

# Region-Aware Wasserstein Distances of Persistence Diagrams and Merge Trees

Mathieu Pont and Christoph Garth

**Abstract**—This paper presents a generalization of the Wasserstein distance for both persistence diagrams and merge trees [20], [66] that takes advantage of the regions of their topological features in the input domain. Specifically, we redefine the comparison of topological features as a distance between the values of their extrema-aligned regions. It results in a more discriminative metric than the classical Wasserstein distance and generalizes it through an input parameter adjusting the impact of the region properties in the distance. We present two strategies to control both computation time and memory storage of our method by respectively enabling the use of subsets of the regions in the computation, and by compressing the regions’ properties to obtain low-memory representations. Extensive experiments on openly available ensemble data demonstrate the efficiency of our method, with running times on the orders of minutes on average. We show the utility of our contributions with two applications. First, we use the assignments between topological features provided by our method to track their evolution in time-varying ensembles and propose the *temporal persistence curves* to facilitate the understanding of how these features appear, disappear and change over time. Second, our method allows to compute a distance matrix of an ensemble that can be used for dimensionality reduction purposes and visually represent in 2D all its members, we show that such distance matrices also allow to detect key phases in the ensemble. Finally, we provide a C++ implementation that can be used to reproduce our results.

**Index Terms**—Topological data analysis, ensemble data, merge trees, persistence diagrams.

## I. INTRODUCTION

As acquisition devices become more precise and computational power continues to grow, modern datasets, acquired or simulated, are becoming increasingly detailed and complex. Such rising complexity poses a challenge for interactive exploration, making data analysis and interpretation more difficult for end users. This has motivated the development of intermediary representations which summarize the data in a concise and informative manner. In that direction, Topological Data Analysis (TDA) methods [20], [97] have been proven useful, such as in the visualization community [39], by providing representations that extract the features of interest within the data while having memory storage requirements orders of magnitude lower than the data themselves. Such methods have been successfully applied to several domains [39], for example in turbulent combustion [11], [32], [47], material sciences [24], [34], [36], nuclear energy [53], fluid dynamics [44], [57], bioimaging [2], [9], [14], quantum chemistry [7], [28], [58] or astrophysics [71], [75]. Popular instances in the visualization community of representations studied in TDA include the

persistence diagram (Fig. 2, top insets) [4], [20], [21], [31] and the merge tree (Fig. 2, middle insets) [9], [11], [12], [14].

In modern applications, alongside the growing complexity of datasets, a significant challenge arises with the notion of *ensemble datasets* [23], [40], [85]. Instead of representing a phenomenon with a single dataset, these consist of multiple ones, each of them referred as an *ensemble member*. A common approach in TDA is to compute a topological abstraction, such as a persistence diagram or a merge tree, for each ensemble member in order to encode its features of interest in a compact manner. Then, a strategy consists in analyzing the given ensemble of datasets through the resulting ensemble of topological representations. While this has several benefits, like the direct processing of features of interests and a reduced memory footprint, it necessitates the development of analysis methods for these specific objects.

Several works explored this direction, notably with the computation of an average topological representation [66], [82], [83], [91] that is representative of an ensemble and can be used for clustering tasks, or with variability analysis methods for ensembles of topological representations [65], [67]. The basic and main building block of all these methods is the notion of distances between topological representations. While practically useful and well studied metrics have been developed for such objects, persistence diagrams and merge trees themselves can suffer from a lack of specificity, where geometrically different datasets have the same topological representation, preventing to discriminate them (Fig. 3). By encoding features too compactly, from a topological point of view, they introduce a limitation that causes important geometric details to be lost. This lack of geometric specificity impacts the subsequent data analysis pipeline after their computation and can prevent the identification of important feature differences in an ensemble. This can lead to wrong scientific conclusions where datasets seem to be identical from a topological point of view but are not from a geometrical one. To tackle this limitation, topological representations must be extended with geometric properties and for their analysis, a metric being computable in practice needs to be defined. Earlier works enrich topological descriptors with only coarse and aggregated geometric properties, particularly with size and position [48], [69], [74], [93], therefore overlooking the fine-grained and vertex-level inner structure of each feature.

This paper addresses this limitation by first associating to each topological feature encoded in persistence diagrams or merge trees a detailed characterization of the geometry of its inner structure. Specifically, we consider for each topological feature a vertex-level characterization of its region in the input

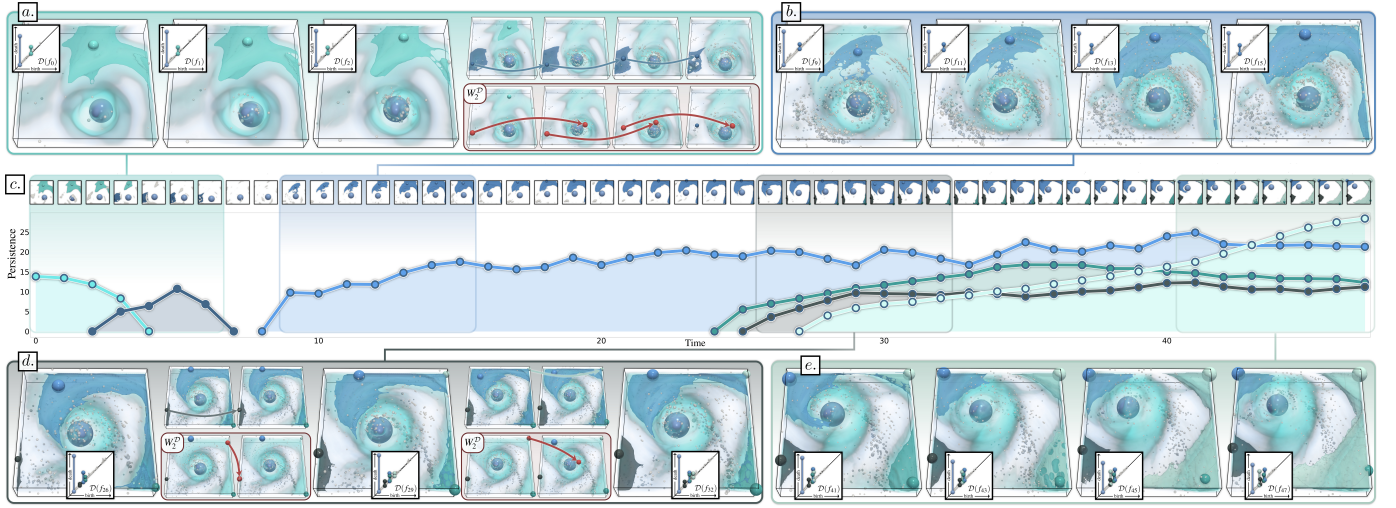


Fig. 1. Visual analysis of the 48 time steps of the *Isabel* ensemble (wind velocity) with our novel region-aware Wasserstein distance. Members at different points in time are shown in (a), (b), (d) and (e). The persistence diagram of each member (black insets) allows to identify its features of interest, their maxima are visualized in the input domain with a sphere whose size is proportionnal to their persistence (a notion of importance) and the regions of the highest persistent features are shown with a surface. Our method provides assignments between the features of consecutive members, allowing to track their evolution over time (with matched features having the same color). In this example, it correctly matches the same features together over the whole ensemble while the classical Wasserstein distance  $W_2^D$  provide some mismatches (red arrows in red insets of (a) and (d)) preventing their correct tracking. The *temporal persistence curves* in (c) represent the evolution over time of the persistence of the matched features with our method, only curves for the highest persistent features are shown with the same color than in the input domain. This allows to visually convey the changes in the ensemble in one view and permits the user to identify and select important time steps to be visualized. For instance, in the second time window (b) the blue feature starts to appear and grow, while in the last time window (e) the white feature overtake the other ones.

domain alongwith all its scalar values. We then introduce a distance metric for both descriptors that takes advantage of this geometric information. Our approach is a natural generalization of the Wasserstein distance between persistence diagrams and merge trees [20], [66], that we extend to incorporate such geometric properties. We show that the proposed metric is more discriminative than the original Wasserstein distance that eventually corresponds to a special case of our method. We also propose and evaluate strategies, subject to input parameters, to control both the computation time and memory footprint of our method (Sec. IV), by respectively allowing to use more or less the geometric information in the computation and by compressing the geometric characterizations to reduce memory storage. We show the utility of our contributions with two visualization tasks. First, we use the assignments between the features provided by our distance to track their evolution in time-varying datasets (Sec. V-B) and propose a visualization tool to easily see how they evolve over time, allowing to detect phases and gain insights about the ensemble. Second, the distance matrix of an ensemble can be used for dimensionality reduction (Sec. V-C) in order to visually represent in 2D all the members of the ensemble and how they relate to each other.

#### A. Related work

The literature related to our work can be classified into two groups: (i) topological representations and (ii) their metrics.

(i) **Topological representations:** Methods from computational topology [20] have been explored and extended by the visualization community for a diverse set of applications [94]. The persistence diagram [4], [20], [21], [31] (Sec. II-B and Fig. 2 top insets) and the merge tree [12], [78] (Sec. II-C and Fig. 2

middle insets) intuitively segment the domain into regions (Fig. 6), each of them being a feature of interest. It can be computed through matrix reduction [4], [20] or using discrete morse theory [31]. The merge tree additionally encodes the connectivity of the features in a hierarchical manner and efficient algorithms for its computation (and its generalization, the contour tree) have been documented [1], [12], [13], [29], [51], [52], [78]. Other examples of topological representations include the Reeb graph [8], [30], [60], [62], [68] or the Morse-Smale complex [10], [19], [33], [35], [37], [70].

(ii) **Metrics for topological representations:** Using concepts from optimal transport [43], [54], the Wasserstein distance between persistence diagrams [20] (Sec. II-D) is based on a bipartite assignment problem for which exact computations [56] or fast approximations [6], [45] are available in open-source [81]. It was proven to be stable under perturbations of the input data [15]. However, persistence diagrams may lack in specificity and can represent significantly different datasets in the same way, failing to discriminate those that differ in feature connectivity (a simple example can be found in [66] Figure 3). This has motivated the development of metrics for more advanced descriptors, including merge trees.

A stable distance between merge trees has been proposed [55] having however an exponential time complexity, therefore not tractable for real-life datasets. A variant has been developed using pre-existing correspondence labels between merge trees nodes [27], since manual labeling is impractical for real-life datasets, some heuristics need to be considered [95]. Apart from that, Beketayev et al. [5] focused on a dual representation, the *branch decomposition tree* (BDT, see Sec. II-C), by minimizing a term over all possible branch decompositions and Wetzels et al. [89] provided a similar but



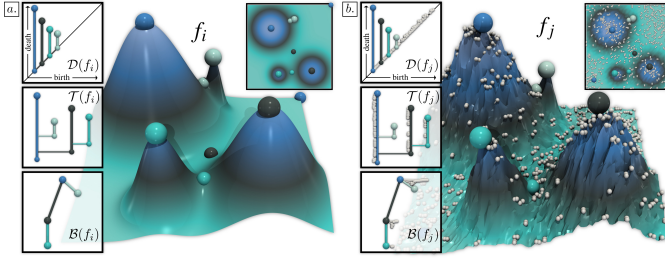


Fig. 2. Illustration of the topological descriptors studied in this work, computed on a 2D minimal example (a) and a noisy version of it (b). For all descriptors, the color code represents the persistence of the corresponding saddle-maximum pair. Critical points are visualized with spheres, the larger ones indicating maxima. Persistence diagrams, merge trees, and branch decomposition trees (BDTs) are shown in the top, center, and bottom insets, respectively. In both datasets, the four primary features (the most prominent hills) are highlighted with salient pairs in the diagram and the merge tree. To maintain clarity in the visualization, pairs with low persistence (below 10% of the function range) are displayed with small white arcs and spheres, while the larger and colored ones correspond to those with higher persistence.

faster approach based on edit distances. They also proposed a constrained edit distance based on deformations of the trees [87] that provide meaningful matchings. Both methods have however a quartic time complexity allowing to only process merge trees with a very small number of nodes. An unconstrained version of the latter was shown to be stable but NP-complete with exponential time complexity [86] for which some heuristics have been explored to choose a trade-off between the quartic and exponential time complexity [90]. Sridharamurthy et al. [76] adapted an efficient algorithm for the computation of constrained edit distances between trees [96] to merge trees by defining edit costs related to these objects. It results in an efficient distance with quadratic time complexity that can be easily computed for real-life datasets and whose stability can be controlled thanks to an input parameter. Based on this, Pont et al. [66] generalized the Wasserstein distance between extremum persistence diagrams to merge trees, allowing to choose a trade-off between both representations given an input parameter. One benefit of this framework, in addition to its efficiency, is that an analysis method developed using this distance can trivially process both persistence diagrams and merge trees without additional effort.

Some works have taken an initial step toward incorporating geometric information into distance computation by focusing on coarse and aggregated properties. It usually involves the addition of a term to the distance between two topological features, originally defined as the distance between their minimum and maximum values respectively (their *birth* and their *death*, see Sec. II-B). For instance, the distance between the coordinates in the domain of the extrema of two topological features can be used [48], [74], [93]. Note that while Yan et al. [93] or Mingzhe et al. [48] propose a framework to incorporate geometric information into distance computation, they mainly consider coordinates in the domain. The size, i.e. the number of vertices, of the segmented region corresponding to a feature can also be considered [69]. Inspired by [80], the geodesic distance between the extrema of a feature in the hyperplane plot of the scalar function can provide some geometric insights

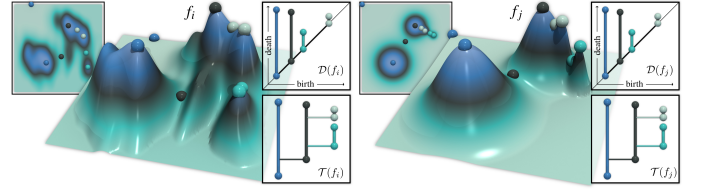


Fig. 3. The persistence diagram,  $\mathcal{D}(f_i)$ , and the merge tree,  $\mathcal{T}(f_i)$ , provide a visual summary of the number, data range, and salience of the features of interest present in the data. However, they represent each feature without considering its geometric properties. This results in a lack of specificity for these objects, which can yield identical data representations for significantly distinct datasets. In this example, the four features of each dataset are identically encoded in the corresponding topological representation whereas their respective geometry differs. These datasets can therefore not be discriminated using the celebrated Wasserstein distance for topological representations.

about topological features. *Decorated* approaches (for merge trees [17] and Reeb spaces [18]) attach auxiliary signals, for example with local persistence diagrams. However these methods have several limitations, first they only add coarse information about the geometric properties of the features, they are still missing detailed characterizations of the features geometry related to their shape etc. Second, they extend the comparison of the scalar values of topological features with a term of a different nature and scale (distances in the domain, differences in size etc.), challenging the correct weighting and mixture of both. In contrast, our approach uses a more detailed characterization of the features. It takes advantage of all their scalar values in order to precisely characterize their geometry and extends the usual distance between topological features to not only compare their minimum and maximum values, but up to all their values given an input parameter controlling the amount of information to use.

## B. Contributions

This paper makes the following new contributions:

- 1) *A region-aware metric for persistence diagrams and merge trees:* We generalize recent works on Wasserstein distances for persistence diagrams and merge trees [20], [66] to introduce a new metric that incorporates and uses richer and more detailed geometric characterizations of the regions of the topological features. It makes them special cases of our method, by allowing to control how much of this detailed characterization should be used depending on the applications. It can be computed efficiently and we prove theoretical properties such as it is metric, stable under some conditions and more discriminative than the original metrics it is based on.
- 2) *Time and memory optimization strategies:* We present and evaluate strategies to control both the computation time and the memory footprint of our method. First, we propose a way to simplify the comparisons of the features geometry inspired by subsampling. Second, we show how traditional compression methods can be used to compress the detailed characterization of the features to reduce memory storage and how it impacts our metric.
- 3) *An application to feature tracking:* The matchings between topological features provided by our novel metric

can be used to track their evolution in time-varying ensembles (Sec. V-B). We also introduce the *Temporal Persistence Curves* that visually convey how topological features evolve over time allowing to easily detect when they appear, disappear or change.

- 4) *An application to dimensionality reduction*: Our method provides distance matrices that can be used by traditional dimensionality reduction algorithms to embed each topological representation as a 2D point and visually see how they relate to each other (Sec. V-C). We also show that such distance matrices allow to detect when important changes occur in an ensemble.
- 5) *Implementation*: We provide a C++ implementation of our algorithms for reproduction purposes.

## II. PRELIMINARIES

This section introduces the various concepts upon which our work is based. We refer to the reference books [20], [97] for a detailed introduction to computational topology.

### A. Input data

We consider an ensemble of  $N$  piecewise linear (PL) scalar fields  $f_i : \mathcal{M} \rightarrow \mathbb{R}$ , with  $i \in \{1, \dots, N\}$ , defined on a PL  $d$ -manifold  $\mathcal{M}$  being a simplicial complex with  $d \leq 3$  in our applications (like triangulated 2D images or 3D volumes). Each  $f_i$  is defined on its vertices and linearly interpolated on the simplices of higher dimension. A topological representation can be computed for each ensemble member  $f_i$  to summarize its features of interest in a compact manner. We focus in this work on the extremum *Persistence Diagram* (PD), noted  $\mathcal{D}(f_i)$ , as well as a dual representation of the *Merge Tree* (MT), called the *Branch Decomposition Tree* (BDT), noted  $\mathcal{B}(f_i)$ .

### B. Persistence diagrams

In the case of extremum persistence diagrams, each topological feature of a PL scalar field  $f : \mathcal{M} \rightarrow \mathbb{R}$  can be intuitively seen as a subset of  $\mathcal{M}$  containing one minimum, respectively a maximum, and values that increase, respectively decrease, in a monotonic manner from either side of it (Fig. 6). Each feature can be represented as a 2D point  $(f(c), f(c'))$  described by its minimum value  $f(c)$  and its maximum one  $f(c')$ . All topological features of  $f$  can then be summarized as a set of 2D points called the extremum persistence diagram  $\mathcal{D}(f)$ .

The topological features involving minima of  $f$  can be uniquely identified using its sub-level set  $f_{-\infty}^{-1}(w) = \{p \in \mathcal{M} \mid f(p) \leq w\}$ , defined as the inverse of  $(-\infty, w]$  by  $f$ , i.e. all the points of  $\mathcal{M}$  having a value  $f$  below or equal to  $w$ . When  $w$  is increased from the lowest values to the highest ones, topological changes, such as the appearance or the merge of connected components, occur when  $f_{-\infty}^{-1}(w)$  reaches specific vertices of  $\mathcal{M}$  called the critical points of  $f$  [3]. For instance, for a PL  $d$ -manifold  $\mathcal{M}$  with  $d > 1$ , connected components appear at minima and merge at saddles. For such merges, the Elder rule [20] states that the *younger* component (created last) *dies* in favor of the *older* one. The dying component can then be associated with a unique pair

of critical points, the minimum  $c$  at which it appeared and the saddle  $c'$  at which it merges with another component. Their respective value  $f(c)$  and  $f(c')$ , with  $f(c) < f(c')$ , are respectively called its *birth* and its *death*. The 2D point  $(f(c), f(c'))$  summarizes this topological feature and is called a minimum-saddle persistence pair. Its *persistence* is given by  $f(c') - f(c)$  and is often seen as a measure of importance of the given feature. Note that the first connected component that appeared, at the global minimum, never dies since it never merges with another one, in practice its death value is cropped to the global maximum. A similar idea of this whole process can be developed for saddle-maximum pairs using the super-level set  $f_{+\infty}^{-1}(w) = \{p \in \mathcal{M} \mid f(p) \geq w\}$ .

We consider  $\mathcal{M}$  as a simplicial complex (for example the triangulation of a regular grid [38], [42]) and  $f$  is considered to be injective on its vertices (easily achieved with symbolic perturbations [22]), this sub-level set process is done through the so-called *lower star filtration* [20], [31] which intuitively processes each vertex one at a time in function order (i.e. in the increasing order of its function value).

A persistence pair is often visualized with a segment connecting the values of its two critical points in the y-axis, whose length corresponds to its persistence (Fig. 2 top insets). The persistence diagram of a PL scalar field then provides a visual summary of its features of interest, where salient features stand out from the diagonal (Fig. 2b top inset, colored pairs) while pairs likely corresponding to noise are located near the diagonal (Fig. 2b top inset, white pairs). The diagonal in this 2D birth/death space consists of features with zero persistence, i.e. non-existent structures of the data. A persistence diagram is defined as the set of persistence pairs alongside an infinite number of points in the diagonal with infinite multiplicity.

### C. Merge trees

In addition to encoding the presence and range of the topological features of  $f$ , as persistence diagrams do, merge trees also capture their connectivity. The minimum-saddle pairs of  $f$  are captured by its *join* tree while the saddle-maximum pairs by its *split* tree (Fig. 2, center insets). Each of these two trees is called a *merge tree*, noted  $\mathcal{T}(f)$ . The nodes of such a tree are critical points of  $f$ , the root of a join tree is the global maximum and its leaves are all the local minima, while the split tree has the global minimum as a root and all local maxima as leaves. For a PL  $d$ -manifold  $\mathcal{M}$  with  $d > 1$ , the inner nodes of a merge tree are the saddle critical points of the persistence pairs it represents. The edges of the join tree inform about which connected components of the sub-level set  $f_{-\infty}^{-1}(w)$  merged with each other as  $w$  increases (similarly for the split tree with the super-level set  $f_{+\infty}^{-1}(w)$  as  $w$  decreases).

The join tree  $\mathcal{T}^-(f)$  is defined in a continuous manner as the quotient space  $\mathcal{T}^- = \mathcal{M} / \sim$  by the equivalence relation  $\sim$  stating that two points  $p_1$  and  $p_2$  in  $\mathcal{M}$  are equivalent if  $f(p_1) = f(p_2)$  and if  $p_1$  and  $p_2$  belong to the same connected component of the sub-level set  $f_{-\infty}^{-1}(f(p_1))$ . The split tree  $\mathcal{T}^+(f)$  is defined in a similar manner using the super-level set  $f_{+\infty}^{-1}$ . Their discretization corresponds to the objects that were described in the previous paragraph.

Merge trees are often visualized using a persistence-driven *branch decomposition* [61] representing the persistence pairs captured by these trees. A *persistent branch* is a monotone path on the tree connecting the two nodes of a same persistence pair. Then, each persistent branch is represented as a vertical segment and their ancestor relations with a horizontal one (Fig. 2, center insets). The persistence-driven *branch decomposition tree* (BDT), noted  $\mathcal{B}$ , is a dual representation of the merge tree where each persistent branch is collapsed to a node and its edges encode their hierarchy (Fig. 2, bottom insets). Specifically, if one of the two nodes of a persistent branch  $b_i$  is itself on the path of another persistent branch  $b_j$  (the other node of  $b_i$  being a leaf), then  $b_i$  will be a child of  $b_j$  in  $\mathcal{B}$ . Each node of a BDT corresponds therefore to a persistence pair and is associated with its birth and death value.

To tackle instabilities called *saddle swaps*, merge trees can be pre-processed [76] to merge adjacent interior nodes (being saddle points if  $d > 1$ ) if the difference in their function value is below a threshold determined by the input parameter  $\varepsilon_1 \in [0, 1]$ . This has no effect when  $\varepsilon_1 = 0$ , and as  $\varepsilon_1$  increases, more nodes are merged. In the BDT, this parameter has the effect to bring up pairs in the tree hierarchy.

#### D. Wasserstein metric space

We now define the Wasserstein distance between persistence diagrams [20] and its generalization to merge trees [66].

The Wasserstein distance between two persistence diagrams seeks an assignment between their points. A point of a diagram can either be assigned to one point, and exactly one, in the other diagram, representing two features that are matched together, or to its projection on the diagonal of the other diagram, intuitively representing the deletion or insertion of a feature. In practice, when computing the distance between two diagrams  $\mathcal{D}(f_i)$  and  $\mathcal{D}(f_j)$ , they are first pre-processed to augment each diagram with the diagonal projection of the off-diagonal points of the other diagram:

$$\begin{aligned}\mathcal{D}'(f_i) &= \mathcal{D}(f_i) \cup \{\Delta(p_j) \mid p_j \in \mathcal{D}(f_j)\} \\ \mathcal{D}'(f_j) &= \mathcal{D}(f_j) \cup \{\Delta(p_i) \mid p_i \in \mathcal{D}(f_i)\},\end{aligned}$$

where  $\Delta(p_i) = (\frac{x_i+y_i}{2}, \frac{x_i+y_i}{2})$  is the diagonal projection of an off-diagonal point  $p_i = (x_i, y_i)$ . This procedure simulates the fact that persistence diagrams have an infinite number of diagonal points (dummy features with zero persistence) and allow off-diagonal points to possibly be matched to their diagonal projection (i.e. their closest point among all points on the diagonal). It also guarantees that the two diagrams now have the same number of points ( $|\mathcal{D}'(f_i)| = |\mathcal{D}'(f_j)|$ ).

Two points  $p_i = (x_i, y_i) \in \mathcal{D}'(f_i)$  and  $p_j = (x_j, y_j) \in \mathcal{D}'(f_j)$ , can be assigned to each other with a cost corresponding to their distance in the birth/death space, called the *ground metric*:

$$d_q(p_i, p_j) = (|x_i - x_j|^q + |y_i - y_j|^q)^{1/q} = \|p_i - p_j\|_q, \quad (1)$$

with  $q \geq 1$ . In the case where both  $p_i$  and  $p_j$  are located on the diagonal (i.e.  $x_i = y_i$  and  $x_j = y_j$ ),  $d_q(p_i, p_j)$  is set to zero (such that these dummy features do not contribute in the distance).

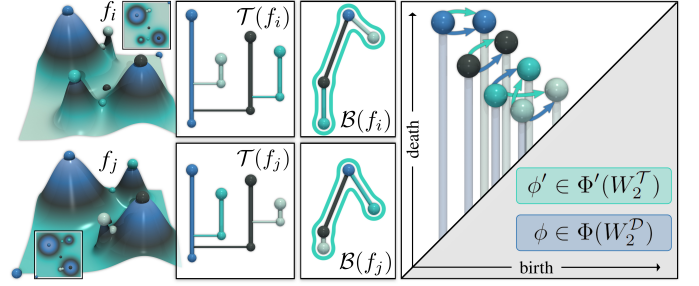


Fig. 4. The Wasserstein distance  $W_2^T$  between two BDTs  $\mathcal{B}(f_i)$  and  $\mathcal{B}(f_j)$  optimizes an assignment between their pairs in the birth/death space (green arrows, right) with a search space constrained by the structure of the trees. It amounts to search for a common contiguous structure between the BDTs (cyan halo). When  $\varepsilon_1 = 1$ , the tree structures do not constrain the search space anymore and  $W_2^T$  get back to the Wasserstein distance between persistence diagrams  $W_2^D$  (black arrows, right).

The  $L^q$ -Wasserstein distance  $W_q^D$  is then defined as:

$$W_q^D(\mathcal{D}(f_i), \mathcal{D}(f_j)) = \min_{\phi \in \Phi} \left( \sum_{p_i \in \mathcal{D}'(f_i)} d_q(p_i, \phi(p_i))^q \right)^{1/q}, \quad (2)$$

where  $\Phi$  is the set of all possible assignments  $\phi$  mapping each point  $p_i \in \mathcal{D}'(f_i)$  to exactly one point  $p_j \in \mathcal{D}'(f_j)$ , being either an off-diagonal point or the diagonal projection of  $p_i$  (i.e.  $\phi(p_i) = \Delta(p_i) = p_j \in \mathcal{D}'(f_j)$ ), indicating the deletion of the corresponding feature. Similarly, when  $p_i$  is a diagonal point being matched to an off-diagonal point of  $\mathcal{D}'(f_j)$  it indicates its appearance. The computation of Eq. 2 can be done using traditional assignment problem solvers [6], [45], [56].

The Wasserstein distance between persistence diagrams was later generalized to BDTs [66]. The expression of this distance, noted  $W_2^T(\mathcal{B}(f_i), \mathcal{B}(f_j))$ , is the same as Eq. 2 with  $q = 2$ . The important difference relies on its possibly smaller search space of assignments, noted  $\Phi' \subseteq \Phi$ , that can be constrained by the structure of the trees. The search space  $\Phi'$  between  $\mathcal{B}(f_i)$  and  $\mathcal{B}(f_j)$ , called the set of rooted partial isomorphisms [66], corresponds to all their possible contiguous common subtrees including at least their root. It intuitively ensures that nodes in a same subtree of the first tree can only be mapped to nodes of a same subtree in the second one. The best assignment of  $\Phi'$  that minimizes the distance (cyan halo on the BDTs of Fig. 4), aims to find the largest common contiguous subtree having the lowest total cost given the matchings, deletions and insertions of this assignment. Its computation is done by recursively solving local and small assignment problems in the tree hierarchies. It is said to generalize the Wasserstein distance between persistence diagrams thanks to the parameter  $\varepsilon_1$  (Sec. II-C) that controls how much of the BDT structures should be taken into account in the distance. When  $\varepsilon_1 = 1$ , all saddles in a merge tree are collapsed and results in all the nodes of its BDT to be moved up in the hierarchy and become children of the root. In that case, the various local and small assignment problems become one large and we have  $W_2^T(\mathcal{B}(f_i), \mathcal{B}(f_j)) = W_2^D(\mathcal{D}(f_i), \mathcal{D}(f_j))$ . This parameter allows to choose a trade-off between the stability of persistence diagrams and the discriminative power of merge trees.

### III. REGION-AWARE WASSERSTEIN DISTANCES

This section introduces our generalization of the Wasserstein distance for extremum persistence diagrams and merge trees [20], [66]. It extends them by describing features not only with their minimum and maximum (i.e. their birth and their death) but with a vertex-level characterization of their regions in the input domain. An input parameter controls the impact of the regions' characterization in the distance, making possible to not use them at all, getting back to the original Wasserstein distances therefore being special cases of our method.

In the following, since persistence diagrams and merge trees can be represented as BDTs we will use this term to represent both objects. We first formulate how persistence pairs can be augmented with their respective regions to define region-aware BDTs. Then, we introduce a new ground metric between region-aware persistence pairs, being at the core of our new distance for which we show theoretical properties (metric, stable and discriminative). In Sec. IV we present our algorithm for its computation as well as two strategies to control the computation time and the memory footprint of the method.

#### A. Motivation

As shown in Fig. 3, persistence diagrams and merge trees can lack in specificity, where two significantly different datasets, from a geometric point of view, have the same topological representation. In that case, their Wasserstein distance, being 0, prevents to discriminate them.

Another failure case of current methods is illustrated in Fig. 5 showing an incorrect matching of the Wasserstein distance. Two features being significantly different from a geometric point of view (the black and green features have respectively a circle and a star shape, Fig. 5) will be incorrectly matched together by the Wasserstein distance (red dashes, Fig. 5) because they have similar birth and death values.

Overall, this motivates the use of additional information on top of existing topological representations and the development of a metric taking that information into account to allow for the discrimination of geometrically different datasets. For instance, in Fig. 5, our novel distance metric uses a vertex-level characterization of the geometry of the features, that goes beyond the simple comparison of birth and death values, and therefore provides the correct matching.

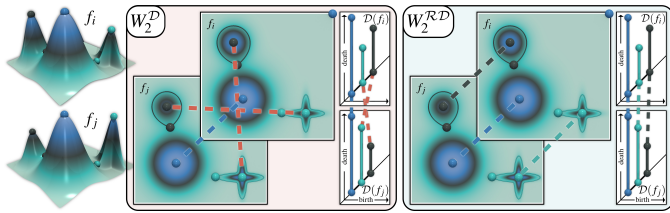


Fig. 5. The Wasserstein distance  $W_2^D$  matches topological features having overall close birth and death values. In this example, the geometry of the features with black and green spheres are significantly different (circle shape and star shape). For these two datasets, the death value of these features have been swapped resulting in their incorrect matching according  $W_2^D$  (red dashes). In contrast, our metric  $W_2^{RD}$  takes into account their geometry through their regions in the input domain and provide a correct matching.

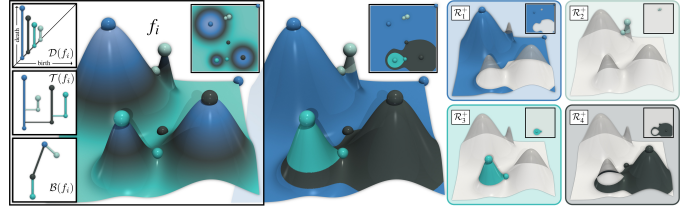


Fig. 6. The computation of the persistence diagram and the merge tree allows to uniquely segment the domain of the function  $f_i$  into regions, each corresponding to a topological feature. This figure shows the maximum regions  $\mathcal{R}_i^+$  with  $i \in \{1, \dots, 4\}$ , each being a subset of the domain, containing a unique maximum and values monotonically decreasing outward from it. The color code of the regions is the same than for the saddle-maximum persistence pairs in the topological representations.

#### B. Region-aware branch decomposition trees

We first formalize how persistence diagrams and merge trees allow to segment the domain into regions, each associated to a persistence pair. Since we work on extremum persistence diagrams and merge trees, we will focus the explanation on minimum-saddle and saddle-maximum persistence pairs. Then, we define the *region-aware* persistence pairs, an extension of persistence pairs with a vertex-level characterization of their region, and use them to define region-aware BDTs.

Given  $f : \mathcal{M} \rightarrow \mathbb{R}$ , the computation of  $\mathcal{B}(f)$  naturally segments the input domain  $\mathcal{M}$  through the following mapping:

$$\pi_f : \mathcal{M} \rightarrow \mathcal{B}(f), \quad p \mapsto b_i = \pi_f(p),$$

which assigns every vertex  $p \in \mathcal{M}$  to a unique persistence pair  $b_i \in \mathcal{B}(f)$  representing the connecting component to which  $p$  belongs to during the filtration. For each pair  $b_i \in \mathcal{B}(f)$ , we define its region  $\mathcal{R}_i$  as the set-valued pre-image of  $\pi$ :

$$\mathcal{R}_i = \pi_f^{-1}(b_i) = \{p \in \mathcal{M} \mid \pi_f(p) = b_i\}.$$

Because  $\pi_f$  is surjective and assigns exactly one persistence pair to every point, the family  $\{\mathcal{R}_i\}_i$  forms a partition of the domain:  $\bigcup_i \mathcal{R}_i = \mathcal{M}$  and  $\mathcal{R}_{i_1} \cap \mathcal{R}_{i_2} = \emptyset$  for  $i_1 \neq i_2$ . Each region  $\mathcal{R}_i$  contains precisely one extremum  $e_i$  and all its function values vary monotonically away from  $e_i$ .  $\mathcal{R}_i$  is a *maximum region*, noted  $\mathcal{R}_i^+$ , if  $b_i$  is a saddle-maximum persistence pair (Fig. 6) or a *minimum region*, noted  $\mathcal{R}_i^-$ , if  $b_i$  is a minimum-saddle persistence pair.

Given a persistence pair  $b_i$  of the BDT  $\mathcal{B}(f)$ , associated with a pair of values  $(x_i, y_i)$  (its birth and its death), let  $s_i$  the value of its saddle critical point (i.e. either its death  $y_i$  if it is a minimum-saddle pair or its birth  $x_i$  if it is a saddle-maximum one), and  $e_i$  its value at the extremum (the other value than  $s_i$ ), we define its region-aware generalization:

$$b_i^{\mathcal{R}} = (\mathcal{R}_i, f|_{\mathcal{R}_i}, s_i),$$

where  $f|_{\mathcal{R}_i}$  corresponds to the function  $f$  restricted to the domain of  $\mathcal{R}_i$ . Note that by definition, the value at the extremum  $e_i$  of  $b_i$  lies in the image of  $f|_{\mathcal{R}_i}$ , since its corresponding vertex belongs to  $\mathcal{R}_i$ , but not its saddle value  $s_i$ . This generalization captures the feature's spatial support  $\mathcal{R}_i$  with its values  $f|_{\mathcal{R}_i}$  and the saddle value  $s_i$ , providing a strictly richer characterization than the classical birth and death pair. Then, the region-aware



BDT  $\mathcal{B}^{\mathcal{R}}(f)$  is defined as the BDT  $\mathcal{B}(f)$  where each of its pair  $b_i$  is replaced by its region-aware generalization  $b_i^{\mathcal{R}}$ .

### C. Definition

After augmenting BDTs with a vertex-level characterization of the regions of their features, we aim to define a distance that leverages this information. For this, we generalize the original Wasserstein distances [20], [66] (Sec. II-D) to make them able to use such geometric properties in order to overcome their limitations regarding their lack of geometric considerations (see failure cases in Fig. 3 and Fig. 5).

We start by defining a novel ground metric allowing to compare region-aware persistence pairs. For this, we generalize the classical ground metric used for persistence diagrams and merge trees (Eq. 1) to use up to all the region values of such persistence pairs rather than just their minimum and maximum. We then define the region-aware Wasserstein distance for which we show theoretical properties (metric, stable and discriminative). We additionally propose a parameter adjusting the influence of the region properties on the distance (Sec. III-D3), allowing to ignore them entirely, reverting to the classical case. One benefit of this approach is its adaptability to different applications, depending on the relevance of region information. Moreover, as we will see in Sec. IV-A, it also allows for controlling the computation time of the method.

1) **Region-aware ground metric:** Let two BDTs  $\mathcal{B}(f_1)$  and  $\mathcal{B}(f_2)$  and two persistence pairs, one from each BDT,  $b_i = (e_i, s_i) \in \mathcal{B}(f_1)$  and  $b_j = (e_j, s_j) \in \mathcal{B}(f_2)$ . The classical ground metric (Eq. 1) corresponds to a two-term  $L_q$  norm, one comparing the saddle values ( $s_i$  and  $s_j$ ) and the other the values of the extrema ( $e_i$  and  $e_j$ ). The region-aware generalizations of  $b_i$  and  $b_j$ , i.e.  $b_i^{\mathcal{R}} = (\mathcal{R}_i, f_{1|\mathcal{R}_i}, s_i)$  and  $b_j^{\mathcal{R}} = (\mathcal{R}_j, f_{2|\mathcal{R}_j}, s_j)$ , replace the values at their respective extrema ( $e_i$  and  $e_j$ ) by a vertex-level characterization of their regions ( $\mathcal{R}_i, f_{1|\mathcal{R}_i}$  and  $\mathcal{R}_j, f_{2|\mathcal{R}_j}$ ). Recall that no information is lost since  $e_i \in \text{Im}(f_{1|\mathcal{R}_i})$  and  $e_j \in \text{Im}(f_{2|\mathcal{R}_j})$  by definition. Accordingly, we generalize the classical ground metric to region-aware persistence pairs by replacing the term comparing the values at their extremum by one taking into account the characterization of the regions. We call this new term the *regional discrepancy*, noted  $C_q^{\mathcal{R}}$ , and we define the region-aware ground metric as:

$$d_q^{\mathcal{R}}(b_i^{\mathcal{R}}, b_j^{\mathcal{R}}) = (|s_i - s_j|^q + C_q^{\mathcal{R}}(b_i^{\mathcal{R}}, b_j^{\mathcal{R}}))^{1/q}. \quad (3)$$

If the regional discrepancy  $C_q^{\mathcal{R}}$  is defined such that only the values of the extremum of  $b_i^{\mathcal{R}}$  and  $b_j^{\mathcal{R}}$  are compared (i.e.  $C_q^{\mathcal{R}}(b_i^{\mathcal{R}}, b_j^{\mathcal{R}}) = |e_i - e_j|^q$ ) then  $d_q^{\mathcal{R}}$  reduces to the original ground metric. We now define it in order to use up to all the regions' information provided by region-aware persistence pairs, keeping the comparison of only the values of their extrema as a special case thanks to an input parameter (Sec. III-D3).

To compare the regions while preserving the comparison of the values of their extrema, we first *align* (i.e. center) the regions at their respective extrema so that they coincide (Fig. 7b), then we compare the values of the centered regions (Fig. 7c). We note  $\mathcal{R}'_i$  and  $\mathcal{R}'_j$  the centered regions and  $f'_{1|\mathcal{R}'_i}$  and  $f'_{2|\mathcal{R}'_j}$  their functions. The intersection of the centered

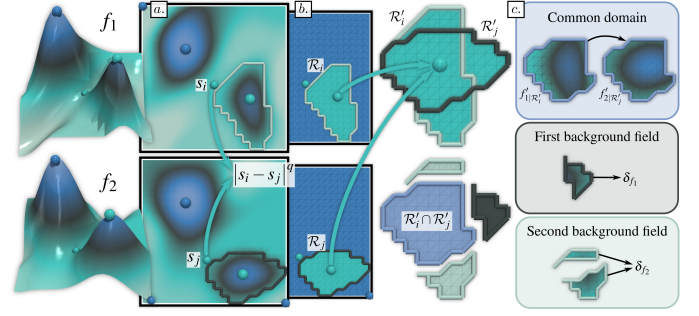


Fig. 7. The computation of the region-aware ground metric  $d_q^{\mathcal{R}}$  consists in the comparison of the saddle values of two persistence pairs (a), then, their regions  $\mathcal{R}_i$  and  $\mathcal{R}_j$  are aligned at their extrema (b) and a distance is computed on their common domain  $\mathcal{R}'_i \cap \mathcal{R}'_j$  (c). An additional cost is used for the points outside the common domain (i.e. in  $\mathcal{R}'_i \setminus \mathcal{R}'_j$  and  $\mathcal{R}'_j \setminus \mathcal{R}'_i$ ) using the background fields  $\delta_{f_1}$  and  $\delta_{f_2}$  to make these comparisons well-defined.

regions  $\mathcal{R}'_i \cap \mathcal{R}'_j$  (Fig. 7b, violet area) contains at least their extrema. On this overlap, both functions are defined allowing a pointwise comparison. For each point outside this overlap, i.e. on the symmetric difference  $\mathcal{R}'_i \triangle \mathcal{R}'_j = (\mathcal{R}'_i \setminus \mathcal{R}'_j) \cup (\mathcal{R}'_j \setminus \mathcal{R}'_i)$  (Fig. 7b, white and black areas), only one region (either one of them depending on the point) provides a value with no counterpart in the other. To make the comparisons well-defined everywhere, we embed each function  $f'_{1|\mathcal{R}'_i}$  and  $f'_{2|\mathcal{R}'_j}$  into the union domain  $\mathcal{R}'_i \cup \mathcal{R}'_j$  by extending them with a *background* field, respectively  $\delta_{f_1}$  and  $\delta_{f_2}$ , outside their original support. This allows for points outside the common domain to also contribute. We therefore define the regional discrepancy as:

$$C_q^{\mathcal{R}}(b_i^{\mathcal{R}}, b_j^{\mathcal{R}}) = \sum_{v \in \mathcal{R}'_i \cap \mathcal{R}'_j} |f'_{1|\mathcal{R}'_i}(v) - f'_{2|\mathcal{R}'_j}(v)|^q \quad (4)$$

$$+ \sum_{x \in \mathcal{R}'_i \setminus \mathcal{R}'_j} |f'_{1|\mathcal{R}'_i}(x) - \delta_{f_2}(x)|^q \quad (5)$$

$$+ \sum_{y \in \mathcal{R}'_j \setminus \mathcal{R}'_i} |\delta_{f_1}(y) - f'_{2|\mathcal{R}'_j}(y)|^q, \quad (6)$$

intuitively measuring the cost of transforming one aligned region to another. Line 4 compares their functions pointwise where both aligned regions coincide. Line 5 penalizes the parts of the first region that must be deleted and line 6 the parts of the second region that must be inserted. These two lines penalize unmatched support via the background fields  $\delta_{f_1}$  and  $\delta_{f_2}$ . A natural background is the null field,  $\delta_f(x) = 0$  for all  $x$ , which encodes the notion that a feature contributes no signal outside its own region and yields the intended interpretation of lines 5 and 6 as deletion and insertions of nonshared points. Another possibility is the background as the data itself,  $\delta_f(x) = f(x)$  for all  $x$ , which allows to show the stability of our new metric (Sec. III-D2), intuitively comparing the points outside the region overlap to what the other function is at those locations. When regions degenerate to their extrema, the symmetric difference terms vanish and the overlap term reduces to  $|e_i - e_j|^q$ , recovering the classical case.

2) **Region-aware projection:** Before defining a Wasserstein distance using the region-aware ground metric, we first generalize to region-aware persistence pairs the possibility of match-

ing a persistence pair to the diagonal (representing a deletion or insertion). In the classical case (Sec. II-D), a persistence pair  $b_i = (e_i, s_i)$  can be matched to its diagonal projection  $(m_i, m_i)$  with  $m_i = \frac{e_i + s_i}{2}$ . For a region-aware persistence pair  $b_i^{\mathcal{R}} = (\mathcal{R}_i, f|_{\mathcal{R}_i}, s_i)$  we consider a constant function  $f|_{\mathcal{R}_i}^{\Delta}$  that maps all points of  $\mathcal{R}_i$  to the diagonal projection value  $m_i$  and define the region-aware projection:

$$\Delta^{\mathcal{R}}(b_i^{\mathcal{R}}) = (\mathcal{R}_i, f|_{\mathcal{R}_i}^{\Delta}, m_i),$$

The cost of matching  $b_i^{\mathcal{R}}$  to the diagonal (i.e. deleting or inserting it) is then  $d_q^{\Delta}(b_i^{\mathcal{R}}) = d_q^{\Delta}(b_i^{\mathcal{R}}, \Delta^{\mathcal{R}}(b_i^{\mathcal{R}}))$ . When  $\mathcal{R}_i$  collapses to its extremum the cost reduces to the classical point-to-diagonal distance, consistent with the classical metric.

3) **Region-aware Wasserstein distance:** Based on the definition of the Wasserstein distance between BDTs [66] we now define the region-aware Wasserstein distance between two region-aware BDTs  $\mathcal{B}^{\mathcal{R}}(f_1)$  and  $\mathcal{B}^{\mathcal{R}}(f_2)$ . It consists in the search of a mapping  $\phi'$  from the nodes of  $\mathcal{B}^{\mathcal{R}}(f_1)$  to the nodes of  $\mathcal{B}^{\mathcal{R}}(f_2)$  having the lowest total cost measured by the region-aware ground metric  $d_q^{\mathcal{R}}$  (Eq. 3). Let  $B_1$  the nodes in  $\mathcal{B}^{\mathcal{R}}(f_1)$  being mapped by  $\phi'$ , let  $\overline{B}_1$  (the complement of  $B_1$ ) and  $\overline{B}_2$  the nodes being not mapped in respectively  $\mathcal{B}^{\mathcal{R}}(f_1)$  and  $\mathcal{B}^{\mathcal{R}}(f_2)$ . The region-aware Wasserstein distance is defined as:

$$\begin{aligned} W_q^{\mathcal{R}\mathcal{T}}(\mathcal{B}^{\mathcal{R}}(f_1), \mathcal{B}^{\mathcal{R}}(f_2)) &= \min_{\phi' \in \Phi'} \left( \sum_{b_i^{\mathcal{R}} \in B_1} d_q^{\mathcal{R}}(b_i^{\mathcal{R}}, \phi'(b_i^{\mathcal{R}}))^q \right. \\ &\quad + \sum_{b_i^{\mathcal{R}} \in \overline{B}_1} d_q^{\Delta}(b_i^{\mathcal{R}})^q \\ &\quad \left. + \sum_{b_j^{\mathcal{R}} \in \overline{B}_2} d_q^{\Delta}(b_j^{\mathcal{R}})^q \right)^{1/q}, \end{aligned} \quad (7) \quad (8) \quad (9)$$

where  $\Phi'$  is the search space associated with the original Wasserstein distance between BDTs (Sec. II-D). When  $\varepsilon_1 = 1$ , getting back to the persistence diagrams case, we note it  $W_q^{\mathcal{R}\mathcal{T}}$ .

#### D. Properties

1) **Metric property:** We show in Appendix A that the region-aware ground distance  $d_q^{\mathcal{R}}$  (Eq. 3) is indeed a metric for the space of region-aware persistence pairs (after centering their regions and extending their centered functions with their background fields). We summarize here after the main elements of the proof in an intuitive manner and invite the reader to read Appendix A for more details. Non-negativity and symmetry are satisfied by construction because, respectively,  $d_q^{\mathcal{R}}$  is the  $q^{\text{th}}$  root of the sum of absolute values raised to a power of  $q$  with  $q \geq 1$ , and deletion and insertion costs (lines 5 and 6) are defined in a symmetric manner. Identity of indiscernibles follows naturally after centering regions at their extrema and extending their functions with their background fields, pairs are considered identical if and only if their saddle values are equals as well as the extension of their functions on the union of their domain. For the triangle inequality, we consider an  $L_q$  norm on the common domain of three region-aware persistence pairs with their centered functions being extended by their background fields, allowing to use the Minkowski's inequality. After showing that  $d_q^{\mathcal{R}}$  is a metric, we

then show that  $W_2^{\mathcal{R}\mathcal{T}}$  (and therefore  $W_2^{\mathcal{R}\mathcal{D}}$ ) is indeed a metric for region-aware branch decomposition trees (Appendix A).

2) **Stability property:** We study in Appendix C the stability of the region-aware Wasserstein distance for persistence diagrams  $W_q^{\mathcal{R}\mathcal{D}}$  when excluding extremum-swap instabilities (when the extrema of the input functions don't change order, as described in details in [88]). We show that, with data background, it can be bounded by above by the  $L_q$  norm of the difference of the input scalar fields up to a constant independent of them. Specifically, we show that  $W_q^{\mathcal{R}\mathcal{D}}(\mathcal{D}^{\mathcal{R}}(f), \mathcal{D}^{\mathcal{R}}(g)) \leq 2^{1/q} \|f - g\|_q$ . For this, we take inspiration of the proof by Skraba and Turner [73] for the original Wasserstein distance between persistence diagrams that we extend to take into account the region properties.

3) **Generalization property:** We now introduce an operator controlling the impact of the regions properties of the features on the distance. Given a region-aware persistence pair  $b_i^{\mathcal{R}} = (\mathcal{R}_i, f|_{\mathcal{R}_i}, s_i)$ , we define the operator  $\mathcal{S}(b_i^{\mathcal{R}}, \lambda)$  with  $\lambda \in [0, 1]$  returning a modified version of  $b_i^{\mathcal{R}}$  where the size of its region decreases as long as  $\lambda$  increases. Specifically, when  $\lambda = 0$  the region of  $b_i^{\mathcal{R}}$  is not modified. Then, the more  $\lambda$  increases, the more points of its region will be removed. Finally, when  $\lambda = 1$ , only the extremum of the region is kept. More formally, let  $\mathcal{R}_i^{\lambda}$  the modified version of  $\mathcal{R}_i$  by  $\mathcal{S}(b_i^{\mathcal{R}}, \lambda)$ , then the following properties must be satisfied:  $\mathcal{R}_i^{\lambda_1} \supseteq \mathcal{R}_i^{\lambda_2}$  for  $\lambda_1 < \lambda_2$  and  $\mathcal{R}_i^{\lambda} = \{e\}$  when  $\lambda = 1$  where  $e$  is the extremum  $b_i^{\mathcal{R}}$ . We provide an example of such operator in Sec. IV-A inspired by subsampling. This operator is computed in a pre-processing step before computing the distance. By definition, when  $\lambda = 1$  only the extremum remains in a region, therefore  $d_q^{\mathcal{R}}(\mathcal{S}(b_i^{\mathcal{R}}, \lambda), \mathcal{S}(b_j^{\mathcal{R}}, \lambda))$  will only compare the extrema and saddle values as in the simple case, getting back to the original ground metric of Eq. 1. In that case, we have  $W_2^{\mathcal{R}\mathcal{T}}(\mathcal{B}^{\mathcal{R}}(f_1), \mathcal{B}^{\mathcal{R}}(f_2)) = W_2^{\mathcal{T}}(\mathcal{B}(f_1), \mathcal{B}(f_2))$ .

4) **Discriminative property:** Using the operator  $\mathcal{S}(b_i^{\mathcal{R}}, \lambda)$  we now show that the region-aware Wasserstein distance  $W_2^{\mathcal{R}\mathcal{T}}$  is more discriminative than the original Wasserstein distance  $W_2^{\mathcal{T}}$ . Let  $b_{i,\lambda}^{\mathcal{R}} = \mathcal{S}(b_i^{\mathcal{R}}, \lambda)$  and  $\mathcal{R}_i^{\lambda}$  its region. Recall that the more  $\lambda$  decreases the more  $\mathcal{R}_i^{\lambda}$  contains points, i.e.  $\mathcal{R}_i^{\lambda_1} \supseteq \mathcal{R}_i^{\lambda_2}$  for  $\lambda_1 < \lambda_2$ . When comparing two pairs with the region-aware ground metric (Eq. 3), it implies that every sum in the regional discrepancy  $C_q^{\mathcal{R}}$  for  $\lambda_2$  is taken over a subset of the indices used for  $\lambda_1$ , since all terms are added and non-negative it implies that  $C_q^{\mathcal{R}}(b_{i,\lambda_1}^{\mathcal{R}}, b_{j,\lambda_1}^{\mathcal{R}}) \geq C_q^{\mathcal{R}}(b_{i,\lambda_2}^{\mathcal{R}}, b_{j,\lambda_2}^{\mathcal{R}})$ . The saddle values of the two pairs being unchanged we therefore have  $d_q^{\mathcal{R}}(b_{i,\lambda_1}^{\mathcal{R}}, b_{j,\lambda_1}^{\mathcal{R}}) \geq d_q^{\mathcal{R}}(b_{i,\lambda_2}^{\mathcal{R}}, b_{j,\lambda_2}^{\mathcal{R}})$ . Let  $\mathcal{B}_{\lambda}^{\mathcal{R}}(f)$  the region-aware BDT  $\mathcal{B}^{\mathcal{R}}(f)$  on which the operator  $\mathcal{S}(\cdot, \lambda)$  has been applied to all its pairs. For any matching  $\phi' \in \Phi'$ , let  $\mathcal{C}_{\lambda}(\phi')$  its associated cost between  $\mathcal{B}_{\lambda}^{\mathcal{R}}(f_1)$  and  $\mathcal{B}_{\lambda}^{\mathcal{R}}(f_2)$ . We have  $\mathcal{C}_{\lambda_1}(\phi') \geq \mathcal{C}_{\lambda_2}(\phi')$ , taking the minima over  $\phi' \in \Phi'$  yields:

$$W_2^{\mathcal{R}\mathcal{T}}(\mathcal{B}_{\lambda_1}^{\mathcal{R}}(f_1), \mathcal{B}_{\lambda_1}^{\mathcal{R}}(f_2)) \geq W_2^{\mathcal{R}\mathcal{T}}(\mathcal{B}_{\lambda_2}^{\mathcal{R}}(f_1), \mathcal{B}_{\lambda_2}^{\mathcal{R}}(f_2)).$$

Taking  $\lambda_2 = 1$  (getting back to the original Wasserstein distance, Sec. III-D3) and  $\lambda_1 = 0$  (being the region-aware Wasserstein distance without preprocessing by  $\mathcal{S}$ ) we get:  $W_2^{\mathcal{R}\mathcal{T}}(\mathcal{B}^{\mathcal{R}}(f_1), \mathcal{B}^{\mathcal{R}}(f_2)) \geq W_2^{\mathcal{T}}(\mathcal{B}(f_1), \mathcal{B}(f_2))$ .

#### IV. ALGORITHM

We compute  $W_q^{\mathcal{RT}}$  by extending the dynamic-programming scheme of Pont et al. [66], which efficiently searches over rooted partial isomorphisms of two input BDTs. In our setting, the input trees are the region-aware BDTs (Sec. III-B), possibly preprocessed with  $\mathcal{S}(\cdot, \lambda)$  (Sec. III-D3), every comparison of two pairs is done with the region-aware ground metric  $d_q^{\mathcal{R}}$  (Sec. III-C1) and deletions/insertions are handled via the region-aware projection  $\Delta^{\mathcal{R}}(\cdot)$  (Sec. III-C2). Since  $d_q^{\mathcal{R}}$  is a metric (III-D1), the optimality of the algorithm of Pont et al. [66] holds and will indeed find the assignments minimizing the region-aware Wasserstein distance (Eq. 7). The algorithm takes  $\mathcal{O}(|\mathcal{B}|^2)$  steps in practice [66], with  $|\mathcal{B}|$  the number of nodes in the input BDTs. The computation of  $d_q^{\mathcal{R}}$  necessitates  $\mathcal{O}(n+m)$  steps in the worst case, where  $n$  and  $m$  are respectively the number of vertices in the two regions to be compared. The overall complexity of  $W_q^{\mathcal{RT}}$  is therefore  $\mathcal{O}(|\mathcal{B}|^2(N+M))$  with  $N$  and  $M$  the total number of vertices in the input scalar fields. For a region-aware BDT, the values of all its regions are stored without duplication in a single structure corresponding to the original data. In this work we focus on regular grids, each region-aware persistence pair then stores the maximal bounds of its region within the original data, it allows to easily convert a local index of the region to a global index in the original data. It additionally stores a boolean mask indicating whether or not a specific index in its maximal bounds indeed belongs to the region.

We then present two strategies to control both computation time and memory footprint of the method. First, in Sec. IV-A, we provide an example of the operator  $\mathcal{S}$  (Sec. III-C) allowing to choose how much of the regions properties should be taken into account, having a direct impact on execution time. Second, in Sec. IV-B, we present compression strategies in order to obtain low-memory representations of the regions.

##### A. Subsampled regions comparison

We now define an operator  $\mathcal{S}(b_i^{\mathcal{R}}, \lambda)$  (Sec. III-C) based on subsampling, allowing to use only a subset of a region  $\mathcal{R}_i$  of a region-aware persistence pair  $b_i^{\mathcal{R}}$  in the computation. When the input domain is a regular grid, the subsampling-based operator  $\mathcal{S}(b_i^{\mathcal{R}}, \lambda)$  selects a subset of points with a stride of  $n$  in each dimension (one point over  $n$ ) starting from the extremum of  $\mathcal{R}_i$  ensuring its inclusion. We use  $n = \lambda m + 1$ , rounded to the nearest integer where  $m$  is the maximum number of points in the dimension containing the most. When  $\lambda = 0$  the stride of 1 ensures that indeed all the points in the region are kept (as defined in Sec. III-C for this operator). When  $\lambda = 1$  the stride of  $m + 1$  ensures that no points except the extremum of the region is kept. All the ensembles considered in this work are regular grids and the members within a same ensemble have the same number of points in each dimension. This ensures that the points of two subsampled regions being aligned at their respective extremum will also be aligned.

##### B. Regions compression

The region-aware Wasserstein distance requires access to the feature regions in the original data. However, a key

advantage of persistence diagrams and merge trees is their significantly lower memory footprint compared to the data themselves, enabling analysis tasks without storing the full dataset. To compute the region-aware Wasserstein distance without retaining the original data on disk, compression methods can be considered. Specifically, we propose to compress the original data (to avoid to fully store it on disk) given an input parameter  $\tau \in [0, 1]$  such that the size of the compressed data is a ratio  $\tau$  of the original data size. When  $\tau = 1$  the original data is kept. The more  $\tau$  decreases, the more the size of the compressed data decreases as well. When  $\tau = 0$ , no parameters are allowed for the compressed data and therefore no information about the regions are considered. We additionally store the region membership of each original vertex, such that the decompressed data has the same than the original one. The compression parameter  $\tau$  allows to choose the memory overhead of our method compared to the one of the original persistence diagrams and merge trees. By doing so, the user can choose how much memory storage to use depending on its capacity and its use-cases.

We therefore focus on compression methods that allow to easily control the size of the compressed representation. Most compression methods focus however on an important but different use-case, where the error is controlled, and not necessarily the size. Still, some methods can be noted including ZFP [49], neural fields [50], and B-splines [64]. We show in Appendix D how given  $\tau$ , the parameters of these models need to be defined in order to be as close as possible to the target size. In Sec. V-D we show how this compression affects our distance and propose guidelines to choose between the different models given user requirements such as compression size and execution time.

#### V. RESULTS

This section presents experimental results obtained on a computer with two AMD EPYC 7453 CPUs (2.75 Ghz, 28 cores each) and 512 GB of RAM. The input persistence diagrams and merge trees were computed with FTM [29] and pre-processed to remove low persistent features (with a threshold of 0.5 % of the data range). We have implemented our method in C++ as an extension of the Wasserstein distance module of persistence diagrams and merge trees in TTK [81].

Our experiments were performed on a variety of simulated and acquired 2D and 3D ensembles inspired by [66]. They used ensembles extracted from various SciVis contests [25], [26], [40], [59], [63], [79], [85], [92] or from previous works [23]. In this work, we have extended the SciVis contests ensembles used in [66] in order to create more challenging ones (described in Tab. I). For most of them, it consists of all the available data (for a same set of simulation parameters when applicable) instead of only a small subset of it.

1) *Baseline methods and parameter selection:* We compare our method to the classical Wasserstein distances  $W_2^{\mathcal{D}}$  and  $W_2^{\mathcal{T}}$  [20], [66] and to two methods that took a first step in the direction of using geometrical information by focusing on coarse properties. First, the so-called *geometrical lifting* [48], [74], [93], i.e. the use of the coordinates in the domain of the

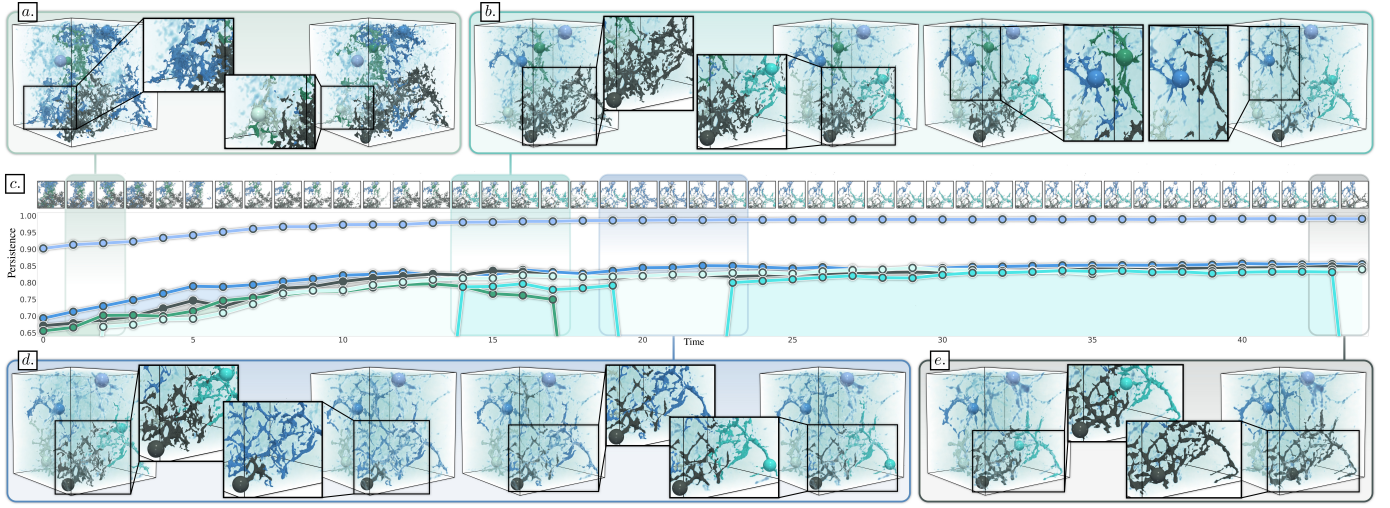


Fig. 8. Visual analysis of the *Dark Sky* ensemble (dark matter density) with the region-aware Wasserstein distance of merge trees. Members at different points in time are shown in (a), (b), (d) and (e). We track the evolution of the topological features using the assignments of our method for consecutive time steps (matched features having the same color). The temporal persistence curves in (c) allow to visually detect important changes in the ensemble by representing the evolution of the persistence of the matched features over time (only curves for the highest persistent features are shown with same coloring than in the input domain). In this example it allows to track the appearance, the split and the merge of the filamentary structures of the cosmic web. For instance, in the first time window (a), we see the white feature emerging from the blue one. In (d), both the cyan and blue features merges (first zoom), and split later (second zoom), the same color is used after split. Finally, in (e) it merges with the black one.

critical points in the distance computation, that we note  $W_2^{\mathcal{L}D}$  and  $W_2^{\mathcal{L}T}$  for respectively persistence diagrams and merge trees. Second, the use of the volume of the regions [69], i.e. the number of vertices, that we respectively note  $W_2^{\mathcal{V}D}$  and  $W_2^{\mathcal{V}T}$ . We formally define these methods in Appendix E and use the implementations available in TTK [81]. The last two methods are subject to parameters controlling the impact of the geometrical properties (Appendix E), the weights  $w_{\mathcal{L}}$  for the coordinates distance and  $w_{\mathcal{V}}$  for the volume difference. Our method is subject to the subsampling parameter  $\lambda$  (Sec. IV-A). We use the values  $w_{\mathcal{L}} = 0.5$ ,  $w_{\mathcal{V}} = 0.2$  and  $\lambda = 0.1$  since they gave the best results for our applications averaged over all the ensembles (see Sec. V-C). When processing merge trees, we use the recommended value of  $\varepsilon_1 = 0.05$  given by Pont et al. [66]. Finally, for the ground metric (Eq. 1 and Eq. 3), we use  $q = 2$ , the standard Euclidean metric, consistent with prior works for visualization applications [65]–[67], [72], [83].

#### A. Time performance

Tab. I evaluates the time performance for both persistence diagrams and merge trees of our method compared to the three baselines. We observe that the execution time of each method is indeed a function of the number of ensemble members ( $N$ ) and the average number of persistence pairs ( $|\mathcal{B}|$ ).  $W_2^{\mathcal{R}D}$  and  $W_2^{\mathcal{R}T}$  are 1.3 slower in average than  $W_2^{\mathcal{L}D}$  and  $W_2^{\mathcal{L}T}$ , it is an expected result since they use the regions' properties in the computation and compare more values than only the birth and the death. Overall, the lifting distances  $W_2^{\mathcal{L}D}$  and  $W_2^{\mathcal{L}T}$  and the volume distances  $W_2^{\mathcal{V}D}$  and  $W_2^{\mathcal{V}T}$  have execution times between those of the classical and region-aware Wasserstein distances. Averaged over all ensembles, it takes approximately 3 minutes to compute the distance matrix of an ensemble using  $W_2^{\mathcal{R}D}$  and  $W_2^{\mathcal{R}T}$ . Note that the same topological representations are used as input for

TABLE I  
COMPARISON OF THE RUNNING TIMES (IN SECONDS), BETWEEN OUR METHOD AND THE BASELINES, OF THE DISTANCE MATRIX COMPUTATION FOR EACH ENSEMBLE USING PERSISTENCE DIAGRAMS AND MERGE TREES.

| Dataset                 | $N$ | $ \mathcal{B} $ | Persistence Diagrams |                      |                      |                             | Merge Trees         |                      |                      |                             |
|-------------------------|-----|-----------------|----------------------|----------------------|----------------------|-----------------------------|---------------------|----------------------|----------------------|-----------------------------|
|                         |     |                 | $W_2^{\mathcal{D}}$  | $W_2^{\mathcal{L}D}$ | $W_2^{\mathcal{V}D}$ | $W_2^{\mathcal{R}D}$ (ours) | $W_2^{\mathcal{T}}$ | $W_2^{\mathcal{L}T}$ | $W_2^{\mathcal{V}T}$ | $W_2^{\mathcal{R}T}$ (ours) |
| Asteroid Impact (3D)    | 68  | 854             | 33.3                 | 31.7                 | 34.2                 | 35.9                        | 20.4                | 19.8                 | 20.5                 | 23.2                        |
| Cloud Processes (2D)    | 91  | 2880            | 419.5                | 387.9                | 435.2                | 503.4                       | 350.7               | 312.3                | 359.8                | 416.7                       |
| Viscous Fingering (3D)  | 120 | 48              | 13.2                 | 9.9                  | 10.2                 | 5.8                         | 7.2                 | 8.6                  | 6.1                  | 7.3                         |
| Dark Matter (3D)        | 99  | 4646            | 973.9                | 984.4                | 1,187.9              | 1,492.1                     | 759.9               | 755.2                | 789.7                | 1,138.8                     |
| Volcanic Eruptions (2D) | 200 | 828             | 102.2                | 99.2                 | 109.9                | 143.1                       | 45.2                | 49.2                 | 54.1                 | 66.3                        |
| Ionization Front (2D)   | 200 | 68              | 20.8                 | 23.8                 | 20.4                 | 17.4                        | 15.4                | 19.8                 | 18.1                 | 18.1                        |
| Ionization Front (3D)   | 200 | 1228            | 218.2                | 236.5                | 235.8                | 354.8                       | 236.9               | 265.2                | 262.4                | 298.5                       |
| Earthquake (3D)         | 212 | 614             | 56.4                 | 61.8                 | 61.4                 | 95.6                        | 36.8                | 44.6                 | 40.1                 | 64.3                        |
| Isabel (3D)             | 48  | 1832            | 66.6                 | 58.9                 | 67.2                 | 68.4                        | 54.6                | 52.9                 | 56.6                 | 62.3                        |
| Starting Vortex (2D)    | 12  | 39              | 0.1                  | 0.1                  | 0.1                  | 0.1                         | 0.1                 | 0.1                  | 0.1                  | 0.1                         |
| Sea Surface Height (2D) | 48  | 1453            | 40.9                 | 37.9                 | 41.6                 | 45.3                        | 30.6                | 28.4                 | 29.7                 | 32.2                        |
| Vortex Street (2D)      | 45  | 9               | 0.2                  | 0.2                  | 0.3                  | 0.2                         | 0.4                 | 0.4                  | 0.5                  | 0.4                         |
| Heated Cylinder (2D)    | 60  | 12              | 0.9                  | 0.5                  | 0.6                  | 0.6                         | 0.5                 | 0.9                  | 0.6                  | 0.7                         |

all methods.  $W_2^{\mathcal{L}D}$  and  $W_2^{\mathcal{L}T}$  additionally use the coordinates of the extrema,  $W_2^{\mathcal{V}D}$  and  $W_2^{\mathcal{V}T}$  the volume of the features and  $W_2^{\mathcal{R}D}$  and  $W_2^{\mathcal{R}T}$  the regions' properties. The subsampling parameter  $\lambda$  (Sec. IV-A) and the compression parameter  $\tau$  (Sec. IV-B) only impacts the additional information used by our method and not the input topological representations.

#### B. Topological matching for feature tracking

The computation of our distance consists in the optimization of assignments between the persistence pairs of input BDTs. Since each persistence pair corresponds to a feature in the input data, these assignments allow to establish relationships between them. When a time-varying ensemble is considered, a distance can be computed for each consecutive time steps and the global evolution of each feature can be tracked over the whole ensemble. Overall, the matchings provide visual hints to the users to help them relate features with each other.

An example of such matchings is illustrated in Fig. 9 where we compare the matchings provided by the three baseline methods and ours on the cloud ensemble. While each column



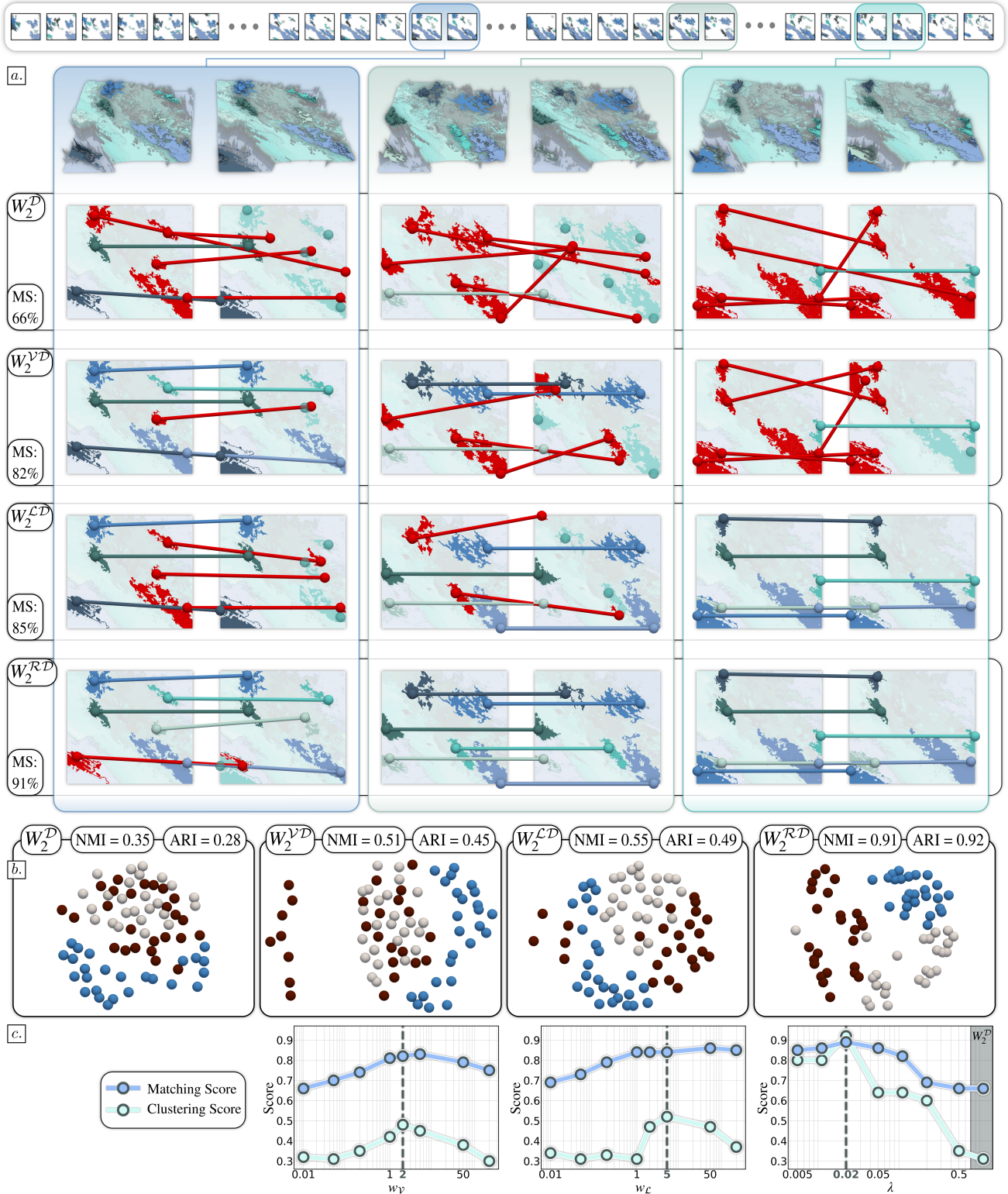


Fig. 9. Pairwise feature tracking and dimensionality reduction of the Cloud ensemble (cloud pressure in Pa), using the baseline methods ( $W_2^D$ ,  $W_2^{VD}$  and  $W_2^{LD}$ ) and ours ( $W_2^{RD}$ ). In (a), three matchings are selected to highlight where  $W_2^{VD}$ ,  $W_2^{RD}$  and  $W_2^{LD}$  (respectively for each column) perform the best comparing to the other methods. No matchings of  $W_2^D$  manage to outperform the other methods, hence no column specifically for this method. We see that  $W_2^{LD}$  can fail when the extrema of a same region at different time steps change position (first column in (a)). We also evaluate the accuracy of the matchings, *matching score* (MS), given a ground-truth of the features that should be matched together, our method provides a 91% accuracy, outperforming the baseline methods. In (b), we show the MDS planar embeddings of the four methods, our method outperforms greatly the others with a NMI and ARI of 0.91 and 0.92 respectively and have clusters that are visually more separated. In (c), we show the matching score and the clustering score (average of NMI and ARI) given the parameter that needs to be tuned for each method (none for  $W_2^D$ ). For each method, the parameter was selected to maximize both matching and clustering scores. Note that the chosen  $w_L$  value for  $W_2^{LD}$  can be not high enough to match some features together (second column in (a)) while it still provides the best results overall. For  $W_2^{RD}$  and  $\lambda = 1$  we get back to  $W_2^D$  (gray area in the chart), for any other values of  $\lambda$  we see that  $W_2^{RD}$  outperforms  $W_2^D$ .

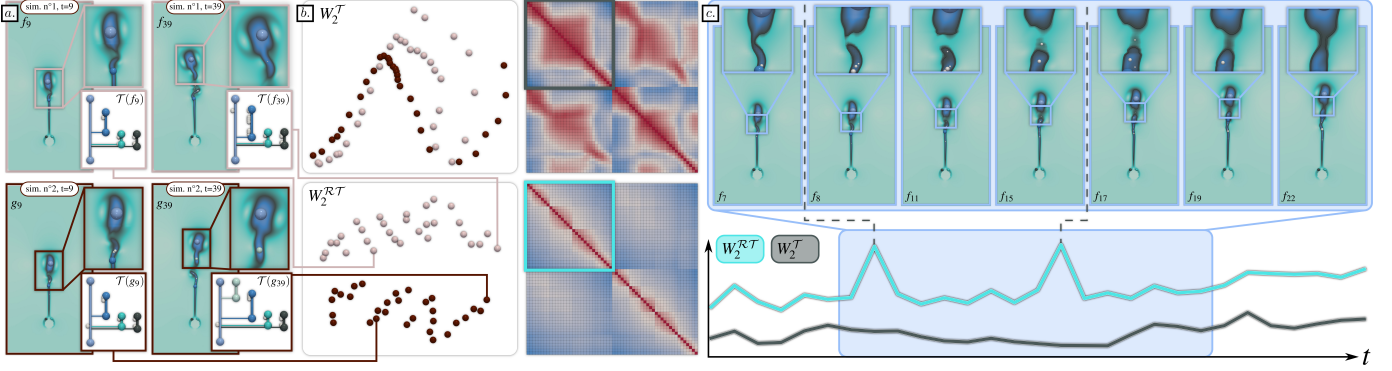


Fig. 10. Dimensionality reduction and phases detection of the heated cylinder ensemble. It consists of two time-varying simulations that differ regarding the geometry of the *front* of their heated fluid, one being tilted to the right, the other to the left (examples of each are shown in (a)). Since merge trees encode each feature independently of its geometry, the Wasserstein distance  $W_2^T$  can not differentiate between the two simulations. This can be seen in the 2D MDS embedding of the distance matrix of  $W_2^T$  ((b), top) where the points of the two simulations overlap (the color code indicating the simulation membership). On the contrary, the region-aware Wasserstein distance  $W_2^{RT}$  can identify this specificity and correctly differentiate and separate the points of the two simulations ((b), bottom). In (c) the distances between consecutive time steps of the first simulation are shown for both  $W_2^T$  (black) and  $W_2^{RT}$  (green). The latter allows to detect important changes in the simulation (peaks in the green distance curve in (c)) that  $W_2^T$  can not. These changes correspond to the moments when the tail of the heated fluid separates from the front (first peak) and when it stops moving away to starts merging again with it (second peak).

in Fig. 9a corresponds to the best result for respectively each method (except for  $W_2^D$  that does not outperform any other method), our distance still provides visually meaningful matchings for each of them. Moreover, we quantitatively evaluate such matchings according a ground-truth of the features that should be matched together, in average our method increases by 17% the matchings accuracy on this example compared to the other methods considered.

We introduce the *temporal persistence curves*, a plot of the evolution of the persistence of the features being matched together given time. It consists of various curves, one for each feature, visually representing how its persistence evolves over time, permitting to detect when it appears, grows or disappears. Such curves act as a visual summary of an ensemble, enabling the visual identification of important changes within it in a single static view. It can help the user to select pertinent time steps to be visualized and therefore facilitate data exploration. A related but different object, the *vineyards*, has been introduced by Cohen-Steiner et al. [16]. They represent a series of persistence diagrams as 3D curves, each representing the birth and the death of features being matched together over time. Temporal persistence curves instead are 2D curves encoding the evolution of persistence, facilitating their visualization on a plane, providing a high-level and concise visual summary of a time-series for change detection and time-step selection.

The use of temporal persistence curves is illustrated in Fig. 1 with the *Isabel* ensemble (wind velocity of a hurricane), and in Fig. 8 with the *Dark Sky* ensemble (dark matter density in a cosmology simulation), with our distances  $W_2^D$  and  $W_2^{RT}$ . Since persistence diagrams and merge trees encode features regardless of their geometry, the Wasserstein distances for these objects,  $W_2^D$  and  $W_2^T$ , can produce mismatches between the features (red arrows in the red insets of Fig. 1). It prevents their correct tracking over time and impacts the subsequent interpretation of the curves. In contrast, our method correctly tracks the features over the whole ensemble and the resulting curves allow to automatically detect key changes within it.

### C. Distance matrices for dimensionality reduction

Given an ensemble, our method can be used to compute a distance matrix for all its members. Such distance matrices are then processed by typical dimensionality reduction algorithms such as MDS [46]. This method produces an embedding in  $d$  dimensions such that the distance matrix between the embedded points is as close as possible than the given distance matrix. In order to visually represent the members of an ensemble in a plane, we have chosen  $d = 2$ . Overall, such embeddings provide visual hints to the users to help to understand the relation between members of a same ensemble.

We show some examples of such embeddings, first in Fig. 10b where the classical Wasserstein distance  $W_2^T$  fails to separate the two simulations (embeddings overlap), whereas our method  $W_2^{RT}$  cleanly splits them, demonstrating sensitivity to geometric differences that topology alone cannot capture. Second, in Fig. 11 where we illustrate some examples where our method is the only one among all the methods considered that provides embeddings where the classes are correctly separated (given a ground-truth classification). We additionally show that the distance matrices generated by our method

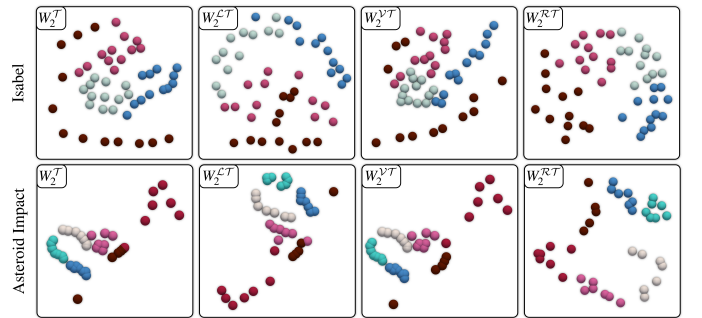


Fig. 11. Comparison of the MDS planar embeddings of all methods for two input ensembles. The color encodes the classification ground-truth. On these examples, only our method ( $W_2^{RT}$ , right column) provides embeddings where the classes are correctly separated ( $NMI = ARI = 1$ ).



TABLE II  
COMPARISON OF MDS PLANAR EMBEDDINGS QUALITY SCORES  
AVERAGED OVER *all* ENSEMBLES FOR PERSISTENCE DIAGRAMS AND  
MERGE TREES, BOLD: BEST VALUES.

| Indicator | PDs Ensembles Average |                      |                      |                                | MTs Ensembles Average |                      |                      |                                |
|-----------|-----------------------|----------------------|----------------------|--------------------------------|-----------------------|----------------------|----------------------|--------------------------------|
|           | $W_2^D$               | $W_2^{\mathcal{L}D}$ | $W_2^{\mathcal{V}D}$ | $W_2^{\mathcal{R}D}$<br>(ours) | $W_2^T$               | $W_2^{\mathcal{L}T}$ | $W_2^{\mathcal{V}T}$ | $W_2^{\mathcal{R}T}$<br>(ours) |
| NMI       | 0.66                  | 0.69                 | 0.69                 | <b>0.78</b>                    | 0.66                  | 0.69                 | 0.69                 | <b>0.79</b>                    |
| ARI       | 0.59                  | 0.63                 | 0.62                 | <b>0.72</b>                    | 0.60                  | 0.66                 | 0.64                 | <b>0.73</b>                    |

permit to detect key changes in a time-varying ensemble that the classical Wasserstein distance can not (Fig. 10c).

Tab. II compares with quantitative indicators the MDS embeddings for all the ensembles considered in this work using our method and the baselines. As indicators, we consider the normalized mutual information (NMI) [77], [84] and the adjusted rand index (ARI) [41] that evaluate how much the clusters are well preserved in the 2D embedding compared to a ground truth classification associated with each ensemble (such as the distinct phases of a time-varying phenomenon, the different simulation parameters etc.). We define these indicators in Appendix F. Tab. II averages these scores over all persistence diagrams and merge trees ensembles. We have optimized the parameters of the various methods in order to get the best scores for each of them. Specifically, we have optimized the weight  $w_{\mathcal{L}}$  for the geometrical lifting [48], [74], [93], the weight  $w_{\mathcal{V}}$  for the volume difference [69] (Appendix E) and the subsampling parameter  $\lambda$  of our method (Sec. IV-A). In practice, we use  $w_{\mathcal{L}} = 0.5$ ,  $w_{\mathcal{V}} = 0.2$  and  $\lambda = 0.1$ . The use of the coordinates distance ( $W_2^{\mathcal{L}D}$  and  $W_2^{\mathcal{L}T}$ ) [48], [74], [93] and the volume difference ( $W_2^{\mathcal{V}D}$  and  $W_2^{\mathcal{V}T}$ ) [69] slightly improve the scores compared to the classical Wasserstein distances. In contrast, our method provides a way better improvement of these scores with an approximately 21% increase compared to the classical Wasserstein distances and an approximately 16% increase compared to all the baselines.

Another example is shown in Fig. 9 where we compare the embeddings provided by the three baseline methods and ours on the cloud ensemble. We quantitatively evaluate such embeddings according a ground-truth classification, in average our method increases by 109% the scores on this example compared to the other methods considered.

## D. Framework quality

1) *Empirical stability*: We empirically evaluate the stability of the region-aware Wasserstein distance  $W_q^{\mathcal{R}D}$  in Fig. 12 using null background (Sec. III-C). Recall that in Appendix C

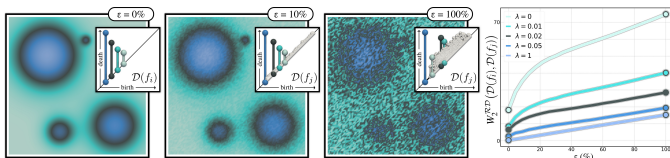


Fig. 12. Empirical stability evaluation of  $W_2^{\mathcal{R}D}$ . A noisy version version  $f_j$  of an input scalar field  $f_i$  is created by adding a noise of increasing amplitude  $\epsilon$ . The evolution of  $W_2^{\mathcal{R}D}(\mathcal{D}(f_i), \mathcal{D}(f_j))$  given  $\epsilon$  and for different values of the subsampling parameter  $\lambda$  shows no sudden spikes or drops that would be related to high instabilities, but rather a smooth evolution.

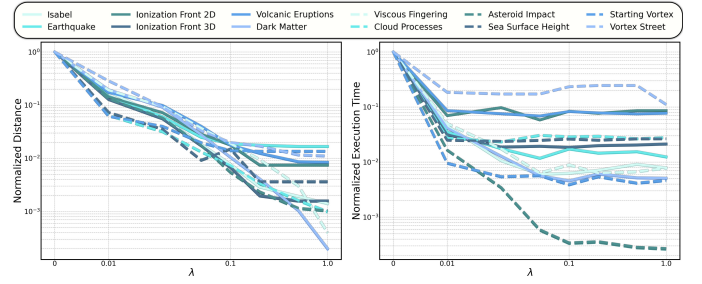


Fig. 13. Evolution of the normalized distance and execution time given the subsampling parameter  $\lambda$  for all persistence diagram ensembles.

we have proven the stability of  $W_q^{\mathcal{R}D}$  using data background (Sec. III-C) when excluding extremum-swap instabilities. Given a scalar field  $f_i$ , we create noisy variants  $f_j$  of it with a noise of increasing amplitude  $\epsilon$ , i.e.  $\|f_i - f_j\|_{\infty} \leq \epsilon$ . Then, we compute the distance  $W_2^{\mathcal{R}D}(\mathcal{D}^{\mathcal{R}}(f_i), \mathcal{D}^{\mathcal{R}}(f_j))$  and observe its evolution given  $\epsilon$  for different values of the subsampling parameter  $\lambda$ . When  $\lambda = 1$  we have  $W_2^{\mathcal{R}D}(\mathcal{D}^{\mathcal{R}}(f_i), \mathcal{D}^{\mathcal{R}}(f_j)) = W_2^D(\mathcal{D}(f_i), \mathcal{D}(f_j))$  which is proven to be stable [15] and we indeed observe its evolution to be linear (blue line at the bottom in Fig. 12). The more  $\lambda$  decreases, the more we see  $W_2^{\mathcal{R}D}$  increasing as more values of the regions are considered in the distance. In that case, we can observe that  $W_2^{\mathcal{R}D}$  increases rapidly for low noise amplitude until arriving to a similar linear slope, this shift happens for around 20% of noise amplitude for  $\lambda = 0.1$ . Overall, for the various values of  $\lambda$  we observe a smooth evolution of the distance and no sudden spikes or drops that would be related to high instabilities.

2) *Impact of subsampling*: We then evaluate in Fig. 13 how the subsampling parameter  $\lambda$  (Sec. IV-A) impacts the region-aware Wasserstein distance and its execution time. Interestingly, all normalized distance curves (one for each ensemble) follows the same kind of trend, they decrease approximately linearly as  $\lambda$  increases in the log-log plot, implying an approximately power-law relationship between the normalized distance and  $\lambda$ . Overall, this shows that there is a smooth transition between the classical Wasserstein distance at  $\lambda = 1$  and the region-aware variant at  $\lambda = 0$ , following a common pattern. As expected, the execution time decreases as  $\lambda$  increases. In that case, less values of the regions are used in the computation, naturally decreasing the execution time.

3) *Impact of compression*: In Fig. 14 we evaluate the impact of the compression of the regions on our method. We aim to better understand the benefits and drawbacks of each

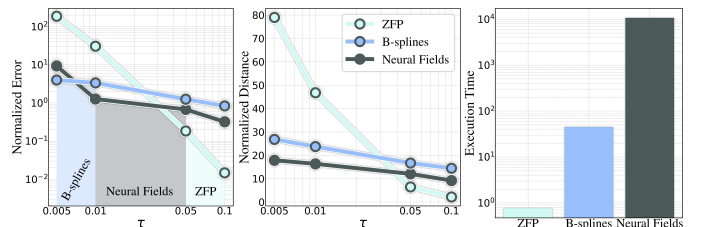


Fig. 14. Evolution of the normalized compression error (left) and distance (middle) given the compression parameter  $\tau$  averaged over *all* ensembles, and average execution time (right) for each compression method.

compression method considered in our context. We evaluate each of them for different values  $\tau \in \{0.005, 0.01, 0.05, 0.1\}$  (Sec. IV-B), ranging compression ratios from 200 : 1 to 10 : 1. We explain in Appendix D how the parameters of each method are set to give the best results. We first evaluate the compression error of each method averaged over all ensembles for each  $\tau$  value (Fig. 14, left). We observe that B-splines gave the best results for high compression ratios ( $\tau = 0.005$ ), close to neural fields that are more effective for intermediary compression ratios ( $\tau = 0.01$ ). Above these values, ZFP seems to outperform the other two methods. We then evaluate the impact on our distance  $W_2^{\mathcal{R}\mathcal{D}}$  (Fig. 14, center) by computing a distance between an original region-aware persistence diagram  $\mathcal{D}^{\mathcal{R}}(f_i)$  and the same diagram with its regions compressed. We see that the distance follows a same kind of trend than the compression error. Finally, the execution time of each method averaged over all ensembles and all  $\tau$  values (Fig. 14, right) indicates very fast compressions for ZFP, with less than 1 second in average per member, to few minutes for B-splines, and is relatively slow for neural fields (few hours).

### E. Limitations

Our method necessitates the points of extrema-aligned regions to coincide with each other, allowing their fast comparison. We have considered in this work regular grids, which fit this property, however, our method can not be directly applied to unstructured grid as it is. For this, an interpolation method would be needed, in order to determine for a point of a region which point in another region should be considered (using barycentric coordinates of neighbors for instance). This results however in more computation, increasing the running time.

Persistence diagrams and merge trees can be visualized in the plane to effectively summarize the topological features of a scalar field. It is harder to do such visualizations of the regions of these features, especially for 3D ones. In our work, we have decided to tackle this drawback by directly visualizing the features in the original scalar field (Fig. 1 and Fig. 8). We leave for future work an investigation of concise and expressive visualizations representing the features geometry.

Simply aligning the regions at their extrema can possibly induce a high cost for the matching of features that have been rotated or stretched from each other. While computing a ground metric that take into account such transformations would results in better comparisons of the regions, it would also necessitates a higher execution time.

Finally, the proven stability of  $W_q^{\mathcal{R}\mathcal{D}}$  (Sec. III-D2 and Appendix C) needs to exclude extremum-swap instabilities due to the alignment of regions to their extrema when comparing them. To tackle this, the development of a new metric that aligns regions not simply on their extrema but to minimize a mismatch term is a promising direction.

## VI. CONCLUSION

In this paper, we presented a generalization of the Wasserstein distances for persistence diagrams and merge trees [20], [66] that uses a richer characterization of the topological features and redefines their comparison as a distance between

their extrema-aligned regions. A parameter is defined to control the amount of information about the regions to use, making the original Wasserstein distances special cases of our method. We have shown some theoretical properties: it is a metric, it is stable under certain conditions and it is more discriminative than the original distances it generalizes. We have presented two strategies to control both computation time and memory footprint of the method by respectively using only subsets of the regions in the computation and to compress the regions' properties to obtain low-memory representations.

A natural direction for future work is the definition of more advanced analysis methods using this novel distance metric such as barycenters [66] (to compute a region-aware BDT representative of a set) or variance analysis methods like principal geodesic analysis [67]. An orthogonal direction than the latter consists in the improvement of the region-aware ground metric. Instead of simply aligning the regions at their extrema, they could be deformed to minimize a fitting criterion. Overall, this work is a first step towards the extension and generalization of the Wasserstein distances for persistence and merge trees by incorporating region-aware comparisons.

## REFERENCES

- [1] A. Acharya and V. Natarajan. A parallel and memory efficient algorithm for constructing the contour tree. In *IEEE PacificViz*, 2015. doi: 10.1109/PACIFICVIS.2015.7156387
- [2] K. Anderson, J. Anderson, S. Palande, and B. Wang. Topological data analysis of functional MRI connectivity in time and space domains. In *MICCAI Workshop on Connectomics in NeuroImaging*, pp. 67–77. Springer, Cham, 2018. doi: 10.1007/978-3-030-00755-3\_8
- [3] T. F. Banchoff. Critical points and curvature for embedded polyhedral surfaces. *The American Mathematical Monthly*, 45(1):245–256, 1967. doi: 10.1080/00029890.1970.11992523
- [4] U. Bauer, M. Kerber, and J. Reininghaus. Distributed computation of persistent homology. In *Algorithm Engin. and Exp.*, pp. 31–38. SIAM, Philadelphia, 2014. doi: 10.1137/1.9781611973198.4
- [5] K. Beketayev, D. Yeliussizov, D. Morozov, G. H. Weber, and B. Hamann. Measuring the distance between merge trees. In *TopoInVis*. 2014. doi: 10.1007/978-3-319-04099-8\_10
- [6] D. P. Bertsekas. A new algorithm for the assignment problem. *Mathematical Programming*, 21(1):152–171, 1981. doi: 10.1007/BF01584237
- [7] H. Bhatia, A. G. Gyulassy, V. Lordi, J. E. Pask, V. Pascucci, and P.-T. Bremer. Topoms: Comprehensive topological exploration for molecular and condensed-matter systems. *J. of Computational Chemistry*, 39(16):936–952, 2018. doi: 10.1002/JCC.25181
- [8] S. Biasotti, D. Giorgio, M. Spagnuolo, and B. Falcidieno. Reeb graphs for shape analysis and applications. *TCS*, 392(1-3):5–22, 2008. doi: 10.1016/J.TCS.2007.10.018
- [9] A. Bock, H. Doraiswamy, A. Summers, and C. T. Silva. TopoAngler: Interactive Topology-Based Extraction of Fishes. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):812–821, 2018. doi: 10.1109/TVCG.2017.2743980
- [10] P. Bremer, H. Edelsbrunner, B. Hamann, and V. Pascucci. A topological hierarchy for functions on triangulated surfaces. *IEEE Transactions on Visualization and Computer Graphics*, 10(4):385–396, 2004. doi: 10.1109/TVCG.2004.3
- [11] P. Bremer, G. Weber, J. Tierny, V. Pascucci, M. Day, and J. Bell. Interactive exploration and analysis of large scale simulations using topology-based data segmentation. *IEEE Transactions on Visualization and Computer Graphics*, 17(9):1307–1324, 2011. doi: 10.1109/TVCG.2010.253
- [12] H. Carr, J. Snoeyink, and U. Axen. Computing contour trees in all dimensions. In *Symp. on Dis. Alg.*, 9 pages, pp. 918–926. SIAM, Philadelphia, 2000. doi: 10.1016/S0925-7721(02)00093-7
- [13] H. Carr, G. Weber, C. Sewell, and J. Ahrens. Parallel peak pruning for scalable SMP contour tree computation. In *IEEE LDAV*, pp. 75–84. IEEE, Baltimore, 2016. doi: 10.1109/LDAV.2016.7874312



- [14] H. A. Carr, J. Snoeyink, and M. van de Panne. Simplifying Flexible Isosurfaces Using Local Geometric Measures. In *VIS*, pp. 497–504. IEEE, Austin, 2004. doi: 10.1109/VISUAL.2004.96
- [15] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. In *Symposium on Computational Geometry*, 2005.
- [16] D. Cohen-Steiner, H. Edelsbrunner, and D. Morozov. Vines and vineyards by updating persistence in linear time. In *Proceedings of the Twenty-Second Annual Symposium on Computational Geometry*, SCG '06, 8 pages, p. 119–126. Association for Computing Machinery, New York, NY, USA, 2006. doi: 10.1145/1137856.1137877
- [17] J. Curry, H. Hang, W. Mio, T. Needham, and O. B. Okutan. Decorated merge trees for persistent topology. *Journal of Applied and Computational Topology*, 6(3):371–428, Sep 2022. doi: 10.1007/s41468-022-00089-3
- [18] J. Curry, W. Mio, T. Needham, O. B. Okutan, and F. Russold. Stability and Approximations for Decorated Reeb Spaces. In W. Mulzer and J. M. Phillips, eds., *40th International Symposium on Computational Geometry (SoCG 2024)*, vol. 293 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 44:1–44:17. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2024. doi: 10.4230/LIPIcs.SocG.2024.44
- [19] L. De Floriani, U. Fugacci, F. Iuricich, and P. Magillo. Morse complexes for shape segmentation and homological analysis: discrete models and algorithms. *Computer Graphics Forum*, 34(2):761–785, 2015.
- [20] H. Edelsbrunner and J. Harer. *Computational Topology: An Introduction*. American Mathematical Society, 2009.
- [21] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological Persistence and Simplification. *Discrete Computational Geometry*, 28(4):511–533, 2002. doi: 10.1007/S00454-002-2885-2
- [22] H. Edelsbrunner and E. P. Mücke. Simulation of simplicity: a technique to cope with degenerate cases in geometric algorithms. *ACM Transactions on Graphics*, 9(1):66–104, 1990. doi: 10.1145/77635.77639
- [23] G. Favelier, N. Faraj, B. Summa, and J. Tierny. Persistence Atlas for Critical Point Variability in Ensembles. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):1152–1162, 2018.
- [24] G. Favelier, C. Gueunet, and J. Tierny. Visualizing ensembles of viscous fingers. In *IEEE SciVis Contest*, 2016.
- [25] D. Feng, B. Hentschel, S. Griessbach, L. Hoffmann, and M. von Hobe. The IEEE SciVis Contest. <http://scviscontest.ieeevis.org/2014/>, 2014.
- [26] C. Garth, B. Geveci, B. Hentschel, J. Kuhnert, I. Michel, T.-M. Rhyne, and S. Schröder. The IEEE SciVis Contest. <http://scviscontest.ieeevis.org/2016/>, 2016.
- [27] E. Gasparovic, E. Munch, S. Oudot, K. Turner, B. Wang, and Y. Wang. Intrinsic interleaving distance for merge trees. *La Matematica*, 4(1):40–65, Mar 2025. doi: 10.1007/s44007-024-00143-9
- [28] D. Guenther, R. Alvarez-Boto, J. Contreras-Garcia, J.-P. Piquemal, and J. Tierny. Characterizing Molecular Interactions in Chemical Systems. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2476–2485, 2014.
- [29] C. Gueunet, P. Fortin, J. Jomier, and J. Tierny. Task-Based Augmented Contour Trees with Fibonacci Heaps. *IEEE TPDS*, 30(8):1889–1905, 2019. doi: 10.1109/TPDS.2019.2898436
- [30] C. Gueunet, P. Fortin, J. Jomier, and J. Tierny. Task-based Augmented Reeb Graphs with Dynamic ST-Trees. In *EGPGV*, pp. 27–37. EG, Eindhoven, 2019. doi: 10.2312/PGV.20191107
- [31] P. Guillou, J. Vidal, and J. Tierny. Discrete Morse Sandwich: Fast Computation of Persistence Diagrams for Scalar Data – An Algorithm and A Benchmark. *IEEE Transactions on Visualization and Computer Graphics*, 30(4):1897–1915, 2023. doi: 10.1109/TVCG.2023.3238008
- [32] A. Gyulassy, P. Bremer, R. Grout, H. Kolla, J. Chen, and V. Pascucci. Stability of dissipation elements: A case study in combustion. *Computer Graphics Forum*, 33(3):51–60, 2014. doi: 10.1111/CGF.12361
- [33] A. Gyulassy, P. Bremer, and V. Pascucci. Shared-Memory Parallel Computation of Morse-Smale Complexes with Improved Accuracy. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):1183–1192, 2019. doi: 10.1109/TVCG.2018.2864848
- [34] A. Gyulassy, M. A. Duchaineau, V. Natarajan, V. Pascucci, E. Bringa, A. Higginbotham, and B. Hamann. Topologically Clean Distance Fields. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1432–1439, 2007. doi: 10.1109/TVCG.2007.70603
- [35] A. Gyulassy, D. Guenther, J. A. Levine, J. Tierny, and V. Pascucci. Conforming Morse-Smale complexes. *IEEE Transactions on Visualization and Computer Graphics*, 2014.
- [36] A. Gyulassy, A. Knoll, K. Lau, B. Wang, P. Bremer, M. Papka, L. A. Curtiss, and V. Pascucci. Interstitial and Interlayer Ion Diffusion Geometry Extraction in Graphitic Nanosphere Battery Materials. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):916–925, 2016. doi: 10.1109/TVCG.2015.2467432
- [37] A. Gyulassy, V. Natarajan, V. Pascucci, P. Bremer, and B. Hamann. A topological approach to simplification of three-dimensional scalar functions. *IEEE Transactions on Visualization and Computer Graphics*, 2006. doi: 10.1109/TVCG.2006.57
- [38] H. Freudenthal. Simplicialzerlegungen von beschränkter Flachheit. *Ann. of Math.*, 43(3):580–583, 1942. doi: 10.48550/arXiv.2302.11922
- [39] C. Heine, H. Leitte, M. Hlawitschka, F. Iuricich, L. De Floriani, G. Scheuermann, H. Hagen, and C. Garth. A survey of topology-based methods in visualization. *Computer Graphics Forum*, 35(3):643–667, 2016. doi: 10.1111/CGF.12933
- [40] B. Hentschel, B. Geveci, M. Turk, and S. Skillman. The IEEE SciVis Contest. <http://scviscontest.ieeevis.org/2015/>, 2015.
- [41] L. Hubert and P. Arabie. Comparing partitions. *Journal of Classification*, 2(1):193–218, Dec 1985. doi: 10.1007/BF01908075
- [42] H.W. Kuhn. Some combinatorial lemmas in topology. *IBM JoRD*, 4(5):518–524, 1960. doi: 10.1147/RD.45.0518
- [43] L. Kantorovich. On the translocation of masses. *AS USSR*, 1942.
- [44] J. Kasten, J. Reininghaus, I. Hotz, and H. Hege. Two-dimensional time-dependent vortex regions based on the acceleration magnitude. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2080–2087, 2011. doi: 10.1109/TVCG.2011.249
- [45] M. Kerber, D. Morozov, and A. Nigmatov. Geometry helps to compare persistence diagrams. *ACM J. of Experimental Algorithmics*, 22:1–20, 2017. doi: 10.1145/3064175
- [46] J. B. Kruskal and M. Wish. Multidimensional Scaling. In *SUPS*, 1978.
- [47] D. E. Laney, P. Bremer, A. Mascarenhas, P. Miller, and V. Pascucci. Understanding the structure of the turbulent mixing layer in hydrodynamic instabilities. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):1053–1060, 2006.
- [48] M. Li, X. Yan, L. Yan, T. Needham, and B. Wang. Flexible and probabilistic topology tracking with partial optimal transport. *IEEE Transactions on Visualization and Computer Graphics*, 31(10):7951–7969, 2025. doi: 10.1109/TVCG.2025.3561300
- [49] P. Lindstrom. Fixed-rate compressed floating-point arrays. *IEEE Transactions on Visualization and Computer Graphics*, 20, 08 2014. doi: 10.1109/TVCG.2014.2346458
- [50] Y. Lu, K. Jiang, J. A. Levine, and M. Berger. Compressive neural representations of volumetric scalar fields. *Computer Graphics Forum*, 40(3), 06 2021. doi: 10.1111/cgf.14295
- [51] J. Lukaszczuk, M. Will, F. Wetzels, G. H. Weber, and C. Garth. Extremum: Scalable augmented merge tree computation via extremum graphs. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):1085–1094, 2024. doi: 10.1109/TVCG.2023.3326526
- [52] S. Maadasamy, H. Doraiswamy, and V. Natarajan. A hybrid parallel algorithm for computing and tracking level set topology. In *2012 19th International Conference on High Performance Computing*, pp. 1–10, 2012. doi: 10.1109/HiPC.2012.6507496
- [53] D. Maljovec, B. Wang, P. Rosen, A. Alfonsi, G. Pastore, C. Rabiti, and V. Pascucci. Topology-inspired partition-based sensitivity analysis and visualization of nuclear simulations. In *IEEE PacificVis*, pp. 64–71. IEEE, Taipei, 2016. doi: 10.1109/PACIFICVIS.2016.7465252
- [54] G. Monge. Mémoire sur la théorie des déblais et des remblais. *Académie Royale des Sciences de Paris*, 1781.
- [55] D. Morozov, K. Beketayev, and G. H. Weber. Interleaving distance between merge trees. In *TopoInVis*. 2014.
- [56] J. Munkres. Algorithms for the assignment and transportation problems. *J. of SIAM*, 5(1):32–38, 1957. doi: 10.1137/0105003
- [57] F. Nauleau, F. Vivodtzev, T. Bridel-Bertomeu, H. Beaugendre, and J. Tierny. Topological Analysis of Ensembles of Hydrodynamic Turbulent Flows An Experimental Study. In *2022 IEEE 12th Symposium on Large Data Analysis and Visualization (LDAV)*, pp. 1–11, Oct. 2022. doi: 10.1109/LDAV57265.2022.9966403
- [58] M. Olejniczak, A. S. P. Gomes, and J. Tierny. A Topological Data Analysis Perspective on Non-Covalent Interactions in Relativistic Calculations. *IJQC*, 120(8):e26133, 2019. doi: 10.1002/qua.26133
- [59] K. Olsen, S. Day, B. Minster, R. Moore, Y. Cui, A. Chourasia, M. Thiebaux, H. Francoeur, P. Maechling, S. Cutchin, and K. Nunes. The IEEE SciVis Contest. <http://scviscontest.ieeevis.org/2006/>, 2006.
- [60] S. Parsa. A deterministic  $O(m \log m)$  time algorithm for the reeb graph. In *Proceedings of the Twenty-Eighth Annual Symposium on Computational Geometry*, SoCG '12, 8 pages, p. 269–276. Association for Computing Machinery, New York, NY, USA, 2012. doi: 10.1145/2261250.2261289
- [61] V. Pascucci, K. Cole-McLaughlin, and G. Scorzelli. Multi-resolution computation and presentation of contour trees. In *IASTED*, 2004.

- [62] V. Pascucci, G. Scorzelli, P. T. Bremer, and A. Mascarenhas. Robust on-line computation of Reeb graphs: simplicity and speed. *ACM Transactions on Graphics*, 26(3):58, 2007. doi: 10.1145/1276377.1276449
- [63] J. Patchett and G. R. Gislis. The IEEE SciVis Contest. <http://sciviscontest.ieeevis.org/2018/>, 2018.
- [64] T. Peterka, Y. S. G. Nashed, I. Grindeanu, V. S. Mahadevan, R. Yeh, and X. Tricoche. Foundations of multivariate functional approximation for scientific data. In *2018 IEEE 8th Symposium on Large Data Analysis and Visualization (LDAV)*, pp. 61–71, 2018. doi: 10.1109/LDAV.2018.8739195
- [65] M. Pont and J. Tierny. Wasserstein Auto-Encoders of Merge Trees (and Persistence Diagrams). *IEEE Transactions on Visualization and Computer Graphics*, 30(9):6390–6406, 2024. doi: 10.1109/TVCG.2023.3334755
- [66] M. Pont, J. Vidal, J. Delon, and J. Tierny. Wasserstein distances, geodesics and barycenters of merge trees. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):291–301, 2022. doi: 10.1109/TVCG.2021.3114839
- [67] M. Pont, J. Vidal, and J. Tierny. Principal geodesic analysis of merge trees (and persistence diagrams). *IEEE Transactions on Visualization and Computer Graphics*, 29(2):1573–1589, 2023. doi: 10.1109/TVCG.2022.3215001
- [68] G. Reeb. Sur les points singuliers d’une forme de Pfaff complètement intégrable ou d’une fonction numérique. *Comptes Rendus des séances de l’Académie des sciences*, 222(847-849):76, 1946.
- [69] H. Saikia, H.-P. Seidel, and T. Weinkauff. Extended branch decomposition graphs: Structural comparison of scalar data. *Computer Graphics Forum*, 33(3):41–50, 2014. doi: 10.1111/cgf.12360
- [70] N. Shivashankar and V. Natarajan. Parallel Computation of 3D Morse-Smale Complexes. *Computer Graphics Forum*, 31(3):965–974, 2012. doi: 10.1111/J.1467-8659.2012.03089.X
- [71] N. Shivashankar, P. Pranav, V. Natarajan, R. van de Weygaert, E. P. Bos, and S. Rieder. Felix: A topology based framework for visual exploration of cosmic filaments. *IEEE Transactions on Visualization and Computer Graphics*, 22(6):1745–1759, 2016. doi: 10.1109/TVCG.2015.2452919
- [72] K. Sisouk, J. Delon, and J. Tierny. Wasserstein Dictionaries of Persistence Diagrams. *IEEE Transactions on Visualization & Computer Graphics*, 30(02):1638–1651, Feb. 2024. doi: 10.1109/TVCG.2023.3330262
- [73] P. Skrabar and K. Turner. Wasserstein stability for persistence diagrams. *arXiv: Algebraic Topology*, 2020.
- [74] M. Soler, M. Plainchault, B. Conche, and J. Tierny. Lifted Wasserstein Matcher for Fast and Robust Topology Tracking. In *2018 IEEE 8th Symposium on Large Data Analysis and Visualization (LDAV)*, pp. 23–33, Oct. 2018. doi: 10.1109/LDAV.2018.8739196
- [75] T. Sousbie. The Persistent Cosmic Web and its Filamentary Structure: Theory and Implementations. *Royal Astronomical Society*, 414:384–403, 2011. doi: 10.1111/j.1365-2966.2011.18394.x
- [76] R. Sridharamurthy, T. B. Masood, A. Kamakshidasan, and V. Natarajan. Edit distance between merge trees. *IEEE Transactions on Visualization and Computer Graphics*, 26(3):1518–1531, 2020. doi: 10.1109/TVCG.2018.2873612
- [77] A. Strehl and J. Ghosh. Cluster ensembles — a knowledge reuse framework for combining multiple partitions. *J. Mach. Learn. Res.*, 3(null):583–617, 35 pages, Mar. 2003. doi: 10.1162/153244303321897735
- [78] S. P. Tarasov and M. N. Vyalyi. Construction of contour trees in 3d in  $(n \log n)$  steps. In *Proceedings of the Fourteenth Annual Symposium on Computational Geometry*, SCG ’98, 8 pages, p. 68–75. Association for Computing Machinery, New York, NY, USA, 1998. doi: 10.1145/276884.276892
- [79] R. Taylor, A. Chourasia, D. Whalen, and M. L. Norman. The IEEE SciVis Contest. <http://sciviscontest.ieeevis.org/2008/>, 2008.
- [80] D. M. Thomas and V. Natarajan. Detecting symmetry in scalar fields using augmented extremum graphs. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2663–2672, 2013. doi: 10.1109/TVCG.2013.148
- [81] J. Tierny, G. Favelier, J. A. Levine, C. Gueunet, and M. Michaux. The Topology ToolKit. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):832–842, 2017. doi: 10.1109/TVCG.2017.2743938
- [82] K. Turner, Y. Mileyko, S. Mukherjee, and J. Harer. Fréchet Means for Distributions of Persistence Diagrams. *Discrete Computational Geometry*, 52(1):44–70, 2014. doi: 10.1007/s00454-014-9604-7
- [83] J. Vidal, J. Budin, and J. Tierny. Progressive Wasserstein Barycenters of Persistence Diagrams. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):151–161, 2020. doi: 10.1109/TVCG.2019.2934256
- [84] N. X. Vinh, J. Epps, and J. Bailey. Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *J. Mach. Learn. Res.*, 11:2837–2854, 18 pages, Dec. 2010.
- [85] W. Wang, C. Bruyere, B. Kuo, and T. Scheitlin. The IEEE SciVis Contest. <http://sciviscontest.ieeevis.org/2004/>, 2004.
- [86] F. Wetzel, M. Anders, and C. Garth. Taming horizontal instability in merge trees: On the computation of a comprehensive deformation-based edit distance. In *2023 Topological Data Analysis and Visualization (TopoInVis)*, pp. 82–92, 2023. doi: 10.1109/TopoInVis60193.2023.00015
- [87] F. Wetzel and C. Garth. A deformation-based edit distance for merge trees. In *2022 Topological Data Analysis and Visualization (TopoInVis)*, pp. 29–38, 2022. doi: 10.1109/TopoInVis57755.2022.00010
- [88] F. Wetzel and C. Garth. Categorizing merge tree edit distances by stability using minimal vertex perturbation, 2025.
- [89] F. Wetzel, H. Leitte, and C. Garth. Branch decomposition-independent edit distances for merge trees. *Computer Graphics Forum*, 41(3):367–378, June 2022. doi: 10.1111/cgf.14547
- [90] F. Wetzel, H. Leitte, and C. Garth. Accelerating computation of stable merge tree edit distances using parameterized heuristics. *IEEE Transactions on Visualization and Computer Graphics*, 31(6):3706–3718, 2025. doi: 10.1109/TVCG.2025.3567120
- [91] F. Wetzel, M. Pont, J. Tierny, and C. Garth. Merge Tree Geodesics and Barycenters with Path Mappings. *IEEE Transactions on Visualization & Computer Graphics*, 30(01):1095–1105, Jan. 2024. doi: 10.1109/TVCG.2023.3326601
- [92] T. Wischgoll, A. Chourasia, K. Gorges, M. Bruck, and N. Rober. The IEEE SciVis Contest. <http://sciviscontest.ieeevis.org/2017/>, 2017.
- [93] L. Yan, T. B. Masood, F. Rasheed, I. Hotz, and B. Wang. Geometry-Aware Merge Tree Comparisons for Time-Varying Data With Interleaving Distances. *IEEE Transactions on Visualization & Computer Graphics*, 29(08):3489–3506, Aug. 2023. doi: 10.1109/TVCG.2022.3163349
- [94] L. Yan, T. B. Masood, R. Sridharamurthy, F. Rasheed, V. Natarajan, I. Hotz, and B. Wang. Scalar field comparison with topological descriptors: Properties and applications for scientific visualization. *Computer Graphics Forum*, 40(3):599–633, 2021. doi: 10.1111/cgf.14331
- [95] L. Yan, Y. Wang, E. Munch, E. Gasparovic, and B. Wang. A structural average of labeled merge trees for uncertainty visualization. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):832–842, 2020. doi: 10.1109/TVCG.2019.2934242
- [96] K. Zhang. A constrained edit distance between unordered labeled trees. *Algorithmica*, 15(3):205–222, 18 pages, Mar. 1996. doi: 10.1007/BF01975866
- [97] A. J. Zomorodian. Topology for computing. In M. J. Atallah and M. Blanton, eds., *Algorithms and Theory of Computation Handbook (Second Edition)*, chap. 3, pp. 82–112. CRC Press, Boca Raton, 2010. doi: 10.1017/CBO9780511546945

## VII. BIOGRAPHY SECTION



**Mathieu Pont** received the Ph.D. degree in Computer Science from Sorbonne Université in 2023. He has now a post-doctoral position at the RPTU Kaiserslautern-Landau. He is an active contributor to the Topology ToolKit (TTK), an open source library for topological data analysis. His notable contributions to TTK include distances, geodesics and barycenters of merge trees, for feature tracking and ensemble clustering.



**Christoph Garth** received the PhD degree in computer science from Technische Universität (TU) Kaiserslautern in 2007. After four years as a post-doctoral researcher with the University of California, Davis, he rejoined TU Kaiserslautern where he is currently a full professor of computer science. His research interests include large scale data analysis and visualization, in situ visualization, topology-based methods, and interdisciplinary applications of visualization.

Given a region-aware persistence pair  $b_i^{\mathcal{R}} = (\mathcal{R}_i, f_{|\mathcal{R}_i}, s_i)$ , let  $\mathcal{R}'_i$  the centered region at the extremum of  $b_i^{\mathcal{R}}$  and  $f'_{|\mathcal{R}'_i}$  its function. We define  $\tilde{f}_i$  as the function  $f'_{|\mathcal{R}'_i}$  extended with the background field  $\delta_f$  outside  $\mathcal{R}'_i$ . We call  $\tilde{f}_i$  the *extended* function, defined as:

$$\tilde{f}_i(v) = \begin{cases} f'_{|\mathcal{R}'_i}(v) & \text{if } v \in \mathcal{R}'_i \\ \delta_f(v) & \text{if } v \notin \mathcal{R}'_i. \end{cases} \quad (14)$$

We adopt the convention that whenever two region-aware persistence pairs  $b_i^{\mathcal{R}} = (\mathcal{R}_i, f_{|\mathcal{R}_i}, s_i) \in \mathcal{B}^{\mathcal{R}}(f_1)$  and  $b_j^{\mathcal{R}} = (\mathcal{R}_j, f_{|\mathcal{R}_j}, s_j) \in \mathcal{B}^{\mathcal{R}}(f_2)$  are compared, their extended functions are understood to be defined on the common domain  $\Omega = \mathcal{R}'_i \cup \mathcal{R}'_j$ . All equalities  $\tilde{f}_i = \tilde{f}_j$  are to be interpreted on  $\Omega$ . In this section we show that  $d_q^{\mathcal{R}}$  is a metric for the space of region-aware persistence pairs with the equivalence relation stating that  $b_i^{\mathcal{R}} \sim b_j^{\mathcal{R}} \iff s_i = s_j$  and  $\tilde{f}_i = \tilde{f}_j$  on  $\Omega = \mathcal{R}'_i \cup \mathcal{R}'_j$ .

**(i) Non-negativity:** each term of the definition of  $d_q^{\mathcal{R}}$  is non-negative, each being an absolute value raising to a power of  $q$ . The  $q^{\text{th}}$  root (with  $q \geq 1$ ) of the sum of non-negative numbers is non-negative therefore  $d_q^{\mathcal{R}}(b_i^{\mathcal{R}}, b_j^{\mathcal{R}}) \geq 0$ .

**(ii) Identity of indiscernibles:** as stated before, we identify region-aware persistence pairs modulo  $\sim$ , we write  $b_i^{\mathcal{R}} = b_j^{\mathcal{R}}$  to mean  $b_i^{\mathcal{R}} \sim b_j^{\mathcal{R}}$  and we keep the same symbol  $d_q^{\mathcal{R}}$  for the induced metric on the quotient. We want to show that two points are indistinguishable in distance if and only if they are actually the same, i.e.  $d_q^{\mathcal{R}}(b_i^{\mathcal{R}}, b_j^{\mathcal{R}}) = 0 \iff b_i^{\mathcal{R}} = b_j^{\mathcal{R}}$ .

First, we show that if  $b_i^{\mathcal{R}} = b_j^{\mathcal{R}}$  then  $d_q^{\mathcal{R}}(b_i^{\mathcal{R}}, b_j^{\mathcal{R}}) = 0$ . If two pairs are identical then their saddle values satisfy  $s_i = s_j$  therefore  $|s_i - s_j|^q = 0$ . The functions of their aligned regions  $f'_{1|\mathcal{R}'_i}(v_i)$  and  $f'_{2|\mathcal{R}'_j}(v_i)$  coincide on the intersection making  $|f'_{1|\mathcal{R}'_i}(v_i) - f'_{2|\mathcal{R}'_j}(v_i)|^q = 0$  for all  $v$ . Moreover, since the regions are identical, there are no points outside the intersection domain, i.e.  $\mathcal{R}'_i \setminus \mathcal{R}'_j = \mathcal{R}'_j \setminus \mathcal{R}'_i = \emptyset$ .

Second, we want to show that if  $d_q^{\mathcal{R}}(b_i^{\mathcal{R}}, b_j^{\mathcal{R}}) = 0$  then  $b_i^{\mathcal{R}} = b_j^{\mathcal{R}}$ . The term  $|s_i - s_j|^q = 0$  forces  $s_i = s_j$ . The sum over the intersection domain  $\mathcal{R}'_i \cap \mathcal{R}'_j$  being 0 implies that for every point  $v \in \mathcal{R}'_i \cap \mathcal{R}'_j$  then  $f'_{1|\mathcal{R}'_i}(v) = f'_{2|\mathcal{R}'_j}(v)$ . For any point  $x \in \mathcal{R}'_i \setminus \mathcal{R}'_j$ , the deletion cost  $|f'_{1|\mathcal{R}'_i}(x) - \delta_{f_2}(x)|^q = 0$  forces  $f'_{1|\mathcal{R}'_i}(x) = \delta_{f_2}(x)$ . Similarly, for any point  $y \in \mathcal{R}'_j \setminus \mathcal{R}'_i$ , the function  $f'_{2|\mathcal{R}'_j}(y)$  must be equal to  $\delta_{f_1}(y)$ . By considering the extended functions  $\tilde{f}_i$  and  $\tilde{f}_j$  (Eq. 14), we see that the two functions agree everywhere on the combined domain. Combined with the equality  $s_i = s_j$  and the fact that the alignment procedure makes the comparison of regions well-defined, we conclude that the two region-aware persistence pairs are identical.

Therefore,  $d_q^{\mathcal{R}}(b_i^{\mathcal{R}}, b_j^{\mathcal{R}}) = 0 \iff b_i^{\mathcal{R}} = b_j^{\mathcal{R}}$ .

**(iii) Symmetry:** The absolute difference  $|s_i - s_j|$  is symmetric. The intersection  $\mathcal{R}'_i \cap \mathcal{R}'_j$  is symmetric with respect to the two regions. The deletion cost for  $b_i^{\mathcal{R}}$  relative to  $b_j^{\mathcal{R}}$  and the insertion cost for  $b_j^{\mathcal{R}}$  relative to  $b_i^{\mathcal{R}}$  swap roles if we reverse the order. In other words, when we exchange  $b_i^{\mathcal{R}}$  and  $b_j^{\mathcal{R}}$ , the sums in lines for deletion and insertion exchange but their contributions remain the same. Thus,  $d_q^{\mathcal{R}}(b_i^{\mathcal{R}}, b_j^{\mathcal{R}}) = d_q^{\mathcal{R}}(b_j^{\mathcal{R}}, b_i^{\mathcal{R}})$ .

## APPENDIX A

### METRIC PROOF USING COMMON BACKGROUND FIELD

We show in this section that the region-aware ground distance  $d_q^{\mathcal{R}}$  proposed in this work is indeed a metric and that therefore the region-aware Wasserstein distance  $W_q^{\mathcal{R}, \mathcal{T}}$  (and  $W_q^{\mathcal{R}, \mathcal{D}}$ ) also is.

We first recall the expression of  $d_q^{\mathcal{R}}$  (with  $q \geq 1$ ):

$$d_q^{\mathcal{R}}(b_i^{\mathcal{R}}, b_j^{\mathcal{R}}) = (|s_i - s_j|^q + C_q^{\mathcal{R}}(b_i^{\mathcal{R}}, b_j^{\mathcal{R}}))^{1/q}, \quad (10)$$

with  $C_q^{\mathcal{R}}$ , the regional discrepancy, defined as:

$$C_q^{\mathcal{R}}(b_i^{\mathcal{R}}, b_j^{\mathcal{R}}) = \sum_{v \in \mathcal{R}'_i \cap \mathcal{R}'_j} |f'_{1|\mathcal{R}'_i}(v) - f'_{2|\mathcal{R}'_j}(v)|^q \quad (11)$$

$$+ \sum_{x \in \mathcal{R}'_i \setminus \mathcal{R}'_j} |f'_{1|\mathcal{R}'_i}(x) - \delta_{f_2}(x)|^q \quad (12)$$

$$+ \sum_{y \in \mathcal{R}'_j \setminus \mathcal{R}'_i} |\delta_{f_1}(y) - f'_{2|\mathcal{R}'_j}(y)|^q, \quad (13)$$

where  $b_i^{\mathcal{R}} = (\mathcal{R}_i, f_{|\mathcal{R}_i}, s_i) \in \mathcal{B}^{\mathcal{R}}(f_1)$  and  $b_j^{\mathcal{R}} = (\mathcal{R}_j, f_{|\mathcal{R}_j}, s_j) \in \mathcal{B}^{\mathcal{R}}(f_2)$  are two region-aware persistence pairs in two BDTs.  $\mathcal{R}'_i$  and  $\mathcal{R}'_j$  are the centered regions at the respective extremum of  $b_i^{\mathcal{R}}$  and  $b_j^{\mathcal{R}}$ , and  $f'_{1|\mathcal{R}'_i}$  and  $f'_{2|\mathcal{R}'_j}$  their functions.  $\delta_{f_1}$  and  $\delta_{f_2}$  are the background fields against which the values outside the intersection domain are compared. We will restrict the explanation to common background field, i.e.  $\delta_{f_1} = \delta_{f_2}$ , such as the null background (Sec. III-C1) and show the general case in Appendix B.

We now want to show that it is a metric. For this, we need to show four properties: (i) non-negativity, (ii) identity of indiscernibles, (iii) symmetry and (iv) triangle inequality.

(iv) **Triangle inequality:** Let  $b_i^{\mathcal{R}}$ ,  $b_j^{\mathcal{R}}$  and  $b_k^{\mathcal{R}}$  be three region-aware persistence pairs. We want to show that the direct path between two points is never longer than going through an intermediate point, i.e.  $d_q^{\mathcal{R}}(b_i^{\mathcal{R}}, b_k^{\mathcal{R}}) \leq d_q^{\mathcal{R}}(b_i^{\mathcal{R}}, b_j^{\mathcal{R}}) + d_q^{\mathcal{R}}(b_j^{\mathcal{R}}, b_k^{\mathcal{R}})$ . For this, we will build on the definition of extended functions (Eq. 14) but this time for three region-aware persistence pairs, i.e. we define the three extended functions  $\tilde{f}_i$ ,  $\tilde{f}_j$  and  $\tilde{f}_k$  like in Eq. 14 but this time on  $\Omega = \mathcal{R}_i' \cup \mathcal{R}_j' \cup \mathcal{R}_k'$ . Next, we can use the  $L_q(\Omega)$  norm to measure the difference between these extended functions:

$$\|\tilde{f}_i - \tilde{f}_j\|_{L_q(\Omega)}^q = \sum_{v \in \Omega} |\tilde{f}_i(v) - \tilde{f}_j(v)|^q, \quad (15)$$

allowing to rewrite the region-aware ground distance as:

$$d_q^{\mathcal{R}}(b_i^{\mathcal{R}}, b_j^{\mathcal{R}})^q = |s_i - s_j|^q + \|\tilde{f}_i - \tilde{f}_j\|_{L_q(\Omega)}^q. \quad (16)$$

The  $L_q$  norm (for  $q \geq 1$ ) satisfies the triangle inequality by Minkowski's inequality:  $\|\tilde{f}_i - \tilde{f}_k\|_{L_q(\Omega)} \leq \|\tilde{f}_i - \tilde{f}_j\|_{L_q(\Omega)} + \|\tilde{f}_j - \tilde{f}_k\|_{L_q(\Omega)}$ . Similarly, the absolute difference on the real numbers satisfies the triangle inequality:  $|s_i - s_k| \leq |s_i - s_j| + |s_j - s_k|$ . This separation of the contribution of both terms leads to the definition of a vector in  $\mathbb{R}^2$ :

$$v_{ij} = (|s_i - s_j|, \|\tilde{f}_i - \tilde{f}_j\|_{L_q(\Omega)}), \quad (17)$$

then, the overall distance is simply the  $L_q$  norm of this vector:

$$d_q^{\mathcal{R}}(b_i^{\mathcal{R}}, b_j^{\mathcal{R}}) = \|v_{ij}\| = (|s_i - s_j|^q + \|\tilde{f}_i - \tilde{f}_j\|_{L_q(\Omega)}^q)^{1/q}. \quad (18)$$

Using the Minkowski inequality for the  $L_q$  norm in  $\mathbb{R}^2$ , we know that  $\|v_{ik}\| \leq \|v_{ij}\| + \|v_{jk}\|$ . Translating back, this inequality becomes:

$$\begin{aligned} (|s_i - s_k|^q + \|\tilde{f}_i - \tilde{f}_k\|_{L_q(\Omega)}^q)^{1/q} &\leq (|s_i - s_j|^q + \|\tilde{f}_i - \tilde{f}_j\|_{L_q(\Omega)}^q)^{1/q} \\ &\quad + (|s_j - s_k|^q + \|\tilde{f}_j - \tilde{f}_k\|_{L_q(\Omega)}^q)^{1/q}, \end{aligned}$$

using the definition of  $d_q^{\mathcal{R}}$  in Eq. 18, it directly implies the triangle inequality for  $d_q^{\mathcal{R}}$ , i.e.  $d_q^{\mathcal{R}}(b_i^{\mathcal{R}}, b_k^{\mathcal{R}}) \leq d_q^{\mathcal{R}}(b_i^{\mathcal{R}}, b_j^{\mathcal{R}}) + d_q^{\mathcal{R}}(b_j^{\mathcal{R}}, b_k^{\mathcal{R}})$ .

Finally, we want to show that the region-aware Wasserstein distance  $W_q^{\mathcal{RT}}$  (and its special case for persistence diagrams  $W_q^{\mathcal{RD}}$ ) is a metric. It consists of the Wasserstein distance between merge trees  $W_2^{\mathcal{T}}$  using the region-aware ground metric  $d_q^{\mathcal{R}}(b_i^{\mathcal{R}}, b_j^{\mathcal{R}})$ . In their metric proof, Pont et al. [66] shows that their distance is indeed a metric if the ground distance is also a metric. Specifically, the mapping edit cost  $\gamma(b_i \rightarrow \phi'(b_i))$  in their proof must be replaced by  $d_q^{\mathcal{R}}(b_i^{\mathcal{R}}, b_j^{\mathcal{R}})$ , and the deletion and insertion costs,  $\gamma(b_i \rightarrow \emptyset)$  and  $\gamma(\emptyset \rightarrow b_j)$ , by respectively  $d_q^{\mathcal{R}}(b_i^{\mathcal{R}}, \Delta^{\mathcal{R}}(b_i^{\mathcal{R}}))$  and  $d_q^{\mathcal{R}}(\Delta^{\mathcal{R}}(b_j^{\mathcal{R}}), b_j^{\mathcal{R}})$ . Therefore  $W_q^{\mathcal{RT}}$  (and the special case  $W_q^{\mathcal{RD}}$ ) is a metric.

## APPENDIX B

### METRIC PROOF USING ANY BACKGROUND FIELD

In this section, we extend the proof of Appendix A for the use of non-common background fields, i.e.  $\delta_{f_1} \neq \delta_{f_2}$  given two region-aware BDTs  $\mathcal{B}^{\mathcal{R}}(f_1)$  and  $\mathcal{B}^{\mathcal{R}}(f_2)$ . In that case the triangle inequality shown in Appendix A can break. Specifically, Eq. 16 is only valid for common background. Indeed, recall that in Eq. 16 we have  $\Omega = \mathcal{R}_i' \cup \mathcal{R}_j' \cup \mathcal{R}_k'$  and rewriting the ground distance (Eq. 10) as Eq. 16 necessitates that for every  $x \in \Omega' = \Omega \setminus (\mathcal{R}_i' \cup \mathcal{R}_j')$  we have  $\|\tilde{f}_i - \tilde{f}_j\|_{L_q(\Omega')}^q = 0$  which is not the case if  $\delta_{f_1} \neq \delta_{f_2}$ .

(i) **Triangle inequality counter-example in the base case:** Considering the data background fields (i.e.  $\delta_f(x) = f(x)$ ) and three scalar fields  $f = [m, a, \varepsilon, b, 2\varepsilon, 0, d]$ ,  $g = [m, a, \varepsilon, b, 0, 2\varepsilon, d]$ ,  $h = [m, a, \varepsilon, b, -c, 2\varepsilon, d]$ , with  $m < -c < 0 < \varepsilon < b < d < a$  and  $\varepsilon$  very small, near 0. We also consider  $q = 1$  for simplicity but the example works for every valid values of  $q$ . There are three regions with same saddle values and with region values  $f|_{\mathcal{R}_1} = [b, 2\varepsilon]$ ,  $g|_{\mathcal{R}_2} = [b]$  and  $h|_{\mathcal{R}_3} = [b]$ , let respectively  $b_f^{\mathcal{R}}$ ,  $b_g^{\mathcal{R}}$  and  $b_h^{\mathcal{R}}$  the region-aware persistence pairs associated with these region values. When comparing  $b_f^{\mathcal{R}}$  to  $b_h^{\mathcal{R}}$ , the cost is  $|2\varepsilon + c|$  (the vertex having value  $2\varepsilon$  in  $f$  is part of the symmetric difference and is compared, due to the use of the data background fields, to  $-c$  in  $h$ ). However, from  $b_f^{\mathcal{R}}$  to  $b_g^{\mathcal{R}}$  we have a cost of  $2\varepsilon$  and for  $b_g^{\mathcal{R}}$  to  $b_h^{\mathcal{R}}$  a cost of 0. Since  $|2\varepsilon + c|$  can be arbitrarily large compared to  $2\varepsilon$  we have  $d_1^{\mathcal{R}}(b_f^{\mathcal{R}}, b_h^{\mathcal{R}}) > d_1^{\mathcal{R}}(b_f^{\mathcal{R}}, b_g^{\mathcal{R}}) + d_1^{\mathcal{R}}(b_g^{\mathcal{R}}, b_h^{\mathcal{R}})$ .

(ii) **Triangle inequality for any background field:** Since the region-aware ground distance uses a partner-dependent domain, it can therefore breaks the triangle inequality when the values outside this common domain do not coincide (i.e. for non-common background fields). Specifically, let three region-aware persistence pairs  $b_f^{\mathcal{R}}$ ,  $b_g^{\mathcal{R}}$  and  $b_h^{\mathcal{R}}$  and their respective regions  $\mathcal{R}_1$ ,  $\mathcal{R}_2$  and  $\mathcal{R}_3$ . When comparing  $b_f^{\mathcal{R}}$  to  $b_h^{\mathcal{R}}$  the domain used is  $\mathcal{R}_1 \cup \mathcal{R}_3$ , while being  $\mathcal{R}_1 \cup \mathcal{R}_2$  for  $b_f^{\mathcal{R}}$  to  $b_g^{\mathcal{R}}$  and  $\mathcal{R}_2 \cup \mathcal{R}_3$  for  $b_g^{\mathcal{R}}$  to  $b_h^{\mathcal{R}}$ . For a vertex  $v \in \mathcal{R}_1 \cup \mathcal{R}_3$ , it is possible that either  $x \notin \mathcal{R}_1 \cup \mathcal{R}_2$  or  $x \notin \mathcal{R}_2 \cup \mathcal{R}_3$  (but not both), therefore, the Minkowski's inequality can not be applied.

To solve this, we redefine the comparison of two regions to be done on a fixed-size domain  $\Lambda$ . Let a region-aware persistence pair  $b_i^{\mathcal{R}} = (\mathcal{R}_i, f|_{\mathcal{R}_i}, s_i) \in \mathcal{B}^{\mathcal{R}}(f)$  for a scalar field  $f : \mathcal{M} \rightarrow \mathbb{R}$ . We rewrite its extended function (Eq. 14) as:

$$\tilde{f}_i(v) = \begin{cases} f|_{\mathcal{R}_i'}(v) & \text{if } v \in \mathcal{R}_i' \\ \delta_f(v) & \text{if } v \notin \mathcal{R}_i' \text{ and } v \in \mathcal{M}' \\ 0 & \text{if } v \notin \mathcal{M}', \end{cases} \quad (19)$$

with  $\tilde{f}_i : \Lambda \rightarrow \mathbb{R}$  and  $\mathcal{M}'$  the full domain aligned at the extremum of  $b_i^{\mathcal{R}}$ . Note the need of defining a default value for vertices not in the  $\mathcal{M}'$  (third line). Indeed, some vertices of the common fixed-size domain  $\Lambda$  around the extremum could possibly not belong to  $\mathcal{M}'$  and not be an preimage under  $\delta_f$ .

Additionally, since  $\Lambda$  can be arbitrarily large (possibly being the full domain in practice), vertices being far from a region can be part of  $\Lambda$  when centering the region to its extremum. In order to restore a notion of locality about the regions, the values of the extended function can be weighted (such as a



decaying weight given the distance to the extremum), giving the weighted extended function  $\Psi_i^f(v) = w_i(v)\tilde{f}_i(v)$ . Moreover, to avoid the multiple contribution of a vertex (being part of many  $\Lambda$  after aligning various regions at their extremum), the weights of a same vertex  $v$  can be normalized to enforce a partition of unity such that  $\sum_i w_i(v) = 1$  for all regions  $\mathcal{R}_i$ .

Let  $b_i^{\mathcal{R}} = (\mathcal{R}_i, f_1|_{\mathcal{R}_i}, s_i) \in \mathcal{B}^{\mathcal{R}}(f_1)$  and  $b_j^{\mathcal{R}} = (\mathcal{R}_j, f_2|_{\mathcal{R}_j}, s_j) \in \mathcal{B}^{\mathcal{R}}(f_2)$  two region-aware persistence pairs and  $\Psi_i^{f_1}$  the weighted extended function for the pair  $i$  given  $f_1$  (same for  $\Psi_j^{f_2}$ ). We can rewrite Eq. 11 as  $C_q^{\mathcal{R}}(b_i^{\mathcal{R}}, b_j^{\mathcal{R}}) = \|\Psi_i^{f_1} - \Psi_j^{f_2}\|_{L_q(\Lambda)}^q$  and the region-aware ground metric (Eq. 10) as:

$$d_q^{\mathcal{R}}(b_i^{\mathcal{R}}, b_j^{\mathcal{R}}) = (|s_i - s_j|^q + \|\Psi_i^{f_1} - \Psi_j^{f_2}\|_{L_q(\Lambda)}^q)^{1/q}. \quad (20)$$

From this, the triangle inequality can be proven in the same way than in Appendix A starting from Eq. 17. Moreover, non-negativity, identity of indiscernibles and symmetry can be proven in the same way than in Appendix A.

#### APPENDIX C STABILITY OF $W_q^{\mathcal{RD}}$

In this section we study the stability of the region-aware Wasserstein distance for persistence diagrams  $W_q^{\mathcal{RD}}$ . We manage to prove, by excluding extremum-swap instabilities and by using data background, that  $W_q^{\mathcal{RD}}$  can be bounded by above by the  $L_q$  norm of the difference of the input scalar fields multiplied by a constant independent on the input data, specifically  $W_q^{\mathcal{RD}}(\mathcal{D}^{\mathcal{R}}(f), \mathcal{D}^{\mathcal{R}}(g)) \leq 2^{1/q} \|f - g\|_q$ .

For this, we take inspiration of the proof by Skraba and Turner [73] for the stability of the original Wasserstein between persistence diagrams  $W_q^{\mathcal{D}}$ , that we extend to take into account the region properties. The main idea of their proof is to consider a straight line homotopy between the input scalar fields  $f$  and  $g$  and to partition it into finitely many sub-intervals over which the vertex ordering induced by the scalar function remains unchanged. In this way, they establish a local result on each sub-interval that can be combined together to arrive at the final inequality  $W_q^{\mathcal{D}}(\mathcal{D}(f), \mathcal{D}(g)) \leq \|f - g\|_q$ .

##### A. Homotopy and partition

Let  $f : \mathcal{M} \rightarrow \mathbb{R}$  and  $g : \mathcal{M} \rightarrow \mathbb{R}$ , two input scalar fields defined on the vertices of a simplicial complex  $\mathcal{M}$  and linearly interpolated on the simplices of higher dimension. Without loss of generality, they are considered to be injective on their vertices using symbolic perturbations [22] that break ties. Let respectively  $\mathcal{D}^{\mathcal{R}}(f)$  and  $\mathcal{D}^{\mathcal{R}}(g)$  their region-aware persistence diagrams computed using the lower-star filtration. Let a straight line homotopy between the input scalar fields  $f$  and  $g$  be  $h_t = (1-t)f + tg$  with  $t \in [0, 1]$ , i.e. their linear interpolation as  $t$  varies. There are finitely many times  $0 < t_1 < t_2 < \dots < t_n < 1$  at which two vertex values become equal, i.e. the order changes by a *vertex swap*. Specifically, let  $t_0 = 0$  and  $t_{n+1} = 1$ , for each  $t_k$  with  $k \in \{1, \dots, n\}$  there exist two vertices  $\sigma_i$  and  $\sigma_j$  such that  $h_t(\sigma_i) < h_t(\sigma_j)$  (or respectively  $h_t(\sigma_i) < h_t(\sigma_j)$ ) for  $t \in (t_{k-1}, t_k)$ ,  $h_t(\sigma_i) = h_t(\sigma_j)$  for  $t = t_k$  and  $h_t(\sigma_i) > h_t(\sigma_j)$  (or respectively  $h_t(\sigma_i) < h_t(\sigma_j)$ ) for  $t \in (t_k, t_{k+1})$ . On each open interval  $I_k = (t_k, t_{k+1})$  the vertex order is fixed for all  $h_t$  with  $t \in I_k$ .

##### B. Stability of $W_q^{\mathcal{RD}}$ for functions with same vertex order

We first consider the simple case where functions with same vertex order are considered, such as functions within the same interval  $I_k = (t_k, t_{k+1})$  of the straight line homotopy  $h_t$ . Let  $a \in I_k$  and  $b \in I_k$  and  $a < b$ , given that  $h_a$  and  $h_b$  have the same vertex order, the critical vertices, the persistence pairing and the region partition  $\{\mathcal{R}_i\}_i$  are the same for  $\mathcal{D}^{\mathcal{R}}(h_a)$  and  $\mathcal{D}^{\mathcal{R}}(h_b)$ , they both contain the same number of region-aware persistence pairs  $b_i^{\mathcal{R}}$  for all  $i \in \{1, \dots, m\}$  with  $m = |\mathcal{D}^{\mathcal{R}}(h_a)| = |\mathcal{D}^{\mathcal{R}}(h_b)|$ . Consider the assignment  $\phi'$  mapping each  $b_i^{\mathcal{R}} = (\mathcal{R}_i, h_a|_{\mathcal{R}_i}, s_i) \in \mathcal{D}^{\mathcal{R}}(h_a)$  to its equivalent pair  $b_j^{\mathcal{R}} = (\mathcal{R}_j, h_b|_{\mathcal{R}_j}, s_j) \in \mathcal{D}^{\mathcal{R}}(h_b)$ , i.e.  $\phi'(b_i^{\mathcal{R}}) = b_j^{\mathcal{R}}$  and  $b_i^{\mathcal{R}}$  and  $b_j^{\mathcal{R}}$  have the same region partition ( $\mathcal{R}_i = \mathcal{R}_j$ ) and critical vertices, only their values can differ (but not their order).

The region-aware Wasserstein distance is the minimum over all possible assignments and therefore bounded by above by the associated cost of  $\phi'$  (line 21). Since the matched pairs given  $\phi'$  have the same region, the regional discrepancy  $C_q^{\mathcal{R}}$  (line 22) consists only in the  $L_q$  norm on the common domain of the matched pairs (line 23), i.e. there are no vertices to compare in their symmetric difference. To simplify notation, let  $A = \mathcal{D}^{\mathcal{R}}(h_a)$  and  $\mathcal{W}_{ab}^q = W_q^{\mathcal{RD}}(\mathcal{D}^{\mathcal{R}}(h_a), \mathcal{D}^{\mathcal{R}}(h_b))^q$ .

$$\mathcal{W}_{ab}^q \leq \sum_{b_i^{\mathcal{R}} \in A} d_q^{\mathcal{R}}(b_i^{\mathcal{R}}, \phi'(b_i^{\mathcal{R}}))^q \quad (21)$$

$$= \sum_{b_i^{\mathcal{R}} \in A} (|s_i - s_j|^q + C_q^{\mathcal{R}}(b_i^{\mathcal{R}}, \phi'(b_i^{\mathcal{R}}))) \quad (22)$$

$$= \sum_{b_i^{\mathcal{R}} \in A} (|s_i - s_j|^q + \sum_{v \in \mathcal{R}_i'} |h'_a|_{\mathcal{R}_i'}(v) - h'_{b|\mathcal{R}_j'}(v)|^q) \quad (23)$$

$$= \sum_{b_i^{\mathcal{R}} \in A} |s_i - s_j|^q + \sum_{b_i^{\mathcal{R}} \in A} \sum_{v \in \mathcal{R}_i'} |h'_a|_{\mathcal{R}_i'}(v) - h'_{b|\mathcal{R}_j'}(v)|^q. \quad (24)$$

Since every saddle is a vertex in  $\mathcal{M}$  and that the matched saddles given  $\phi'$  are the same vertices of  $\mathcal{M}$  then:

$$\sum_{b_i^{\mathcal{R}} \in A} |s_i - s_j|^q \leq \sum_{v \in \mathcal{M}} |h_a(v) - h_b(v)|^q. \quad (25)$$

Then, since each vertex lies in exactly one region  $\mathcal{R}_i$  and that the regions of all  $b_i^{\mathcal{R}}$  form a partition of  $\mathcal{M}$  we have:

$$\begin{aligned} \sum_{b_i^{\mathcal{R}} \in A} \sum_{v \in \mathcal{R}_i'} |h'_a|_{\mathcal{R}_i'}(v) - h'_{b|\mathcal{R}_j'}(v)|^q &= \sum_{b_i^{\mathcal{R}} \in A} \sum_{v \in \mathcal{R}_i} |h_a(v) - h_b(v)|^q \\ &= \sum_{v \in \mathcal{M}} |h_a(v) - h_b(v)|^q. \end{aligned} \quad (26)$$

Note that the term is first rewritten by using the vertices of the region without alignment using the original scalar field. Using Eq. 25 and Eq. 26 in Eq. 24 gives:

$$\mathcal{W}_{ab}^q \leq 2 \sum_{v \in \mathcal{M}} |h_a(v) - h_b(v)|^q = 2 \|h_a - h_b\|_q^q. \quad (27)$$

Hence:

$$W_q^{\mathcal{RD}}(\mathcal{D}^{\mathcal{R}}(h_a), \mathcal{D}^{\mathcal{R}}(h_b)) \leq 2^{1/q} \|h_a - h_b\|_q, \quad (28)$$

for any two functions  $h_a$  and  $h_b$  having the same vertex order.

(i) **Fixed-size domain and vertex weights:** We now prove the stability of  $W_q^{\mathcal{RD}}$  for functions with same vertex order when the region-aware ground metric uses a fixed-size domain and when vertex weights are considered (as described in Appendix B). We will consider weights that decay given the distance to the extremum of a region and are greater than or equal to 0. Since  $h_a$  and  $h_b$  have the same vertex order they have the same extrema vertices, therefore, the weight for each vertex for each region are equal, i.e.  $w_i^{h_a}(v) = w_i^{h_b}(v) = w_i(v)$ .

We consider the assignment  $\phi'$  described in Sec. C-B. Let  $A = \mathcal{D}^{\mathcal{R}}(h_a)$ . To simplify notation, we write the sum of all regional discrepancy terms as  $\mathcal{C}_{\phi'} = \sum_{b_i^{\mathcal{R}} \in A} C_q^{\mathcal{R}}(b_i^{\mathcal{R}}, \phi'(b_i^{\mathcal{R}}))$  and  $v'_i$  the vertex  $v \in \mathcal{M}$  after aligning the domain to the extremum of the region  $\mathcal{R}_i$ . Since the matched pairs given  $\phi'$  shared the same extrema, the vertices of  $\Lambda$  being compared using their extended functions are therefore a subset of the vertices of  $\mathcal{M}$  allowing to bound line 30 by line 31. Then, since  $\sum_{b_i^{\mathcal{R}} \in A} w_i(v) = 1$ , that each  $w_i(v) \geq 0$  and that  $q \geq 1$  then  $\sum_{b_i^{\mathcal{R}} \in A} w_i(v)^q \leq 1$  allowing to bound line 32 by line 33.

$$\mathcal{C}_{\phi'} = \sum_{b_i^{\mathcal{R}} \in A} \|\Psi_i^{h_a} - \Psi_j^{h_b}\|_{L_q(\Lambda)}^q \quad (29)$$

$$= \sum_{b_i^{\mathcal{R}} \in A} \sum_{v'_i \in \Lambda} |w_i(v) \tilde{h}_{a_i}(v'_i) - w_i(v) \tilde{h}_{b_j}(v'_i)|^q \quad (30)$$

$$\leq \sum_{b_i^{\mathcal{R}} \in A} \sum_{v \in \mathcal{M}} w_i(v)^q |h_a(v) - h_b(v)|^q \quad (31)$$

$$= \sum_{v \in \mathcal{M}} \sum_{b_i^{\mathcal{R}} \in A} w_i(v)^q |h_a(v) - h_b(v)|^q \quad (32)$$

$$\leq \sum_{v \in \mathcal{M}} |h_a(v) - h_b(v)|^q. \quad (33)$$

Hence:

$$\sum_{b_i^{\mathcal{R}} \in A} C_q^{\mathcal{R}}(b_i^{\mathcal{R}}, \phi'(b_i^{\mathcal{R}})) \leq \sum_{v \in \mathcal{M}} |h_a(v) - h_b(v)|^q. \quad (34)$$

Writing  $\mathcal{W}_{ab}^q = W_q^{\mathcal{RD}}(\mathcal{D}^{\mathcal{R}}(h_a), \mathcal{D}^{\mathcal{R}}(h_b))^q$ , we have:

$$\mathcal{W}_{ab}^q \leq \sum_{b_i^{\mathcal{R}} \in A} d_q^{\mathcal{R}}(b_i^{\mathcal{R}}, \phi'(b_i^{\mathcal{R}}))^q \quad (35)$$

$$= \sum_{b_i^{\mathcal{R}} \in A} (|s_i - s_j|^q + C_q^{\mathcal{R}}(b_i^{\mathcal{R}}, \phi'(b_i^{\mathcal{R}}))) \quad (36)$$

$$= \sum_{b_i^{\mathcal{R}} \in A} |s_i - s_j|^q + \sum_{b_i^{\mathcal{R}} \in A} C_q^{\mathcal{R}}(b_i^{\mathcal{R}}, \phi'(b_i^{\mathcal{R}})). \quad (37)$$

Using Eq. 25 and Eq. 34 in Eq. 37 gives:

$$\mathcal{W}_{ab}^q \leq 2 \sum_{v \in \mathcal{M}} |h_a(v) - h_b(v)|^q = 2 \|h_a - h_b\|_q^q. \quad (38)$$

Hence:

$$W_q^{\mathcal{RD}}(\mathcal{D}^{\mathcal{R}}(h_a), \mathcal{D}^{\mathcal{R}}(h_b)) \leq 2^{1/q} \|h_a - h_b\|_q, \quad (39)$$

for any two functions  $h_a$  and  $h_b$  having the same vertex order when using a fixed-size domain and vertex weights.

### C. Stability of $W_q^{\mathcal{RD}}$ for any functions

We have shown in the previous section a local stability result for scalar fields having the same vertex order, such as those within a same interval  $I_k = (t_k, t_{k+1})$  of the partition of the homotopy  $h_t$ . To prove the stability of  $W_q^{\mathcal{RD}}$  for any functions (when no extremum-swap instabilities are considered) we have to *combine* together all the local results over all intervals  $I_k$  with  $k \in \{0, \dots, n\}$  of the homotopy  $h_t$ .

Recall that all scalar fields  $h_t$  with values of  $t$  in the same interval  $I_k = (t_k, t_{k+1})$  share a common vertex order. Moreover, since  $f$  and  $g$  are considered to be injective on the vertices of  $\mathcal{M}$ , the common vertex order within  $I_k$  is a *strict total* order (an order without equalities), because ties can only occur at the endpoints  $t_k$  and  $t_{k+1}$ . At these endpoints, the value of two vertices become equal and the vertex orders of  $h_{t_k}$  and  $h_{t_{k+1}}$  become *partial* (i.e. contain equalities). A partial order can be extended to various strict total orders depending on how ties are broken. Specifically, since  $h_{t_k}$  and  $h_{t_{k+1}}$  are the endpoints of  $I_k$ , the common strict total order within  $I_k$  is one possible (*linear*) extension of the partial orders of  $h_{t_k}$  and  $h_{t_{k+1}}$ . Let  $\mathcal{D}_k^{\mathcal{R}}(h_{t_k})$  and  $\mathcal{D}_k^{\mathcal{R}}(h_{t_{k+1}})$  the diagram of respectively  $h_{t_k}$  and  $h_{t_{k+1}}$  computed using the common order within  $I_k$ . Since they share the same order we can use the result of Sec. C-B (i.e. Eq. 28 or Eq. 39) for these diagrams:

$$\begin{aligned} W_q^{\mathcal{RD}}(\mathcal{D}_k^{\mathcal{R}}(h_{t_k}), \mathcal{D}_k^{\mathcal{R}}(h_{t_{k+1}})) &\leq 2^{1/q} \|h_{t_k} - h_{t_{k+1}}\|_q \\ &= 2^{1/q} (t_{k+1} - t_k) \|f - g\|_q. \end{aligned} \quad (40)$$

The last line follows from rewriting the term  $h_{t_k} - h_{t_{k+1}}$  as  $(t_{k+1} - t_k)(f - g)$  and due to  $\|cu\|_q = |c| \|u\|_q$  for any scalar  $c$ .

Now that we have a local stability result for the endpoints of an interval  $I_k$ , to prove the stability of  $W_q^{\mathcal{RD}}$  we aim to combine such result over all intervals of the homotopy. Specifically, recall that  $h_{t_0} = f$  and  $h_{t_{n+1}} = g$ , we will use the triangle inequality to bound by above  $W_q^{\mathcal{RD}}(\mathcal{D}^{\mathcal{R}}(f), \mathcal{D}^{\mathcal{R}}(g))$  by the sum of the distances between the scalar fields of all consecutive  $t_k$  with  $k \in \{0, \dots, n+1\}$  that will be shown to be itself bounded by above by the  $L_q$  norm of the difference between  $f$  and  $g$  (up to a factor independent of  $f$  and  $g$ ).

Given two intervals  $I_k = (t_k, t_{k+1})$  and  $I_{k+1} = (t_{k+1}, t_{k+2})$ , the distance between  $\mathcal{D}_k^{\mathcal{R}}(h_{t_k})$  and  $\mathcal{D}_{k+1}^{\mathcal{R}}(h_{t_{k+2}})$ , i.e. the diagram of the endpoints of these two intervals given the common order of respectively  $I_k$  and  $I_{k+1}$ , can be bounded by above (using the triangle inequality) by:

$$W_q^{\mathcal{RD}}(\mathcal{D}_k^{\mathcal{R}}(h_{t_k}), \mathcal{D}_{k+1}^{\mathcal{R}}(h_{t_{k+2}})) \leq W_q^{\mathcal{RD}}(\mathcal{D}_k^{\mathcal{R}}(h_{t_k}), \mathcal{D}_k^{\mathcal{R}}(h_{t_{k+1}})) \quad (41)$$

$$+ W_q^{\mathcal{RD}}(\mathcal{D}_k^{\mathcal{R}}(h_{t_{k+1}}), \mathcal{D}_{k+1}^{\mathcal{R}}(h_{t_{k+1}})) \quad (42)$$

$$+ W_q^{\mathcal{RD}}(\mathcal{D}_{k+1}^{\mathcal{R}}(h_{t_{k+1}}), \mathcal{D}_{k+1}^{\mathcal{R}}(h_{t_{k+2}})), \quad (43)$$

i.e. by the distance between the diagrams of the endpoints of  $I_k$  given its common order (line 41), plus the distance between the diagrams of the endpoints of  $I_{k+1}$  given its common order (line 43), plus a *bridge* term corresponding to the distance

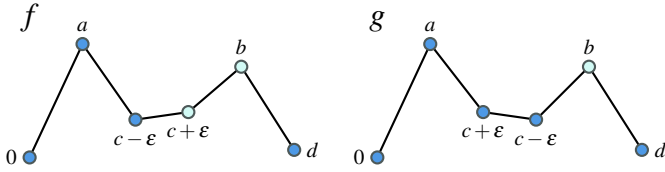


Fig. 15. The scalar fields considered for the counter-example that shows the impossibility for the null background to have a stability bound for  $W_q^{\mathcal{RD}}$ . The scalar fields  $f$  and  $g$  are defined on the same one-dimensional domain. The labels indicate the values of the corresponding vertices and their position in the y-axis indicate their value order (for instance  $a > b$  and  $\varepsilon$  is very small). The nodes color represent the (maximum) region membership of the vertices to the two regions having respectively  $a$  and  $b$  as extremum.

between the diagram of  $h_{t_{k+1}}$  given the common order within  $I_k$  and the one given the common order within  $I_{k+1}$  (line 42).

**Remark:** the choice of how ties are broken does not affect the resulting (classical) persistence diagram [73], i.e. for any strict total order extending a given partial order, the birth–death pairs (intervals) are the same, even if the specific critical vertices can differ. Consequently, in the classical case, the bridge term at a swap time  $t_k$  is 0, and the proof closes using Eq. 44 alone. By contrast, for region-aware persistence diagrams, region memberships do depend on the tie-breaking (a vertex could change region depending on the order choice), so the diagrams at the swap time may differ. We therefore need to control this bridge term in what follows.

(i) **No possible bound using null background:** We first show with a counter-example that  $W_q^{\mathcal{RD}}$  with the null background can not be bounded by above by the  $L_q$  norm of the difference between the two input scalar fields (up to a factor).

Let  $f$  and  $g$  two scalar fields described in Fig. 15 having respectively the same two extrema  $a$  and  $b$ . We will use  $q = 1$  for simplicity but the example works for every valid values of  $q$ . The  $L_1$  norm of the difference between  $f$  and  $g$  is  $\|f - g\|_1 = 4\varepsilon$  (with  $\varepsilon$  very small). They only differ by the swap of two vertices having very close values. When comparing the regions of  $b$  in respectively  $f$  and  $g$ , the vertex of the region  $b$  in  $f$  with value  $c + \varepsilon$  is part of the symmetric difference (it has no correspondent vertex in the region of  $b$  in  $g$ ), due to the null background, it is therefore compared to 0. The difference  $|c + \varepsilon - 0|$  can be arbitrarily large compared to  $4\varepsilon$  hence there is no constant  $C$  such that  $W_1^{\mathcal{RD}}(\mathcal{D}^{\mathcal{R}}(f), \mathcal{D}^{\mathcal{R}}(g)) \leq C\|f - g\|_1$ .

Coming back to prove the stability of  $W_q^{\mathcal{RD}}$ , recall that at a swap time  $t_{k+1}$  a vertex  $x$  could change region, i.e.  $x \in \mathcal{R}$  in  $\mathcal{D}_k^{\mathcal{R}}(h_{t_{k+1}})$  but  $x \notin \mathcal{R}$  in  $\mathcal{D}_{k+1}^{\mathcal{R}}(h_{t_{k+1}})$  for a region  $\mathcal{R}$  in both diagrams. When comparing both  $\mathcal{R}$  regions in both diagrams, the contribution of  $x$ , coming from the symmetric difference, is  $|h_{t_{k+1}}(x) - 0|^q$  corresponding to a comparison to 0 that can be arbitrarily large. Hence, the bridge term (line 42) do not vanish and can not be bounded by  $C\|f - g\|_q$  for some constant  $C$ . Consequently, summing the local interval bounds cannot yield a global stability bound for  $W_q^{\mathcal{RD}}$  using the null background.

(ii) **Stability using data background:** Recall that no extremum-swap instabilities are considered, therefore, at a swap time  $t_{k+1}$ , the extrema vertices of  $\mathcal{D}_k^{\mathcal{R}}(h_{t_{k+1}})$  and  $\mathcal{D}_{k+1}^{\mathcal{R}}(h_{t_{k+1}})$  are the same (even if the region membership of various vertices can change due to the swap of the values of

two saddle, and/or regular, vertices). A notable exception is the possible appearance or disappearance (depending on the common order within  $I_k$  and  $I_{k+1}$ ) of a zero-persistence pair formed by the two vertices that tie at  $t_{k+1}$  (having equal value).

Consider the assignment  $\phi'$  that maps each pair  $b_i^{\mathcal{R}} = (\mathcal{R}_i, h_{t_{k+1}}|_{\mathcal{R}_i}, s_i) \in \mathcal{D}_k^{\mathcal{R}}(h_{t_{k+1}})$  to the pair having the same extremum vertex  $b_j^{\mathcal{R}} = (\mathcal{R}_j, h_{t_{k+1}}|_{\mathcal{R}_j}, s_j) \in \mathcal{D}_{k+1}^{\mathcal{R}}(h_{t_{k+1}})$ . If applicable  $\phi'$  also maps to the diagonal the zero-persistence pair of the two vertices that tie with a cost of 0. Let  $\Psi_i^{h_{t_{k+1}}}$  and  $\Psi_j^{h_{t_{k+1}}}$  the weighted extended functions using the data background (Appendix B) of respectively  $b_i^{\mathcal{R}}$  and  $b_j^{\mathcal{R}}$ . Since the extrema vertices are the same and the scalar values also are, then  $\Psi_i^{h_{t_{k+1}}}$  and  $\Psi_j^{h_{t_{k+1}}}$  coincide everywhere and the regional discrepancy is  $\|\Psi_i^{h_{t_{k+1}}} - \Psi_j^{h_{t_{k+1}}}\|_{L_q(\Lambda)}^q = 0$ . Moreover, for each pair, either its saddle is unchanged for both orders or it is one of the two vertices that tie, in either case the saddle values  $s_i$  and  $s_j$  are also equal and  $|s_i - s_j|^q = 0$ . We conclude that  $d_q^{\mathcal{R}}(b_i^{\mathcal{R}}, b_j^{\mathcal{R}}) = 0$  for all pairs, hence  $W_q^{\mathcal{RD}}(\mathcal{D}_k^{\mathcal{R}}(h_{t_{k+1}}), \mathcal{D}_{k+1}^{\mathcal{R}}(h_{t_{k+1}})) = 0$ .

Coming back to the combination of two intervals (Eq. 41), the bridge term (line 42) being 0, we can rewrite Eq. 41 as:

$$W_q^{\mathcal{RD}}(\mathcal{D}_k^{\mathcal{R}}(h_{t_k}), \mathcal{D}_{k+1}^{\mathcal{R}}(h_{t_{k+2}})) \leq W_q^{\mathcal{RD}}(\mathcal{D}_k^{\mathcal{R}}(h_{t_k}), \mathcal{D}_k^{\mathcal{R}}(h_{t_{k+1}})) + W_q^{\mathcal{RD}}(\mathcal{D}_{k+1}^{\mathcal{R}}(h_{t_{k+1}}), \mathcal{D}_{k+1}^{\mathcal{R}}(h_{t_{k+2}})).$$

We can now generalize this result to not only over 2 intervals  $I_k$  and  $I_{k+1}$  but over all intervals  $I_k$  with  $k \in \{0, \dots, n\}$ :

$$W_q^{\mathcal{RD}}(\mathcal{D}_0^{\mathcal{R}}(h_{t_0}), \mathcal{D}_n^{\mathcal{R}}(h_{t_{n+1}})) \leq \sum_{k=0}^n W_q^{\mathcal{RD}}(\mathcal{D}_k^{\mathcal{R}}(h_{t_k}), \mathcal{D}_k^{\mathcal{R}}(h_{t_{k+1}})).$$

Note that here  $\mathcal{D}_0^{\mathcal{R}}(h_{t_0})$  correspond to the diagram of  $h_{t_0}$  computed using the common order in  $I_0$ , to not confuse with the diagram of dimension 0 (same remark for  $\mathcal{D}_n^{\mathcal{R}}(h_{t_{n+1}})$ ).

Recall that  $h_{t_0} = f$  and  $h_{t_{n+1}} = g$ , we can therefore write  $\mathcal{D}^{\mathcal{R}}(f) = \mathcal{D}_0^{\mathcal{R}}(h_{t_0})$  and  $\mathcal{D}^{\mathcal{R}}(g) = \mathcal{D}_n^{\mathcal{R}}(h_{t_{n+1}})$  (since vertex orders agree). Moreover, to simplify notation, we write  $W_q^{\mathcal{RD}}(\mathcal{D}^{\mathcal{R}}(h_{t_k}), \mathcal{D}^{\mathcal{R}}(h_{t_{k+1}})) = W_q^{\mathcal{RD}}(\mathcal{D}_k^{\mathcal{R}}(h_{t_k}), \mathcal{D}_k^{\mathcal{R}}(h_{t_{k+1}}))$ . Then, using Eq. 40, we have:

$$\begin{aligned} W_q^{\mathcal{RD}}(\mathcal{D}^{\mathcal{R}}(f), \mathcal{D}^{\mathcal{R}}(g)) &\leq \sum_{k=0}^n W_q^{\mathcal{RD}}(\mathcal{D}_k^{\mathcal{R}}(h_{t_k}), \mathcal{D}_k^{\mathcal{R}}(h_{t_{k+1}})) \\ &\leq \sum_{k=0}^n 2^{1/q} \|h_{t_k} - h_{t_{k+1}}\|_q \\ &= \sum_{k=0}^n 2^{1/q} (t_{k+1} - t_k) \|f - g\|_q \\ &= 2^{1/q} \|f - g\|_q, \end{aligned} \quad (44)$$

Hence:

$$W_q^{\mathcal{RD}}(\mathcal{D}^{\mathcal{R}}(f), \mathcal{D}^{\mathcal{R}}(g)) \leq 2^{1/q} \|f - g\|_q, \quad (45)$$

for any two functions  $f$  and  $g$  when no extremum-swap instabilities are considered.

## APPENDIX D

### REGIONS COMPRESSION

In this section we briefly present the compression methods considered in this work, and how given the compression parameter  $\tau$  (Sec. IV-B of the main manuscript) these models need to be defined in order to provide a compressed size as close as possible to the target one.

Given  $\tau$ , we compute the desired number of parameters in the compressed size as  $p = \lfloor \tau n \rfloor$  with  $n$  the number of values in the input data to compress (we consider 32 bits float data).

**ZFP** [49] is a lossy compression method that operates by dividing data into small blocks, transforming them into a fixed-point representation, and applying a customized encoding scheme that balances precision and compression efficiency. ZFP supports a fixed-rate mode allowing to bound the compressed size given an input parameter. Specifically, we need to provide a rate  $r$  defined as the number of bits per value that we allow. Considering 32 bits float values, the total target bit budget is  $32p$ . The entire dataset has  $n$  values, therefore the average number of bits allowed per value is  $r = \frac{32p}{n}$ .

**Neural fields** [50] are neural networks applied to compression, it consists in the use of an architecture whose size is lower than the input data, the network is then be trained to predict the scalar value given the input coordinates and the architecture is itself the compressed representation. Inspired by Lu et al. [50] we use an architecture with a total of  $l$  layers consisting of an input layer of  $d$  neurons ( $d$  being the dimension of the data to compress),  $l-2$  hidden layers of  $k$  neurons each and an output layer of 1 neuron (to predict a scalar field). The total number of parameters (weights and biases) is given by:

$$\underbrace{dk + k}_{\text{input}} + \underbrace{(l-2)(k^2 + k)}_{\text{hidden}} + \underbrace{k + 1}_{\text{output}}, \quad (46)$$

We consider that the total number of layers  $l$  is an input parameter of the method, we now want to solve for  $k$ , the number of neurons in each hidden layer:

$$\begin{aligned} dk + k + (l-2)(k^2 + k) + k + 1 &= p \\ \Leftrightarrow dk + k + (l-2)(k^2 + k) + k + 1 - p &= 0 \\ \Leftrightarrow k^2(l-2) + k(d+1+l-2+1) + 1 - p &= 0 \\ \Leftrightarrow k^2(l-2) + k(d+l) + 1 - p &= 0. \end{aligned} \quad (47)$$

We solve this using the quadratic formula, we first set  $a = l-2$ ,  $b = d+l$  and  $c = 1-p$ , we then compute  $\Delta = b^2 - 4ac$  and finally get  $k = \text{round}(\frac{-b+\sqrt{\Delta}}{2a})$ .

In practice we use  $l = 18$  with therefore 16 hidden layers, we also use residual connections for the hidden layers. We found that the following parameters gave the best results averaged over all ensembles: a learning rate of 0.005, a batch size of 1024, a number of epochs of 1000, a learning rate scheduling multiplying by 0.9 the learning rate every 20 epochs and the leaky ReLU activation function. We optimize the model using a NVIDIA GeForce RTX 4090 GPU with 24 GB of RAM.

**B-splines** [64] consist in the optimization of a grid of control points to best fit the input data. Each point value

in the input data is approximated by weighted sum of the optimized control points where the weights correspond to the so-called *basis functions*. The grid of control points acts as the compressed representation. We define the grid such that its aspect ratio is as close as possible to the one of the input data. It amounts to set the number of points in each dimension of the grid to be proportionnal to the extent of that dimension in the input data. Specifically, let  $e_d$  the extend of the  $d^{\text{th}}$  dimension in the input data (i.e. the difference between the maximum and the minimum coordinate for this specific dimension over all the points). Let  $g_d$  the number of control points in the  $d^{\text{th}}$  dimension in the grid. In 2D we want that:

$$\begin{cases} \frac{g_0}{g_1} = \frac{e_0}{e_1} \end{cases} \quad (48)$$

$$g_0 g_1 = p \quad (49)$$

We can therefore write Eq. 48 as  $g_0 = \frac{e_0}{e_1} g_1$ , then replacing  $g_0$  in Eq. 49 by this new expression gives:

$$\begin{aligned} \frac{e_0}{e_1} g_1^2 &= p \\ \Leftrightarrow g_1^2 &= p \frac{e_1}{e_0} \\ \Leftrightarrow g_1 &= \sqrt{p \frac{e_1}{e_0}} \end{aligned} \quad (50)$$

In practice, we therefore set the number of control points in the first dimension as  $g_0 = \text{round}(\frac{e_0}{e_1} \sqrt{p \frac{e_1}{e_0}})$  and as  $g_1 = \text{round}(\sqrt{p \frac{e_1}{e_0}})$  for the second dimension.

In 3D we want that:

$$\begin{cases} \frac{g_0}{g_2} = \frac{e_0}{e_2} \end{cases} \quad (51)$$

$$\begin{cases} \frac{g_1}{g_2} = \frac{e_1}{e_2} \end{cases} \quad (52)$$

$$g_0 g_1 g_2 = p \quad (53)$$

We rewrite Eq. 51 as  $g_0 = \frac{e_0}{e_2} g_2$ . Then, we rewrite Eq. 52 as  $g_1 = \frac{e_1}{e_2} g_2$ . Replacing these expressions in Eq. 53 gives:

$$\begin{aligned} \frac{e_0}{e_2} \frac{e_1}{e_2} g_2^3 &= p \\ \Leftrightarrow g_2^3 &= p \frac{e_2^2}{e_0 e_1} \\ \Leftrightarrow g_2 &= \sqrt[3]{p \frac{e_2^2}{e_0 e_1}} \end{aligned} \quad (54)$$

In practice, we therefore set the number of control points in the first dimension as  $g_0 = \text{round}(\frac{e_0}{e_2} \sqrt[3]{p \frac{e_2^2}{e_0 e_1}})$ , as  $g_1 = \text{round}(\frac{e_1}{e_2} \sqrt[3]{p \frac{e_2^2}{e_0 e_1}})$  for the second dimension and finally as  $g_2 = \text{round}(\sqrt[3]{p \frac{e_2^2}{e_0 e_1}})$  for the third dimension.

## APPENDIX E

### GLOBAL GEOMETRIC PROPERTIES DEFINITIONS

We define in this section the global geometric properties that were studied in previous works [48], [69], [74], [93] and



consist of the baselines with which we compare our method to (in addition to the classical Wasserstein distances [20], [66]). These two methods extend the classical ground metric (Eq. 1 of the main manuscript) to not only compare the birth and the death of two persistence pairs but also some global geometric properties about them.

First, the so-called *geometrical lifting* [48], [74], [93] uses the distance between the coordinates in the domain of the extrema of the persistence pairs. We will define it for 3D domains but its general definition is not restricted to this specific dimension. Let two persistence pairs  $p_i = (b_i, d_i)$  and  $p_j = (b_j, d_j)$  and their 3D coordinates  $(x_i, y_i, z_i)$  and  $(x_j, y_j, z_j)$ , the geometrical lifting ground metric is defined as:

$$d_q^{\mathcal{L}}(p_i, p_j) = \left( |b_j - b_i|^q + |d_j - d_i|^q + w_{\mathcal{L}}(|x_j - x_i|^q + |y_j - y_i|^q + |z_j - z_i|^q) \right)^{1/q}, \quad (55)$$

where  $w_{\mathcal{L}}$  allows to weight the contribution of the coordinates distance in the final distance. As suggested by Soler et al. [74], the coordinates can be normalized by the bounds of the input data. We note  $W_2^{\mathcal{L}\mathcal{D}}$  and  $W_2^{\mathcal{L}\mathcal{T}}$  the Wasserstein distance for respectively persistence diagrams and merge trees using  $d_q^{\mathcal{L}}(p_i, p_j)$  as a ground metric.

Second, we define the volume difference [69]. Let two persistence pairs  $p_i = (b_i, d_i)$  and  $p_j = (b_j, d_j)$  and the volume, i.e. the number of vertices, of their regions in the domain  $v_i$  and  $v_j$ , the volume difference ground metric is defined as:

$$d_q^{\mathcal{V}}(p_i, p_j) = \left( |b_j - b_i|^q + |d_j - d_i|^q + w_{\mathcal{V}}|v_j - v_i|^q \right)^{1/q}, \quad (56)$$

where  $w_{\mathcal{V}}$  allows to weight the contribution of the volume difference in the final distance. As suggested by Saikia et al. [69], the volume of a persistence pair can be normalized by the global volume of the input data. We note  $W_2^{\mathcal{V}\mathcal{D}}$  and  $W_2^{\mathcal{V}\mathcal{T}}$  the Wasserstein distance for respectively persistence diagrams and merge trees using  $d_q^{\mathcal{V}}(p_i, p_j)$  as a ground metric.

## APPENDIX F

### DIMENSIONALITY REDUCTION INDICATORS

To quantify how well 2D embedding preserves the original cluster structure, we use the *Normalized Mutual Information* (NMI) [77], [84] and the *Adjusted Rand Index* (ARI) [41], two widely-used, label-permutation-invariant measures for comparing a predicted partition  $\mathcal{C}$  with a ground-truth partition  $\mathcal{G}$ .

Both metrics (i) are *label-invariant*, (ii) account for chance agreement, and (iii) are recommended in comparative studies of clustering-quality indices [84]. Together they offer complementary perspectives—information-theoretic fidelity (NMI) and pairwise correctness (ARI)—on how faithfully the 2D embedding preserves the high-dimensional clusters.

**Normalized Mutual Information (NMI).** For two partitions  $\mathcal{C} = \{\mathcal{C}_k\}_{k=1}^K$  and  $\mathcal{G} = \{\mathcal{G}_\ell\}_{\ell=1}^L$  of the same set of  $n$  points, let  $p_k = |\mathcal{C}_k|/n$  and  $q_\ell = |\mathcal{G}_\ell|/n$  be the empirical cluster-label

probabilities, and  $p_{k\ell} = |\mathcal{C}_k \cap \mathcal{G}_\ell|/n$  the joint probabilities. The *Shannon entropy* of a partition is:

$$H(\mathcal{C}) = - \sum_{k=1}^K p_k \log p_k, \quad H(\mathcal{G}) = - \sum_{\ell=1}^L q_\ell \log q_\ell.$$

The *mutual information* between the two partitions is:

$$I(\mathcal{C}; \mathcal{G}) = \sum_{k=1}^K \sum_{\ell=1}^L p_{k\ell} \log \frac{p_{k\ell}}{p_k q_\ell},$$

which measures the reduction in uncertainty about  $\mathcal{C}$  given  $\mathcal{G}$  (and vice versa). The *Normalized Mutual Information* is the symmetric normalization:

$$\text{NMI}(\mathcal{C}, \mathcal{G}) = \frac{2I(\mathcal{C}; \mathcal{G})}{H(\mathcal{C}) + H(\mathcal{G})}, \quad (57)$$

taking values in  $[0, 1]$ , with 1 for perfect agreement and 0 for statistical independence between the partitions. Because it derives from information theory, NMI is invariant to cluster-label permutations and rewards embeddings that preserve the full *information content* of the original clustering.

**Adjusted Rand Index (ARI).** Let  $\mathcal{C} = \{\mathcal{C}_i\}_{i=1}^K$  and  $\mathcal{G} = \{\mathcal{G}_j\}_{j=1}^L$  be two partitions of  $n$  points. Let the  $K \times L$  *contingency table*  $[n_{ij}]$ , where:

$$n_{ij} = |\mathcal{C}_i \cap \mathcal{G}_j|$$

counts how many points are simultaneously assigned to cluster  $\mathcal{C}_i$  and to cluster  $\mathcal{G}_j$ . The row sums  $n_{i\cdot} = \sum_j n_{ij}$  and column sums  $n_{\cdot j} = \sum_i n_{ij}$  are the cluster sizes in  $\mathcal{C}$  and  $\mathcal{G}$ , respectively.

The *chance agreement baseline* is defined as:

$$E = \frac{\sum_i \binom{n_{i\cdot}}{2} \sum_j \binom{n_{\cdot j}}{2}}{\binom{n}{2}},$$

i.e. the expected number of agreeing pairs under a random matching that preserves the given row and column totals. Then, the *Adjusted Rand Index* is:

$$\text{ARI}(\mathcal{C}, \mathcal{G}) = \frac{\sum_{i,j} \binom{n_{ij}}{2} - E}{\frac{1}{2} \left[ \sum_i \binom{n_{i\cdot}}{2} + \sum_j \binom{n_{\cdot j}}{2} \right] - E}. \quad (58)$$

ARI equals 1 for perfect agreement, 0 for random agreement, and can be negative when the partitions disagree more than would be expected by chance.