Structure and Stability of the 1-Dimensional Mapper*

Mathieu Carrière[1] and Steve Oudot[2]

1 **DataShape, Inria Saclay, Palaiseau, France**
mathieu.carriere@inria.fr
2 **DataShape, Inria Saclay, Palaiseau, France**
steve.oudot@inria.fr

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 relating the structure of the Mapper to that 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.

1998 ACM Subject Classification F.2.2 [Nonnumerical Algorithms and Problems] Geometrical problems and computations

Keywords and phrases Mapper, Reeb Graph, Extended Persistence, Topological Data Analysis

Digital Object Identifier 10.4230/LIPIcs.SoCG.2016.25

1 Introduction

The *Mapper*[1] was introduced in [22] as a new mathematical object to summarize the topological structure of a continuous map $f : X \to \mathbb{R}^d$. 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^{-1} gives an open cover $\mathcal{U}$ of X. Splitting each element of $\mathcal{U}$ into its various connected components yields a connected cover $\mathcal{V}$, whose nerve is the Mapper (which thus has one k-simplex per non-empty $(k+1)$-fold intersection of elements of $\mathcal{V}$). The Mapper can be thought of as a *pixelized version* of the Reeb space, where the resolution is prescribed by the cover $\mathcal{I}$. In practice, its construction from point cloud data is easy to describe and implement, requiring only to build a neighborhood graph whose size is at worst quadratic in the size of the point cloud.

Since its introduction, the Mapper has aroused the interest of practitioners in the data sciences, with several success stories [1, 21], due to its ability to deal with very general functions and datasets. Nevertheless, 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, the Mapper is a pixelized version of the Reeb graph, so it should

* Note: The proofs are omitted in this extended abstract and can be found in the full version [9].
[1] In this article we call *Mapper* the mathematical object, not the algorithm used to build it.

32nd International Symposium on Computational Geometry (SoCG 2016).
Editors: Sándor Fekete and Anna Lubiw; Article No. 25; pp. 25:1–25:16
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

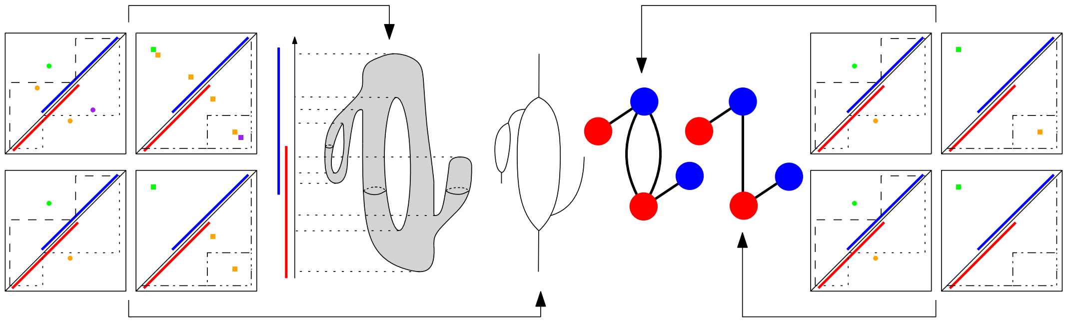


Figure 1 From left to right: a 2-manifold equipped with the height function; the corresponding Reeb graph, MultiNerve Mapper, and Mapper. For each object, we display the persistence diagrams of dimension 0 (green), 1 (orange) and 2 (purple). Extended points are squares while ordinary and relative points 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.

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 and quantify the stability of the structure of the Mapper when f is scalar? These are the questions addressed here.

Contributions. Assuming f is scalar, 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. The connection happens through an intermediate object, called the *MultiNerve Mapper*, which we define as the *multinerve* [16] of the connected pullback cover.

Given a pair (X, f) with $f : X \to \mathbb{R}$ continuous, and an interval cover $\mathcal{I}$ of $\text{im}(f)$, we show that the MultiNerve Mapper itself is a Reeb graph, for a perturbed pair (X', f') (Theorem 5.3). 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'). More precisely, we can match the quotient maps' persistence diagrams $\text{Dg}(\tilde{f})$ and $\text{Dg}(\tilde{f}')$ with each other (Theorem 5.2), and thus draw a correspondence between the features of the MultiNerve Mapper and the ones of the Reeb graph of (X, f). This correspondence is oblivious to the actual layouts of the 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 a persistence diagram. The points in this diagram are in one-to-one correspondence with the features (branches, holes) in the (MultiNerve) Mapper. Thus, like $\text{Dg}(\tilde{f})$, our diagram for the (MultiNerve) Mapper serves as a bag-of-features type descriptor.

An interesting property of our descriptor is to be predictable[2] given the persistence diagram of the quotient 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 descriptor 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 5.4). See Figure 1.

[2] As a byproduct, we also clarify the relationship between the diagram of $\tilde{f}$ and the one of f (Theorem 2.5).

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 6.1) follow easily from the general stability theorem for extended persistence [15]. Similar guarantees hold when the domain X or the cover $\mathcal{I}$ is perturbed. These stability guarantees can be exploited in practice to approximate the descriptors of the Mapper and MultiNerve Mapper from point cloud data efficiently. The details are given in Section 7 of the full version of the paper [9].

Our main proof technique consists in perturbing the so-called *telescope* [7] corresponding to the pair (X, f). We introduce a set of elementary perturbations and study their effects on the persistence diagram. By performing these perturbations in sequence, we can track the points in the diagram while the pair (X, f) is being modified. We believe these elementary perturbations are of an independent interest (see Section 4).

Related work. Reeb graphs are now well understood and have been used in a wide range of applications. We refer the interested reader to [4, 5, 6] for a comprehensive list of references. In a recent study, even more structure has been given to the Reeb graphs by categorifying them [17].

Several variants of these graphs have 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 [22] is one of them. Chazal et al. [14] introduced the λ-*Reeb graph*, which is another type of Reeb graph pixelization with intervals. The authors can derive upper bounds on the Gromov-Hausdorff distance between the space and its Reeb or λ-Reeb graph. This is too much asking in general; as a result, the hypothesis are very strong.

Joint Contour Nets [8, 11] 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. Munch and Wang [20] recently showed that, as the lengths of the intervals in the cover $\mathcal{I}$ go to zero uniformly, the Joint Contour Net and the Mapper itself converge to the continuous Reeb space in the so-called *interleaving distance* [17]. Their result holds in the general case of vector-valued functions. Here we restrict the focus to real-valued functions but are able to make non-asymptotic claims (Corollary 5.4).

On another front, Stovner [23] proposed a categorified version of the Mapper, seen as a covariant functor from the covered topological spaces to the simplicial complexes. Dey et al. [18] pointed out the inherent instability of the Mapper and proposed a multiscale variant that 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. Babu [2] characterized the Mapper with zigzag persistent homology. Here, we do not coarsen a zigzag module but rather identify specific areas of an extended persistence diagram corresponding to features that disappear in the Mapper. By doing so, we answer two open questions from [18], introducing a signature that describes the set of features of the Mapper completely, 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. In the following, “connected” stands for “pathconnected", and "cc" stands for "connected component(s)". 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

▶ **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 $\mathrm{Crit}(f) = \{a_1 < ... < a_n\} \subset \mathbb{R}$, called the set of *critical values*, s.t. 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 $\mu_i : Y_i \times (a_i, a_{i+1}) \to X^{(a_i,a_{i+1})}$ s.t. $\forall i = 0, ..., n, f|_{X^{(a_i,a_{i+1})}} = \pi_2 \circ \mu_i^{-1}$, 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.

All Morse functions on a smooth manifold are of Morse type. However, the converse is 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.

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 $\mathbb{R}^{\mathrm{op}} = \{\tilde{x} \mid x \in \mathbb{R}\}$, 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. First, 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:

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

where we have identified the space X with the pair of spaces $(X, \emptyset)$. The subfamily $\{F_\alpha\}_{\alpha\in\mathbb{R}}$ is called the *ordinary* part of the filtration, while $\{F_{\tilde{\alpha}}\}_{\tilde{\alpha}\in\mathbb{R}^{\mathrm{op}}}$ is called the *relative* part.

Applying the homology functor H_* to this filtration gives the so-called *extended persistence module* $\mathbb{V}$ of f, which is a sequence of vector spaces connected by linear maps 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 half-open *interval modules*—see e.g. [12]: $\mathbb{V} \simeq \bigoplus_{k=1}^n \mathbb{I}[b_k, d_k)$, where each summand $\mathbb{I}[b_k, d_k)$ is made of copies of the field of coefficients at every 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* (cc, 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* (PD) of f, denoted $\mathrm{Dg}(f)$. The distinction between ordinary and relative parts of the filtration allows us to classify the points in $\mathrm{Dg}(f)$ as follows:

- $p = (x, y)$ is called an *ordinary* point if $x, y \in \mathbb{R}$;
- $p = (x, y)$ is called a *relative* point if $x, y \in \mathbb{R}^{\mathrm{op}}$;
- $p = (x, y)$ is called an *extended* point if $x \in \mathbb{R}, y \in \mathbb{R}^{\mathrm{op}}$;

Note that ordinary points lie strictly above the diagonal $\Delta = \{(x, x) \mid x \in \mathbb{R}\}$ and relative points lie strictly below Δ, while extended points can be located anywhere, including on Δ (e.g. when a cc lies inside a single critical level, see Section 2.3). It is common to partition $\mathrm{Dg}(f)$ 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.

Stability. An important property of extended PDs is to be stable in the so-called *bottleneck distance* d_{b}^{∞}. Given two PDs D, D', a *partial matching* between D and D' is a subset Γ of $D \times D'$ where for every $p \in D$ there is at most one $p' \in D'$ such that $(p, p') \in \Gamma$, and conversely, for every $p' \in D'$ there is at most one $p \in D$ such that $(p, p') \in \Gamma$. Furthermore, Γ must match points of the same type (ordinary, relative, extended) and of the same homological dimension only. The *cost* of Γ is: $\mathrm{cost}(\Gamma) = \max\{\max_{p \in D} \delta_D(p), \max_{p' \in D'} \delta_{D'}(p')\}$, where $\delta_D(p) = \|p - p'\|_\infty$ if p is matched to some $p' \in D'$ and $\delta_D(p) = d_\infty(p, \Delta)$ if p is unmatched – same for $\delta_{D'}(p')$.

▶ **Definition 2.2.** Let D, D' be two PDs. The *bottleneck distance* between D and D' is $d_{\mathrm{b}}^{\infty}(D, D') = \inf_\Gamma \ \mathrm{cost}(\Gamma)$, where Γ ranges over all partial matchings between D and D'.

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

▶ **Theorem 2.3** (Stability [15])**.** *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 [15], the theorem can be strengthened to apply to each subdiagram $\mathrm{Ord}, \mathrm{Ext}^+, \mathrm{Ext}^-, \mathrm{Rel}$ and to each homological dimension individually.

2.3 Reeb Graphs

▶ **Definition 2.4.** 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 $x \sim_f y$ if and only if $f(x) = f(y)$ and x, y belong to the same cc of $f^{-1}(f(x)) = f^{-1}(f(y))$. The *Reeb graph* $\mathrm{R}_f(X)$ is the quotient space $X/\sim_f$.

As f is constant on equivalence classes, there is an induced quotient map $\tilde{f} : \mathrm{R}_f(X) \to \mathbb{R}$ with $f = \tilde{f} \circ \pi$, where π is the projection $X \to \mathrm{R}_f(X)$ induced by $\sim_f$. If f is a function of Morse type, then the pair (X, f) is an $\mathbb{R}$-*constructible space* in the sense of [17]. This ensures that the Reeb graph is a multigraph, whose nodes are in one-to-one correspondence with the cc of the critical level sets of f. We can equip this multigraph with a metric by assigning the length $l(v_i, v_j) = |f(v_i) - f(v_j)|$ to each edge (v_i, v_j). In the following, the combinatorial version of the Reeb graph is denoted by $\mathcal{C}(\mathrm{R}_f(X))$.

Connection to the extended persistence. 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, 15] 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 cc 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 downward branches are given by the points in $\mathrm{Ord}_0(\tilde{f})$;
- The vertical spans of the upward branches 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 PD of the quotient map $\tilde{f}$. Note that it is a bag-of-features type of descriptor, taking an inventory of all the features together with their vertical spans, but leaving aside the actual layout of the features. As a consequence, it is an incomplete descriptor: two Reeb graphs with the same PD may not be isomorphic.

The following theorem summarizes the known connections between $\mathrm{Dg}(\tilde{f})$ and $\mathrm{Dg}(f)$. It formalizes the intuition that the Reeb graph captures part of the topological structure of f, the missing features being either "inessential", or "horizontal", or "higer-dimensional". We provide a proof in the full version [9] for completeness.

▶ **Theorem 2.5.** *Let X be a topological space and $f : X \to \mathbb{R}$ a function of Morse type. Then, $\mathrm{Dg}(\tilde{f}) \subseteq \mathrm{Dg}(f)$. More precisely:*

$$\mathrm{Dg}_0(\tilde{f}) = \mathrm{Dg}_0(f); \quad \mathrm{Dg}_1(\tilde{f}) = \mathrm{Dg}_1(f) \setminus (\mathrm{Ext}_1^+(f) \cup \mathrm{Ord}_1(f)); \quad \mathrm{Dg}_p(\tilde{f}) = \emptyset \ \forall p \geq 2.$$

2.4 Covers and Nerves

Let Z be a topological space. A *cover* of Z is a family $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$ of subsets of Z, 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 a k-simplex per $(k+1)$-fold intersection of elements of $\mathcal{U}$:

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

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 $\tilde{U}_\alpha = U_\alpha \setminus \bigcup_{\alpha' \neq \alpha \in A} U_{\alpha'}$. The cc of $\tilde{U}_\alpha$ are called the *proper subsets* of U_α. $\mathcal{U}$ is called *generic* if no proper subset 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 2.6.** *If $\mathcal{U}$ is a minimal open interval cover of $Z \subseteq \mathbb{R}$, then no more than two elements of $\mathcal{U}$ can intersect at a time. Moreover, if Z is $\mathbb{R}$ itself or a compact subset thereof, then any cover $\mathcal{U}$ of Z has a minimal subcover.*

From now on, unless otherwise stated, all covers of $Z \subseteq \mathbb{R}$ will be generic, open, minimal, interval covers (*gomic* for short). An immediate consequence of Lemma 2.6 is that every element U of a gomic $\mathcal{U}$ has exactly one proper subset $\tilde{U}$. More precisely, U partitions into three subintervals: $U = U_\cap^- \sqcup \tilde{U} \sqcup U_\cap^+$, 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).

2.5 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 cc: $V_\alpha = \bigsqcup_{i\in\{1\dots c(\alpha)\}} V_\alpha^i$, where $c(\alpha)$ is the number of cc of V_α. Then, $\mathcal{V} = \{V_\alpha^i\}_{\alpha\in A, i\in\{1\dots c(\alpha)\}}$ is a connected cover of X. It is called the *connected pullback cover*, and its nerve $\mathcal{N}(\mathcal{V})$ is the Mapper.

▶ **Definition 2.7.** Let X, Z be topological spaces, $f : X \to Z$ a continuous function, $\mathcal{U}$ a cover of $\mathrm{im}(f)$ and $\mathcal{V}$ the associated cover of X. The *Mapper* of $(f, \mathcal{U})$ is $\mathrm{M}_f(X, \mathcal{U}) = \mathcal{N}(\mathcal{V})$.

See Figure 1 for an illustration. Note that, when $Z = \mathbb{R}$ and $\mathcal{U}$ is a gomic of $\mathrm{im}(f)$, the Mapper has a natural 1-dimensional stratification since no more than two intervals can intersect at a time by Lemma 2.6. Hence, in this case, it has the structure of a (possibly infinite) simple graph and therefore has trivial homology in dimension 2 and above. When $\mathcal{U}$ is not a gomic, the Mapper may not be a graph nor have trivial homology in dimension 2.

3 MultiNerve Mapper

Given a cover $\mathcal{U} = \{U_\alpha\}_{\alpha\in A}$ of a topological space X, it is possible to extend the concept of nerve to a *simplicial poset* called the *multinerve*:

▶ **Definition 3.1** ([16])**.** The *multinerve* $\mathcal{M}(\mathcal{U})$ is the simplicial poset defined by:

$$\mathcal{M}(\mathcal{U}) = \{(\{\alpha_0, \cdots, \alpha_k\}, C) \ \mid \ \bigcap_{i=0,\cdots,k} U_{\alpha_i} \neq \emptyset \text{ and } C \text{ is a cc of } \bigcap_{i=0,\cdots,k} U_{\alpha_i}\}.$$

The proof that this set, together with the least element $(\emptyset, \bigcup_{\mathcal{U}})$ and equipped with the partial order $(F, C) \preceq (F', C') \iff F \subseteq F'$ and $C' \subseteq C$, is a simplicial poset, can be found in [16]. We extend the concept of Mapper by using the multinerve of the connected pullback cover instead of its nerve:

▶ **Definition 3.2.** Let X, Z be topological spaces, $f : X \to Z$ a continuous function, $\mathcal{U}$ a cover of $\mathrm{im}(f)$ and $\mathcal{V}$ the associated cover of X. The *MultiNerve Mapper* of X is $\overline{\mathrm{M}}_f(X, \mathcal{U}) = \mathcal{M}(\mathcal{V})$.

See Figure 1 for an illustration. Again, when $Z = \mathbb{R}$ and $\mathcal{U}$ is a gomic of $\mathrm{im}(f)$, the MultiNerve Mapper is a (possibly infinite) multigraph and therefore has trivial homology in dimension 2 and above. Contrary to the Mapper, it also takes the cc of the intersections into account. As we shall see in Section 5, the MultiNerve Mapper is able to capture the same features as the Mapper, even with coarser gomics, and is more naturally related to the Reeb graph.

The connection between the Mapper and the MultiNerve Mapper is induced by the connection between nerves and multinerves [16]: $\mathrm{M}_f(X, \mathcal{U}) = \pi_1(\overline{\mathrm{M}}_f(X, \mathcal{U}))$, where $\pi_1 : (F, C) \mapsto F$ is the projection of the simplices $(\{\alpha_0, ..., \alpha_k\}, C)$ of the multinerve $\overline{\mathrm{M}}_f(X, \mathcal{U})$ onto their first coordinate. 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, π_1 induces a surjective homomorphism in homology.

4 Telescope

In this section we introduce the telescopes, which are our main objects of study when we relate the structure of the MultiNerve Mapper to the one of the Reeb graph of a perturbation of the pair (X, f).

▶ **Definition 4.1** (Telescope [7]). A *telescope* is an adjunction space of the following form:

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

where $a_0 = -\infty$ and $a_{n+1} = +\infty$ by convention 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.

A telescope comes equipped with π_1 and π_2, which are the projections onto the first factor and second factor respectively. Given any interval I, we let $T^I = \pi_1 \circ \pi_2^{-1}(I)$.

A function of Morse type $f : X \to \mathbb{R}$ naturally induces a telescope T_X defined with $\mathrm{Crit}(T) = \mathrm{Crit}(f)$, $X_i = f^{-1}(\{a_i\})$, $Y_i = \pi_1 \circ \mu_i^{-1} \circ f^{-1}(a_i, a_{i+1})$, $\phi_i = (\bar{\mu}_i|_{Y_i \times \{a_i\}}, \mathrm{id})$ and $\psi_i = (\bar{\mu}_i|_{Y_i \times \{a_{i+1}\}}, \mathrm{id})$. One can define a homeomorphism $\mu : X \to T_X$ such that $f = \pi_2 \circ \mu$, so that $\mathrm{Dg}(\pi_2) = \mathrm{Dg}(f)$. We refer the reader e.g. to [17] for more details.

Operations

We will now present three kinds of perturbations on telescopes that preserve the MultiNerve Mapper, namely: *Merge*, *Split*, and *Shift*. For this we will use generalized attaching maps:

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

▶ **Definition 4.2** (Merge). Let T be a telescope. Let $a \le b$. If $[a, b]$ contains at least one critical value, i.e. $a_{i-1} < a \le a_i \le a_j \le b < a_{j+1}$, then the Merge on T between a, b is the telescope $T' = \mathrm{Merge}_{a,b}(T)$ given by:

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

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. $a_{i-1} < a \le b < a_i$, then $\mathrm{Merge}_{a,b}(T)$ is given by:

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

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

See Figure 2 for an illustration.

Similarly, we define the Merge between a, b on a diagram D as the diagram $\mathrm{Merge}_{a,b}(D)$ given by:

$$\mathrm{Merge}_{a,b}(x, y) = (\bar{x}, \bar{y}) \text{ where } \bar{x} = \begin{cases} x \text{ if } x \notin [a, b] \\ \bar{a} \text{ otherwise} \end{cases} \text{ and similarly for } y.$$

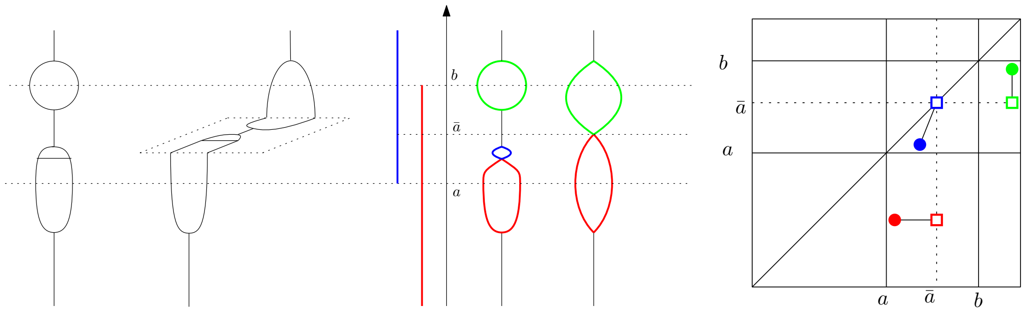


Figure 2 Left: Effect of a Merge. Middle: Effect on the corresponding Reeb graph. Right: Effect on the corresponding extended PD of dimension 1.

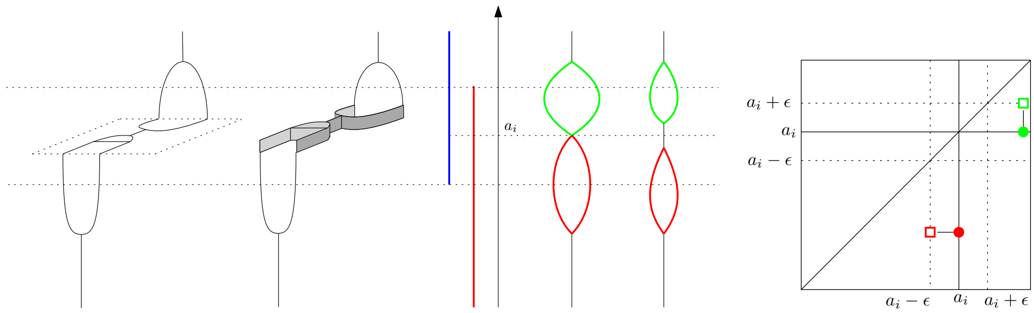


Figure 3 Left: Effect of a Split. Middle: Effect on the corresponding Reeb graph. Right: Effect on the corresponding extended PD of dimension 1.

▶ **Lemma 4.3.** *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,* $\mathrm{Dg}(\pi_2') = \mathrm{Merge}_{a,b}(\mathrm{Dg}(\pi_2))$.

▶ **Definition 4.4** (Split)**.** Let T be a telescope. Let $a_i \in \mathrm{Crit}(T)$ and ϵ s.t. $0 \leq \epsilon < \min\{a_{i+1} - a_i, a_i - a_{i-1}\}$. The ϵ-Split on T at a_i is the telescope $T' = \mathrm{Split}_{\epsilon,a_i}(T)$ given by:

$$...(Y_{i-1} \times [a_{i-1}, a_i]) \cup_{\psi_{i-1}} (X_i \times \{a_i\}) \cup_{\phi_i} (Y_i \times [a_i, a_{i+1}])...$$
$$\updownarrow$$
$$...(Y_{i-1} \times [a_{i-1}, a_i - \epsilon]) \cup_{\psi_{i-1}^{a_i-\epsilon}} (X_i \times \{a_i - \epsilon\}) \cup_{\mathrm{id}} (X_i \times [a_i - \epsilon, a_i + \epsilon]) \cup_{\mathrm{id}} (X_i \times \{a_i + \epsilon\}) \cup_{\phi_i^{a_i+\epsilon}} (Y_i \times [a_i + \epsilon, a_{i+1}])...$$

See Figure 3 for an illustration.

Similarly, we define the ϵ-Split at a_i on a diagram D as the diagram $\mathrm{Split}_{\epsilon,a_i}(D)$ given by:

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

Note that the definition of $\mathrm{Split}_{\epsilon,a_i}(D)$ assumes implicitly that D 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 4.4.

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 an homeomorphism. The new attaching maps introduced by the Split are identity maps, hence the following lemma:

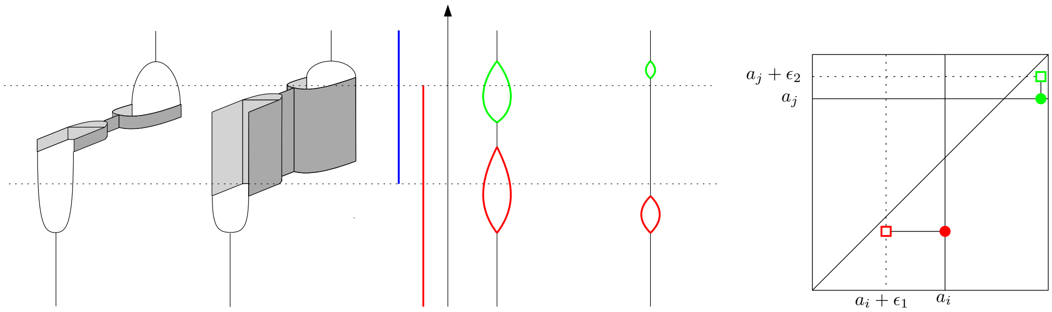


Figure 4 Left: Effect of a double Shift with amplitudes $\epsilon_1 < 0 < \epsilon_2$. Middle: Effect on the corresponding Reeb graph. Right: Effect on the corresponding extended PD of dimension 1.

▶ **Lemma 4.5.** *The new critical values $a_i - \epsilon$ and $a_i + \epsilon$ created after a Split are down- and up-forks respectively.*

▶ **Lemma 4.6.** *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, $\mathrm{Dg}(\pi_2') = \mathrm{Split}_{\epsilon,a_i}(\mathrm{Dg}(\pi_2))$.*

▶ **Definition 4.7** (Shift)**.** Let T be a telescope. Let $a_i \in \mathrm{Crit}(T)$ and ϵ s.t. $0 \leq |\epsilon| < \min\{a_{i+1} - a_i, a_i - a_{i-1}\}$. The ϵ-Shift on T at a_i is the telescope $T' = \mathrm{Shift}_{\epsilon,a_i}(T)$ given by:

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

See Figure 4 for an illustration. Similarly, we define the ϵ-Shift at a_i on a diagram D as the diagram $\mathrm{Shift}_{\epsilon,a_i}(D)$ given by:

$$\mathrm{Shift}_{\epsilon,a_i}(x,y) = (\bar{x},\bar{y}) \text{ where } \bar{x} = \begin{cases} x \text{ if } x \neq a_i \\ a_i + \epsilon \text{ otherwise} \end{cases} \text{ and similarly for } y$$

Note that the definition of $\mathrm{Shift}_{\epsilon,a_i}(D)$ assumes implicitly that D 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 4.7.

▶ **Lemma 4.8.** *Let $a_i \in \mathrm{Crit}(T)$, ϵ s.t. $0 < |\epsilon| < \min\{a_{i+1} - a_i, a_i - a_{i-1}\}$, $T' = \mathrm{Shift}_{\epsilon,a_i}(T)$ and $\pi_2' : T' \to \mathbb{R}$ the projection onto the second factor. Then, $\mathrm{Dg}(\pi_2') = \mathrm{Shift}_{\epsilon,a_i}(\mathrm{Dg}(\pi_2))$.*

Invariance

The above operations leave the (MultiNerve) Mapper unchanged under certain conditions:

▶ **Proposition 4.9.** *Let T be a telescope and $\mathcal{I}$ be a gomic of $\mathrm{im}(\pi_2)$.*

(i) *Let $a \leq b$ s.t. a, b belong to the same intersection $I \cap J$ or proper interval $\tilde{I}$. 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 $\epsilon < \min\{a_i - a, b - a_i\}$, then $\overline{\mathrm{M}}_{\pi_2}(\mathrm{Split}_{\epsilon,a_i}(T), \mathcal{I})$ 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 $c - a_i < \epsilon < \min\{d, a_{i+1}\} - a_i$, then $\overline{\mathrm{M}}_{\pi_2}(\mathrm{Shift}_{\epsilon,a_i}(T), \mathcal{I})$ 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* $\max\{a, a_{i-1}\} - a_i < \epsilon < b - a_i$*, then* $\overline{\mathrm{M}}_{\pi_2}(\mathrm{Shift}_{\epsilon,a_i}(T), \mathcal{I})$ *is isomorphic to* $\overline{\mathrm{M}}_{\pi_2}(T, \mathcal{I})$*.*

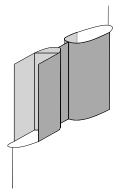
5 Structure of the MultiNerve Mapper

Let $f : X \to \mathbb{R}$ be of Morse type, and let $\mathcal{I}$ be a gomic of $\mathrm{im}(f)$. Let T_X be the corresponding telescope. In this section, we move out all critical values 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 $[a_{I\cap J}, b_{I\cap J}]$. We define

$$T'_X := \mathrm{Merge}'_{\mathcal{I}} \circ \mathrm{Shift}_{\mathcal{I}} \circ \mathrm{Split}_{\mathcal{I}} \circ \mathrm{Merge}_{\mathcal{I}}(T_X), \tag{1}$$

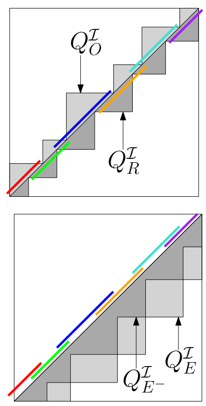
where each operation is defined individually as follows:

- $\mathrm{Merge}_{\mathcal{I}}$ 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 $\mathrm{Split}_{\epsilon,\bar{a}}$ with $\bar{a}$ a critical value after the first $\mathrm{Merge}_{\mathcal{I}}$ (therefore not an interval endpoint) and $\epsilon > 0$ satisfying the assumptions of Prop. 4.9 (ii).
- $\mathrm{Shift}_{\mathcal{I}}$ is the composition of all the $\mathrm{Shift}_{\epsilon,\bar{a}_+}$ with $\bar{a}_+$ an up-fork critical value after the $\mathrm{Split}_{\mathcal{I}}$ and $\epsilon > 0$ such that the assumptions of Prop. 4.9 (iii) are satisfied, and of all the $\mathrm{Shift}_{\epsilon,\bar{a}_-}$ with $\bar{a}_-$ a down-fork critical value after the $\mathrm{Split}_{\mathcal{I}}$ and $\epsilon < 0$ such that the assumptions of Prop. 4.9 (iv) are satisfied. After $\mathrm{Shift}_{\mathcal{I}}$ there are no more critical values located in the intersections of consecutive intervals of $\mathcal{I}$.
- $\mathrm{Merge}'_{\mathcal{I}}$ is the composition of all the $\mathrm{Merge}_{a_{\tilde{I}}, b_{\tilde{I}}}$, $I \in \mathcal{I}$.



Combine Figures 2, 3 and 4 with the figure opposite for an illustration of this sequence of transformations. We let $\pi'_2 : T'_X \to \mathbb{R}$ be the projection onto the second factor. In the following, we identify the pair (T_X, π_2) with (X, f). since they are isomorphic in the category of $\mathbb{R}$-constructible spaces. We also rename π'_2 into f' for convenience. Let $\tilde{f}' : \mathrm{R}_{f'}(T'_X) \to \mathbb{R}$ be the induced quotient map.

▶ **Lemma 5.1.** *For* T'_X *defined as in (1),* $\overline{\mathrm{M}}_{f'}(T'_X, \mathcal{I})$ *is isomorphic to* $\overline{\mathrm{M}}_f(X, \mathcal{I})$ *as a combinatorial multigraph.*



The effect of (1) on the extended PD of f is illustrated in Figure 5. 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 the critical values $\bar{a}$ introduced by the $\mathrm{Merge}_{\mathcal{I}}$. In the following, we use the term *cell* to designate a 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 5 correspond to squares associated to a proper subinterval (blue) or intersection (green).

Our first structure theorem (Theorem 5.2 below) involves certain unions of colored cells from Figure 5, called the *staircases* and defined as follows. Given an interval I (indifferently open, closed or half-open),

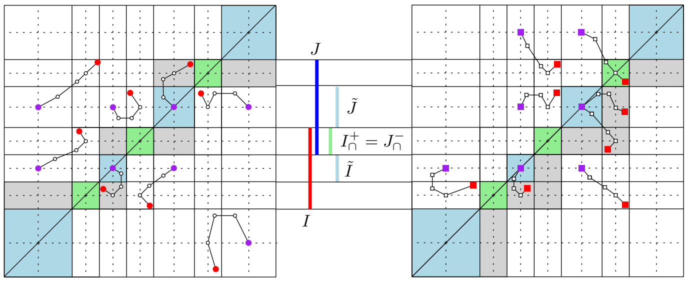


Figure 5 The left panel displays the trajectories of points in Ord (disks above the diagonal) and Rel (disks under the diagonal) while the right panel displays the trajectories of points in Ext. For both diagrams, the original point is red, the final point is purple and intersection and proper intervals are colored in green and light blue respectively.

let $Q_I^+ = \{(x,y) \in \mathbb{R}^2 \mid x \leq y \in I\}$ be the half-square above the diagonal, and $Q_I^- = \{(x,y) \in \mathbb{R}^2 \mid y < x \in I\}$ the half-square strictly below the diagonal. Decompose now each interval $I \in \mathcal{I}$ as $I = I_\cap^- \sqcup \tilde{I} \sqcup I_\cap^+ \in \mathcal{I}$, then let $Q_O^{\mathcal{I}} = \bigcup_{I\in\mathcal{I}} Q^+_{\tilde{I}\cup I_\cap^+}$, $Q_R^{\mathcal{I}} = \bigcup_{I\in\mathcal{I}} Q^-_{\tilde{I}\cup I_\cap^-}$ and $Q_{E^-}^{\mathcal{I}} = \bigcup_{I\in\mathcal{I}} Q_I^-$ be the staircases obtained by taking the unions of these half-squares, as illustrated in the figures on the right. Our structure theorem is stated as follows (recall that we have identified the pairs (X,f) and (T_X,π_2) and renamed π_2' into f'):

▶ **Theorem 5.2.** *For T_X' defined as in (1), for every $p \geq 0$ there is a perfect matching between:*
(i) $\mathrm{Ord}_p(f')$ *and* $\mathrm{Ord}_p(f) \setminus Q_O^{\mathcal{I}}$*,*
(ii) $\mathrm{Rel}_p(f')$ *and* $\mathrm{Rel}_p(f) \setminus Q_R^{\mathcal{I}}$*,*
(iii) $\mathrm{Ext}_p^-(f')$ *and* $\mathrm{Ext}_p^-(f) \setminus Q_{E^-}^{\mathcal{I}}$*,*
(iv) $\mathrm{Ext}_p^+(f')$ *and* $\mathrm{Ext}_p^+(f) \cup (\mathrm{Ext}_p^-(f) \cap Q_{E^-}^{\mathcal{I}})$*.*

The proof follows the tracking strategy illustrated in Figure 5. For each point of $\mathrm{Dg}(f)$, the results of Section 4 allow us to recreate its trajectory through the various operations of (1). Then, a series of simple observations (detailed in [9]) leads to the conclusion.

Theorems 2.5 and 5.2 together induce the following perfect matching between the persistence diagrams of the quotient maps[3]:

$$\mathrm{Ord}(\tilde{f}') \text{ and } \mathrm{Ord}(\tilde{f}) \setminus Q_O^{\mathcal{I}}; \quad \mathrm{Ext}(\tilde{f}') \text{ and } \mathrm{Ext}(\tilde{f}) \setminus Q_{E^-}^{\mathcal{I}}; \quad \mathrm{Rel}(\tilde{f}') \text{ and } \mathrm{Rel}(\tilde{f}) \setminus Q_R^{\mathcal{I}}. \quad (2)$$

Now we relate $\tilde{f}'$ to the MultiNerve Mapper through the Reeb graph of f', using the property that f' has exactly one critical value inside each proper interval and none outside:

▶ **Theorem 5.3.** *For T_X' defined as in (1), $\overline{\mathrm{M}}_f(X,\mathcal{I})$ is isomorphic to $\mathcal{C}(\mathrm{R}_{f'}(T_X'))$ as a combinatorial multigraph.*

A signature for MultiNerve Mapper. Theorem 5.3 means that the dictionary introduced in Section 2.3 can be used to describe the structure of the MultiNerve Mapper from the extended persistence diagram of the perturbed quotient function $\tilde{f}'$. $\mathrm{Dg}(\tilde{f}')$ is obtained from $\mathrm{Dg}(\tilde{f})$ as in (2). This suggests using the off-staircase part of $\mathrm{Dg}(\tilde{f})$ as a descriptor for the

[3] Note that $\mathrm{Ext}_0^-(g) = \emptyset$ for any Morse-type function g, including $g = \tilde{f}$, $g = f$, $g = f'$, and $g = \tilde{f}'$.

structure of the MultiNerve Mapper:

$$\begin{aligned}\mathrm{Dg}(\overline{\mathrm{M}}_f(X,\mathcal{I})) &= \mathrm{Ord}(\tilde{f}) \setminus Q_O^{\mathcal{I}} \cup \mathrm{Ext}(\tilde{f}) \setminus Q_{E^-}^{\mathcal{I}} \cup \mathrm{Rel}(\tilde{f}) \setminus Q_R^{\mathcal{I}} \\ &= \mathrm{Ord}_0(f) \setminus Q_O^{\mathcal{I}} \cup (\mathrm{Ext}_0^+(f) \cup \mathrm{Ext}_1^-(f)) \setminus Q_{E^-}^{\mathcal{I}} \cup \mathrm{Rel}_1(f) \setminus Q_R^{\mathcal{I}}, \end{aligned} \tag{3}$$

where the second equality comes from Theorem 2.5. We call this descriptor the *persistence diagram* of the MultiNerve Mapper. Note that this descriptor is not computed by applying persistence to some function defined on the multinerve, but it is rather a pruned version of the persistence diagram of $\tilde{f}$. As for Reeb graphs, it serves as a bag-of-features descriptor of the structure of $\overline{\mathrm{M}}_f(X,\mathcal{I})$. The fact that $\mathrm{Dg}(\overline{\mathrm{M}}_f(X,\mathcal{I})) \subseteq \mathrm{Dg}(\tilde{f})$ formalizes the intuition that the MultiNerve Mapper is a *pixelized version* of the Reeb graph, in which some of the features disappear due to the staircases (prescribed by the cover). The following convergence result is a direct consequence of our theorems:

▶ **Corollary 5.4.** *Suppose the granularity of the gomic $\mathcal{I}$ is at most ε. Then,*

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

Thus, the features (branches, holes) of the Reeb graph that are missing in the MultiNerve Mapper have spans at most ε. Moreover, the two signatures become equal when ε is smaller than the smallest ℓ^∞-distance of the points of $\mathrm{Dg}(\tilde{f})$ to the diagonal. For even smaller ε, $\overline{\mathrm{M}}_f(X,\mathcal{I})$ and $\mathcal{C}(\mathrm{R}_f(X))$ become isomorphic as combinatorial multigraphs, up to vertex splits and edge subdivisions – see [9]. Note that convergence occurs before ε goes to zero.

Induced signature for Mapper. Recall from Section 3 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: $Q_E^{\mathcal{I}} = \bigcup_{I\cup J \text{ s.t. } I,J\in\mathcal{I} \text{ and } I\cap J\neq\emptyset} Q_{I\cup J}^-$. The other staircases remain unchanged. Hence the following descriptor for the Mapper:

$$\begin{aligned}\mathrm{Dg}(\mathrm{M}_f(X,\mathcal{I})) &= \mathrm{Ord}(\tilde{f}) \setminus Q_O^{\mathcal{I}} \cup \mathrm{Ext}(\tilde{f}) \setminus Q_E^{\mathcal{I}} \cup \mathrm{Rel}(\tilde{f}) \setminus Q_R^{\mathcal{I}} \\ &= \mathrm{Ord}_0(f) \setminus Q_O^{\mathcal{I}} \cup (\mathrm{Ext}_0^+(f) \cup \mathrm{Ext}_1^-(f)) \setminus Q_E^{\mathcal{I}} \cup \mathrm{Rel}_1(f) \setminus Q_R^{\mathcal{I}}. \end{aligned} \tag{4}$$

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. Moreover, the convergence result stated in Corollary 5.4 holds the same with the Mapper. See Figure 1 for an example that summarizes the different spaces and signatures.

6 Stability

Intuitively, for a point in the descriptor $\mathrm{Dg}(\overline{\mathrm{M}}_f(X,\mathcal{I}))$, the ℓ^∞-distance to its corresponding staircase ($Q_O^{\mathcal{I}}$, $Q_{E^-}^{\mathcal{I}}$ or $Q_R^{\mathcal{I}}$, depending on the type of the point) 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 descriptor $\mathrm{Dg}(\tilde{f})$ that is not in the MultiNerve Mapper's descriptor (i.e. that lies inside the 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 aim here is to translate this intuition into stability results. 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 as well for the Mapper (with $Q_{E^-}^{\mathcal{I}}$ replaced by $Q_E^{\mathcal{I}}$).

Stability w.r.t. perturbations of the function. Let Θ be a subset of $\mathbb{R}^2$. Given a partial matching Γ between two persistence diagrams D, D', the Θ-*cost* of Γ is: $\mathrm{cost}_\Theta(\Gamma) = \max\{\max_{p\in D}\ \delta_D(p),\ \max_{p'\in D'}\ \delta_{D'}(p')\}$, where $\delta_D(p) = \|p-p'\|_\infty$ if $\exists p' \in D'$ s.t. $(p,p') \in \Gamma$ and $\delta_D(p) = d_\infty(p,\Theta)$ otherwise — same for $\delta_{D'}(p')$. The bottleneck distance becomes: $d^\infty_{\mathrm{b},\Theta}(D,D') = \inf_\Gamma\ \mathrm{cost}_\Theta(\Gamma)$, where Γ ranges over all partial matchings between D and D'. This is again a pseudometric and not a metric. To avoid heavy notations, we write d_Θ instead of $d^\infty_{\mathrm{b},\Theta}$. Note that the usual bottleneck distance is obtained by taking Θ to be the diagonal $\Delta = \{(x,x) \mid x \in \mathbb{R}\}$. 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 descriptors as follows, where the notation D_* stands for the subdiagram of D of the right type (Ordinary, Extended or Relative):

$$d_\mathcal{I}(D,D') = \max\left\{d_{Q^\mathcal{I}_O}(D_O,D'_O),\ d_{Q^\mathcal{I}_{E^-}}(D_E,D'_E),\ d_{Q^\mathcal{I}_R}(D_R,D'_R)\right\}. \tag{5}$$

▶ **Theorem 6.1.** *Given a gomic $\mathcal{I}$, for any Morse-type functions $f,g : X \to \mathbb{R}$,*

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

This result follows easily from the definition of the signature in (3) and from the stability of extended persistence diagrams (Theorem 2.3). It can be extended likewise to perturbations of the domain using the framework of [10] and the stability result therein, see [9]. Note that the classical bottleneck distance d_Δ is unstable in this context.

Stability w.r.t. perturbations of the cover. Let us now fix the pair (X,f) and consider varying gomics. We aim for a quantification of the extent to which the structure of the (MultiNerve) Mapper 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.

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. Given a persistence diagram D and two gomics $\mathcal{I}, \mathcal{J}$, we consider the quantity:

$$d_D(\mathcal{I},\mathcal{J}) = \max_{*\in\{O,E^-,R\}}\ \sup_{p\in D_*\cap(Q^\mathcal{I}_*\triangle Q^\mathcal{J}_*)}\ \max\left\{d_\infty(p,Q^\mathcal{I}_*),\ d_\infty(p,Q^\mathcal{J}_*)\right\}, \tag{6}$$

where $\triangle$ denotes the symmetric difference, where D_* stands for the subdiagram of D of the right type (Ordinary, Extended or Relative), and where we adopt the convention that $\sup_{p\in\emptyset}$ ... is zero instead of infinite. Deriving an upper bound on $d_D(\mathcal{I},\mathcal{J})$ in terms of the Hausdorff distance between the staircases is straightforward, since the supremum in (6) is taken over points that lie in the symmetric difference between the staircases:

▶ **Theorem 6.2.** *Given a Morse-type function $f : X \to \mathbb{R}$, for any gomics $\mathcal{I}, \mathcal{J}$,*

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

where $\tilde f$ is the quotient map defined on the Reeb graph $\mathrm{R}_f(X)$.

Moreover, we have

$$\mathrm{Dg}_*(\tilde f)\cap(Q^\mathcal{I}_*\triangle Q^\mathcal{J}_*) = (\mathrm{Dg}_*(\tilde f)\cap Q^\mathcal{I}_*)\triangle(\mathrm{Dg}_*(\tilde f)\cap Q^\mathcal{J}_*) = \mathrm{Dg}_*(\overline{\mathrm{M}}_f(X,\mathcal{I}))\triangle\mathrm{Dg}_*(\overline{\mathrm{M}}_f(X,\mathcal{J})),$$

where the second equality follows from the definition of the descriptor of the MultiNerve Mapper given in (3). Thus, $d_{\mathrm{Dg}(\tilde f)}(\mathcal{I},\mathcal{J})$ quantifies the proximity of each descriptor 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.

7 Discrete Case

Building the signatures of $\overline{\mathrm{M}}_f(X,\mathcal{I})$ and $\mathrm{M}_f(X,\mathcal{I})$ requires to compute the critical values of f exactly, which may not always be possible. However, the stability properties presented in Section 6 can be exploited in practice to approximate the signatures from point cloud data. The approach boils down to applying known scalar field analysis techniques [13], then pruning the obtained persistence diagrams using the staircases. If one wants to further guarantee that the approximate signatures do correspond to some perturbed Mapper or MultiNerve Mapper, then one can use known techniques for Reeb graph approximation instead [19]. The details are given in the full version of the paper [9], together with a glimpse at other questions related to the discrete approximation of the Mapper (such as the statistical computation of confidence intervals and convergence rates to the Reeb graph) which we are currently working on.

Acknowledgements. We thank the anonymous referees for their insightful comments. This work was supported by ERC grant Gudhi and by ANR project TopData.

References

1. M. Alagappan. From 5 to 13: Redefining the Positions in Basketball. MIT Sloan Sports Analytics Conference, 2012.
2. A. Babu. Zigzag Coarsenings, Mapper Stability and Gene-network Analyses. 2013.
3. V. Barra and S. Biasotti. 3D Shape Retrieval and Classification using Multiple Kernel Learning on Extended Reeb Graphs. *The Visual Computer*, 30(11):1247–1259, 2014.
4. U. Bauer, X. Ge, and Y. Wang. Measuring Distance Between Reeb Graphs. In *Proc. 30th Sympos. Comput. Geom.*, pages 464–473, 2014.
5. U. Bauer, E. Munch, and Y. Wang. Strong Equivalence of the Interleaving and Functional Distortion Metrics for Reeb Graphs. In *Proc. 31st Sympos. Comput. Geom.*, 2015.
6. S. Biasotti, D. Giorgi, M. Spagnuolo, and B. Falcidieno. Reeb Graphs for Shape Analysis and Applications. *Theor. Comput. Sci.*, 392(1-3):5–22, 2008.
7. G. Carlsson, V. de Silva, and D. Morozov. Zigzag Persistent Homology and Real-valued Functions. In *Proc. 25th Sympos. Comput. Geom.*, pages 247–256, 2009.
8. H. Carr and D. Duke. Joint Contour Nets. *IEEE Trans. Vis. Comput. Graph.*, 20(8):1100–1113, 2014.
9. M. Carrière and S. Oudot. Structure and Stability of the 1-Dimensional Mapper. arXiv 1511.05823, 2015.
10. M. Carrière, S. Oudot, and M. Ovsjanikov. Stable Topological Signatures for Points on 3D Shapes. In *Proc. 13th Sympos. Geom. Proc.*, 2015.
11. A. Chattopadhyay, H. Carr, D. Duke, Z. Geng, and O. Saeki. Multivariate Topology Simplification. arXiv 1509.04465, 2015.
12. F. Chazal, V. de Silva, M. Glisse, and S. Oudot. The Structure and Stability of Persistence Modules. arXiv 1207.3674, 2012.
13. F. Chazal, L. Guibas, S. Oudot, and P. Skraba. Analysis of scalar fields over point cloud data. In *Proc. 20th Sympos. Discr. Algo.*, pages 1021–1030, 2009.

14 F. Chazal and J. Sun. Gromov-Hausdorff Approximation of Filament Structure Using Reeb-type Graph. In *Proc. 30th Sympos. Comput. Geom.*, 2014.
15 D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Extending persistence using Poincaré and Lefschetz duality. *Found. Comput. Math.*, 9(1):79–103, 2009.
16 É. Colin de Verdière, G. Ginot, and X. Goaoc. Multinerves and helly numbers of acyclic families. In *Proc. 28th Sympos. Comput. Geom.*, pages 209–218, 2012.
17 V. de Silva, E. Munch, and A. Patel. Categorified Reeb Graphs. arXiv 1501.04147, 2015.
18 T. Dey, F. Mémoli, and Y. Wang. Mutiscale Mapper: A Framework for Topological Summarization of Data and Maps. arXiv 1504.03763, 2015.
19 T. Dey and Y. Wang. Reeb graphs: Approximation and persistence. *Discr. Comput. Geom.*, 49(1):46–73, 2013.
20 E. Munch and B. Wang. Convergence between Categorical Representations of Reeb Space and Mapper. In *Proc. 32nd Sympos. Comput. Geom.*, 2016.
21 M. Nicolau, A. Levine, and G. Carlsson. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proc. National Acad. Sci.*, 108(17):7265–7270, 2011.
22 G. Singh, F. Mémoli, and G. Carlsson. Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition. In *Sympos. PB Graphics*, 2007.
23 R. B. Stovner. On the Mapper Algorithm. 2012.