $U_{\alpha} \subseteq V_{\xi(\alpha)}$ for $\alpha \in A$. Let $\widehat{U}_{\alpha,i}, i \in \{1, \ldots, n_{\alpha}\}$ denote the connected components of $f^{-1}(U_{\alpha})$ and $\widehat{V}_{\beta,j}, j \in \{1, \ldots, m_{\beta}\}$ denote the connected components of $f^{-1}(V_{\beta})$. Then, the map of covers $f^{*}(\xi)$ from $f^{*}(\mathcal{U})$ to $f^{*}(\mathcal{V})$ is given by requiring that each set $\widehat{U}_{\alpha,i}$ is sent to the *unique* set of the form $\widehat{V}_{\xi(\alpha),j}$ so that $\widehat{U}_{\alpha,i} \subseteq \widehat{V}_{\xi(\alpha),j}$.

Furthermore, observe that if $\mathcal{U} \xrightarrow{\xi} \mathcal{V} \xrightarrow{\zeta} \mathcal{W}$ are three different covers of a topological space with the intervening maps of covers between them, then $f^*(\zeta \circ \xi) = f^*(\zeta) \circ f^*(\xi)$.

The above result for three covers easily extends to multiple covers and their pullbacks. The sequence of pullbacks connected by cover maps and the corresponding sequence of nerves connected by simplicial maps define multiscale mappers. Recall the definition of towers to designate a sequence of *objects* connected with maps. Let $\mathfrak{U} = {\mathfrak{U}_a \xrightarrow{u_{a,a'}} \mathfrak{U}_{a'}}_{r \le a \le a'}$ denote a tower, where $r = \operatorname{res}(\mathfrak{U})$ refers to its resolution. The objects here can be covers, simplicial complexes, or vector spaces. The notion of resolution and the variable *a* intuitively specify the granularity of the covers and the simplicial complexes induced by them.

The pullback property given by Proposition 13 makes it possible to take the pullback of a given tower of covers of a space via a given continuous function into another space as stated in proposition below.

Proposition 14. Let \mathfrak{U} be a cover tower of Z and $f : X \to Z$ be a continuous function. Then, $f^*(\mathfrak{U})$ is a cover tower of X.

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

Definition 5 (Multiscale Mapper). Let X and Z be topological spaces and $f : X \to Z$ be a continuous map. Let \mathfrak{U} be a cover tower of Z. Then, the *multiscale mapper* is defined to be the simplicial tower obtained by the nerve of the pullback:

$$\mathsf{MM}(\mathfrak{U}, f) := N(f^*(\mathfrak{U}))$$

where the simplicial maps are induced by the respective cover maps. See Figure 13.9 for an illustration.

Consider for example a sequence $res(\mathfrak{U}) \leq a_1 < a_2 < \ldots < a_n$ of *n* distinct real numbers. Then, the definition of multiscale mapper MM(\mathfrak{U}, f) gives rise to the following simplicial tower:

$$N(f^*(\mathcal{U}_{a_1})) \to N(f^*(\mathcal{U}_{a_2})) \to \dots \to N(f^*(\mathcal{U}_{a_n})).$$
(13.3)

which is a sequence of simplicial complexes connected by simplicial maps.

Applying to them the homology functor $H_p(\cdot)$, p = 0, 1, 2, ..., with coefficients in a field, one obtains a persistence module: tower of vector spaces connected by linear maps.

$$\mathrm{H}_{p}(N(f^{*}(\mathcal{U}_{a_{1}}))) \to \dots \to \mathrm{H}_{p}(N(f^{*}(\mathcal{U}_{a_{n}}))).$$
(13.4)

Given our assumptions that the covers are finite and that the function f is well-behaved, we obtain that the homology groups of all nerves have finite dimensions. Thus, we get a persistence



Figure 13.9: Illustrating construction of multiscale mapper from a cover tower; CT and ST denote cover and simplicial towers respectively, that is, $CT(Z) = \mathfrak{U}$, $CT(X) = f^*(\mathfrak{U})$, and $ST(X) = N(f^*(\mathfrak{U}))$.

module which is p.f.d.(see Topic 6 on towers). Now one can summarize the persistence module induced by $MM(\mathfrak{U}, f)$ with its persistent diagram $Dgm_pMM(\mathfrak{U}, f)$ for each dimension $p \in \mathbb{N}$. The diagram $Dgm_pMM(\mathfrak{U}, f)$ can be viewed as a topological summary of f through the lens of \mathfrak{U} .

13.3.2 Persistence of H₁-classes in mapper and multiscale mapper

To apply the results for nerves in Section 13.2 to mappers and multiscale mappers, we need a 'size' measure on X. For this, we assume that Z is a metric space and we pull back the metric to X via $f : X \to Z$. Assuming that X is path connected, let $\Gamma_X(x, x')$ denote the set of all continuous paths $\gamma : [0, 1] \to X$ between any two given points $x, x' \in X$ so that $\gamma(0) = x$ and $\gamma(1) = x'$.

Definition 6 (Pullback metric). Given a metric space (Z, d_Z) , we define its *pullback metric* as the following pseudometric d_f on X: for $x, x' \in X$,

$$\mathsf{d}_f(x,x') := \inf_{\gamma \in \Gamma_X(x,x')} \operatorname{diam}_Z(f \circ \gamma).$$

Consider the Lebesgue number of the pullback covers of *X*. The following observation in this respect is useful.

Proposition 15. Let \mathcal{U} be a cover for the codomain Z and \mathcal{U}' be its restriction to f(X). Then, the pullback cover $f^*\mathcal{U}$ has the same Lebesgue number as that of \mathcal{U}' ; that is $\lambda(f^*\mathcal{U}) = \lambda(\mathcal{U}')$.

PROOF. First, observe that, for any path connected cover of X, a subset of X that realizes the Lebesgue number can be taken as path connected because, if not, this subset can be connected by a path entirely lying within the cover element containing it. Let $X' \subseteq X$ be any subset where $s(X') \leq \lambda(\mathcal{U}')$. Then, $f(X') \subseteq Z$ has a diameter at most $\lambda(\mathcal{U}')$ by the definitions of size (Definition 3) and pullback metric. Therefore, by the definition of Lebesgue number, f(X') is contained in a cover element $U' \in \mathcal{U}'$. Since X' is path connected, a path connected component of $f^{-1}(U')$ contains X'. It follows that there is a cover element in $f^*\mathcal{U}$ that contains X'. Since X' was chosen

as an arbitrary path connected subset of size at most $\lambda(\mathcal{U}')$, we have $\lambda(f^*\mathcal{U}) \ge \lambda(\mathcal{U}')$. At the same time, it is straightforward from the definition of size that each cover element in $f^{-1}(U')$ has at most the size of U' for any $U' \in \mathcal{U}'$. Combining with the fact that \mathcal{U}' is the restriction of \mathcal{U} to f(X), we have $\lambda(f^*\mathcal{U}) \le \lambda(\mathcal{U}')$, establishing the equality as claimed.

Given a cover \mathcal{U} of Z, consider the mapper $N(f^*\mathcal{U})$. Let z_1, \ldots, z_g be a set of optimal cycle basis for $H_1(X)$ where the metric used to define optimality is the pullback metric d_f . Then, as a consequence of Theorem 10 we have:

Theorem 16. Let $f : X \to Z$ be a map from a path connected space X to a metric space Z equipped with a cover \mathcal{U} (*i* and *ii* below) or a tower of covers { \mathcal{U}_a } (*iii* below). Let \mathcal{U}' be the restriction of \mathcal{U} to f(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 $\phi_{\mathcal{U}*}[z_j] = 0$ 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 $\{\phi_{U*}[z_j]\}_{j=\ell',\dots,g}$ are linearly independent where $s(z_{\ell'}) > 4s_{max}(U)$.
- *iii* Consider a H₁-persistence module of a multiscale mapper induced by a tower of path connected covers:

$$H_1(N(f^*\mathcal{U}_{a_0})) \xrightarrow{s_{1*}} H_1(N(f^*\mathcal{U}_{a_1})) \xrightarrow{s_{2*}} \cdots \xrightarrow{s_{n*}} H_1(N(f^*\mathcal{U}_{a_n}))$$
(13.5)

Let $\hat{s}_{i*} = s_{i*} \circ s_{(i-1)*} \circ \cdots \circ \bar{\phi}_{\mathcal{U}_{a_0}*}$. Then, the assertions in (i) and (ii) hold for $H_1(N(f^*\mathcal{U}_{a_i}))$ with the map $\hat{s}_{i*} : X \to N(f^*\mathcal{U}_{a_i})$.

13.4 Stability

To be useful in practice, the multiscale mapper should be stable against the perturbations in the maps and the covers. we show that such a stability is enjoyed by the multiscale mapper under some natural condition on the tower of covers. Recall that previous stability results for towers were drawn on the notion of interleaving. We identify compatible notions of interleaving for cover towers as a way to measure the "closeness" between two cover towers.

13.4.1 Interleaving of cover towers and multiscale mappers

In this section we consider cover and simplicial towers indexed over \mathbb{R} . In practice, we often have a cover tower $\mathfrak{U} = {\mathcal{U}_a \xrightarrow{u_{a,a'}} \mathcal{U}_{a'}}_{a \le a'}$ indexed by a discrete set in $A \subset \mathbb{R}$. Any such tower can be extended to a cover tower indexed over \mathbb{R} by taking $\mathcal{U}_{\varepsilon} = \mathcal{U}_a$ for each index $\varepsilon \in (a, a')$ where a, a'are any two consecutive indices in the ordered set A.

Definition 7 (Interleaving of cover towers). Let $\mathfrak{U} = {\mathcal{U}_a}$ and $\mathfrak{V} = {\mathcal{V}_a}$ be two cover towers of a topological space X so that $\operatorname{res}(\mathfrak{U}) = \operatorname{res}(\mathfrak{V}) = r$. Given $\eta \ge 0$, we say that \mathfrak{U} and \mathfrak{V} are

 η -interleaved if one can find cover maps $\zeta_a : \mathcal{U}_a \to \mathcal{V}_{a+\eta}$ and $\xi_{a'} : \mathcal{V}_{a'} \to \mathcal{U}_{a'+\eta}$ for all $a, a' \ge r$; see the diagram below.



Analogously, if we replace the operator '+' by the multiplication '.' in the above definition, then we say that \mathfrak{U} and \mathfrak{V} are *multiplicatively* η *-interleaved*.

Proposition 17. (i) If \mathfrak{U} and \mathfrak{B} are (multiplicative) η_1 -interleaved and \mathfrak{B} and \mathfrak{W} are (multiplicative) η_2 -interleaved, then, \mathfrak{U} and \mathfrak{W} are (multiplicative ($\eta_1\eta_2$)-) ($\eta_1 + \eta_2$)-interleaved. (ii) Let $f: X \to Z$ be a continuous function and \mathfrak{U} and \mathfrak{B} be two (multiplicative) η -interleaved tower of covers of Z. Then, $f^*(\mathfrak{U})$ and $f^*(\mathfrak{B})$ are also (multiplicative) η -interleaved.

Note that in the definition of interleaving cover towers, we do not have explicit requirement that maps need to make sub-diagrams commute unlike the the interleaving between simplicial towers. However, it follows from Proposition 2 that interleaving cover towers lead to interleaving between simplicial towers for $N(\mathfrak{U})$ and $N(\mathfrak{V})$ as shown in the proposition below.

Proposition 18. Let \mathfrak{U} and \mathfrak{V} be two (multiplicatively) η -interleaved cover towers of X with $res(\mathfrak{U}) = res(\mathfrak{V})$. Then, $N(\mathfrak{U})$ and $N(\mathfrak{V})$ are also (multiplicatively) η -interleaved.

PROOF. We prove the proposition for additive interleaving. Replacing the '+' operator with '.' gives the proof for multiplicative interleaving. Let *r* denote the common resolution of \mathfrak{U} and \mathfrak{B} . Write $\mathfrak{U} = {\mathcal{U}_a \xrightarrow{u_{a,d'}} \mathcal{U}_{a'}}_{r \le a \le a'}$ and $\mathfrak{B} = {\mathcal{V}_a \xrightarrow{v_{a,d'}} \mathcal{V}_{a'}}_{r \le a \le a'}$, and for each $a \ge r$ let $\zeta_a : \mathcal{U}_a \to \mathcal{V}_{a+\eta}$ and $\xi_a : \mathcal{V}_a \to \mathcal{U}_{a+\eta}$ be given as in Definition 7. To define interleaving between the towers of nerves arising out of covers, we consider similar diagrams to tower interleaving at the level of covers involving covers of the form \mathcal{U}_a and \mathcal{V}_a , and apply the nerve construction. This operation yields diagrams identical to those for towers where for every a, a' where $a' \ge a \ge r$:

- $K_a := N(\mathcal{U}_a), L_a := N(\mathcal{V}_a),$
- $x_{a,a'} := N(u_{a,a'})$, for $r \le a \le a'$; $y_{a,a'} := N(v_{a,a'})$, for $r \le a \le a'$; $\varphi_a := N(\zeta_a)$, and $\psi_a := N(\zeta_a)$.

To satisfy Definition of interleaving, it remains to verify conditions (i) to (iv). We only verify (i), since the proof of the others follows the same arguments. For this, notice that both the composite map $\xi_{a+\eta} \circ \zeta_a$ and $u_{a,a+2\eta}$ are maps of covers from \mathcal{U}_a to $\mathcal{U}_{a+2\eta}$. By Proposition 2 we then have that $N(\xi_{a+\eta} \circ \zeta_a)$ and $N(u_{a,a+2\eta}) = f_{a,a+2\eta}$ are contiguous. But, by the properties of the nerve construction $N(\xi_{a+\eta} \circ \zeta_a) = N(\xi_{a+\eta}) \circ N(\zeta_a) = \psi_{a+\eta} \circ \varphi_a$, which completes the claim.

Combining Proposition 17 and Proposition 18, we get that the two multiscale mappers under cover perturbations stay stable, which is the first part of Corollary 19. Recall from Chapter on General Persistence that, for a finite simplicial tower S and $p \in \mathbb{N}$, we denote by $\text{Dgm}_p(S)$ the *p*-th persistence diagram of the tower S with coefficients in a fixed field. Using Proposition 18, we have a stability result for $\text{Dgm}_p\text{MM}(\mathfrak{U}, f)$ when f is kept fixed but the cover tower \mathfrak{U} is perturbed, which is the second part of the corollary below.

Corollary 19. For $\eta \ge 0$, let \mathfrak{U} and \mathfrak{B} be two finite cover towers of Z with $\operatorname{res}(\mathfrak{U}) = \operatorname{res}(\mathfrak{B}) > 0$. Let $f : X \to Z$ be well-behaved and \mathfrak{U} and \mathfrak{B} be η -interleaved. Then, $\operatorname{MM}(\mathfrak{U}, f)$ and $\operatorname{MM}(\mathfrak{B}, f)$ are η -interleaved. In particular, the bottleneck distance between the persistence diagrams $\operatorname{Dgm}_p \operatorname{MM}(\mathfrak{U}, f)$ and $\operatorname{Dgm}_p \operatorname{MM}(\mathfrak{B}, f)$ is at most η for all $p \in \mathbb{N}$.

13.4.2 (*c*, *s*)-good covers

Although $Dgm_pMM(\mathfrak{U}, f)$ is stable under perturbations of the covers \mathfrak{U} as we showed, it is not necessarily stable under perturbations of the map f. To address this issue, we introduce a special family of covers called (c,s)-good covers. To define these covers, we use the index value of the covers to denote their scales. The notation ε for indexing is chosen to emphasize this meaning.

Definition 8 ((*c*, *s*)-good cover tower). Given a cover tower $\mathfrak{U} = {\mathcal{U}_{\varepsilon}}_{\varepsilon \ge s}$, we say that it is (*c*,*s*)good if for any $\varepsilon \ge s$, we have that (i) $s_{max}(\mathcal{U}_{\varepsilon}) \le \varepsilon$ and (ii) $\lambda(\mathcal{U}_{c\varepsilon}) \ge \varepsilon$.

As an example, consider the cover tower $\mathfrak{U} = {\mathfrak{U}_{\varepsilon}}_{\varepsilon \ge s}$ with $\mathfrak{U}_{\varepsilon} := {B_{\varepsilon/2}(z) \mid z \in Z}$. It is a (2, *s*)-good cover tower of the metric space (*Z*, *d_Z*).

We now characterize the persistent homology of multiscale mappers induced by (c,s)-good cover towers. Theorem 20 states that the multiscale-mappers induced by any two (c, s)-good cover towers 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 cover towers contain roughly the same information.

Theorem 20. Given a map $f : X \to Z$, let $\mathfrak{U} = \{\mathfrak{U}_{\varepsilon} \xrightarrow{u_{\varepsilon,\varepsilon'}} \mathfrak{U}_{\varepsilon'}\}_{\varepsilon \leq \varepsilon'}$ and $\mathfrak{V} = \{\mathcal{V}_{\varepsilon} \xrightarrow{v_{\varepsilon,\varepsilon'}} \mathcal{V}_{\varepsilon'}\}_{\varepsilon \leq \varepsilon'}$ be two (c, s)-good cover towers of Z. Then the corresponding multiscale mappers MM (\mathfrak{U}, f) and MM (\mathfrak{V}, f) are multiplicatively c-interleaved.

PROOF. First, we make the following observation.

Claim 2. Any two (c, s)-good cover towers \mathfrak{U} and \mathfrak{V} are multiplicatively *c*-interleaved.

PROOF. It follows easily from the definitions of (c, s)-good cover tower. Specifically, first we construct $\zeta_{\varepsilon} : \mathcal{U}_{\varepsilon} \to \mathcal{V}_{c\varepsilon}$. For any $U \in \mathcal{U}_{\varepsilon}$, we have that $\operatorname{diam}(U) \leq \varepsilon$. Furthermore, since \mathfrak{V} is (c, s)-good, there exists $V \in \mathcal{V}_{c\varepsilon}$ such that $U \subseteq V$. Set $\zeta_{\varepsilon}(U) = V$; if there are multiple choice of V, we can choose an arbitrary one. We can construct $\xi_{\varepsilon'} : \mathcal{V}_{\varepsilon'} \to \mathcal{U}_{c\varepsilon'}$ in a symmetric manner, and the claim then follows.

This claim, combined with Propositions 17 and 18, prove the theorem.

We also need the following definition in order to state the stability results precisely.

Definition 9. Given a tower of covers $\mathfrak{U} = {\mathcal{U}_{\varepsilon}}$ and $\varepsilon_0 \ge \operatorname{res}(\mathfrak{U})$, we define the ε_0 -truncation of \mathfrak{U} as the tower $\operatorname{Tr}_{\varepsilon_0}(\mathfrak{U}) := {\mathcal{U}_{\varepsilon}}_{\varepsilon_0 < \varepsilon}$. Observe that, by definition $\operatorname{res}(\operatorname{Tr}_{\varepsilon_0}(\mathfrak{U})) = \varepsilon_0$.

Proposition 21. Let X be a compact topological space, (Z, d_Z) be a compact path connected metric space, and $f, g : X \to Z$ be two continuous functions such that for some $\delta \ge 0$ one has that $\delta = \max_{x \in X} \mathsf{d}_Z(f(x), g(x))$. Let \mathfrak{W} be any (c, s)-good cover tower of Z. Let $\varepsilon_0 = \max(1, s)$. Then, the ε_0 -truncations of $f^*(\mathfrak{W})$ and $g^*(\mathfrak{W})$ are multiplicatively $(2c \max(\delta, s) + c)$ -interleaved.

PROOF. For notational convenience write $\eta := 2c \max(\delta, s) + c$, $\{\mathcal{U}_t\} = \mathfrak{U} := f^*(\mathfrak{W})$, and $\{\mathcal{V}_t\} = \mathfrak{V} := g^*(\mathfrak{W})$. With regards to satisfying the definition of interleaving for \mathfrak{U} and \mathfrak{V} , for each $\varepsilon \ge \varepsilon_0$ we need only exhibit maps of covers $\zeta_{\varepsilon} : \mathcal{U}_{\varepsilon} \to \mathcal{V}_{\eta\varepsilon}$ and $\xi_{\varepsilon} : \mathcal{V}_{\varepsilon} \to \mathcal{U}_{\eta\varepsilon}$. We first establish the following, where recall that the offset O^r is defined as $O^r := \{z \in Z \mid \mathsf{d}_Z(z, O) \le r\}$.

Claim 3. For all $O \subset Z$, and all $\delta' \ge \delta$, $f^{-1}(O) \subseteq g^{-1}(O^{\delta'})$.

PROOF. Let $x \in f^{-1}(O)$, then $d_Z(f(x), O) = 0$. Thus,

$$\mathsf{d}_Z(g(x), O) \le \mathsf{d}_Z(f(x), O) + \mathsf{d}_Z(g(x), f(x)) \le \delta,$$

which implies the claim.

Now, pick any $\varepsilon \ge \varepsilon_0$, any $U \in \mathcal{U}_{\varepsilon}$, and fix $\delta' := \max(\delta, s)$. Then, there exists $W \in \mathcal{W}_{\varepsilon}$ such that $U \in \operatorname{cc}(f^{-1}(W))$, where $\operatorname{cc}(Y)$ stands for the set of path connected components of Y. Claim 3 implies that $f^{-1}(W) \subseteq g^{-1}(W^{\delta'})$. Since \mathfrak{W} is a (c, s)-good cover of the connected space Z and $s \le \max(\delta, s) \le 2\delta' + \varepsilon$, there exists at least one set $W' \in \mathcal{W}_{c(2\delta'+\varepsilon)}$ such that $W^{\delta'} \subseteq W'$. This means that U is contained in some element of $\operatorname{cc}(g^{-1}(W'))$ where $W' \in \mathcal{W}_{c(2\delta'+\varepsilon)}$. But, also, since $c(2\delta' + \varepsilon) \le c(2\delta' + 1)\varepsilon$ for $\varepsilon \ge \varepsilon_0 \ge 1$, there exists $W'' \in \mathcal{W}_{c(2\delta'+1)\varepsilon}$ such that $W' \subseteq W''$. This implies that U is contained in some element of $\operatorname{cc}(g^{-1}(W'))$ where $W'' \in \mathcal{W}_{c(2\delta'+1)\varepsilon}$. This process, when applied to all $U \in \mathcal{U}_{\varepsilon}$, all $\varepsilon \ge \varepsilon_0$, defines a map of covers $\zeta_t : \mathcal{U}_t \to \mathcal{V}_{(2c\delta'+c)\varepsilon}$. A similar observation produces for each $\varepsilon \ge \varepsilon_0$ a map of covers ξ_{ε} from $\mathcal{V}_{\varepsilon}$ to $\mathcal{V}_{(2c\delta'+c)\varepsilon}$.

So we have in fact proved that ε_0 -truncations of \mathfrak{U} and \mathfrak{B} are multiplicatively η -interleaved. \Box

Applying Proposition 21, Proposition 18, we get the following result, where Dgm_{log} stands for the persistence diagram at the log-scale (of coordinates).

Corollary 22. Let \mathfrak{W} be a (c, s)-good cover tower of the compact connected metric space Z and let $f, g: X \to Z$ be any two well-behaved continuous functions such that $\max_{x \in X} \mathsf{d}_Z(f(x), g(x)) = \delta$. Then, the bottleneck distance between the persistence diagrams

 $\mathsf{d}_b(\mathrm{Dgm}_{\log}\mathrm{MM}(\mathfrak{W},f),\mathrm{Dgm}_{\log}\mathrm{MM}(\mathfrak{W},g)) \leq \log(2c\max(s,\delta)+c) + \max(0,\log\frac{1}{s}).$

PROOF. We use the notation of Proposition 21. Let $\mathfrak{U} = f^*(\mathfrak{W})$ and $\mathfrak{V} = g^*(\mathfrak{W})$. If max(1, s) = s, then \mathfrak{U} and \mathfrak{V} are multiplicatively $(2c \max(s, \delta) + c)$ -interleaved by Proposition 21 which gives a bound on the bottleneck distance of $\log(2c \max(s, \delta) + c)$ between the corresponding persistence diagrams at the log-scale. In the case when s < 1, the bottleneck distance remains the same only for the 1-truncations of \mathfrak{U} and \mathfrak{V} . Shifting the starting point of the two families to the left by at most *s* can introduce barcodes of lengths at most $\log \frac{1}{s}$ or can stretch the existing barcodes to the left by at most $\log \frac{1}{s}$ for the respective persistence modules at the log-scale. To see this, consider the persistence module below where $\varepsilon_1 = s$:

$$H_k(N(f^*(\mathcal{U}_{\varepsilon_1}))) \to H_k(N(f^*(\mathcal{U}_{\varepsilon_2}))) \to \cdots \to H_k(N(f^*(\mathcal{U}_1))) \to \cdots \to H_k(N(f^*(\mathcal{U}_{\varepsilon_n})))$$

A homology class born at any index in the range [s, 1) either dies at or before the index 1 or is mapped to a homology class of $H_k(N(f^*(U_1)))$. In the first case we have a bar code of length at most $|\log s| = \log \frac{1}{s}$ at the log-scale. In the second case, a bar code of the persistence module

 $H_k(N(f^*(\mathcal{U}_{\varepsilon_1}))) \to \cdots \to H_k(N(f^*(\mathcal{U}_{\varepsilon_n})))$

starting at index 1 gets stretched to the left by at most $|\log s| = \log \frac{1}{s}$. The same conclusion can be drawn for the persistence module induced by \mathfrak{V} . Therefore the bottleneck distance between the respective persistence diagrams at log-scale changes by at most $\log \frac{1}{s}$.

13.4.3 Relation to intrinsic Čech filtration

In Section 13.3.2, we have seen that given a tower of covers \mathfrak{U} and a map $f: X \to Z$ there exists a natural pull-back pseudo-metric d_f defined on the input domain X (Definition 6). With such a pseudo-metric on X, we can now construct the standard (intrinsic) Čech filtration $\mathcal{C}(X) = \{\mathbb{C}_{\varepsilon}(X)\}_{\varepsilon}$ (or Rips filtration) in X directly, instead of computing the Nerve complex of the pull-back covers as required by mapper. The resulting filtration $\mathcal{C}(X)$ is connected by inclusion maps instead of simplicial maps. This is easier for computational purposes even though one has a method to compute the persistence diagram of a tower involving arbitrary simplicial maps. Furthermore, it turns out that the resulting sequence of Cech complexes C interleaves with the sequence of complexes MM(\mathfrak{U}, f), implying that their corresponding persistence diagrams approximate each other. Specifically, in Theorem 23, we show that when the codomain of the function $f: X \to Z$ is a metric space (Z, d_Z) , the multiscale mapper induced by any (c, s)-good cover tower interleaves (at the homology level) with an intrinsic Čech filtration of X defined below. We have already considered Čech filtrations before. However, we considered only a finite subset of a metric space to define the Cech complex. Here we redefine it again to account for the fact that each point of the (pseudo-)metric space is considered and call it intrinsic Čech complex (see an earlier example of intrinsic Cech complex when we analyzed graphs).

Definition 10. Given a (pseudo)metric space (Y, d_Y) , its *intrinsic Čech complex* $\mathbb{C}^r(Y)$ at scale r is defined as the nerve complex of the set of intrinsic r-balls $\{B(y; r)\}_{y \in Y}$ defined using (pseudo)metric d_Y .

The above definition gives way to defining a Čech filtration.

Definition 11 (Intrinsic Čech filtration). The *intrinsic* Čech filtration of the (pseudo)metric space (Y, d_Y) is

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

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

Recall the definition of the pseudometric d_f on X (Definition 6) induced from a metric on Z. Applying Definition 10 on the pseudometric space (X, d_f) , we obtain its intrinsic Čech complex $\mathbb{C}^r(X)$ at scale r and then its Čech filtration $\mathcal{C}_s(X)$.

Theorem 23. Let $\mathcal{C}_s(X)$ be the intrinsic Čech filtration of (X, d_f) starting with resolution s. Let $\mathfrak{U} = \{ \mathfrak{U}_{\varepsilon} \xrightarrow{u_{\varepsilon,\varepsilon'}} \mathfrak{U}_{\varepsilon'} \}_{s \le \varepsilon \le \varepsilon'}$ be a (c, s)-good cover tower of the compact connected metric space Z. Then the multiscale mapper MM(\mathfrak{U}, f) and $\mathcal{C}_s(X)$ are multiplicatively 2*c*-interleaved.

By a property (see the book) of multiplicative interleaving, the following result is deduced immediately from Theorem 23.

Corollary 24. Given a continuous map $f : X \to Z$ and a (c, s)-good cover tower \mathfrak{U} of Z, let $\mathrm{Dgm}_{\log}\mathrm{MM}(\mathfrak{U}, f)$ and $\mathrm{Dgm}_{\log}\mathbb{C}_s$ denote the log-scaled persistence diagram of the persistence modules induced by $\mathrm{MM}(\mathfrak{U}, f)$ and by the intrinsic Čech filtration \mathbb{C}_s of (X, d_f) respectively. We have that

 $\mathsf{d}_b(\mathrm{Dgm}_{\mathrm{log}}\mathrm{MM}(\mathfrak{U}, f), \mathrm{Dgm}_{\mathrm{log}}\mathfrak{C}_s) \leq 2c.$

13.5 Exact Computation for PL-functions on simplicial domains

The stability result in Theorem 23 further motivates us to design efficient algorithms for constructing multiscale mapper or its approximation in practice. A priori, the construction of the mapper and multiscale mapper may seem clumsy. Even for PL-functions defined on a simplicial complex, the standard algorithm needs to determine for each simplex the subset (partial simplex) on which the function value falls within a certain range. We observe that for such an input, it is sufficient to consider the restriction of the function to the 1-skeleton of the complex for computing the mapper and the multiscale mapper. Since the 1-skeleton (a graph) is typically much smaller in size than the full complex, this helps improving the time efficiency of computing the mapper and multiscale mapper.

Consider one of the most common types of input in practice, a real-valued PL-function f: $|K| \rightarrow \mathbb{R}$ defined on the underlying space |K| of a simplicial complex K given as a vertex function. In what follows, we consider this PL setting, and show that interestingly, if the input function satisfies a mild "minimum diameter" condition, then we can compute both mapper and multiscale mapper from simply the 1-skeleton (graph structure) of K. This makes the computation of the multiscale mapper from a PL-function significantly faster and simpler as its time complexity depends on the size of the 1-skeleton of K, which is typically orders of magnitude smaller than the total number of simplices (such as triangles, tetrahedra, etc) in K.

Recall that K^1 denote the 1-skeleton of a simplicial complex K: that is, K^1 contains the set of vertices and edges of K. Define $\tilde{f} : |K^1| \to \mathbb{R}$ to be the restriction of f to $|K^1|$; that is, \tilde{f} is the PL function on $|K^1|$ induced by function values at vertices.

Condition 1 (Minimum diameter condition). For a cover tower \mathfrak{W} of a compact connected metric space (Z, d_Z) , let

 $\kappa(\mathfrak{W}) := \inf\{\operatorname{diam}(W); W \in \mathcal{W} \in \mathfrak{W}\}\$

denote the minimum diameter of any element of any cover of the tower \mathfrak{W} . Given a simplicial complex *K* with a function $f : |K| \to Z$ and a tower of covers \mathfrak{W} of the metric space *Z*, we say that (K, f, \mathfrak{W}) satisfies the *minimum diameter* condition if diam $(f(\sigma)) \le \kappa(\mathfrak{W})$ for every simplex $\sigma \in K$.

In our case, f is a PL-function, and thus satisfying the minimum diameter condition means that for every edge $e = (u, v) \in K^1$, $|f(u) - f(v)| \le \kappa(\mathfrak{W})$. In what follows we assume that K is connected. We do not lose any generality by this assumption because the arguments below can be applied to each connected component of K.

Definition 12 (Isomorphic simplicial towers). Two simplicial towers $S = \{S_{\varepsilon} \xrightarrow{S_{\varepsilon,\varepsilon'}} S_{\varepsilon'}\}$ and $\mathcal{T} = \{T_{\varepsilon} \xrightarrow{t_{\varepsilon,\varepsilon'}} T_{\varepsilon'}\}$ are *isomorphic*, denoted $S \cong \mathcal{T}$, if $\operatorname{res}(S) = \operatorname{res}(\mathcal{T})$, and there exist simplicial isomorphisms η_{ε} and $\eta_{\varepsilon'}$ such that the diagram below commutes for all $\operatorname{res}(S) \le \varepsilon \le \varepsilon'$.



Our main result in this section is the following theorem which enables us to compute the mapper, multiscale mapper, as well as the persistence diagram for the multiscale mapper of a PL function f from its restriction \tilde{f} to the 1-skeleton of the respective simplicial complex.

Theorem 25. Given a PL-function $f : |K| \to \mathbb{R}$ and a tower of covers \mathfrak{W} of the image of f with (K, f, \mathfrak{W}) satisfying the minimum diameter condition, we have that $MM(\mathfrak{W}, f) \cong MM(\mathfrak{W}, \tilde{f})$, where \tilde{f} is the restriction of f to $|K^1|$.

We show in Proposition 26 that the two mapper outputs $M(\mathcal{W}, f)$ and $M(\mathcal{W}, \tilde{f})$ are identical up to a relabeling of their vertices (hence simplicially isomorphic) for every $\mathcal{W} \in \mathfrak{W}$. Also, since the simplicial maps in the filtrations $MM(\mathfrak{W}, f)$ and $MM(\mathfrak{W}, \tilde{f})$ are induced by the pullback of the same tower of covers \mathfrak{W} , they are identical again up to the same relabeling of the vertices. This then establishes the theorem.

In what follows, for clarity of exposition, we use *X* and *X*¹ to denote the underlying space |K| and $|K^1|$ of *K* and K^1 , respectively. Also, we do not distinguish between a simplex $\sigma \in K$ and its image $|\sigma| \subseteq X$ and thus freely say $\sigma \subseteq X$ when it actually means that $|\sigma| \subseteq X$ for a simplex $\sigma \in K$.

Proposition 26. If (K, f, \mathfrak{W}) satisfies the minimum diameter condition, then for every $W \in \mathfrak{W}$, M(W, f) is identical to $M(W, \tilde{f})$ up to relabeling of the vertices.

PROOF. Let $\mathcal{U} = f^*\mathcal{W}$ and $\tilde{\mathcal{U}} = \tilde{f}^*\mathcal{W}$. By definition of \tilde{f} , each $\tilde{U} \in \tilde{\mathcal{U}}$ is a connected component of some $U \cap X^1$ for some $U \in \mathcal{U}$. In Proposition 27, we show that $U \cap X^1$ is connected for every $U \in \mathcal{U}$. Therefore, for every element $U \in \mathcal{U}$, there is a unique element $\tilde{U} = U \cap X^1$ in $\tilde{\mathcal{U}}$ and vice versa. It is not hard to show that $\bigcap_{i=1}^k U_i \neq \emptyset$ if and only if $\bigcap_{i=1}^k \tilde{U}_i \neq \emptyset$. This finishes the proof. \Box

Proposition 27. If (X, f, \mathfrak{W}) satisfies the minimum diameter condition, then for every $W \in \mathfrak{W}$ and every $U \in f^*(W)$, the set $U \cap X^1$ is connected.

PROOF. Fix $U \in f^*(W)$. If $U \cap X^1$ is not connected, let C_1, \ldots, C_k denote its $k \ge 2$ connected components. First, we show that each C_i contains at least one vertex of X^1 . Let e = (u, v) be any edge of X^1 that intersects U. If both ends u and v lie outside U, then $|f(u) - f(v)| > |\max_U f - \min_U f| \ge \kappa(\mathfrak{W})$. But, this violates the minimum diameter condition. Thus, at least one vertex of e is contained in U. It immediately follows that C_i contains at least one vertex of X^1 .

Let Δ be the set of all simplices $\sigma \subseteq X$ so that $\sigma \cap U \neq \emptyset$. Fix $\sigma \in \Delta$ and let *x* be any point in $\sigma \cap U$. We defer the proof of the following claim as an exercise.

Claim 4. There exists a point y in an edge of σ so that f(x) = f(y).

Since σ contains an edge *e* that is intersected by *U*, it contains a vertex of *e* that is contained in *U*. This means every simplex $\sigma \in \Delta$ has a vertex contained in *U*. For each i = 1, ..., k let $\Delta_i := \{ \sigma \subseteq X | V(\sigma) \cap C_i \neq \emptyset \}$ Since every simplex $\sigma \in \Delta$ has a vertex contained in *U*, we have $\Delta = \bigcup_i \Delta_i$. We argue that the sets $\Delta_1, \ldots, \Delta_k$ are disjoint from each other. Otherwise, there exist $i \neq j$ and a simplex σ with a vertex u in Δ_i and another vertex v in Δ_j . Then, the edge (u, v) must be in U because f is PL. But, this contradicts that C_i and C_j are disjoint. This establishes that each Δ_i is disjoint from each other and hence Δ is not connected contradicting that U is connected. Therefore, our initial assumption that $U \cap X^1$ is disconnected is wrong.

13.6 Approximating multiscale mapper for general maps

While results in the previous section concern real-valued PL-functions, we now provide a significant generalization for the case where f maps the underlying space of K into an arbitrary compact metric space Z. We present a "combinatorial" version of the (multiscale) mapper where each connected component of a pullback $f^{-1}(W)$ for any cover W in the cover of Z consists of only vertices of K. Hence, the construction of the Nerve complex for this modified (multiscale) mapper is purely combinatorial, simpler, and more efficient to implement. But we lose the "exactness", that is, in contrast with the guarantees provided by Theorem 25, the combinatorial mapper only *approximates* the actual multiscale mapper at the homology level. Also, it requires a (c, s)good tower of covers of Z. One more caveat is that the towers of simplicial complexes arising in this case do not interleave in the (strong) sense but in a weaker sense (see the PCD chapter). This limitation worsens the approximation result by a factor of 3.

In what follows, as before, cc(O) for a set O denotes the set of all path connected components of O.

Given a map $f : |K| \to Z$ defined on the underlying space |K| of a simplicial complex K, to construct the mapper and multiscale mapper, one needs to compute the pullback cover $f^*(W)$ for a cover W of the compact metric space Z. Specifically, for any $W \in W$ one needs to compute the preimage $f^{-1}(W) \subset |K|$ and shatter it into connected components. Even in the setting adopted in 13.5, where we have a PL function $\tilde{f} : |K^1| \to \mathbb{R}$ defined on the 1-skeleton K^1 of K, the connected components in $cc(\tilde{f}^{-1}(W))$ may contain vertices, edges, and also *partial edges*: say for an edge $e \in K^1$, its intersection $e_W = e \cap f^{-1}(W) \subseteq e$, that is, $f(e_W) = f(e) \cap W$, is a partial edge. See Figure 13.10 for an example. In general for more complex maps, $\sigma \cap f^{-1}(W)$ for any k-simplex σ may be partial triangles, tetrahedra, etc., which can be nuisance for computations. The combinatorial version of mapper and multiscale mapper sidesteps this problem by ensuring that each connected component in the pullback $f^{-1}(W)$ consists of *only vertices of* K. It is thus simpler and faster to compute.

13.6.1 Combinatorial mapper and multiscale mapper

Let G be a graph with vertex set V(G) and edge set E(G). Suppose we are given a map f: V(G) $\rightarrow Z$ and a finite open cover $\mathcal{W} = \{W_{\alpha}\}_{\alpha \in A}$ of the metric space (Z, d_Z) . For any $W_{\alpha} \in \mathcal{W}$, the preimage $f^{-1}(W_{\alpha})$ consists of a set of vertices which is shattered into subsets by the connectivity of the graph G. These subsets are taken as connected components. We now formalize this:

Definition 13 (*G*-induced connected component). Given a set of vertices $O \subseteq V(G)$, the set of connected components of O induced by G, denoted by $cc_G(O)$, is the partition of O into a maximal subset of vertices connected in $G_O \subseteq G$, the subgraph spanned by vertices in O. We refer to each



Figure 13.10: Partial thickened edges belong to the two connected components in $f^{-1}(W)$. Note that each set in $cc_G(f^{-1}(W))$ contains only the set of vertices of a component in $cc(f^{-1}(W))$.

Algorithm 1 MMAPPER (f, K, \mathfrak{W})

Input:

 $f: |K| \to Z$ given by $f_V: V(K) \to Z$, a cover tower $\mathfrak{W} = \{\mathcal{W}_1, \dots, \mathcal{W}_t\}$

Output:

Persistence diagram $Dgm_*(MM_{K^1}(\mathfrak{W}, f_V))$ induced by the combinatorial MM of f w.r.t. \mathfrak{W}

- 1: **for** i = 1, ..., t **do**
- 2: compute $V_W \subseteq V(K)$ where $f(V_W) = f(V(K)) \cap W$ and $\{V_W^j\}_j = cc_{K^1}(V_W), \forall W \in \mathcal{W}_i;$
- 3: compute Nerve complex $N_i = N(\{V_W^j\}_{j,W})$.
- 4: end for
- 5: compute the filtration \mathcal{F} : { $N_i \rightarrow N_{i+1}$, $i \in [1, t-1]$ }
- 6: compute $\text{Dgm}_*(\mathcal{F})$.

such maximal subset of vertices as a *G*-induced connected component of *O*. We define $f^{*_G}(W)$, the *G*-induced pull-back via the function *f*, as the collection of all *G*-induced connected components $cc_G(f^{-1}(W_\alpha))$ for all $\alpha \in A$.

Definition 14. (*G*-induced multiscale mapper) Similar to the mapper construction, we define the *G*-induced mapper $M_G(W, f)$ as the nerve complex $N(f_G^*(W))$.

Given a tower of covers $\mathfrak{W} = \{\mathcal{W}_{\varepsilon}\}$ of Z, we define the *G*-induced multiscale mapper $MM_G(\mathfrak{W}, f)$ as the tower of *G*-induced nerve complexes $\{N(f_G^*(\mathcal{W}_{\varepsilon})) \mid \mathcal{W}_{\varepsilon} \in \mathfrak{W}\}.$

Given a map $f : |K| \to Z$ defined on the underlying space |K| of a simplicial complex K, let $f_V : V(K) \to \mathbb{R}$ denote the restriction of f to the vertices of K. Consider the 1-skeleton graph K^1 that provides the connectivity information for vertices in V(K). Given any cover tower \mathfrak{W} of the metric space Z, the K^1 -induced multiscale mapper $MM_{K^1}(\mathfrak{W}, f_V)$ is called the *combinatorial multiscale mapper of f w.r.t.* \mathfrak{W} .

13.6.2 Advantage of combinatorial multiscale mapper

A simple description of the computation of the combinatorial mapper is in Algorithm 1. For the simple PL example in Figure 13.10, $f^{-1}(W)$ contains two connected components, one consists of the set of white dots, while the other consists of the set of black dots. More generally, the construction of the pullback cover needs to inspect only the 1-skeleton K^1 of K, which is typically of significantly smaller size. Furthermore, the construction of the Nerve complex N_i as in Algorithm 1 is also much simpler: We simply remember, for each vertex $v \in V(K)$, the set I_v of ids of connected components $\{V_W^j\}_{j,W\in W_i}$ which contain it. Any subset of I_v gives rise to a simplex in the Nerve complex N_i .

Let MM(\mathfrak{W} , f) denote the standard multiscale mapper as introduced in 13.3.1. Our main result in this section is that if \mathfrak{W} is a (c, s)-good cover tower of Z, then the resulting two simplicial towers, MM(\mathfrak{W} , f) and MM_{K1}(\mathfrak{W} , f_V) weakly interleave (Definition in the book), and admits a bounded distance between their respective persistence diagrams as a consequence of the weakinterleaving result of [3]. This weaker setting of interleaving worsens the approximation by a factor of 3.

Theorem 28. Assume that (Z, d_Z) is a compact and connected metric space. Given a map $f : |K| \to Z$, let $f_V : V(K) \to Z$ be the restriction of f to the vertex set V(K) of K.

Given a (c, s)-good cover tower \mathfrak{W} of Z such that (K, f, \mathfrak{W}) satisfies the minimum diameter condition (cf. Condition 1), the bottleneck distance between the persistence diagrams $Dgm_{log}MM(\mathfrak{W}, f)$ and $Dgm_{log}MM_{K^1}(\mathfrak{W}, f_V)$ is at most $3\log(3c) + 3\max(0, \log \frac{1}{s})$ for all $k \in \mathbb{N}$.

13.7 Notes and Exercises

A corollary of the nerve theorem is that the space and the nerve have isomorphic homology groups if all intersections of cover elements are homotopically trivial. This chapter studies a case when covers do not necessarily satisfy this property. The result that for path connected covers, no new 1-dimensional homology class is created in the nerve is proved in [6]. The materials in sections 13.1 and 13.2 are taken from there. This result can be generalized for other dimensions; see Exercise 5.

The concept of mapper was introduced by Singh, Mémoli, and Carlsson [15], and has since been used in diverse applications, e.g [9, 10, 13, 16]. The authors of [15] showed for the first time that a cover for the codomain in addition to domains can be useful for data analysis. The mapper in some sense is connected to Reeb graphs (spaces) where the cover elements degenerate to points in the codomain, see [11] for example. The structure and stability of 1-dimensional mapper is studied in great details by Carrièr and Oudot in [2]. They showed that given a real valued function $f: X \to \mathbb{R}$ and an appropriate cover \mathcal{U} , the extended persistence diagram of a mapper $M(\mathcal{U}, f)$ is a subset of the same of the Reeb graph R_f . Furthermore, they characterized the features of the Reeb graph that may disappear from the mapper. The mapper (for a real-valued function f) can also be viewed as a Reeb graph $R_{f'}$ of a perturbed function $f': X' \to \mathbb{R}$. It is shown in [2] how one can track the changes between R_f and $R_{f'}$. In [1], the author established a convergence result between Mapper for a real valued f and the Reeb graph R_f . Specifically, the mapper is characterized with a zigzag persistence module that is a coarsening of the zigzag persistence module for R_f . It is shown that the mapper converges to R_f in the bottleneck distance of the corresponding zigzag persistence diagrams as the lengths of the intervals in the cover approaches zero. Munch and Wang [11] showed a similar convergence in interleaving distance using sheaf theory [4].

The multiscale mapper which work on the notion of a filtration of covers was developed in [5]. Most of the materials in this chapter are taken from this paper. The results on the class of 1-cycles that persist through multiscale mapper are taken from [6].

Exercises

- 1. For a simplicial complex K, simplices with no cofacet are called maximal simplices. Consider a closed cover of |K| with the closures of the maximal simplices as the cover elements. Let N(K) denote the nerve of this cover. Prove that N(N(K)) is isomorphic to a subcomplex of K.
- 2. A vertex v in K is called dominated by a vertex v' if every maximal simplex containing v also contains v'. We say K collapses strongly to a complex L if L is obtained by a series of deletions of dominated vertices with all their incident simplices. Show that K strongly collapses to N(N(K)).
- 3. We say a cover \mathcal{U} of a metric space (Y, d) is (α, β) -cover if $\alpha \leq \lambda(\mathcal{U})$ and $\beta \geq s_{max}(\mathcal{U})$.
 - Consider a δ-sample P of Y, that is, every metric ball B(y; δ), y ∈ Y, contains a point in P. Prove that the cover U = {B(p; 2δ)}_{p∈P} is a (δ, 4δ)-cover of Y.
 - Prove that the infinite cover $\mathcal{U} = \{B(y; \delta)\}_{y \in Y}$ is a $(\delta, 2\delta)$ -cover of *Y*.
- 4. Theorem 5 requires that the cover to be path connected. Show that this condition is necessary by presenting a counterexample otherwise.
- 5. One may generalize Theorem 5 as follows: If for any $k \ge 0$, *t*-wise intersections of cover elements for all t > 0 have trivial reduced homology for H_{k-t} , then the nerve map induces a surjection in H_k . Prove or disprove it.
- 6. Consider a function $f : X \to Z$ from a path connected space X to a metric space Z. Definite the equivalence relation \sim_f such that $x \sim_f x'$ holds 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 Reeb space R_f is the quotient of X under this equivalence relation.
 - Prove that the quotient map $q: X \to \mathsf{R}_f$ is surjective and also induces a surjection $q_*: \mathsf{H}_1(X) \to \mathsf{H}_1(\mathsf{R}_f).$
 - Call a class $[c] \in H_1(X)$ vertical if and only if there is no $c' \in C_1(X)$ so that [c] = [c']and $f \circ \sigma$ is constant for every $\sigma \in c'$. Show that $q_*([c]) \neq 0$ if and only if *c* is vertical.
 - Let z₁,..., z_g be an optimal cycle basis of H₁(X) defined with respect to the pseudometric d_f. Let l ∈ [1, g] be the smallest integer so that s(z_l) ≠ 0. Prove that if no such l exists, H₁(R_f) is trivial, otherwise, {[q(z_i)]}_{i=l,...g} is a basis for H₁(R_f).

7. Let us 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

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

Prove that $\tilde{\mathsf{d}}_f$ is a pseudo-metric.

- 8. Prove Proposition 17.
- 9. Prove Theorem 23.
- 10. Prove Claim 4.

Bibliography

- [1] Aravindakshan Babu. Zigzag coarsenings, mapper stability and gene-network analyses, 2013. PhD thesis, Stanford University.
- [2] Mathieu Carrière and Steve Oudot. Structure and stability of the one-dimensional mapper. *Found. Comput. Math.*, 18(6):1333–1396, 2018.
- [3] Frédéric Chazal, David Cohen-Steiner, Marc Glisse, Leonidas J. Guibas, and Steve Oudot. Proximity of persistence modules and their diagrams. In *Proc. 25th ACM Sympos. on Comput. Geom.*, pages 237–246, 2009.
- [4] Justin Curry. Sheaves, cosheaves and applications, 2014.
- [5] Tamal K. Dey, Facundo Mémoli, and Yusu Wang. Multiscale mapper: Topological summarization via codomain covers. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 997–1013, 2016.
- [6] Tamal K. Dey, Facundo Mémoli, and Yusu Wang. Topological analysis of nerves, reeb spaces, mappers, and multiscale mappers. In 33rd International Symposium on Computational Geometry, SoCG 2017, July 4-7, 2017, Brisbane, Australia, pages 36:1–36:16, 2017.
- [7] Allen Hatcher. Algebraic Topology. Cambridge University Press, Cambridge, 2002.
- [8] Jean Leray. Sur la forme des espaces topologiques et sur les points fixes des représentations. *J. Math. Pure Appl.*, 24:95–167, 1945.
- [9] Li Li, Wwi-Yi Cheng, Benjamin S. Glicksberg, Omri Gottesman, Ronald Tamler, Rong Chen, Erwin P. Bottinger, and Joel T. Dudley. Identification of type 2 diabetes subgroups through topological analysis of patient similarity. *Science Translational Medicine*, 7(311):311ra174, 2015.
- [10] P.Y. Lum, G. Singh, A. Lehman, T. Ishkhanikov, M. Vejdemo-Johansson, M. Alagap-pan, J. Carlsson, and G. Carlsson. Extracting insights from the shape of complex datausing topology. *Scientific reports*, 3, 2013.
- [11] Elizabeth Munch and Bei Wang. Convergence between categorical representations of reeb space and mapper. In Sándor P. Fekete and Anna Lubiw, editors, 32nd International Symposium on Computational Geometry, SoCG 2016, June 14-18, 2016, Boston, MA, USA, vol-

ume 51 of *LIPIcs*, pages 53:1–53:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2016.

- [12] James R. Munkres. *Elements of Algebraic Topology*. Addison–Wesley Publishing Company, Menlo Park, 1984.
- [13] Monica Nicolau, Arnold J. Levine, and Gunnar Carlsson. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proc. National Acad. Sci.*, 108.17:7265–7270, 2011.
- [14] Victor V. Prasolov. *Elements of combinatorial and differential topology*, volume 74. Amer. Math. Soc., 2006.
- [15] Gurjeet Singh, Facundo Mémoli, and Gunnar Carlsson. Topological methods for the analysis of high dimensional data sets and 3d object recognition. In *Proc. Eurographics Symposium* on *Point-Based Graphics* (2007), pages 91–100, 2007.
- [16] Brenda Y. Torres, Jose H. M. Oliveira, Ann Thomas Tate, Poonam Rath, Katherine Cumnock, and David S. Schneider. Tracking resilience to infections by mapping disease space. *Plos Biology*, 14(4):1–19, 2016.