Structure and Stability of the One-Dimensional Mapper

Mathieu Carriere, Steve Y. Oudot

Structure and Stability of the 1-Dimensional Mapper

Mathieu Carrière, Steve Oudot

Abstract

Given a continuous function $f : X \to \mathbb{R}$ and a cover $\mathcal{I}$ of its image by intervals, the Mapper is the nerve of a refinement of the pullback cover $f^{-1}(\mathcal{I})$. Despite its success in applications, little is known about the structure and stability of this construction from a theoretical point of view. As a pixelized version of the Reeb graph of f, it is expected to capture a subset of its features (branches, holes), depending on how the interval cover is positioned with respect to the critical values of the function. Its stability should also depend on this positioning. We propose a theoretical framework that relates the structure of the Mapper to the one of the Reeb graph, making it possible to predict which features will be present and which will be absent in the Mapper given the function and the cover, and for each feature, to quantify its degree of (in-)stability. Using this framework, we can derive guarantees on the structure of the Mapper, on its stability, and on its convergence to the Reeb graph as the granularity of the cover $\mathcal{I}$ goes to zero.

1 Introduction

Many data sets nowadays come in the form of point clouds with function values attached to the points. Such data may come either from direct measurements (e.g. think of a sensor field measuring some physical quantity like temperature or humidity), or as a byproduct of some data analysis pipeline (e.g. think of a word function in the quantization phase of the bag-of-words model). There is a need for summarizing such data and for uncovering their inherent structure, to enhance further processing steps and to ease interpretation.

One way of characterizing the structure of a scalar field $f : X \to \mathbb{R}$ is to look at the evolution of the topology of its *level sets*—i.e. sets of the form $f^{-1}(\{\alpha\})$, for α ranging over $\mathbb{R}$. This information is summarized in a mathematical object called the *Reeb graph* of the pair (X, f), denoted by $\mathrm{R}_f(X)$ and defined as the quotient space obtained by identifying the points of X that lie in the same connected component of the same level set of f [31]. The Reeb graph is known to be a graph (technically, a multi-graph) when X is a smooth manifold and f is a Morse function, or more generally when f is of *Morse type* (see Definition 2.1). Moreover, since the map f is constant over equivalence classes, there is a well-defined induced map on $\mathrm{R}_f(X)$.

The connection between the topology of the Reeb graph and the one of its originating pair (X, f) has been the object of much study in the past and is now well understood. It has gained increasing interest in the recent years, with the introduction of persistent homology [25] and its extended version [18]. Indeed, the *extended persistence diagram* of the induced map $\tilde{f}$ (see Section 2 for a formal definition) describes the structure of the Reeb graph in the following sense. It takes the form of a multiset of points in the plane, where each point is matched with a *feature* (branch or hole) of the Reeb graph in a one-to-one manner. Furthermore, the coordinates of the point characterize the *span* of the feature, that is, the interval of $\mathbb{R}$ spanned by its image under $\tilde{f}$. The vertical distance of the point to the diagonal $\Delta = \{(x, x) : x \in \mathbb{R}\}$ measures the length of that interval and thereby quantifies the prominence of the feature. Thus, the extended persistence diagram plays the role of a "bag-of-features" type signature, summarizing the Reeb graph through its list of features together with their spans, and forgetting about the actual layout of those features.

One issue with the Reeb graph is its computation. Indeed, when the pair (X, f) is known only through a finite set of measurements, the graph can only be approximated within a certain error. Quantifying this error, in particular finding the right metric in which to measure it, has been the object of intense investigation in the recent years. A simple choice is to use the *bottleneck distance* between the extended persistence diagram associated to the Reeb graph and the one associated to its approximation [17]. This pseudometric treats the Reeb graph and its approximation as bags of features, and it measures the differences between their

respective sets of features. In particular, it is oblivious to the layouts of the features in each of the graphs. Other distances have been proposed recently, to capture a greater part of this layout [4, 5, 20].

Building approximations from finite point samples with scalar values is a problem in its own right. A natural approach is to build a simplicial complex (for instance the *Rips complex*) on top of the point samples, to serve as a proxy for the underlying continuous space; then, to extend the scalar values at the vertices to a piecewise-linear (PL) function over the simplicial complex by linear interpolation; finally, to apply some exact computation algorithm for PL functions. This is the approach advocated by Dey and Wang [23], who rely on the $O(n \log n)$ expected time algorithm of Harvey, Wenger and Wang [26] for the last step. The drawbacks of this approach are:

- Its relative complexity: the Reeb graph computation from the PL function is based on collapses of its simplicial domain that may break the complex structure temporarily and therefore require some repairs.
- Its overall computational cost: here, n is not the number of data points, but the number of vertices, edges and triangles of the Rips complex, which, in principle, can be up to cubic in the number of data points. Indeed, triangles are needed to compute an approximation of the Reeb graph, in the same way as they are to compute 1-dimensional homology.

The *Mapper*[1] was introduced by Singh, Mémoli and Carlsson [32] as a new mathematical object to summarize the topological structure of a pair $(X, f : X \to \mathbb{R}^d)$. Its construction depends on the choice of a cover $\mathcal{I}$ of the image of f by open sets. Pulling back $\mathcal{I}$ through f gives an open cover of the domain X. This cover may have some elements that are disconnected, so it is refined into a connected cover by splitting each element into its various connected components. Then, the Mapper is defined as the nerve of the connected cover, having one vertex per element, one edge per pair of intersecting elements, and more generally, one k-simplex per non-empty $(k+1)$-fold intersection. From a philosophical point of view, the Mapper can be thought of as a *pixelized version* of the Reeb space, where the resolution is prescribed by the cover $\mathcal{I}$. From a practical point of view, its construction from point cloud data is very easy to describe and to implement, using standard graph traversals to detect connected components. Furthermore, it only requires to build the 1-skeleton graph of the Rips complex, whose size scales up at worst quadratically (and not cubically) with the size of the input point cloud.

As a simple alternative to the Reeb space, the Mapper has been the object of much interest by practitioners in the data sciences. It has played a key role in several success stories, such as the identification of a new subgroup of breast cancers [30], or the elaboration of a new classification of player positions in the NBA [1], due to its ability to deal with very general functions and datasets. Meanwhile, it has become the flagship component in the software suite developed by Ayasdi, a data analytics company founded in the late 2000's whose interest is to promote the use of topological methods in the data sciences. Somewhat surprisingly, despite this success, very little is known to date about the structure of the Mapper and its stability with respect to perturbations of the pair (X, f) or of the cover $\mathcal{I}$. Intuitively, when f is scalar, as a pixelized version of the Reeb graph, the Mapper should capture some of its features (branches, holes) and miss others, depending on how the cover $\mathcal{I}$ is positioned with respect to the critical values of f. How can we formalize this phenomenon? The stability of the structure of the Mapper should also depend on this positioning. How can we quantify it? These are the questions addressed in this article.

Contributions. We draw an explicit connection between the Mapper and the Reeb graph, from which we derive guarantees on the structure of the Mapper and quantities to measure its stability. Specifically:

- The connection happens through an intermediate object, called the *MultiNerve Mapper*, which we define as the *multinerve* of the connected pullback cover in the sense of [19]. The Mapper and its MultiNerve variant are related through the usual Nerve-vs-MultiNerve connection (see Lemma 3.4).

[1] In this article we call *Mapper* the mathematical object, not the algorithm used to build it. Moreover, we focus on the case where the codomain of the function is $\mathbb{R}$.

- Given a pair (X, f) and an interval cover $\mathcal{I}$, we relate the topological structure of the (MultiNerve) Mapper to the one of the Reeb graph. More precisely, we characterize the topology of these objects with particular *zigzag persistence modules* that we relate to each other with the so-called Mayer-Vietoris half-pyramid (Theorem 4.3). This correspondence is oblivious to the actual layouts of the topological features in the two graphs, which in principle could differ.

- The previous connection allows us to derive a signature for the (MultiNerve) Mapper, which takes the form of an extended persistence diagram. The points in this diagram are in one-to-one correspondence with the features (branches, holes) in the (MultiNerve) Mapper. Thus, like the extended persistence diagram of the induced map $\tilde{f}$ for the Reeb graph, our diagram for the (MultiNerve) Mapper serves as a bag-of-features type signature of its structure.

- An interesting property of our signature is to be predictable[2] given the extended persistence diagram of the induced map $\tilde{f}$. Indeed, it is obtained from this diagram by removing the points lying in certain *staircases* that are defined solely from the cover $\mathcal{I}$ and that encode the mutual positioning of the intervals of the cover. Thus, the signature for the (MultiNerve) Mapper is a subset of the one for the Reeb graph, which provides theoretical evidence to the intuitive claim that the Mapper is a pixelized version of the Reeb graph. Then, one can easily derive sufficient conditions under which the bag-of-features structure of the Reeb graph is preserved in the (MultiNerve) Mapper, and when it is not, one can easily predict which features are preserved and which ones disappear (Corollary 4.6).

- The staircases also play a role in the stability of the (MultiNerve) Mapper, since they prescribe which features will (dis-)appear as the function f is perturbed. Stability is then naturally measured by a slightly modified version of the bottleneck distance, in which the staircases play the role of the diagonal. Our stability guarantees (Theorem 5.2) follow easily from the general stability theorem for extended persistence [18]. Similar guarantees hold when the domain X or the cover $\mathcal{I}$ is perturbed (Theorems 5.4 and 6.1).

- These stability guarantees can be exploited in practice to approximate the signatures of the Mapper and MultiNerve Mapper from point cloud data. The approach boils down to applying known scalar field analysis techniques [13] then pruning the obtained extended persistence diagrams using the staircases (Theorem 8.10). The approach becomes more involved if one wants to further guarantee that the approximate signature does correspond to some perturbed Mapper or MultiNerve Mapper (Theorem 8.9).

- We also refine the analysis by showing that the MultiNerve Mapper itself is a Reeb graph, for a perturbed pair (X', f') (Theorem 7.18). Furthermore, we are able to track the changes that occur in the structure of the Reeb graph as we go from the initial pair (X, f) to its perturbed version (X', f'). This allows us to compute the *functional distortion distance* between the (MultiNerve) Mapper and the Reeb graph (Theorem 7.24). Our main proof technique consists in progressively perturbing the so-called *telescope* [7] associated with the pair (X, f). To be more specific, we decompose the perturbation into a sequence of elementary perturbations with a predictable effect on the functional distortion distance. We believe the introduction of these elementary perturbations and the analysis of their effects on persistence diagrams and on the functional distortion distance between telescopes are of an independent interest (see Section 7). In particular, these elementary perturbations have already been used in other works about Reeb graphs and Mappers [9, 10].

Related work. Reeb graphs can be seen as particular types of *skeletonization*, when one tries to recover the geometric structure of data with graphs. Several kinds of graph-like geometric structures have been studied within the past few years, such as *persistent skeletons* [27] or *graph-induced simplicial complexes* [21]. See [27] for a list of references. As mentioned previously, Reeb graphs are now well understood and have been used in a wide range of applications. Algorithms for their computation have been proposed, as well as

[2] As a byproduct, we also clarify the relationship between the extended persistence diagrams of $\tilde{f}$ and f (Theorem 2.8).

metrics for their comparison. We refer the interested reader to the survey [6] and to the introductions of [4] and [5] for a comprehensive list of references. In a recent study, even more structure has been given to the Reeb graphs by categorifying them [20].

A lot of variants of these graphs have also been studied in the last decade to face the common issues that come with the Reeb graphs (complexity and computational cost among others). The Mapper [32] is one of them. Chazal et al. [16] introduced the λ-*Reeb graph*, which is another type of Reeb graph pixelization with intervals. It is the quotient space obtained by identifying the points with the transitive closure of the following relation: $x \sim y \Leftrightarrow x, y$ belong to the same level set and x, y belong to the same element of a given family of intervals. The computation is easier than for the Reeb graph, and the authors can derive upper bounds on the Gromov-Hausdorff distance between the space and its Reeb or λ-Reeb graph. However, this is too much asking in general; as a result, the hypothesis made on the space are very strong (it has to be close to a metric graph in the Gromov-Hausdorff distance already). Moreover, taking the transitive closure makes the structure of the output more difficult to interpret.

Joint Contour Nets [8, 12] and Extended Reeb graphs [3] are Mapper-like objects. The former is the Mapper computed with the cover of the codomain given by rounding the function values, while the latter is the Mapper computed from a partition of the domain with no overlap. Both structures are used for scientific purposes (visualization, shape descriptors, to name a few) and algorithms are proposed for their computation. Babu [2] characterized the Mapper with coarsened levelset zigzag persistence modules and showed that, as the lengths of the intervals in the cover $\mathcal{I}$ go to zero uniformly, the Mapper of a real-valued function converges to the continuous Reeb graph in the bottleneck distance. Similarly, Munch and Wang [29] recently characterized the Mapper with constructible cosheaves and showed the same type of convergence for both the Joint Contour Net and the Mapper in the so-called *interleaving distance* [20]. Their result holds in the general case of vector-valued functions. Differently, we restrict the focus to real-valued functions but are able to make non-asymptotic claims (Corollary 4.6).

On another front, Stovner [33] proposed a categorified version of the Mapper, which is seen as a covariant functor from the category of covered topological spaces to the category of simplicial complexes. Dey et al. [22] pointed out the inherent instability of the Mapper and introduced a multiscale variant, called the *MultiScale Mapper*, which is built by taking the Mapper over a hierarchy of covers of the codomain. They derived a stable signature by considering the persistence diagram of this family. Unfortunately, their construction is hard to relate to the original Mapper. Here we work with the original Mapper directly and answer two open questions from [22], introducing a signature that gives a complete description of the set of features of the Mapper together with a quantification of their stability and a provable way of approximating them from point cloud data.

2 Background

Throughout the paper we work with singular homology with coefficients in the field $\mathbb{Z}_2$, which we omit in our notations for simplicity. We use the term "connected" as a shorthand for "path-connected". Given a real-valued function f on a topological space X, and an interval $I \subseteq \mathbb{R}$, we denote by X_f^I the preimage $f^{-1}(I)$. We omit the subscript f in the notation when there is no ambiguity in the function considered.

2.1 Morse-Type Functions

We restrict our focus to the class of real-valued functions called *Morse-type*. These are generalizations of the classical Morse functions:

Definition 2.1. *A continuous real-valued function f on a topological space X is* of Morse type *if:*

(i) *There is a finite set* $\mathrm{Crit}(f) = \{a_1 < ... < a_n\}$*, called the set of* critical values*, such that over every open interval* $(a_0 = -\infty, a_1), ..., (a_i, a_{i+1}), ..., (a_n, a_{n+1} = +\infty)$ *there is a compact and locally connected space* Y_i *and a homeomorphism* $\mu_i : Y_i \times (a_i, a_{i+1}) \to X^{(a_i, a_{i+1})}$ *such that* $\forall i = 0, ..., n, f|_{X^{(a_i, a_{i+1})}} = \pi_2 \circ \mu_i^{-1}$*, where* π_2 *is the projection onto the second factor;*

(ii) $\forall i = 1, ..., n-1, \mu_i$ *extends to a continuous function* $\bar{\mu}_i : Y_i \times [a_i, a_{i+1}] \to X^{[a_i, a_{i+1}]}$ – *similarly* μ_0 *extends to* $\bar{\mu}_0 : Y_0 \times (-\infty, a_1] \to X^{(-\infty, a_1]}$ *and* μ_n *extends to* $\bar{\mu}_n : Y_n \times [a_n, +\infty) \to X^{[a_n, +\infty)}$*;*

(iii) *Each levelset* X^t *has a finitely-generated homology.*

Morse functions are known to be of Morse type while the converse is clearly not true. In fact, Morse-type functions do not have to be differentiable, and their domain does not have to be a smooth manifold nor even a manifold at all. Furthermore, it is possible to find Morse-type functions that are not Morse even though they satisfy the previous assumptions (think of the Gaussian curvature on a torus, for instance).

2.2 Extended Persistence

Let f be a real-valued function on a topological space X. The family $\{X^{(-\infty,\alpha]}\}_{\alpha\in\mathbb{R}}$ of sublevel sets of f defines a *filtration*, that is, it is nested w.r.t. inclusion: $X^{(-\infty,\alpha]} \subseteq X^{(-\infty,\beta]}$ for all $\alpha \leq \beta \in \mathbb{R}$. The family $\{X^{[\alpha,+\infty)}\}_{\alpha\in\mathbb{R}}$ of superlevel sets of f is also nested but in the opposite direction: $X^{[\alpha,+\infty)} \supseteq X^{[\beta,+\infty)}$ for all $\alpha \leq \beta \in \mathbb{R}$. We can turn it into a filtration by reversing the real line. Specifically, let $\mathbb{R}^{\mathrm{op}} = \{\tilde{x} : x \in \mathbb{R}\}$, ordered by $\tilde{x} \leq \tilde{y} \Leftrightarrow x \geq y$. We index the family of superlevel sets by $\mathbb{R}^{\mathrm{op}}$, so now we have a filtration: $\{X^{[\tilde{\alpha},+\infty)}\}_{\tilde{\alpha}\in\mathbb{R}^{\mathrm{op}}}$, with $X^{[\tilde{\alpha},+\infty)} \subseteq X^{[\tilde{\beta},+\infty)}$ for all $\tilde{\alpha} \leq \tilde{\beta} \in \mathbb{R}^{\mathrm{op}}$.

Extended persistence connects the two filtrations at infinity as follows. Replace each superlevel set $X^{[\tilde{\alpha},+\infty)}$ by the pair of spaces $(X, X^{[\tilde{\alpha},+\infty)})$ in the second filtration. This maintains the filtration property since we have $(X, X^{[\tilde{\alpha},+\infty)}) \subseteq (X, X^{[\tilde{\beta},+\infty)})$ for all $\tilde{\alpha} \leq \tilde{\beta} \in \mathbb{R}^{\mathrm{op}}$. Then, let $\mathbb{R}_{\mathrm{Ext}} = \mathbb{R} \cup \{+\infty\} \cup \mathbb{R}^{\mathrm{op}}$, where the order is completed by $\alpha < +\infty < \tilde{\beta}$ for all $\alpha \in \mathbb{R}$ and $\tilde{\beta} \in \mathbb{R}^{\mathrm{op}}$. This poset is isomorphic to $(\mathbb{R}, \leq)$. Finally, define the *extended filtration* of f over $\mathbb{R}_{\mathrm{Ext}}$ by:

$$\begin{aligned} F_\alpha &= X^{(-\infty,\alpha]} && \text{for } \alpha \in \mathbb{R} \\ F_{+\infty} &= X \equiv (X, \emptyset) \\ F_{\tilde{\alpha}} &= (X, X^{[\tilde{\alpha},+\infty)}) && \text{for } \tilde{\alpha} \in \mathbb{R}^{\mathrm{op}}, \end{aligned}$$

where we have identified the space X with the pair of spaces $(X, \emptyset)$. This is a well-defined filtration since we have $X^{(-\infty,\alpha]} \subseteq X \equiv (X, \emptyset) \subseteq (X, X^{[\tilde{\beta},+\infty)})$ for all $\alpha \in \mathbb{R}$ and $\tilde{\beta} \in \mathbb{R}^{\mathrm{op}}$. The subfamily $\{F_\alpha\}_{\alpha\in\mathbb{R}}$ is called the *ordinary* part of the filtration, and the subfamily $\{F_{\tilde{\alpha}}\}_{\tilde{\alpha}\in\mathbb{R}^{\mathrm{op}}}$ is called the *relative* part. See Figure 1 for an illustration.

Applying the homology functor H_* to this filtration gives the so-called *extended persistence module* $\mathrm{EP}(f)$:

$$\begin{aligned} \mathrm{EP}(f)_\alpha &= H_*(F_\alpha) = H_*(X^{(-\infty,\alpha]}) && \text{for } \alpha \in \mathbb{R} \\ \mathrm{EP}(f)_{+\infty} &= H_*(F_{+\infty}) = H_*(X) \cong H_*(X, \emptyset) \\ \mathrm{EP}(f)_{\tilde{\alpha}} &= H_*(F_{\tilde{\alpha}}) = H_*(X, X^{[\tilde{\alpha},+\infty)}) && \text{for } \tilde{\alpha} \in \mathbb{R}^{\mathrm{op}}, \end{aligned}$$

and where the linear maps between the spaces are induced by the inclusions in the extended filtration.

For functions of Morse type, the extended persistence module can be decomposed as a finite direct sum of closed-open *interval modules*—see e.g. [14]:

$$\mathrm{EP}(f) \simeq \bigoplus_{k=1}^{n} \mathbb{I}[b_k, d_k),$$

where each summand $\mathbb{I}[b_k, d_k)$ is made of copies of the field of coefficients at each index $\alpha \in [b_k, d_k)$, and of copies of the zero space elsewhere, the maps between copies of the field being identities. Each summand represents the lifespan of a *homological feature* (connected component, hole, void, etc.) within the filtration. More precisely, the *birth time* b_k and *death time* d_k of the feature are given by the endpoints of the interval. Then, a convenient way to represent the structure of the module is to plot each interval in the decomposition as a point in the extended plane, whose coordinates are given by the endpoints. Such a plot is called the *extended persistence diagram* of f, denoted $\mathrm{Dg}(f)$. The distinction between ordinary and relative parts of the filtration allows to classify the points in $\mathrm{Dg}(f)$ in the following way:

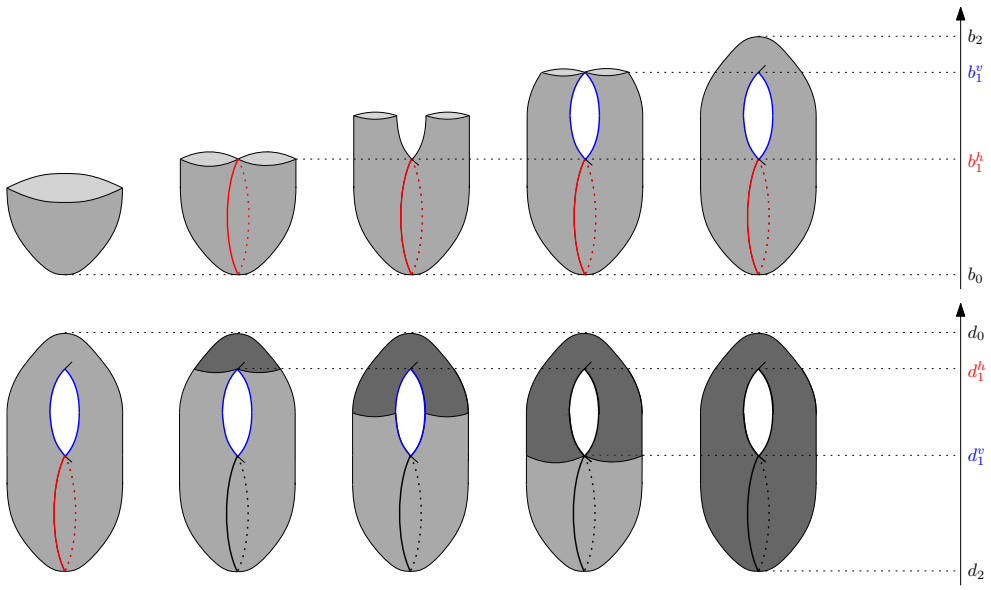


Figure 1: The extended filtration of the height function on a torus. The upper row displays the ordinary part of the filtration while the lower row displays the relative part. The red and blue cycles both correspond to extended points in dimension 1. The point corresponding to the red cycle is located above the diagonal ($d_1^h > b_1^h$), while the point corresponding to the blue cycle is located below the diagonal ($d_1^v > b_1^v$).

- points whose coordinates both belong to $\mathbb{R}$ are called *ordinary* points; they correspond to homological features being born and then dying in the ordinary part of the filtration;
- points whose coordinates both belong to $\mathbb{R}^{\mathrm{op}}$ are called *relative* points; they correspond to homological features being born and then dying in the relative part of the filtration;
- points whose abscissa belongs to $\mathbb{R}$ and whose ordinate belongs to $\mathbb{R}^{\mathrm{op}}$ are called *extended* points; they correspond to homological features being born in the ordinary part and then dying in the relative part of the filtration.

Note that ordinary points lie strictly above the diagonal $\Delta = \{(x, x) : x \in \mathbb{R}\}$ and relative points lie strictly below Δ, while extended points can be located anywhere, including on Δ, e.g. connected components that lie inside a single critical level—see Section 2.4. It is common to decompose $\mathrm{Dg}(f)$ according to this classification:

$$\mathrm{Dg}(f) = \mathrm{Ord}(f) \sqcup \mathrm{Rel}(f) \sqcup \mathrm{Ext}^+(f) \sqcup \mathrm{Ext}^-(f),$$

where by convention $\mathrm{Ext}^+(f)$ includes the extended points located on the diagonal Δ.

Persistence measure. From an extended persistence module $\mathrm{EP}(f)$ we derive a measure on the set of rectangles in the plane, called the *persistence measure* and denoted μ_{EP}. Given a rectangle $R = [a, b] \times [c, d]$ with $a < b \leq c < d$, we let

$$\mu_{\mathrm{EP}}(R) = r_b^c - r_b^d + r_a^d - r_a^c, \tag{1}$$

where r_x^y denotes the rank of the linear map between the vector spaces indexed by $x, y \in \mathbb{R}_{\mathrm{Ext}}$ in $\mathrm{EP}(f)$. When $\mathrm{EP}(f)$ has a well-defined persistence diagram, $\mu_{\mathrm{EP}}(R)$ equals the total multiplicity of the diagram within the rectangle R [14].

Stability. An important property of extended persistence diagrams is to be stable in the so-called *bottleneck distance* d_{b}^{∞}. The definition of this distance is based on partial matchings between the diagrams. Given two persistence diagrams D, D', a *partial matching* between D and D' is a subset Γ of $D \times D'$ such that:

$$\forall p \in D, \text{ there is at most one } p' \in D' \text{ s.t. } (p, p') \in \Gamma,$$

$$\forall p' \in D', \text{ there is at most one } p \in D \text{ s.t. } (p, p') \in \Gamma.$$

Furthermore, Γ must match points of the same type (ordinary, relative, extended) and of the same homological dimension only. The *cost* of Γ is:

$$\mathrm{cost}(\Gamma) = \max\left\{\max_{p\in D} \delta_D(p), \max_{p'\in D'} \delta_{D'}(p')\right\},$$

where

$$\delta_D(p) = \|p - p'\|_\infty \text{ if } \exists p' \in D' \text{ s.t. } (p, p') \in \Gamma \text{ and } d_\infty(p, \Delta) = \inf_{q\in\Delta} \|p - q\|_\infty \text{ otherwise,}$$

$$\delta_{D'}(p') = \|p - p'\|_\infty \text{ if } \exists p \in D \text{ s.t. } (p, p') \in \Gamma \text{ and } d_\infty(p', \Delta) = \inf_{q\in\Delta} \|p' - q\|_\infty \text{ otherwise.}$$

Definition 2.2. *Let D, D' be two persistence diagrams. The* bottleneck distance *between D and D' is:*

$$d_{\mathrm{b}}(D, D') = \inf_{\Gamma} \mathrm{cost}(\Gamma),$$

where Γ ranges over all partial matchings between D and D'.

Note that d_{b} is only a pseudometric, not a true metric, because points lying on Δ can be left unmatched at no cost.

Theorem 2.3 (Stability Theorem in [18])**.** *For any Morse-type functions $f, g : X \to \mathbb{R}$,*

$$d_{\mathrm{b}}^{\infty}(\mathrm{Dg}(f), \mathrm{Dg}(g)) \leq \|f - g\|_\infty.$$

Moreover, as pointed out in [18], the theorem can be strengthened to apply to each subdiagram Ord, Ext^+, Ext^-, Rel and to each homological dimension individually.

2.3 Zigzag persistence

Let $f : X \to \mathbb{R}$ be a Morse-type function, and let $\mathrm{Crit}(f) = \{a_1, \cdots, a_n\}$ be its set of critical values. Let $-\infty = a_0 < s_0 < a_1 < s_1 < a_2 < \cdots < s_{n-1} < a_n < s_n < a_{n+1} = +\infty$. Then, for any $1 \leq i \leq j \leq n$, we define $X_i^j = X^{[s_i, s_j]}$, and the *levelset zigzag* as the following sequence of $2n + 1$ nodes:

$$X_0^0 \hookrightarrow X_0^1 \hookleftarrow X_1^1 \hookrightarrow X_1^2 \hookleftarrow \cdots \hookrightarrow X_{n-1}^n \hookleftarrow X_n^n,$$

where each arrow is the canonical inclusion. Applying the homology functor H_* to the levelset zigzag gives the so-called *levelset zigzag persistence module* $\mathrm{LZZ}(f)$, where the linear maps between the spaces are induced by the inclusions. For functions of Morse type, the levelset zigzag persistence module decomposes as a finite direct sum of closed interval modules:

$$\mathrm{LZZ}(f) \simeq \bigoplus_{k=1}^{m} \mathbb{I}[X_i^{i'}, X_j^{j'}],$$

where i' is either i or $i + 1$, and similarly for j'. Hence, the classification given by Table 1.

Moreover, each summand $\mathbb{I}[X_i^{i'}, X_j^{j'}]$ is made of copies of the field of coefficients for each space between $X_i^{i'}$ and $X_j^{j'}$ and of copies of the zero space elsewhere, the maps between copies of the field being identities. The disjoint union of all of these intervals is called the *levelset zigzag persistence barcode* $\mathrm{LBc}(f)$.

Table 1: Classification of intervals in a levelset zigzag persistence barcode.

Type	I	II	III	IV
	$i'_\alpha = i_\alpha + 1$ $j'_\alpha = j_\alpha$	$i'_\alpha = i_\alpha$ $j'_\alpha = j_\alpha + 1$	$i'_\alpha = i_\alpha + 1$ $j'_\alpha = j_\alpha + 1$	$i'_\alpha = i_\alpha$ $j'_\alpha = j_\alpha$

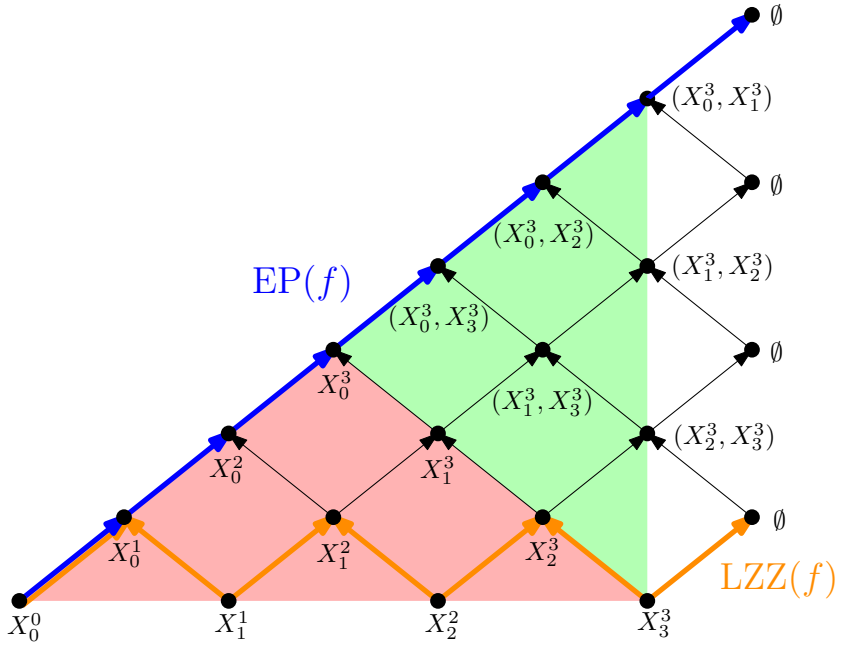


Figure 2: Mayer-Vietoris half-pyramid when the Morse-type function has three critical values. It is composed of two faces of the full Mayer-Vietoris pyramid: the south face (red) and the east face (green). The extended persistence module EP(f) is in blue and the levelset zigzag persistence module LZZ(f) is in orange.

Mayer-Vietoris half-pyramid. The Mayer-Vietoris half-pyramid is the diagram of topological spaces and inclusions displayed in Figure 2. We refer the reader to [7] for more details. Any zigzag within the Mayer-Vietoris half-pyramid that stretches from the left boundary (i.e. the node X_0^0) to the right boundary without backtracking is called *monotone*. Theorem 2.4 below relates any two monotone zigzags.

Theorem 2.4 (Pyramid Theorem in [7])**.** *For any Morse-type function f, there exists a bijection between the barcodes of any pair of monotone zigzag persistence modules in the Mayer-Vietoris half-pyramid.*

Since the extended persistence module of f is a monotone zigzag persistence module—more precisely the principal diagonal—of the Mayer-Vietoris half-pyramid, we have the following corollary.

Corollary 2.5 (Table 1 in [7])**.** *For any Morse-type function f, there exists a bijection between* Dg(f) *and* LBc(f)*, which is described in Table 2.*

In particular, the bottleneck distance, as well as stability results, can be derived for levelset zigzag persistence barcodes using this correspondence and Theorem 2.3.

2.4 Reeb Graphs

Definition 2.6. *Given a topological space X and a continuous function $f : X \to \mathbb{R}$, we define the equivalence relation $\sim_f$ between points of X by:*

$$x \sim_f y \iff \left[f(x) = f(y) \text{ and } x, y \text{ belong to the same connected component of } f^{-1}(f(x)) = f^{-1}(f(y))\right].$$

Table 2: This table gives the correspondences between the points of $\mathrm{Dg}(f)$ and the intervals of $\mathrm{LBc}(f)$. The minus sign on some intervals of $\mathrm{LBc}(f)$ means that the homological dimension of that interval is equal to the dimension of its corresponding point in $\mathrm{Dg}(f)$ minus 1.

Type	Ord	Rel	Ext^+	Ext^-
$\mathrm{Dg}(f)$	$[a_i, a_j)$	$[\tilde{a}_j, \tilde{a}_i)$	$[a_i, \tilde{a}_j)$	$[a_j, \tilde{a}_i)$
$\mathrm{LBc}(f)$	$[X_{i-1}^i, X_{j-1}^{j-1}]$	$[X_i^i, X_{j-1}^j]^-$	$[X_{i-1}^i, X_{j-1}^j]$	$[X_i^i, X_{j-1}^{j-1}]^-$
Type	I	II	III	IV

The Reeb graph $\mathrm{R}_f(X)$ *is the quotient space* $X/\sim_f$.

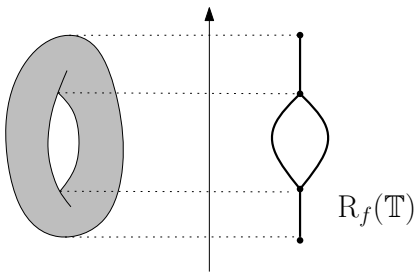


Figure 3: We consider the height function of the torus $\mathbb{T}$. Note how the critical points induce changes on the graph.

As f is constant on equivalence classes, there is an induced map $\tilde{f} : \mathrm{R}_f(X) \to \mathbb{R}$ such that $f = \tilde{f} \circ \pi$, where π is the quotient map $X \to \mathrm{R}_f(X)$:

$$\begin{array}{ccc} X & \xrightarrow{\pi} & \mathrm{R}_f(X) \\ & {\scriptstyle f}\searrow \quad \swarrow{\scriptstyle \tilde{f}} & \\ & \mathbb{R} & \end{array} \tag{2}$$

If f is a function of Morse type, then the pair (X, f) is an $\mathbb{R}$-*constructible space* in the sense of [20]. This ensures that the Reeb graph is a multigraph, whose nodes are in one-to-one correspondence with the connected components of the critical level sets of f. In that case, computing the Reeb graph of a Reeb graph preserves all information, as stated in the following remark.

Remark 2.7. *Let f be a Morse-type function. Then there is a bijection $b : \mathrm{R}_{\tilde{f}}(\mathrm{R}_f(X)) \to \mathrm{R}_f(X)$ such that $\tilde{f} \circ b = \tilde{\tilde{f}}$. In other words, computing the Reeb graph is an idempotent operation.*

In the following, the combinatorial version of the Reeb graph (where each critical point is turned into a node) is denoted by $\mathcal{C}\mathrm{R}_f(X)$.

Persistence-based bag-of-features signature. There is a nice interpretation of $\mathrm{Dg}(\tilde{f})$ in terms of the structure of $\mathrm{R}_f(X)$. We refer the reader to [4] and the references therein for a full description as well as formal definitions and statements. Orienting the Reeb graph vertically so $\tilde{f}$ is the height function, we can see each connected component of the graph as a trunk with multiple branches (some oriented upwards, others oriented downwards) and holes. Then, one has the following correspondences, where the *vertical span* of a feature is the span of its image by $\tilde{f}$:

- The vertical spans of the trunks are given by the points in $\mathrm{Ext}_0^+(\tilde{f})$;

- The vertical spans of the branches that are oriented downwards are given by the points in $\mathrm{Ord}_0(\tilde{f})$;
- The vertical spans of the branches that are oriented upwards are given by the points in $\mathrm{Rel}_1(\tilde{f})$;
- The vertical spans of the holes are given by the points in $\mathrm{Ext}_1^-(\tilde{f})$.

The rest of the diagram of $\tilde{f}$ is empty. These correspondences provide a dictionary to read off the structure of the Reeb graph from the extended persistence diagram of the induced map $\tilde{f}$. Note that it is a bag-of-features type signature, taking an inventory of all the features (trunks, branches, holes) together with their vertical spans, but leaving aside the actual layout of the features. As a consequence, it is an incomplete signature: two Reeb graphs with the same persistence diagram may not be isomorphic, as illustrated in Figure 4.

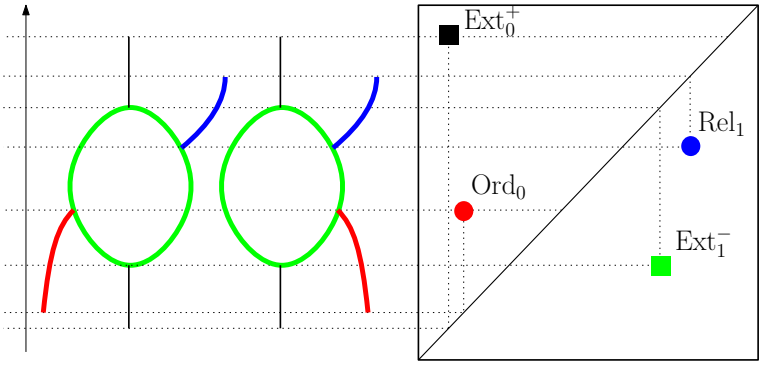


Figure 4: Two Reeb graphs with the same set of features but not the same layout.

Connection to the extended and zigzag persistence of f. We now show that the topological structure of $\mathrm{R}_f(X)$ is actually nothing but a simplification of the one of f. This can be phrased using the extended persistence diagrams of f and $\tilde{f}$:

Theorem 2.8. *Let X be a topological space and $f : X \to \mathbb{R}$ be a function of Morse type. Then, the levelset zigzag persistence barcodes of f and $\tilde{f}$ in dimension 0 are the same: $\mathrm{LBc}_0(f) = \mathrm{LBc}_0(\tilde{f})$, and the extended persistence diagram of $\tilde{f}$ is included in the one of f: $\mathrm{Dg}(\tilde{f}) \subseteq \mathrm{Dg}(f)$. More precisely:*

$$\begin{aligned}
\mathrm{Dg}_0(\tilde{f}) &= \mathrm{Dg}_0(f) \\
\mathrm{Dg}_1(\tilde{f}) &= \mathrm{Dg}_1(f) \setminus (\mathrm{Ext}_1^+(f) \cup \mathrm{Ord}_1(f)) \\
\mathrm{Dg}_p(\tilde{f}) &= \emptyset \text{ if } p \geq 2
\end{aligned}$$

Note that $\mathrm{Ext}_0^-(\tilde{f}) = \emptyset$ because every essential 0-dimensional feature corresponds to some connected component of the domain, and it is born at the minimum function value and killed at the maximum function value over that connected component, hence it belongs to Ext_0^+. Similarly, $\mathrm{Rel}_0(\tilde{f}) = \emptyset$ because no 0-dimensional homology class (i.e. connected component) can be created in the relative part of the extended filtration of f. Hence, the structure of a Reeb graph can be read off from the levelset zigzag persistence module of $\tilde{f}$. Indeed, since $\mathrm{Ext}_1^+(\tilde{f})$, $\mathrm{Ord}_1(\tilde{f}), \mathrm{Ext}_0^-(\tilde{f}), \mathrm{Rel}_0(\tilde{f})$ and $\mathrm{Dg}_p(\tilde{f})$ for $p \geq 2$ are empty, it follows from Corollary 2.5 that there is a bijection preserving types between $\mathrm{Dg}_0(\tilde{f}) \cup \mathrm{Dg}_1(\tilde{f})$ and $\mathrm{LBc}_0(\tilde{f})$. This is because all intervals in the 1-dimensional extended persistence module of $\tilde{f}$ are either of type Rel or Ext^-, and thus their analogues in the levelset zigzag persistence module of $\tilde{f}$ have homological dimension 0 according to Table 2.

We provide a proof for completeness, as we have not seen this result stated formally in the literature. First, note that $\mathrm{Crit}(f) = \{a_1, \cdots, a_n\} = \mathrm{Crit}(\tilde{f})$. Hence, given $i \leq j$ and $[s_i, s_j]$ as in Section 2.3, we recall that X_i^j denote $X^{[s_i,s_j]} = f^{-1}([s_i, s_j])$ and $\mathrm{R}_f(X)_i^j$ denote $\mathrm{R}_f(X)^{[s_i,s_j]} = \tilde{f}^{-1}([s_i, s_j])$.

Lemma 2.9. *Let π denote the quotient map $X \to \mathrm{R}_f(X)$. Let $i \leq j$, and $I = [s_i, s_j]$, as defined in Section 2.3. Then the morphism $\pi_* : H_0(X^I) \to H_0(\mathrm{R}_f(X)^I)$ is an isomorphism.*

The proof of Lemma 2.9 is simpler when π admits *continuous sections*, i.e. when there exist continuous maps $\sigma : \mathrm{R}_f(X) \to X$ such that $\pi \circ \sigma = \mathrm{id}_{\mathrm{R}_f(X)}$. Below we give the proof under this hypothesis, deferring the general case of Morse-type functions to Appendix A. The hypothesis holds for instance when X is a compact smooth manifold and f is a Morse function, or when X is a simplicial complex and f is piecewise-linear.

Lemma 2.9. Since π is surjective, proving the result boils down to showing that x, y are connected in X_i^j if and only if $\pi(x), \pi(y)$ are connected in $\mathrm{R}_f(X)_i^j$.

- If x, y are connected in X_i^j, then $\pi(x), \pi(y)$ are connected in $\mathrm{R}_f(X)_i^j$ by continuity of π and commutativity of (2).
- If $\pi(x), \pi(y)$ are connected in $\mathrm{R}_f(X)_i^j$, then choose a path γ connecting $\pi(x)$ to $\pi(y)$. By definition of σ, we have $\pi \circ \sigma \circ \pi(x) = \pi(x)$, thus $\sigma \circ \pi(x)$ and x lie in the same connected component of $f^{-1}(f(x))$. Let γ_x be a path connecting x to $\sigma \circ \pi(x)$. Similarly, let γ_y be a path connecting $\sigma \circ \pi(y)$ to y. Then, $\gamma_y \circ \sigma(\gamma) \circ \gamma_x$ is a path between x and y in X_i^j.

□ □

Theorem 2.8. We first show that $\mathrm{LBc}_0(f) = \mathrm{LBc}_0(\tilde{f})$. Let π denote the quotient map $X \to \mathrm{R}_f(X)$. Since π is continuous, it induces a morphism in homology π_*. We will show that π_* induces an isomorphism between $\mathrm{LZZ}(f)$ and $\mathrm{LZZ}(\tilde{f})$ in dimension 0. First, note that $\mathrm{Crit}(f) = \{a_1, \cdots, a_n\} = \mathrm{Crit}(\tilde{f})$. Hence both $\mathrm{LZZ}(f)$ and $\mathrm{LZZ}(\tilde{f})$ have $2n+1$ nodes. Now, let $1 \leq i \leq n$.

- According to Lemma 2.9, $\pi_* : H_0(X_i^i) \to H_0(\mathrm{R}_f(X)_i^i)$ is an isomorphism, and the same holds for $\pi_* : H_0(X_i^{i+1}) \to H_0(\mathrm{R}_f(X)_i^{i+1})$. Hence π_* induces a pointwise isomorphism in dimension 0 between $\mathrm{LZZ}(f)$ and $\mathrm{LZZ}(\tilde{f})$.
- Let $\iota : X_i^i \to X_i^{i+1}$ and $\iota^{\mathrm{R}} : \mathrm{R}_f(X)_i^i \to \mathrm{R}_f(X)_i^{i+1}$ be canonical inclusions. Then, we have $\pi \circ \iota = \iota^{\mathrm{R}} \circ \pi$ by definition of ι^{R}. Hence, the following diagram commutes:

$$\begin{array}{ccc} H_0(X_i^i) & \xrightarrow{\iota_*} & H_0(X_i^{i+1}) \\ \downarrow{\scriptstyle \pi_*} & & \downarrow{\scriptstyle \pi_*} \\ H_0(\mathrm{R}_f(X)_i^i) & \xrightarrow[\iota_*^{\mathrm{R}}]{} & H_0(\mathrm{R}_f(X)_i^{i+1}) \end{array}$$

  and the same is true for the canonical inclusions $X_{i-1}^i \hookleftarrow X_i^i$ and $\mathrm{R}_f(X)_{i-1}^i \hookleftarrow \mathrm{R}_f(X)_i^i$.

Hence, the induced pointwise isomorphism is an isomorphism between $\mathrm{LZZ}_0(f)$ and $\mathrm{LZZ}_0(\tilde{f})$.

Now, recall that there is a bijection b_1 preserving types between $\mathrm{Dg}(\tilde{f})$ and $\mathrm{LBc}_0(\tilde{f})$. Since there is also a bijection b_2 preserving types between $\mathrm{LBc}_0(\tilde{f})$ and $\mathrm{LBc}_0(f)$ and a bijection b_3 preserving types between $\mathrm{LBc}_0(f)$ and $\mathrm{Ord}_0(f) \cup \mathrm{Ext}_0^+(f) \cup \mathrm{Rel}_1(f) \cup \mathrm{Ext}_1^-(f)$ from Corollary 2.5, the result follows by considering the bijection $b_3 \circ b_2 \circ b_1$.

□ □

A distance between Reeb graphs. We now give the definition of the *functional distortion distance* [4] between Reeb graphs. Note that any Reeb graph $\mathrm{R}_f(X)$ can be equipped with a canonical metric: $d_f(x, x') = \min_{\pi : x \to x'} \{\max_{t \in [0,1]} \tilde{f} \circ \pi(t) - \min_{t \in [0,1]} \tilde{f} \circ \pi(t)\}$, where $\pi : [0,1] \to \mathrm{R}_f(X)$ ranges over the continuous paths from x to x' ($\pi(0) = x$ and $\pi(1) = x'$). Then, given a pair of Reeb graphs, the functional distortion distance measures the *distortion* of their corresponding metrics. Hence, it is very similar to the Gromov-Hausdorff distance. We use this distance in Section 7.4 to provide a convergence result of the (MultiNerve) Mapper to the Reeb graph.

Definition 2.10. *Let X, Y be topological spaces and $f : X \to \mathbb{R}$ and $g : Y \to \mathbb{R}$ be continuous scalar functions. The* functional distortion distance *between $\mathrm{R}_f(X)$ and $\mathrm{R}_g(Y)$ is:*

$$d_{\mathrm{FD}}(\mathrm{R}_f(X), \mathrm{R}_g(Y)) = \inf_{\phi,\psi} \max\left\{\frac{1}{2}D(\phi,\psi),\ \|\tilde{f} - \tilde{g}\circ\phi\|_\infty,\ \|\tilde{f}\circ\psi - \tilde{g}\|_\infty\right\}, \tag{3}$$

where:

- *$\phi : \mathrm{R}_f(X) \to \mathrm{R}_g(Y)$ and $\psi : \mathrm{R}_g(Y) \to \mathrm{R}_f(X)$ are continuous maps,*
- *$D(\phi,\psi) = \sup\ \{|d_f(x,x') - d_g(y,y')| : (x,y),(x',y') \in C(\phi,\psi)\}$,*
- *$C(\phi,\psi) = \{(x,\phi(x)) : x \in \mathrm{R}_f(X)\} \cup \{(\psi(y),y) : y \in \mathrm{R}_g(Y)\}$.*

The functional distortion distance enjoys the following stability theorem:

Theorem 2.11 (Theorem 4.1 in [4])**.** *Let X be a topological space and let $f, g : X \to \mathbb{R}$ be two Morse-type functions with continuous sections. Then:*

$$d_{\mathrm{FD}}(\mathrm{R}_f(X), \mathrm{R}_g(X)) \leq \|f - g\|_\infty.$$

Since d_{FD} can be quite hard to compute and to interpret, we also study the bottleneck distance between the extended persistence diagrams $d_{\mathrm{b}}(\mathrm{Dg}(\tilde{f}), \mathrm{Dg}(\tilde{g}))$ in Section 4. Recall that d_{b} is only a pseudometric—see Figure 4. However, it can be computed efficiently, it allows for interpretation (recall that extended persistence diagrams act as bag-of-feature signatures) and it has been proven [10] that d_{b} and d_{FD} are actually equivalent for close Reeb graphs.

2.5 Covers and Nerves

Let Z be a topological space. A *cover* of Z is a family $\mathcal{U}$ of subsets of Z, $\mathcal{U} = \{U_\alpha\}_{\alpha\in A}$, such that $Z = \bigcup_{\alpha\in A} U_\alpha$. It is *open* if all its elements are open subspaces of Z. It is *connected* if all its elements are connected subspaces of Z. Its *nerve* is the abstract simplicial complex $\mathcal{N}(\mathcal{U})$ that has one k-simplex per $(k+1)$-fold intersection of elements of $\mathcal{U}$:

$$\{\alpha_0, ..., \alpha_k\} \in \mathcal{N}(\mathcal{U}) \iff \bigcap_{i=0,...,k} U_{\alpha_i} \neq \emptyset.$$

When $\mathcal{V}$ itself is a cover of Z, it is called a *subcover* of $\mathcal{U}$. It is *proper* if it is not equal to $\mathcal{U}$. Finally, $\mathcal{U}$ is called *minimal* if it admits no proper subcover or, equivalently, if it has no element included in the union of the other elements. Given a minimal cover $\mathcal{U} = \{U_\alpha\}_{\alpha\in A}$, for every $\alpha \in A$ we let

$$\tilde{U}_\alpha = U_\alpha \setminus \bigcup_{\alpha'\neq\alpha\in A} U_{\alpha'},$$

be the *proper subset* of U_α, that is the maximal subset of U_α that has an empty intersection with the other elements of $\mathcal{U}$. $\mathcal{U}$ is called *generic* if no connected component of the proper subsets of its elements is a singleton.

Consider now the special case where Z is a subset of $\mathbb{R}$, equipped with the subspace topology. A subset $U \subseteq Z$ is an *interval of Z* if there is an interval I of $\mathbb{R}$ such that $U = I \cap Z$. Note that U is open in Z if and only if I can be chosen open in $\mathbb{R}$. A cover $\mathcal{U}$ of Z is an *interval cover* if all its elements are intervals. In this case, $\mathrm{End}(\mathcal{U})$ denotes the set of all of the interval endpoints. Finally, the *granularity* of $\mathcal{U}$ is the supremum of the lengths of its elements, i.e. it is the quantity $\sup_{U\in\mathcal{U}} |U|$ where $|U| = \sup(U) - \inf(U) \in \mathbb{R} \cup \{+\infty\}$.

Lemma 2.12. *No more than two elements of a minimal open interval cover can intersect at a time.*

Proof. Assume for a contradiction that there are $k \geq 3$ elements of $\mathcal{U}$: $U_1, \cdots, U_k$, that have a non-empty common intersection. For every i, fix an open interval I_i of $\mathbb{R}$ such that $U_i = I_i \cap Z$. Up to a reordering of the indices, we can assume without loss of generality that I_1 has the smallest lower bound and I_2 has the largest upper bound. Since $I_1 \cap I_2 \supseteq U_1 \cap U_2 \neq \emptyset$, the remaining intervals satisfy $I_i \subseteq I_1 \cup I_2$. In particular, we have $U_3 = I_3 \cap Z \subseteq (I_1 \cup I_2) \cap Z = (I_1 \cap Z) \cup (I_2 \cap Z) = U_1 \cup U_2$, so the cover $\mathcal{U}$ is not minimal. □ □

Lemma 2.13. *If Z is $\mathbb{R}$ itself or a compact subset thereof, then any cover $\mathcal{U}$ of Z has a minimal subcover.*

Proof. When Z is compact, there exists a subcover $\mathcal{V}$ of $\mathcal{U}$ that has finitely many elements. Any subcover of $\mathcal{V}$ with the minimum number of elements is then a minimal cover of Z.

When $Z = \mathbb{R}$, the same argument applies to any subset of the form $[-n, n]$, $n \in \mathbb{N}$. Then, a simple induction on n allows us to build a minimal subcover of $\mathcal{U}$. □ □

From now on, unless otherwise stated, all covers of $Z \subseteq \mathbb{R}$ will be generic, open, minimal, interval covers (*gomic* for short). Given such a cover $\mathcal{U}$, the proper subset $\tilde{U}$ of any interval $U \in \mathcal{U}$ is itself an interval of Z since $\mathcal{U}$ is generic, therefore we call it the *proper subinterval* of U. Moreover, Lemma 2.12 yields a total order on the intervals of $\mathcal{U}$, so each one of them partitions into subintervals as follows:

$$U = U_\cap^- \sqcup \tilde{U} \sqcup U_\cap^+, \tag{4}$$

where $U_\cap^-$ is the intersection of U with the element right below it in the cover ($U_\cap^- = \emptyset$ if that element does not exist), and where $U_\cap^+$ is the intersection of U with the element right above it ($U_\cap^+ = \emptyset$ if that element does not exist).

2.6 Mapper

Let $f : X \to Z$ be a continuous function. Consider a cover $\mathcal{U}$ of $\mathrm{im}(f)$, and pull it back to X via f^{-1}. Then, decompose every $V_\alpha = f^{-1}(U_\alpha) \subseteq X$ into its connected components: $V_\alpha = \bigsqcup_{i \in \{1, \cdots, c(\alpha)\}} V_\alpha^i$, where $c(\alpha)$ is the number of connected components of V_α. Then, $\mathcal{V} = \{V_\alpha^i\}_{\alpha \in A, i \in \{1, \cdots, c(\alpha)\}}$ is a connected cover of X. It is called the *connected pullback cover*, and its nerve $\mathcal{N}(\mathcal{V})$ is the Mapper.

Definition 2.14. *Let X, Z be topological spaces, $f : X \to Z$ be a continuous function, $\mathcal{U}$ be a cover of $\mathrm{im}(f)$ and $\mathcal{V}$ be the associated connected pullback cover. The* Mapper *of X is $\mathrm{M}_f(X, \mathcal{U}) = \mathcal{N}(\mathcal{V})$.*

See Figure 5 for an illustration. Note that, when $Z = \mathbb{R}$ and $\mathcal{U}$ is a gomic, the Mapper has a natural 1-dimensional stratification since no more than two intervals can intersect at a time by Lemma 2.12. Hence, in this case, it has the structure of a (possibly infinite) simple graph and therefore has trivial homology in dimension 2 and above.

3 MultiNerve Mapper

In this section, we explain how to extend the construction of Mapper by using a slight modification of the nerve, called the *multinerve*, and whose definition relies on *simplicial posets* [19].

3.1 Simplicial Posets and MultiNerves

Definition 3.1. *A* simplicial poset *is a partially ordered set $(P, \preceq)$, whose elements are called* simplices*, and which satisfies the two following properties:*

(i) *P has a least element called 0 such that $\forall p \in P$, $0 \preceq p$;*

(ii) *$\forall p \in P$, $\exists d \in \mathbb{N}$ such that the lower segment $[0, p] = \{q \in P : q \preceq p\}$ is* isomorphic *to the set of simplices of the standard d-simplex with the inclusion as partial order, where an isomorphism between posets is a bijective and order-preserving function.*

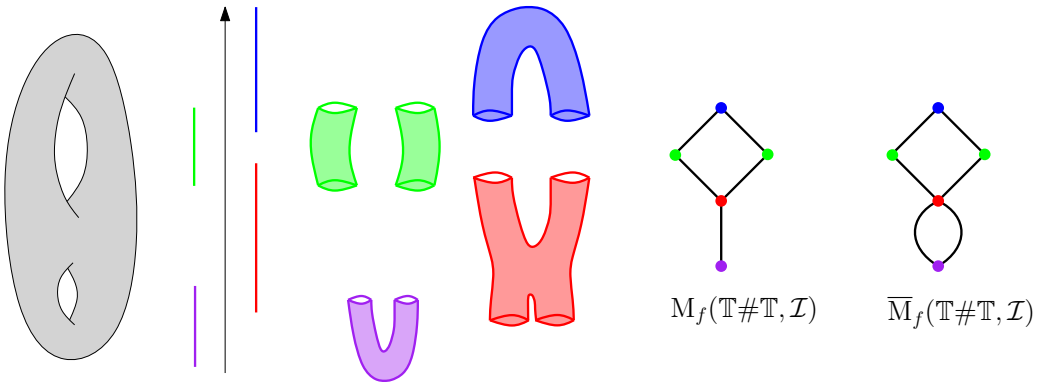


Figure 5: Example of the Mapper and the MultiNerve Mapper computed on the double torus $\mathbb{T}\#\mathbb{T}$ with the height function f. The cover $\mathcal{I}$ of $\mathrm{im}(f)$ has four intervals (red, green, blue and purple), and the cover of the double torus has five connected components (one is blue, one is red, one is purple and the other two are green). The Mapper and the MultiNerve Mapper are displayed on the right.

Simplicial posets are extensions of simplicial complexes: while every simplicial complex is also a simplicial poset (with inclusion as partial order and $\emptyset$ as least element), the converse is not always true as different simplices may have the same set of vertices. However, these simplices cannot be faces of the same higher-dimensional simplex, otherwise (ii) would be false. See Figure 6 for an example of a simplicial poset that is not a simplicial complex.

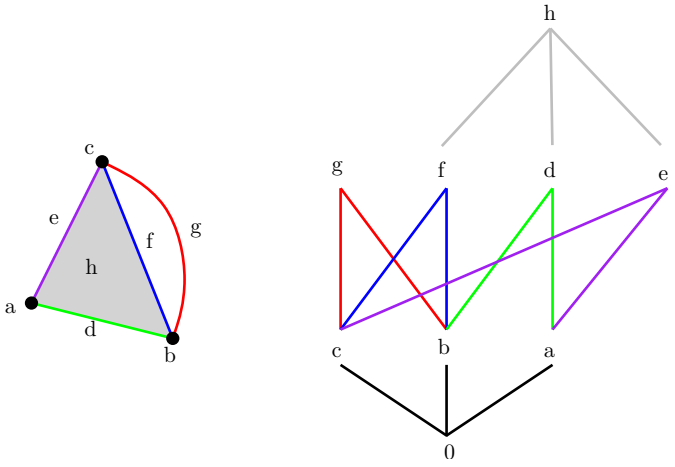


Figure 6: Left: A simplicial poset that is not a simplicial complex. Indeed, edges f and g have the same vertices (b and c). Right: The corresponding Hasse diagram showing the partial order on the simplices. Note that f, g cannot be part of the same 2-cell.

Given a cover $\mathcal{U}$ of a topological space X, the nerve is extended to a simplicial poset as follows:

Definition 3.2. *Let $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$ be a cover of a topological space X. The* multinerve $\mathcal{M}(\mathcal{U})$ *is the simplicial poset defined by:*

$$\mathcal{M}(\mathcal{U}) = \left\{ (\{\alpha_0, \cdots, \alpha_k\}, C) \ : \ \bigcap_{i=0}^{k} U_{\alpha_i} \neq \emptyset \text{ and } C \text{ is a connected component of } \bigcap_{i=0}^{k} U_{\alpha_i} \right\}.$$

The proof that this set, together with the least element $(\emptyset, \bigcup_{\alpha \in A} U_\alpha)$ and the partial order $(F, C) \preceq$

$(F', C') \Leftrightarrow F \subseteq F'$ and $C' \subseteq C$, is a simplicial poset, can be found in [19]. Given a simplex (F, C) in the multinerve of a cover, its *dimension* is $|F| - 1$. The dimension of the multinerve of a cover is the maximal dimension of its simplices. Given two simplices $(F, C), (F', C')$, we say that (F, C) is a *face* of (F', C') if $(F, C) \preceq (F', C')$.

Given a connected pullback cover $\mathcal{V}$, we extend the Mapper by using the multinerve $\mathcal{M}(\mathcal{V})$ instead of $\mathcal{N}(\mathcal{V})$. This variant will be referred to as the MultiNerve Mapper in the following.

Definition 3.3. *Let X, Z be topological spaces, $f : X \to Z$ be a continuous function, $\mathcal{U}$ be a cover of* $\mathrm{im}(f)$ *and $\mathcal{V}$ be the associated connected pullback cover. The* MultiNerve Mapper *of X is $\overline{\mathrm{M}}_f(X, \mathcal{U}) = \mathcal{M}(\mathcal{V})$.*

See Figure 5 for an illustration. For the same reasons as Mapper, when $Z = \mathbb{R}$ and $\mathcal{U}$ is a gomic of $\mathrm{im}(f)$, the MultiNerve Mapper is a (possibly infinite) multigraph having trivial homology in dimension 2 and above. Contrarily to the Mapper, the MultiNerve Mapper also takes the connected components of the intersections into account in its construction. As we shall see in Section 4, it is able to capture the same features as the Mapper but with coarser gomics, and it is more naturally related to the Reeb graph.

3.2 Connection to Mapper

The connection between the Mapper and the MultiNerve Mapper is induced by the following connection between nerves and multinerves:

Lemma 3.4 ([19])**.** *Let X be a topological space and $\mathcal{U}$ be a cover of X. Let $\pi_1 : (F, C) \mapsto F$ be the projection of the simplices of $\mathcal{M}(\mathcal{U})$ onto the first coordinate. Then, $\pi_1(\mathcal{M}(\mathcal{U})) = \mathcal{N}(\mathcal{U})$.*

Corollary 3.5. *Let X, Z be topological spaces and $f : X \to Z$ continuous. Let $\mathcal{U}$ be a cover of* $\mathrm{im}(f)$*. Then,* $\mathrm{M}_f(X, \mathcal{U}) = \pi_1(\overline{\mathrm{M}}_f(X, \mathcal{U}))$.

Thus, when $Z = \mathbb{R}$ and $\mathcal{U}$ is a gomic, the Mapper is the simple graph obtained by gluing the edges that have the same endpoints in the MultiNerve Mapper. In this special case it is even possible to embed $\mathrm{M}_f(X, \mathcal{U})$ as a subcomplex of $\overline{\mathrm{M}}_f(X, \mathcal{U})$. Indeed, both objects are multigraphs over the same set of nodes since they are built from the connected pullback cover. Then, it is enough to map each edge of $\mathrm{M}_f(X, \mathcal{U})$ to one of its copies in $\overline{\mathrm{M}}_f(X, \mathcal{U})$, chosen arbitrarily, to get a subcomplex. This mapping serves as a simplicial section for the projection π_1, therefore:

Lemma 3.6. *When $Z = \mathbb{R}$ and $\mathcal{U}$ is a gomic, π_1 induces a surjective homomorphism in homology.*

Note that this is not true in general when $\overline{\mathrm{M}}_f(X, \mathcal{U})$ has a higher dimension. See Figure 7 for an example.

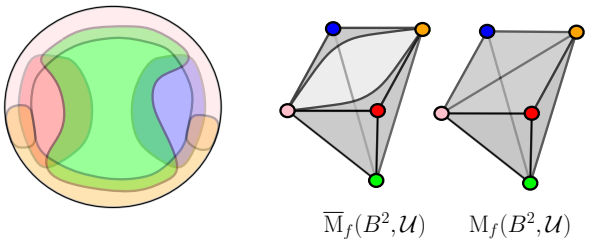


Figure 7: The domain is the disk B^2, and we consider the identity function f, as well as a generic open minimal cover $\mathcal{U}$ with five elements. The MultiNerve Mapper is homeomorphic to the disk B^2 and the Mapper is homeomorphic to the sphere S^2. Then, $H_2(\mathrm{M}_f(B^2, \mathcal{U})) \neq 0$ while $H_2(\overline{\mathrm{M}}_f(B^2, \mathcal{U})) = 0$.

4 Structure of the MultiNerve Mapper

In this section, we study and characterize the topological structure of the (MultiNerve) Mapper computed on a non discrete topological space. More precisely, we show that this topological structure can be read off from the extended persistence diagram of the Reeb graph. To prove this, we show that the MultiNerve Mapper $\overline{\mathrm{M}}_f(X,\mathcal{I})$ is actually isomorphic (as a combinatorial multigraph) to a specific Reeb graph, whose extended persistence diagram is related to the extended persistence diagram $\mathrm{Dg}(\tilde{f})$ of $\mathrm{R}_f(X)$.

4.1 Topology of the MultiNerve Mapper

In order to show that the MultiNerve Mapper is a specific Reeb graph, we first show that (MultiNerve) Mappers can be equipped with functions.

Definition 4.1. *Let $\mathcal{I} = \{I_\alpha\}_{\alpha\in A}$ be a gomic of* $\mathrm{im}(f)$ *and $\mathcal{V} = \{V_\alpha^i\}_{1\leq i\leq c(\alpha),\alpha\in A}$ be the associated connected pullback cover. Then we define $\bar{\mathsf{m}}_{\mathcal{I}} : \overline{\mathrm{M}}_f(X,\mathcal{I}) \to \mathbb{R}$ as the piecewise-linear extension of the function defined on the nodes of $\overline{\mathrm{M}}_f(X,\mathcal{I})$ by $V_\alpha^i \mapsto \mathrm{mid}(\tilde{I}_\alpha)$, where $\mathrm{mid}(\tilde{I}_\alpha)$ is the midpoint of the proper subinterval $\tilde{I}_\alpha$ of I_α. The definition of $\mathsf{m}_{\mathcal{I}} : \mathrm{M}_f(X,\mathcal{I}) \to \mathbb{R}$ is similar.*

Hence, Reeb graphs can be computed from $\overline{\mathrm{M}}_f(X,\mathcal{I})$ and $\mathrm{M}_f(X,\mathcal{I})$, once they are equipped with $\bar{\mathsf{m}}_{\mathcal{I}}$ and $\mathsf{m}_{\mathcal{I}}$ respectively. Let us call them $\mathrm{R}_{\bar{\mathsf{m}}_{\mathcal{I}}}(\overline{\mathrm{M}}_f(X,\mathcal{I}))$ and $\mathrm{R}_{\mathsf{m}_{\mathcal{I}}}(\mathrm{M}_f(X,\mathcal{I}))$, with corresponding induced maps $\tilde{\bar{\mathsf{m}}}_{\mathcal{I}} : \mathrm{R}_{\bar{\mathsf{m}}_{\mathcal{I}}}(\overline{\mathrm{M}}_f(X,\mathcal{I})) \to \mathbb{R}$ and $\tilde{\mathsf{m}}_{\mathcal{I}} : \mathrm{R}_{\mathsf{m}_{\mathcal{I}}}(\mathrm{M}_f(X,\mathcal{I})) \to \mathbb{R}$. The following lemma, which states that (MultiNerve) Mappers are isomorphic to their Reeb graphs, is a simple consequence of Remark 2.7.

Lemma 4.2. *Let X be a topological space and $f : X \to \mathbb{R}$ be a Morse-type function. Let $\mathcal{I}$ be a gomic of* $\mathrm{im}(f)$. *Then $\overline{\mathrm{M}}_f(X,\mathcal{I})$ and $\mathcal{C}\mathrm{R}_{\bar{\mathsf{m}}_{\mathcal{I}}}(\overline{\mathrm{M}}_f(X,\mathcal{I}))$ are isomorphic as combinatorial multigraphs. The same is true for $\mathrm{M}_f(X,\mathcal{I})$ and $\mathcal{C}\mathrm{R}_{\mathsf{m}_{\mathcal{I}}}(\mathrm{M}_f(X,\mathcal{I}))$.*

Hence, by a slight abuse of notation, we rename $\tilde{\mathsf{m}}_{\mathcal{I}}$ and $\tilde{\bar{\mathsf{m}}}_{\mathcal{I}}$ into $\mathsf{m}_{\mathcal{I}}$ and $\bar{\mathsf{m}}_{\mathcal{I}}$ for convenience.

We now state the main result of this section, which ensures that the extended persistence diagram $\mathrm{Dg}(\bar{\mathsf{m}}_{\mathcal{I}})$, i.e. the bag-of-features signature of $\mathrm{R}_{\bar{\mathsf{m}}_{\mathcal{I}}}(\overline{\mathrm{M}}_f(X,\mathcal{I}))$ and $\overline{\mathrm{M}}_f(X,\mathcal{I})$, is nothing but a simplification of $\mathrm{Dg}(\tilde{f})$, i.e. the bag-of-features signature of $\mathrm{R}_f(X)$.

Theorem 4.3. *Let X be a topological space and $f : X \to \mathbb{R}$ be a Morse-type function. Let $\mathrm{R}_f(X)$ be the corresponding Reeb graph and $\tilde{f} : \mathrm{R}_f(X) \to \mathbb{R}$ be the induced map. Let $\mathcal{I}$ be a gomic of* $\mathrm{im}(f)$. *There are bijections between:*

(i) $\mathrm{Ord}_0(\bar{\mathsf{m}}_{\mathcal{I}})$ *and* $\mathrm{Ord}_0(\tilde{f}) \setminus Q_O^{\mathcal{I}}$ (iii) $\mathrm{Ext}_1^-(\bar{\mathsf{m}}_{\mathcal{I}})$ *and* $\mathrm{Ext}_1^-(\tilde{f}) \setminus Q_{E^-}^{\mathcal{I}}$

(ii) $\mathrm{Rel}_1(\bar{\mathsf{m}}_{\mathcal{I}})$ *and* $\mathrm{Rel}_1(\tilde{f}) \setminus Q_R^{\mathcal{I}}$ (iv) $\mathrm{Ext}_0^+(\bar{\mathsf{m}}_{\mathcal{I}})$ *and* $\mathrm{Ext}_0^+(\tilde{f})$

where $Q_O^{\mathcal{I}} = \bigcup_{I\in\mathcal{I}} Q^+_{\tilde{I}\cup I_\cap^+}$, $Q_R^{\mathcal{I}} = \bigcup_{I\in\mathcal{I}} Q^-_{\tilde{I}\cup I_\cap^-}$, and $Q_{E^-}^{\mathcal{I}} = \bigcup_{I\in\mathcal{I}} Q_I^-$, and where, for any interval I with endpoints $a \leq b$, we let $Q_I^+ = \{(x,y) \in \mathbb{R}^2 : a \leq x \leq y \leq b\}$ be the corresponding half-square above the diagonal, and $Q_I^- = \{(x,y) \in \mathbb{R}^2 : a \leq y < x \leq b\}$ be the half-square strictly below the diagonal. See Figure 8 for an illustration.

The remaining of Section 4.1 is devoted to the proof of Theorem 4.3. In order to state the proof, we first introduce *cover zigzag persistence modules*.

Definition 4.4. *Let X be a topological space and $f : X \to \mathbb{R}$ be a Morse-type function. Let $\mathcal{I} = \{I_\alpha\}_{1\leq\alpha\leq m}$ be a gomic of* $\mathrm{im}(f)$, *sorted by the natural order defined in Section 2.5.*

Let $\mathrm{Crit}(f) = \{-\infty = a_0, a_1, ..., a_n, a_{n+1} = +\infty\}$. *For any open interval I with left endpoint a, we define the integers $l(I)$, $r(I)$ by $l(I) = \max\{i : a_i \leq a\}$ and $r(I) = \max\{l(I), \max\{i : a_i \in I\}\}$. Then, we define the* cover zigzag persistence module $\mathrm{CZZ}(f,\mathcal{I})$ *by*

$$\mathrm{CZZ}(f,\mathcal{I}) = H_*\left(X^{r(I_1)}_{l(I_1)} \hookleftarrow X^{r(I_1\cap I_2)}_{l(I_1\cap I_2)} \hookrightarrow X^{r(I_2)}_{l(I_2)} \hookleftarrow X^{r(I_2\cap I_3)}_{l(I_2\cap I_3)} \hookrightarrow \cdots \hookleftarrow X^{r(I_{m-1}\cap I_m)}_{l(I_{m-1}\cap I_m)} \hookrightarrow X^{r(I_m)}_{l(I_m)}\right),$$

where the X_i^j spaces are as in Section 2.3. We also let $\mathrm{CBc}(f,\mathcal{I})$ *denote the barcode of this module.*

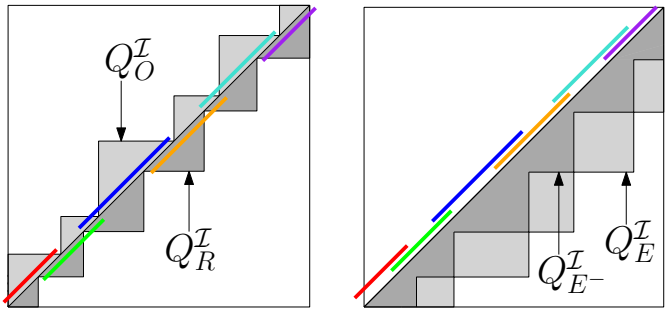


Figure 8: Left: Staircases of ordinary (light grey) and relative (dark grey) types. Right: Staircases of extended types—$Q_{E^-}^{\mathcal{I}}$ is in dark grey while $Q_E^{\mathcal{I}}$ is the union of $Q_{E^-}^{\mathcal{I}}$ with the light grey area.

Note that cover zigzag persistence modules can be isometrically embedded (with the bottleneck distance) into the south face of the Mayer-Vietoris half-pyramid. Indeed, each node of $\mathrm{CZZ}(f,\mathcal{I})$ belongs to this south face. The only difficulty is that $\mathrm{CZZ}(f,\mathcal{I})$ may include the same node several times consecutively when there is a sequence of consecutive intervals in the gomic that are all included between two consecutive critical values of f, i.e. for which $l(I) = r(I)$. However, in that case, the corresponding arrows in the module are isomorphisms. Thus, composing these arrows leaves the resulting barcode unchanged.

Lemma 4.5. *Let X be a topological space and $f : X \to \mathbb{R}$ be a Morse-type function. Let $\mathcal{I}$ be a gomic of* $\mathrm{im}(f)$. *Then, there is a bijection between* $\mathrm{Dg}(\bar{\mathsf{m}}_{\mathcal{I}})$ *and* $\mathrm{CBc}_0(f,\mathcal{I})$.

Proof. Recall from Corollary 2.5 that it suffices to show that $\mathrm{LZZ}_0(\bar{\mathsf{m}}_{\mathcal{I}})$ and $\mathrm{CZZ}_0(f,\mathcal{I})$ are isomorphic as zigzag persistence modules. Assume without loss of generality that $\mathcal{I}$ has m elements, with $m \in \mathbb{N}^*$. First, note that $\mathrm{card}(\mathrm{Crit}(\bar{\mathsf{m}}_{\mathcal{I}}))$ is equal to m. Hence, both $\mathrm{LZZ}(\bar{\mathsf{m}}_{\mathcal{I}})$ and $\mathrm{CZZ}(f,\mathcal{I})$ have exactly $2m+1$ nodes. Moreover, since the MultiNerve Mapper tracks the connected components of the interval and intersection preimages of f, each element of $\mathrm{LZZ}_0(\bar{\mathsf{m}}_{\mathcal{I}})$ is of the form $H_0(f^{-1}(I))$, $I \in \mathcal{I}$, or $H_0(f^{-1}(I \cap J))$, I, J consecutive in $\mathcal{I}$.

Let $I \in \mathcal{I}$. Since f is Morse-type, $X_{l(I)}^{r(I)}$ and $X^I = f^{-1}(I)$ have the same homotopy type. Indeed, recall from Section 2.3 that there exist $s_{l(I)}$ and $s_{r(I)}$ such that $X_{l(I)}^{r(I)} = f^{-1}\left(\left[s_{l(I)}, s_{r(I)}\right]\right)$ and $s_{l(I)}$ (resp. $s_{r(I)}$) and the left (resp. right) endpoint of I are located between the same consecutive critical values of f. In particular, $X_{l(I)}^{r(I)}$ and X^I have the same number of connected components, meaning that $H_0(X^I)$ and $H_0(X_{l(I)}^{r(I)})$ are isomorphic groups. The same is also true for any $I \cap J$, $I, J \in \mathcal{I}$.

Hence, we define a canonical pointwise isomorphism Ψ in dimension 0 as follows: for each node, send each connected component of one preimage, or equivalently each generator of one homology group, to the connected component of the other preimage which intersects it (there is only one since the preimages have the same number of connected components). By definition of the MultiNerve Mapper, Ψ commutes with the canonical inclusion. Hence, $\mathrm{LZZ}_0(\bar{\mathsf{m}}_{\mathcal{I}})$ and $\mathrm{CZZ}_0(f,\mathcal{I})$ are isomorphic. □ □

Finally, we relate the cover zigzag persistence barcode to the extended persistence diagram of the Reeb graph. Namely, we show that a specific simplification of this extended persistence diagram encodes the same information as the cover zigzag persistence barcode.

Theorem 4.3. Again, recall from Corollary 2.5 that $\mathrm{Dg}(\tilde{f})$ encodes the same information as $\mathrm{LBc}_0(\tilde{f})$. Hence, since $\mathrm{Dg}(\bar{\mathsf{m}}_{\mathcal{I}})$ and $\mathrm{CBc}_0(f,\mathcal{I})$ are equivalent from Lemma 4.5, we focus on the relation between $\mathrm{LBc}_0(\tilde{f})$ and $\mathrm{CBc}_0(f,\mathcal{I})$. As mentioned after Definition 4.4, the cover zigzag persistence module $\mathrm{CZZ}(f,\mathcal{I})$ can be isometrically embedded in the south face of the Mayer-Vietoris half-pyramid. Hence, we can assume without loss of generality that the set of nodes of $\mathrm{CZZ}(f,\mathcal{I})$ is a subset of the nodes of a monotone zigzag module $\overline{\mathrm{CZZ}}(f,\mathcal{I})$ that can be drawn along the south face of the Mayer-Vietoris half-pyramid by interpolating the

elements of $\mathrm{CZZ}(f,\mathcal{I})$. Thus, it suffices by Theorem 2.4 to study which intervals disappear when going from $\mathrm{LBc}_0(\tilde{f})$ to $\overline{\mathrm{CBc}}_0(f,\mathcal{I})$ and then to $\mathrm{CBc}_0(f,\mathcal{I})$ using the pyramid rules recalled in Figure 9.

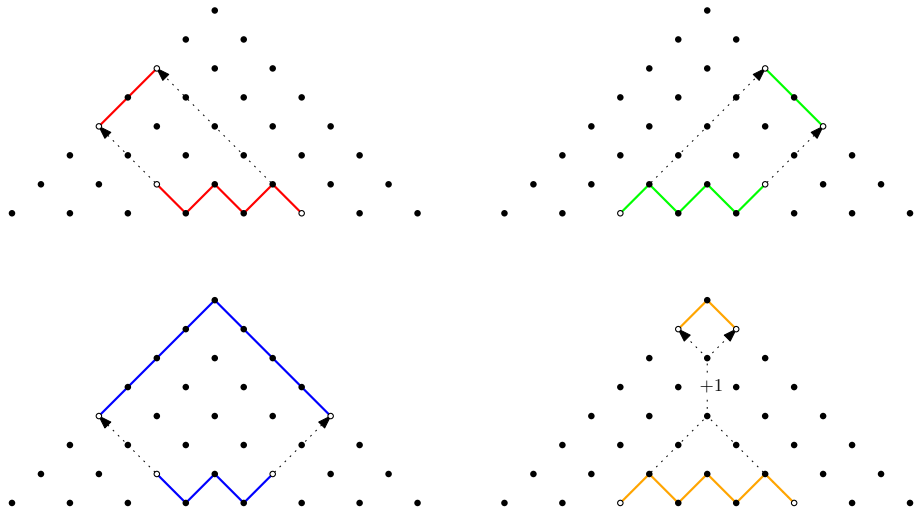


Figure 9: (From [7]) We show the axis of travel of birth and death endpoints of intervals of $\mathrm{LZZ}(f)$ to the up-down zigzag persistence module bounding the south face of the Mayer-Vietoris half-pyramid for interval modules that correspond to type I intervals (upper-left, red), type II intervals (upper-right, green), type III intervals (down-left, blue), and type IV intervals (down-right, orange). The +1 in the down-right figure means that the homological dimension is increased by one.

We first give analogues of staircases for zigzag persistence. For any $I = I_\cap^- \sqcup \tilde{I} \sqcup I_\cap^+ \in \mathcal{I}$, we define:

- $\mathrm{supp}_O(I)$ as the set of nodes of $\mathrm{LZZ}(f)$ that are located strictly between $X_{l(\tilde{I}\cup I_\cap^+)}^{l(\tilde{I}\cup I_\cap^+)}$ and $X_{r(\tilde{I}\cup I_\cap^+)-1}^{r(\tilde{I}\cup I_\cap^+)}$,
- $\mathrm{supp}_R(I)$ as the set of nodes of $\mathrm{LZZ}(f)$ that are located strictly between $X_{l(I_\cap^-\cup\tilde{I})}^{l(I_\cap^-\cup\tilde{I})+1}$ and $X_{r(I_\cap^-\cup\tilde{I})}^{r(I_\cap^-\cup\tilde{I})}$,
- $\mathrm{supp}_{E^-}(I)$ as the set of nodes of $\mathrm{LZZ}(f)$ that are located strictly between $X_{l(I)}^{l(I)+1}$ and $X_{r(I)-1}^{r(I)}$.

There are two possible ways for an interval of $\mathrm{LBc}_0(f)$ to disappear in $\mathrm{CBc}_0(f,\mathcal{I})$: either its homological dimension is shifted by 1, or its intersection with the set of nodes of $\mathrm{CZZ}(f,\mathcal{I})$ is empty after being projected onto $\overline{\mathrm{CBc}}_0(f,\mathcal{I})$—see Figure 10. According to the pyramid rules, we have that:

- Projections of type III intervals of $\mathrm{LBc}_0(f)$ onto $\overline{\mathrm{CBc}}_0(f,\mathcal{I})$ always intersect with the nodes of $\mathrm{CZZ}(f,\mathcal{I})$ and their homological dimensions cannot be shifted. Hence, none of them disappears. This proves (iv).
- Projections of type IV intervals of $\mathrm{LBc}_0(f)$ onto $\overline{\mathrm{CBc}}_0(f,\mathcal{I})$ always intersect with the nodes of $\mathrm{CZZ}(f,\mathcal{I})$. However, their homological dimensions can be shifted by 1. This happens when the endpoints collide in the south face of the Mayer-Vietoris half-pyramid. Hence, only those intervals whose support is included in $\mathrm{supp}_{E^-}(I)$ for some $I \in \mathcal{I}$ go through such a shift before getting to $\overline{\mathrm{CBc}}_0(f,\mathcal{I})$. This proves (iii).
- Homological dimensions of type I intervals in $\mathrm{LBc}_0(f)$ cannot be shifted, but their projections onto $\overline{\mathrm{CBc}}_0(f,\mathcal{I})$ may not always intersect with the nodes of $\mathrm{CZZ}(f,\mathcal{I})$. This happens for those intervals whose support is included in $\mathrm{supp}_O(I)$ for some $I \in \mathcal{I}$, thus proving (i).
- Homological dimensions of type II intervals in $\mathrm{LBc}_0(f)$ cannot be shifted, but their projections onto $\overline{\mathrm{CBc}}_0(f,\mathcal{I})$ may not always intersect with the nodes of $\mathrm{CZZ}(f,\mathcal{I})$. This happens for those intervals whose support is included in $\mathrm{supp}_R(I)$ for some $I \in \mathcal{I}$, thus proving (ii).

□ □

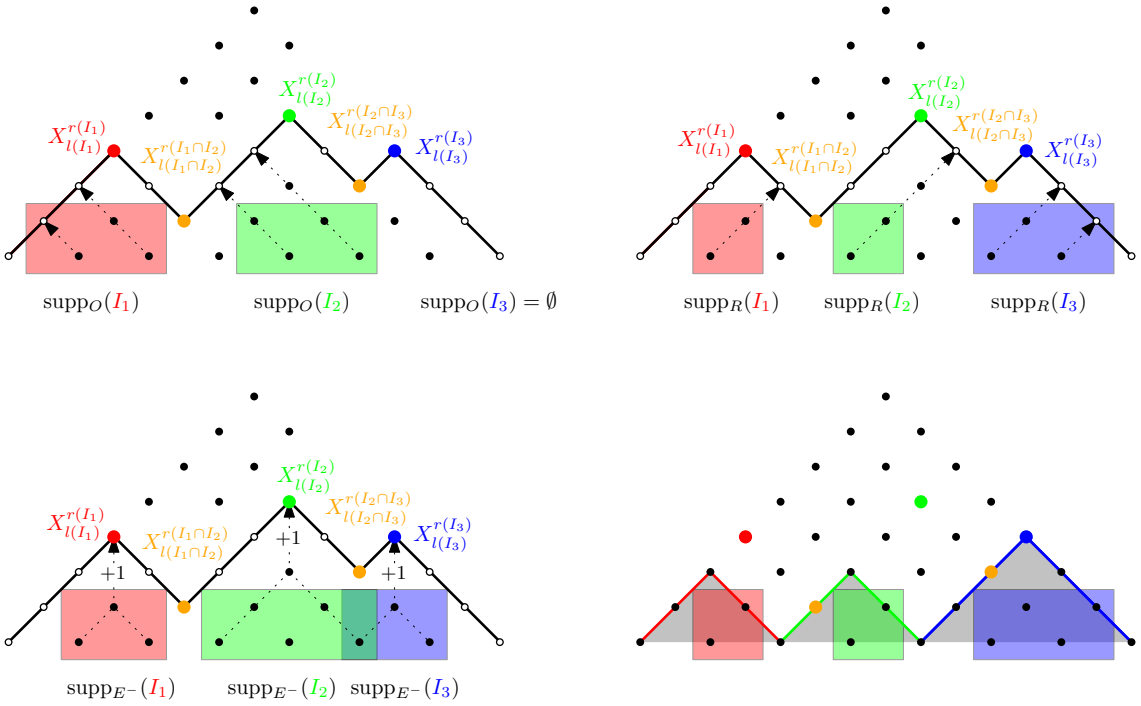


Figure 10: The black path in the south face of the Mayer-Vietoris half-pyramid represents the monotone zigzag persistence module $\overline{\mathrm{CZZ}}(f,\mathcal{I})$ for a gomic $\mathcal{I}$ with three intervals. The white disks on this path are the nodes that do not intersect the set of nodes of the cover zigzag persistence module $\mathrm{CZZ}(f,\mathcal{I})$, which are colored according to the interval of $\mathcal{I}$ they represent (and are colored orange if they represent an intersection). The boxes outline the support of the intervals of $\mathrm{LBc}_0(f)$ that disappear in the MultiNerve Mapper depending on their types (upper-left for type I intervals, upper-right for type II intervals and down-left for type IV intervals). We also show (down-right) the analogue, drawn in grey color, of $Q^{\mathcal{I}}_R$ on the south face of the Mayer-Vietoris half-pyramid.

4.2 A signature for MultiNerve Mapper

Theorem 4.3 means that the dictionary introduced in Section 2.4 can be used to describe the structure of the MultiNerve Mapper from the extended persistence diagram of the induced function $\tilde{f}$. Indeed, the topological features of $\overline{\mathrm{M}}_f(X,\mathcal{I})$ are in bijection with the points of $\mathrm{Dg}(\tilde{f})$ minus the ones that fall into the various staircases ($Q^{\mathcal{I}}_O$, $Q^{\mathcal{I}}_{E^-}$, $Q^{\mathcal{I}}_R$) corresponding to their type. Moreover, by Theorem 2.8, $\mathrm{Dg}(\tilde{f})$ itself is obtained from $\mathrm{Dg}_0(f)$ and $\mathrm{Dg}_1(f)$ by removing the points of $\mathrm{Ext}^+_1(f)$ and $\mathrm{Ord}_1(f)$. Hence, we use the off-staircase part of $\mathrm{Dg}(\tilde{f})$ as a signature for the structure of the MultiNerve Mapper[3]:

$$\begin{aligned}\mathrm{Dg}(\overline{\mathrm{M}}_f(X,\mathcal{I})) &= (\mathrm{Ord}(\tilde{f})\setminus Q^{\mathcal{I}}_O)\cup(\mathrm{Ext}(\tilde{f})\setminus Q^{\mathcal{I}}_{E^-})\cup(\mathrm{Rel}(\tilde{f})\setminus Q^{\mathcal{I}}_R)\\ &= (\mathrm{Ord}_0(f)\setminus Q^{\mathcal{I}}_O)\cup((\mathrm{Ext}^+_0(f)\cup\mathrm{Ext}^-_1(f))\setminus Q^{\mathcal{I}}_{E^-})\cup(\mathrm{Rel}_1(f)\setminus Q^{\mathcal{I}}_R).\end{aligned} \tag{5}$$

We call this signature the *extended persistence diagram* of the MultiNerve Mapper. Note that this signature is not computed by applying persistence to some function defined on the multinerve, but it is rather a pruned version of the extended persistence diagram of $\tilde{f}$. As for Reeb graphs, it serves as a bag-of-features type signature of the structure of $\overline{\mathrm{M}}_f(X,\mathcal{I})$. Moreover, the fact that $\mathrm{Dg}(\overline{\mathrm{M}}_f(X,\mathcal{I}))\subseteq\mathrm{Dg}(\tilde{f})$ formalizes the intuition that the MultiNerve Mapper should be viewed as a *pixelized version* of the Reeb graph, in which some of the features disappear due to the staircases (prescribed by the cover). For instance, in Figure 11 we

[3]Recall that $\mathrm{Ext}^-_0(f)=\mathrm{Rel}_0(f)=\emptyset$.

show a double torus equipped with the height function, together with its associated Reeb graph, MultiNerve Mapper, and Mapper. We also show the corresponding extended persistence diagrams. In each case, the points in the diagram represent the features of the object: the extended points represent the holes (dimension 1 and above) and the trunks (dimension 0) while the ordinary and relative points represent the branches.

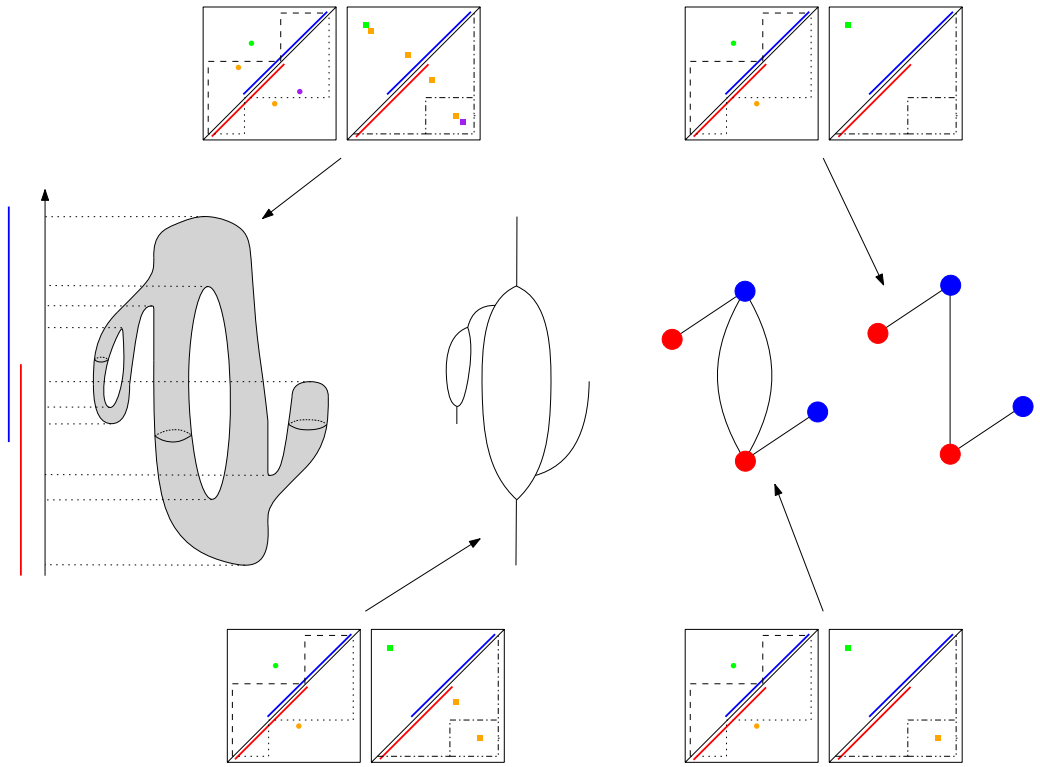


Figure 11: From left to right: a 2-manifold equipped with the height function; the corresponding Reeb graph, MultiNerve Mapper, and Mapper. For each object, we display the extended persistence diagrams of dimension 0 (green points), 1 (orange points) and 2 (purple points). Extended points are squares while ordinary and relative points are disks (above and below the diagonal respectively). The staircases are represented with dashed ($Q_O^{\mathcal{I}}$), dotted ($Q_{E^-}^{\mathcal{I}}$), dash-dotted ($Q_R^{\mathcal{I}}$), and dash-dot-dotted ($Q_E^{\mathcal{I}}$) lines. One can see how to go from the extended persistence diagram of the height function to the one of the induced map (remove the points in dimension 2 and the points in dimension 1 above the diagonal), then to the one of the MultiNerve Mapper (remove the points inside the staircases corresponding to their type), and finally, to the one of the Mapper (remove the extended points in $Q_E^{\mathcal{I}}$).

Convergence of the signature. The following convergence result (which is in fact non-asymptotic) is a direct consequence of our previous results:

Corollary 4.6. *Suppose the granularity of the gomic $\mathcal{I}$ is at most ε. Then,*

$$\mathrm{Dg}(\tilde{f}) \setminus \{(x,y) \,:\, |y-x| \leq \varepsilon\} \subseteq \mathrm{Dg}(\overline{\mathrm{M}}_f(X,\mathcal{I})) \subseteq \mathrm{Dg}(\tilde{f}).$$

Thus, the features (branches, holes) of the Reeb graph that are missing in the MultiNerve Mapper have spans at most ε. In particular, we have $d_{\mathrm{b}}(\mathrm{Dg}(\overline{\mathrm{M}}_f(X,\mathcal{I})), \mathrm{Dg}(\tilde{f})) \leq \varepsilon/2$. Moreover, the two signatures become equal when ε becomes smaller than the smallest vertical distance of the points of $\mathrm{Dg}(\tilde{f})$ to the diagonal. Finally, $\overline{\mathrm{M}}_f(X,\mathcal{I})$ and $\mathrm{R}_f(X)$ themselves become isomorphic as combinatorial graphs up to one-step vertex

splits and edge subdivisions (which are topologically trivial modifications) when ε becomes smaller than the smallest absolute difference between distinct critical values of f.

We show a similar convergence result in the functional distortion distance in Section 7. Note that building the signature $\mathrm{Dg}(\overline{\mathrm{M}}_f(X,\mathcal{I}))$ requires computing the critical values of f exactly, which may not always be possible. However, as for Reeb graphs, the signature can be approximated efficiently and with theoretical guarantees under mild sampling conditions using existing work on scalar fields analysis, as we will see in Section 8.

4.3 Induced signature for Mapper

Recall from Lemma 3.6 that the projection $\pi_1 : \overline{\mathrm{M}}_f(X,\mathcal{I}) \to \mathrm{M}_f(X,\mathcal{I})$ induces a surjective homomorphism in homology. Thus, the Mapper has a simpler structure than the MultiNerve Mapper. To be more specific, π_1 identifies all the edges connecting the same pair of vertices. This eliminates the corresponding holes in $\overline{\mathrm{M}}_f(X,\mathcal{I})$. Since the two vertices lie in successive intervals of the cover, the corresponding diagram points lie in the following extended staircase (see the staircase $Q_E^{\mathcal{I}}$ displayed on the right in Figure 8):

$$Q_E^{\mathcal{I}} = \bigcup_{I\cup J \text{ such that } I\cap J\neq\emptyset} Q^-_{I\cup J}.$$

The other staircases remain unchanged. Hence the following signature:

$$\begin{aligned}\mathrm{Dg}(\mathrm{M}_f(X,\mathcal{I})) &= (\mathrm{Ord}(\tilde{f}) \setminus Q_O^{\mathcal{I}}) \cup (\mathrm{Ext}(\tilde{f}) \setminus Q_E^{\mathcal{I}}) \cup (\mathrm{Rel}(\tilde{f}) \setminus Q_R^{\mathcal{I}}) \\ &= (\mathrm{Ord}_0(f) \setminus Q_O^{\mathcal{I}}) \cup ((\mathrm{Ext}_0^+(f) \cup \mathrm{Ext}_1^-(f)) \setminus Q_E^{\mathcal{I}}) \cup (\mathrm{Rel}_1(f) \setminus Q_R^{\mathcal{I}}).\end{aligned} \tag{6}$$

The interpretation of this signature in terms of the structure of the Mapper follows the same rules as for the MultiNerve Mapper and Reeb graph—see again Figure 11. Moreover, the convergence result stated in Corollary 4.6 holds for the Mapper as well.

5 Stability in the bottleneck distance

Intuitively, for a point in the signature $\mathrm{Dg}(\overline{\mathrm{M}}_f(X,\mathcal{I}))$, the ℓ^∞-distance to its corresponding staircase[4] measures the amount by which the function f or the cover $\mathcal{I}$ must be perturbed in order to eliminate the corresponding feature (branch, hole) in the MultiNerve Mapper. Conversely, for a point in the Reeb graph's signature $\mathrm{Dg}(\tilde{f})$ that is not in the MultiNerve Mapper's signature (i.e. that lies inside its corresponding staircase), the ℓ^∞-distance to the boundary of the staircase measures the amount by which f or $\mathcal{I}$ must be perturbed in order to create a corresponding feature in the MultiNerve Mapper. Our goal here is to formalize this intuition. For this we adapt the bottleneck distance so that it takes the staircases into account. Our results are stated for the MultiNerve Mapper, they hold the same for the Mapper with the staircase $Q_{E^-}^{\mathcal{I}}$ replaced by its extension $Q_E^{\mathcal{I}}$.

An extension of the bottleneck distance. Let Θ be a subset of $\mathbb{R}^2$. Given a partial matching Γ between two extended persistence diagrams $\mathrm{Dg}, \mathrm{Dg}'$, the Θ-*cost* of Γ is:

$$\mathrm{cost}_\Theta(\Gamma) = \max\left\{\max_{p\in\mathrm{Dg}} \delta_{\mathrm{Dg}}(p),\ \max_{p'\in\mathrm{Dg}'} \delta_{\mathrm{Dg}'}(p')\right\},$$

where:

$$\delta_{\mathrm{Dg}}(p) = \|p-p'\|_\infty \text{ if } \exists p' \in \mathrm{Dg}' \text{ such that } (p,p') \in \Gamma \text{ and } d_\infty(p,\Theta) \text{ otherwise,}$$

$$\delta_{\mathrm{Dg}'}(p') = \|p-p'\|_\infty \text{ if } \exists p \in \mathrm{Dg} \text{ such that } (p,p') \in \Gamma \text{ and } d_\infty(p',\Theta) \text{ otherwise.}$$

[4] $Q_O^{\mathcal{I}}$, $Q_{E^-}^{\mathcal{I}}$ or $Q_R^{\mathcal{I}}$, depending on the type of the point.

The bottleneck distance becomes:

$$d_{\mathrm{b},\Theta}(\mathrm{Dg},\mathrm{Dg}') = \inf_{\Gamma}\ \mathrm{cost}_{\Theta}(\Gamma),$$

where Γ ranges over all partial matchings between Dg and Dg$'$. This is again a pseudometric and not a metric. Note that the usual bottleneck distance is obtained by taking Θ to be the diagonal Δ. Given a gomic $\mathcal{I}$, we choose different sets Θ depending on the types of the points in the two diagrams. More precisely, we define the distance between signatures as follows:

Definition 5.1. *Given a gomic $\mathcal{I}$, we define the distance $d_{\mathcal{I}}$ between extended persistence diagrams* Dg, Dg$'$ *as:*

$$d_{\mathcal{I}}(\mathrm{Dg},\mathrm{Dg}') = \max\left\{d_{\mathrm{b},\mathrm{Q}^{\mathcal{I}}_{\mathrm{O}}}(\mathrm{Ord},\mathrm{Ord}'),\ d_{\mathrm{b},\mathrm{Q}^{\mathcal{I}}_{\mathrm{E}^-}}(\mathrm{Ext},\mathrm{Ext}'),\ d_{\mathrm{b},\mathrm{Q}^{\mathcal{I}}_{\mathrm{R}}}(\mathrm{Rel},\mathrm{Rel}')\right\}. \tag{7}$$

5.1 Stability with respect to perturbations of the function

The distance $d_{\mathcal{I}}$ *stabilizes* the (MultiNerve) Mappers, as stated in the following theorem:

Theorem 5.2. *Given a topological space X, Morse-type functions $f, g : X \to \mathbb{R}$ and a gomic $\mathcal{I}$ of granularity at most $\epsilon > 0$, the following stability inequality holds:*

$$d_{\mathcal{I}}(\mathrm{Dg}(\mathrm{M}_f(X,\mathcal{I})),\mathrm{Dg}(\mathrm{M}_g(X,\mathcal{I}))) \leq d_{\mathcal{I}}(\mathrm{Dg}(\overline{\mathrm{M}}_f(X,\mathcal{I})),\mathrm{Dg}(\overline{\mathrm{M}}_g(X,\mathcal{I}))) \leq \|f-g\|_\infty. \tag{8}$$

Moreover, $d_{\mathcal{I}}$ and d_{b} are related as follows:

$$d_{\mathrm{b}}(\mathrm{Dg}(\overline{\mathrm{M}}_f(X,\mathcal{I})),\mathrm{Dg}(\overline{\mathrm{M}}_g(X,\mathcal{I}))) \leq \frac{\varepsilon}{2} + d_{\mathcal{I}}(\mathrm{Dg}(\overline{\mathrm{M}}_f(X,\mathcal{I})),\mathrm{Dg}(\overline{\mathrm{M}}_g(X,\mathcal{I}))). \tag{9}$$

$$d_{\mathrm{b}}(\mathrm{Dg}(\mathrm{M}_f(X,\mathcal{I})),\mathrm{Dg}(\mathrm{M}_g(X,\mathcal{I}))) \leq \varepsilon + d_{\mathcal{I}}(\mathrm{Dg}(\mathrm{M}_f(X,\mathcal{I})),\mathrm{Dg}(\mathrm{M}_g(X,\mathcal{I}))). \tag{10}$$

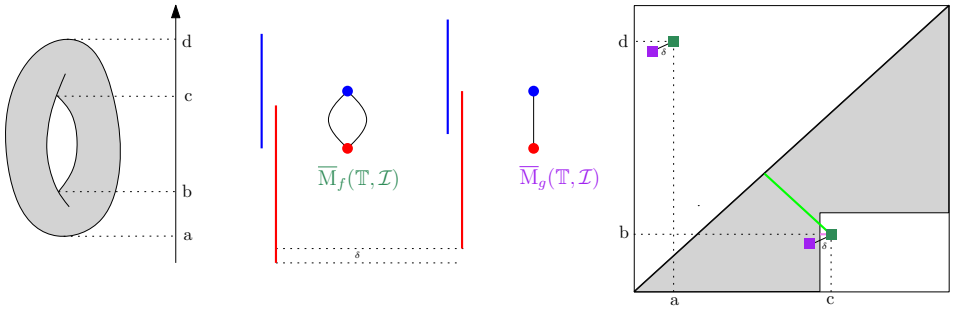


Figure 12: We compute the MultiNerve Mapper of the height function f on the torus $\mathbb{T}$, given a gomic $\mathcal{I}$ with two intervals. We also compute the MultiNerve Mapper of a perturbed function g such that $\|f-g\|_\infty \leq \delta$. We plot the extended persistence diagrams of $\tilde{f}$ (dark green) and $\tilde{g}$ (purple). Note that the signature of $\overline{\mathrm{M}}_g(\mathbb{T},\mathcal{I})$ is obtained by removing the purple point beneath the diagonal since it belongs to a staircase, while the signature of $\overline{\mathrm{M}}_f(\mathbb{T},\mathcal{I})$ is equal to $\mathrm{Dg}(\tilde{f})$. If we used the bottleneck distance to compare the two signatures, their distance would be equal to the distance to the diagonal of the dark green point beneath Δ (green segment), which can be arbitrarily large, while, using $d_{\mathcal{I}}$, their distance becomes the distance of the same point to the staircase (tiny pink segment), which is bounded by δ.

The proof of Theorem 5.2 relies on the following monotonicity property, which is immediate:

Lemma 5.3. *Let $\Theta \subseteq \mathbb{R}^2$ be in the closure of $\Theta' \subseteq \mathbb{R}^2$. Then,*

$$d_{\Theta'}(\mathrm{Dg},\mathrm{Dg}') \leq d_{\Theta}(\mathrm{Dg},\mathrm{Dg}') \leq d_{\Theta'}(\mathrm{Dg},\mathrm{Dg}') + d_{\mathrm{H}}(\Theta,\Theta'),$$

where d_{H} denotes the Hausdorff distance in the ℓ_∞-norm.

Theorem 5.2. Equation (9) and (10) are direct applications of Lemma 5.3. Equation (8) is proven by the following sequence of (in)equalities:

$$\begin{aligned} d_{\mathcal{I}}(\mathrm{Dg}(\overline{\mathrm{M}}_f(X,\mathcal{I})), \mathrm{Dg}(\overline{\mathrm{M}}_g(X,\mathcal{I}))) &= d_{\mathcal{I}}(\mathrm{Dg}(\tilde{f}), \mathrm{Dg}(\tilde{g})) \\ &\leq d_{\mathrm{b},\Delta}(\mathrm{Dg}(\tilde{f}), \mathrm{Dg}(\tilde{g})) = d_{\mathrm{b}}(\mathrm{Dg}(\tilde{f}), \mathrm{Dg}(\tilde{g})) \\ &\leq d_{\mathrm{b}}(\mathrm{Dg}(f), \mathrm{Dg}(g)) \\ &\leq \|f-g\|_\infty. \end{aligned}$$

The first equality comes from the observation that the points of $\mathrm{Dg}(\tilde{f}) \sqcup \mathrm{Dg}(\tilde{g})$ that lie inside their corresponding staircase can be left unmatched and have a zero cost in the matching, so removing them as in (5) does not change the bottleneck cost. The first inequality follows from Lemma 5.3 since the diagonal Δ is included in the closure of each of the staircases. The second inequality follows from Theorem 2.8 and the fact that the matchings only match points of the same type (ordinary, extended, relative) and of the same homological dimension. The last inequality comes from Theorem 2.3. □ □

Interpretation of the stability. Note that the bottleneck distance d_{b} is unstable in this context—see Figure 12. The theorem allows us to make some interesting claims. For instance, denoting by $Q_p^{\mathcal{I}}$ the staircase corresponding to the type of a diagram point p, the quantity

$$d_{\mathcal{I}}(\mathrm{Dg}, \emptyset) = \max_{p \in \mathrm{Dg}} d_\infty(p, Q_p^{\mathcal{I}})$$

measures the amount by which the diagram Dg must be perturbed in the metric $d_{\mathcal{I}}$ in order to bring all its points to the staircase. Hence, by Theorem 5.2, given a pair (X, f), the quantity

$$d_{\mathcal{I}}(\mathrm{Dg}(\overline{\mathrm{M}}_f(X,\mathcal{I})), \emptyset) = \max_{p \in \mathrm{Dg}(\overline{\mathrm{M}}_f(X,\mathcal{I}))} d_\infty(p, Q_p^{\mathcal{I}})$$

is a lower bound on the amount by which f must be perturbed in the supremum norm in order to remove all the features (branches and holes) from the MultiNerve Mapper. Conversely,

$$\min_{p \in \mathrm{Dg}(\overline{\mathrm{M}}_f(X,\mathcal{I}))} d_\infty(p, Q_p^{\mathcal{I}})$$

is a lower bound on the maximum amount of perturbation allowed for f if one wants to preserve all the features in the MultiNerve Mapper no matter what. Note that this does not prevent other features from appearing. The quantity that controls those is related to the points of $\mathrm{Dg}(\tilde{f})$ (including diagonal points) that lie in the staircases. More precisely, the quantity

$$\min_{p \in \mathrm{Dg}(\tilde{f}) \cup \Delta} d_\infty(p, \partial Q_p^{\mathcal{I}} \setminus \Delta)$$

is a lower bound on the maximum amount by which f can be perturbed if one wants to preserve the structure (set of features) of the MultiNerve Mapper no matter what. Note that this lower bound is in fact zero since $\partial Q_O^{\mathcal{I}} \setminus \Delta$ and $\partial Q_R^{\mathcal{I}} \setminus \Delta$ come arbitrarily close to the diagonal Δ (recall Figure 8). This means that, as small as the perturbation of f may be, it can always make new branches appear in the MultiNerve Mapper. However, it will not impact the set of holes if its amplitude is less than

$$\min_{p \in \mathrm{Ext}(\tilde{f}) \cup \Delta} d_\infty(p, \partial Q_{E^-}^{\mathcal{I}} \setminus \Delta).$$

From this discussion we derive the following rule of thumb: having small overlaps between the intervals of the gomic helps capture more features (branches and holes) of the Reeb graph in the (MultiNerve) Mapper; conversely, having large overlaps helps prevent new holes from appearing in the (MultiNerve) Mapper under small perturbations of the function. This is an important trade-off to consider in applications.

5.2 Stability with respect to perturbations of the domain

More generally, we can derive a stability result for perturbations of the pair (X, f), provided we make some extra assumptions on the regularity of the domain and function. Typically, we will assume X to be a compact Riemannian manifold (or, more generally, a compact length space with curvature bounded above) and f to be Lipschitz-continuous. To measure the amount of perturbation of the domain we use the concept of *correspondence* from metric geometry: given another pair (Y, g), a correspondence is a subset C of the product space $X \times Y$ such that the canonical projections $C \to X$ and $C \to Y$ are surjective. We consider the *functional distortion* associated with C, which is the quantity:

$$\varepsilon_{\mathfrak{f}}(C) = \sup_{(x,y)\in C} |f(x) - g(y)|.$$

Similarly, writing respectively d_X and d_Y for the intrinsic metrics of X and Y, we consider the *metric distortion* of C:

$$\varepsilon_{\mathfrak{m}}(C) = \sup_{(x,y)\in C,(x',y')\in C} |d_X(x, x') - d_Y(y, y')|.$$

The *Gromov-Hausdorff distance* between X and Y is then:

$$d_{\mathrm{GH}}(X, Y) = \frac{1}{2} \inf_C \varepsilon_{\mathfrak{m}}(C),$$

where C ranges over all correspondences between X and Y. Now we can derive a stability guarantee for the signatures of MultiNerve Mappers in this context, using a variant of Theorem 2.3 proven in [11]:

Theorem 5.4. *Fix a gomic $\mathcal{I}$. Let X and Y be two compact Riemannian manifolds or length spaces with curvature bounded above. Denote by $\rho(X)$ and $\rho(Y)$ their respective convexity radii (i.e. the smallest radius for which any geodesic ball is convex). Let $f : X \to \mathbb{R}$ and $g : Y \to \mathbb{R}$ be Lipschitz-continuous Morse-type functions, with Lipschitz constants c_f and c_g respectively. Assume $d_{\mathrm{GH}}(X, Y) \leq \frac{1}{20} \min(\rho(X), \rho(Y))$. Then, for any correspondence $C \in \mathcal{C}(X, Y)$ such that $\epsilon_{\mathfrak{m}}(C) < \frac{1}{10} \min(\rho(X), \rho(Y))$,*

$$d_{\mathcal{I}}(\mathrm{Dg}(\overline{\mathrm{M}}_f(X, \mathcal{I})), \mathrm{Dg}(\overline{\mathrm{M}}_g(Y, \mathcal{I}))) \leq (9(c_f + c_g) + \min\{c_f, c_g\})\epsilon_{\mathfrak{m}}(C) + \epsilon_{\mathfrak{f}}(C).$$

Proof. The proof is the same sequence of (in)equalities as for Theorem 5.2, except the last inequality is replaced by $d_\Delta(\mathrm{Dg}(f), \mathrm{Dg}(g)) \leq (9(c_f + c_g) + \min\{c_f, c_g\})\epsilon_{\mathfrak{m}}(C) + \epsilon_{\mathfrak{f}}(C)$, which comes[5] from Theorem 3.4 in [11]. □ □

This result brings about the same discussion as in Section 5, with f replaced by the pair (X, f).

6 Stability with respect to perturbations of the cover

Let us now fix the pair (X, f) and consider varying gomics. For each choice of gomic, Eqs. (5)-(6) tell which points of the diagram $\mathrm{Dg}(f)$ end up in the diagram of the (MultiNerve) Mapper and thus participate in its structure. We aim for a quantification of the extent to which this structure may change as the gomic is perturbed. For this we adopt the dual point of view: for any two choices of gomics, we want to use the points of the diagram $\mathrm{Dg}(f)$ to assess the degree by which the gomics differ. This is a reversed situation compared to Section 5, where the gomic was fixed and was used to assess the degree by which the persistence diagrams of two functions differed.

[5]Note that Theorem 3.4 in [11] is stated only for the ordinary part of the persistence diagrams, however its analysis extends to the full extended filtrations at no extra cost.

A distance between gomics. The diagram points that discriminate between the two gomics are the ones located in the symmetric difference of the staircases, since they witness that the symmetric difference is non-empty. Moreover, their ℓ^∞-distances to the staircase of the other gomic provide a lower bound on the Hausdorff distance between the two staircases and thus quantify the extent to which the two covers differ. We formalize this intuition as follows: given a persistence diagram Dg and two gomics $\mathcal{I}, \mathcal{J}$, we consider the quantity:

$$d_{\mathrm{Dg}}(\mathcal{I}, \mathcal{J}) = \max_{*\in\{O,E^-,R\}} \left\{ \sup_{p\in \mathrm{Dg}^*\cap(Q_*^{\mathcal{I}} \triangle Q_*^{\mathcal{J}})} \max\left\{d_\infty(p, Q_*^{\mathcal{I}}),\, d_\infty(p, Q_*^{\mathcal{J}})\right\} \right\}, \tag{11}$$

where $\triangle$ denotes the symmetric difference, where Dg^* stands for the subdiagram of Dg of the right type (Ord, Ext or Rel), and where we adopt the convention that $\sup_{p\in\emptyset} ...$ is zero instead of infinite. Note that there is always one of the two terms in (11) that is zero since the supremum is taken over all points that lie in the symmetric difference of the staircases. Deriving an upper bound on $d_{\mathrm{Dg}}(\mathcal{I}, \mathcal{J})$ in terms of the Hausdorff distances between the staircases is straightforward, since the supremum in (11) is taken over points that lie in the symmetric difference between the staircases:

$$d_{\mathrm{Dg}}(\mathcal{I}, \mathcal{J}) \leq \max_{*\in\{O,E^-,R\}} d_{\mathrm{H}}(Q_*^{\mathcal{I}},\, Q_*^{\mathcal{J}}),$$

where d_{H} stands for the Hausdorff distance in the ℓ^∞-norm. The connection to the MultiNerve Mapper appears when we take Dg to be the persistence diagram of the induced map $\tilde{f}$ defined on the Reeb graph $\mathrm{R}_f(X)$. Indeed, we have

$$\mathrm{Ord}(\tilde{f}) \cap (Q_O^{\mathcal{I}} \triangle Q_O^{\mathcal{J}}) = (\mathrm{Ord}(\tilde{f}) \cap Q_O^{\mathcal{I}}) \triangle (\mathrm{Ord}(\tilde{f}) \cap Q_O^{\mathcal{J}}) = \mathrm{Ord}(\overline{\mathrm{M}}_f(X, \mathcal{I})) \triangle \mathrm{Ord}(\overline{\mathrm{M}}_f(X, \mathcal{J})),$$

where the second equality follows from the definition of the signature of the MultiNerve Mapper given in (5). Similar equalities can be derived with Ext and Rel. Thus, $d_{\mathrm{Dg}(\tilde{f})}(\mathcal{I}, \mathcal{J})$ quantifies the proximity of each signature to the other staircase. In particular, having $d_{\mathrm{Dg}(\tilde{f})}(\mathcal{I}, \mathcal{J}) = 0$ means that there are no diagram points in the symmetric difference, so the two gomics are equivalent from the viewpoint of the structure of the MultiNerve Mapper. Differently, having $d_{\mathrm{Dg}(\tilde{f})}(\mathcal{I}, \mathcal{J}) > 0$ means that the structures of the two MultiNerve Mappers differ, and the value of $d_{\mathrm{Dg}(\tilde{f})}(\mathcal{I}, \mathcal{J})$ quantifies by how much the covers should be perturbed to make the two multigraphs isomorphic. Furthermore, we have the following upper bound on this quantity:

Theorem 6.1. *Given a Morse-type function $f : X \to \mathbb{R}$, for any gomics $\mathcal{I}, \mathcal{J}$,*

$$d_{\mathrm{Dg}(\tilde{f})}(\mathcal{I}, \mathcal{J}) \leq \max_{*\in\{O,E^-,R\}} d_{\mathrm{H}}(Q_*^{\mathcal{I}},\, Q_*^{\mathcal{J}}).$$

Tightness. It is easy to build examples where the upper bound is tight, for instance by placing a diagram point at a corner of one of the staircases[6]. On the other hand, there are obvious cases where the bound is not tight, for instance we have $d_{\mathrm{Dg}(\tilde{f})}(\mathcal{I}, \mathcal{J}) = 0$ as soon as there are no diagram points in the symmetric difference, whereas the symmetric difference itself may not be empty. What the upper bound measures depends on the subdiagram. For instance, for $* = E^-$, we defined $Q_{E^-}^{\mathcal{I}}$ to be the set $\bigcup_{(a,b)\in\mathcal{I}}\{(x,y) \in \mathbb{R}^2 : a \leq y < x \leq b\}$, so $d_{\mathrm{H}}(Q_{E^-}^{\mathcal{I}},\, Q_{E^-}^{\mathcal{J}})$ measures the supremum of the differences between the intervals in one cover to their closest interval in the other cover:

$$d_{\mathrm{H}}(Q_{E^-}^{\mathcal{I}},\, Q_{E^-}^{\mathcal{J}}) = \max\left\{ \sup_{(a,b)\in\mathcal{I}} \inf_{(c,d)\in\mathcal{J}} \max\{|a-c|,\, |b-d|\},\ \sup_{(c,d)\in\mathcal{J}} \inf_{(a,b)\in\mathcal{I}} \max\{|a-c|,\, |b-d|\} \right\}.$$

Similar formulas can be derived for the other subdiagrams.

[6]Which is easily done by choosing suitable critical values as coordinates for this point.

7 Convergence in the functional distortion distance

Since d_{b} is merely a pseudometric, the relationship between the (MultiNerve) Mapper and the Reeb graph is only partially explained by Theorem 4.3. In this section, we bound the *functional distortion distance* d_{FD} (a true distance between metric graphs equipped with continuous functions) between the (MultiNerve) Mapper and the Reeb graph, and we provide an alternative proof of Theorem 4.3 as a byproduct. To this end, we connect the (MultiNerve) Mapper and the Reeb graph through a sequence of metric spaces on which we can control the functional distortion distance. This connection has an interest in its own right, as it was leveraged in other contributions on Mappers and Reeb graphs recently—see e.g. [9, 10].

7.1 Telescopes and Operators

In this section we introduce the telescopes, which are our main objects of study when we relate the MultiNerve Mapper to the Reeb graph.

Recall that, given topological spaces X and $A \subseteq Y$ together with a continuous map $f : A \to X$, the *adjunction space* $X \cup_f Y$ (also denoted $Y \cup_f X$) is the quotient of the disjoint union $X \amalg Y$ by the equivalence relation induced by the identifications $\{f(a) \sim a\}_{a \in A}$.

Definition 7.1 (Telescope [7])**.** *A* telescope *is an adjunction space of the following form:*

$$T = (Y_0 \times (a_0, a_1]) \cup_{\psi_0} (X_1 \times \{a_1\}) \cup_{\phi_1} (Y_1 \times [a_1, a_2]) \cup_{\psi_1} \ldots \cup_{\phi_n} (Y_n \times [a_n, a_{n+1})),$$

where $-\infty = a_0 < a_1 < \cdots < a_n < a_{n+1} = +\infty$*, and where the* $\phi_i : Y_i \times \{a_i\} \to X_i \times \{a_i\}$ *and* $\psi_i : Y_i \times \{a_{i+1}\} \to X_{i+1} \times \{a_{i+1}\}$ *are continuous maps. The* a_i *are called the* critical values *of* T *and their set is denoted by* $\mathrm{Crit}(T)$*, the* ϕ_i *and* ψ_i *are called* attaching maps*, the* Y_i *are compact and locally connected spaces called the* cylinders *and the* X_i *are topological spaces called the* critical slices*. Moreover, all* Y_i *and* X_i *have finitely-generated homology.*

Extended persistence diagram. A telescope comes equipped with functions π_1 and π_2, which are the projections onto the first factor and second factor respectively. From now on, given any interval I, we let T^I denote $\pi_1 \circ \pi_2^{-1}(I)$. Then, the extended persistence diagram $\mathrm{Dg}(\pi_2)$ can be described using the following Lemma.

Lemma 7.2. *Since* ϕ_i *and* ψ_i *are continuous,*

$$\forall \alpha \in [a_i, a_{i+1}),\ T^{(-\infty,\alpha]} \text{ deform retracts onto } T^{(-\infty,a_i]}$$
$$\forall \alpha \in (a_{i-1}, a_i],\ T^{[\alpha,+\infty)} \text{ deform retracts onto } T^{[a_i,+\infty)},$$

where a topological space X *is said to* deform retract *onto* $Y \subseteq X$ *if there exists a continuous function* $F : X \times [0,1] \to X$ *such that* $F(\cdot, 0) = \mathrm{id}_X$*,* $F|_{Y \times \{\alpha\}}(\cdot, \alpha) = \mathrm{id}_Y$ *for any* $\alpha \in [0,1]$*, and* $F(X, 1) \subseteq Y$*. In particular, this means that the inclusion* $Y \hookrightarrow X$ *is a homotopy equivalence.*

Corollary 7.3. *The following inclusion holds:* $\mathrm{Dg}(\pi_2) \subseteq \mathrm{Crit}(T) \times \mathrm{Crit}(T)$.

Construction from a Morse-type function. One can build telescopes from the domain of Morse-type functions—see Definition 2.1. Indeed, a function $f : X \to \mathbb{R}$ of Morse type naturally induces a telescope $T(X, f)$ with

- $\mathrm{Crit}(T(X, f)) = \mathrm{Crit}(f)$,
- $X_i = f^{-1}(a_i)$,
- $Y_i = \pi_1 \circ \mu_i^{-1} \circ f^{-1}((a_i, a_{i+1}))$,
- $\phi_i : (y, a_i) \mapsto (\bar{\mu}_i|_{Y_i \times \{a_i\}}(y, a_i), a_i)$, $\forall y \in Y_i$, $\forall i \in \{1, ..., n\}$,

- $\psi_i : (y, a_{i+1}) \mapsto (\bar{\mu}_i|_{Y_i \times \{a_{i+1}\}}(y, a_{i+1}), a_{i+1})$, $\forall y \in Y_i$, $\forall i \in \{0, ..., n-1\}$,

$T(X, f)$ is well-defined thanks to the following Lemma:

Lemma 7.4. $\mathrm{im}(\phi_i) \subseteq f^{-1}(a_i) \times \{a_i\}$ *and* $\mathrm{im}(\psi_i) \subseteq f^{-1}(a_{i+1}) \times \{a_{i+1}\}$.

Proof. Let $(y, a_{i+1}) \in Y_i \times \{a_{i+1}\}$. Consider the sequence $(y, v_n)_{n\in\mathbb{N}}$, for an arbitrary $(v_n)_{n\in\mathbb{N}} \in (a_i, a_{i+1})^{\mathbb{N}}$ that converges to a_{i+1}. Then, $(f \circ \bar{\mu}_i(y, v_n))_{n\in\mathbb{N}}$ converges to $f \circ \bar{\mu}_i(y, a_{i+1})$ by continuity of $f \circ \bar{\mu}$. Moreover, for all $n \in \mathbb{N}$ we have $f \circ \bar{\mu}_i(y, v_n) = f \circ \mu_i(y, v_n) = v_n$ since $f|_{f^{-1}(a_i, a_{i+1})} = \pi_2 \circ \mu_i^{-1}$. Therefore, $(f \circ \bar{\mu}_i(y, v_n))_{n\in\mathbb{N}}$ converges also to a_{i+1}. By uniqueness of the limit, we have $f \circ \bar{\mu}_i(y, a_{i+1}) = a_{i+1}$, meaning that $\bar{\mu}_i(y, a_{i+1}) \in f^{-1}(a_{i+1})$. Thus, $\mathrm{im}(\psi_i) \subseteq f^{-1}(a_{i+1}) \times \{a_{i+1}\}$. The same argument applies to show that $\mathrm{im}(\phi_i) \subseteq f^{-1}(a_i) \times \{a_i\}$. □ □

Correspondence between X and $T(X, f)$. We now exhibit a homeomorphism between $T(X, f)$ and X. Let $\mu : T(X, f) \to X$ be defined by:

$$\mu(y, z) = \begin{cases} y \text{ if } (y, z) \in X_i \times \{a_i\} \text{ for some } i; \\ \mu_i(y, z) \text{ if } (y, z) \in Y_i \times (a_i, a_{i+1}) \text{ for some } i. \end{cases}$$

The map μ is bijective as every μ_i is. It is also continuous as every $\bar{\mu}_i$ is. Since every continuous bijection from a compact space to a Hausdorff space is a homeomorphism (see e.g. Proposition 13.26 in [34]), μ defines a homeomorphism between $T(X, f)$ and X. Moreover, $\pi_2 = f \circ \mu$ so $\mathrm{Dg}(f) = \mathrm{Dg}(\pi_2)$.

Operators on telescopes

The decomposition of telescopes into cylinders can be used to define simple operators that modify the telescope structures in a predictable way. Specifically, we detail three types of operators, corresponding to the cases where one asks for either removal of critical values (Merge operator), duplication of critical values (Split operator), or translation of critical values (Shift operator). To formalize this, we use *generalized attaching maps*:

$$\begin{array}{lllll} \phi_i^a : & Y_i \times \{a\} \to & X_i \times \{a\}; & (y, a) \mapsto (\pi_1 \circ \phi_i(y, a_i), a), \\ \psi_i^a : & Y_i \times \{a\} \to & X_{i+1} \times \{a\}; & (y, a) \mapsto (\pi_1 \circ \psi_i(y, a_{i+1}), a). \end{array}$$

Merge. Merge operators merge all critival values located in $[a, b]$ into a single critical value $\bar{a} = \frac{a+b}{2}$.

Definition 7.5 (Merge)**.** *Let T be a telescope. Let $a \leq b$. If $[a, b]$ contains at least one critical value, i.e. $\exists i, j \in \mathbb{N}$ such that $a_{i-1} < a \leq a_i \leq a_j \leq b < a_{j+1}$, then the* Merge *on T between a, b is the telescope $T' = \mathrm{Merge}_{a,b}(T)$ given by:*

$$\begin{gathered} ...(Y_{i-1} \times [a_{i-1}, a_i]) \cup_{\psi_{i-1}} (X_i \times \{a_i\}) \cup_{\phi_i} ... \cup_{\psi_{j-1}} (X_j \times \{a_j\}) \cup_{\phi_j} (Y_j \times [a_j, a_{j+1}])... \\ \downarrow \\ ...(Y_{i-1} \times [a_{i-1}, \bar{a}]) \cup_{f_{i-1}} (T^{[a,b]} \times \{\bar{a}\}) \cup_{g_j} (Y_j \times [\bar{a}, a_{j+1}])... \end{gathered}$$

where $\bar{a} = \frac{a+b}{2}$, where $f_{i-1} = \psi_{i-1}^{\bar{a}}$ if $a = a_i$ and $f_{i-1} = \mathrm{id}_{Y_{i-1} \times \{\bar{a}\}}$ otherwise, and where $g_j = \phi_j^{\bar{a}}$ if $b = a_j$ and $g_j = \mathrm{id}_{Y_j \times \{\bar{a}\}}$ otherwise.

If $[a, b]$ contains no critical value, i.e. $a_{i-1} < a \leq b < a_i$, then $\mathrm{Merge}_{a,b}(T)$ *is given by:*

$$\begin{gathered} ...(X_{i-1} \times \{a_{i-1}\}) \cup_{\phi_{i-1}} (Y_{i-1} \times [a_{i-1}, a_i]) \cup_{\psi_{i-1}} (X_i \times \{a_i\})... \\ \downarrow \\ ... \cup_{\phi_{i-1}} (Y_{i-1} \times [a_{i-1}, \bar{a}]) \cup_{f_{i-1}} (T^{[a,b]} \times \{\bar{a}\}) \cup_{g_{i-1}} (Y_{i-1} \times [\bar{a}, a_i]) \cup_{\psi_{i-1}} ... \end{gathered}$$

where $\bar{a} = \frac{a+b}{2}$, and where $f_{i-1} = g_{i-1} = \mathrm{id}_{Y_{i-1} \times \{\bar{a}\}}$.

See the left panel of Figure 13 for an illustration.

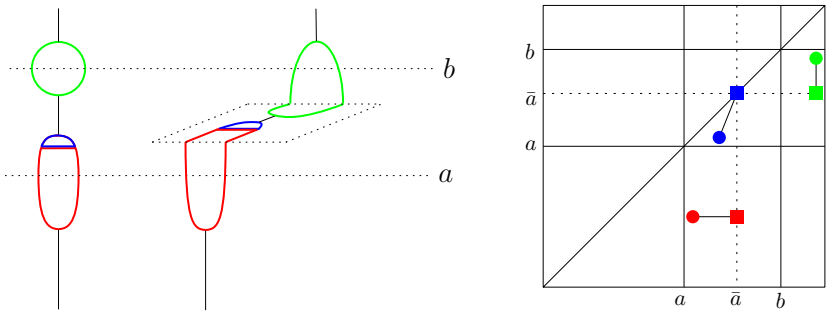


Figure 13: Left: Effect of a Merge on a telescope. Right: Effect on the corresponding extended persistence diagram. Points before the Merge are disks while points after the Merge are squares.

Merge for persistence diagrams. Similarly, we define the Merge between a, b on an extended persistence diagram Dg as the diagram $\mathrm{Merge}_{a,b}(\mathrm{Dg})$ given by $\mathrm{Merge}_{a,b}(x, y) = (\bar{x}, \bar{y})$, where:

$$\bar{x} = \begin{cases} x \text{ if } x \notin [a, b] \\ \bar{a} \text{ otherwise} \end{cases} \quad \text{and } \bar{y} = \begin{cases} y \text{ if } y \notin [a, b] \\ \bar{a} \text{ otherwise} \end{cases}$$

Points in the strips $x \in [a, b]$, $y \in [a, b]$ are snapped to the lines $x = \bar{a}$ and $y = \bar{a}$ respectively. See the right panel of Figure 13. See also the first intermediate points along the trajectories of the red points in Figure 20 for another illustration on extended persistence diagrams.

Commutativity of the operators. We now prove that extended persistent homology commutes with this operator, i.e. $\mathrm{Dg}(\mathrm{Merge}) = \mathrm{Merge}(\mathrm{Dg})$.

Lemma 7.6. *Let $a \leq b$ and $T' = \mathrm{Merge}_{a,b}(T)$. Let $\pi_2' : T' \to \mathbb{R}$ be the projection onto the second factor. Then, $\mathrm{Dg}(\pi_2') = \mathrm{Merge}_{a,b}(\mathrm{Dg}(\pi_2))$.*

Proof. We only study the sublevel sets of the functions, which means that we only prove the result for the ordinary part of the diagrams. The proof is symmetric for superlevel sets, leading to the result for the extended and the relative parts.

Assume $a_{i-1} < a \leq a_i \leq a_j \leq b < a_{j+1}$. Given $x \leq y$, we let $\Pi_{x,y} : H_*(T^{(-\infty,x]}) \to H_*(T^{(-\infty,y]})$ and $\Pi'_{x,y} : H_*((T')^{(-\infty,x]}) \to H_*((T')^{(-\infty,y]})$ be the homomorphisms induced by inclusions. Since f is of Morse type, Lemma 7.2 relates Π' to Π as follows (see Figure 14):

$$\Pi'_{x,y} = \begin{cases} \Pi_{x,y} \text{ if } x, y \notin [a, b] \text{ (green)} & \Pi_{x,a_{i-1}} \text{ if } x < a, y \in [a, \bar{a}) \text{ (pink)} \\ \Pi_{a_{i-1},y} \text{ if } x \in [a, \bar{a}), y > b \text{ (blue)} & \Pi_{a_{i-1},a_j} \text{ if } x \in [a, \bar{a}), y \in [\bar{a}, b] \text{ (orange)} \\ \Pi_{a_j,y} \text{ if } x \in [\bar{a}, b], y > b \text{ (grey)} & \mathrm{id}^*_{Y_{i-1}} \text{ if } x, y \in [a, \bar{a}) \text{ (brown)} \\ \Pi_{x,a_j} \text{ if } x < a, y \in [\bar{a}, b] \text{ (turquoise)} & \mathrm{id}^*_{Y_j} \text{ if } x, y \in [\bar{a}, b] \text{ (purple)} \end{cases} \tag{12}$$

The equality between the diagrams follows from these relations and the inclusion-exclusion formula (1). Consider for instance the case where the point $(x, y) \in \mathrm{Dg}(\pi_2)$ belongs to the union A of the pink and the turquoise areas. One can select two abscissae $x_1 < x < x_2$ and an arbitrarily small $\epsilon > 0$. Then, the total multiplicity of the corresponding rectangle R in $\mathrm{Dg}(\pi_2')$ (displayed in the right panel of Figure 14) is given by:

$$\mathrm{mult}(R) = \mathrm{rank}\ \Pi'_{x_2,a-\epsilon} - \mathrm{rank}\ \Pi'_{x_2,b+\epsilon} + \mathrm{rank}\ \Pi'_{x_1,b+\epsilon} - \mathrm{rank}\ \Pi'_{x_1,a-\epsilon}.$$

The first relation in (12) shows that R has exactly the same multiplicity in $\mathrm{Dg}(\pi_2)$, since all its corners belong to the green area. As this is true for arbitrarily small $\epsilon > 0$, it means that $R' = R \cap A$ also has the same multiplicity in $\mathrm{Dg}(\pi_2)$ as in $\mathrm{Dg}(\pi_2')$. Now, if we pick a point inside R' with an ordinate different than $\bar{a}$,

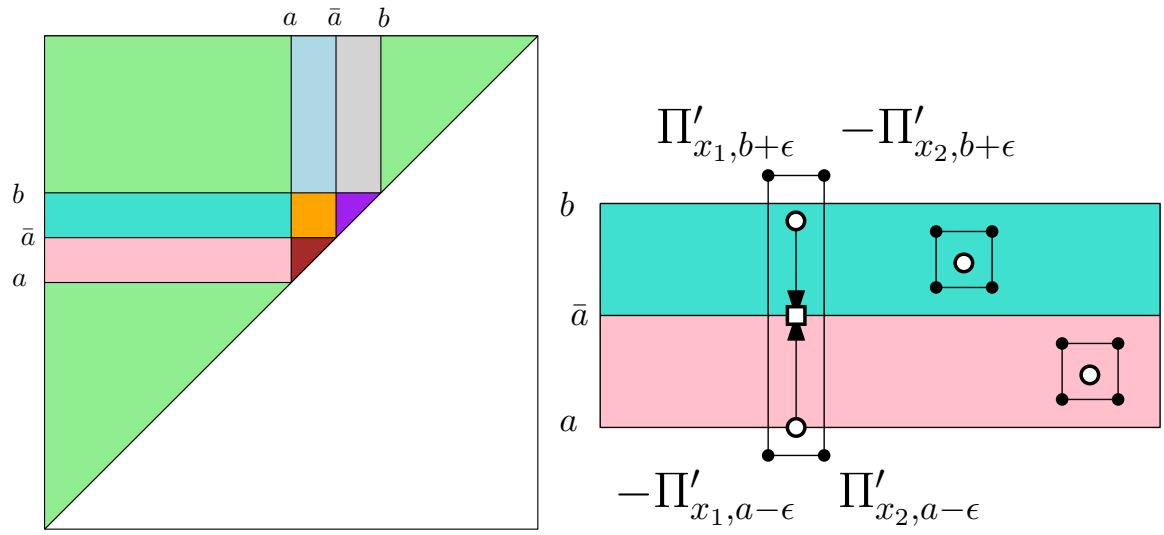


Figure 14: Left: Areas of the extended persistence diagram used in the proof. Right: Examples of the boxes we use to prove the result (circles represent points before the Merge, squares represent points after the Merge).

we can compute its multiplicity in $\mathrm{Dg}(\pi'_2)$ by surrounding it with a box included in the turquoise area (if the ordinate is bigger than $\bar{a}$) or in the pink area (if it is smaller). Boxes in the turquoise area have multiplicity $\mathrm{rank}\,\Pi'_{x_2,y_1} - \mathrm{rank}\,\Pi'_{x_2,y_2} + \mathrm{rank}\,\Pi'_{x_1,y_2} - \mathrm{rank}\,\Pi'_{x_1,y_1} = \mathrm{rank}\,\Pi_{x_2,a_j} - \mathrm{rank}\,\Pi_{x_2,a_j} + \mathrm{rank}\,\Pi_{x_1,a_j} - \mathrm{rank}\,\Pi_{x_1,a_j} = 0$. Similarly, boxes in the pink area also have multiplicity zero. Thus, all points of R' in $\mathrm{Dg}(\pi'_2)$ have ordinate $\bar{a}$. Again, as it is true for x_1, x_2 as close to each other as we want, it means that (x, y) is snapped to $(x, \bar{a})$ in $\mathrm{Dg}(\pi'_2)$. The treatment of the other areas in the plane is similar.

Now, if $[a, b]$ contains no critical values, then $\Pi' = \Pi$, so the result is clear. □ □

Split. Split operators split a critical value a_i into two different ones $a_i - \varepsilon$ and $a_i + \varepsilon$.

Definition 7.7 (Split)**.** *Let T be a telescope. Let $a_i \in \mathrm{Crit}(T)$ and ϵ such that*

$$0 \leq \epsilon < \min\{a_{i+1} - a_i, a_i - a_{i-1}\}.$$

The ϵ-Split on T at a_i is the telescope $T' = \mathrm{Split}_{\epsilon,a_i}(T)$ given by:

$$\begin{array}{c} \ldots(Y_{i-1} \times [a_{i-1}, a_i]) \cup_{\psi_{i-1}} (X_i \times \{a_i\}) \cup_{\phi_i} (Y_i \times [a_i, a_{i+1}])\ldots \\ \downarrow \\ \ldots(Y_{i-1} \times [a_{i-1}, a_i - \epsilon]) \cup_{\psi_{i-1}^{a_i-\epsilon}} (X_i \times \{a_i - \epsilon\}) \cup_{\mathrm{id}} (X_i \times [a_i - \epsilon, a_i + \epsilon]) \cup_{\mathrm{id}} (X_i \times \{a_i + \epsilon\}) \cup_{\phi_i^{a_i+\epsilon}} (Y_i \times [a_i + \epsilon, a_{i+1}])\ldots \end{array}$$

See the left panel of Figure 15 for an illustration.

Down- and up-forks. Splits create particular critical values called *down-* and *up-forks*. Intuitively, Split operations allow to distinguish between all possible types of changes in 0- and 1-dimensional homology of the sublevel and superlevel sets, namely: union of two connected components, creation of a connected component, destruction of a connected component, and separation of a connected component. Unions and creations occur at down-forks while separations and destructions occur at up-forks. See Figure 16 for an illustration. We formalize and prove this intuition in Lemma 7.11.

Definition 7.8. *A critical value $a_i \in \mathrm{Crit}(T)$ is called an* up-fork *if ψ_{i-1} is an homeomorphism, and it is called a* down-fork *if ϕ_i is a homeomorphism.*

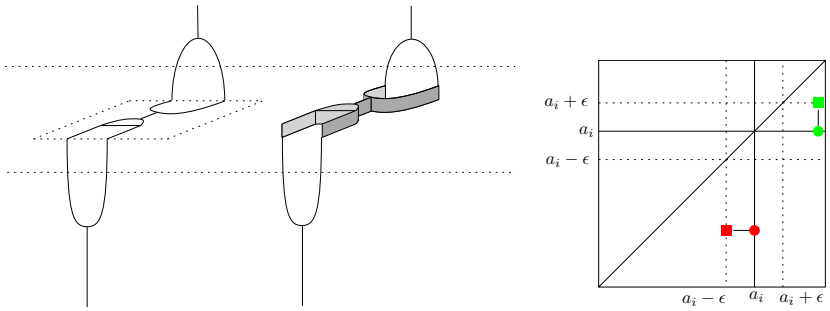


Figure 15: Left: Effect of a Split on a telescope. Right: Effect on the corresponding extended persistence diagram. Points before the Split are disks while points after the Split are squares.

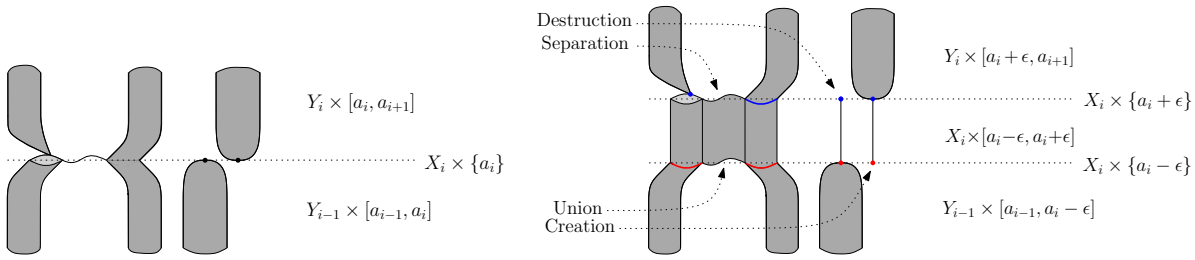


Figure 16: Left and right panels display the space before and after a Split respectively. Subsets of X_i that are colored in red and blue correspond to $\mathrm{im}(\pi_1 \circ \psi_{i-1})$ and $\mathrm{im}(\pi_1 \circ \phi_i)$ respectively.

Since the attaching maps introduced by the Split are identity maps, we have the following lemma:

Lemma 7.9. *The critical values $a_i - \epsilon$ and $a_i + \epsilon$ created with* Split *are down- and up-forks respectively.*

The next lemma is a direct consequence of the existence and continuity of ϕ_i^{-1} (resp. ψ_{i-1}^{-1}) when $a_i \in \mathrm{Crit}(T)$ is a down-fork (resp. up-fork):

Lemma 7.10. *Let $a_i \in \mathrm{Crit}(T)$. If a_i is an up-fork, then $T^{(-\infty,a_i]}$ deform retracts onto $T^{(-\infty,\alpha]}$ for all $\alpha \in (a_{i-1}, a_i]$. If a_i is a down-fork, then $T^{[a_i,+\infty)}$ deform retracts onto $T^{[\alpha,+\infty)}$ for all $\alpha \in [a_i, a_{i+1})$.*

Now we can prove the previous intuition concerning down- and up-forks correct:

Lemma 7.11. *Let $a_i \in \mathrm{Crit}(T)$. If a_i is an up-fork, then it can only be the birth time of relative cycles and the death time of relative and extended cycles in $\mathrm{Dg}(\pi_2)$. If a_i is a down-fork, then it can only be the birth time of ordinary and extended cycles and the death time of ordinary cycles in $\mathrm{Dg}(\pi_2)$.*

Proof. Let $0 \leq \epsilon, \epsilon' < \min\{a_{i+1} - a_i, a_i - a_{i-1}\}$. Consider the extended persistence module of π_2:

$$\begin{array}{ccccccccc} \ldots & \longrightarrow & H_*(T^{(-\infty,a_i-\epsilon]}) & \longrightarrow & H_*(T^{(-\infty,a_i]}) & \longrightarrow & H_*(T^{(-\infty,a_i+\epsilon']}) & \longrightarrow & \ldots \\ \ldots & \longrightarrow & H_*(T, T^{[a_i+\epsilon',+\infty)}) & \longrightarrow & H_*(T, T^{[a_i,+\infty)}) & \longrightarrow & H_*(T, T^{[a_i-\epsilon,+\infty)}) & \longrightarrow & \ldots \end{array}$$

If a_i is an up-fork, then the composition $H_*(T^{(-\infty,a_i-\epsilon]}) \to H_*(T^{(-\infty,a_i+\epsilon']})$ is an isomorphism since $T^{(-\infty,a_i+\epsilon']}$ deform retracts onto $T^{(-\infty,a_i-\epsilon]}$ by Lemmas 7.2 and 7.10. As ϵ, ϵ' can be chosen arbitrarily small, there cannot be any creation of ordinary or extended cycle at a_i. There also cannot be any destruction of ordinary cycle.

Similarly, if a_i is a down-fork, then the composition $H_*(T, T^{[a_i+\epsilon',+\infty)}) \to H_*(T, T^{[a_i-\epsilon,+\infty)})$ is an isomorphism since $T^{[a_i-\epsilon,+\infty)}$ deform retracts onto $T^{[a_i+\epsilon',+\infty)}$. Again, there cannot be any destruction of extended or relative cycle at a_i. There also cannot be any creation of relative cycle. □ □

Split for persistence diagrams. Similarly, we define the ϵ-Split at a_i on a diagram Dg as the diagram $\mathrm{Split}_{\epsilon,a_i}(\mathrm{Dg})$ given by $\mathrm{Split}_{\epsilon,a_i}(x,y) = (\bar{x},\bar{y})$, where:

$$\bar{x} = \begin{cases} x \text{ if } x \neq a_i \\ a_i + \epsilon \text{ if } x = a_i \text{ and } (x,y) \in \mathrm{Rel} \\ a_i - \epsilon \text{ if } x = a_i \text{ and } (x,y) \notin \mathrm{Rel} \end{cases} \quad \text{and } \bar{y} = \begin{cases} y \text{ if } y \neq a_i \\ a_i - \epsilon \text{ if } y = a_i \text{ and } (x,y) \in \mathrm{Ord} \\ a_i + \epsilon \text{ if } y = a_i \text{ and } (x,y) \notin \mathrm{Ord} \end{cases}$$

Points located on the lines $x, y = a_i$ are snapped to the lines $x, y = a_i \pm \epsilon$ according to their type. Note that the definition of $\mathrm{Split}_{\epsilon,a_i}(\mathrm{Dg})$ assumes implicitly that Dg contains no point within the horizontal and vertical bands $[a_i - \epsilon, a_i) \times \mathbb{R}$, $(a_i, a_i + \epsilon] \times \mathbb{R}$, $\mathbb{R} \times [a_i - \epsilon, a_i)$ and $\mathbb{R} \times (a_i, a_i + \epsilon]$, which is the case under the assumptions of Definition 7.7. See the right panel of Figure 15 for an illustration. See also the second intermediate points along the trajectories of the red points in Figure 20 for another illustration on extended persistence diagrams.

Commutativity of the operators. We now prove that extended persistent homology commutes with this operator, i.e. $\mathrm{Dg}(\mathrm{Split}) = \mathrm{Split}(\mathrm{Dg})$.

Lemma 7.12. *Let $a_i \in \mathrm{Crit}(T)$. Let $0 < \epsilon < \min\{a_{i+1} - a_i, a_i - a_{i-1}\}$, $T' = \mathrm{Split}_{\epsilon,a_i}(T)$ and $\pi_2' : T' \to \mathbb{R}$ the projection onto the second factor. Then, $\mathrm{Dg}(\pi_2') = \mathrm{Split}_{\epsilon,a_i}(\mathrm{Dg}(\pi_2))$.*

Proof. Note that $T = \mathrm{Merge}_{a_i-\epsilon,a_i+\epsilon}(T')$. Hence, by Lemma 7.6, $\mathrm{Dg}(\pi_2)$ can be obtained from $\mathrm{Dg}(\pi_2')$ with $\mathrm{Dg}(\pi_2) = \mathrm{Merge}_{a_i-\epsilon,a_i+\epsilon}(\mathrm{Dg}(\pi_2'))$. Note also that π_2' has no critical value within the open interval $(a_i - \epsilon, a_i + \epsilon)$, so $\mathrm{Dg}(\pi_2')$ has no point within the horizontal and vertical bands $\mathbb{R} \times (a_i - \varepsilon, a_i + \varepsilon)$ and $(a_i - \varepsilon, a_i + \varepsilon) \times \mathbb{R}$. Finally, Lemma 7.9 ensures that $a_i + \epsilon, a_i - \epsilon$ are up- and down-forks respectively, so Lemma 7.11 tells us exactly where the preimages of the points of $\mathrm{Dg}(\pi_2)$ through the Merge are located depending on their type. □ □

Shift. Shift operators translate critical values.

Definition 7.13 (Shift)**.** *Let T be a telescope. Let $a_i \in \mathrm{Crit}(T)$ and ϵ such that*

$$0 \leq |\epsilon| < \min\{a_{i+1} - a_i, a_i - a_{i-1}\}.$$

The ϵ-Shift on T at a_i is the telescope $T' = \mathrm{Shift}_{\epsilon,a_i}(T)$ given by:

$$\begin{array}{c} ...(Y_{i-1} \times [a_{i-1}, a_i]) \cup_{\psi_{i-1}} (X_i \times \{a_i\}) \cup_{\phi_i} (Y_i \times [a_i, a_{i+1}])... \\ \downarrow \\ ...(Y_{i-1} \times [a_{i-1}, a_i + \epsilon]) \cup_{\psi_{i-1}^{a_i+\epsilon}} (X_i \times \{a_i + \epsilon\}) \cup_{\phi_i^{a_i+\epsilon}} (Y_i \times [a_i + \epsilon, a_{i+1}])... \end{array}$$

See the left panel of Figure 17 for an illustration.

Shift for persistence diagrams. Similarly, we define the ϵ-Shift at a_i on a diagram Dg as the diagram $\mathrm{Shift}_{\epsilon,a_i}(\mathrm{Dg})$ given by $\mathrm{Shift}_{\epsilon,a_i}(x,y) = (\bar{x},\bar{y})$ where:

$$\bar{x} = \begin{cases} x \text{ if } x \neq a_i \\ a_i + \epsilon \text{ otherwise} \end{cases} \quad \text{and } \bar{y} = \begin{cases} y \text{ if } y \neq a_i \\ a_i + \epsilon \text{ otherwise} \end{cases}$$

Points located on the lines $x, y = a_i$ are snapped to the lines $x, y = a_i + \epsilon$. Note that the definition of $\mathrm{Shift}_{\epsilon,a_i}(\mathrm{Dg})$ assumes implicitly that Dg contains no point within the horizontal and vertical bands delimited by a_i and $a_i + \varepsilon$, which is the case under the assumptions of Definition 7.13. See the right panel of Figure 17 for an illustration. See also the third intermediate points along the trajectories of the red points in Figure 20 for another illustration on extended persistence diagrams.

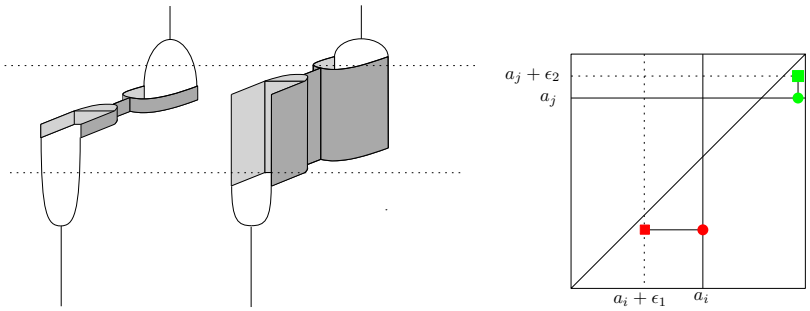


Figure 17: Left: Effect of a double Shift with amplitudes $\epsilon_1 < 0 < \epsilon_2$. Right: Effect on the corresponding extended persistence diagram. Points before the Shift are disks while points after the Shift are squares.

Commutativity of the operators. We now prove that extended persistent homology commutes with this operator, i.e. $\mathrm{Dg}(\mathrm{Shift}) = \mathrm{Shift}(\mathrm{Dg})$.

Lemma 7.14. *Let $a_i \in \mathrm{Crit}(T)$, $\epsilon \in (a_{i-1} - a_i,\ a_{i+1} - a_i)$, $T' = \mathrm{Shift}_{\epsilon,a_i}(T)$ and $\pi_2' : T' \to \mathbb{R}$ the projection onto the second factor. Then, $\mathrm{Dg}(\pi_2') = \mathrm{Shift}_{\epsilon,a_i}(\mathrm{Dg}(\pi_2))$.*

Proof. Again, the following relations coming from Lemma 7.2:

$$\Pi'_{x,y} = \begin{cases} \Pi_{x,y} \text{ if } x, y \notin (a_{i-1}, a_{i+1}) \text{ (green)} & \Pi_{a_i,y} \text{ if } x \in [a_i + \epsilon, a_{i+1}), y \geq a_{i+1} \text{ (grey)} \\ \Pi_{x,a_{i-1}} \text{ if } x \leq a_{i-1}, y \in (a_{i-1}, a_i + \epsilon) \text{ (pink)} & \Pi_{a_{i-1},y} \text{ if } x \in (a_{i-1}, a_i + \epsilon), y \geq a_{i+1} \text{ (blue)} \\ \Pi_{x,a_i} \text{ if } x \leq a_{i-1}, y \in [a_i + \epsilon, a_{i+1}) \text{ (turquoise)} & \mathrm{id}^*_{Y_{i-1}} \text{ if } x, y \in (a_{i-1}, a_i + \epsilon) \text{ (brown)} \\ \Pi_{a_{i-1},a_i} \text{ if } x \in (a_{i-1}, a_i + \epsilon), y \in [a_i + \epsilon, a_{i+1}) \text{ (orange)} & \mathrm{id}^*_{Y_i} \text{ if } x, y \in [a_i + \epsilon, a_{i+1}) \text{ (purple)} \end{cases}$$

allow us to prove the result similarly to Lemma 7.6—see Figure 18. For instance, one can choose a box that intersects the lines $y = a_i + \epsilon$ and $y = a_i$, show that the total multiplicity is preserved, then choose another small box that does not intersect $y = a_i + \epsilon$ inside the first box, and show that its multiplicity is zero.

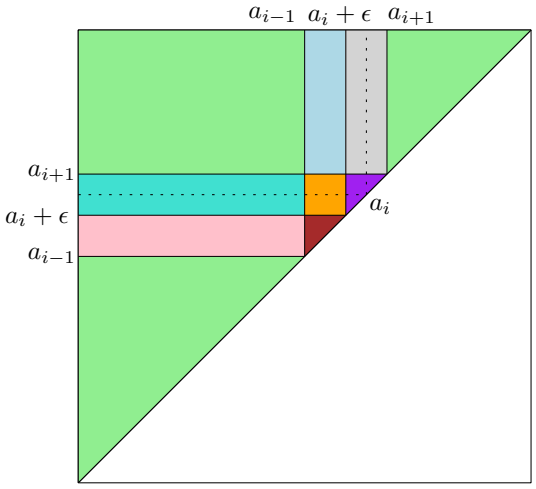


Figure 18: Areas of the extended persistence diagram used in the proof, with $\varepsilon < 0$.

□ □

7.2 Operators on MultiNerve Mapper

We first provide invariance results for MultiNerve Mappers computed on telescopes as defined in Section 7.1. The result is stated in a way that is adapted to its use in the following sections. The conclusion would still hold under somewhat weaker assumptions.

Proposition 7.15. *Let T be a telescope, π_2 be the projection onto the second coordinate, and $\mathcal{I}$ be a gomic of* $\mathrm{im}(\pi_2)$. *Let* $\mathrm{End}(\mathcal{I})$ *denote the set of endpoints of intervals of $\mathcal{I}$, sorted in ascending order. All isomorphisms mentioned in the following items are in the category of combinatorial multigraphs.*

(i) *Let $a \leq b$ such that there exists an interval $I \in \mathcal{I}$ for which a, b belong to either $I_\cap^-$, $\tilde{I}$ or $I_\cap^+$. Then,* $\overline{\mathrm{M}}_{\pi_2}(\mathrm{Merge}_{a,b}(T), \mathcal{I})$ *is isomorphic to* $\overline{\mathrm{M}}_{\pi_2}(T, \mathcal{I})$.

(ii) *Let $a_i \in \mathrm{Crit}(T) \setminus \mathrm{End}(\mathcal{I})$, and $a < a_i < b$ with a, b consecutive in* $\mathrm{End}(\mathcal{I})$. *If $a_{i-1} < a < b < a_{i+1}$ and $0 < \varepsilon < \min\{a_i - a, b - a_i\}$, then* $\overline{\mathrm{M}}_{\pi_2}(\mathrm{Split}_{\varepsilon,a_i}(T), \mathcal{I})$ *is isomorphic to* $\overline{\mathrm{M}}_{\pi_2}(T, \mathcal{I})$.

(iii) *Let $a_i \in \mathrm{Crit}(T) \setminus \mathrm{End}(\mathcal{I})$, and $b < a_i < c < d$ with b, c, d consecutive in* $\mathrm{End}(\mathcal{I})$. *If a_i is an up-fork, $(b, c) = I \cap J$ is an intersection, and $c - a_i < \varepsilon < \min\{d, a_{i+1}\} - a_i$, then* $\overline{\mathrm{M}}_{\pi_2}(\mathrm{Shift}_{\varepsilon,a_i}(T), \mathcal{I})$ *is isomorphic to* $\overline{\mathrm{M}}_{\pi_2}(T, \mathcal{I})$.

(iv) *Let $a_i \in \mathrm{Crit}(T) \setminus \mathrm{End}(\mathcal{I})$, and $a < b < a_i < c$ with a, b, c consecutive in* $\mathrm{End}(\mathcal{I})$. *If a_i is a down-fork, $(b, c) = I \cap J$ is an intersection, and* $\max\{a, a_{i-1}\} - a_i < \varepsilon < b - a_i$, *then* $\overline{\mathrm{M}}_{\pi_2}(\mathrm{Shift}_{\varepsilon,a_i}(T), \mathcal{I})$ *is isomorphic to* $\overline{\mathrm{M}}_{\pi_2}(T, \mathcal{I})$.

Proof. Under the assumptions given by each item, the connected components in every intersection $I \cap J$, $I, J \in \mathcal{I}$ and in every element $I \in \mathcal{I}$ remain the same after each operation. Given any intersection $K = I \cap J$, $I, J \in \mathcal{I}$, or interval $K = I \in \mathcal{I}$, we recall that T^K denotes $\pi_1 \circ \pi_2^{-1}(K)$. Then, we have:

(i) - (ii) T^K deform retracts onto $(\mathrm{Merge}_{a,b}(T))^K$ and $(\mathrm{Split}_{\varepsilon,a_i}(T))^K$ deform retracts onto T^K;

(iii) - (iv) The Shifts move the up-fork to the upper proper subinterval, and the down-fork to the lower proper subinterval, which preserves the connected components in each of the two intervals as well as in their intersection.

Thus, the MultiNerve Mapper is not changed by any of the aforementioned operations. □ □

7.3 Connection between the (MultiNerve) Mapper and the Reeb graph

In this section, we describe a sequence of metric spaces linking the MultiNerve Mapper and the Reeb graph. Let $f : X \to \mathbb{R}$ be of Morse type, and let $\mathcal{I}$ be a gomic of $\mathrm{im}(f)$. Let $T(X, f)$ be the corresponding telescope. The idea is to move all critical values out of the intersection preimages $f^{-1}(I \cap J)$, so that the MultiNerve Mapper and the Reeb graph become isomorphic. For any interval $I \in \mathcal{I}$, we let $a_{\tilde{I}} < b_{\tilde{I}}$ be the endpoints of its proper subinterval $\tilde{I}$, so we have $\tilde{I} = [a_{\tilde{I}}, b_{\tilde{I}}]$. For any non-empty intersection $I \cap J$, we fix a subinterval $[a_{I\cap J}, b_{I\cap J}] \subset I \cap J$ such that every critical value within $I \cap J$ falls into $[a_{I\cap J}, b_{I\cap J}]$ (which is possible because f is of Morse type hence has finitely many critical values). We then define three different operations individually as follows:

- $\mathrm{Merge}_{\mathcal{I}}$ is the composition of all the $\mathrm{Merge}_{a_{\tilde{I}}, b_{\tilde{I}}}$, $I \in \mathcal{I}$, and of all the $\mathrm{Merge}_{a_{I\cap J}, b_{I\cap J}}$, $I, J \in \mathcal{I}$ and $I \cap J \neq \emptyset$. All these functions commute, so their composition is well-defined. The same holds for the following compositions.

- $\mathrm{Split}_{\mathcal{I}}$ is the composition of all the $\mathrm{Split}_{\varepsilon,\bar{a}}$ with $\bar{a}$ a critical value after $\mathrm{Merge}_{\mathcal{I}}$ (therefore not an interval endpoint) and $\varepsilon > 0$ such that the assumptions of Proposition 7.15 (ii) are satisfied.

- $\text{Shift}_{\mathcal{I}}$ is the composition of all the $\text{Shift}_{\varepsilon,\bar{a}_+}$ with $\bar{a}_+$ an up-fork critical value after the $\text{Split}_{\mathcal{I}}$ and $\varepsilon > 0$ such that the assumptions of Proposition 7.15 (iii) are satisfied, and of all the $\text{Shift}_{\varepsilon,\bar{a}_-}$ with $\bar{a}_-$ a down-fork critical value after the $\text{Split}_{\mathcal{I}}$ and $\varepsilon < 0$ such that the assumptions of Proposition 7.15 (iv) are satisfied. After $\text{Shift}_{\mathcal{I}}$ there are no more critical values located in the intersections of consecutive intervals of $\mathcal{I}$.
- $\text{Merge}'_{\mathcal{I}}$ is the composition of all the $\text{Merge}_{a_{\tilde{I}},b_{\tilde{I}}}$, $I \in \mathcal{I}$.

We can now define our sequence of intermediate spaces:

Definition 7.16. *Let X be a topological space, $f : X \to \mathbb{R}$ be a Morse-type function, and $\mathcal{I}$ be a gomic of $\text{im}(f)$. Let $T(X, f)$ be the telescope associated to f. We define the telescope $\bar{T}_{\mathcal{I}}$ with:*

$$\bar{T}_{\mathcal{I}}(X, f) = \text{Merge}'_{\mathcal{I}} \circ \text{Shift}_{\mathcal{I}} \circ \text{Split}_{\mathcal{I}} \circ \text{Merge}_{\mathcal{I}}(T(X, f)).$$

We also let $\bar{f}_{\mathcal{I}}$ denote the projection of $\bar{T}_{\mathcal{I}}$ onto the second factor.

See Figure 19 for an illustration of this sequence of transformations. When often write $\bar{T}_{\mathcal{I}}$ instead of $\bar{T}_{\mathcal{I}}(X, f)$ when the pair (X, f) is clear from the context. In the following, we identify the pair (T, π_2) with (X, f) since they are isomorphic in the category of $\mathbb{R}$-constructible spaces. We also let $\tilde{\bar{f}}_{\mathcal{I}} : \text{R}_{\bar{f}_{\mathcal{I}}}(\bar{T}_{\mathcal{I}}) \to \mathbb{R}$ denote the induced map defined on the Reeb graph of $\bar{T}_{\mathcal{I}}$.

Thanks to Proposition 7.15 and the choice of the $a_{\tilde{I}}, b_{\tilde{I}}, a_{I\cap J}, b_{I\cap J}, \epsilon$ in the definitions of $\text{Merge}_{\mathcal{I}}$, $\text{Split}_{\mathcal{I}}, \text{Shift}_{\mathcal{I}}$ and $\text{Merge}'_{\mathcal{I}}$, we provide Lemma 7.17 below, which states that the MultiNerve Mapper is not affected by this sequence of transformations.

Lemma 7.17. *For $(\bar{T}_{\mathcal{I}}, \bar{f}_{\mathcal{I}})$ defined as in Definition 7.16, $\overline{\text{M}}_{\bar{f}_{\mathcal{I}}}(\bar{T}_{\mathcal{I}}, \mathcal{I})$ and $\overline{\text{M}}_f(X, \mathcal{I})$ are isomorphic as combinatorial multigraphs.*

This allows us to prove the following result, which states that the MultiNerve Mapper $\overline{\text{M}}_f(X, \mathcal{I})$ is actually the same object than the perturbed Reeb graph $\text{R}_{\bar{f}_{\mathcal{I}}}(\bar{T}_{\mathcal{I}})$.

Theorem 7.18. *For $(\bar{T}_{\mathcal{I}}, \bar{f}_{\mathcal{I}})$ defined as in Definition 7.16, $\overline{\text{M}}_f(X, \mathcal{I})$ and $\mathcal{C}\text{R}_{\bar{f}_{\mathcal{I}}}(\bar{T}_{\mathcal{I}})$ are isomorphic as combinatorial multigraphs.*

We know from Lemma 7.17 that $\overline{\text{M}}_f(X, \mathcal{I})$ and $\overline{\text{M}}_{\bar{f}_{\mathcal{I}}}(\bar{T}_{\mathcal{I}}, \mathcal{I})$ are isomorphic as combinatorial multigraphs. Theorem 7.18 is then a consequence of the following result, whose hypothesis is satisfied by the $\bar{T}_{\mathcal{I}}$ of Definition 7.16:

Lemma 7.19. *Let T be a telescope and let $\pi_2 : T \to \mathbb{R}$ be the projection onto the second factor. Suppose that every proper subinterval $\tilde{I}$ in the cover $\mathcal{I}$ contains exactly one critical value of π_2, and that the intersections $I \cap J$ contain none. Then, $\overline{\text{M}}_{\pi_2}(T, \mathcal{I})$ and $\mathcal{C}\text{R}_{\pi_2}(T)$ are isomorphic as combinatorial multigraphs.*

Proof. The nodes of $\mathcal{C}\text{R}_{\pi_2}(T)$ represent the connected components of the preimages of all critical values of π_2, while the nodes of $\overline{\text{M}}_{\pi_2}(T, \mathcal{I})$ represent the connected components of the preimages of all $I \in \mathcal{I}$. The hypothesis of the lemma implies that there is exactly one critical value per interval $I \in \mathcal{I}$, hence the nodes of $\overline{\text{M}}_{\pi_2}(T, \mathcal{I})$ and of $\mathcal{C}\text{R}_{\pi_2}(T)$ are in bijection. Meanwhile, the edges of $\mathcal{C}\text{R}_{\pi_2}(T)$ are given by the connected components of the $Y_i \times [a_i, a_{i+1}]$. Since the proper subintervals contain one critical value each and the $I \cap J$ contain none, the pullbacks of all intersections of consecutive intervals also span the $Y_i \times [a_i, a_{i+1}]$. Hence, the edges of $\overline{\text{M}}_{\pi_2}(T, \mathcal{I})$ are in bijection with the ones of $\mathcal{C}\text{R}_{\pi_2}(T)$. Moreover, their endpoints are defined in both cases by the ϕ_i and ψ_i. Hence the multigraph isomorphism. □ □

In passing, it is interesting to study the behavior of the MultiNerve Mapper as the hypothesis of the lemma is weakened. For instance:

Lemma 7.20. *Let T be a telescope and let $\pi_2 : T \to \mathbb{R}$ be the projection onto the second factor. Suppose that every interval I in the cover $\mathcal{I}$ contains at most one critical value of π_2. Then, $\overline{\text{M}}_{\pi_2}(T, \mathcal{I})$ is obtained from $\mathcal{C}\text{R}_{\pi_2}(T)$ by splitting some vertices into two and by subdividing some edges once.*

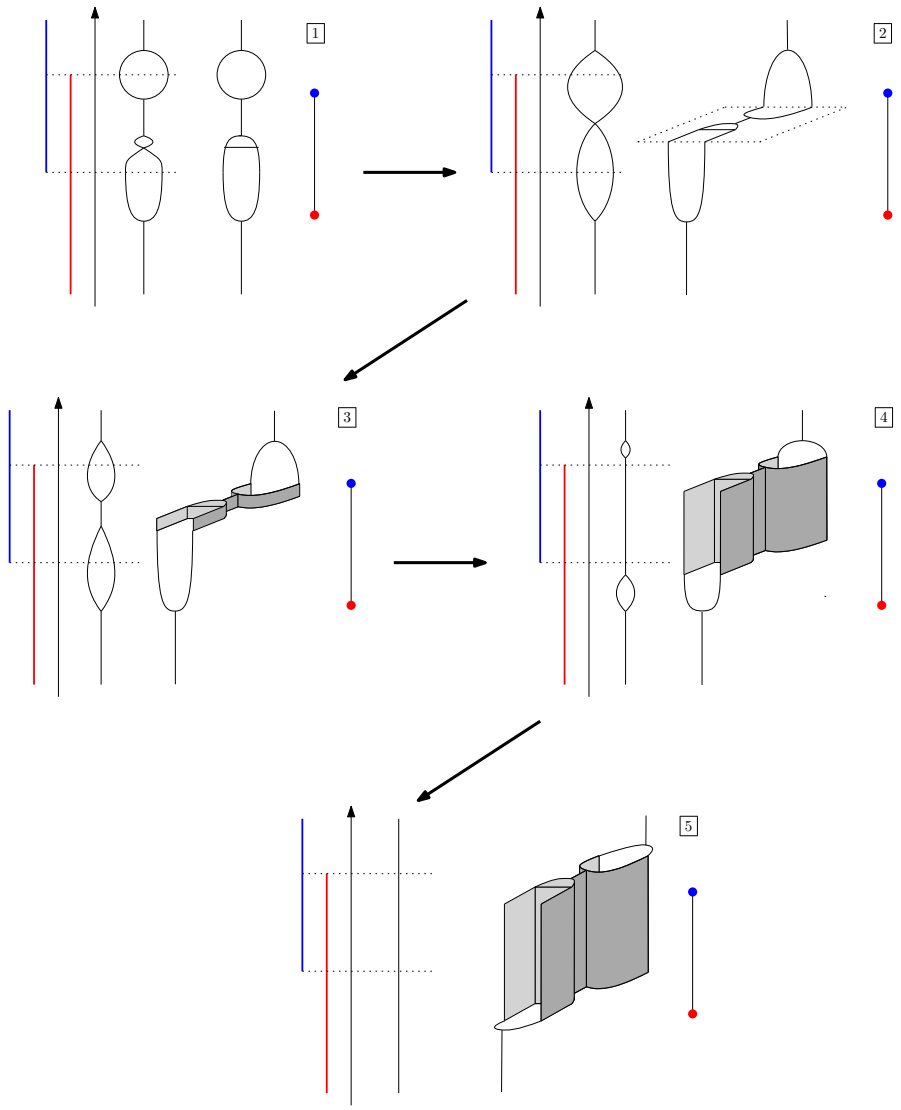


Figure 19: Illustration of the sequence of transformations in (7.16) on the features located in an interval intersection. For each figure, we display the original space (middle), its Reeb graph (left) and its MultiNerve Mapper (right).

Thus, the MultiNerve Mapper may non longer be 'exactly' isomorphic to the combinatorial Reeb graph (counter-examples are easy to build, by making some of the critical values fall into intersections of intervals in the cover), however it is still isomorphic to it up to vertex splits and edge subdivisions, which are topologically trivial modifications.

Lemma 7.20. The proof is constructive and it proceeds in 3 steps:
1. For every interval $I \in \mathcal{I}$ that does not contain a critical value, add a dummy critical value (with identities as connecting maps) in the proper subinterval $\tilde{I}$. The effect on the Mapper is null, while the effect on the Reeb graph is to subdivide once each edge crossing the dummy critical value. At this stage, every interval of $\mathcal{I}$ contains exactly one critical value. For simplicity we identify T with the new telescope.
2. For every interval $I \in \mathcal{I}$ whose corresponding critical value does not lie in the proper subinterval $\tilde{I}$ but rather in some intersection $I \cap J$ (defined uniquely since $\mathcal{I}$ is a gomic), merge I and J into a single interval

$I \cup J$. The coarser cover $\mathcal{J}$ thus obtained is still a gomic and it has the extra property that every proper subinterval contains exactly one critical value and every intersection contains none. Then, by Lemma 7.19, the MultiNerve Mapper $\overline{\mathrm{M}}_{\pi_2}(T, \mathcal{J})$ is isomorphic to the combinatorial Reeb graph $\mathcal{C}\mathrm{R}_{\pi_2}(T)$.
3. There remains to study the differences between $\overline{\mathrm{M}}_{\pi_2}(T, \mathcal{I})$ and $\overline{\mathrm{M}}_{\pi_2}(T, \mathcal{J})$. The only difference between the two covers is that some isolated pairs of intervals (I, J) have been merged because their intersection $I \cap J$ contained a critical value a_i. For every such pair, there are as many connected components in the preimage $\pi_2^{-1}(I)$ as in $\pi_2^{-1}(J)$ as in $\pi_2^{-1}(I \cap J)$ as in $\pi_2^{-1}(I \cup J)$ because $I \cup J$ contains no critical value other than a_i. Hence, every vertex of $\overline{\mathrm{M}}_{\pi_2}(T, \mathcal{J})$ corresponding to a connected component of $\pi_2^{-1}(I \cup J)$ is split into two in $\overline{\mathrm{M}}_{\pi_2}(T, \mathcal{I})$. Moreover, the two copies are connected by a single edge, given by the corresponding connected component of $\pi_2^{-1}(I \cap J)$. Now, assuming without loss of generality that J lies above I, we have $(I \cup J)^+_\cap = J^+_\cap$, which by assumption contains no critical value, so the connections between the vertex copy corresponding to $\pi_2^{-1}(J)$ and the vertices lying above it in $\overline{\mathrm{M}}_{\pi_2}(T, \mathcal{I})$ are the same as the connections between the original vertex and the vertices lying above it in $\overline{\mathrm{M}}_{\pi_2}(T, \mathcal{J})$. Similarly, $(I \cup J)^-_\cap = I^-_\cap$ contains no critical value by assumption, so the connections between the vertex copy corresponding to $\pi_2^{-1}(I)$ and the vertices lying below it in $\overline{\mathrm{M}}_{\pi_2}(T, \mathcal{I})$ are the same as the connections between the original vertex and the vertices lying below it in $\overline{\mathrm{M}}_{\pi_2}(T, \mathcal{J})$. □ □

Extension to the Mapper. Due to the simple relation between the Mapper and the MultiNerve Mapper given by Corollary 3.5, Theorem 7.18 can be extended for Mappers.

Definition 7.21. *Let X be a topological space, and $f : X \to \mathbb{R}$ be a Morse-type function. Let $(\bar{T}_\mathcal{I}, \bar{f}_\mathcal{I})$ be defined as in Definition 7.16. Let* $\mathrm{Cyl}(\bar{T}_\mathcal{I})$ *be the set of the connected components of the cylinders of $\bar{T}_\mathcal{I}$. We define the equivalence relation $\sim$ between elements of* $\mathrm{Cyl}(\bar{T}_\mathcal{I})$ *as:*

$$C \sim C' \Leftrightarrow \begin{cases} C, C' \text{ are connected components of the same cylinder} \\ \phi_i(C \times \{a_i\}) \text{ and } \phi_i(C' \times \{a_i\}) \text{ belong to the same connected component} \\ \psi_i(C \times \{a_{i+1}\}) \text{ and } \psi_i(C' \times \{a_{i+1}\}) \text{ belong to the same connected component} \end{cases}$$

Then, we define $T_\mathcal{I}$ as $\bar{T}_\mathcal{I}/\sim$, equipped with the projection onto the second factor that we call $f_\mathcal{I}$.

Intuitively, we glue the pairs C, C' of connected components of the same cylinder whose images under the attaching maps are in the same connected component of the critical slice, i.e. those that induce edges with the same endpoints in the multinerve. Hence, we obtain the following corollary using Corollary 3.5:

Corollary 7.22. *$\mathcal{C}\mathrm{R}_{f_\mathcal{I}}(T_\mathcal{I})$ and $\mathrm{M}_f(X, \mathcal{I})$ are isomorphic as combinatorial multigraphs.*

7.4 Convergence results

Recall that the d_{FD} compares metric graphs, whereas the (MultiNerve) Mappers are combinatorial graphs. However, since $\overline{\mathrm{M}}_f(X, \mathcal{I})$ and $\mathrm{R}_{\bar{f}_\mathcal{I}}(\bar{T}_\mathcal{I})$ are essentially the same according to Theorem 7.18, we can use $\mathrm{R}_{\bar{f}_\mathcal{I}}(\bar{T}_\mathcal{I})$ as a metric graph representation of $\overline{\mathrm{M}}_f(X, \mathcal{I})$, when computing the functional distortion distance. Note that we could also use $\mathrm{R}_{\bar{\mathsf{m}}_\mathcal{I}}(\overline{\mathrm{M}}_f(X, \mathcal{I}))$ since it is isomorphic to $\overline{\mathrm{M}}_f(X, \mathcal{I})$ as well according to Lemma 4.2, but its connection to $\mathrm{R}_f(X)$ is unclear. On the opposite, even though d_{FD} is most of the time untractable, its computation is possible with $\mathrm{R}_{\bar{f}_\mathcal{I}}(\bar{T}_\mathcal{I})$ thanks to the sequence of transformations of Definition 7.16. We will see at the end of the section that $\bar{\mathsf{m}}_\mathcal{I}$ and $\bar{\tilde{f}}_\mathcal{I}$ actually coincide on $\overline{\mathrm{M}}_f(X, \mathcal{I})$.

Theorem 7.24 below shows that $\mathrm{R}_{\bar{f}_\mathcal{I}}(\bar{T}_\mathcal{I})$ is close to $\mathrm{R}_f(X)$ if $\mathcal{I}$ has a small granularity. To prove it, we use the following lemma, whose proof is just a simple extension of the one of Proposition 3.1 in [10], and is deferred to Appendix B:

Lemma 7.23. *Let S be a set of pairwise disjoint bounded open intervals, and let* Merge_S *be defined as the composition of all* $\mathrm{Merge}_{a,b}$*, $(a, b) \in S$. Let R_g be a Reeb graph such that* $\mathrm{Crit}(\tilde{g}) \subset \bigcup_{I \in S} I$ *and let $\mathrm{R}_{g'}$ be the Reeb graph of the telescope* $\mathrm{Merge}_S(\mathrm{R}_g)$*. Then $d_{\mathrm{FD}}(\mathrm{R}_g, \mathrm{R}_{g'}) \leq \sup\{\mathrm{length}(I) : I \in S\}$.*

Given a gomic $\mathcal{I}$, we let $\epsilon_1(\mathcal{I}) = \sup\{\text{length}(\tilde{I}) \,:\, I \in \mathcal{I}\}$ and $\epsilon_2(\mathcal{I}) = \sup\{\text{length}(I \cap J) \,:\, I, J \in \mathcal{I}\}$. Note that $\epsilon_1(\mathcal{I})$ and $\epsilon_2(\mathcal{I})$ can be thought of as different types of granularity measures of $\mathcal{I}$. They are both bounded from above by the granularity of $\mathcal{I}$ as defined in Section 2.5.

Theorem 7.24. *Suppose the granularity of the gomic $\mathcal{I}$ is at most ε. For $(\bar{T}_{\mathcal{I}}, \bar{f}_{\mathcal{I}})$ defined as in Definition 7.16, we have $d_{\mathrm{FD}}(\mathrm{R}_{\bar{f}_{\mathcal{I}}}(\bar{T}_{\mathcal{I}}), \mathrm{R}_f(X)) \leq \epsilon_1(\mathcal{I}) + \epsilon_2(\mathcal{I}) + \max\{\epsilon_1(\mathcal{I}), \epsilon_2(\mathcal{I})\} \leq 3\epsilon$.*

Moreover, for $(T_{\mathcal{I}}, f_{\mathcal{I}})$ defined as in Definition 7.21, we have $d_{\mathrm{FD}}(\mathrm{R}_{f_{\mathcal{I}}}(T_{\mathcal{I}}), \mathrm{R}_f(X)) \leq 7\varepsilon/2$.

Proof. We start with the MultiNerve Mapper. By the triangle inequality, it suffices to bound the functional distortion distance for each of the four operations $\mathrm{Merge}_{\mathcal{I}}$, $\mathrm{Shift}_{\mathcal{I}}$, $\mathrm{Split}_{\mathcal{I}}$ and $\mathrm{Merge}'_{\mathcal{I}}$ individually. Let R_1 be the Reeb graph of the telescope $\mathrm{Merge}_{\mathcal{I}}(\mathrm{R}_f(X))$. Similarly, let R_2 be the Reeb graph of $\mathrm{Split}_{\mathcal{I}}(\mathrm{R}_1)$, R_3 be the Reeb graph of $\mathrm{Shift}_{\mathcal{I}}(\mathrm{R}_2)$ and $\mathrm{R}_4 = \mathrm{R}_{\bar{f}_{\mathcal{I}}}(\bar{T}_{\mathcal{I}})$ be the Reeb graph of $\mathrm{Merge}'_{\mathcal{I}}(\mathrm{R}_3)$. Examples of such Reeb graphs can be seen in the left parts of Figure 19.

Then we have $d_{\mathrm{FD}}(\mathrm{R}_{\bar{f}_{\mathcal{I}}}(\bar{T}_{\mathcal{I}}), \mathrm{R}_f(X)) \leq d_{\mathrm{FD}}(\mathrm{R}_f(X), \mathrm{R}_1) + d_{\mathrm{FD}}(\mathrm{R}_1, \mathrm{R}_2) + d_{\mathrm{FD}}(\mathrm{R}_2, \mathrm{R}_3) + d_{\mathrm{FD}}(\mathrm{R}_3, \mathrm{R}_4)$.

- By Lemma 7.23, we have $d_{\mathrm{FD}}(\mathrm{R}_f(X), \mathrm{R}_1) \leq \max\{\epsilon_1(\mathcal{I}), \epsilon_2(\mathcal{I})\}$ and $d_{\mathrm{FD}}(\mathrm{R}_3, \mathrm{R}_4) \leq \epsilon_1(\mathcal{I})$.
- Assume without loss of generality that $\mathrm{Split}_{\mathcal{I}}$ is the composition of all $\mathrm{Split}_{\alpha,\bar{a}}$, where $\bar{a}$ is a critical value of R_1. Since R_1 is obtained from R_2 by taking the composition of all $\mathrm{Merge}_{\bar{a}-\alpha,\bar{a}+\alpha}$, it follows from Lemma 7.23 that $d_{\mathrm{FD}}(\mathrm{R}_1, \mathrm{R}_2) \leq 2\alpha$.
- Since the assumptions of Prop. 7.15 (iii) and Prop. 7.15 (iv) are satisfied by $\mathrm{Shift}_{\mathcal{I}}$, it follows that R_2 and R_3 are isomorphic, because the number, the types and the ordering of the critical values of R_2 are preserved when transformed into R_3. It is then straightforward that the functional distortion distance between R_2 and R_3 is the maximal amplitude of the Shift operations involved. According to the assumptions of Proposition 7.15 (iii) and Proposition 7.15 (iv), these amplitudes are all bounded by $\epsilon_2(\mathcal{I})$.

The result follows by letting $\alpha \to 0$.

Concerning the Mapper, the result is obtained by adding an extra $\epsilon/2$ to the previous upper bound, which corresponds to the functional distortion distance cost of gluing edges with the same endpoints. □ □

Note that a similar result can be obtained directly by using the convergence result of the so-called *interleaving distance*—see Theorem 4.1 in [29], and the strong equivalence between the functional distortion distance and this interleaving distance—see Theorem 14 in [5]. However, the upper bound gets larger (7ϵ) and there is no clear intuition on the Reeb graph used to represent the Mapper (also called geometric Mapper) in [29].

Finally, Theorems 7.24 and 2.11 allow us to derive the following result with the triangle inequality:

Corollary 7.25. *Let X be a topological space, and let $f, g : X \to \mathbb{R}$ be two Morse-type functions with continuous sections. Let $(\bar{T}_{\mathcal{I}}(X, f), \bar{f}_{\mathcal{I}})$ (resp. $(\bar{T}_{\mathcal{I}}(X, g), \bar{g}_{\mathcal{I}})$) denote the pair computed with function f (resp. g) as in Definition 7.16. Finally, let $\mathcal{I}$ be a gomic of granularity at most ϵ. Then:*

$$d_{\mathrm{FD}}(\mathrm{R}_{\bar{f}_{\mathcal{I}}}(\bar{T}_{\mathcal{I}}(X, f)), \mathrm{R}_{\bar{g}_{\mathcal{I}}}(\bar{T}_{\mathcal{I}}(X, g))) \leq \|f - g\|_\infty + 6\epsilon.$$

Moreover, for $(T_{\mathcal{I}}(X, f), f_{\mathcal{I}})$ and $(T_{\mathcal{I}}(X, g), g_{\mathcal{I}})$ computed as in Definition 7.21, we have:

$$d_{\mathrm{FD}}(\mathrm{R}_{f_{\mathcal{I}}}(T_{\mathcal{I}}(X, f)), \mathrm{R}_{g_{\mathcal{I}}}(T_{\mathcal{I}}(X, g))) \leq \|f - g\|_\infty + 7\epsilon.$$

7.5 An alternative proof of Theorem 4.3

We can prove Theorem 4.3 again by studying the effect of the transformation defined in Definition 7.16 on the extended persistence diagram of f. These effect is illustrated in Figure 20. There are two grids in this figure: the one with solid lines is defined by the interval endpoints, while the one with dotted lines is defined by midpoints of proper subintervals and intersections. In the following, we use the term *cell* to designate a

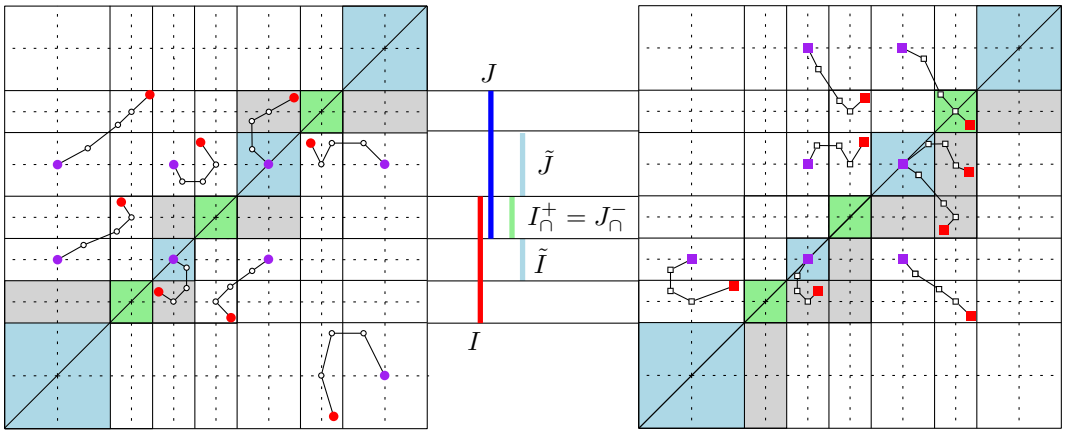


Figure 20: The left panel displays the trajectories of points in Ord (disks above the diagonal) and Rel (disks below the diagonal) while the right panel displays the trajectories of points in Ext. For both diagrams, the original point is red, the final point is purple, and intersections and proper subintervals are colored in green and light blue respectively.

rectangle of the first grid. Cells are closed if they correspond to proper subintervals for both coordinates, they are open if they correspond to intersections for both coordinates, and they are neither closed nor open otherwise. Blue and green cells in Figure 20 correspond to squares associated to a proper subinterval (blue) or intersection (green). We can now interpret the effects of the transformations in (7.16) on the persistence diagram visually:

- $\mathrm{Merge}_{\mathcal{I}}$ snaps all the points to the second grid by Lemma 7.6.
- $\mathrm{Split}_{\mathcal{I}}$ moves the points to one of the four possible quarters inside their cell, depending on the point's type by Lemma 7.12. More precisely, ordinary points are moved to the down-left quarter, extended points are moved to the up-left quarter, and relative points are moved to the up-right quarter.
- Then, $\mathrm{Shift}_{\mathcal{I}}$ moves the points to a neighboring cell by Lemma 7.14. This neighboring cell is given by the point's type (as in the case of $\mathrm{Split}_{\mathcal{I}}$) and by the coordinates of the point. For instance, an extended point (x, y) lying in the row of a green cell and in the column of another green cell, has coordinates that both belong to interval intersections. Then, this point is moved to the upper-left neighboring cell. Differently, an extended point whose abscissa (resp. ordinate) is in an intersection and whose ordinate (resp. abscissa) is not, is only moved to the left (resp. upper) cell. The same can be said for ordinary (resp. relative) points by changing up-left to down-left (resp. up-right). Points whose coordinates both belong to proper subintervals are not moved by $\mathrm{Shift}_{\mathcal{I}}$, regardless of their type.
- Finally, $\mathrm{Merge}'_{\mathcal{I}}$ re-snaps the points to the second grid by Lemma 7.6.

Thus, each point of $\mathrm{Dg}(f)$ can be tracked through the successive operations of (7.16), and this tracking leads to the following elementary observations:

1. The points of $\mathrm{Ord}(f)$ or $\mathrm{Rel}(f)$ that end their course on the diagonal after the sequence of operations of Definition 7.16 disappear in $\mathrm{Dg}(\bar{f}_{\mathcal{I}})$. This is because ordinary and relative points cannot be located on the diagonal. The rest of the points of $\mathrm{Ord}(f)$ or $\mathrm{Rel}(f)$ are preserved with the same type in $\mathrm{Dg}(\bar{f}_{\mathcal{I}})$.
2. Differently, all the points of $\mathrm{Ext}(f)$ are preserved with the same type (Ext) in $\mathrm{Dg}(\bar{f}_{\mathcal{I}})$. However, some of the points of $\mathrm{Ext}^-(f)$ may end their course on or across the diagonal after the sequence of operations (7.16), thus switching from $\mathrm{Ext}^-(f)$ to $\mathrm{Ext}^+(\bar{f}_{\mathcal{I}})$.

3. All the points lie in the rows and columns of blue cells after $\text{Shift}_{\mathcal{I}}$. Therefore, the points that end their course on the diagonal after the sequence of operations of Definition 7.16 are the ones located in blue cells after $\text{Shift}_{\mathcal{I}}$.

4. Since transfers between cells occur only during $\text{Shift}_{\mathcal{I}}$, a point p that is not in a blue or green cell initially ends up in a blue cell B after $\text{Shift}_{\mathcal{I}}$ if and only if:

   - p is extended and it is in the down, right, or down-right neighboring cell of B (grey cells in the right diagram of Figure 20), or
   - p is ordinary and it is in the up neighboring cell of B (grey cells above the diagonal in the left diagram of Figure 20), or
   - p is relative and it is in the down neighboring cell of B (grey cells below the diagonal in the left diagram of Figure 20).

5. Points that belong to a blue or green cell initially are snapped to the diagonal by $\text{Merge}_{\mathcal{I}}$. Among them, those that belong to a blue cell stay there until the end, whereas those that belong to a green cell eventually leave it—they end up in a blue cell after $\text{Shift}_{\mathcal{I}}$ if they are ordinary or relative, while they end up in a white cell above the diagonal if they are extended.

The outcome of these observations is the following one. Observations 1, 3, 4, 5 imply that the points of $\text{Ord}(f)$ that disappear in $\text{Dg}(\bar{f}_{\mathcal{I}})$ are the ones located initially in the staircase made of the blue, green and grey areas above the diagonal in the left panel of Figure 20, which is nothing but $Q^{\mathcal{I}}_O$. Similarly, the points of $\text{Rel}(f)$ that disappear in $\text{Dg}(\bar{f}_{\mathcal{I}})$ are the ones located initially in the staircase made of the blue, green and grey areas below the diagonal in the left panel of Figure 20, which is nothing but $Q^{\mathcal{I}}_R$. Finally, observations 2, 3, 4, 5 imply that the points of $\text{Ext}^-(f)$ that switch to $\text{Ext}^+(\bar{f}_{\mathcal{I}})$ are the ones located initially in the staircase made of the blue, green and grey areas below the diagonal in the right panel of Figure 20, which is nothing but $Q^{\mathcal{I}}_{E^-}$. The rest of the points of $\text{Dg}(f)$ are preserved (albeit shifted) with the same type (Ord, Rel, Ext^+, Ext^-) in $\text{Dg}(\bar{f}_{\mathcal{I}})$. Hence, there is a perfect matching between:

(i) $\text{Ord}(\bar{f}_{\mathcal{I}})$ and $\text{Ord}(f) \setminus Q^{\mathcal{I}}_O$
(ii) $\text{Rel}(\bar{f}_{\mathcal{I}})$ and $\text{Rel}(f) \setminus Q^{\mathcal{I}}_R$
(iii) $\text{Ext}^-(\bar{f}_{\mathcal{I}})$ and $\text{Ext}^-(f) \setminus Q^{\mathcal{I}}_{E^-}$
(iv) $\text{Ext}^+(\bar{f}_{\mathcal{I}})$ and $\text{Ext}^+(f) \cup (\text{Ext}^-(f) \cap Q^{\mathcal{I}}_{E^-})$

This, combined with Theorem 2.8, is equivalent to Theorem 4.3. This matching also shows that the critical points of $\bar{f}_{\mathcal{I}}$ and $f_{\mathcal{I}}$ are located at the midpoints of proper subintervals of the gomic's elements. Hence, $\tilde{\bar{f}}_{\mathcal{I}}$ and $\tilde{f}_{\mathcal{I}}$ actually coincide with $\bar{\mathsf{m}}_{\mathcal{I}}$ and $\mathsf{m}_{\mathcal{I}}$, which allows us to state this final result:

Theorem 7.26. *Let X be a topological space and $f : X \to \mathbb{R}$ be a Morse-type function. Let $\mathcal{I}$ be a gomic with granularity at most ϵ. Let $\bar{\mathsf{m}}_{\mathcal{I}}$ and $\mathsf{m}_{\mathcal{I}}$ be as in Definition 4.1. Then:*

$$d_{\text{b}}(\text{Dg}(\bar{\mathsf{m}}_{\mathcal{I}}), \text{Dg}(\tilde{f})) \leq \epsilon/2, \tag{13}$$

$$d_{\text{b}}(\text{Dg}(\mathsf{m}_{\mathcal{I}}), \text{Dg}(\tilde{f})) \leq \epsilon. \tag{14}$$

Moreover, in both cases, the matching achieving the distance is actually a bijection preserving types.

8 Discrete Case

In this section we discuss the approximation of the (MultiNerve) Mapper and of its signature (5) when the pair (X, f) is known only through a finite set of sample points equipped with function values. Throughout the section, X is a compact Riemannian manifold of $\mathbb{R}^D$, $f : X \to \mathbb{R}$ a Morse-type function, $\mathcal{I}$ a gomic, and X_n a point cloud in X with n points.

8.1 Approximation tools

Rips complex. All constructions take a neighborhood graph as input, such as for instance the 1-skeleton graph of the Rips complex, defined as follows:

Definition 8.1. *Let $\delta \geq 0$ be a scale parameter. The* Rips complex *of X_n of parameter δ is the simplicial complex* $\mathrm{Rips}_\delta(X_n)$ *defined by:*

$$\{x_{i_0}, ..., x_{i_k}\} \in \mathrm{Rips}_\delta(X_n) \Leftrightarrow \|x_{i_p} - x_{i_q}\| \leq \delta \text{ for any } 0 \leq p, q \leq k, .$$

Its 1-skeleton graph is called the Rips graph *of parameter δ and denoted by* $\mathrm{Rips}^1_\delta(X_n)$.

Moreover, given a geometric realization $|\mathrm{Rips}_\delta(X_n)|$, we let $f^{\mathrm{PL}} : |\mathrm{Rips}_\delta(X_n)| \to \mathbb{R}$ denote the piecewise-linear interpolation of f along the simplices of $\mathrm{Rips}_\delta(X_n)$*, and we let* $\mathrm{R}_{f^{\mathrm{PL}}}(|\mathrm{Rips}_\delta(X_n)|)$ *denote its Reeb graph, with induced function $\tilde{f}^{\mathrm{PL}}$.*

Geometric quantities. Two geometric quantities that assess the smoothness of topological spaces will be used in the hypotheses of the results in this section.

- *Reach.* The *medial axis* of $X \subset \mathbb{R}^D$ is the set of points in $\mathbb{R}^D$ with at least two nearest neighbors in X:

  $$\mathrm{med}(X) = \{y \in \mathbb{R}^D \,:\, \mathrm{card}\{x \in X \,:\, \|y - x\| = \|y, X\|\} \geq 2\},$$

  where $\|y, X\| = \inf\{\|y - x\| \,:\, x \in X\}$. The *reach* of X, denoted by $\mathrm{rch}(X)$, is the distance of X to its medial axis:

  $$\mathrm{rch}(X) = \inf\{\|x - m\| \,:\, x \in X, m \in \mathrm{med}(X)\}.$$

- *Convexity radius.* A set $Y \subseteq X$ is said to be *convex* whenever every geodesic path in X between two points of Y stays in Y. The *convexity radius* of X is the smallest radius ρ for which every geodesic ball in X of radius less than ρ is convex.

Regularity of the function. Intuitively, approximating a Reeb graph computed with a function f that has large variations is more difficult than for a smooth function, for some notion of regularity that we now specify. Our result is given in a general setting by considering the *modulus of continuity of f*.

Definition 8.2. *Let $f : X \to \mathbb{R}$ be a Morse-type function. The* modulus of continuity *ω_f of f is:*

$$\omega_f : \left\{ \begin{array}{ccc} \mathbb{R}_+ & \to & \mathbb{R}_+ \\ \delta & \mapsto & \sup\{|f(x) - f(x')| \,:\, x, x' \in X,\ \|x - x'\| \leq \delta\} \end{array} \right.$$

It follows from the Definition 8.2 that ω_f satisfies :

1. $\omega_f(\delta) \to \omega(0) = 0$ as $\delta \to 0$;
2. ω_f is nonnegative and non-decreasing on $\mathbb{R}^+$;
3. ω_f is subadditive: $\omega_f(\delta_1 + \delta_2) \leq \omega_f(\delta_1) + \omega_f(\delta_2)$ for any δ_1, $\delta_2 > 0$;
4. ω_f is continous on $\mathbb{R}^+$.

Modulus of continuity. More generally, we say that a function ω defined on $\mathbb{R}_+$ is *a modulus of continuity* if it satisfies the four properties above, and we say that it is *a modulus of continuity for f* if, in addition, we have $\omega \geq \omega_f$.

8.2 Reeb graph

The following theorem states that the Rips complex of a point cloud can be used as a proxy for the original space X. Hence, it is possible to approximate the Reeb graph of X, in the bottleneck distance, by computing the Reeb graph of the Rips complex built on top of the point cloud.

Theorem 8.3 (Theorem 4.6 and Remark 2 in [23]). *Assume X has positive reach* rch *and convexity radius ρ. Let $\delta \geq 0$ be a scale parameter, and let ω be a modulus of continuity for f.*

If $4d_{\mathrm{H}}(X, X_n) \leq \delta \leq \min\left\{\frac{1}{4}\mathrm{rch}, \frac{1}{4}\rho\right\}$, then:

$$d_{\mathrm{b}}(\mathrm{Ext}_1^-(\tilde{f}), \mathrm{Ext}_1^-(\tilde{f}^{\mathrm{PL}})) \leq 2\omega(\delta).$$

Note that the original version of this theorem is only proven for Lipschitz functions in [23], but it extends at no cost, i.e. with the same proof, to functions with modulus of continuity.

Theorem 8.4 (Theorem 2 in [15]). *Assume X has convexity radius $\rho > 0$. Let $\delta > 0$ be a scale parameter, and let ω be a modulus of continuity for f.*

If $4d_{\mathrm{H}}(X, X_n) \leq \delta \leq \rho$, then:

$$\max\{d_{\mathrm{b}}(\mathrm{Ord}_0(\tilde{f}), \mathrm{Ord}_0(\tilde{f}^{\mathrm{PL}})), d_{\mathrm{b}}(\mathrm{Ext}_0^+(\tilde{f}), \mathrm{Ext}_0^+(\tilde{f}^{\mathrm{PL}})), d_{\mathrm{b}}(\mathrm{Rel}_1(\tilde{f}), \mathrm{Rel}_1(\tilde{f}^{\mathrm{PL}}))\} \leq \omega(\delta).$$

Again, the original version of this theorem is only proven for Lipschitz functions in [15], but it extends at no cost to functions with modulus of continuity. Moreover, three more remarks need to be made. Firstly, this theorem is originally stated only for the ordinary part of the persistence diagrams but its proof extends to the full extended filtrations at no extra cost. Secondly, it is stated for a nested pair of Rips complexes, however, as pointed out in Section 4.3 in [15], in 0-dimensional homology a single Rips graph is sufficient for the theorem to hold. Thirdly, its approximation function is piecewise-constant and not piecewise-linear as in this article. However, the filtrations induced by the sublevel sets and the superlevel sets of the piecewise-constant function are actually lower- and upper-star filtrations, and it is known in that case that piecewise-linear and piecewise-constant functions induce the same persistence diagram. See Section 2.5 of [28] for a proof of this statement.

Combining the two theorems gives the following complete approximation result:

Theorem 8.5. *Under the assumptions of Theorem 8.3, we have $d_{\mathrm{b}}(\mathrm{Dg}(\tilde{f}), \mathrm{Dg}(\tilde{f}^{\mathrm{PL}})) \leq 2\omega(\delta)$.*

8.3 (MultiNerve) Mapper

We report three possible constructions for the (MultiNerve) Mapper from the pair (X_n, f): the first is from the original Mapper paper [32], the second is inspired from the graph-induced complex paper [21], and the third is simply the Rips complex approximation. Given a choice of neighborhood parameter δ and the corresponding Rips graph, the construction from [32] uses the vertices as witnesses for the connected components of the pullback cover on $\mathrm{Rips}_\delta^1(X_n)$ and for their pairwise intersections. Differently, the construction from [21] uses the edges as witnesses for the pairwise intersections. Thus, both constructions have the same vertex set but potentially different edge sets.

Vertex-based connectivity. Given an arbitrary interval I in $\mathbb{R}$, the preimage of I in X_n is defined to be $X_n \cap f^{-1}(I)$, and its connected components are defined to be the connected components of the induced subgraph $\mathrm{Rips}_\delta^1(X_n \cap f^{-1}(I))$. Then, the vertices in the (MultiNerve) Mapper are the connected components of the preimages of the intervals $I \in \mathcal{I}$. Given two intersecting intervals I, J of $\mathcal{I}$, given a connected component C_I in the preimage of I and a connected component C_J in the preimage of J, the corresponding vertices are connected by an edge in the Mapper if there is a connected component in the preimage of $I \cap J$ that is contained in both C_I and C_J; in the MultiNerve Mapper, there are as many copies of this edge as there are connected components in the preimage of $I \cap J$ that are contained in $C_I \cap C_J$. We denote these two constructions by $\mathrm{M}_f^\bullet(\mathrm{Rips}_\delta^1(X_n), \mathcal{I})$ and $\overline{\mathrm{M}}_f^\bullet(\mathrm{Rips}_\delta^1(X_n), \mathcal{I})$ respectively. Moreover, functions $\mathsf{m}_{\mathcal{I}}^\bullet$ and $\bar{\mathsf{m}}_{\mathcal{I}}^\bullet$ can be defined on these constructions exactly like in Definition 4.1.

Edge-based connectivity. The vertex set of the (MultiNerve) Mapper is the same as in the previous construction. Now, for any intersecting intervals I, J of $\mathcal{I}$, we redefine the preimage of the intersection $I \cap J$ to be the subset of $\mathrm{Rips}_\delta^1(X_n)$ spanned not only by the points of $X_n \cap f^{-1}(I \cap J)$ and the graph edges connecting them, but also by the relative interiors of the edges of $\mathrm{Rips}_\delta^1(X_n)$ that have one vertex in $X_n \cap f^{-1}(I)$ and the other in $X_n \cap f^{-1}(J)$. Then, given a connected component C_I in the preimage of I and a connected component C_J in the preimage of J, we connect the corresponding vertices by an edge in the Mapper if there is a connected component of the redefined preimage of $I \cap J$ that connects[7] C_I and C_J in $\mathrm{Rips}_\delta^1(X_n)$; in the MultiNerve Mapper, we add as many copies of this edge as there are connected components in the redefined preimage of $I \cap J$ that connect C_I and C_J. We denote these two constructions by $\mathrm{M}_f^{\triangle}(\mathrm{Rips}_\delta^1(X_n), \mathcal{I})$ and $\overline{\mathrm{M}}_f^{\triangle}(\mathrm{Rips}_\delta^1(X_n), \mathcal{I})$ respectively. Again, we also define functions $\mathsf{m}_{\mathcal{I}}^{\triangle}$ and $\bar{\mathsf{m}}_{\mathcal{I}}^{\triangle}$ on these constructions.

Rips complex. As for Reeb graphs, one can compute (MultiNerve) Mappers $\overline{\mathrm{M}}_{f^{\mathrm{PL}}}(|\mathrm{Rips}_\delta(X_n)|, \mathcal{I})$ and $\mathrm{M}_{f^{\mathrm{PL}}}(|\mathrm{Rips}_\delta(X_n)|, \mathcal{I})$ from a geometric realization of a Rips complex built on top of X_n with parameter δ, using the piecewise-linear extension f^{PL}. Seeing $|\mathrm{Rips}_\delta(X_n)|$ as a topological space, we also define the telescopes $\bar{T}_{\mathcal{I}}(|\mathrm{Rips}_\delta(X_n)|, f^{\mathrm{PL}})$ and $T_{\mathcal{I}}(|\mathrm{Rips}_\delta(X_n)|, f^{\mathrm{PL}})$ with Definition 7.16 and Definition 7.21, with corresponding projections onto the second factor $\bar{f}_{\mathcal{I}}^{\mathrm{PL}}$ and $f_{\mathcal{I}}^{\mathrm{PL}}$. We recall that $\mathcal{C}\mathrm{R}_{\bar{f}_{\mathcal{I}}^{\mathrm{PL}}}(\bar{T}_{\mathcal{I}}(|\mathrm{Rips}_\delta(X_n)|, f^{\mathrm{PL}}))$ and $\mathcal{C}\mathrm{R}_{f_{\mathcal{I}}^{\mathrm{PL}}}(T_{\mathcal{I}}(|\mathrm{Rips}_\delta(X_n)|, f^{\mathrm{PL}}))$ are isomorphic to $\overline{\mathrm{M}}_{f^{\mathrm{PL}}}(|\mathrm{Rips}_\delta(X_n)|, \mathcal{I})$ and $\mathrm{M}_{f^{\mathrm{PL}}}(|\mathrm{Rips}_\delta(X_n)|, \mathcal{I})$ respectively, according to Theorem 7.18 and Corollary 7.22. Moreover, the induced maps $\tilde{\bar{f}}_{\mathcal{I}}^{\mathrm{PL}} = \bar{\mathsf{m}}_{\mathcal{I}}^{\mathrm{PL}}$ and $\tilde{f}_{\mathcal{I}}^{\mathrm{PL}} = \mathsf{m}_{\mathcal{I}}^{\mathrm{PL}}$ are related to $\tilde{f}^{\mathrm{PL}}$ according to Theorem 7.26. See Figure 21 for an illustration of these functions.

8.4 Relationships between the constructions

In each of the three constructions detailed above, the Mapper is included in the MultiNerve Mapper by definition. Moreover, the preimages of the intersections in the second construction are supersets of the preimages in the first construction, and two different connected components in the same preimage in the first construction cannot be connected in the second construction, therefore the (MultiNerve) Mapper from the first construction is included in its counterpart from the second construction. Hence the following diagram of inclusions:

$$\begin{array}{ccc} \mathrm{M}_f^{\bullet}(\mathrm{Rips}_\delta^1(X_n), \mathcal{I}) & \longrightarrow & \overline{\mathrm{M}}_f^{\bullet}(\mathrm{Rips}_\delta^1(X_n), \mathcal{I}) \\ \downarrow & & \downarrow \\ \mathrm{M}_f^{\triangle}(\mathrm{Rips}_\delta^1(X_n), \mathcal{I}) & \longrightarrow & \overline{\mathrm{M}}_f^{\triangle}(\mathrm{Rips}_\delta^1(X_n), \mathcal{I}) \end{array} \tag{15}$$

The vertical inclusions become equalities when there are no *intersection-crossing edges* in the Rips graph, defined as follows:

Definition 8.6. *An edge $[u, v]$ of the Rips graph is* interval-crossing *if there is an interval $I \in \mathcal{I}$ such that $I \subseteq (\min\{f(u), f(v)\}, \max\{f(u), f(v)\})$. It is* intersection-crossing *if there is a pair of intervals $I, J \in \mathcal{I}$ such that $\emptyset \neq I \cap J \subseteq (\min\{f(u), f(v)\}, \max\{f(u), f(v)\})$.*

Indeed, in the absence of intersection-crossing edges, each connected component in the preimage of an interval intersection in the second construction contains a vertex and therefore deform retracts onto the corresponding connected component in the first construction. Hence:

Lemma 8.7. *If there are no intersection-crossing edges in $\mathrm{Rips}_\delta^1(X_n)$, then $\mathrm{M}_f^{\bullet}(\mathrm{Rips}_\delta^1(X_n), \mathcal{I})$ is isomorphic to $\mathrm{M}_f^{\triangle}(\mathrm{Rips}_\delta^1(X_n), \mathcal{I})$ as combinatorial multigraphs. The same is true for $\overline{\mathrm{M}}_f^{\bullet}(\mathrm{Rips}_\delta^1(X_n), \mathcal{I})$ and $\overline{\mathrm{M}}_f^{\triangle}(\mathrm{Rips}_\delta^1(X_n), \mathcal{I})$.*

Concerning the relation between the first two constructions and the third one, we have:

[7] By which we mean that the closure of the connected component in $\mathrm{Rips}_\delta^1(X_n)$ contains points from C^I and from C^J.

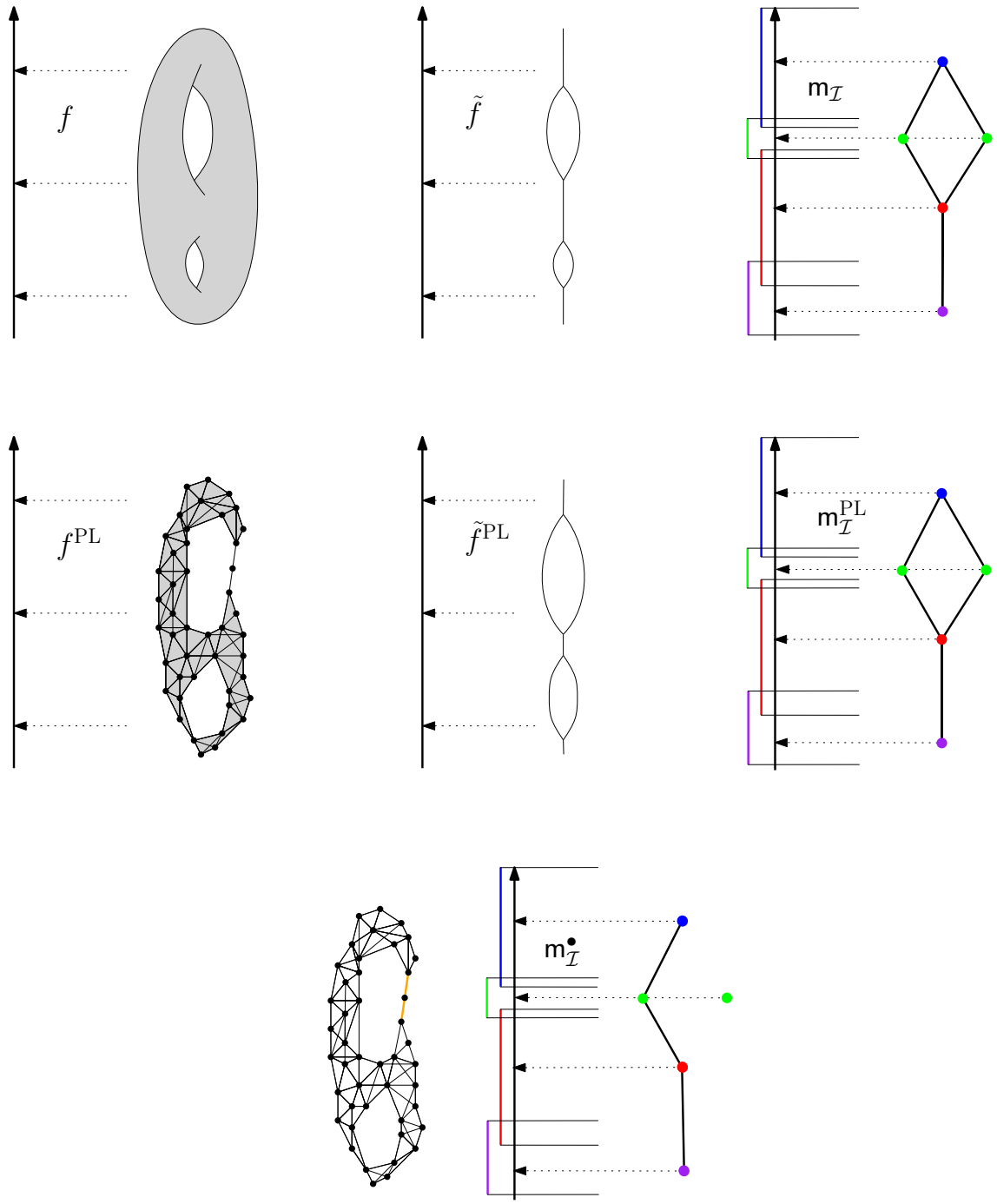


Figure 21: Examples of the function defined on the original space (left column), its induced function defined on the Reeb graph (middle column) and the function defined on the Mapper (right column). Note that the Mapper computed from the geometric realization of the Rips complex (middle row, right) is not isomorphic to the standard Mapper (last row), since there are two intersection-crossing edges in the Rips complex (outlined in orange).

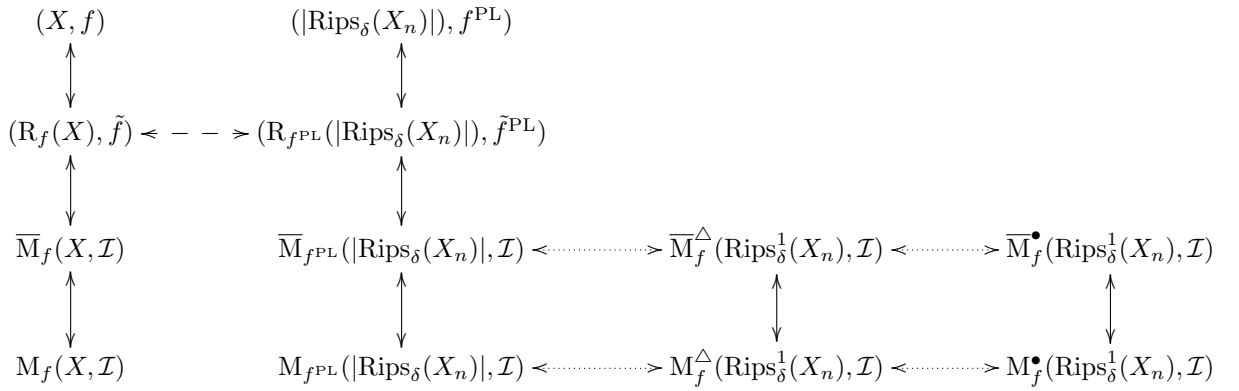
Lemma 8.8. *If there are no interval-crossing edges in* $\mathrm{Rips}_\delta^1(X_n)$*, then* $\mathrm{M}_{f^{\mathrm{PL}}}(|\mathrm{Rips}_\delta(X_n)|, \mathcal{I})$ *is isomorphic to* $\mathrm{M}_f^{\triangle}(\mathrm{Rips}_\delta^1(X_n), \mathcal{I})$ *as combinatorial multigraphs. The same is true for* $\overline{\mathrm{M}}_{f^{\mathrm{PL}}}(|\mathrm{Rips}_\delta(X_n)|, \mathcal{I})$ *and* $\overline{\mathrm{M}}_f^{\triangle}(\mathrm{Rips}_\delta^1(X_n), \mathcal{I})$*.*

Proof. Note that $\mathrm{M}_{f^{\mathrm{PL}}}(|\mathrm{Rips}_\delta(X_n)|, \mathcal{I})$ and $\overline{\mathrm{M}}_{f^{\mathrm{PL}}}(|\mathrm{Rips}_\delta(X_n)|, \mathcal{I})$ are the same as $\mathrm{M}_{f^{\mathrm{PL}}}(|\mathrm{Rips}_\delta^1(X_n)|, \mathcal{I})$ and $\overline{\mathrm{M}}_{f^{\mathrm{PL}}}(|\mathrm{Rips}_\delta^1(X_n)|, \mathcal{I})$ respectively, since only the connected component of the preimages of intervals are involved in the construction of the (MultiNerve) Mapper. Hence, for the rest of the proof we set the domain of f^{PL} to be $|\mathrm{Rips}_\delta^1(X_n)|$. Every connected component in the preimage through f^{PL} of an interval of $\mathcal{I}$ must contain a vertex, therefore it deform retracts onto the corresponding preimage through f. Hence the vertex sets of the aforementioned simplicial posets are the same. Every connected component in the preimage through f^{PL} of an interval intersection $I \cap J$ either contains a vertex, in which case it deform retracts onto the corresponding preimage through f in the vertex-based connectivity, or it does not contain any vertex, in which case the edge of the Rips graph that contains the connected component creates an edge in the (MultiNerve) Mapper in the edge-based connectivity. □ □

See Figure 22 for an example showing the importance of the hypothesis in the lemma.

8.5 Relationships between the signatures

Relationships between the (MultiNerve) Mapper constructions. The following diagram summarizes the relationships between the various (MultiNerve) Mapper constructions:



(16)

The vertical arrows between the first and second rows are provided by Theorem 2.8. The ones between the second, third and fourth rows are given by Eqs. (5) and (6). The dotted horizontal arrows are provided by Lemmas 8.7 and 8.8. Finally, the dashed horizontal arrow is given by Theorem 8.5.

Approximation of the (MultiNerve) Mapper. We then derive from (16) the following approximation guarantee:

Theorem 8.9. *Under the assumptions of Theorem 8.3, and given a gomic* $\mathcal{I}$*, we have:*

$$d_{\mathcal{I}}\left(\mathrm{Dg}(\overline{\mathrm{M}}_f(X, \mathcal{I})), \mathrm{Dg}(\overline{\mathrm{M}}_{f^{\mathrm{PL}}}(|\mathrm{Rips}_\delta(X_n)|, \mathcal{I}))\right) \leq 2\omega(\delta).$$

If furthermore there are no interval-crossing edges, then $\overline{\mathrm{M}}_{f^{\mathrm{PL}}}(|\mathrm{Rips}_\delta(X_n)|, \mathcal{I})$ *and* $\overline{\mathrm{M}}_f^{\triangle}(\mathrm{Rips}_\delta^1(X_n), \mathcal{I})$ *are isomorphic as combinatorial multigraphs.*

If there are no intersection-crossing edges either, then $\overline{\mathrm{M}}_{f^{\mathrm{PL}}}(|\mathrm{Rips}_\delta(X_n)|, \mathcal{I})$ *and* $\overline{\mathrm{M}}_f^{\bullet}(\mathrm{Rips}_\delta^1(X_n), \mathcal{I})$ *are also isomorphic as combinatorial multigraphs.*

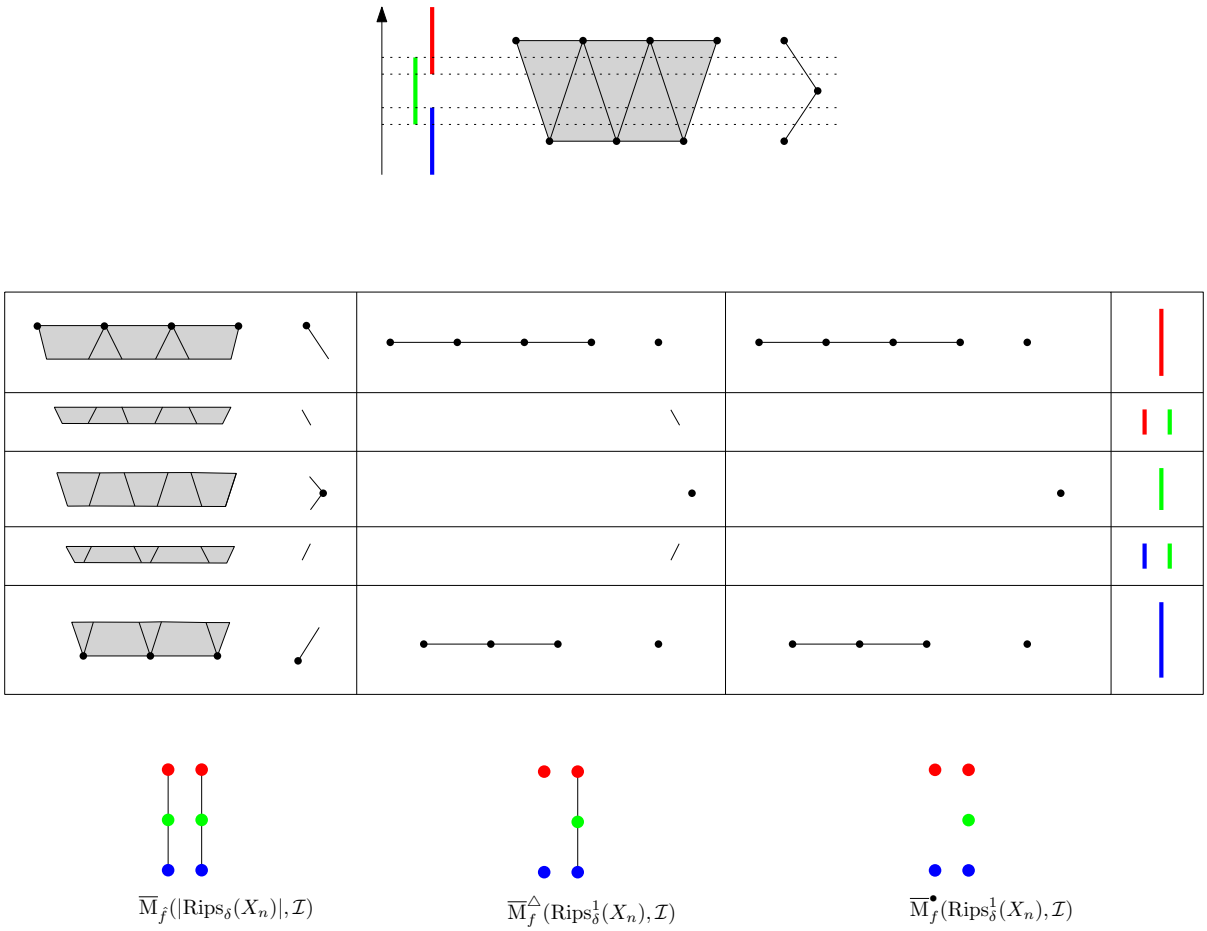


Figure 22: We study a Rips complex $\mathrm{Rips}_\delta(X_n)$ built on top of a point cloud X_n with ten points. This complex has two connected components. We also compute (MultiNerve) Mappers with the height function, whose image is covered by three intervals. We display the preimages of the intervals and their intersections for $\overline{\mathrm{M}}_{f^{\mathrm{PL}}}(|\mathrm{Rips}_\delta(X_n)|,\mathcal{I})$, $\overline{\mathrm{M}}_f^{\triangle}(\mathrm{Rips}_\delta^1(X_n),\mathcal{I})$ and $\overline{\mathrm{M}}_f^{\bullet}(\mathrm{Rips}_\delta^1(X_n),\mathcal{I})$. The edges of the right connected component are intersection-crossing but not interval-crossing, so $\overline{\mathrm{M}}_f^{\triangle}(\mathrm{Rips}_\delta^1(X_n),\mathcal{I})$ recovers it correctly while $\overline{\mathrm{M}}_f^{\bullet}(\mathrm{Rips}_\delta^1(X_n),\mathcal{I})$ fails to. The edges of the left connected component are interval-crossing, so both $\overline{\mathrm{M}}_f^{\bullet}(\mathrm{Rips}_\delta^1(X_n),\mathcal{I})$ and $\overline{\mathrm{M}}_f^{\triangle}(\mathrm{Rips}_\delta^1(X_n),\mathcal{I})$ fail to recover the connected component.

The same result holds for $\mathrm{M}_{f^{\mathrm{PL}}}(|\mathrm{Rips}_\delta(X_n)|,\mathcal{I})$, $\mathrm{M}_f^{\triangle}(\mathrm{Rips}_\delta^1(X_n),\mathcal{I})$ and $\mathrm{M}_f^{\bullet}(\mathrm{Rips}_\delta^1(X_n),\mathcal{I})$, provided $d_{\mathcal{I}}$ is replaced by the bottleneck distance with the appropriate extended staircase $Q_E^{\mathcal{I}}$. Thus, we can construct discrete (MultiNerve) Mappers whose signatures approximate the ones of the corresponding continuous structures $\mathrm{M}_f(X,\mathcal{I})$ and $\overline{\mathrm{M}}_f(X,\mathcal{I})$.

Approximation of the (MultiNerve) Mapper signature. In some situations, one is merely interested in approximating the signatures of $\mathrm{M}_f(X,\mathcal{I})$ and $\overline{\mathrm{M}}_f(X,\mathcal{I})$ without actually building corresponding discrete (MultiNerve) Mappers. In such cases, one can simply apply the scalar fields analysis approach of [13] to approximate $\mathrm{Dg}(f)$, then remove the points from $(\mathrm{Ext}_1^+(f) \cup \mathrm{Ord}_1(f))$ as well as the points lying in their corresponding staircases, to get an approximation of the signatures:

Theorem 8.10. *Under the assumptions of Theorem 8.3, and given a gomic $\mathcal{I}$, let* Dg *denote the extended*

persistence diagram computed by the algorithm of [15], and then pruned by removing the points of the Ext^+_1 *and* Ord_1 *subdiagrams as well as the points located in the staircase corresponding to their type. Then this diagram approximates the signature of* $\overline{\mathrm{M}}_f(X,\mathcal{I})$ *as follows:*

$$d_{\mathcal{I}}\left(\mathrm{Dg}(\overline{\mathrm{M}}_f(X,\mathcal{I})),\,\mathrm{Dg}\right) \leq 2\omega(\delta).$$

The same bound applies for the approximation of $\mathrm{Dg}(\mathrm{M}_f(X,\mathcal{I}))$*, provided the staircase* $Q^{\mathcal{I}}_{E^-}$ *is replaced by its extended version* $Q^{\mathcal{I}}_{E}$ *in the definitions of* $d_{\mathcal{I}}$ *and* Dg*.*

Note that this result holds much more generally than Theorem 8.9, however there may be no discrete (MultiNerve) Mapper construction associated with the approximate diagram Dg.

9 Conclusion

Here we proposed a theoretical framework for the analysis of the structure and stability of the Mapper. In particular, we showed that the topological structure of the (MultiNerve) Mapper is nothing but a simplification of the one of the Reeb graph. From this we derived a bag-of-features type signature for the (MultiNerve) Mapper, which gives a complete description of the features present in the graph (trunks, branches, holes). We related this signature to the persistence of the Reeb graph, then to the persistence of the pair (X, f) itself, using various staircases induced by the interval cover $\mathcal{I}$. This allowed us to determine the structure of the Mapper from the corresponding persistence diagrams. We also gave stability guarantees to the signature, with respect to perturbations of the function, of its domain, or of the interval cover. These guarantees make it possible to predict the structural changes that may occur in the Mapper under such perturbations. Then, we refined the analysis by showing that the MultiNerve Mapper is actually the Reeb graph of a perturbed pair (X', f') that can be obtained from the original pair (X, f) through a sequence of controllable elementary perturbations. These perturbations allowed us to bound the functional distortion distance between the (MultiNerve) Mapper and the Reeb graph, and are of an independent interest [9, 10]. Finally, we discussed the construction of the Mapper and the approximation of its signature from point cloud data, relating the structure of the discrete Mapper to the one of its continuous version.

We plan to investigate two directions for future work:

- The first direction deals with the use of the (MultiNerve) Mapper in Machine Learning. Is it possible to feed such graphs to Machine Learning methods (using e.g. graph kernels) while taking into account the specificities of these graphs?
- The second direction concerns vector-valued functions. Indeed, the Mapper's definition is not bound to the codomain being $\mathbb{R}$ in any fundamental way. However, the stratification of the domain X by the critical level sets in the general setting is less well understood. Chattopadhyay et al. [12] studied the local configuration of the Reeb space around singular values for bivariate functions. In a different context, assuming X is a combinatorial manifold and f is piecewise-linear, Edelsbrunner et al. [24] showed that the corresponding Reeb space is a stratified space, and they characterized its coarsest stratification. We believe part of our framework could be generalized to these settings, leading to convergence and stability results that would refine the current state of knowledge [29].

References

[1] Muthu Alagappan. From 5 to 13: Redefining the Positions in Basketball. MIT Sloan Sports Analytics Conference, 2012.

[2] Aravindakshan Babu. Zigzag Coarsenings, Mapper Stability and Gene-network Analyses. PhD Thesis, 2013.

[3] Vincent Barra and Silvia Biasotti. 3D Shape Retrieval and Classification using Multiple Kernel Learning on Extended Reeb graphs. *The Visual Computer*, 30(11):1247–1259, 2014.

[4] Ulrich Bauer, Xiaoyin Ge, and Yusu Wang. Measuring Distance Between Reeb Graphs. In *Proceedings of the 30th Symposium on Computational Geometry*, pages 464–473, 2014.

[5] Ulrich Bauer, Elizabeth Munch, and Yusu Wang. Strong Equivalence of the Interleaving and Functional Distortion Metrics for Reeb Graphs. In *Proceedings of the 31st Symposium on Computational Geometry*, 2015.

[6] Silvia Biasotti, Daniela Giorgi, Michela Spagnuolo, and Bianca Falcidieno. Reeb Graphs for Shape Analysis and Applications. *Theoretical Computer Science*, 392:5–22, 2008.

[7] Gunnar Carlsson, Vin de Silva, and Dmitriy Morozov. Zigzag Persistent Homology and Real-valued Functions. In *Proceedings of the 25th Symposium on Computational Geometry*, pages 247–256, 2009.

[8] Hamish Carr and David Duke. Joint Contour Nets. *IEEE Transaction on Visualization and Computer Graphics*, 20(8):1100–1113, 2014.

[9] Mathieu Carrière, Bertrand Michel, and Steve Oudot. Statistical Analysis and Parameter Selection for Mapper. *CoRR*, abs/1706.00204, 2017.

[10] Mathieu Carrière and Steve Oudot. Local Equivalence and Induced Metrics for Reeb Graphs. In *Proceedings of the 33rd Symposium on Computational Geometry*, 2017.

[11] Mathieu Carrière, Steve Oudot, and Maks Ovsjanikov. Local Signatures using Persistence Diagrams. *HAL preprint*, 2015.

[12] Amit Chattopadhyay, Hamish Carr, David Duke, Zhao Geng, and Osamu Saeki. Multivariate Topology Simplification. *Computational Geometry*, 58:1–24, 2016.

[13] Frédéric Chazal, David Cohen-Steiner, Marc Glisse, Leonidas Guibas, and Steve Oudot. Proximity of Persistence Modules and their Diagrams. In *Proceedings of the 25th Symposium on Computational Geometry*, pages 237–246, 2009.

[14] Frédéric Chazal, Vin de Silva, Marc Glisse, and Steve Oudot. *The Structure and Stability of Persistence Modules*. Springer, 2016.

[15] Frédéric Chazal, Leonidas Guibas, Steve Oudot, and Primoz Skraba. Analysis of scalar fields over point cloud data. In *Proceedings of the 20th Symposium on Discrete Algorithm*, pages 1021–1030, 2009.

[16] Frédéric Chazal and Jian Sun. Gromov-Hausdorff Approximation of Filament Structure Using Reeb-type Graph. In *Proceedings of the 30th Symposium on Computational Geometry*, 2014.

[17] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of Persistence Diagrams. *Discrete and Computational Geometry*, 37(1):103–120, 2007.

[18] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Extending persistence using Poincaré and Lefschetz duality. *Foundation of Computational Mathematics*, 9(1):79–103, 2009.

[19] Éric Colin de Verdière, Grégory Ginot, and Xavier Goaoc. Multinerves and Helly numbers of acyclic families. In *Proceedings of the 28th Symposium on Computational Geometry*, pages 209–218, 2012.

[20] Vin de Silva, Elizabeth Munch, and Amit Patel. Categorified Reeb Graphs. *Discrete and Computational Geometry*, 55:854–906, 2016.

[21] Tamal Dey, Fengtao Fan, and Yusu Wang. Graph Induced Complex on Point Data. In *Proceedings of the 29th Symposium on Computational Geometry*, pages 107–116, 2013.

[22] Tamal Dey, Facundo Mémoli, and Yusu Wang. Multiscale Mapper: Topological Summarization via Codomain Covers. In *Proceedings of the 27th Symposium on Discrete Algorithms*, pages 997–1013, 2016.

[23] Tamal Dey and Yusu Wang. Reeb Graphs: Approximation and Persistence. *Discrete and Computational Geometry*, 49(1):46–73, 2013.

[24] Herbert Edelsbrunner, John Harer, and Amit Patel. Reeb Spaces of Piecewise Linear Mappings. In *Proceedings of the 24th Symposium on Computational Geometry*, pages 242–250, 2008.

[25] Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological Persistence and Simplification. *Discrete and Computational Geometry*, 28:511–533, 2002.

[26] William Harvey, Yusu Wang, and Rephael Wenger. A randomized O(m log m) time algorithm for computing Reeb graphs of arbitrary simplicial complexes. In *Proceedings of the 26th Symposium on Computational Geometry*, pages 267–276, 2010.

[27] Vitaliy Kurlin. A one-dimensional homologically persistent skeleton of an unstructured point cloud in any metric space. In *Proceedings of the 13th Symposium on Geometry Processing*, 2015.

[28] Dmitriy Morozov. *Homological Illusions of Persistence and Stability.* Ph.D. dissertation, Department of Computer Science, Duke University, 2008.

[29] Elizabeth Munch and Bei Wang. Convergence between Categorical Representations of Reeb Space and Mapper. In *Proceedings of the 32nd Symposium on Computational Geometry*, volume 51, pages 53:1–53:16, 2016.

[30] Monica Nicolau, Arnold Levine, and Gunnar Carlsson. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proceedings of the National Academy of Science*, 108(17):7265–7270, 2011.

[31] Georges Reeb. Sur les points singuliers d'une forme de Pfaff complètement intégrable ou d'une fonction numérique. *Compte Rendu de l'Académie des Science de Paris*, 222:847–849, 1946.

[32] Gurjeet Singh, Facundo Mémoli, and Gunnar Carlsson. Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition. In *Symposium on Point Based Graphics*, 2007.

[33] Roar Bakken Stovner. On the Mapper Algorithm. Master Thesis, 2012.

[34] Wilson Sutherland. *Introduction to Metric and Topological Spaces.* Oxford University Press, 2009.

A Proof of Lemma 2.9

In this proof, we use the notations of Definition 2.1. Let $0 < \epsilon < \frac{1}{2} \min_{k=1,\dots,n} \min\{s_k - a_k, a_k - s_{k-1}\}$. The idea of the proof is to replace the right inverse of the projection $\pi : X \to \mathrm{R}_f(X)$ by a continuous map $\sigma : \mathrm{R}_f(X) \to X$ such that the composition $\pi \circ \sigma$ is homotopic to the identity of $\mathrm{R}_f(X)$. In order to make our new σ compatible with the function f, we need to perturb f to some other function g whose preimages of intervals $[s_i, s_j]$, $i \leq j$, are equal to the ones of f.

Let $g : X \to \mathbb{R}$ be defined by:

$$\forall x \in X,\ g(x) = \begin{cases} f(x) \text{ if } \min\limits_{k=1,\dots,n} |f(x) - a_k| > 2\epsilon \\ a_i \text{ otherwise, where } i = \mathrm{argmin}_k |f(x) - a_k| \end{cases}$$

As g is constant on equivalence classes of $\sim_f$, there is an induced quotient map $\tilde{g} : \mathrm{R}_f(X) \to \mathbb{R}$. Moreover, for any $i \leq j$, we have $g^{-1}([s_i, s_j]) = f^{-1}([s_i, s_j])$ by definition of g and ϵ. The same holds for $\tilde{f}$ and $\tilde{g}$.

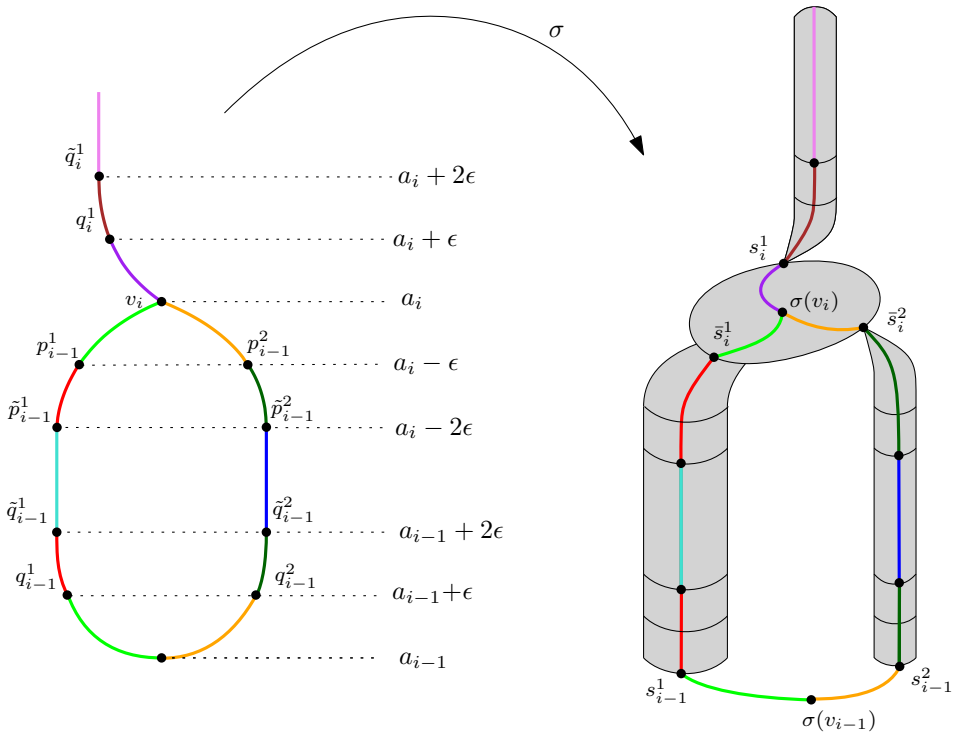


Figure 23: The left panel displays the Reeb graph and the right panel displays the space X itself. σ sends an arc of the Reeb graph to the path with the same color in X.

Now we want to define a continuous map $\sigma : \mathrm{R}_f(X) \to X$ such that the composition with the projection $\pi \circ \sigma$ is homotopic to $\mathrm{id}_{\mathrm{R}_f(X)}$. For any node v_i, if Y_{i-1} has k_i connected components $Y_{i-1}^1, ..., Y_{i-1}^{k_i}$ and Y_i has l_i connected components $Y_i^1, ..., Y_i^{l_i}$, we let $\{(\tilde{p}_{i-1}^k, p_{i-1}^k) \mid k = 1, ..., k_i\}$ and $\{(q_i^l, \tilde{q}_i^l) \mid l = 1, ..., l_i\}$ denote points in $\mathrm{R}_f(X)$ located at levelsets $a_i - 2\epsilon, a_i - \epsilon, a_i + \epsilon, a_i + 2\epsilon$. See Figure 23. For any $i = 1, ..., n$ and any $l = 1, ..., l_i$, we select an arbitrary point $y_i^l \in Y_i^l$ and we let $s_i^l = \phi_i(y_i^l, a_i)$ and $\bar{s}_{i+1}^l = \psi_i(y_i^l, a_{i+1})$.

For any critical value a_i and any vertex v_i of $\mathrm{R}_f(X)$ at that level, we let $\sigma(v_i)$ be an arbitrary point in $\pi^{-1}(v_i)$, $\sigma(q_i^l) = s_i^l$, and $\sigma(p_{i-1}^k) = \bar{s}_i^k$. Moreover, as there exists a path $\gamma_k^{i,-} : [a_i - \epsilon, a_i] \to X$ from $\bar{s}_i^k$ to $\sigma(v_i)$, σ sends the arc $[p_{i-1}^k, v_i]$ to this path $\gamma_k^{i,-}$. Similarly, it sends the arc $[v_i, q_i^l]$ to a path $\gamma_l^{i,+} : [a_i, a_i + \epsilon] \to X$ from $\sigma(v_i)$ to s_i^l. Finally, σ also monotonically reparametrizes the arcs $[\tilde{p}_i^k, p_i^k]$ and $[q_i^l, \tilde{q}_i^l]$. Let $\mathrm{param}_i^+ : [a_i + \epsilon, a_i + 2\epsilon] \to [a_i, a_i + 2\epsilon]$, and $\mathrm{param}_i^- : [a_i - 2\epsilon, a_i - \epsilon] \to [a_i - 2\epsilon, a_i]$ be these reparametrizations. Again, see Figure 23. More formally, let $x \in X$ and assume that $a_i \leq f(x) \leq a_{i+1}$ and that $\pi(x)$ belongs to the l-th edge of the Reeb graph between these two critical values. Then:

- $\sigma \circ \pi(x) = \mu_i(y_i^l, f(x))$ if $a_i + 2\epsilon \leq f(x) \leq a_{i+1} - 2\epsilon$;
- $\sigma \circ \pi(x) = \mu_i(y_i^l, \mathrm{param}_i^+ \circ f(x))$ if $a_i + \epsilon \leq f(x) \leq a_i + 2\epsilon$;
- $\sigma \circ \pi(x) = \mu_i(y_i^l, \mathrm{param}_{i+1}^- \circ f(x))$ if $a_{i+1} - 2\epsilon \leq f(x) \leq a_{i+1} - \epsilon$;
- $\sigma \circ \pi(x) = \gamma_l^{i,+}(f(x))$ if $a_i \leq f(x) \leq a_i + \epsilon$;
- $\sigma \circ \pi(x) = \gamma_l^{i+1,-}(f(x))$ if $a_{i+1} - \epsilon \leq f(x) \leq a_{i+1}$.

By construction we have $g \circ \sigma = \tilde{g}$ and $\tilde{g} \circ \pi = g$ (note that this is not true for f).

Let $i \leq j$ and $I = [s_i, s_j]$. Then we have $\pi(g^{-1}(I)) \subseteq \tilde{g}^{-1}(I)$. Hence, π induces a morphism between $H_0(g^{-1}(I))$ and $H_0(\tilde{g}^{-1}(I))$. Let us show that this morphism is an isomorphism. Since π is surjective, this boils down to showing that x, y are connected in $g^{-1}(I)$ iff $\pi(x), \pi(y)$ are connected in $\tilde{g}^{-1}(I)$.

- If x, y are connected in $g^{-1}(I)$, then so are $\pi(x), \pi(y)$ in $\tilde{g}^{-1}(I)$, by continuity of π and the fact that $\tilde{g} \circ \pi = g$.
- If $\pi(x), \pi(y)$ are connected in $\tilde{g}^{-1}(I)$, then choose a path γ connecting $\pi(x)$ and $\pi(y)$. Now by definition of σ, there exists a path γ_x connecting x and $\sigma \circ \pi(x)$ in $g^{-1}(I)$. Indeed, σ can send $\pi(x)$ to five different locations in $g^{-1}(I)$ according to the value of $f(x)$, as seen above. Assume $f(x) \notin \mathrm{Crit}(f)$. Since there is a path $\tilde{\gamma}$ between x and $\mu_i(y_i^l, f(x))$, one can always find a path γ_x between x and $\sigma \circ \pi(x)$ in $g^{-1}(I)$ with an appropriate combination of $\tilde{\gamma}$, $\mu_i(y_i^l, \cdot)$ and $\gamma_l^{(i,+)/(i+1,-)}$. Now, assume $f(x) \in \mathrm{Crit}(f)$, and let $v_i = \pi(x)$. Then $\sigma(v_i)$ and x both belong to $\pi^{-1}(v_i)$, so they belong to the same connected component of the $g^{-1}(g(x))$ and one can find a path between them in $g^{-1}(I)$. Similarly, there exists a path γ_y connecting $\sigma \circ \pi(y)$ and y in $g^{-1}(I)$. Then $\gamma_y \circ \sigma(\gamma) \circ \gamma_x$ is a path between x and y in $g^{-1}(I)$ by continuity of σ and the fact that $g \circ \sigma = \tilde{g}$. So x, y are connected in $g^{-1}(I)$.

Since $g^{-1}(I) = f^{-1}(I)$ and $\tilde{g}^{-1}(I) = \tilde{f}^{-1}(I)$, we have that π_* is an isomorphism between $H_0(f^{-1}(I))$ and $H_0(\tilde{f}^{-1}(I))$, and the proof is complete.

B Proof of Lemma 7.23

This proof uses the same construction as the proof of Proposition 3.1 in [10].

We first prove the result with the stronger assumption that critical values of R_g are all located inside the interiors of the intervals of S, that is $\mathrm{Crit}(\mathrm{R}_g) \subseteq \bigcup_{I \in S} \mathrm{int}\, I$. We define explicit continuous maps $\phi : \mathrm{R}_g \to \mathrm{R}_{g'}$ and $\psi : \mathrm{R}_{g'} \to \mathrm{R}_g$ as depicted in Figure 24. More precisely, since all critical values of R_g belong to $\bigcup_{I \in S} \mathrm{int}\, I$, we only need to specify ϕ and ψ inside each interval $I \in S$ and then ensure that the piecewise-defined maps are assembled consistently. Let $I = [a, b]$ be such an interval. Since $a, b \notin \mathrm{Crit}(g)$, there exist two levels $a < \alpha \leq \beta < b$ such that R_g is only composed of arcs in $[a, \alpha]$ and $[\beta, b]$ (dashed lines in Figure 24). For any connected component C of $g^{-1}([a, b])$, the map ϕ sends all points of $C \cap g^{-1}([\alpha, \beta])$ to the corresponding critical point $y_C \in \mathrm{R}_{g'}$ created by $\mathrm{Merge}_{a,b}$, and it extends the arcs of $C \cap g^{-1}([a, \alpha])$ (resp. $C \cap g^{-1}([\beta, b])$) into arcs of $(g')^{-1}([a, \bar{a}])$ (resp. $(g')^{-1}([\bar{a}, b])$). In return, the map ψ sends the critical point y_C to an arbitrary point of C. Then, since the Merge operation preserves connected components, for each arc A' of $(g')^{-1}([a, b])$ connected to y_C, there is at least one corresponding path A in R whose endpoint in $g^{-1}(a)$ or $g^{-1}(b)$ matches with the one of A' (see the colors in Figure 24). Hence ψ sends A' to A and the piecewise-defined maps are assembled consistently.

Let us bound the three terms in the $\max\{\cdots\}$ in (3) with this choice of maps ϕ, ψ:

- We first bound $\|g' - g \circ \psi\|_\infty$. Let $x \in \mathrm{R}_{g'}$. Either $g'(x) \notin \bigcup_{I \in S} I$, and in this case we have $g'(x) = g(\psi(x))$ by definition of ψ; or, there is $I = [a, b] \in S$ such that $g'(x) \in \mathring{I}$ and then $g(\psi(x)) \in [a, b]$. In both cases $|g'(x) - g \circ \psi(x)| < b - a$. Hence, $\|g' - g \circ \psi\|_\infty < \max_{I \in S} |I|$.
- Since the previous proof is symmetric in g and g', one also has $\|g - g' \circ \phi\|_\infty < \max_{I \in S} |I|$.
- We now bound $D(\phi, \psi)$:

  Let $(x, \phi(x)), (\psi(y), y) \in C(\phi, \psi)$ (the cases $(x, \phi(x)), (x', \phi(x'))$ and $(\psi(y), y), (\psi(y'), y')$ are similar). Let $\pi_{g'} : [0, 1] \to \mathrm{R}_{g'}$ be a continuous path from $\phi(x)$ to y which achieves $d_{g'}(\phi(x), y)$.

  - Assume $g(x) \notin \bigcup_{I \in S} I$. Then one has $\psi \circ \phi(x) = x$. Hence, $\pi_g := \psi \circ \pi_{g'}$ is a path from x to $\psi(y)$. Moreover, since $\|g' - g \circ \psi\|_\infty < b - a$, it follows that

$$\begin{aligned} &\max\ \mathrm{im}(g \circ \pi_g) < \max\ \mathrm{im}(g' \circ \pi_{g'}) + \max_{I \in S} |I|, \\ &\min\ \mathrm{im}(g \circ \pi_g) > \min\ \mathrm{im}(g' \circ \pi_{g'}) - \max_{I \in S} |I|. \end{aligned} \tag{17}$$

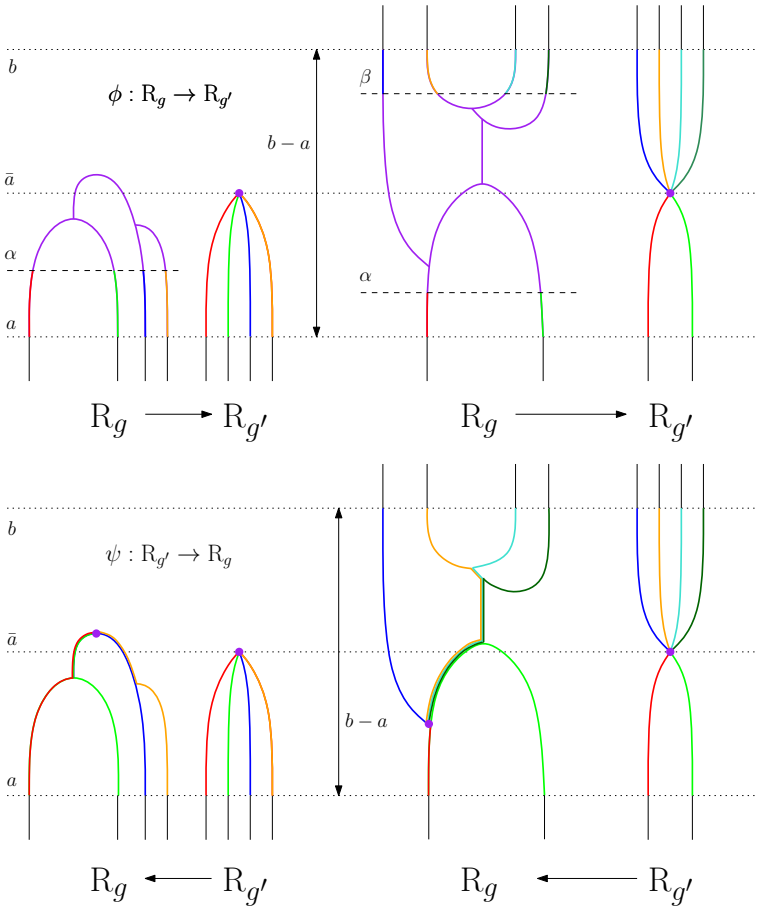


Figure 24: The effects of ϕ and ψ around a specific critical value a_i of f. Segments are matched according to their colors (up to reparameterization).

Hence, one has

$$d_g(x,\psi(y)) \leq \max\ \mathrm{im}(g\circ\pi_g) - \min\ \mathrm{im}(g\circ\pi_g) < d_{g'}(\phi(x),y) + 2\max_{I\in S}|I|,$$
$$-d_g(x,\psi(y)) \geq \min\ \mathrm{im}(g\circ\pi_g) - \max\ \mathrm{im}(g\circ\pi_g) > -d_{g'}(\phi(x),y) - 2\max_{I\in S}|I|.$$

This shows that $|d_g(x,\psi(y)) - d_{g'}(\phi(x),y)| < 2\max_{I\in S}|I|$.

– Assume that there is $I=[a,b]\in S$ such that $g(x)\in(a,b)$. Then, by definition of ϕ,ψ, we have $g'(\phi(x))\in(a,b)$, and, since ϕ and ψ preserve connected components, there is a path $\pi'_g : [0,1]\to \mathrm{R}_g$ from x to $\psi\circ\phi(x)$ within the interval $[a,b]$, which itself is included in the interior of the offset $\mathrm{im}(g'\circ\pi_{g'})^{b-a}$. Let now π_g be the concatenation of π'_g with $\psi\circ\pi_{g'}$, which goes from x to $\psi(y)$. Since $\|g'-g\circ\psi\| < \max_{I\in S}|I|$, it follows that $\mathrm{im}(g\circ\psi\circ\pi_{g'}) \subseteq \mathrm{int}\ \mathrm{im}(g'\circ\pi_{g'})^{\max_{I\in S}|I|}$, and since $\mathrm{im}(g\circ\pi_g) = \mathrm{im}(g\circ\pi'_g)\cup\mathrm{im}(g\circ\psi\circ\pi_{g'})$ by concatenation, one finally has $\mathrm{im}(g\circ\pi_g)\subseteq \mathrm{int}\ \mathrm{im}(g'\circ\pi_{g'})^{\max_{I\in S}|I|}$. Hence, the inequalities of (17) hold, implying that $|d_g(x,\psi(y)) - d_{g'}(\phi(x),y)| < 2\max_{I\in S}|I|$.

Since these inequalities hold for any couples $(x,\phi(x))$ and $(\psi(y),y)$, we deduce that $D(\phi,\psi)\leq 2\max_{I\in S}|I|$.

In the general case where an endpoint of an interval of S can be a critical value of g, one needs to slightly change the definition of ϕ. More precisely, α and β have to be taken outside but arbitrarily close to $[a,b]$. Then, the previous proof extends verbatim.