Reeb Graphs: Approximation and Persistence

Tamal K. Dey[*]   Yusu Wang[*]

Abstract

Given a continuous function $f : \mathsf{X} \to \mathbb{R}$ on a topological space X, its *level set* $f^{-1}(a)$ changes continuously as the real value a changes. Consequently, the connected components in the level sets appear, disappear, split and merge. The Reeb graph of f summarizes this information into a graph structure. Previous work on Reeb graph mainly focused on its efficient computation. In this paper, we initiate the study of two important aspects of the Reeb graph which can facilitate its broader applications in shape and data analysis.

The first one is the approximation of the Reeb graph of a function on a smooth compact manifold M without boundary. The approximation is computed from a set of points P sampled from M. By leveraging a relation between the Reeb graph and the so-called *vertical homology group*, as well as between cycles in M and in a Rips complex constructed from P, we compute the H_1-homology of the Reeb graph from P. It takes $O(n \log n)$ expected time, where n is the size of the 2-skeleton of the Rips complex. As a by-product, when M is an orientable 2-manifold, we also obtain an efficient near-linear time (expected) algorithm for computing the rank of $\mathsf{H}_1(\mathsf{M})$ from point data. The best known previous algorithm for this problem takes $O(n^3)$ time for point data.

The second aspect concerns the definition and computation of the *persistent Reeb graph homology* for a sequence of Reeb graphs defined on a filtered space. For a piecewise-linear function defined on a filtration of a simplicial complex K, our algorithm computes all persistent H_1-homology for the Reeb graphs in $O(nn_e^3)$ time, where n is the size of the 2-skeleton and n_e is the number of edges in K.

1 Introduction

Given a topological space X and a continuous scalar function $f : \mathsf{X} \to \mathbb{R}$, the set $\{x \in \mathsf{X} : f(x) = a\}$ is a *level set* of f for some value $a \in \mathbb{R}$. The level sets of f may have multiple connected components. The Reeb graph of f is obtained by continuously collapsing each connected component in the level set into a single point. Intuitively, as a changes continuously, the connected components in the level sets appear, disappear, split and merge; and the Reeb graph of f tracks such changes. Hence, the Reeb graph provides a simple yet meaningful abstraction of the input scalar field. It has been used in a range of applications in computer graphics and visualization; see, for example, the survey [3] and references therein on applications of Reeb graph.

Our results. Most of the previous work on the Reeb graph focused on its efficient computation. In this paper, we initiate the study of two questions related to Reeb graphs both of which are important in shape and data analysis applications.

The first question is concerned with the approximation of the Reeb graph from a set of points sampled from a hidden manifold. It turns out that the Reeb graph homology is also related to the so-called vertical homology groups. These relations enable us to develop an efficient algorithm to approximate the Reeb graph of the manifold from its point samples.

---

[*]Department of Computer Science and Engineering, The Ohio State University, Columbus, OH 43221.
Emails: { `tamaldey`, `yusu`}@cse.ohio-state.edu.

As a by-product of our approximation result, we also obtain a near-linear time algorithm that computes the first betti number $\beta_1(\mathsf{M})$ of an orientable smooth compact 2-manifold M without boundary from its point samples. This result may be of independent interest even though the correctness of our algorithm needs a slightly stronger condition than the previous best-known approach for computing $\beta_1(\mathsf{M})$ from point data. In particular, it is shown in [1] that $\beta_1(\mathsf{M})$ can be computed as the first betti number of certain Rips complex constructed out of the input data[1]. A straightforward computation of betti numbers of the Rips complex using Smith normal form [23] takes cubic time whereas our algorithm runs in near-linear expected time.

The second question we study concerns with the definition and computation of loops in Reeb graphs which remain "persistent" as its defining domain "grows". We propose a definition of the *persistent Reeb graph homology* for a sequence of Reeb graphs. They are computed for a function defined on a filtered space in the same spirit as the standard persistent homology [19]. Interestingly, this problem does not seem to be easier than computing the standard persistent homology, potentially due to the fact that the domains in question (the sequence of Reeb graphs) do not have an inclusion between them, as was the case for standard persistence homology.

Related work. As mentioned already, most previous work on the Reeb graph focused on its efficient computation. Shinagawa and Kunii [26] presented the first provably correct algorithm to compute Reeb Graphs for a triangulation of a 2-manifold in $\Theta(m^2)$ time where m is the number of vertices in the triangulation. Cole-McLaughlin et al. [11] improved the running time to $O(m \log m)$. Tierny et al. [27] proposed an algorithm that computes the Reeb graph for a 3-manifold with boundary embedded in $\mathbb{R}^3$ in time $O(n \log n + hn)$, where h is number of independent loops in the Reeb graph. For a piecewise-linear function defined on an arbitrary simplicial complex, a simple algorithm is proposed in [15] that runs in time $O(n \log n + L)$, where $L = \Theta(nm)$ is the total complexity of all level-sets passing through critical points. Doraiswamy and Natarajan [16] extended the sweeping idea to compute the Reeb graph in $O(n \log n (\log \log n)^3)$ time from an arbitrary simplicial complex, where n is the size of the 2-skeleton of this simplicial complex. A streaming algorithm was presented in [25] to compute the Reeb graph for an arbitrary simplicial complex in an incremental manner in $\Theta(nm)$ time. Recently, Harvey et al. [20] presented an efficient randomized algorithm to compute the Reeb graph for an arbitrary simplicial complex in $O(n \log m)$ expected running time. The Reeb graph for a time-varying function defined on a 3-dimensional space was studied in [18].

Recently a flurry of research has been initiated on estimating topological information from point data, such as computing ranks of homology groups [8], cut locus [13], and the shortest set of homology loops [14]. In [6], Chazal et al. initiated the study of approximating topological attributes of scalar functions from point data, and showed that the standard persistent diagram induced by a function can be approximated from input points. This result was later used in [7] to produce a clustering algorithm with theoretical guarantees. The results from [6, 7] can be used to approximate *loop-free* Reeb graphs (also called *contour trees*) from point data, thus providing a partial solution to our first question. However, it is unclear how to approximate loops in the Reeb graph which correspond to *a subset of essential loops* in the input domain which represent a subgroup of H_1-homology.

2 Background and notations

Homology. A homology group of a topological space X encodes its topological connectivity. We consider the *simplicial* homology group if X is a simplicial complex, and consider the *singular* homology group otherwise, both denoted with $\mathsf{H}_p(\mathsf{X})$ for the pth homology group. The definitions of these two homology

---

[1]In fact, in [1] Attali et al. show a much stronger result: the Rips complex of a point clouds captures the homotopy type of a compact topological space in Euclidean space under some conditions. An earlier result of Hausmann [22] also shows that Rips complex captures the topology of an input Riemannian manifold, but under much stronger conditions.

groups can be obtained from any standard book on algebraic topology. Here we single out the concepts of p-chains and p-cycles in singular homology whose definitions are not as widely known in computational geometry as their simplicial counterparts. See [21, 23] for detailed discussions on this topic.

A singular p-simplex for a topological space X is a continuous map σ from the standard p-simplex $\Delta^p \subset \mathbb{R}^p$ to X. For example, a 1-simplex σ is a continuous map $\sigma : [0,1] \to \mathsf{X}$. A p-chain is a formal sum of singular p-simplices. A singular p-cycle in X is a p-chain whose boundary is a zero $(p-1)$-chain. Therefore, technically speaking, a p-chain or a p-cycle for X is a formal sum of maps. In this paper we will only deal with 1-chains and 2-chains. Let a *loop* refer to a continuous map $\mathbb{S}^1 \to \mathsf{X}$ or a finite union of such maps. For any 1-cycle $\alpha = \sigma_1 + \cdots + \sigma_k$, there is a corresponding loop ϕ whose image in X coincides with the disjoint union of images $\sigma_i([0,1])$, for $i \in [1,k]$ (see page 108–109 in [21]). We call this loop the *carrier* of α, and that α is *carried* by loop ϕ. All singular 1-cycles carried by the same loop are homologous. Hence, in the remainder of the paper, we sometimes abuse the notations slightly and talk about a loop as if it is a 1-cycle. For example, we will say that two loops are homologous which means that cycles carried by these two loops are homologous.

We assume that X is compact and triangulable. Its simplicial homology defined by a triangulation identifies to its singular homology. We also assume that the homology groups are defined over $\mathbb{Z}_2$ coefficients. Since $\mathbb{Z}_2$ is a field, $\mathsf{H}_p(\mathsf{X})$ is a vector space of dimension p. It will be clear from the context whether we are talking about simplicial or singular homology of X. Unless specified, we assume singular homology for X. Let $\mathbf{Z}_p(\mathsf{X})$ denote the p-th cycle group in X. A continuous map $\Phi : \mathsf{X}_1 \to \mathsf{X}_2$ between two topological spaces induces a map among its chain groups which we denote as $\Phi_\#$. Clearly, $\Phi_\#$ provides a map from the cycle group $\mathbf{Z}_p(\mathsf{X}_1)$ to the cycle group $\mathbf{Z}_p(\mathsf{X}_2)$ which in turn induces a homomorphism $\Phi_* : \mathsf{H}_p(\mathsf{X}_1) \to \mathsf{H}_p(\mathsf{X}_2)$.

Horizontal and Vertical Homology Following [9], we now extend the standard homology to the so-called *horizontal* and *vertical homology with respect to a function* $f : \mathsf{X} \to \mathbb{R}$. Given a continuous function f, its *level sets* and *interval sets* are defined as: $\mathsf{X}_a := f^{-1}(a)$ and $\mathsf{X}_I := f^{-1}(I)$ for $a \in \mathbb{R}$ and for an open or closed interval $I \subseteq \mathbb{R}$, respectively. From now on we sometimes omit the use of f when its choice is clear from the context.

A homology class $\omega \in \mathsf{H}_p(\mathsf{X})$ is *horizontal* if there exists a discrete set of iso-values $\{a_i\}$ such that ω has a pre-image under the map $\mathsf{H}_p(\bigcup_i \mathsf{X}_{a_i}) \to \mathsf{H}_p(\mathsf{X})$ induced by inclusion. The set of horizontal homology classes form a subgroup $\overline{\mathsf{H}}_p(\mathsf{X})$ of $\mathsf{H}_p(\mathsf{X})$ since the trivial homology class is horizontal, and the addition of any two horizontal homology class is still horizontal. We call this subgroup $\overline{\mathsf{H}}_p(\mathsf{X})$ the *horizontal homology group of* X *with respect to* f. The *vertical homology group of* X *with respect to* f is defined as:

$$\breve{\mathsf{H}}_p(\mathsf{X}) := \mathsf{H}_p(\mathsf{X})/\overline{\mathsf{H}}_p(\mathsf{X}), \text{ the quotient of } \mathsf{H}_p(\mathsf{X}) \text{ with } \overline{\mathsf{H}}_p(\mathsf{X}).$$

The coset $\omega + \overline{\mathsf{H}}_p(\mathsf{X})$ for every class $\omega \in \mathsf{H}_p(\mathsf{X})$ provides an equivalence class in $\breve{\mathsf{H}}_p(\mathsf{X})$. We call ω a *vertical homology class* if $\omega + \overline{\mathsf{H}}_p(\mathsf{X})$ is not identity in $\breve{\mathsf{H}}_p(\mathsf{X})$. In other words, $\omega \notin \overline{\mathsf{H}}_p(\mathsf{X})$. Two homology classes ω_1 and ω_2 are *vertically homologous* if $\omega_1 + \omega_2 \in \overline{\mathsf{H}}_p(\mathsf{X})$.

We percolate the definitions from the homology classes to cycles. A cycle α is *horizontal* if $[\alpha]$, the standard homology class represented by α, is a horizontal class. Two cycles α_1 and α_2 are *vertically homologous* if $[\alpha_1]$ and $[\alpha_2]$ are vertically homologous. Obviously, two p-cycles α_1 and α_2 are vertically homologous if and only if there is a $(p+1)$-chain B such that $\partial B + \alpha_1 + \alpha_2$ is a horizontal cycle. See the torus in the right figure for an example, where α_2 is a horizontal cycle as it is homologous to α_3 carried by a loop contained in a connected component of a level set; while α_1 is a vertical cycle, i.e, $[\alpha_1]$ is a vertical homology class. We say that $\{\alpha_1, \ldots, \alpha_k\}$ is *a set of base cycles for* $\mathsf{H}_p(\mathsf{X})$ if $\{[\alpha_1], \ldots, [\alpha_k]\}$ form a basis for $\mathsf{H}_p(\mathsf{X})$. A set of base cycles for $\overline{\mathsf{H}}_p(\mathsf{X})$ and $\breve{\mathsf{H}}_p(\mathsf{X})$ are defined analogously.
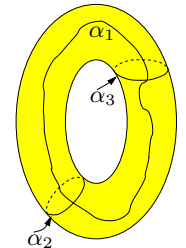
Finally, the *range* of a loop $\gamma \subseteq \mathsf{X}$, denoted by $\mathrm{range}(\gamma)$, is the interval $[\min_{x \in \gamma} f(x), max_{x \in \gamma} f(x)]$. The *height* of this loop, $\mathrm{height}(\gamma)$, is simply the length of $\mathrm{range}(\gamma)$. We extend the definitions of range

and height to cycles by saying that $\mathrm{range}(\alpha) = \mathrm{range}(\gamma)$ and $\mathrm{height}(\alpha) = \mathrm{height}(\gamma)$ where the cycle $\alpha \in \mathbf{Z}_1(X)$ is carried by the loop γ. The *height of a homology class* ω, denoted by $\mathrm{height}(\omega)$, is the minimal height of any cycle in this class. Notice that the height of a horizontal class ω is not necessarily zero since ω may be the addition of multiple height-0 horizontal classes.

Reeb graph. Given a triangulable topological space X and a continuous function $f : X \to \mathbb{R}$, we say that two points $x, y \in X$ are *equivalent*, denoted by $x \sim y$, if and only if x and y belong to the same connected component of X_a for some $a \in \mathbb{R}$. Consider the quotient space $X_\sim$ which is the set of equivalence classes equipped with the quotient topology induced by this equivalence relation; $X_\sim$ is also called the *Reeb graph* of X with respect to f, denoted by $\mathrm{Rb}_f(X)$. See Figure 1 (a) and (b) for an example.
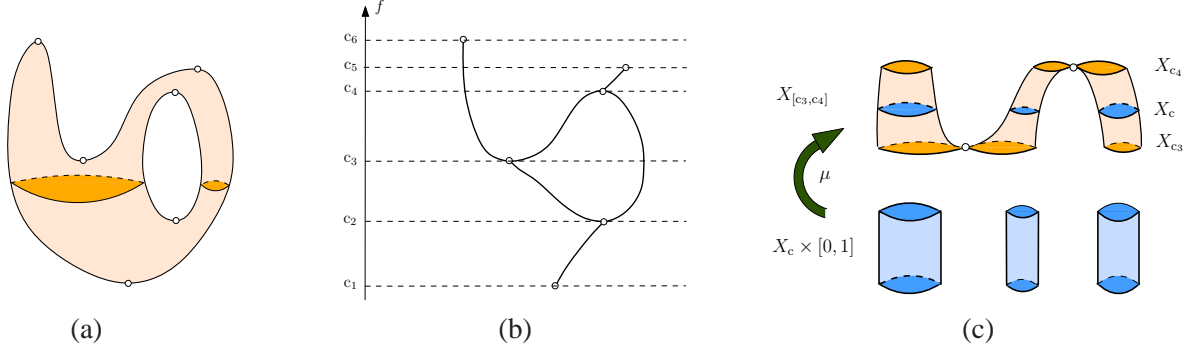

(a) (b) (c)

Figure 1: (a) X is a solid torus and its Reeb graph w.r.t the height function f is shown in (b). (c) f is a levelset-tame function w.r.t. discrete values $\{c_1, \ldots, c_6\}$. There is a continuous map $\mu : X_c \times [0,1] \to X_{[c_3,c_4]}$ whose restriction to the open set $X_c \times (0,1)$ is a homeomorphism. In the top row, there are three disjoint interval-components in $X_{(c_3,c_4)}$ whose closures may intersect in level-sets X_{c_3} and X_{c_4}.

An alternative way to view the Reeb graph is that there is a natural continuous surjection $\Phi : X \to X_\sim$ where $\Phi(x) = \Phi(y)$ *if and only if* x and y come from the same connected component of a level set of f. In this sense, $\mathrm{Rb}_f(X)$ is obtained by continuously identifying each connected component. The map Φ induces a scalar function $\tilde{f} : \mathrm{Rb}_f(X) \to \mathbb{R}$ where $\tilde{f}(p) = f(x)$ if $p = \Phi(x)$. Since $f(x) = f(y)$ whenever $\Phi(x) = \Phi(y)$, the function $\tilde{f}$ is well-defined. Since f is continuous, so is $\tilde{f}$. The range or height of a loop in $\mathrm{Rb}_f(X)$ is measured with respect to this function $\tilde{f}$. In this paper, we also use f to refer to $\tilde{f}$ for simplicity.

3 Reeb graphs and vertical homology

In this section, we show that $H_1(\mathrm{Rb}_f(X))$ and the first vertical homology group $\check{H}_1(X)$ of X are isomorphic. This relation is observed for 2-manifolds in [9], but to the best of our knowledge, it has not been formally introduced and proved anywhere yet for general topological spaces. We include it here for completion.

The surjection $\Phi : X \to \mathrm{Rb}_f(X)$ induces a chain map $\Phi_\#$ from the 1-chains of X to the 1-chains of $\mathrm{Rb}_f(X)$ which eventually induces a homomorphism $\Phi_* : H_1(X) \to H_1(\mathrm{Rb}_f(X))$. For the horizontal subgroup $\overline{H}_1(X)$, we have that $\Phi_*(\overline{H}_1(X)) = \{0\} \subseteq \overline{H}_1(\mathrm{Rb}_f(X))$. Hence Φ_* induces a well-defined homomorphism between the quotient groups

$$\check{\Phi} : \check{H}_1(X) = \frac{H_1(X)}{\overline{H}_1(X)} \longrightarrow \frac{H_1(\mathrm{Rb}_f(X))}{\overline{H}_1(\mathrm{Rb}_f(X))} = H_1(\mathrm{Rb}_f(X)).$$

The right equality above follows from that $\overline{H}_1(\mathrm{Rb}_f(X)) = \{0\}$, which holds because every level set of $\mathrm{Rb}_f(X)$ consists only of a set of disjoint points. In what follows, we show that $\check{\Phi}$ is an isomorphism under
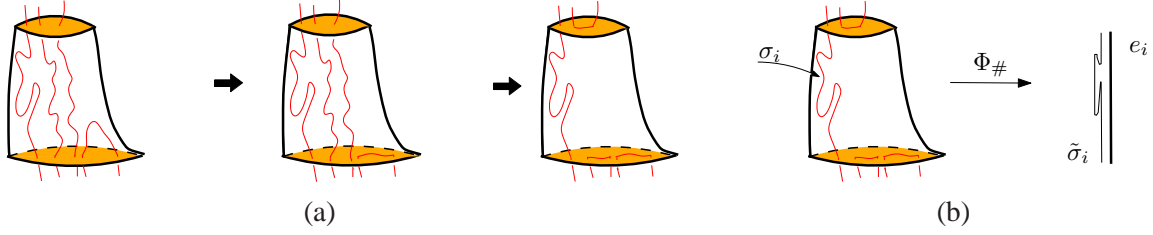
4

Figure 2: (a) Left: the interval-component $\mu(\mathcal{C})$ contains three through-paths and one turning-path. Middle: the turning-path can be deformed to a path contained in the level set X_{c_z}. Right: $\gamma_{\mathcal{C}}$ is modified so that at most one through-path is left. (b) The image of the singular simplex σ_i is the through-path in this interval-component. Its image is a singular simplex $\tilde{\sigma}_i : [0,1] \to |e_i|$ which we draw with the the thin curve slightly off $|e_i|$ for illustration purpose.

some mild conditions. Intuitively, this is not surprising as Φ maps each contour in the level set to a single point, which in turn also collapses every horizontal cycle.

For technical reasons, we consider functions that behave nicely. Specifically, we call a continuous function $f : \mathsf{X} \to \mathbb{R}$ *levelset-tame* if there exist finite number of discrete values $\{c_1, \ldots, c_k\}$ so that the following holds: for any two consecutive c_i and c_{i+1}, (i) there is a homeomorphism $\mu_i : \mathsf{X}_c \times (0,1) \to \mathsf{X}_{(c_i, c_{i+1})}$ for an arbitrary $c \in (c_i, c_{i+1})$; and (ii) the homeomorphism μ_i can be extended to a continuous map $\mu_i : \mathsf{X}_c \times [0,1] \to X_{[c_i, c_{i+1}]}$. In this case, we also say that f is levelset-tame w.r.t. the set of discrete values $\{c_1, \ldots, c_k\}$; note that the choice of c_is and μ_is are not unique. See Figure 1 (c) for an example. It can be shown that Morse functions on a compact smooth manifold and piecewise-linear functions on a finite simplicial complex are both levelset-tame functions.

First, we prove the following result which implies that the map $\check{\Phi} : \check{\mathsf{H}}_1(\mathsf{X}) \to \mathsf{H}_1(\mathsf{Rb}_f(\mathsf{X}))$ as introduced above is injective.

Lemma 3.1 *Let $f : \mathsf{X} \to \mathbb{R}$ be a levelset-tame function, and Φ, Φ_* as defined before. Then we have that $ker(\Phi_*) = \overline{\mathsf{H}}_1(\mathsf{X})$ where $ker(\Phi_*)$ denotes the kernel of Φ_*.*

Proof: Since Φ maps all points in the same connected component in a level set of f into a single point, we have that $\overline{\mathsf{H}}_1(\mathsf{X}) \subseteq \ker(\Phi_*)$. Hence we now focus on proving the opposite direction $\ker(\Phi_*) \subseteq \overline{\mathsf{H}}_1(\mathsf{X})$. That is, for any homology class $h \in \mathsf{H}_1(\mathsf{X})$, if $\Phi_*(h) = 0$, then $h \in \overline{\mathsf{H}}_1(\mathsf{X})$. Specifically, take a loop $\gamma \subseteq \mathsf{X}$ carrying a cycle from the class h. We will show that there exists a loop $\widehat{\gamma}$ which is contained in the union of a discrete set of level sets and which is homologous to γ. This will then imply that h is horizontal.

Assume without loss of generality that γ is the image of a map $\mathbb{S}^1 \to \mathsf{X}$; the case when γ is a finite union of such images can be handled by applying the following proof to each image of $\mathbb{S}^1$.

Let $\{c_1, \ldots, c_k\}$ be a set of discrete values with respect to which f is levelset-tame. Fix an arbitrary interval $[c_z, c_{z+1}]$ and any $c \in (c_z, c_{z+1})$. By definition of a levelset-tame function, there exists a continuous map $\mu : \mathsf{X}_c \times [0,1] \to \mathsf{X}_{[c_z, c_{z+1}]}$ whose restriction to the open set $\mathsf{X}_c \times (0,1)$ is a homeomorphism onto $\mathsf{X}_{(c_z, c_{z+1})}$. The product space $\mathsf{X}_c \times [0,1]$ has several connected components each of which, called a *cylinder*, corresponds to the product between a connected component in the level-set X_c and $[0,1]$. The images of all such cylinders under μ can touch each other only in X_{c_z} or in $\mathsf{X}_{c_{z+1}}$ when μ is no longer a homeomorphism. See Figure 1 (c) for an illustration, where in this example, $\mathsf{X}_c \times [0,1]$ has three cylinders. Let us consider a single cylinder $\mathcal{C} = S \times [0,1]$, where S is the corresponding connected component in X_c. Denote by $\mathcal{C}^o$ the open cylinder $S \times (0,1)$. We call the image $\mu(\mathcal{C}^o)(\subseteq \mathsf{X})$ of every open cylinder $\mathcal{C}^o$ an *interval-component of* X. Note that all interval-components of X are disjoint, and so are their images under the map Φ in the Reeb graph $\mathsf{Rb}_f(\mathsf{X})$.

Next, consider $\gamma_{\mathcal{C}^o} = \gamma \cap \mu(\mathcal{C}^o)$ and $\gamma_{\mathcal{C}} = \gamma \cap \mu(\mathcal{C})$, which are the intersections of γ with the interval-component $\mu(\mathcal{C}^o)$ and with the closure of $\mu(\mathcal{C}^o)$, respectively. Each connected component in $\gamma_{\mathcal{C}^o}$ is a path of the following two types: a *through-path* π where the two endpoints of its closure lie in X_{c_z} and $\mathsf{X}_{c_{z+1}}$, respectively; and a *turning-path* π where the endpoints of its closure either lie both in X_{c_z} or both in $\mathsf{X}_{c_{z+1}}$. The closure of a through-path or a turning-path in $\mu(\mathcal{C}^o)$ is called a through-path or a turning-path in $\mu(\mathcal{C})$. It can be verified that any turning-path π with endpoints p and q can be continuously deformed to a path connecting p and q within the same contour of a level set, using an argument similar to what we invoke below. Therefore we can transform γ to another homologous loop that contains only through-paths in $\gamma_{\mathcal{C}}$. See Figure 2 (a) for an illustration. As such, from now on, we assume that $\gamma_{\mathcal{C}}$ contains only through-paths.

Our arguments consist of two steps. In **Step 1**, we modify γ into another homologous loop γ' which contains at most one through-path within any interval-component of X. In **Step 2**, we show that if $\Phi(\gamma')$ is null-homologous in $\mathsf{Rb}_f(\mathsf{X})$, then γ' must have no through-path in any interval-component of X, implying that γ' is contained only in the level-sets $\bigcup_i \mathsf{X}_{c_i}$. Hence γ' carries a horizontal cycle and h is a horizontal homology class.

Step 1. In this step, we modify $\gamma_{\mathcal{C}}$ so that it contains only portions lying in the two level-sets $\mathsf{X}_{c_z} \cup \mathsf{X}_{c_{z+1}}$, together with at most one through-path in $\mu(\mathcal{C})$. See Figure 2 (a). Specifically, suppose there are more than one through-paths in $\gamma_{\mathcal{C}}$. Then, for any pair of through-paths π_1 and π_2, we show that there exists a 2-chain B such that $\partial B + \pi_1 + \pi_2$ is contained in the two level sets X_{c_z} and $\mathsf{X}_{c_{z+1}}$. Hence, we can convert γ to $\gamma' = \gamma + (\partial B + \pi_1 + \pi_2)$ and the intersection $\gamma'_{\mathcal{C}} = \gamma' \cap \mu(\mathcal{C})$ has two fewer through-paths than $\gamma_{\mathcal{C}}$. Obviously, γ is homologous to γ'. By continuing this process, we can cancel out all pairs of through-paths in $\gamma_{\mathcal{C}}$ till at most one through-path is left, and the resulting loop γ' is homologous to γ.
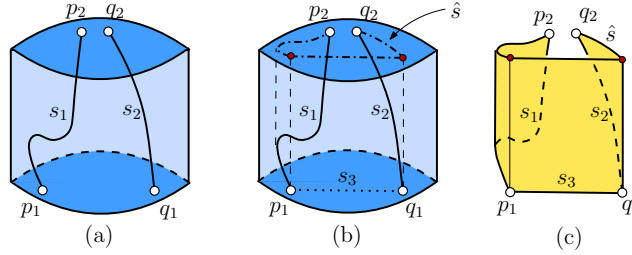

Figure 3: (a) An illustration of the cylinder $\mathcal{C} = S \times [0, 1]$, where each horizontal slice of this cylinder is a copy of S. (b) $\hat{s}$ is the projection of $s = s_1 \circ s_3 \circ s_2$ from the product space onto the slice $\mathcal{C}[1]$. (c) The boundary of the surface B' is $s + \hat{s}$.

We now show how to construct a 2-chain B for a pair of through-paths π_1 and π_2 from $\gamma_{\mathcal{C}}$. Let $\pi_1{}^o$ and $\pi_2{}^o$ denote the interiors of π_1 and π_2, respectively. Note that $\pi_1{}^o$ and $\pi_2{}^o$ are contained in the image $\mu(\mathcal{C}^o) \subseteq \mathsf{X}_{(c_z, c_{z+1})}$ of the open cylinder $\mathcal{C}^o = S \times (0, 1)$. Since the restriction of μ to the open set $\mathcal{C}^o$ is a homeomorphism, $\pi_1{}^o$ and $\pi_2{}^o$ have unique pre-images $s_1{}^o$ and $s_2{}^o$ in $\mathcal{C}^o$ under μ. Let s_1 (resp. s_2) denote the closure of $s_1{}^o$ (resp. $s_2{}^o$) in $\mathcal{C}$, with p_1 and p_2 (resp. q_1 and q_2) being its endpoints. See Figure 3 (a) for an illustration. Notice that $\mu(s_1) = \pi_1$ and $\mu(s_2) = \pi_2$ due to the continuity of μ.

Since the cylinder $\mathcal{C}$ is the product space $S \times [0, 1]$, every point $\mathbf{x} \in \mathcal{C}$ can be represented as $\mathbf{x} = (x, t)$, where $x \in S$ is called its *horizontal coordinate* and $t \in [0, 1]$ is its *vertical coordinate* (or *height*). We use a *slice* $\mathcal{C}[t]$ to refer to one copy of S at height t.

Since each slice $\mathcal{C}[t]$ of the cylinder $\mathcal{C}$ is path-connected, there is a path, say s_3, that connects p_1 and q_1 in $\mathcal{C}[0]$. Let s denote the concatenated curve $s_1 \circ s_3 \circ s_2$; see Figure 3 (b). Now for every point $\mathbf{x} = (x, t_x) \in s$, consider the "vertical line" $l_x = \{(x, t) \mid t \in [t_x, 1]\}$. That is, l_x contains the images of $\mathbf{x}$ in each slice $\mathcal{C}[t]$ with $t \geq t_x$. The union of l_xs for all $\mathbf{x} \in s$ traces out a 2-dimensional surface B'. The boundary of B' is

$\partial B' = s \circ \hat{s}$ where $\hat{s}$ is the image of s in $\mathcal{C}[1]$. See Figure 3 (b) and (c). Through the continuous map μ, we obtain a 2-chain B whose carrier is $\mu(B') \subseteq \mathsf{X}_{[c_z, c_{z+1}]}$ and $\partial \mu(B') = \pi_1 \circ \mu(s_3) \circ \pi_2 \circ \mu(\hat{s})$. Furthermore, $\mu(s_3)$ and $\mu(\hat{s})$ lie in the level-sets $\mathsf{X}_{c_z} \cup \mathsf{X}_{c_{z+1}}$. Hence by taking $\gamma' = \gamma + \partial \mu(B')$, we have reduced a pair of through-paths.

Now we group through-paths in $\gamma_\mathcal{C}$ into pairs, with at most one left unpaired. We construct a 2-chain for every pair, and let $\mathcal{B}$ denote the union of all these 2-chains. Obviously, $\gamma' = \gamma + \partial \mathcal{B}$ is homologous to γ and its intersection $\gamma' \cap \mu(\mathcal{C})$ has at most one through-path. By performing this procedure for all cylinders and for all intervals $[c_z, c_{z+1}]$, $z = 1, \ldots, k-1$, we obtain a loop $\widehat{\gamma}$ which is homologous to γ, and has at most one through-path within each interval-component in X.

Step 2. We now choose a specific 1-cycle $\alpha = \sum_{i=1}^r \sigma_i + \sum_{j=1}^t \rho_j$ carried by $\widehat{\gamma}$ that is of the following form: there are two types of singular simplices in α: a simplex σ_i whose image in X is a through-path and a simplex ρ_j whose image is completely contained within a level set X_{c_z} for some $z \in [1, k]$. Consider the image of α in $\mathbf{Z}_1(\mathsf{Rb}_f(\mathsf{X}))$, $\tilde{\alpha} := \Phi_\#(\alpha) = \sum_{i=1}^r \tilde{\sigma}_i + \sum_{j=1}^t \tilde{\rho}_j$, with $\tilde{\sigma}_i = \Phi_\#(\sigma_i)$ and $\tilde{\rho}_j = \Phi_\#(\rho_j)$. Since the map Φ collapses each connected component in a level set to a single point, each $\tilde{\rho}_j$ is a constant map, and hence $\tilde{\alpha}$ is homologous to $\sum_{i=1}^r \tilde{\sigma}_i$, which we still denote as $\tilde{\alpha}$ for simplicity.

Now insert a set of vertices V to $\mathsf{Rb}_f(\mathsf{X})$, which is the set of points with function value $f(c_i)$ for $i \in [1, k]$. The removal of these vertices from $\mathsf{Rb}_f(\mathsf{X})$ leaves a set of connected components. Since the function $f : \mathsf{X} \to \mathbb{R}$ is levelset-tame w.r.t. $\{c_1, \ldots, c_k\}$, each such connected component is necessarily the image of some continuous bijection $g : (0, 1) \to \mathsf{Rb}_f(\mathsf{X})$, and we call each connected component *an arc of* $\mathsf{Rb}_f(\mathsf{X})$. Indeed, each such connected component is the image of some interval-component of X under the map Φ. Since an interval-component T of X is the evolution of a connected component in a level set without changing its topology, $\Phi(T)$ is necessarily a piece of curve monotone in the function values. Also observe that, by the definition of interval-components, all such arcs are disjoint. Hence we obtain a triangulation K of $\mathsf{Rb}_f(\mathsf{X})$ whose vertices are V and edges are the closures of those arcs defined above.

By the construction of $\widehat{\gamma}$, the image of each singular simplex σ_i is contained in a different interval-component. Hence $\tilde{\sigma}_i([0, 1])$ is contained within the underlying space of a single edge e in K. The boundary of $\tilde{\sigma}_i$ coincides with endpoints of e which are vertices in V. See Figure 2 (b) for an illustration. Given an edge $e \in K$, let $|e| \subseteq |K| = \mathsf{Rb}_f(\mathsf{X})$ denote the underlying space of e. Let $e_i \in K$ denote the edge such that $\tilde{\sigma}_i$ is a map $\tilde{\sigma}_i : [0, 1] \to |e_i|$. Observe that each $\tilde{\sigma}_i$ is mapped to a unique edge e_i.

Finally, consider the singular cycle $\tilde{\alpha} = \Sigma_{i=1}^r \tilde{\sigma}_i$. The carrier for this cycle is homotopic to the carrier of the cycle $h = \Sigma_{i=1}^r(h_i : [0, 1] \to |e_i|)$ where h_i is a homeomorphism. Thus the two cycles h and $\tilde{\alpha}$ are homologous. Consider the simplicial cycle $g = \sum_{i=1}^r e_i$, and let $[g] \in \overline{\mathsf{H}}_1(K)$ denote the simplicial homology class it belongs to. The class $[g]$ identifies to $[h]$ via the standard isomorphism between simplicial homology groups $\mathsf{H}_1(K)$ and the singular homology group $\mathsf{H}_1(|K|)$ (see e.g, page 194 of [23]). Therefore, this standard isomorphism also identifies $[g]$ to $[\tilde{\alpha}]$. On the other hand, in simplicial homology, as there are no 2-simplices in K, g is null-homologous if and only if $g = \emptyset$, which means that the number r of singular simplices in $\tilde{\alpha}$ is necessarily zero if $\tilde{\alpha}$ is null-homologous. This implies that the loop $\widehat{\gamma} \subset \mathsf{X}$ does not contain any through-path, and is completely contained within the union of level-sets $\bigcup_z \mathsf{X}_{c_z}$. Hence $\widehat{\gamma}$ (and thus γ) carries a horizontal cycle and its corresponding homology class h is horizontal. In other words, if $\Phi_*(h) = 0$ then $h \in \overline{\mathsf{H}}_1(\mathsf{X})$, implying $ker(\Phi_*) \subseteq \overline{\mathsf{H}}_1(\mathsf{X})$. Combining this with that $\overline{\mathsf{H}}_1(\mathsf{X}) \subseteq ker(\Phi_*)$ completes our proof.

∎

Claim 3.2 *Let* $f : \mathsf{X} \to \mathbb{R}$ *be a levelset-tame function, and* Φ *as defined before. Given any loop* $\gamma \subseteq \mathsf{Rb}_f(\mathsf{X})$, *there is a loop* $\widehat{\gamma} \subseteq \mathsf{X}$ *such that* $\Phi(\widehat{\gamma}) = \gamma$ *and* $\mathrm{range}(\widehat{\gamma}) = \mathrm{range}(\gamma)$.

Proof: We construct $\widehat{\gamma}$ as follows. Suppose that γ consists of a sequences of k arcs,

$$\gamma[p_1, p_2], \gamma[p_2, p_3], \ldots, \gamma[p_k, p_1],$$

where each p_i is a node of the Reeb graph $\mathsf{Rb}_f(\mathsf{X})$. For each p_i, choose an arbitrary pre-image $q_i \in \mathsf{X}$ from $\Phi^{-1}(p_i)$. Now for each arc $\gamma[p_i, p_{i+1}]$, connect q_i and q_{i+1} within $\mathsf{X}(\gamma[p_i, p_{i+1}])$ by $\widehat{\gamma}[q_i, q_{i+1}]$ arbitrarily. The concatenatin of all $\widehat{\gamma}[q_i, q_{i+1}]$ provides $\widehat{\gamma} \subset \mathsf{X}$. It is easy to check that $\mathrm{range}(\gamma) = \mathrm{range}(\widehat{\gamma})$. ∎

Theorem 3.3 *Given a levelset-tame function $f : \mathsf{X} \to \mathbb{R}$, let $\check{\Phi} : \check{\mathsf{H}}_1(\mathsf{X}) \to \mathsf{H}_1(\mathsf{Rb}_f(\mathsf{X}))$ be the homomorphism induced by the surjection $\Phi : \mathsf{X} \to \mathsf{Rb}_f(\mathsf{X})$ as defined before. The map $\check{\Phi}$ is an isomorphism. Furthermore, for any vertical homology class $\omega \in \check{\mathsf{H}}_1(\mathsf{X})$, we have that $\mathrm{height}(\omega) = \mathrm{height}(\check{\Phi}(\omega))$.*

Proof: First, by Claim 3.2, the homomorphism $\Phi_* : \mathsf{H}_1(\mathsf{X}) \to \mathsf{H}_1(\mathsf{Rb}_f(\mathsf{X}))$ is surjective. It then follows that its induced quotient map $\check{\Phi}$ is also surjective. The injectivity of $\check{\Phi}$ follows from Lemma 3.1. Hence $\check{\Phi}$ is an isomorphism.

For the second part of the theorem, suppose α is a vertical cycle such that $[\alpha] = \omega$ and $\mathrm{height}(\alpha) = \mathrm{height}(\omega)$, i.e., α is *a thinnest cycle* in the vertical homology class ω. Let γ be the loop in $\mathsf{Rb}_f(\mathsf{X})$ that carries a thinnest cycle in the homology class $\check{\Phi}(\omega) \in \mathsf{H}_1(\mathsf{Rb}_f(\mathsf{X}))$. We have that

$$\mathrm{height}(\alpha) = \mathrm{height}(\Phi_\#(\alpha)) \geq \mathrm{height}(\check{\Phi}(\omega)) = \mathrm{height}(\gamma) \tag{1}$$

On the other hand, by Claim 3.2, there is a loop $\widehat{\gamma}$ in X (which is a pre-image of γ under Φ) such that $\Phi(\widehat{\gamma}) = \gamma$ and $\mathrm{height}(\widehat{\gamma}) = \mathrm{height}(\gamma)$. Let $\hat{\alpha}$ be any 1-cycle carried by $\widehat{\gamma}$. By Lemma 3.1, we have $[\hat{\alpha}] = \omega$, as the cycle $\alpha + \hat{\alpha}$ is mapped to a null-homologous cycle in $\mathsf{Rb}_f(\mathsf{X})$. Hence $\mathrm{height}(\gamma) = \mathrm{height}(\widehat{\gamma}) \geq \mathrm{height}(\alpha)$. Combining this with Eqn (1) proves that $\mathrm{height}(\check{\Phi}(\omega)) = \mathrm{height}(\omega)$. ∎

4   Approximating Reeb graphs

Let M be a compact and smooth m-manifold without boundary embedded in $\mathbb{R}^d$. The *reach* $\rho(\mathsf{M})$ of M is the minimal distance from any point $x \in \mathsf{M}$ to the so-called medial axis of M. Given a point $p \in \mathsf{M}$, let $B_\mathsf{M}(p, r)$ denote the open geodesic ball centered at p with radius r. Let r_p be the maximal radius so that $B_\mathsf{M}(p, r_p)$ is convex in the sense that the minimizing geodesics between any two points in $B_\mathsf{M}(p, r_p)$ is contained in $B_\mathsf{M}(p, r_p)$. The *convexity radius of* M is simply $\rho_c(\mathsf{M}) = \inf_{p \in \mathsf{M}} r_p$.

A set of points P is an ε-*sample* [2] of M if $P \subset \mathsf{M}$ and for any point $x \in \mathsf{M}$, there is a point $p \in P$ within ε geodesic distance from x. Given P and a real $r > 0$, the *Čech complex* $\mathcal{C}^r(P)$ is a simplicial complex where a simplex $\sigma \in \mathcal{C}^r(P)$ if and only if the vertices of σ are the centers of d-balls of radius $r/2$ with a non-empty common intersection. Instead of common intersection, if we only require pairwise intersection among the set of d-balls, we obtain the so-called *Vietoris-Rips complex (Rips complex for short)* $\mathcal{R}^r(P)$.

Overview. Consider an ε-sample $P \subset \mathsf{M}$ and a function $f : \mathsf{M} \to \mathbb{R}$ with its value only available at sample points in P. In what follows, we show that for an appropriate r, the Reeb graph of the Rips complex $\mathcal{R}^r(P)$ approximates $\mathsf{Rb}_f(\mathsf{M})$ both in terms of the rank of the first homology group, and in terms of the range and the height of cycles and homology classes. Our precise definition of approximation will be given later. Once the Rips complex is constructed, computing its Reeb graph takes only $O(n \log n)$ expected time [20], where n is the size of the 2-skeleton of $\mathcal{R}^r(P)$. Since f is only available at sample points in P, the approximation quality naturally depends on how well the function $f : \mathsf{M} \to \mathbb{R}$ behaves. We assume that f is Lipschitz with Lipschitz constant Lip_f.

In Section 4.1 we first introduce some relations between cycles of M and those of the geometric realization $|\mathcal{R}^r(P)|$ of the Rips complex $\mathcal{R}^r(P)$. Using these relations, in Section 4.2, we show that there are maps between $\mathsf{H}_1(\mathsf{M})$ and $\mathsf{H}_1(|\mathcal{R}^r(P)|)$ that are not only isomorphic, but also preserve the height / range of a homology class. This, combined with Theorem 3.3, eventually leads to our approximation of $\mathsf{Rb}_f(\mathsf{M})$.

---

[2] Here ε-sample is not defined relative to reach or feature size as commonly done in reconstruction literature [12].

This approximation result can be used to estimate the first betti number of an orientable 2-manifold from its point samples in near-linear expected time.

4.1 Relation between cycles in M and $|\mathcal{R}^r(P)|$

The simplicial complex $\mathcal{R}^r(P)$ as defined is not necessarily embedded in $\mathbb{R}^d$. Consider the embedding $e \colon \mathcal{R}^r(P) \to \Delta^{|P|}$ of $\mathcal{R}^r(P)$ into the standard simplex in $\mathbb{R}^{|P|}$. Let $|\mathcal{R}^r(P)|$ denote the underlying space of the geometric realization $e(\mathcal{R}^r(P))$. A piecewise-linear function f on $\mathcal{R}^r(P)$ defines naturally a piecewise linear function on its geometric realization $|\mathcal{R}^r(P)|$ which we also denote as f. The Reeb graph of a PL-function f on $\mathcal{R}^r(P)$ is in fact the Reeb graph of f on its geometric realization $|\mathcal{R}^r(P)|$. Hence $\mathsf{Rb}_f(\mathcal{R}^r(P)) := \mathsf{Rb}_f(|\mathcal{R}^r(P)|)$. Analogously, the vertical / horizontal homology groups of $\mathcal{R}^r(P)$ with respect to a PL-function f are also defined using $|\mathcal{R}^r(P)|$. In this section, we relate cycles from M and those from $|\mathcal{R}^r(P)|$ via (simplicial) cycles of $\mathcal{R}^r(P)$. We will show how to construct the maps as indicated in Figure 4 below, such that these maps not only induce isomorphisms in the corresponding homology groups, but also preserve height and range of cycles.
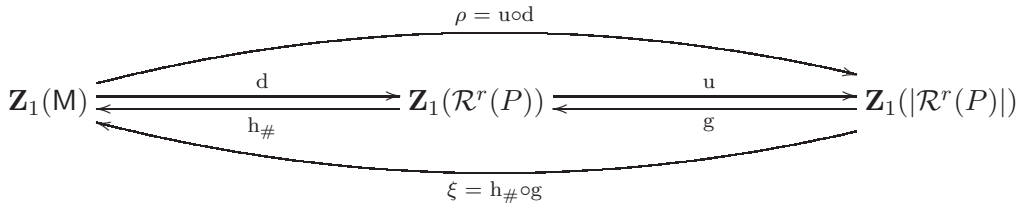

Figure 4: Maps between cycle groups

A general version of the next claim which establishes an isomorphism between the homology groups of M and those of Čech and Rips-complexes is well-known (see, e.g. [24] for Čech-complexes and [22] for Rips-complexes; a variant for compact spaces was also observed by Steve Oudot (personal communications) and a much stronger result showing that Rips-complexes capturing topology of sampled shapes is given in [1]). We include a proof of it for completeness. First, we quote a result from [14], the map of which will be used later as well.

Proposition 4.1 (Proposition 3.3 of [14]) *Let $P \subset M$ be an ε-sample and r a parameter such that $2\varepsilon \le r \le \sqrt{\frac{3}{5}}\rho(M)$. There is a homotopy equivalence $\theta : \mathcal{C}^{2r}(P) \to M$ such that $\theta(p) = p$ for any $p \in P$ and $\theta(\sigma) \subset M \cap (\cup_{p \in Vert(\sigma)} B_M(p, r))$.*

Lemma 4.2 *Let $P \subset M$ be an ε-sample and r a parameter such that $4\varepsilon \le r \le \frac{1}{2}\sqrt{\frac{3}{5}}\rho(M)$. Then,*

$$\mathsf{H}_1(\mathcal{C}^r(P)) \simeq \mathsf{H}_1(\mathcal{R}^r(P)) \simeq \mathsf{H}_1(\mathcal{C}^{2r}(P)) \simeq \mathsf{H}_1(M).$$

The first two isomorphisms are induced by the natural inclusion from $\mathcal{C}^r(P)$ to $\mathcal{R}^r(P)$ and then to $\mathcal{C}^{2r}(P)$. The last isomorphism is induced by the homotopy equivalence θ from Proposition 4.1.

Proof: Consider the following sequence of inclusions:

$$\mathcal{C}^r(P) \overset{i_1}{\hookrightarrow} \mathcal{R}^r(P) \overset{i_2}{\hookrightarrow} \mathcal{C}^{2r}(P).$$

By Proposition 3.4 [14], we know that the inclusion $i = i_2 \circ i_1$ induces an isomorphism $\mathsf{H}_1(\mathcal{C}^r(P)) \simeq \mathsf{H}_1(\mathcal{C}^{2r}(P))$. On the other hand, note that $\mathcal{C}^r(P)$ and $\mathcal{R}^r(P)$ share the same edge set, and $\mathcal{R}^r(P)$ only has

more triangles than $\mathcal{C}^r(P)$. Hence the inclusion i_1 induces a surjective homomorphism from $\mathsf{H}_1(\mathcal{C}^r(P))$ to $\mathsf{H}_1(\mathcal{R}^r(P))$. It then follows that both i_1 and i_2 must induce isomorphisms in the corresponding first homology groups. ∎

Maps d **and** h$_\#$**.** We now define maps as indicated in Figure 4. First, given a cycle $\alpha \in \mathbf{Z}_1(\mathsf{M})$, we map it to a cycle $\mathrm{d}(\alpha) \in \mathbf{Z}_1(\mathcal{R}^r(P))$ using the same Decomposition method [2] as applied in [14]. In particular, use an arbitrary, but fixed, way to break the carrier of α into pieces where each piece has length at most $r - 2\varepsilon$. For each piece with endpoints x_i and x_{i+1}, find the closest sample points p_i and p_{i+1} from P to x_i and x_{i+1}, respectively, and connect p_i and p_{i+1} (which is necessarily an edge in $\mathcal{R}^r(P)$ by triangle inequality). The resulting simplicial 1-cycle in $\mathcal{R}^r(P)$ is $\mathrm{d}(\alpha)$. Later in Lemma 4.3, we will show that this map d indeed takes homologous cycles to homologous cycles, and as such induces a well-defined homomorphism d_* at the homology level.

We define the map $\mathrm{h} : \mathcal{R}^r(P) \to \mathsf{M}$ as the inclusion map $\mathcal{R}^r(P) \hookrightarrow \mathcal{C}^{2r}(P)$ composed with the homotopy equivalence $\theta : \mathcal{C}^{2r}(P) \to \mathsf{M}$ introduced in Proposition 4.1. The corresponding chain map $\mathrm{h}_\#$ induces a homomorphism $\mathrm{h}_* : \mathsf{H}_p(\mathcal{R}^r(P)) \to \mathsf{H}_p(\mathsf{M})$. We restrict h_* only to the first homology group $\mathrm{h}_* : \mathsf{H}_1(\mathcal{R}^r(P)) \to \mathsf{H}_1(\mathsf{M})$. By Lemma 4.2, h_* is an isomorphism.

The following lemma states that $\mathrm{d} \colon \mathbf{Z}_1(\mathsf{M}) \to \mathbf{Z}_1(\mathcal{R}^r(P))$ is in fact the homology-inverse of $\mathrm{h}_\#$. The ranges of mapped cycles are also related. We put the proof of the following lemma in Appendix A to maintain the flow of the presentation. Given two intervals $I_1 = [a, b]$ and $I_2 = [c, d]$, we say that I_1 is *oneside-δ-close to I_2* if $[a, b] \subseteq [c - \delta, d + \delta]$. and I_1 and I_2 are *δ-Hausdorff-close* if the two intervals are oneside-δ-close to each other. In the Lemma below, assume that f is a (Lip$_f$)-Lipschitz function on M and its values for the vertices $P \subset \mathsf{M}$ define a piecewise linear function on $\mathcal{R}^r(P)$ which we also denote as f.

Lemma 4.3 *(i)* $\mathrm{h}_* : \mathsf{H}_1(\mathcal{R}^r(P)) \to \mathsf{H}_1(\mathsf{M})$ *is an isomorphism. The map* d *induces an isomorphism* $\mathrm{d}_* \colon \mathsf{H}_1(\mathsf{M}) \to \mathsf{H}_1(\mathcal{R}^r(P))$ *such that* $\mathrm{h}_* = (\mathrm{d}_*)^{-1}$.

(ii) The range of the cycle $\mathrm{d}(\alpha) \in \mathbf{Z}_1(\mathcal{R}^r(P))$ is oneside-$(r \cdot \mathrm{Lip}_f)$-close to the range of $\alpha \in \mathbf{Z}_1(\mathsf{M})$. Similarly, the range of the cycle $\mathrm{h}_\#(\hat{\alpha}) \in \mathbf{Z}_1(\mathsf{M})$ is oneside-$(r \cdot \mathrm{Lip}_f)$-close to $\hat{\alpha} \in \mathbf{Z}_1(\mathcal{R}^r(P))$.

(iii) The ranges of any homology class $\omega \in \mathsf{H}_1(\mathsf{M})$ (resp. $\hat{\omega} \in \mathsf{H}_1(\mathcal{R}^r(P))$) and its image $\mathrm{d}_*(\omega) \in \mathsf{H}_1(\mathcal{R}^r(P))$ (resp. $\mathrm{h}_*(\hat{\omega}) \in \mathsf{H}_1(\mathsf{M})$), are $(r \cdot \mathrm{Lip}_f)$-Hausdorff-close.

Maps u **and** g**.** The map u is taken as the standard map between the simplicial chain groups of a simplicial complex and the singular chain groups of its underlying space; see e.g, the map μ defined on page 194 of [23].

We now define the map $\mathrm{g} : \mathbf{Z}_1(|\mathcal{R}^r(P)|) \to \mathbf{Z}_1(\mathcal{R}^r(P))$. Recall we have embedded $\mathcal{R}^r(P)$ in the standard simplex $\Delta^{|P|} \subset \mathbb{R}^{|P|}$, and $|\mathcal{R}^r(P)|$ is the underlying space of this geometric realization $\mathrm{e}(\mathcal{R}^r(P))$ of $\mathcal{R}^r(P)$. In particular, Each vertex $p_i \in P$ is mapped to the point $v_i = (0, \ldots, 0, 1, 0, \ldots, 0) \in \mathbb{R}^{|P|}$ with the ith position 1; and a simplex in $\mathcal{R}^r(P)$ with vertices $\{p_{i_0}, \ldots, p_{i_l}\}$ is mapped to the simplex in $\mathbb{R}^{|P|}$ with vertices $\{v_{i_0}, \ldots, v_{i_l}\}$. Consider a cycle α in $|\mathcal{R}^r(P)|$. The carrier of α passes through a sequence of simplices S of $\mathrm{e}(\mathcal{R}^r(P))$ if a point in the carrier is contained in multiple simplices, then keep the one with the minimum dimension. Let $S = \{\sigma_1, \ldots, \sigma_m\}$. Now choose an arbitrary but fixed vertex u_i for each σ_i, and let $p_{u_i} \in P$ denote the unique pre-image of u_i in $\mathcal{R}^r(P)$ under the embedding map u. Notice that for any two consecutive simplices σ_i and σ_{i+1} that the carrier of α passes through, it is necessary that either σ_i is face of σ_{i+1} or σ_{i+1} is a face of σ_i. Hence either $p_{u_i} = p_{u_{i+1}}$ or $p_{u_i} p_{u_{i+1}}$ is an edge in $\mathcal{R}^r(P)$. Therefore, we map α simply to the cycle $\mathrm{g}(\alpha)$ given by the sequence of vertices $(p_{u_1}, \ldots, p_{u_m}, p_{u_1})$ and edges between them. We have the following result about maps u and g.

Lemma 4.4 *(i) Every cycle α in $\mathcal{R}^r(P)$ is mapped to a cycle $\mathrm{u}(\alpha)$ with the same range in $|\mathcal{R}^r(P)|$ under $\mathrm{u} : \mathbf{Z}_1(\mathcal{R}^r(P)) \to \mathbf{Z}_1(|\mathcal{R}^r(P)|)$. The map $\mathrm{u}_* : \mathsf{H}_1(\mathcal{R}^r(P)) \to \mathsf{H}_1(|\mathcal{R}^r(P)|)$ is an isomorphism, and the ranges of any homology class $\omega \in \mathsf{H}_1(\mathcal{R}^r(P))$ and its image $\mathrm{u}_*(\omega) \in \mathsf{H}_1(|\mathcal{R}^r(P)|)$ are also the same.*

(ii) Every cycle α in $|\mathcal{R}^r(P)|$ is mapped to a cycle $\mathrm{g}(\alpha)$ in $\mathcal{R}^r(P)$ whose range is oneside-$(r \cdot \mathrm{Lip}_f)$-close to that of α. The map $\mathrm{g} : \mathbf{Z}_1(|\mathcal{R}^r(P)|) \to \mathbf{Z}_1(\mathcal{R}^r(P))$ induces an isomorphism $\mathrm{g}_* : \mathsf{H}_1(|\mathcal{R}^r(P)|) \to \mathsf{H}_1(\mathcal{R}^r(P))$, and $\mathrm{g}_* = (\mathrm{u}_*)^{-1}$. The ranges of any homology class $\hat{\omega} \in \mathsf{H}_1(|\mathcal{R}^r(P)|)$ and its image $\mathrm{g}_*(\hat{\omega}) \in \mathsf{H}_1(\mathcal{R}^r(P))$ are $(r \cdot \mathrm{Lip}_f)$-Hausdorff close.

Proof: For part (i) of the lemma, note that it is well known that u induces an isomorphism between the respective simplicial and singular homology groups (see e.g, Theorem 34.3 of [23]). Furthermore, since u maps each simplex to a map whose range is its underlying space, u preserves the range of a cycle.

For part (ii) of the lemma, first observe that for any cycle α from $|\mathcal{R}^r(P)|$, we have that $[\mathrm{u} \circ \mathrm{g}(\alpha)] = [\alpha]$. Indeed, by the construction of g, it is easy to verify that $\mathrm{u} \circ \mathrm{g}(\alpha)$ and α are homotopic. Since u induces an isomorphism from $\mathcal{R}^r(P)$ to $|\mathcal{R}^r(P)|$, it follows that g maps homologous cycles in $|\mathcal{R}^r(P)|$ to homologous cycles in $\mathcal{R}^r(P)$. Hence g induces a well-defined homomorphism $\mathrm{g}_* : \mathsf{H}_1(|\mathcal{R}^r|) \to \mathsf{H}_1(\mathcal{R}^r)$. Furthermore, $\mathrm{g} \circ \mathrm{u}(\alpha') = \alpha'$ for any cycle $\alpha' \in \mathcal{R}^r(P)$. It follows that g_* is the inverse of u_* and hence is an isomorphism.

Finally, note that for each simplex $\sigma \in |\mathcal{R}^r(P)|$, the function value difference between any two points $x, y \in \sigma$ is bounded by $r \cdot \mathrm{Lip}_f$. Let γ be the carrier of a cycle α in $|\mathcal{R}^r(P)|$. By the construction of g, for each piece $\gamma \cap \sigma_i$ of γ within the simplex $\sigma_i \in S$, we have that $|f(x) - f(u_i)| \le r \cdot \mathrm{Lip}_f$ for any point $x \in \gamma \cap \sigma_i$. Since $f(u_i) = f(p_{u_i})$, we have:

$$[\min_{i \in [1,m]} f(p_{u_i}) + r \cdot \mathrm{Lip}_f, \ \max_{i \in [1,m]} f(p_{u_i}) - r \cdot \mathrm{Lip}_f] \subseteq \mathrm{range}(\alpha) \subseteq [\min_{i \in [1,m]} f(p_{u_i}) - r \cdot \mathrm{Lip}_f, \ \max_{i \in [1,m]} f(p_{u_i}) + r \cdot \mathrm{Lip}_f].$$

On the other hand, we have that $\mathrm{range}(\mathrm{g}(\alpha)) \subseteq [\min_{i \in [1,m]} f(p_{u_i}), \max_{i \in [1,m]} f(p_{u_i})]$. Hence $\mathrm{range}(\mathrm{g}(\alpha))$ is oneside-$(r \cdot \mathrm{Lip}_f)$-close to $\mathrm{range}(\alpha)$. By a similar argument as in the proof of Lemma 4.3 (iv), the closeness between the corresponding homology classes follows. ∎

Combining Lemma 4.3 and 4.4, we obtain a similar result for maps between $\mathbf{Z}_1(\mathsf{M})$ and $\mathbf{Z}_1(|\mathcal{R}^r(P)|)$.

Theorem 4.5 *Let $P \subset \mathsf{M}$ be an ε-sample and r a parameter such that $4\varepsilon \le r \le \frac{1}{2}\sqrt{\frac{3}{5}}\rho(\mathsf{M})$.*

(i) There is a map $\rho := \mathrm{u}_\# \circ \mathrm{d}$ from $\mathbf{Z}_1(\mathsf{M})$ to $\mathbf{Z}_1(|\mathcal{R}^r(P)|)$ that induces an isomorphism $\rho_* : \mathsf{H}_1(\mathsf{M}) \to \mathsf{H}_1(|\mathcal{R}^r(P)|)$. The range of cycle $\rho(\alpha)$ is oneside-$(r \cdot \mathrm{Lip}_f)$-close to the range of α.

(ii) There is a map $\xi := \mathrm{h}_\# \circ \mathrm{g}$ from $\mathbf{Z}_1(|\mathcal{R}^r(P)|)$ to $\mathbf{Z}_1(\mathsf{M})$ that induces an isomorphism $\xi_* : \mathsf{H}_1(|\mathcal{R}^r(P)|) \to \mathsf{H}_1(\mathsf{M})$. The range of cycle $\xi(\hat{\alpha})$ is oneside-$(2r \cdot \mathrm{Lip}_f)$-close to the range of cycle $\hat{\alpha}$.

(iii) Furthermore, ρ_* is the inverse of ξ_*. The ranges of any homology class $\omega \in \mathsf{H}_1(\mathsf{M})$ (resp. $\hat{\omega} \in \mathsf{H}_1(|\mathcal{R}^r(P)|)$) and its image $\rho_*(\omega) \in \mathsf{H}_1(|\mathcal{R}^r(P)|)$ (resp. $\xi_*(\hat{\omega}) \in \mathsf{H}_1(\mathsf{M})$) are $(2r \cdot \mathrm{Lip}_f)$-Hausdorff-close.

4.2 $\mathrm{Rb}_f(\mathsf{M})$ and $\mathrm{Rb}_f(\mathcal{R}^r(P))$

We now show that under mild conditions on M, the induced isomorphisms ρ_* and ξ_* as defined above in fact map horizontal classes to horizontal classes, and vertical classes to vertical classes.

Set $s = \mathrm{rank}(\overline{\mathsf{H}}_1(\mathsf{M}))$. It turns out that we can find a basis $\{[\alpha_1], \ldots, [\alpha_s]\}$ for the horizontal subgroup $\overline{\mathsf{H}}_1(\mathsf{M})$ such that each class $[\alpha_i]$, $i \in [1, s]$, has height 0; as well as a set of base cycles $\{\alpha_1, \ldots, \alpha_s\}$ corresponding to this basis with $\mathrm{height}(\alpha_i) = 0$ for any $i \in [1, s]$. Such a 0-*height basis* for $\overline{\mathsf{H}}_1(\mathsf{M})$ can be

11

constructed by a simple greedy approach, where at each iteration we take a homology class with smallest height that is independent of all the previous elements in the basis. The details can be found in Appendix B. The corresponding set of base cycles $\{\alpha_1, \ldots, \alpha_s\}$ with $\mathrm{height}(\alpha_i) = 0$ is called a set of 0-*height base cycles* for $\overline{\mathsf{H}}_1(\mathsf{M})$. For a horizontal homology class ω with height 0, the *span of* ω is the length of the maximal interval I such that ω has a pre-image in the levelset X_a for any $a \in I$. Intuitively, this is the interval in function values in which this homology class survives in the level-sets.

Let $\mathbf{s}^*(\mathsf{M})$ denote the smallest span of any 0-height horizontal class of the input manifold M, and $\mathbf{t}^*(\mathsf{M})$ the minimal height of any vertical class of M. We assume that both $\mathbf{s}^*(\mathsf{M})$ and $\mathbf{t}^*(\mathsf{M})$ are positive for our input level-set tame function on M.

Theorem 4.6 *Given a level-set tame function f on a manifold M, let $r > 0$ be such that $\mathbf{s}^*(\mathsf{M}), \mathbf{t}^*(\mathsf{M}) > 2r \cdot \mathrm{Lip}_f$. Let ρ_* and ξ_* be as defined in Theorem 4.5. Then we have that $\rho_*(\overline{\mathsf{H}}_1(\mathsf{M})) = \overline{\mathsf{H}}_1(|\mathcal{R}^r(P)|)$, $\xi_*(\overline{\mathsf{H}}_1(|\mathcal{R}^r(P)|)) = \overline{\mathsf{H}}_1(\mathsf{M})$ and $\breve{\mathsf{H}}_1(\mathsf{M}) \simeq \breve{\mathsf{H}}_1(|\mathcal{R}^r(P)|)$.*

Proof: For simplicity, in this proof let R denote $|\mathcal{R}^r(P)|$. Below we first show that $\rho_*(\overline{\mathsf{H}}_1(\mathsf{M})) = \overline{\mathsf{H}}_1(\mathrm{R})$. Consider a set of 0-height base cycles $\{\alpha_1, \ldots, \alpha_s\}$ for $\overline{\mathsf{H}}_1(\mathsf{M})$ with $s = \mathrm{rank}(\overline{\mathsf{H}}_1(\mathsf{M}))$.

Take an arbitrary α_i for $i \in [1, s]$, and let $[a, b]$ denote the maximal interval[3] such that $[\alpha_i]$ has a preimage in the levelset M_c for any $c \in [a, b]$. The span of $[\alpha_i]$ is $b - a$ and is at least $\mathbf{s}^*(\mathsf{M}) > 2r \cdot \mathrm{Lip}_f$. Take a representative cycle γ_a from M_a and γ_b from M_b of the homology class $[\alpha_i]$. Set $I_a := [a - r \cdot \mathrm{Lip}_f, a + r \cdot \mathrm{Lip}_f]$ and $I_b := [b - r \cdot \mathrm{Lip}_f, b + r \cdot \mathrm{Lip}_f]$. It follows from Theorem 4.5 that the carrier of $\rho(\gamma_a)$ is contained in the interval levelset R_{I_a} while the carrier of $\rho(\gamma_b)$ is contained in R_{I_b}. (Note, $[\rho(\alpha_i)] = [\rho(\gamma_a)] = [\rho(\gamma_b)]$ is a non-trivial homology class in $\mathsf{H}_1(\mathrm{R})$.) Since $b - a > 2r \cdot \mathrm{Lip}_f$, we have $I_a \cap I_b = \emptyset$. A simple application of the Mayer-Vietoris sequence provides that the homology class $[\rho(\alpha_i)]$ has a preimage in the levelset R_c for any $c \in [a + r \cdot \mathrm{Lip}_f, b - r \cdot \mathrm{Lip}_f]$, which in turn implies that $[\rho(\alpha_i)]$ is horizontal. (A similar argument is used in [9].) Since $[\rho(\alpha_i)]$ is horizontal for any $i \in [1, s]$, $\rho_*(\overline{\mathsf{H}}_1(\mathsf{M}))$ is a subgroup of $\overline{\mathsf{H}}_1(\mathrm{R})$.

We now show that the opposite direction $\overline{\mathsf{H}}_1(\mathrm{R}) \subseteq \rho_*(\overline{\mathsf{H}}_1(\mathsf{M}))$ is also true, which would imply that $\rho_*(\overline{\mathsf{H}}_1(\mathsf{M})) = \overline{\mathsf{H}}_1(\mathrm{R})$. Specifically, take a set of 0-height base cycles $\{\beta_1, \ldots, \beta_t\}$ for $\overline{\mathsf{H}}_1(\mathrm{R})$. By Theorem 4.5, their images $\{\xi(\beta_1), \ldots, \xi(\beta_t)\}$ in M is a set of independent cycles such that $\mathrm{height}(\xi(\beta_i)) \leq 2r \cdot \mathrm{Lip}_f$. Since the minimal height of any vertical cycle in M is $\mathbf{t}^*(\mathsf{M}) > 2r \cdot \mathrm{Lip}_f$, each $\xi(\beta_i)$ has to be a horizontal homology cycle. As such, $\xi_*(\overline{\mathsf{H}}_1(R)) \subseteq \overline{\mathsf{H}}_1(\mathsf{M})$, which means that $\overline{\mathsf{H}}_1(R) = \rho_*(\xi_*(\overline{\mathsf{H}}_1(R))) \subseteq \rho_*(\overline{\mathsf{H}}_1(\mathsf{M}))$. It then follows that $\rho_*(\overline{\mathsf{H}}_1(\mathsf{M})) = \overline{\mathsf{H}}_1(R)$. Since the isomorphism ρ_* sends $\overline{\mathsf{H}}_1(\mathsf{M})$ to $\overline{\mathsf{H}}_1(R)$, the induced homomorphism at the quotient level is also an isomorphism; that is, $\breve{\mathsf{H}}_1(\mathsf{M}) \simeq \breve{\mathsf{H}}_1(R)$. ∎

4.3 Putting everything together

We say that a Reeb graph $\mathsf{Rb}_f(A)$ δ-*approximates* another Reeb graph $\mathsf{Rb}_g(B)$ if there is an isomorphism between $\mathsf{H}_1(\mathsf{Rb}_f(A))$ and $\mathsf{H}_1(\mathsf{Rb}_g(B))$ such that the ranges of corresponding pairs of homology classes are δ-Hausdorff-close[4]. Combining Theorems 3.3, 4.5 and 4.6, we have our first main result.

Theorem 4.7 *Let $f : \mathsf{M} \to \mathbb{R}$ be a level-set tame function defined on M with Lipschitz constant Lip_f. Given an ε-sample P of M, let r be a parameter such that $4\varepsilon \leq r < \min\{\frac{1}{4}\rho(\mathsf{M}), \frac{1}{4}\rho_c(\mathsf{M}), \frac{\mathbf{t}^*}{2\mathrm{Lip}_f}, \frac{\mathbf{s}^*}{2\mathrm{Lip}_f}\}$, and $\mathcal{R}^r(P)$ the Rips complex constructed from P using radius $r/2$. Then $\mathsf{Rb}_f(\mathcal{R}^r(P))$ is a $(2r \cdot \mathrm{Lip}_f)$-approximation of $\mathsf{Rb}_f(\mathsf{M})$, and $\mathsf{Rb}_f(\mathcal{R}^r(P))$ can be computed in $O(n \log n)$ expected time [20], where n is the size of the 2-skeleton of $\mathcal{R}^r(P)$.*

---

[3]Such maximal interval can be open. We assume it is closed for simplicity. The case when it is open can be handled similarly.

[4]In fact, we can also require that there is a map from $\mathbf{Z}_1(\mathsf{Rb}_f(A)) \to \mathbf{Z}_1(\mathsf{Rb}_f(B))$ that induces an isomorphism from $\mathsf{H}_1(\mathsf{Rb}_f(A)) \to \mathsf{H}_1(\mathsf{Rb}_g(B))$ where every cycle from $\mathsf{Rb}_f(A)$ is mapped to a cycle in $\mathsf{Rb}_g(B)$ whose range is δ-Hausdorff-close.

Remark 1. Here we provide a brief discussion of why we focus only on the 1-st homology information of the Reeb graph, as well as the intuition behind our definition of a δ-approximate Reeb graph.

The Reeb graph is an abstract graph and contains only the 0- and 1-dimensional topological information. Given a Reeb graph $\mathsf{Rb}_f(\mathsf{M})$, its 0-th homology simply encodes the connected components information of M, and can be approximated from point data easily by returning the number of connected components in an appropriately constructed Rips complex in linear time.

At the same time, compared to general abstract graphs, the Reeb graph has the extra information of the natural function f defined on it. Hence one may also ask what the 0-th *persistent* homology of $\mathsf{Rb}_f(\mathsf{X})$ induced by f is. This turns out to be the same as approximating the 0-th persistent homology for X and can be solved using results from [6, 7].

Therefore, the only remaining issue is to approximate the 1-st homology of a Reeb graph. Similar to the case for the 0-th homology, there are two aspects: (i) computing $\mathsf{H}_1(\mathsf{Rb}_f(\mathsf{M}))$ itself; and (ii) computing the 1-st *persistent* homology of $\mathsf{Rb}_f(\mathsf{M})$ induced by the function f. For (i), our result shows that $\mathsf{H}_1(\mathsf{Rb}_f(\mathcal{R}^r(P)))$ for a certain Rips complex $\mathcal{R}^r(P)$ constructed from the point samples P is isomorphic to $\mathsf{H}_1(\mathsf{Rb}_f(\mathsf{M}))$. For (ii), since every 1-cycle in a Reeb graph is essential, the standard persistence is not able to describe them, and one has to use the extended persistence as introduced in [9], which is determined by the range of essential cycles. Hence our definition of the approximation also requires that ranges of corresponding homology classes (and even cycles) are also close.

Remark 2. One can strengthen Theorem 4.7 slightly to show that if the parameter r does not satisfy the conditions that $r < \frac{\mathbf{t}^*}{2\mathrm{Lip}_f}$ or $r < \frac{\mathbf{s}^*}{2\mathrm{Lip}_f}$, then all homology classes of $\mathsf{H}_1(\mathsf{Rb}_f(\mathsf{M}))$ with height at least $2r \cdot \mathrm{Lip}_f$ are preserved in $\mathsf{H}_1(\mathsf{Rb}_f(\mathcal{R}^r(P)))$ (and vice versa).

Computing $\beta_1(\mathsf{M})$ for orientable 2-manifolds. It was shown in [11] that for a Morse function $f : \mathsf{M} \to \mathbb{R}$ defined on a compact orientable surface M without boundary, one has $\mathrm{rank}(\mathsf{H}_1(\mathsf{M})) = 2 \cdot \mathrm{rank}(\mathsf{H}_1(\mathsf{Rb}_f(\mathsf{M})))$. Hence intuitively, using Theorem 4.7, we can compute $\beta_1(\mathsf{M}) = \mathrm{rank}(\mathsf{H}_1(\mathsf{M}))$ by $2 \cdot \mathrm{rank}(\mathsf{H}_1(\mathsf{Rb}_f(\mathcal{R}^r(P)))$ from an appropriate f and a Rips Complex $\mathcal{R}^r(P)$ constructed from a point sample P of M. Specifically, choose a function $f : \mathsf{M} \to \mathbb{R}$ so that we can evaluate it at points in P. For example, pick a base point $\mathbf{v} \in P$ and define a function $f_{\mathbf{v}}(x)$ to be the Euclidean distance from $x \in \mathsf{M}$ to the base point $\mathbf{v}$. Observe that the Lipschitz constant of this function $f_{\mathbf{v}}$ is at most 1. Our algorithm simply computes the Reeb graph $\mathsf{Rb}_{f_{\mathbf{v}}}(\mathcal{R}^r(P))$ and returns $2 \cdot \mathrm{rank}(\mathsf{H}_1(\mathsf{Rb}_{f_{\mathbf{v}}}(\mathcal{R}^r(P))))$.

Corollary 4.8 *Let M be an orientable smooth compact 2-manifold M without boundary and P an ε-sample of M. The above algorithm computes $\beta_1(\mathsf{M}))$ in $O(n \log n)$ expected time if $\mathbf{t}^*(\mathsf{M})$ and $\mathbf{s}^*(\mathsf{M})$ are positive for the chosen function f, and the parameters satisfy $4\varepsilon \le r < \min\{\frac{1}{4}\rho(\mathsf{M}), \frac{1}{4}\rho_c(\mathsf{M}), \frac{\mathbf{t}^*}{2\mathrm{Lip}_f}, \frac{\mathbf{s}^*}{2\mathrm{Lip}_f}\}.$*

Observe that a Morse function on an orientable 2-manifold provides positive $\mathbf{t}^*$ and $\mathbf{s}^*$. We remark that our algorithm produces a correct answer only under good choices of f and r; while previously, the best algorithm to estimate $\beta_1(\mathsf{M})$ only depends on choosing r small enough. The advantage of our algorithm is its efficiency, as the previous algorithm needs to compute the first-betti number of the simplicial complex $\mathcal{R}^r(P)$ for certain r, which takes $O(n^3)$ time no matter what the intrinsic dimension of M is, where n is the size of the 2-skeleton of $\mathcal{R}^r(P)$.

5 Persistent Reeb graph

Imagine that we have a set of points P sampled from a hidden space X, and $f : \mathsf{X} \to \mathbb{R}$ a function whose values at points in P are available. We wish to study this function f through its Reeb graph. A natural

approach to approximate X from P is to construct a Rips complex $\mathcal{R}^r(P)$ from P. Since it is often unclear what the right value of r should be, it is desirable to compute a series of Reeb graphs from Rips complexes constructed with various r, and then find out which cycles in the Reeb graph persist. This calls for computing *persistent homology groups* for the sequence of Reeb graphs.

Let $K_1 \subseteq K_2 \subseteq \cdots \subseteq K_n$ be a filtration of a simplicial complex K_n. A piecewise linear function $f : |K_n| \to \mathbb{R}$ provides a PL-function for every K_i, $i \in [1, n]$. Let $\mathbf{R}_i := \mathsf{Rb}_f(K_i)$ denote the Reeb graph of f defined on the geometric realization $|K_i|$ of K_i. Below we first show that there is a sequence of homomorphisms $\mathsf{H}_1(\mathbf{R}_i) \to \mathsf{H}_1(\mathbf{R}_{i+1})$ induced by the inclusions $K_i \subset K_{i+1}$. We then present an algorithm to compute the persistent homologies induced by these homomorphisms.

5.1 Persistent Reeb graph homology

Let Φ_i denote the associated quotient map from $|K_i| \to \mathbf{R}_i$, for any $i \in [1, n]$. Since the canonical inclusion $|K_i| \hookrightarrow |K_j|$ respects the the equivalence relation that defines the quotient space $\mathbf{R}_i$, the maps Φ_is, along with inclusions between K_js, induce a well-defined continuous map between the quotient spaces $\xi : \mathbf{R}_i \to \mathbf{R}_j$, for any $i < j$. Let ι_i denote the inclusion map from $|K_i|$ to $|K_{i+1}|$, and ξ_i the induced map from $\mathbf{R}_i$ to $\mathbf{R}_{i+1}$. We have the following diagram that commutes.

$$
\begin{array}{ccccccc}
|K_1| & \xrightarrow{\iota_1} & |K_2| & \xrightarrow{\iota_2} & \cdots\cdots & \xrightarrow{\iota_{n-1}} & |K_n| \\
\downarrow{\Phi_1} & & \downarrow{\Phi_2} & & & & \downarrow{\Phi_n} \\
\mathbf{R}_1 & \xrightarrow{\xi_1} & \mathbf{R}_2 & \xrightarrow{\xi_2} & \cdots\cdots & \xrightarrow{\xi_{n-1}} & \mathbf{R}_n
\end{array}
$$

The sequence of continuous maps ξ_i induces the following sequence of homomorphisms:

$$
\mathsf{H}_1(\mathbf{R}_1) \xrightarrow{\xi_{1*}} \mathsf{H}_1(\mathbf{R}_2) \xrightarrow{\xi_{2*}} \quad \cdots\cdots \xrightarrow{\xi_{(n-1)*}} \mathsf{H}_1(\mathbf{R}_n)
$$

Following [17], we can now define the *persistent homology groups* as the images of maps $\xi_*^{i,j} = \xi_{j*} \circ \cdots \circ \xi_{i*} : \mathsf{H}_1(\mathbf{R}_i) \to \mathsf{H}_1(\mathbf{R}_j)$. In other words, the image $\mathrm{Im}(\xi_*^{i,j})$ consists of homology classes from $\mathsf{H}_1(\mathbf{R}_j)$ that also have pre-images in $\mathsf{H}_1(\mathbf{R}_i)$ (i.e, persist from $\mathsf{H}_1(\mathbf{R}_i)$ to $\mathsf{H}_1(\mathbf{R}_j)$). The *persistent betti numbers* $\beta^{i,j}$ is defined as the rank of the persistent homology group $\mathrm{Im}(\xi_*^{i,j})$. Set

$$
\mu^{i,j} := \beta^{i-1,j} - \beta^{i,j} + \beta^{i,j-1} - \beta^{i-1,j-1}.
$$

Intuitively, $\mu^{i,j}$ is the number of independent loops created upon entering $\mathbf{R}_i$ and destroyed upon leaving $\mathbf{R}_j$. A *persistence pair* (i, j) is recorded if $\mu^{i,j} > 0$, and the value $\mu^{i,j}$ indicates the multiplicity of this pairing.

We focus on persistent H_1-homology for $\mathbf{R}_i$s in this paper. The persistent H_0-homology for $\mathbf{R}_i$s is the same as persistent H_0-homology for K_is, and thus can be easily computed by a union-find data structure in near linear time. We also remark that by Theorem 3.3, persistent H_1-homology for $\mathbf{R}_i$ is isomorphic to persistent *vertical* homology $\check{\mathsf{H}}_1(|K_i|)$ [5].

5.2 Computation

We now present an algorithm to compute the persistent betti number $\beta^{i,j}$. The numbers $\mu^{i,j}$ and the persistence pairs can be computed easily once we have these persistence betti numbers.

---

[5] Apriori, it is not clear how to compute the persistent one dimensional vertical homology. By maintaining the *extended persistence pairings* dynamically as we change from K_i to K_{i+1}, we can maintain the rank of each $\check{\mathsf{H}}_1(K_i)$, but not the persistent homology between them.

Given a filtration $K_1 \subseteq \cdots \subseteq K_n$, assume $K_{i+1} \setminus K_i$ is one simplex. Since the Reeb graph is completely decided by the 2-skeleton of a simplicial complex, we assume that K_is are 2-complexes. Let n_v, n_e and n_t denote the number of vertices, edges and triangles in K_n, and $n = n_v + n_e + n_t$. Observe that the complexity of each Reeb graph $\mathbf{R}_i$, for $i \in [1, n]$, is bounded by $O(n_e)$. The set of Reeb graphs $\mathbf{R}_i$s can be computed in $O(nn_v)$ time using the incremental algorithm from [25]. We use this algorithm as it can also maintain the image of each edge from K_i in $\mathbf{R}_i$ in $O(n_v)$ time at each incremental step, thus providing Φ_i, for $i \in [1, n]$.

Recall that a set of base cycles for $\mathsf{H}_p(\cdot)$ is a set of cycles whose classes form a basis of $\mathsf{H}_p(\cdot)$. For the sake of exposition in this section, we abuse the notation slightly and use a cycle to also refer to its carrier in the Reeb graph. Specifically, we will see later that our algorithm in fact maintains the carriers of a set of base cycles for $\mathsf{H}_1(\mathbf{R}_i)$, which we also call a *cycle-basis*. We say that a set of cycles are *independent* if the set of homology classes these cycles represent are independent.

To compute $\beta^{i,j}$, one can construct a set of base cycles $\{\alpha_1, \ldots, \alpha_r\}$ for $\mathsf{H}_1(\mathbf{R}_i)$ with $r = \mathrm{rank}(\mathsf{H}_1(\mathbf{R}_i))$, and check how many of their images in $\mathbf{R}_j$ remain independent. A straightforward implementation of this approach takes $O(n^2 n_e^3)$ time. Indeed, $r = O(n_e)$ and the complexity of each cycle α_i is bounded by $O(n_v)$ (by representing them as a sequence of vertices). Computing the images of all α_is takes $O(rn_v^2) = O(n_e n_v^2)$ time using the incremental algorithm from [25], and the independence test for these r cycles takes $O(rn_e^2) = O(n_e^3)$ time. Finally, there are n^2 pairs of i and j that we need to test, giving rise to $O(n^2 n_e^3)$ total time complexity. To improve the time complexity, we follow the idea of the standard persistence algorithm [19] and perform only one scan of the sequence of Reeb graphs, while maintaining a set of base cycles at any moment during the course.

Notice that the standard persistence algorithm cannot be directly applied to the sequence of Reeb graphs as there are no inclusions among them. In fact, the underlying spaces of two consecutive Reeb graphs can change dramatically. See Figure 5 for such an example. We also remark that there may not be an inclusion relation between $\mathbf{R}_i$ and $\mathbf{R}_{i+1}$ in either direction, that is, $\mathbf{R}_i \nsubseteq \mathbf{R}_{i+1}$ and $\mathbf{R}_i \nsupseteq \mathbf{R}_{i+1}$: see Case 3 discussed later. Hence while it is possible to model the persistent Reeb graph homology via zigzag persistence theory [4], the efficient algorithm to compute zigzag persistence as developed in [5] cannot yet be applied here.
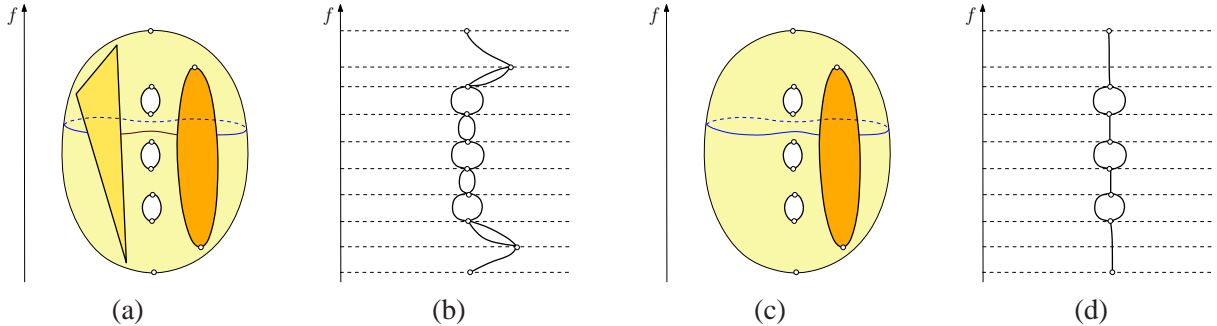

Figure 5: (a) shows a genus-g torus with the two caps missing; $g = 3$ in this case. Darker color regions indicate the two holes (missing caps) on this torus. Its Reeb graph w.r.t. the height function is shown in (b). Now if we fill the left triangle, as shown in (c), then $\Theta(g)$ number of independent vertical homological classes become horizontal, thus killing $\Theta(g)$ number of loops in the Reeb graph, which is shown in (d). In other words, by adding just one simplex (a triangle), the first betti number decreases by $\Theta(g)$.

Consistent base cycles. From now on, let $\mathbf{G}^{(i)}$ denote the cycle-basis of $\mathsf{H}_1(\mathbf{R}_i)$ that we maintain at the i-th step. For each cycle $\gamma \in \mathbf{G}^{(i)}$, we associate with it a *birth-time* $\mathbf{t}(\gamma)$, which is the earliest time (index) $k \leq i$ such that some pre-image of the homology class $[\gamma]$ under the map $\xi_*^{k,i} : \mathsf{H}_1(\mathbf{R}_k) \to \mathsf{H}_1(\mathbf{R}_i)$ exists.

15

In order to extract $\beta^{i,j}$, we wish to maintain the following *consistency condition* between $\mathbf{G}^{(i)}$ and $\mathbf{G}^{(j)}$: let $\mathbf{G}^{(i)} = \{\alpha_1^{(i)}, \alpha_2^{(i)}, \ldots, \alpha_r^{(i)}\}$ and $\mathbf{G}^{(j)} = \{\alpha_1^{(j)}, \ldots, \alpha_s^{(j)}\}$. Consider the set $\widehat{G}$ of images of cycles $\{\alpha_l^{(i)}\}$ in $\mathbf{R}_j$. $\mathbf{G}^{(i)}$ and $\mathbf{G}^{(j)}$ are *consistent* if the cardinality of $\widehat{G} \cap \mathbf{G}^{(j)}$ is *exactly* $\beta^{i,j}$. Notice that there are always $\beta^{i,j}$ number of independent cycles in $\widehat{G}$. However, its intersection with $\mathbf{G}^{(j)}$ may have much smaller cardinality. A sequence of cycle-bases $\{\mathbf{G}^{(i)} \mid i \in [1, n]\}$ is *consistent* if the consistency condition holds for any pair $\mathbf{G}^{(i)}$ and $\mathbf{G}^{(j)}$, $0 \leq i < j \leq n$. The following claim implies that we can read off $\beta^{i,j}$ easily from a consistent sequence of cycle-bases.

Lemma 5.1 *If a sequence of cycle-bases $\{\mathbf{G}^{(i)} \mid i \in [1, n]\}$ is consistent, then for any $1 \leq i < j \leq n$, $\beta^{i,j}$ equals the number of cycles in $\mathbf{G}^{(j)}$ whose birth-time is smaller than or equal to i.*

Proof: Consider a pair of indices $i < j$ and the corresponding cycle-basis $\mathbf{G}^{(i)}$ for $\mathsf{H}_1(\mathbf{R}_i)$ and $\mathbf{G}^{(j)}$ for $\mathsf{H}_1(\mathbf{R}_j)$. Assume that there are k cycles in $\mathbf{G}^{(j)}$ with birth-time smaller than or equal to i. Since all these cycles are independent in $\mathbf{R}_j$ (and thus in $\xi^{i,j}(\mathbf{R}_i)$), we have that $k \leq \beta^{i,j}$. On the other hand, since $\mathbf{G}^{(i)}$ and $\mathbf{G}^{(j)}$ are consistent, we have that $k \geq \beta^{i,j}$, implying that $k = \beta^{i,j}$. ∎

Algorithm description. In light of Lemma 5.1, our goal is to maintain consistent cycle-bases at any moment. We now describe how we update the set of base cycles as we move from K_k to $K_{k+1} = K_k \cup \{\sigma\}$; σ can be a 0-, 1-, or 2-simplex. Set $g_i := \operatorname{rank}(\mathsf{H}_1(\mathbf{R}_i))$ for any $i \in [1, n]$. Assume at k-th step we already have consistent $\{\mathbf{G}^{(i)} \mid i \in [1, k]\}$. For each cycle-basis $\mathbf{G}^{(i)}$, we also maintain the birth-time of each cycle in it. Assume cycles in $\mathbf{G}^{(k)} = \{\gamma_1, \ldots, \gamma_{g_k}\}$ are sorted by their birth-times. At the beginning of the k-th step, we first use the incremental algorithm from [25] to compute the Reeb graph $\mathbf{R}_{k+1}$ from $\mathbf{R}_k$. We next need to update $\mathbf{G}^{(k)}$ to $\mathbf{G}^{(k+1)}$ for $\mathbf{R}_{k+1}$ so that $\mathbf{G}^{(k+1)}$ is consistent with each $\mathbf{G}^{(i)}$ for $i \in [1, k]$. There are three cases.

Case 1: σ is a vertex. A new connected component is created in K_{k+1}, consisting of only σ. Similarly, a new node is created in $\mathbf{R}_{k+1}$. The set of base 1-cycles are not affected, and $\mathbf{G}^{(k+1)} = \mathbf{G}^{(k)}$.

Case 2: $\sigma = pq$ is an edge. Let $\widehat{p} = \Phi_k(p)$ and $\widehat{q} = \Phi_k(q)$ be the images of endpoints p and q of σ in the Reeb graph $\mathbf{R}_k$. Adding σ to K_k creates a new edge $e = \widehat{p}\widehat{q}$ in $\mathbf{R}_{k+1}$. If $\widehat{p}$ and $\widehat{q}$ are not in the same connected component in $\mathbf{R}_k$, then adding e will only reduce the rank of $\mathsf{H}_0(\mathbf{R}_k)$ by 1 and does not affect $\mathsf{H}_1(\mathbf{R}_k)$. In that case $\mathbf{G}^{(k+1)} = \mathbf{G}^{(k)}$. Otherwise, $\widehat{p}$ and $\widehat{q}$ are already connected in $\mathbf{R}_k$. Adding e results in $\operatorname{rank}(\mathsf{H}_1(\mathbf{R}_{k+1})) = \operatorname{rank}(\mathsf{H}_1(\mathbf{R}_k)) + 1$. Let γ be any cycle in $\mathbf{R}_{k+1}$ that contains e (which can be computed easily in linear time). All previous base cycles in $\mathbf{G}^{(k)}$ will remain independent in $\mathbf{R}_{k+1}$, and we simply set $\mathbf{G}^{(k+1)} = \mathbf{G}^{(k)} \cup \{\gamma\}$. The birth-time for γ is $k + 1$.

Case 3: σ is a triangle. The first two cases are simple and similar to the cases of standard persistence algorithm. Case 3 is much more complicated. In particular, unlike the standard persistence algorithm wherein adding a triangle may reduce β_1 by at most 1, the rank of $\mathsf{H}_1(\mathbf{R}_k)$ may decrease by $\Theta(g_k)$. What happens is that even though $\beta_1(K_k)$ is reduced by at most 1, arbitrary number of vertical homology classes can be converted into horizontal homology classes. An example is given in Figure 5.

Let $\sigma = \triangle pqr$, and let $\widehat{p} = \Phi_k(p)$, $\widehat{q} = \Phi_k(q)$ and $\widehat{r} = \Phi_k(r)$ be the images of the three endpoints of σ in $\mathbf{R}_k$, respectively. Assume without loss of generality that $f(p) \leq f(q) \leq f(r)$, and set $e_1 = pq$, $e_2 = qr$ and $e_3 = pr$. First, we compute the image of each e_i in $\mathbf{R}_k$, which is necessarily a monotone path (i.e, monotonic in function values) denoted by $\pi_i = \Phi_k(e_i)$. These images can be computed in $O(n_v)$ time using the incremental algorithm and the data structure of [25]. By our assumption of $f(p) \leq f(q) \leq f(r)$, π_1 and π_2 are disjoint in their interiors, while π_3 may share subcurves with π_1 and π_2. Set $\pi_{1,2} := \pi_1 \circ \pi_2$

to be the concatenation of π_1 and π_2, which is still a monotone path, and note $\pi_{1,2}$ and π_3 share the same two endpoints.

Now if $\pi_{1,2}$ and π_3 coincide in $\mathbf{R}_k$, the addition of triangle σ does not ensue any change, that is, $\mathbf{R}_{k+1} = \mathbf{R}_k$ and $\mathbf{G}^{(k+1)} = \mathbf{G}^{(k)}$. In this case, the vertical homology of K_k remains the same; either σ destroys a horizontal homology class in $\mathsf{H}_1(K_k)$, or it creates a 2-cycle.

Otherwise, the H_1-homology of the Reeb graph changes. Assume the two monotone paths $\pi_{1,2}$ and π_3 form s simple loops between them (see the right figure where $s = 3$). Then, with the addition of σ, each point in π_3 is mapped to the corresponding point in $\pi_{1,2}$ with the same function value. Hence this process collapses all these s independent loops and we have $\mathsf{g}_{k+1} = \mathsf{g}_k - s$.
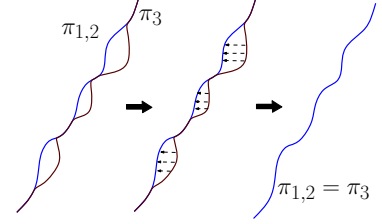
We now describe how to compute $\mathbf{G}^{(k+1)}$ for this case. First, we need to compute the image $\widehat{G} := \xi_k(\mathbf{G}^{(k)})$ of the set of base cycles $\mathbf{G}^{(k)}$ in $\mathbf{R}_{k+1}$. To do this, we need the map ξ_k. Observe that ξ_k maps each edge in $\mathbf{R}_k$ either to the same edge in $\mathbf{R}_{k+1}$, or to a monotone path in $\mathbf{R}_{k+1}$. The latter case can potentially happen only for edges in the paths $\pi_{1,2}$ and π_3 — in particular, for those edges in subcurves from $\pi_{1,2}$ and π_3 that are merged together. Since both $\pi_{1,2}$ and π_3 are monotone, images of edges from $\pi_{1,2}$ and π_3 can be computed in $O(|\pi_{1,2}| + |\pi_3|) = O(n_v)$ time by merging the sorted lists of vertices in $\pi_{1,2}$ and π_3. Hence we can compute the map ξ_k in $O(n_v)$ time.

Once ξ_k is computed, given a simple cycle γ from $\mathbf{R}_k$, we can compute its image in $\mathbf{R}_{k+1}$ in $O(n_v)$ time. This is because (i) there are $O(n_v)$ number of edges in γ; and (ii) the total size of the images of edges from γ in $\mathbf{R}_{k+1}$ has an upper bound $|\gamma| + |\xi_k(\pi_{1,2})| + |\xi_k(\pi_3)| = O(n_v)$. The set of cycles $\widehat{G} := \xi_k(\mathbf{G}^{(k)})$ in $\mathbf{R}_{k+1}$ can then be computed in $O(n_v \mathsf{g}_k)$ time. Let $\widehat{G} = \{\widehat{\gamma}_1, \ldots, \widehat{\gamma}_{\mathsf{g}_k}\}$.

The remaining task is to construct $\mathbf{G}^{(k+1)}$ that is consistent with $\mathbf{G}^{(i)}$ for any $i \le k$. One needs $\mathsf{g}_{k+1} = \mathrm{rank}(\xi_*^{k,k+1})$ independent cycles from $\widehat{G}$ to make $\mathbf{G}^{(k+1)}$ consistent with $\mathbf{G}^{(k)}$. To this end, we perform the following two steps.

(S1). We represent each cycle in $\widehat{G}$ as a linear combination of cycles in a basis for the graph $\mathbf{R}_{k+1}$.

(S2). We check the dependency of cycles in $\widehat{G}$ in order of their birth-times, and remove redundant cycles to obtain $\mathbf{G}^{(k+1)}$.

Step (S1). Since $\mathbf{R}_{k+1}$ is a graph, we compute a canonical basis of cycles, $B = \{\alpha_1, \ldots, \alpha_{\mathsf{g}_{k+1}}\}$, in the following standard way. Construct an arbitrary spanning tree T of $\mathbf{R}_{k+1}$. Let $E = \{e_1, \ldots, e_{\mathsf{g}_{k+1}}\}$ denote the set of non-tree edges in $\mathbf{R}_{k+1}$. Each edge $e_i = pq \in E$ creates a canonical cycle that concatenates edge e_i with the two unique paths in T from p and q to their common ancestor. We set α_i to be this canonical cycle created by e_i. Obviously, each e_i appears exactly once among all cycles in B. Given a cycle $\gamma \in \widehat{G}$, we need to find coefficients c_is such that $\gamma = \sum_{i=1}^{\mathsf{g}_{k+1}} c_i \alpha_i$, where each c_i is either 0 or 1. Since e_i appears only in α_i, we have c_i equal the number of times e_i appears in γ modulo 2. Since γ is a simple curve, c_i is 1 if $e_i \in \gamma$ and 0 otherwise. Hence all c_is for $i \in [1, \mathsf{g}_{k+1}]$ can be computed in $O(n_v)$ time for one curve γ. Computing the coefficients of all cycles in $\widehat{G}$ takes $O(n_v \mathsf{g}_k)$ time.

Step (S2). Recall that cycles in $\mathbf{G}^{(k)} = \{\gamma_1, \ldots, \gamma_{\mathsf{g}_k}\}$ are sorted by increasing order of their birth-times. Note, the birth-time of the cycle $\widehat{\gamma}_i \in \widehat{G}$, which is the image of the cycle $\gamma_i \in \mathbf{G}^{(k)}$ in $\mathbf{R}_{k+1}$, may be smaller than the birth-time of γ_i. Now represent cycles in $\widehat{G}$ with respect to the canonical basis $B = \{\alpha_1, \ldots, \alpha_{\mathsf{g}_{k+1}}\}$ in a matrix M, where the ith column of M, denoted by $\mathrm{col}_M[i]$, contains the coordinates of $\widehat{\gamma}_i$ under basis B; that is, $\widehat{\gamma}_i = \sum_{j=1}^{g_{k+1}} \mathrm{col}_M[i][j]\alpha_j$. Obviously, the matrix M has size $\mathsf{g}_k \times \mathsf{g}_{k+1}$.

Next, we perform a left-to-right reduction of matrix M, which is the same as the reduction of the adjacency matrix used in the standard persistence algorithm [10, 19]. In particular, the only operation that

17

one can use is to add a column to another one on its right. For a column $\text{col}_M[i]$, let its *low-row index* denote the largest index j such that $\text{col}_M[i][j] = 1$. At the end of the reduction, each column is either empty or has a unique low-row index; that is, no other column can have the same low-row index as this one. We set $\mathbf{G}^{(k+1)}$ as the subset of $\widehat{G}$ whose corresponding columns in the reduced matrix M' is not all zeros. The reduction takes time $O(g_{k+1}g_k^2)$. Intuitively, the consistency of $\mathbf{G}^{(k+1)}$ with each $\mathbf{G}^{(i)}$ for $i \in [1, k]$ follows from the left-to-right reduction. It guarantees that if a set of cycles in $\widehat{G}$ are dependent, then only those created earlier (i.e, with smaller birth-time) will be kept.

Lemma 5.2 $\mathbf{G}^{(k+1)}$ *as constructed above provides a basis of* $\mathsf{H}_1(\mathbf{R}_{k+1})$. *Furthermore, if* $\{\mathbf{G}^{(1)}, \ldots, \mathbf{G}^{(k)}\}$ *is consistent, so is* $\{\mathbf{G}^{(1)}, \ldots, \mathbf{G}^{(k+1)}\}$.

Proof: Let M' denote the reduced matrix of M. Recall that $\widehat{G} = \{\widehat{\gamma}_1, \ldots, \widehat{\gamma}_{g_k}\}$ contains the images of cycles from $\mathbf{G}^{(k)}$ in $\mathbf{R}_{k+1}$. Set $\widehat{G}_i = \{\widehat{\gamma}_1, \ldots, \widehat{\gamma}_i\}$, and let G'_i be the set of cycles from $\widehat{G}_i$ whose corresponding column in the reduced matrix M' is non-empty (i.e, not all zeros). In other words, $G'_i = \widehat{G}_i \cap \mathbf{G}^{(k+1)}$ is the intersection between $\widehat{G}_i$ and the set $\mathbf{G}^{(k+1)}$ constructed by our algorithm. By induction on i, it is easy to show that for any $i \in [1, g_k]$, cycles in G'_i generate the same subgroup of $\mathsf{H}_1(\mathbf{R}_{k+1})$ as $\widehat{G}_i$. It then follows that, in the end, cycles in $\mathbf{G}^{(k+1)} = G'_{g_k}$ are all independent in $\mathbf{R}_{k+1}$ and $|\mathbf{G}^{(k+1)}|$ equals the rank of the homology group generated by cycles in $\widehat{G}$, which is $\beta^{k,k+1} = g_{k+1}$. This proves the first part of the claim.

For the second part of the claim, first note that $\mathbf{G}^{(k+1)}$ is consistent with $\mathbf{G}^{(k)}$ as $\widehat{G} \cap \mathbf{G}^{(k+1)} = \mathbf{G}^{(k+1)}$ and has cardinality g_{k+1}. Now consider an arbitrary $\mathbf{G}^{(i)}$ with $i < k$. Since $\{\mathbf{G}^{(1)}, \ldots, \mathbf{G}^{(k)}\}$ are consistent, and cycles $\{\gamma_1, \ldots, \gamma_{g_k}\}$ in $\mathbf{G}^{(k)}$ are sorted by their birth-times, it follows from Lemma 5.1 that the first $s = \beta^{i,k}$ number of cycles $G_s = \{\gamma_1, \ldots, \gamma_s\}$ from $\mathbf{G}^{(k)}$ are images of cycles from $\mathbf{G}^{(i)}$. Hence the image of cycles from $\mathbf{G}^{(i)}$ in $\mathbf{R}_{k+1}$ are exactly the cycles in $\widehat{G}_s$, and classes of cycles in $\widehat{G}_s$ generate the persistent homology group $\xi_*^{i,k+1}(\mathsf{H}_1(\text{Rb}_f(K_i)))$. On the other hand, as mentioned above, classes of cycles in $G'_s = \widehat{G}_s \cap \mathbf{G}^{(k+1)}$ generate the same subgroup of $\mathsf{H}_1(\mathbf{R}_{k+1})$ as $\widehat{G}_s$. Since cycles in G'_s are independent, G'_s has rank $\beta^{i,k+1}$, implying that $\mathbf{G}^{(k+1)}$ is consistent with $\mathbf{G}^{(i)}$, for any $i \in [1, k]$. The second part of the claim then follows. ∎

Finally, for our algorithm to continue into the next iteration, we also need to maintain the birth-times for each cycle in $\mathbf{G}^{(k+1)}$. This is achieved by the following claim.

Claim 5.3 *Let* $\mathbf{G}^{(k+1)} = \{\widehat{\gamma}_{I_1}, \ldots \widehat{\gamma}_{I_{g_{k+1}}}\}$, *where* I_is *are the set of indices of non-zero columns in the reduced matrix* M'. *Then the birth-time of* $\widehat{\gamma}_{I_i}$ *equals to the birth-time of* γ_{I_i} *for any* $i \in [1, g_{k+1}]$.

Proof: Recall that $\mathbf{G}^{(k+1)}$ contains the set of cycles $\widehat{\gamma}_{I_i}$ where $\{I_i\}$ is the set of indices of non-zero columns from the reduced matrix M'. Given a cycle $\alpha \in \mathbf{G}^{(i)}$, let $birthtime(\alpha)$ denote the birth-time of α. Assume that one of the cycles, say $\widehat{\gamma}_m \in \mathbf{G}^{(k+1)}$, has a birth-time that is different from that of $\gamma_m \in \mathbf{G}^{(k)}$. Set $t := birthtime(\widehat{\gamma}_m)$. Since $\widehat{\gamma}_m = \xi^k(\gamma_m)$, we have $t \leq birthtime(\gamma_m)$. Since the two birth-times are different, t must be strictly smaller than the birth-time of γ_m.

Furthermore, there exists a cycle $\alpha \in \mathbf{R}_t$ such that its image $\alpha_1 := \xi^{t,k}(\alpha)$ in $\mathbf{R}_k$ is not homologous to γ_m, while its image $\alpha_2 := \xi^{t,k+1}(\alpha)$ in $\mathbf{R}_{k+1}$ is $\widehat{\gamma}_m$. On the other hand, α_1 can be uniquely written as a linear combination of a subset of cycles from $\mathbf{G}^{(k)}$, say $\alpha_1 = \gamma_{J_1} + \cdots + \gamma_{J_t}$. It is easy to verify that the birth-time of each γ_{J_i} is at most t. Since $t < birthtime(\gamma_m)$, it follows that all indices J_is are strictly smaller than m (as cycles in $\mathbf{G}^{(k)}$ are sorted by their birth-times). However, this is not possible since the resulting m-th column will be all zero at the time when we reduce the m-th column to construct $\mathbf{G}^{(k+1)}$ as $\widehat{\gamma}_m = \sum_i \widehat{\gamma}_{J_i}$. Hence the cycle $\widehat{\gamma}_m$ cannot be chosen as a base cycle in $\mathbf{G}^{(k+1)}$ reaching a contradiction. It follows that $t = birthtime(\gamma_m)$, or more generally, $birthtime(\widehat{\gamma}_{I_i}) = birthtime(\gamma_{I_i})$ for every index I_i of non-zero column in the reduced matrix M'. ∎

Putting everything together, we conclude with the following main result.

Theorem 5.4 *Given a filtration $K_1 \subset \cdots \subset K_n$ of a simplicial complex K_n with a piecewise linear function $f : K_n \to \mathbb{R}$, we can compute all persistent first betti numbers for the induced sequence of Reeb graphs $\mathsf{Rb}_f(K_i)s$ in $O(\sum_{i=1}^{n}(n_v \mathsf{g}_i + \mathsf{g}_i^3)) = O(nn_e^3)$ time, where n_v and n_e are the number of vertices and edges in K_n, respectively, n is the size of 2-skeleton of K_n, and g_i is the first betti number of the Reeb graph $\mathsf{Rb}_f(K_i)$.*

6 Conclusions and discussions

In this paper, we present a simple and efficient algorithm to approximate the Reeb graph $\mathsf{Rb}_f(\mathsf{M})$ of a map $f : \mathsf{M} \to \mathbb{R}$ from point data sampled from a smooth and compact manifold M. Given that Reeb graph is an abstract graph with a function defined on it, we only approximate its topology together with the range information for each loop in it. It will be interesting to see whether the Reeb graph we compute from the point data is also geometrically close to some specific embedding of the Reeb graph $\mathsf{Rb}_f(\mathsf{M})$ in the hidden domain M. To this end, our results in Section 4.1 on mappings between cycles can be useful.

We also study how to compute the "persistence" of loops in a Reeb graph by measuring their life time as the defining domain grows. An immediate question is to see whether the time complexity can be further improved to match that of the standard persistence algorithm in the worst case.

Finally, it will be interesting to explore whether one can leverage the simple structure and efficient computation of the Reeb graph to retrieve topological information for various spaces efficiently. For example, given a 3-manifold with a function f defined on it, its vertical H_1-homology is already encoded in the Reeb graph and can thus be computed in near-linear time. Can we retrieve the horizontal H_1-homology efficiently by tracking the levelsets of f, or by defining another function that is somewhat "orthogonal" to f?

References

[1] D. Attali, A. Lieutier, and D. Salinas. Vietoris-rips complexes also provide topologically correct reconstructions of sampled shapes. In *Proc. 27th Annu. Sympos. Comput. Geom.*, SoCG '11, pages 491–500, 2011.

[2] M. Bernstein, V. de Silva, J. Lanford, and J. Tenenbaum. Graph approximations to geodesics on embedded manifolds. Technical report, Dept. Psychology, Stanford University, USA, 2000.

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

[4] G. Carlsson and V. de Silva. Zigzag persistence. *Foundations of Computational Mathematics*, 10(4):367–405, 2010.

[5] G. Carlsson, V. de Silva, and D. Morozov. Zigzag persistent homology and real-valued functions. In *Proc. 25th Annu. Sympos. Comput. Geom.*, SoCG '09, pages 247–256, 2009.

[6] F. Chazal, L. J. Guibas, S. Oudot, and P. Skraba. Analysis of scalar fields over point cloud data. In *Proc. 20th ACM-SIAM Sympos. Discrete Algorithms*, pages 1021–1030, 2009.

[7] F. Chazal, L. J. Guibas, S. Y. Oudot, and P. Skraba. Persistence-based clustering in Riemannian manifolds. In *Proc. 27th Annu. Sympos. Comput. Geom.*, pages 97–106, 2011.

[8] F. Chazal and S. Oudot. Towards persistence-based reconstruction in Euclidean spaces. In *Proc. 24th ACM Sympos. on Comput. Geom.*, pages 232–241, 2008.

[9] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Extending persistence using Poincaré and Lefschetz duality. *Foundations of Computational Mathematics*, 9(1):79–103, 2009.

[10] D. Cohen-Steiner, H. Edelsbrunner, and D. Morozov. Vines and vineyards by updating persistence in linear time. In *Proc. 22nd Annu. Sympos. Comput. Geom.*, pages 119–126, 2006.

[11] K. Cole-McLaughlin, H. Edelsbrunner, J. Harer, V. Natarajan, and V. Pascucci. Loops in Reeb graphs of 2-manifolds. *Discrete Comput. Geom.*, 32(2):231–244, 2004.

[12] T. K. Dey. *Curve and Surface Reconstruction: Algorithms with Mathematical Analysis*. Cambridge U. Press, New York, NY, USA, 2007.

[13] T. K. Dey and K. Li. Cut locus and topology from surface point data. In *Proc. 25th Annu. Sympos. Comput. Geom.*, pages 125–134, 2009.

[14] T. K. Dey, J. Sun, and Y. Wang. Approximating cycles in a shortest basis of the first homology group from point data. *Inverse Problems*, 2012. To appear.

[15] H. Doraiswamy and V. Natarajan. Efficient output-sensitive construction of Reeb graphs. In *Proc. 19th Internat. Sym. Alg. and Comput.*, pages 556–567, 2008.

[16] H. Doraiswamy and V. Natarajan. Efficient algorithms for computing Reeb graphs. *Computational Geometry: Theory and Applications*, 42:606–616, 2009.

[17] H. Edelsbrunner and J. Harer. Persistent homology — a survey. In J. E. Goodman, J. Pach, and R. Pollack, editors, *Surveys on Discrete and Computational Geometry. Twenty Years Later*, pages 257–282. Amer. Math. Soc., Providence, Rhode Island, 2008. Contemporary Mathematics 453.

[18] H. Edelsbrunner, J. Harer, A. Mascarenhas, V. Pascucci, and J. Snoeyink. Time-varying Reeb graphs for continuous space-time data. *Comput. Geom.*, 41(3):149–166, 2008.

[19] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discrete Comput. Geom.*, 28:511–533, 2002.

[20] W. Harvey, R. Wenger, and Y. Wang. A randomized $O(m \log m)$ time algorithm for computing Reeb graph of arbitrary simplicial complexes. In *Proc. 26th Annu. Sympos. Compu. Geom.*, pages 267–276, 2010.

[21] A. Hatcher. *Algebraic Topology*. Cambridge U. Press, New York, 2002.

[22] J.-C. Hausmann. On the Vietoris-Rips complexes and a cohomology theory for metric spaces. In *Prospects in Topology: Proc. Conf. in Honour of William Browder*, Ann. Math. Stud., 138, pages 175–188. Princeton Univ. Press, 1995.

[23] J. R. Munkres. *Elements of Algebraic Topology*. Westview Press, 1996.

[24] P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom.*, 39:419–441, 2008.

[25] V. Pascucci, G. Scorzelli, P.-T. Bremer, and A. Mascarenhas. Robust on-line computation of Reeb graphs: simplicity and speed. *ACM Trans. Graph.*, 26(3):58, 2007.

[26] Y. Shinagawa and T. L. Kunii. Constructing a Reeb graph automatically from cross sections. *IEEE Comput. Graph. Appl.*, 11(6):44–51, 1991.

[27] J. Tierny, A. Gyulassy, E. Simon, and V. Pascucci. Loop surgery for volumetric meshes: Reeb graphs reduced to contour trees. *IEEE Trans. Vis. Comput. Graph.*, 15(6):1177–1184, 2009.

A   Proof for Lemma 4.3

Proof of Claim (i).   Lemma 4.2 implies that $h_\#$ induces an isomorphism $h_* : H_1(\mathcal{R}^r(P)) \to H_1(M)$. We now prove the second part of claim (i).

First, we show the following two results: (C-1) given a cycle α from $\mathbf{Z}_1(M)$, $[\alpha] = [h_\#(d(\alpha))]$; and (C-2) given a cycle $\hat{\alpha} \in \mathbf{Z}_1(\mathcal{R}^r(P))$, we have that $[\hat{\alpha}] = [d(h_\#(\hat{\alpha}))]$. Note that cycles in $\mathbf{Z}_1(\mathcal{R}^r(P))$ are simplicial cycles, while cycles in $\mathbf{Z}_1(M)$ are singular cycles.

These two results (C-1) and (C-2) imply that d maps homologous cycles from $\mathbf{Z}_1(M)$ to homologous cycles in $\mathbf{Z}_1(\mathcal{R}^r(P))$. Hence it indeed induces a homomorphism $d_* : H_1(M) \to H_1(\mathcal{R}^r(P))$. Furthermore, these two results mean that d_* is the inverse of the map h_*. Since h_* is an isomorphism, so is d_*. This proves Claim (i) of the lemma.

Proof of (C-1):

We show that given a cycle α from $\mathbf{Z}_1(M)$, $[\alpha] = [h_\#(d(\alpha))]$. Let $\gamma \subseteq M$ be the carrier of α. To map α to $d(\alpha)$, suppose that its carrier γ is broken into k pieces as described earlier using the Decomposition method. For the i-th piece with endpoints x_i and x_{i+1}, let p_i and p_{i+1} be their closest point in P, respectively; recall that $d(\alpha)$ is the concatenation of all edges $p_i p_{i+1}$ for $i \in [1, k]$.

Now consider the cycle $h_\#(d(\alpha))$ in M: its carrier γ' is the concatenation of $h(p_i p_{i+1}) \subset M$ for all $i \in [1, k]$. By Proposition 3.3 of [14], each curve $h(p_i p_{i+1})$ has endpoints p_i and p_{i+1}, and it is contained in the union of the two Euclidean balls of radius r centered at p_i and at p_{i+1}. Since p_i and p_{i+1} are within r Euclidean distance, $h(p_i p_{i+1})$ is contained in the Euclidean balls of radius $2r$ centered at p_i and at p_{i+1}. Notice that the geodesic distance and the Euclidean distance between two points $x, y \in M$ approximate each other when x and y are close enough (see e.g, Proposition 1.2 from [14]). It follows that when r is smaller than $\rho(M)/4$, $h(p_i p_{i+1})$ is contained in both geodesic balls of radius $3r$ centered at $p_i \in M$ and $p_{i+1} \in M$.

Let $\gamma[x_i, x_{i+1}]$ denote the subcurve of γ from x_i to x_{i+1}. Since the length of $\gamma[x_i, x_{i+1}]$ is less than r by construction, the curve $\gamma[x_i, x_{i+1}]$ is contained in the geodesic balls of radius r centered at x_i and x_{i+1}. This implies that the curve $\gamma[x_i, x_{i+1}]$ is contained in the *geodesic tubular neighborhood*

$$\mathrm{Tub}_r(\pi_g(x_i, x_{i+1})) := \{y \in M \mid d(y, \pi_g(x_i, x_{i+1})) \leq r\}$$

where $\pi_g(x, y)$ denote a minimizing geodesic between two points $x, y \in M$. By Proposition 3.7 of [14], $\mathrm{Tub}_r(\pi_g(x_i, x_{i+1}))$ is contractible and hence $\gamma[x_i, x_{i+1}]$ is homotopy equivalent to $\pi_g(x_i, x_{i+1})$.

On the other hand, due to the sampling condition, the geodesic distances between x_i and p_i, and between x_{i+1} and p_{i+1}, are both bounded by r. Combining this with the fact that $h(p_i p_{i+1})$ lies within the geodesic balls of radius $3r$ centered at both p_i and p_{i+1}, we have that any point in $h(p_i p_{i+1})$ is within geodesic distance $4r$ to both x_i and to x_{i+1}. Hence $h(p_i p_{i+1})$ lies within the geodesic tubular neighborhood $\mathrm{Tub}_{4r}(\pi_g(x_i, x_{i+1}))$. Again by Proposition 3.7 of [14], when $r \leq \rho_c(M)/4$, the curve $\pi_g(x_i, p_i) \circ h(p_i p_{i+1}) \circ \pi_g(p_{i+1}, x_{i+1})$ is homotopy equivalent to $\pi_g(x_i, x_{i+1})$ and thus homotopy equivalent to $\gamma[x_i, x_{i+1}]$. In fact, one can find a homotopy h_i that keeps p_i and p_{i+1} on the geodesics $\pi_g(x_i, p_i)$ and $\pi_g(x_{i+1}, p_{i+1})$

respectively so that two maps h_i and h_{i+1}, for $i \in [1, k-1]$, are consistent in mapping the common endpoints x_{i+1}. Therefore, we can combine all such h_i's to obtain a homotopy between γ' (which is the carrier of $h_\#(d(\alpha))$) and γ (which is the carrier of α). It follows that $[h_\# \circ d(\alpha)] = [\alpha]$.

Proof of (C-2):

We now show that given any (simplicial) cycle $\hat{\alpha} \in \mathbf{Z}_1(\mathcal{R}^r(P))$, we have that $[\hat{\alpha}] = [d \circ h_\#(\hat{\alpha})]$. First, consider the image $\gamma = h(\hat{\alpha})$ of $\hat{\alpha}$ in M; γ is the carrier of the cycle $h_\#(\hat{\alpha}) \in \mathbf{Z}_1(\mathsf{M})$. By construction, γ is the concatenation of $h(e)$s for every edge $e = pq$ in the simplicial cycle $\hat{\alpha}$. By Proposition 3.3 in [14], each curve $h(e)$ is contained inside $\mathsf{M} \cap (B_r(p) \cup B_r(q))$. Hence, for any point $x \in h(e)$, its geodesic distance to p and to q is bounded by $3r$.

Now consider mapping the cycle $h_\#(\hat{\alpha})$ carried by γ back to $\mathcal{R}^r(P)$ using the decomposition method described earlier. Consider the set of breaking points x_i's in the subcurve $h(e) \subset \gamma$ — assume for simplicity that the endpoints of $h(e)$, that is p and q, are also break points. Each break point x_i in $h(e)$ will be mapped to its nearest point $p_i \in P$ and the geodesic distance between x_i and p_i is at most r. Hence p_i is within $3r + r = 4r$ geodesic distance to both endpoints p and q of the edge $e \subset \hat{\alpha}$. This means that both $p_i p$ and $p_i q$ are edges in the Rips complex $\mathcal{R}^{4r}(P)$. Hence the concatenation of arcs $p_i p_{i+1}$ is homotopy equivalent to the edge e in the simplicial complex $\mathcal{R}^{4r}(P)$. Combining this homotopy equivalent map for every edge $e \in \hat{\alpha}$, we have that $d(h_\#(\hat{\alpha}))$ is homotopy equivalent, and thus homologous, to $\hat{\alpha}$ in $\mathcal{R}^{4r}(P)$. Finally, since the inclusion map from $\mathcal{R}^r(P)$ to $\mathcal{R}^{4r}(P)$ induces an isomorphism in the first homology groups when r is small, we have that $d(h_\#(\hat{\alpha}))$ is homologous to $\hat{\alpha}$ in $\mathcal{R}^r(P)$ as well. Thus $[\hat{\alpha}] = [d(h_\#(\hat{\alpha}))]$.

Proving Claim (ii). Claim (ii) follows easily from the constructions of d and h. In particular, consider a cycle $\alpha \in \mathbf{Z}_1(\mathsf{M})$ and $d(\alpha) \in \mathbf{Z}_1(\mathcal{R}^r(P))$. (The case for $\hat{\alpha}$ from $\mathbf{Z}_1(\mathcal{R}^r(P))$ and $h_\#(\hat{\alpha})$ from $\mathbf{Z}_1(\mathsf{M})$ can be similarly argued.) Let $\gamma \subset \mathsf{M}$ be the carrier of α. The Decomposition method breaks γ into k pieces $\gamma(x_i, x_{i+1})$s, for $i \in [0, k]$. Each piece $\gamma(x_i, x_{i+1})$ is mapped to the edge $p_i p_{i+1}$ where p_i is the closest point of x_i in P. Since P is an ε-sample of M, and since the length of $\gamma(x_i, x_{i+1})$ is at most $r - 2\varepsilon$, any point x in $\gamma(x_i, x_{i+1})$ is within r *geodesic distance* to the point $p_i \in \mathsf{M}$. Hence by the Lipschitz condition of f, we have $|f(x_i) - f(p_i)| \leq r \cdot \mathrm{Lip}_f$. It follows that

$$[\min_{i \in [0,k]} f(p_i) + r \cdot \mathrm{Lip}_f, \quad \max_{i \in [0,k]} f(p_i) - r \cdot \mathrm{Lip}_f] \subseteq \mathrm{range}(\gamma) = \mathrm{range}(\alpha)$$

$$\subseteq [\min_{i \in [0,k]} f(p_i) - r \cdot \mathrm{Lip}_f, \quad \max_{i \in [0,k]} f(p_i) + r \cdot \mathrm{Lip}_f].$$

On the other hand, note that under $\mathbb{Z}_2$ coefficient, $\mathrm{range}(d(\alpha)) \subseteq [\min_{i \in [0,k]} f(p_i), \max_{i \in [0,k]} f(p_i)]$ (and it can be much smaller than this interval). It then follows that $\mathrm{range}(d(\alpha))$ is oneside-$(r \cdot \mathrm{Lip}_f)$-close to $\mathrm{range}(\alpha)$.

Proof of Claim (iii). Consider any homology class $\omega \in \mathsf{H}_1(\mathsf{M})$. By Claim (iii) we have that the range of $d_*(\omega)$ is oneside-$(r \cdot \mathrm{Lip}_f)$-close to the range of ω: indeed, choose the thinnest cycle α of ω, we have $\mathrm{range}(d(\alpha))$ is oneside-$(r \cdot \mathrm{Lip}_f)$-close to $\mathrm{range}(\alpha) = \mathrm{range}(\omega)$. Since $\mathrm{range}(d_*(\alpha)) \subseteq \mathrm{range}(d(\gamma))$, $\mathrm{range}(d_*(\alpha))$ is also oneside-$(r \cdot \mathrm{Lip}_f)$-close to $\mathrm{range}(\omega)$.

Now map $d_*(\omega)$ back to $\mathsf{H}_1(\mathsf{M})$, we have that the range of $h_*(d_*(\omega))$ is also oneside-$(r \cdot \mathrm{Lip}_f)$-close to the range of $d_*(\omega)$ by Claim (iii). Since $h_*(d_*(\omega)) = \omega$, the ranges of ω and $d_*(\omega)$ are $(r \cdot \mathrm{Lip}_f)$-Hausdorff-close.

The statement for $\hat{\omega}$ and $h_*(\hat{\omega})$ can be argued similarly.

B   Existence of a set of 0-Height Basis for $\overline{\mathsf{H}}_1(\mathsf{M})$

Recall that $s = \mathrm{rank}(\overline{\mathsf{H}}_1(\mathsf{M}))$. We now show how to construct a basis $\{h_1, \ldots, h_s\}$ for the horizontal subgroup $\overline{\mathsf{H}}_1(\mathsf{M}) \subset \mathsf{H}_1(\mathsf{M})$ such that $\mathrm{height}(h_i) = 0$ for any $i \in [1, s]$.

Suppose we have already constructed a partial basis $H_{k-1} := \{h_1, \ldots, h_{k-1}\}$, each of which has 0-height. Our goal is to find a 0-height class h_k that is independent of elements in H_{k-1}. In particular, let h be any horizontal homology class that is independent of H_{k-1}. By definition, we can find a representative cycle $\gamma = \sum_{i=1}^{r} \gamma_i$ of h such that each γ_i is contained in a distinct level set of M. Take γ_1: if $[\gamma_1]$ is independent of H_{k-1}, then simply set $h_k = [\gamma_1]$. Otherwise, consider the cycle $\gamma' = \sum_{i=2}^{r}$ and obviously $[\gamma']$ is necessarily independent of classes in H_{k-1}. Now set h to be γ' and repeat the above process. Either we terminate when some γ_i which is independent of classes in H_{k-1}, in which case we set $h_k = [\gamma_i]$ which is of height zero. Or the process ends when h can be represented by a cycle γ which itself is contained in a level set. In this case, set $h_k = h$, which is again of height 0.

Perform the above procedure for s rounds. In the end, H_s gives the desired 0-height basis for $\overline{\mathsf{H}}_1(\mathsf{M})$. Note that by this construction, we also obtain a corresponding set of 0-height base cycles for $\overline{\mathsf{H}}_1(\mathsf{M})$.