Computational Topology and 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 9: Point cloud, homology inference

In this chapter, we focus on topological analysis of point cloud data (PCD), a common type of input data across a broad range of applications. Often, there is a hidden space of interest, and the PCD we obtain is only observations / samples from that hidden space. If the sample is sufficiently dense, it should carry information about the hidden space. We are interested in topological information in particular. However, discrete points themselves do not have interesting topology. To impose a connectivity that mimics that of the hidden space, we construct a simplicial complex such as the Rips or Čech complex using the points as vertices. Then, an appropriate filtration is considered as a proxy for the same on the topological space that the PCD presumably samples. This provides topological summaries such as the persistence diagrams induced by the filtrations. Figure 9.1 shows an example application of this approach. The PCD in this case represents atomic configurations of silica in three different states: liquid, glass, and crystal states. Each atomic configuration can be viewed as a set of weighted points, where each point represents the center of an atom and its weight is the radius of the atom. The persistence diagrams for the three states show distinctive features which can be used for further analysis of the phase transitions. The persistence diagrams can also be viewed as a signature of the input PCD and can be used to compare shapes (e.g, [6]) or provide other analysis.

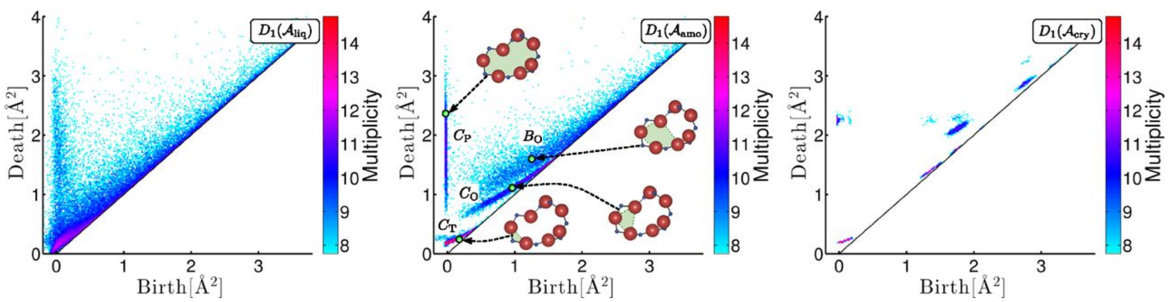


Figure 9.1: Persistence diagrams of silica in liquid (left), glass (middle), and crystal (right) states. Image taken from [21].

We mainly focus on PCD consisting of a set of points $P \subseteq (Z, \mathsf{d}_Z)$ embedded in some metric space Z equipped with a metric d_Z. One of the most common choices for (Z, d_Z) in practice is the d-dimensional Euclidean space $\mathbb{R}^d$ equipped with the standard L_p-distance. We review the relevant concepts of constructing Rips and Čech complexes, their filtrations, and describe the properties of the resulting persistence diagrams in Section 9.1. In practice, the size of a filtration can be prohibitively large. In Section 9.2, we discuss data sparsification strategies to approximate topological summaries much more efficiently and with theoretical guarantees.

As we have mentioned, a PCD can be viewed as a window through which we can peek at topological properties of the hidden space. In particular, we can infer about the hidden homological information using the PCD at hand if it samples the hidden space sufficiently densely. In Section 9.3, we provide such inference results for the cases when the hidden space is a manifold or is a compact set embedded in the Euclidean space. To obtain theoretical guarantees, we also need to introduce the language of sampling conditions to describe the quality of point samples. Finally, in Section 9.4, we focus on the inference of scalar field topology from a set of point samples P, as well as function values available at these samples. More precisely, we wish to estimate the persistent homology of a real-valued function $f : X \to \mathbb{R}$ from a set of discrete points $P \subset X$ as well as the values of f over P.

9.1 Persistence for Rips and Čech filtrations

Suppose we are given a finite set of points P in a metric space (Z, d_Z). Consider a closed ball $B_Z(p, r)$ with radius r (r-radius ball) centered at each point $p \in P$ and consider the space $P^r := \cup_{p \in P} B_Z(p, r)$. The Čech complex w.r.t. P and a parameter $r \geq 0$ is defined as

$$\mathbb{C}^r(P) = \{\sigma = \{p_0, \ldots, p_k\} \mid \cap_{i \in [0,k]} B_Z(p, r) \neq \emptyset\}. \tag{9.1}$$

As mentioned before, the Čech complex $\mathbb{C}^r(P)$ is the *nerve* of the union of balls P^r. If the metric balls centered at points in P in the metric space (Z, d_Z) are convex, then the Nerve Theorem gives the following corollary.

Corollary 1. *For a fixed $r \geq 0$, if the metric ball $B_Z(x, r)$ is convex for every $x \in P$, then $\mathbb{C}^r(P)$ is homotopy equivalent to P^r, and thus $\mathsf{H}_k(\mathbb{C}^r(P)) \cong \mathsf{H}_k(P^r)$ for any dimension $k \geq 0$.*

The above result justifies the utility of Čech complexes. For example, if $P \subseteq \mathbb{R}^d$ and d_Z is the standard L_p-distance for $p > 0$, then the Čech complex $\mathbb{C}^r(P)$ becomes homotopy equivalent to the union of r-radius balls centering points in P. Later in this chapter, we will also see an example where the points P are taken from a Riemannian manifold X equipped with the Riemannian metric d_X. When the radius r small enough, the intrinsic metric balls also become convex. In both cases, the resulting Čech complex captures information of the union of r-balls P^r.

In general, it is not clear at which scale (radius r) one should inspect the input PCD. Varying the scale parameter r, we thus obtain a filtration of spaces $\mathcal{P} := \{P^\alpha \hookrightarrow P^{\alpha'}\}_{\alpha \leq \alpha'}$ as well as a filtered sequence of simplicial complexes $\mathcal{C}(P) := \{\mathbb{C}^\alpha(P) \hookrightarrow \mathbb{C}^{\alpha'}(P)\}_{\alpha \leq \alpha'}$. The homotopy equivalence between P^r and $\mathbb{C}^r$, if holds, further induces an isomorphism between persistence modules obtained from these two filtrations.

Proposition 2 ([9])**.** *If the metric ball $B(x, r)$ is convex for every $x \in P$ and all $r \geq 0$, then the persistent module $\mathsf{H}_k\mathcal{P}$ is isomorphic to the persistent module $\mathsf{H}_k\mathcal{C}(P)$. This also implies that their corresponding persistence diagrams are identical:* $\mathrm{Dgm}_k\mathcal{P} = \mathrm{Dgm}_k\mathcal{C}(P)$*, for any dimension $k \geq 0$.*

A related persistence-based topological invariant is given by the *Vietoris-Rips filtration* $\mathbb{R}\mathrm{ips}(P) = \{\mathbb{VR}^\alpha(P) \hookrightarrow \mathbb{VR}^{\alpha'}(P)\}_{\alpha \leq \alpha'}$, where *Vietoris-Rips complex* $\mathbb{VR}^r(P)$ for a finite subset $P \subseteq (Z, \mathsf{d}_Z)$ at scale r is defined as

$$\mathbb{VR}^r(P) = \{\sigma = \{p_0, \ldots, p_k\} \mid \mathsf{d}_Z(p_i, p_j) \leq 2r \text{ for any } i, j \in [0, k]\}. \tag{9.2}$$

Recall from previous topics that the Čech filtration and Vietoris-Rips filtration are multiplicatively 2-interleaved, meaning that their persistence modules are 2-interleaved at the log-scale, and

$$\mathsf{d}_b(\mathrm{Dgm}_{\log}\mathcal{C}(P), \mathrm{Dgm}_{\log}\mathbb{R}\mathrm{ips}(P)) \leq \log 2, \quad \text{(Corollary ??)}. \tag{9.3}$$

Finite metric spaces. The above definitions of Čech or Rips complexes assume that P is embedded in an ambient metric space (Z, d_Z). It is possible that $Z = P$ and we simply have a discrete metric space spanned by points in P, which we denote by (P, d_P). Obviously, the construction of Čech and Rips complexes can be extended to this case. In particular, the Čech complex $\mathbb{C}^r_P(P)$ is now defined as

$$\mathbb{C}^r_P(P) = \{\sigma = \{p_0, \ldots, p_k\} \mid \cap_{i\in[0,k]} B_P(p;r) \neq \emptyset\}, \tag{9.4}$$

where $B_P(p,r) := \{q \in P \mid \mathsf{d}_P(p,q) \leq r\}$. However, note that when $P \subset Z$ and d_P is the restriction of the metric d_Z to points in P, the Čech complex $\mathbb{C}^r_P(P)$ defined above can be *different* from the Čech complex $\mathbb{C}^r_Z(P)$, as the metric balls (B_P vs. B_Z) are different. In particular, in this case, we have the following relation between the two types of Čech complexes:

$$\mathbb{C}^r_P(P) \subseteq \mathbb{C}^r_Z(P) \subseteq \mathbb{C}^{2r}_P(P). \tag{9.5}$$

On the other hand, in this setting, the two Rips complexes are the same because the definition of Rips complex involves only pairwise distance between input points, not metric balls. In what follows, we often omit the subscript P or Z for the Čech complex when its choice is clear.

The persistence diagrams induced by the Čech and the Rips filtrations can be used as topological summaries for the input PCD P. We can then for example, compare input PCDs by comparing these persistence diagram summaries.

Definition 1 (Čech, Rips distance)**.** Given two finite point sets P and Q, equipped with appropriate metrics, the *Čech distance* between them is a pseudo-distance defined as:

$$\mathsf{d}_{Cech}(P,Q) = \max_k \mathsf{d}_B(\mathrm{Dgm}_k \mathcal{C}(P), \mathrm{Dgm}_k \mathcal{C}(Q))).$$

Similarly, the *Rips distance* between P and Q is a pseudo-distance defined as:

$$\mathsf{d}_{Rips}(P,Q) = \max_k \mathsf{d}_B(\mathrm{Dgm}_k \mathbb{R}\mathrm{ips}(P), \mathrm{Dgm}_k \mathbb{R}\mathrm{ips}(Q))).$$

These distances are stable with respect to the Hausdorff or the Gromov-Hausdorff distance between P and Q depending on whether they are embedded in a common metric space or are viewed as two discrete metric spaces (P, d_P) and (Q, d_Q)). We introduce the Hausdorff and Gromov-Hausdorff distances now. Given a point x and a set A from a metric space (X, d), let $\mathsf{d}(x, A) := \inf_{a\in A} \mathsf{d}(x, a)$ denote the closest distance from x to any point in A.

Definition 2 (Hausdorff distance)**.** Given two compact sets $A, B \subseteq (Z, \mathsf{d}_Z)$, the *Hausdorff distance* between them is defined as:

$$\mathsf{d}_H(A,B) = \max\{\ \max_{a\in A} \mathsf{d}_Z(a,B),\quad \max_{b\in B} \mathsf{d}_Z(b,A)\ \}.$$

Note that the Hausdorff distance requires the input objects to be embedded in a common ambient space. In case they are not embedded in any common ambient space, we use Gromov-Hausdorff distance, which intuitively measures how much two input metric spaces differ from being isometric.

Definition 3 (Gromov-Hausdorff distance)**.** Given two metric spaces (X, d_X) and (Y, d_Y), a *correspondence* C is a subset $C \subseteq X \times Y$ so that (i) for every $x \in X$, there exists some $(x, y) \in C$; and (ii) for every $y' \in Y$, there exists some $(x', y') \in C$. The *distortion induced by* C is

$$\mathrm{distort}_C(X, Y) := \frac{1}{2} \sup_{(x,y),(x',y') \in C} |\mathsf{d}_X(x, x') - \mathsf{d}_Y(y, y')|.$$

The *Gromov-Hausdorff distance between* (X, d_X) *and* (Y, d_Y) is the smallest distortion possible by any correspondence; that is,

$$\mathsf{d}_{GH}(X, Y) := \inf_{C \subseteq X \times Y} \mathrm{distort}_C(X, Y).$$

Theorem 3. *Čech- and Rips-distances satisfy the following stability statements:*

1. *Given two finite sets $P, Q \subseteq (Z, \mathsf{d}_Z)$, we have*

$$\mathsf{d}_{Cech}(P, Q) \leq \mathsf{d}_H(P, Q); \quad \textit{and} \quad \mathsf{d}_{Rips}(P, Q) \leq \mathsf{d}_H(P, Q).$$

2. *Given two finite metric spaces (P, d_P) and (Q, d_Q), we have*

$$\mathsf{d}_{Cech}(P, Q) \leq 2\mathsf{d}_{GH}((P, \mathsf{d}_P), (Q, \mathsf{d}_Q)), \quad \textit{and}\ \mathsf{d}_{Rips}(P, Q) \leq \mathsf{d}_{GH}((P, \mathsf{d}_P), (Q, \mathsf{d}_Q)).$$

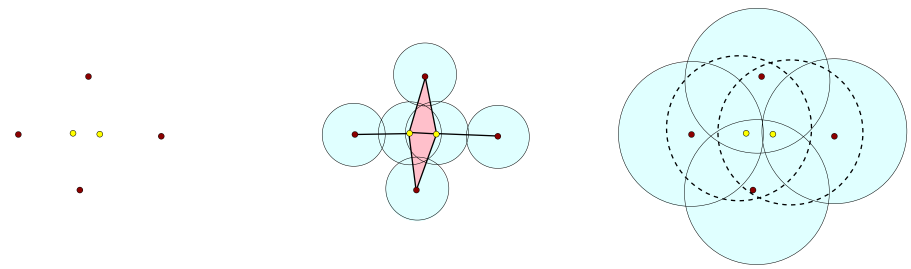
Note that the bound on $\mathsf{d}_{Cech}(P, Q)$ in statement (2) of the above theorem has an extra factor of 2, which comes due to the difference in metric balls – see the discussions after Eqn (9.4). We also remark that (2) in the above theorem can be extended to the so-called *totally bounded metric spaces* (which are not necessarily finite) (P, d_P) and (Q, d_Q) defined as follows. First, recall that *an ε-sample* of a metric space (Z, d_Z) is a finite set $S \subseteq Z$ so that for every $z \in Z$, $\mathsf{d}_Z(z, S) \leq \varepsilon$. A metric space (Z, d_Z) is *totally bounded* if there exists a finite ε-sample for every $\varepsilon > 0$. Intuitively, such a metric space can be approximated by a finite metric space for any resolution.

9.2 Approximation via data sparsification

One issue with using the Vietoris-Rips or Čech filtrations in practice is that their size can become huge, even for moderate number of points. For example, when the scale r is larger than the diameter of a point set P, the Čech and the Vietoris-Rips complexes of P contain every simplex spanned by points in P, in which case the size of d-skeleton of $\mathbb{C}^r(P)$ or $\mathbb{VR}^r(P)$ is $\Theta(n^{d+1})$ for $n = |P|$.

On the other hand, as shown in Figure 9.2, as the scale r increases, certain points could become "redundant", e.g, having no or little contribution to the underlying space of the union of all r-radius balls. Based on this observation, one can approximate these filtrations with *sparsified filtrations* of much smaller size. In particular, as the scale r increases, the point set P with which one constructs a complex is gradually sparsified keeping the total number of simplicies in the complex linear in the input size of P where the dimension of the embedding space is assumed to be fixed.

We describe two data sparsification schemes in Sections 9.2.1 and 9.2.2, respectively. We focus on the Vietoris-Rips filtration for points in a Euclidean space $\mathbb{R}^d$ equipped with the standard Euclidean distance d.



(a) (b) (c)

Figure 9.2: Vietoris-Rips complex: (b) at small scale, the Rips complex of points shown in (a) requires the two yellow points; (c) the two yellow points become redundant at larger scale.

9.2.1 Data sparsification for Rips filtration via reweighting

Most of the concepts presented in this section apply to general finite metric spaces though we describe them for finite point sets equipped with an Euclidean metric. The reason for this choice is that the complexity analysis draws upon the specific property of Euclidean space. The reader is encouraged to think about generalizing the definitions and the technique to other metric spaces.

Definition 4 (Nets and net-tower)**.** Given a finite set of points $P \subset (\mathbb{R}^d, \mathsf{d})$ and $\gamma \geq 0, \gamma' \geq 0$, a subset $Q \subseteq P$ is a (γ, γ')-*net of* P if the following two conditions hold:

Covering condition: Q is a γ-sample for (P, d), i.e., for every $p \in P$, $\mathsf{d}(p, Q) \leq \gamma$.

Packing condition: Q is also γ-sparse, i.e., for every $q \neq q' \in Q$, $\mathsf{d}(q, q') \geq \gamma'$.

If $\gamma = \gamma'$, we also refer to Q as a γ-*net of* P.

A single-parameter family of nets $\{N_\gamma\}_\gamma$ is called a *net-tower* of P if (i) there is a constant $c > 0$ so that for all $\gamma \in \mathbb{R}$, N_γ is a $(\gamma, \gamma/c)$-net for P, and (ii) $N_\gamma \supseteq N_{\gamma'}$ for any $\gamma \leq \gamma'$.

Intuitively, a γ-net approximates a PCD P at resolution γ (Covering condition), while also being sparse (Packing condition). A net-tower provides a sequence of increasingly sparsified approximation of P.

Net-tower via farthest point sampling. We now introduce a specific net-tower constructed via the classical strategy of *farthest point sampling*, also called greedy permutation e.g. in [3, 4]. Given a point set $P \subset (\mathbb{R}^d, \mathsf{d})$, choose an arbitrary point p_1 from P and set $P_1 = \{p_1\}$. Pick p_i recursively as $p_i \in \mathrm{argmax}_{p \in P \setminus P_{i-1}} \mathsf{d}(p, P_{i-1})$[1], and set $P_i = P_{i-1} \cup \{p_i\}$. Now set $\mathsf{t}_{p_i} = \mathsf{d}(p_i, P_{i-1})$, which we refer to as the *exit-time of* p_i. Based on this exit-times, we construct the following two families of sets:

$$\textit{Open net-tower } \mathcal{N} = \{N_\gamma\}_{\gamma \in \mathbb{R}} \text{ where } N_\gamma := \{p \in P \mid \mathsf{t}_p > \gamma\}. \tag{9.6}$$

$$\textit{Closed net-tower } \overline{\mathcal{N}} = \{\overline{N}_\gamma\}_{\gamma \in \mathbb{R}} \text{ where } \overline{N}_\gamma := \{p \in P \mid \mathsf{t}_p \geq \gamma\}. \tag{9.7}$$

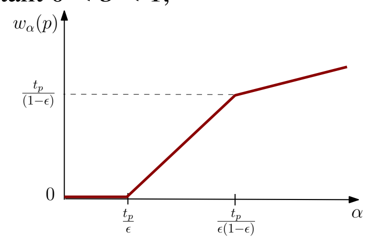
[1] Note that there may be multiple points that maximize $\mathsf{d}(p, P_{i-1})$ making $\mathrm{argmax}_{p \in P \setminus P_{i-1}} \mathsf{d}(p, P_{i-1})$ a set. We can choose p_i to be any point in this set.

It is easy to verify that both N_γ and $\overline{N}_\gamma$ are γ-nets, and the families $\mathcal{N}$ and $\overline{\mathcal{N}}$ are indeed two net-towers as γ increases. As γ increases, N_γ and $\overline{N}_\gamma$ can only change when $\gamma = t_p$ for some $p \in P$. Hence the sequence of subsets $P = P_n \supset P_{n-1} \supseteq \cdots \supseteq P_2 \supseteq P_1$ contain all the distinct sets in the open and close net-towers $\{N_\gamma\}$ and $\{\overline{N}_\gamma\}$.

In what follows, we discuss a sparsification strategy for the Rips filtration of P using the above net-towers. The approach can be extended to other net-towers, such as the net-tower constructed using the net-tree data structure of [20].

Weights, weighted distance, and sparse Rips filtration. Given the exit-time t_ps for all points $p \in P$, we now associate a *weight* $w_p(\alpha)$ for each point p at a scale α as follows (the graph of this weight function is shown on the right): for some constant $0 < \varepsilon < 1$,

$$w_p(\alpha) = \begin{cases} 0 & \frac{\mathsf{t}_p}{\varepsilon} \geq \alpha \\ \alpha - \frac{\mathsf{t}_p}{\varepsilon} & \frac{\mathsf{t}_p}{\varepsilon} < \alpha < \frac{\mathsf{t}_p}{\varepsilon(1-\varepsilon)} \\ \varepsilon\alpha & \frac{\mathsf{t}_p}{\varepsilon(1-\varepsilon)} \leq \alpha \end{cases}$$



Claim 1. *The weight function w_p is a continuous, 1-Lipschitz, and non-decreasing function on α.*

The parameter ε controls the resolution of the sparsification. The *net-induced distance at scale α* between input points is defined as:

$$\widehat{\mathsf{d}}_\alpha(p,q) := \mathsf{d}(p,q) + w_p(\alpha) + w_q(\alpha). \tag{9.8}$$

Definition 5 (Sparse (Vietoris-)Rips)**.** Given a set of points $P \subset \mathbb{R}^d$, a constant $0 < \varepsilon < 1$, and the open net-tower $\{N_\gamma\}$ as well as the closed net-tower $\{\overline{N}_\gamma\}$ for P as introduced above, the *open sparse-Rips complex at scale α* is defined as

$$\mathsf{Q}^\alpha := \{\sigma \subseteq N_{\varepsilon(1-\varepsilon)\alpha} \mid \forall p,q \in \sigma,\ \widehat{\mathsf{d}}_\alpha(p,q) \leq 2\alpha\}; \tag{9.9}$$

while the *closed sparse-Rips at scale α* is defined as

$$\overline{\mathsf{Q}}^\alpha := \{\sigma \subseteq \overline{N}_{\varepsilon(1-\varepsilon)\alpha} \mid \forall p,q \in \sigma,\ \widehat{\mathsf{d}}_\alpha(p,q) \leq 2\alpha\}. \tag{9.10}$$

Set $\mathsf{S}^\alpha := \cup_{\beta \leq \alpha} \overline{\mathsf{Q}}^\alpha$, called the *cumulative complex at scale α*. The *(ε-)sparse Rips filtration* then refers to the $\mathbb{R}$-indexed filtration $\mathcal{S} = \{\mathsf{S}^\alpha \hookrightarrow \mathsf{S}^\beta\}_{\alpha \leq \beta}$.

Obviously, $\mathsf{Q}^\alpha \subseteq \overline{\mathsf{Q}}^\alpha$. Note that for $\alpha < \beta$, Q^α is not necessarily included in Q^β (neither is $\overline{\mathsf{Q}}^\alpha$ in $\overline{\mathsf{Q}}^\beta$); while the inclusion $\mathsf{S}^\alpha \subseteq \mathsf{S}^\beta$ always holds.

In what follows, we show that the sparse Rips filtration approximates the standard Vietoris-Rips filtration $\{\mathbb{VR}^r(P)\}$ defined over P, and that the size of the sparse Rips filtration is only linear in n for any fixed dimension d which is assumed to be constant. The main results are summarized in the following theorem.

Theorem 4. *Let $P \subset \mathbb{R}^d$ be a set of n points where d is a constant, and $\mathbb{R}\mathrm{ips}(P) = \{\mathbb{VR}^r(P)\}$ be the Vietoris-Rips filtration over P. Given net-towers $\{N_\gamma\}$ and $\{\overline{N}_\gamma\}$ induced by exit-times $\{\mathsf{t}_p\}_{p\in P}$, let $\mathcal{S}(P) = \{\mathsf{S}^\alpha\}$ be its corresponding ε-sparse Rips filtration as defined in Definition 5. Then, for a fixed* $0 < \varepsilon < \frac{1}{3}$, *(i)* $\mathbb{S}(P)$ *and* $\mathbb{R}\mathrm{ips}(P)$ *are multiplicatively* $\frac{1}{1-\varepsilon}$*-interleaved at the homology level. Thus, for any* $k \geq 0$*, the persistence diagram* $\mathrm{Dgm}_k \mathbb{S}(P)$ *is a* $\log \frac{1}{1-\varepsilon}$*-approximation of* $\mathrm{Dgm}_k \mathbb{R}\mathrm{ips}(P)$ *at the log-scale.*

(ii) For any fixed dimension $k \geq 0$*, the total number of* k*-simplices ever appeared in* $\mathbb{S}(P)$ *is* $\Theta((\frac{1}{\varepsilon})^{kd} n)$*.*

9.2.2 Approximation via simplicial tower

We now describe a different sparsification strategy by directly building a simplicial tower of Rips complexes connected by simplicial maps whose persistent homology also approximates that of the standard Rips-filtration. This sparsification is conceptually simpler, but its approximation quality is worse than the one introduced in the previous section.

Given a set of points $P \subset \mathbb{R}^d$, $\alpha > 0$, and some $0 < \varepsilon < 1$, we consider the filtration (which is a subsequence of the standard Rips filtration)

$$\mathbb{VR}^{\alpha}(P) \hookrightarrow \mathbb{VR}^{\alpha(1+\varepsilon)}(P) \hookrightarrow \mathbb{VR}^{\alpha(1+\varepsilon)^2}(P) \hookrightarrow \cdots \hookrightarrow \mathbb{VR}^{\alpha(1+\varepsilon)^m}(P). \tag{9.11}$$

We construct a sparsified sequence by seting $P_0 := P$, building a sequence of point sets P_k, $k = 0, 1, \ldots, m$ where P_{k+1} is a $\frac{\alpha\varepsilon}{2}(1+\varepsilon)^{k-1}$-net of P_k, and terminating the process when P_m is of constant size.

Consider the following vertex map $\pi_k : P_k \to P_{k+1}$, for any $k \in [0, m-1]$, where $\pi_k(v)$ is the nearest neighbor of $v \in P_k$ in P_{k+1}. Define $\widehat{\pi}_k : P_0 \to P_{k+1}$ as $\widehat{\pi}_k := \pi_k \circ \cdots \circ \pi_0$. Based on the fact that P_{k+1} is a $\frac{\alpha\varepsilon}{2}(1+\varepsilon)^{k-1}$-net of P_k, it can be verified that π_k induces a simplicial map

$$\pi_k : \mathbb{VR}^{\alpha(1+\varepsilon)^k}(P_k) \to \mathbb{VR}^{\alpha(1+\varepsilon)^{k+1}}(P_{k+1})$$

which further gives rise to a simplicial map $\widehat{\pi}_k : \mathbb{VR}^{\alpha}(P_0) \to \mathbb{VR}^{\alpha(1+\varepsilon)^{k+1}}(P_{k+1})$. We thus have the following tower of simplicial complexes:

$$\widehat{\mathbb{S}} : \mathbb{VR}^{\alpha}(P_0) \xrightarrow{\pi_0} \mathbb{VR}^{\alpha(1+\varepsilon)}(P_1) \xrightarrow{\pi_1} \cdots \xrightarrow{\pi_{m-1}} \mathbb{VR}^{\alpha(1+\varepsilon)^m}(P_m). \tag{9.12}$$

Claim 2. *For any fixed* $\alpha \geq 0$*,* $\varepsilon \geq 0$*, and any integer* $k \geq 0$*, each triangle in the following diagram commutes at the homology level:*

$$\begin{tikzcd}
\mathbb{VR}^{\alpha(1+\varepsilon)^k}(P_0) \arrow[r, hook, "i_k"] \arrow[dr, "\widehat{\pi}_k"] & \mathbb{VR}^{\alpha(1+\varepsilon)^{k+1}}(P_0) \\
\mathbb{VR}^{\alpha(1+\varepsilon)^k}(P_k) \arrow[u, hook, "j_k"] \arrow[r, "\pi_k"] & \mathbb{VR}^{\alpha(1+\varepsilon)^{k+1}}(P_{k+1}) \arrow[u, hook, "j_{k+1}"']
\end{tikzcd}$$

Here, the maps i_k*s and* j_k*s are canonical inclusions.*

The above result implies that at the homology level, the sequence in Eqn (9.12) and the sequence Eqn (9.11) are *weakly* $(1+\varepsilon)$*-interleaved* in a multiplicative manner. In particular, different from the interleaving introduced before, here the interleaving relations only hold at *discrete index values* of the filtrations.

Definition 6 (Weakly interleaving of vector space towers)**.** Let $\mathbb{U} = \{\mathrm{U}_a \xrightarrow{u_{a,b}} \mathrm{U}_b\}_{a_0 \le a \le b}$ and $\mathbb{V} = \{\mathrm{V}_a \xrightarrow{v_{a,b}} \mathrm{V}_b\}_{a_0 \le a \le b}$ be two vector space towers over an index set $A = \{a \in \mathbb{R} \mid a \ge a_0\}$ with resolution $a_0 \ge 0$. For some real number $\varepsilon \ge 0$, we say that they are *weakly ε-interleaved* if there are two families of linear maps $\phi_i : \mathrm{U}_{a_0+i\varepsilon} \to \mathrm{V}_{a_0+(i+1)\varepsilon}$, and $\psi_i : \mathrm{V}_{a_0+i\varepsilon} \to \mathrm{U}_{a_0+(i+1)\varepsilon}$, for any integer $i \ge 0$, such that any subdiagram of the following diagram commutes:

$$\begin{tikzcd}
\mathbb{U}: & \mathrm{U}_{a_0} \arrow[r] \arrow[dr] & \mathrm{U}_{a_0+\varepsilon} \arrow[r] \arrow[dr] & \mathrm{U}_{a_0+2\varepsilon} \arrow[r] \arrow[dr] & \cdots\cdots \arrow[r] & \mathrm{U}_{a_0+m\varepsilon} \arrow[r] \arrow[dr] & \cdots \\
\mathbb{V}: & \mathrm{V}_{a_0} \arrow[r] \arrow[ur] & \mathrm{V}_{a_0+\varepsilon} \arrow[r] \arrow[ur] & \mathrm{V}_{a_0+2\varepsilon} \arrow[r] \arrow[ur] & \cdots\cdots \arrow[r] & \mathrm{V}_{a_0+m\varepsilon} \arrow[r] \arrow[ur] & \cdots
\end{tikzcd} \tag{9.13}$$

It turns out that to verify the commutability conditions of the diagram in Eqn (9.13), it is sufficient to verify it for all subdiagrams (rectangular and triangular commutativity) as we have seen before for interleaving of persistence modules. Furthermore, ε-weakly interleaved persistence modules also have bounded bottleneck distances between their persistence diagrams [10] though the distance bound is relaxed to 3ε, that is, if $\mathbb{U}$ and $\mathbb{V}$ are weakly-ε interleaved, then $\mathsf{d}_b(\mathrm{Dgm}\mathbb{U}, \mathrm{Dgm}\mathbb{V}) \le 3\varepsilon$. Analogous results hold for multiplicative setting. Finally, using a similar packing argument as before, one can also show that the total number of k-simplices that ever appear in the simplicial-map based sparsification $\widehat{\mathbb{S}}$ is linear in n (assuming that k and the dimension d are both constant). To summarize:

Theorem 5. *Given a set of n points $P \subset \mathbb{R}^d$, we can $3\log(1+\varepsilon)$-approximate the persistence diagram of the discrete Rips filtration in Eqn (9.11) by that of the filtration in Eqn (9.12) at the log-scale. The number of k-simplices that ever appear in the filtration in Eqn (9.12) is $O((\frac{1}{\varepsilon})^{O(kd)}n)$.*

9.3 Homology inference from PCDs

So far, we considered the problem of approximating the persistence diagram of a filtration created out of a given PCD. Now we consider the problem of inferring certain homological structure of a (hidden) domain where the input PCD presumably is sampled from. More specifically, the problem we consider is: Given a finite set of points $P \subset \mathbb{R}^d$, residing on or around a hidden domain $X \subseteq \mathbb{R}^d$ of interest, compute or approximate the rank of $\mathsf{H}_*(X)$ using input PCD P. Later in this chapter, X is assumed to be either a smooth Riemannian manifold embedded in $\mathbb{R}^d$, or simply a compact set of $\mathbb{R}^d$.

Main ingredients. Since points themselves do not have interesting topology, we first construct a certain simplicial complex K, typically a Čech or a Vietoris-Rips complex from P. Next, we compute the homological information of K as a proxy for the same of X. Of course, the approximation becomes faithful only when the given sample P is sufficiently dense and the parameters used for building the complexes are appropriate. The high level approach works as follows.

Input: A finite point set $P \subset \mathbb{R}^d$ "approximating" a hidden space $X \subset \mathbb{R}^d$.

Step 1. Compute the Čech complex $\mathbb{C}^\alpha(P)$, or a pair of Rips complexes $\mathbb{VR}^\alpha(P)$ and $\mathbb{VR}^{\alpha'}(P)$ for some appropriate $0 < \alpha < \alpha'$.

Step 2. In the case of Čech complex, return $\dim(\mathsf{H}_*(\mathbb{C}^\alpha(P)))$ as an approximation of $\dim(\mathsf{H}_*(X))$. In the case of Rips complex, return rank (im (i_*)), where the homomorphism $i_* : \mathsf{H}_*(\mathbb{VR}^\alpha(P)) \to \mathsf{H}_*(\mathbb{VR}^{\alpha'}(P))$ is induced by the inclusion $\mathbb{VR}^\alpha(P) \subseteq \mathbb{VR}^{\alpha'}(P)$.

To provide quantitative statements on the approximation quality of the outcome of the above approach, we need to describe first what the quality of the input PCD P is, often referred to as the *sampling conditions*. Intuitively, a better approximation in homology is achieved if the input points P "approximates" / "samples" X better. The quality of input points is often measured by the Hausdorff distance measured with Euclidean distances between PCD P and the hidden domain X of interest (Definition 2), such as requiring that $\mathsf{d}_H(P, X) \leq \varepsilon$ for some $\varepsilon > 0$. Note that points in P do not necessarily lie in X. The approximation guarantee for $\dim(\mathsf{H}_*(X))$ relies on relating the distance fields induced by X and by the sample P. We describe the distance field and feature sizes of X in Section 9.3.1. We present how to infer homology for smooth manifolds and compact sets from data in Section 9.3.2 and Section 9.3.3 respectively. In Section 9.4, we discuss inferring the persistent homology induced by a scalar function $f : X \to \mathbb{R}$ on X.

9.3.1 Distance field and feature sizes

To describe how well P samples X, we introduce two notions of the so-called "feature size" of X: the local feature size and the weak feature size, both related to the distance field d_X w.r.t. X.

Definition 7 (Distance field)**.** Given a compact set $X \subset \mathbb{R}^d$, the *distance field (w.r.t. X)* is

$$\mathsf{d}_X : \mathbb{R}^d \to \mathbb{R},\ x \mapsto \mathsf{d}(x, X).$$

The *α-offset of* X is defined as $X^\alpha := \{x \in \mathbb{R}^d \mid \mathsf{d}_X(x) \leq \alpha\}$, which is simply the sub-level set $\mathsf{d}_X^{-1}((-\infty, \alpha])$ of d_X.

Given $x \in \mathbb{R}^d$, let $\Pi(x) \in X$ denote the set of closest points of x in X; that is,

$$\Pi(x) = \{y \in X \mid \mathsf{d}(x, y) = \mathsf{d}_X(x)\}.$$

The *medial axis of* X, denoted by $\mathcal{M}_X$, is the closure of the set of points with more than one closest point in X; that is,

$$\mathcal{M}_X = \text{closure}\{x \in \mathbb{R}^d \ \mid\ |\Pi(x)| \geq 2\}.$$

Intuitively, $|\Pi(x)| \geq 2$ implies that the maximal Euclidean ball ceneterd at x whose interior is free of points in X meets X in more than one point on its boundary. Hence, $\mathcal{M}_X$ is the closure of the centers of such maximal empty balls.

Definition 8 (Local feature size and reach)**.** For a point $x \in X$, the *local feature size at* x, denoted by $\text{lfs}(x)$, is defined as the minimum distance to the medial axis $\mathcal{M}_X$; that is,

$$\text{lfs}(x) := \mathsf{d}(x, \mathcal{M}_X).$$

The *reach of* X, denoted by $\rho(X)$, is the minimum local feature size of any point in X.

The concept has been primarily developed for the case when X is a smooth manifold embedded in $\mathbb{R}^d$. Indeed, the local feature size can be zero at a non-smooth point: consider a planar polygon; its medial axis intersects its vertices, and the local feature size at a vertex is thus zero. The reach of a smoothly embedded manifold could also be zero; see Section 1.2 of [15] for an example. Next, we describe a "weaker" notion of feature size [8, 13], which is more suitable for compact subsets of $\mathbb{R}^d$.

Critical points of distance field. The distance function d_X introduced above is not everywhere differentiable. Its gradient is defined on $\mathbb{R}^d \setminus \{X \cup \mathcal{M}_X\}$. However, one can still define the following vector which extends the notion of gradient of d_X to include the medial axis $\mathcal{M}_X$: Given any point $x \in \mathbb{R}^d \setminus X$, there exists a unique closed ball with minimal radius that encloses $\Pi(x)$ [22]. Let $c(x)$ denote the center of this minimal enclosing ball, and $r(x)$ its radius. It is easy to see that for any $x \in \mathbb{R}^d \setminus \mathcal{M}_X$, this ball and $c(x)$ degenerates to the unique point in $\Pi(x)$.

Definition 9 (Generalized vector field)**.** Define the following vector field $\nabla_d : \mathbb{R}^d \setminus X \to \mathbb{R}^d$ where the *(generalized) gradient vector* at $x \in \mathbb{R}^d \setminus X$ is:

$$\nabla_{\mathsf{d}}(x) = \frac{x - c(x)}{\mathsf{d}_X(x)}.$$

The *critical points* of ∇_{d} are points x for which $\nabla_{\mathsf{d}}(x) = 0$.

This generalized gradient field ∇_{d} coincides with the gradient of the distance function d_X for points in $\mathbb{R}^d \setminus \{X \cup \mathcal{M}_X\}$. We also call the critical points of ∇_{d} the critical points of the distance function d_X. The distance field (distance function) and its critical points were previously studied in e.g., [19], and have played an important role in sampling theory and homology inference. In general, a point x is a critical point if and only if $x \in \mathbb{R}^d \setminus X$ is contained in the convex hull of $\Pi(x)$ (The convex hull of a compact set $A \subset \mathbb{R}^d$ is the smallest convex set that contains A). It is necessary that all critical points of ∇_{d} belong to the medial axis $\mathcal{M}_X$ of X. For the case where X is a finite set of points in $\mathbb{R}^d$, the critical points of d_X are the non-empty intersections of the Delaunay simplices with their dual Voronoi cells (if exist) [15].

Definition 10 (Weak feature size)**.** Let C denote the set of critical points of ∇_{d}. The *weak feature size of* X, denoted by $\mathrm{wfs}(X)$, is the distance between X and C; that is,

$$\mathrm{wfs}(X) = \min_{x \in X} \inf_{c \in C} \mathsf{d}(x, c).$$

Proposition 6. *If* $0 < \alpha < \alpha'$ *are such that there is no critical value of* d_X *in the closed interval* $[\alpha, \alpha']$*, then* $X^{\alpha'}$ *deformation retracts onto* X^{α}*. In particular, this implies that* $\mathsf{H}_*(X^\alpha) \cong \mathsf{H}(X^{\alpha'})$.

In the homology inference frameworks, the reach is usually used for the case when X is a smooth embedded manifold, while the weak feature size is used for general compact spaces.

9.3.2 Data on manifold

We now consider the problem of homology inference from a point sample of a manifold. We first state a standard result from linear algebra (see also the Sandwich Lemma from [9]), which we use several times in homology inference.

Fact 1. *Given a sequence $A \to B \to C \to D \to E \to F$ of homomorphisms (linear maps) between finite-dimensional vector spaces over some field, if* $\mathrm{rank}\,(A \to F) = \mathrm{rank}\,(C \to D)$*, then this quantity also equals* $\mathrm{rank}\,(B \to E)$.

Specifically, if $A \to B \to C \to E \to F$ is a sequence of homomorphisms such that $\mathrm{rank}\,(A \to F) = \dim C$*, then* $\mathrm{rank}\,(B \to E) = \dim C$.

Let P be a point set sampled from a manifold $X \subset \mathbb{R}^d$. We construct either the Čech complex $\mathbb{C}^\alpha(P)$, or a pair of Rips complexes $\mathbb{VR}^\alpha(P) \hookrightarrow \mathbb{VR}^{2\alpha}(P)$ for some parameter $\alpha > 0$. The homology groups of these spaces are related as follows.

$$\mathsf{H}(X) \xleftrightarrow{\text{Prop. 7}} \mathsf{H}(P^\alpha) \xleftrightarrow{\text{Nerve Lemma}} \mathsf{H}(\mathbb{C}^\alpha(P)) \xleftrightarrow{\text{Fact 1}} \mathrm{image}(\mathsf{H}(\mathbb{VR}^\alpha) \to \mathsf{H}(\mathbb{VR}^{2\alpha})) \tag{9.14}$$

Specifically, recall that A^r is the r-offset of A which also equals the union of balls $\cup_{a \in A} B(a, r)$. The connection between the discrete samples P and the manifold X is made through the union of balls P^α. The following result is a variant of a result by Niyogi, Smale, Weinberger [23][2].

Proposition 7. *Let P be a finite point set be such that $\mathsf{d}_H(X, P) \le \varepsilon$ where $X \subset \mathbb{R}^d$ is a smooth manifold with reach $\rho(X)$. If $3\varepsilon \le \alpha \le \frac{3}{4}\sqrt{\frac{3}{5}}\rho(X)$, then $\mathsf{H}_*(P^\alpha)$ is isomorphic to $\mathsf{H}_*(X)$.*

The Čech complex $\mathbb{C}^\alpha(P)$ is the nerve complex for the set of balls $\{B(p, \alpha), p \in P\}$. As Euclidean balls are convex, Nerve Lemma implies that $\mathbb{C}^\alpha(P)$ is homotopy equivalent to P^α. It follows that we can use the Čech complex $\mathbb{C}^\alpha(P)$, for an appropriate α, to infer homology of X using the isomorphisms $\mathsf{H}_*(X) \cong \mathsf{H}_*(P^\alpha) \cong \mathsf{H}_*(\mathbb{C}^\alpha(P))$. The first isomorphism follows from Proposition 7 and the second one from the homotopy equivalence between the nerve and space.

A stronger statement in fact holds: For any $\alpha \le \beta$, the following diagram commutes:

$$\begin{tikzcd} \mathsf{H}_*(P^\alpha) \arrow[r, "i_*"] \arrow[d, "h_*"] & \mathsf{H}_*(P^\beta) \arrow[d, "h_*"] \\ \mathsf{H}_*(\mathbb{C}^\alpha(P)) \arrow[r, "i_*"] & \mathsf{H}_*(\mathbb{C}^\beta(P)) \end{tikzcd} \tag{9.15}$$

Here, i_* stands for the homomorphism induced by inclusions, and h_* is the homomorphism induced by the homotopy equivalence $h : P^\alpha \to \mathbb{C}^\alpha(P)$. We can now state the following theorem on estimating $\mathsf{H}_*(X)$ from a pair of Rips complexes.

Theorem 8. *Given a smooth manifold X embedded in $\mathbb{R}^d$, let $\rho(X)$ be its reach. Let $P \subset \mathbb{R}^d$ be finite sample such that $\mathsf{d}_H(P, X) \le \varepsilon$. For any $3\varepsilon \le \alpha \le \frac{3}{16}\sqrt{\frac{3}{5}}\rho(X)$, let $i_* : \mathsf{H}_*(\mathbb{VR}^\alpha) \to \mathsf{H}_*(\mathbb{VR}^{2\alpha})$ be the homomorphism induced by the inclusion $i : \mathbb{VR}^\alpha \to \mathbb{VR}^{2\alpha}$. We have that*

$$\mathrm{rank}\,(\mathrm{im}\,(i_*)) = \dim(\mathsf{H}_*(\mathbb{C}^\alpha(P))) = \dim(\mathsf{H}_*(X)).$$

Proof. By Eqn (9.15) and Proposition 7, we have that for $3\varepsilon \le \alpha \le \beta \le \frac{3}{4}\sqrt{\frac{3}{5}}\rho(X)$,

$$\mathsf{H}_*(X) \cong \mathsf{H}_*(P^\alpha) \cong \mathsf{H}_*(\mathbb{C}^\alpha(P)) \cong \mathsf{H}_*(\mathbb{C}^\beta(P)), \tag{9.16}$$

[2] The result of [23] assumes that $P \subseteq X$, but it then shows that P^α deformation retracts to X. In our statement P is not necessarily from X, and the isomorphism follows easily from the result of [23] and Fact 1.

where the last isomorphism is induced by inclusion. On the other hand, recall the interleaving relation between the Čech and the Rips complexes:

$$\cdots \mathbb{C}^{\alpha}(P) \subseteq \mathbb{VR}^{\alpha}(P) \subseteq \mathbb{C}^{2\alpha}(P) \subseteq \mathbb{VR}^{2\alpha}(P) \subseteq \mathbb{C}^{4\alpha}(P) \cdots .$$

We thus have the following sequence of homomorphisms induced by inclusion:

$$\mathsf{H}_*(\mathbb{C}^{\alpha}(P)) \to \mathsf{H}_*(\mathbb{VR}^{\alpha}(P)) \to \mathsf{H}_*(\mathbb{C}^{2\alpha}(P)) \to \mathsf{H}_*(\mathbb{VR}^{2\alpha}(P)) \to \mathsf{H}_*(\mathbb{C}^{4\alpha}(P)).$$

We have $\mathsf{H}_*(\mathbb{C}^{\alpha}(P)) \cong \mathsf{H}_*(\mathbb{C}^{2\alpha}(P)) \cong \mathsf{H}_*(\mathbb{C}^{4\alpha}(P))$ by Eqn (9.16). It follows that

$$(\mathsf{H}_*(\mathbb{C}^{\alpha}(P)) \to \mathsf{H}_*(\mathbb{C}^{4\alpha}(P))) = \dim(\mathsf{H}_*(\mathbb{C}^{\alpha}(P))).$$

The theorem then follows from the second part of Fact 1. □

9.3.3 Data on a compact set

We now consider the case when we are given a finite set of points P sampling a compact subset $X \subset \mathbb{R}^d$. It is known that an offset X^{α} for any $\alpha > 0$ may not be homotopy equivalent to X for every compact set X. In fact, there exist compact sets so that $\mathsf{H}_*(X^{\lambda})$ is not isomorphic to $\mathsf{H}_*(X)$ no matter how small $\lambda > 0$ is (see Figure 4 of [13]). So, in this case we aim to recover the homology groups of an offset X^{λ} of X for a sufficiently small $\lambda > 0$.

The high level framework is in Eqn (9.17). Here we have $0 < \lambda < \mathrm{wfs}(X)$, while $\mathbb{C}^{\alpha}$ and $\mathbb{VR}^{\alpha}$ stand for the Čech and Rips complexes $\mathbb{C}^{\alpha}(P)$ and $\mathbb{VR}^{\alpha}(P)$ over the point set P. For any $0 < \lambda < \mathrm{wfs}(X)$:

$$\mathsf{H}_*(X^{\lambda}) \xleftrightarrow{\text{Prop. 9}} \text{image}(\mathsf{H}_*(\mathbb{C}^{\alpha} \to \mathsf{H}_*(\mathbb{C}^{2\alpha})) \xleftrightarrow[\textit{Fact 1}]{\text{Eqn (9.20)}} \text{image}(\mathsf{H}_*(\mathbb{VR}^{\alpha}) \to \mathsf{H}_*(\mathbb{VR}^{4\alpha})). \tag{9.17}$$

It is similar to Eqn (9.14) for the manifold case. However, we no longer have the isomorphism between $\mathsf{H}_*(P^{\alpha})$ and $\mathsf{H}_*(X)$. To overcome this difficulty, we leverage Proposition 6. This in turn requires us to consider a pair of Čech complexes to infer homology of X^{λ}, instead of a single Čech complex as in the case of manifolds.

More specifically, suppose that the point set P satisfies that $\mathsf{d}_H(P, X) \leq \varepsilon$; then we have the following nested sequence for $\alpha > \varepsilon$ and $\alpha' \geq \alpha + 2\varepsilon$:

$$X^{\alpha-\varepsilon} \subseteq P^{\alpha} \subseteq X^{\alpha+\varepsilon} \subseteq P^{\alpha'} \subseteq X^{\alpha'+\varepsilon}. \tag{9.18}$$

By Proposition 6, we know that if it also holds that $\alpha' + \varepsilon < \mathrm{wfs}(X)$, then the inclusions between $X^{\alpha-\varepsilon} \subseteq X^{\alpha+\varepsilon} \subseteq X^{\alpha'+\varepsilon}$ induce isomorphisms between their homology groups, which are also isomorphic to $\mathsf{H}_*(X^{\lambda})$ for $\lambda \in (0, \mathrm{wfs}(X))$. It then follows from the second part of Fact 1 that, for $\alpha, \alpha' \in (\varepsilon, \mathrm{wfs}(X) - \varepsilon)$ and $\alpha' - \alpha \geq 2\varepsilon$, we have

$$\mathsf{H}_*(X^{\lambda}) \cong \mathrm{im}\,(i_*), \text{ where } i_* : \mathsf{H}_*(P^{\alpha}) \to \mathsf{H}_*(P^{\alpha'}) \text{ is induced by inclusion } i : P^{\alpha} \subseteq P^{\alpha'}. \tag{9.19}$$

Combining the above with the commutative diagram in Eqn (9.15), we obtain the following result on inferring homology of X^{λ} using a pair of Čech complexes.

Proposition 9. *Let X be a compact set in $\mathbb{R}^d$ and $P \subset \mathbb{R}^d$ a finite set of points with $\mathsf{d}_H(X, P) < \varepsilon$ for some $\varepsilon < \frac{1}{4}\mathrm{wfs}(X)$. Then, for all $\alpha, \alpha' \in (\varepsilon, \mathrm{wfs}(X) - \varepsilon)$ such that $\alpha' - \alpha \geq 2\varepsilon$, and any $\lambda \in (0, \mathrm{wfs}(X))$, we have $\mathsf{H}_*(X^\lambda) \cong \mathrm{im}\,(i_*)$, where $i_* : \mathsf{H}_*(\mathbb{C}^\alpha(P)) \to \mathsf{H}_*(\mathbb{C}^{\alpha'}(P))$ is the homomorphism between homology groups induced by the inclusion $i : \mathbb{C}^\alpha(P) \hookrightarrow \mathbb{C}^{\alpha'}(P)$.*

Finally, to perform homology inference with the Rips complexes, we again resort to the interleaving relation between Čech and Rips complexes, and apply the first part of Fact 1 to the following sequence

$$\mathsf{H}_*(\mathbb{C}^{\alpha/2}(P)) \to \mathsf{H}_*(\mathbb{VR}^{\alpha/2}(P)) \to \mathsf{H}_*(\mathbb{C}^{\alpha}(P)) \to \mathsf{H}_*(\mathbb{C}^{2\alpha}(P)) \to \mathsf{H}_*(\mathbb{VR}^{2\alpha}(P)) \to \mathsf{H}_*(\mathbb{C}^{4\alpha}(P)) \quad (9.20)$$

If $2\varepsilon \leq \alpha \leq \frac{1}{4}(\mathrm{wfs} - \varepsilon)$, both $\mathsf{H}_*(\mathbb{C}^{\alpha/2}(P)) \to \mathsf{H}_*(\mathbb{C}^{4\alpha}(P))$ and $\mathsf{H}_*(\mathbb{C}^{\alpha}(P)) \to \mathsf{H}_*(\mathbb{C}^{2\alpha}(P))$ have ranks equal to $\dim(\mathsf{H}_*(X^\lambda))$ by Proposition 9. Applying Fact 1, we then obtain the following result.

Theorem 10. *Let X be a compact set in $\mathbb{R}^d$ and P a finite point set with $\mathsf{d}_H(X, P) < \varepsilon$ for some $\varepsilon < \frac{1}{9}\mathrm{wfs}(X)$. Then, for all $\alpha \in (2\varepsilon, \frac{1}{4}(\mathrm{wfs}(X) - \varepsilon))$ and all $\lambda \in (0, \mathrm{wfs}(X))$, we have $\mathsf{H}_*(X_\lambda) \cong \mathrm{im}\,(j_*)$, where j_* is the homomorphism between homology groups induced by the inclusion $j : \mathbb{VR}^{\alpha/2}(P) \hookrightarrow \mathbb{VR}^{2\alpha}(P)$.*

9.4 Homology inference for scalar fields

Suppose we are only given a finite sample $P \subset X$ from a smooth manifold $X \subset \mathbb{R}^d$ together with a potentially noisy version $\hat{f}$ of a smooth function $f : X \to \mathbb{R}$ presented as a vertex function $\hat{f} : P \to \mathbb{R}$. We are interested in recovering the persistent homology of the sub-level filtration of f from $\hat{f}$. That is, the goal is to approximate the persistent homology induced by f from the discrete sample P and function values $\hat{f}$ on it.

9.4.1 Problem setup

Set $F_\alpha = f^{-1}(-\infty, \alpha] = \{x \in X \mid f(x) \leq \alpha\}$ as the sublevel set of f w.r.t. α. The sublevel set filtration of X induced by f, denoted by $\mathcal{F}_f = \{F_\alpha;\ i^{\alpha,\beta}\}_{\alpha\leq\beta}$, is a family of sets F_α totally ordered by inclusion map $i^{\alpha,\beta} : F_\alpha \hookrightarrow F_\beta$ for any $\alpha \leq \beta$. This filtration induces the following persistence module:

$$\mathsf{H}_p\mathcal{F} \;=\; \{\mathsf{H}_p(F_\alpha) \xrightarrow{i_*^{\alpha,\beta}} \mathsf{H}_p(F_\beta)\,\}_{\alpha\leq\beta}, \;\text{ where } i_*^{\alpha,\beta} \text{ is induced by inclusion map } i^{\alpha,\beta}. \quad (9.21)$$

For simplicity, we often write the filtration and the corresponding persistence module as $\mathcal{F}_f = \{F_\alpha\}_{\alpha\in\mathbb{R}}$ and $\mathsf{H}_p\mathcal{F} = \{\mathsf{H}(F_\alpha)\}_{\alpha\in\mathbb{R}}$, when the choices of maps connecting their elements are clear.

Our goal is to approximate the persistence diagram $\mathrm{Dgm}_p(\mathcal{F})$ from point samples P and $\hat{f} : P \to \mathbb{R}$. Intuitively, we construct a specific Čech (or Rips) complex $\mathbb{C}^r(P)$, use $\hat{f}$ to induce a filtration of $\mathbb{C}^r(P)$, and then use its persistent homology to approximate $\mathrm{Dgm}_p(\mathcal{F})$. More specifically, we need to consider *nested pair filtration* for either $\mathbb{C}^r(P)$ or $\mathbb{VR}^r(P)$.

Nested pair filtration. Let $P_\alpha = \{p \in P \mid \hat{f}(p) \le \alpha\}$ be the set of sample points with the function value for $\hat{f}$ at most α, which presumably samples the sublevel set F_α of X w.r.t. f. To estimate the topology of F_α from these discrete sample P_α, we consider either the Čech complex $\mathbb{C}^r(P_\alpha)$ or the Rips complex $\mathbb{VR}^r(P_\alpha)$. For the time being, consider $\mathbb{VR}^r(P_\alpha)$. As we already saw in previous sections, the topological information of F_α can be inferred from a pair of nested complexes $\mathbb{VR}^r(P_\alpha) \overset{j_\alpha}{\hookrightarrow} \mathbb{VR}^{r'}(P_\alpha)$ for some appropriate $r < r'$. Fixing r and r', for any $\alpha \le \beta$, consider the following commutative diagram induced by inclusions:

$$\begin{tikzcd} \mathsf{H}_*(\mathbb{VR}^r(P_\alpha)) \arrow[r] \arrow[d, "i_{\alpha*}"] & \mathsf{H}_*(\mathbb{VR}^r(P_\beta)) \arrow[d, "i_{\beta*}"] \\ \mathsf{H}_*(\mathbb{VR}^{r'}(P_\alpha)) \arrow[r, "f_{\alpha*}^\beta"] & \mathsf{H}_*(\mathbb{VR}^{r'}(P_\beta)) \end{tikzcd} \tag{9.22}$$

Set $\phi_\alpha^\beta : \mathrm{im}\,(i_{\alpha*}) \to \mathrm{im}\,(i_{\beta*})$ to be $\phi_\alpha^\beta = f_{\alpha*}^\beta|_{\mathrm{im}\,(i_{\alpha*})}$, that is, the restriction of $f_{\alpha*}^\beta$ to $\mathrm{im}\,(i_{\alpha*})$. This map is well-defined as the diagram above commutes. This gives rise to a persistence module $\{\mathrm{im}\,(i_{\alpha*}); \phi_\alpha^\beta\}_{\alpha\le\beta}$, that is, a family of totally ordered vector spaces $\{\mathrm{im}\,(i_\alpha)\}$ with commutative homomorphisms ϕ_α^β between any two elements. We formalize and generalize the above construction below.

Definition 11 (Nested pair filtration)**.** A *nested pair filtration* is a sequence of pairs of complexes $\{AB_\alpha = (A_\alpha, B_\alpha)\}_{\alpha\in\mathbb{R}}$ where (i) $A_\alpha \overset{i_\alpha}{\hookrightarrow} B_\alpha$ is inclusion for every α and (ii) $AB_\alpha \hookrightarrow AB_\beta$ for $\alpha \le \beta$ is given by $A_\alpha \hookrightarrow A_\beta$ and $B_\alpha \overset{f_\alpha^\beta}{\hookrightarrow} B_\beta$. The p-th persistence module of the filtration $\{AB_\alpha\}_{\alpha\in\mathbb{R}}$ is given by the homology module $\{\mathrm{im}\,(\mathsf{H}_p(A_\alpha) \to \mathsf{H}_p(B_\alpha)); \phi_\alpha^\beta\}_{\alpha\le\beta}$ where ϕ_α^β is the restriction of $f_{\alpha*}^\beta$ on the im $i_{\alpha*}$. For simplicity, we say the module is induced by the nested pair filtration $\{A_\alpha \hookrightarrow B_\alpha\}$.

The high level approach of inferring persistent homology of a scalar field $f : X \to \mathbb{R}$ from a set of points equipped with $\hat{f} : P \to \mathbb{R}$ involves the following steps:

Step 1. Sort all points of P in non-decreasing $\hat{f}$-values, $P = \{p_1, \ldots, p_n\}$. Set $a_i = \hat{f}(p_i)$ for $i \in [1, n]$.

Step 2. Compute the persistence diagram induced by the filtration of nested pairs $\{\mathbb{VR}^r(P_{\alpha_i}) \hookrightarrow \mathbb{VR}^{r'}(P_{\alpha_i})\}_{i\in[1,n]}$ (or $\{\mathbb{C}^r(P_{\alpha_i}) \hookrightarrow \mathbb{C}^{r'}(P_{\alpha_i})\}_{i\in[1,n]}$) for appropriate parameters $0 < r < r'$.

To obtain an approximation guarantee for the above approach, we consider an intermediate object defined by the intrinsic Riemannian metric on the manifold X. Indeed, note that the filtration of X w.r.t. f is intrinsic in the sense that it is independent of how X is embedded in $\mathbb{R}^d$. Hence it is more natural to approximate its persistent homology with another object defined intrinsically for X.

Given a compact Riemannian manifold $X \subset \mathbb{R}^d$ embedded in $\mathbb{R}^d$, let d_X be the Riemannian metric of X inherited from the Euclidean metric d_E of $\mathbb{R}^d$. Let $B_X(x, r) := \{y \in X \mid \mathsf{d}_X(x, y) \le r\}$ be the *geodesic ball on X* centered at x and with radius r. In contrast, $B_E(x, r)$ (or simply $B(x, r)$) denotes the Euclidean ball in $\mathbb{R}^d$. A ball $B_X(x, r)$ is *strongly convex* if for every pair $y, y' \in B_X(x, r)$, there exists a unique minimizing geodesic between y and y' whose interior is contained within $B_X(x, r)$. For details on these concepts, see [5, 18].

Definition 12 (Strong convexity)**.** For $x \in X$, let $\rho_c(x; X)$ denote the supreme of radius r such that the geodesic ball $B_X(x, r)$ is strongly convex. The *strong convexity radius of* (X, d_X) is defined as $\rho_c(X) := \inf_{x \in X} \rho_c(x; X)$.

Let $\mathsf{d}_X(x, P) := \inf_{p \in P} \mathsf{d}_X(x, p)$ denote the closest geodesic distance between x and the set $P \subseteq X$.

Definition 13 (ε-geodesic sample)**.** A point set $P \subset X$ is an *ε-geodesic sample of* (X, d_X) if for all $x \in X$, $\mathsf{d}_X(x, P) \leq \varepsilon$.

Recall that P_α is the set of points in P with $\hat{f}$-value at most α. The union of geodesic balls $P_\alpha^{\delta;X} = \bigcup_{p \in P_\alpha} B_X(p, \delta)$ is intuitively the "δ-thickening" of P_α within the manifold X. We use two kinds of Čech and Rips complexes. One is defined with the metric d_E of the ambient Euclidean space which we call *(extrinsic) Čech complex* $\mathbb{C}^\delta(P_\alpha)$ and *(extrinsic) Rips complex* $\mathbb{VR}^\delta(P_\alpha)$. The other is *intrinsic Čech complex* $\mathbb{C}_X^\delta(P_\alpha)$ and *intrinsic Rips complex* $\mathbb{VR}_X^\delta(P_\alpha)$ that are defined with the intrinsic metric d_X. Note that $\mathbb{C}_X^\delta(P_\alpha)$ is the nerve complex of the union of geodesic balls forming $P_\alpha^{\delta;X}$. Also the interleaving relation between the Čech and Rips complexes remains the same as for general metric spaces; that is, $\mathbb{C}_X^\delta(P_\alpha) \subseteq \mathbb{VR}_X^\delta(P_\alpha) \subseteq \mathbb{C}_X^{2\delta}(P_\alpha)$ for any α and δ.

9.4.2 Inference guarantees

Recall that two ε-interleaved filtrations lead to ε-interleaved persistence modules, which further mean that the bottleneck distance between their persistence diagrams are bounded by ε. Here we first relate the space filtration with the intrinsic Čech filtrations and then relate these intrinsic ones with the extrinsic Čech or Rips filtrations of nested pairs as illustrated in Eqn 9.23 below.

$$\{F_\alpha\} \longleftrightarrow \{P_\alpha^{r;X}\} \longleftrightarrow \{\mathbb{C}_X^r(P_\alpha)\} \longleftrightarrow \{\mathbb{C}^r(P_\alpha) \hookrightarrow \mathbb{C}^{r'}(P_\alpha)\} \text{ or } \{\mathbb{VR}^r(P_\alpha) \hookrightarrow \mathbb{VR}^{r'}(P_\alpha)\} \quad (9.23)$$

Proposition 11. *Let $X \subset \mathbb{R}^d$ be a compact Riemmanian manifold with intrinsic metric d_X, and let $f : X \to \mathbb{R}$ be a C-Lipschitz function. Suppose $P \subset X$ is an ε-geodesic sample of X, equipped with $\hat{f} : P \to \mathbb{R}$ so that $\hat{f} = f|_P$. Then, for any fixed $\delta \geq \varepsilon$, the filtration $\{F_\alpha\}_\alpha$ and the filtration $\{P_\alpha^{\delta;X}\}_\alpha$ are ($C\delta$)-interleaved w.r.t. inclusions.*

The intrinsic Čech complex $\mathbb{C}_X^\delta(P_\alpha)$ is the nerve complex for $\{B_X(p, \delta)\}_{p \in P_\alpha}$. Furthermore, for $\delta < \rho_c(X)$, the family of geodesic balls in $\{B_X(p, \delta)\}_{p \in P_\alpha}$ form a cover of the union $P_\alpha^{\delta;X}$ that satisfies the condition of the Nerve Theorem. Hence, there is a homotopy equivalence between the nerve complex $\mathbb{C}_X^\delta(P_\alpha)$ and $P_\alpha^{\delta;X}$. Furthermore, using the same argument for showing that diagram in Eqn (9.15) commutes (Lemma 3.4 of [9]), one can show that the following diagram commutes for any $\alpha \leq \beta \in \mathbb{R}$ and $\delta \leq \xi < \rho_c(X)$:

$$\begin{tikzcd}
\mathsf{H}_*(P_\alpha^{\delta;X}) \arrow[r, "i_*"] \arrow[d, "h_*"] & \mathsf{H}_*(P_\beta^{\xi;X}) \arrow[d, "h_*"] \\
\mathsf{H}_*(\mathbb{C}_X^\delta(P_\alpha)) \arrow[r, "i_*"] & \mathsf{H}_*(\mathbb{C}_X^\xi(P_\beta))
\end{tikzcd} \quad (9.24)$$

Here the horizontal homomorphisms are induced by inclusions, and the vertical ones are isomorphisms induced by the homotopy equivalence between a union of geodesic balls and its nerve complex. Setting $\delta = \xi$ in the above diagram, this leads to the following result (see Lemma 2 of [12] for details):

Corollary 12. *Let X, f, and P be as in Proposition 11 (although f does not need to be C-Lipschitz). For any $\delta < \rho_c(X)$, $\{P_\alpha^{\delta;X}\}_{\alpha\in\mathbb{R}}$ and $\{\mathbb{C}_X^\delta(P_\alpha)\}_{\alpha\in\mathbb{R}}$ are 0-interleaved. Hence they induce isomorphic persistence modules which have identical persistence diagrams.*

Combining with Proposition 11, this implies that the filtration $\{\mathbb{C}_X^r(P_\alpha)\}_\alpha$ and the filtration $\{F_\alpha\}_\alpha$ are Cr-interleaved for $\varepsilon \le r < \rho_c(X)$.

However, we cannot access the intrinsic metric d_X of the manifold X. It turns out that for points that are sufficiently close, their Euclidean distance forms a constant factor approximation of the geodesic distance between them on X.

Proposition 13. *Let $X \subset \mathbb{R}^d$ be an embedded Riemannian manifold with reach ρ_X. For any two points $x, y \in X$ with $\mathsf{d}_E(x,y) \le \rho_X/2$, we have that:*

$$\mathsf{d}_E(x,y) \le \mathsf{d}_X(x,y) \le \left(1 + \frac{4d_E^2(x,y)}{3\rho_X^2}\right)\mathsf{d}_E(x,y) \le \frac{4}{3}\mathsf{d}_E(x,y).$$

This implies the following nested relation between the extrinsic and intrinsic Čech complexes:

$$\mathbb{C}_X^{\frac{3}{4}\delta}(P_\alpha) \subseteq \mathbb{C}^\delta(P_\alpha) \subseteq \mathbb{C}_X^\delta(P_\alpha) \subseteq \mathbb{C}^{\frac{4}{3}\delta}(P_\alpha) \subseteq \mathbb{C}_X^{\frac{4}{3}\delta}(P_\alpha); \;\; \text{for any } \delta < \frac{3}{8}\rho_X. \tag{9.25}$$

Note that a similar relation also holds between the intrinsic Čech filtration and the extrinsic Rips complexes due to the nested relation between extrinsic Čech and Rips complexes. To infer persistent homology from nested pairs filtrations for complexes constructed under the Euclidean metric, we use the following key lemma from [12], which can be thought of as a persistent version as well as a generalization of Fact 1.

Proposition 14. *Let X, f, and P be as in Proposition 11. Suppose that there exist $\varepsilon' \le \varepsilon'' \in [\varepsilon, \rho_c(X))$ and two filtrations $\{G_\alpha\}_\alpha$ and $\{G'_\alpha\}_\alpha$, so that*

$$\text{for all } \alpha \in \mathbb{R}, \quad \mathbb{C}_X^\varepsilon(P_\alpha) \subseteq G_\alpha \subseteq \mathbb{C}_X^{\varepsilon'}(P_\alpha) \subseteq G'_\alpha \subseteq \mathbb{C}_X^{\varepsilon''}(P_\alpha).$$

Then the persistence module induced by the filtraiton $\{F_\alpha\}_\alpha$ for f and that induced by the nested pairs of filtrations $\{G_\alpha \hookrightarrow G'_\alpha\}_\alpha$ are $C\varepsilon''$-interleaved, where f is C-Lipschitz.

Combining this proposition with the sequences in Eqn (9.25), we obtain the following results on inferring the persistent homology induced by a function $f : X \to \mathbb{R}$.

Theorem 15. *Let $X \subset \mathbb{R}^d$ be a compact Riemannian manifold with intrinsic metric d_X, and $f : X \to \mathbb{R}$ a C-Lipschitz function on X. Let ρ_X and $\rho_c(X)$ be the reach and the strong convexity radius of (X, d_X) respectively. Suppose $P \subset X$ is an ε-geodesic sample of X, equipped with $\hat{f} : P \to \mathbb{R}$ such that $\hat{f} = f|_P$. Then:*

(i) for any fixed r such that $\varepsilon \le r \le \min\{\frac{9}{16}\rho_c(X), \frac{9}{32}\rho_X\}$, the persistent homology module induced by the sublevel-set filtration of $f : X \to \mathbb{R}$ and that induced by the filtration of nested pairs $\{\mathbb{C}^r(P_\alpha) \hookrightarrow \mathbb{C}^{\frac{4}{3}r}(P_\alpha)\}_\alpha$ are $\frac{16}{9}Cr$-interleaved; and

(ii) for any fixed r such that $2\varepsilon \leq r \leq \min\{\frac{9}{32}\rho_c(X), \frac{9}{64}\rho_X\}$, the persistent homology module induced by the sublevel set filtration of f and that induced by the filtration of nested pairs $\{\mathbb{VR}^r(P_\alpha) \hookrightarrow \mathbb{VR}^{\frac{8}{3}r}(P_\alpha)\}_\alpha$ are $\frac{32}{9}Cr$-interleaved.

In particular, in each case above, the bottleneck distance between their respective persistence diagrams is bounded by the stated interleaving distance between persistence modules.

9.5 Notes and Exercises

Part of Theorem 3 has been proven in [10, 6]. A complete proof as well as a thorough treatment for geometric complexes such as Rips and Čech complexes can be found in [11]. The first work on data sparsification for Rips filtrations is proposed by Sheehy [26]. The presentation of Chapter 9.2.1 combines the treatments of sparsification of [27, 3]; in particular, in [27], a net-tower created via net-tree data structure (e.g, [20]) is used for constructing sparse Rips filtration. Extension of such sparsification to Čech complexes and a geometric interpretation are provided in [4]. The Rips sparsification is extended to handle weighted Rips complexes derived from distance to measures in [3]. Sparsification via towers is introduced in [16]. This is an application of the algorithm we presented before for computing persistent homology for a simplicial tower. Simplicial maps allow batch-collapse of vertices and leads to more aggressive sparsification. However, in practice it is observed that it also has the over-connection issues as one collapses the vertices. This issue is addressed in [17]. In particular, the SimBa algorithm of [17] exploits the simplicial maps for sparsification, but connects vertices at sparser levels based on a certain distance between two sets (each of which intuitively is the set of original points mapped to this vertex at this level). While SimBa has similar approximation guarantees in sparsification, in practice, the sparsified sequence of complexes has much smaller size compared to prior approaches.

Much of the materials in Section 9.3 are taken from [9, 23, 12, 11]. We remark that there have been different variations of the medial axis in the literature. We follow the notation from [15]. We also note that there exists a robust version of the medial axis, called the λ-medial axis, proposed in [8]. The concept of the local feature size was originally proposed in [25] in the context of mesh generation and a different version that we describe in this chapter was introduced in [1] in the context of curve/surface reconstruction. The local feature size has been widely used in the field of surface reconstruction and mesh generation; see the books [15, 14]. Critical points of the distance field were originally studied in [19]. See [8, 13, 22] for further studies as well as the development on weak feature sizes.

In homology inference for manifolds, we note that Niyogi, Smale and Weinberger in [23] provide two deformnation retract results from union of balls over P to a manifold X; Proposition 3.1 holds for the case when $P \subset \mathsf{X}$, while Proposition 7.1 holds when P is within a tubular neighborhood of X. The latter has much stronger requirement on the radius α. In our presentation, our Proposition 7 uses a corollary of Proposition 3.1 of [23] to obtain an isomorphism between the homology groups of union of balls and of X to allow a better range of the parameter α – however, we lose the deformation retraction here; see the footnote above Proposition 7. Results in Chapter 9.4 are mostly based on work from [12].

This chapter focuses presenting the main framework behind homology (or persistent homology) inference from point cloud data. The current theoretical guarantees hold when input points sample the hidden domain well within Hausdorff distance. See [3, 7, 24] for data sparsification or homology inference for points corrupted with more general noise, and [2] for persistent homology inference under more general noise for input scalar fields.

Exercise

1. Prove Part (i) of Theorem 3.
2. Prove the bound on the Rips pseudo-distance $\mathsf{d}_{Rips}(P, Q)$ in Part (ii) of Theorem 3.
3. Given two finite sets of points $P, Q \subset \mathbb{R}^d$, let d_P and d_Q denote the restriction of the Euclidean metric over P and Q respectively. Consider the Hausdorff distance $\delta_H = \mathsf{d}_H(P, Q)$ between P and Q, as well as the Gromov-Hausdorff distance $\delta_{GH} = \mathsf{d}_{GH}((P, \mathsf{d}_P), (Q, \mathsf{d}_Q))$.
   (i) Prove that $\delta_{GH} \leq 2\delta_H$.
   (ii) Give an example of $P, Q \subset \mathbb{R}^2$ such that δ_H is much larger than δ_{GH}, say $\delta_H \geq 10\delta_{GH}$. [In fact, this can hold for any fixed constant.]
4. Consider the greedy permutation approach introduced in Chapter 9.2, and the assignment of exit-times for points $p \in P$. Construct the open tower $\{N_\gamma\}$ and closed tower $\{\overline{N}_\gamma\}$ as described in the chapter. Prove that both N_γ and $\overline{N}_\gamma$ are γ-nets for P.
5. Suppose we are given $P_0 \supset P_1$ sampled from a metric space (Z, d) where P_1 is an γ-net of P_0. Define $\pi : P_0 \to P_1$ as $\pi(p) \mapsto \mathrm{argmin}_{q \in P_1} \mathsf{d}(p, q)$.
   (a) Prove that the vertex map π induces a simplicial map $\pi : \mathbb{VR}^\alpha(P_0) \to \mathbb{VR}^{\alpha+\gamma}(P_1)$.
   (b) Consider the following diagram. Prove that the map $j \circ \pi$ is contiguous to the inclusion map i.

$$\begin{tikzcd} \mathbb{VR}^\alpha(P_0) \arrow[r, hook, "i"] \arrow[dr, "\pi"'] & \mathbb{VR}^{\alpha+\gamma}(P_0) \\ & \mathbb{VR}^{\alpha+\gamma}(P_1) \arrow[u, hook, "j"'] \end{tikzcd} \tag{9.26}$$

6. Let P be a set of points in $\mathbb{R}^d$. Let d_2 and d_1 denote the distance metric under L_2 norm and under L_1 norm respectively. Let $\mathbb{C}_2(P)$ and $\mathbb{C}_1(P)$ be the Čech filtration over P induced by d_2 and d_1 respectively. Show the relation between the log-scaled version of persistence diagrams $\mathrm{Dgm}_{\log}\mathbb{C}_2(P)$ and $\mathrm{Dgm}_{\log}\mathbb{C}_1(P)$, that is, bound $\mathsf{d}_B(\mathrm{Dgm}_{\log}\mathbb{C}_2(P), \mathrm{Dgm}_{\log}\mathbb{C}_1(P))$.
7. Prove Proposition 11. Using the fact that Diagram 9.24 commutes, prove Corollary 12.

Bibliography

[1] Nina Amenta, Marshall W. Bern, and David Eppstein. The crust and the beta-skeleton: Combinatorial curve reconstruction. *Graphical Models and Image Processing*, 60(2):125–135, 1998.

[2] M. Buchet, F. Chazal, T. K. Dey, F. Fan, S. Oudot, and Y. Wang. Topological analysis of scalar fields with outliers. In *Proc. 31rd Annu. Sympos. Comput. Geom. (SoCG)*, pages 827–841, 2015.

[3] Mickaël Buchet, Frédéric Chazal, Steve Y. Oudot, and Donald Sheehy. Efficient and robust persistent homology for measures. In *Proc. 26th Ann. ACM-SIAM Sympos. Discrete Algorithms*, pages 168–180, 2015.

[4] Nicholas J. Cavanna, Mahmoodreza Jahanseir, and Donald R. Sheehy. A geometric perspective on sparse filtrations. In *Proceedings of the Canadian Conference on Computational Geometry*, 2015.

[5] Isaac Chavel. *Riemannian Geometry: A Modern Introduction, 2nd Ed.* Cambridge university press, 2006.

[6] F. Chazal, D. Cohen-Steiner, L. J. Guibas, F. Mémoli, and S. Y. Oudot. Gromov-hausdorff stable signatures for shapes using persistence. *Computer Graphics Forum*, 28(5):1393–1403, 2009.

[7] F. Chazal, D. Cohen-Steiner, and Q. Mérigot. Geometric inference for probability distributions. *Found. Comput. Math.*, 11(6):733–751, 2011.

[8] F. Chazal and A. Lieutier. Weak feature size and persistent homology: computing homology of solids in $\mathbb{R}^n$ from noisy data samples. In *Proc. 21st Ann. Sympos. Comput. Geom.*, pages 255–262, 2005.

[9] F. Chazal and S.Y. Oudot. Towards persistence-based reconstruction in Euclidean spaces. In *SCG '08: Proc. 24th Ann. Sympos. Comput. Geom.*, pages 232–241, 2008.

[10] Frédéric Chazal, David Cohen-Steiner, Marc Glisse, Leonidas Guibas, and Steve Oudot. Proximity of persistence modules and their diagrams. In *Proceedings of the Twenty-fifth Annual Symposium on Computational Geometry*, SCG '09, pages 237–246, 2009.

[11] Frédéric Chazal, Vin de Silva, and Steve Oudot. Persistence stability for geometric complexes. *Geometriae Dedicata*, 173(1):193–214, Dec 2014.

[12] Frédéric Chazal, Leonidas J. Guibas, Steve Oudot, and Primoz Skraba. Analysis of scalar fields over point cloud data. *Discrete Comput. Geom.*, 46(4):743–775, 2011.

[13] Frédéric Chazal and André Lieutier. Stability and computation of topological invariants of solids in $\mathbb{R}^n$. *Discrete Comput. Geom.*, 37(4):601–617, 2007.

[14] S.-W. Cheng, T. K. Dey, and J. R. Shewchuk. *Delaunay Mesh Generation*. CRC Press, 2012.

[15] T. K. Dey. *Curve and Surface Reconstruction: Algorithms with Mathematical Analysis*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2006.

[16] Tamal K. Dey, Fengtao Fan, and Yusu Wang. Computing topological persistence for simplicial maps. In *Proceedings of the Thirtieth Annual Symposium on Computational Geometry*, SOCG'14, pages 345:345–345:354, New York, NY, USA, 2014. ACM.

[17] Tamal K. Dey, Dayu Shi, and Yusu Wang. SimBa: An Efficient Tool for Approximating Rips-Filtration Persistence via Simplicial Batch-Collapse. In *24th Annual European Symposium on Algorithms (ESA 2016)*, volume 57 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 35:1–35:16, 2016.

[18] S. Gallot, D. Hulin, and J. Lafontaine. *Riemannian Geometry*. Springer-Verlag, 2nd edition, 1993.

[19] Karsten Grove. Critical point theory for distance functions. *Proc. Symp. Pure. Math.*, 54(3):357–385, 1993.

[20] Sariel Har-Peled and Manor Mendel. Fast construction of nets in low-dimensional metrics and their applications. *SIAM Journal on Computing*, 35(5):1148–1184, 2006.

[21] Yasuaki Hiraoka, Takenobu Nakamura, Akihiko Hirata, Emerson G. Escolar, Kaname Matsue, and Yasumasa Nishiura. Hierarchical structures of amorphous solids characterized by persistent homology. *Proceedings of the National Academy of Sciences*, 113(26):7035–7040, 2016.

[22] André Lieutier. Any open bounded subset of R^{n} has the same homotopy type as its medial axis. *Computer-Aided Design*, 36(11):1029–1046, 2004.

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

[24] Partha Niyogi, Stephen Smale, and Shmuel Weinberger. A topological view of unsupervised learning from noisy data. *SIAM J. Comput.*, 40(3):646–663, 2011.

[25] J. Ruppert. A Delaunay refinement algorithm for quality 2-dimensional mesh generation. *J. Algorithms*, 18:548–585, 1995.

[26] D. Sheehy. Linear-size approximations to the Vietoris-Rips filtration. In *Proc. 28th. Annu. Sympos. Comput. Geom.*, pages 239–248, 2012.

[27] D. R. Sheehy. Linear-size approximations to the Vietoris-Rips filtration. *Discrete Comput. Geom.*, 49:778–796, 2013.