

# Robust Barycenters of Persistence Diagrams

Keanu Sisouk, Eloi Tanguy, Julie Delon and Julien Tierny

**Abstract**—This short paper presents a general approach for computing robust Wasserstein barycenters [2], [78], [79] of persistence diagrams. The classical method consists in computing assignment arithmetic means after finding the optimal transport plans between the barycenter and the persistence diagrams. However, this procedure only works for the transportation cost related to the  $q$ -Wasserstein distance  $W_q$  when  $q=2$ . We adapt an alternative fixed-point method [74] to compute a barycenter diagram for generic transportation costs ( $q > 1$ ), in particular those robust to outliers,  $q \in (1, 2)$ . We show the utility of our work in two applications: (i) the clustering of persistence diagrams on their metric space and (ii) the dictionary encoding of persistence diagrams [71]. In both scenarios, we demonstrate the added robustness to outliers provided by our generalized framework. Our Python implementation is available at this address: <https://github.com/Keanu-Sisouk/RobustBarycenter>.

**Index Terms**—Topological data analysis, ensemble data, persistence diagrams, Wasserstein barycenter.

## I. INTRODUCTION

WITH measurement devices and numerical techniques being more and more precise, the resulting datasets have become more and more complex geometrically. This complexity hinders the users during any exploration and analysis to interpret them. Those challenges motivate the conception of expressive, informative, concise and simple data abstractions, encoding the main features and patterns of interest of the data.

Topological Data Analysis (TDA) [25] is a family of tools developed to address this challenge. It aims to provide simple objects, called topological representations, describing the main topological structures in a dataset. TDA has become a staple for analyzing scalar data, its efficiency and robustness being shown in numerous visualization tasks [41]. It has proven to be successful in many applications, several examples include turbulent combustion [11], [37], [49], material sciences [29], [39], [40], [72], nuclear energy [52], fluid dynamics [44], [56], bioimaging [3], [9], [16], chemistry [6], [34], [57], [58] or astrophysics [70], [73] to name a few.

Among the different topological representations, such as the merge and contour trees [1], [14], [15], [35], [51], [75], the Reeb graph [7], [24], [36], [60], [61], [76], or the Morse-Smale complex [10], [22], [26], [27], [33], [38], [66], [69], the Persistence Diagram (Fig. 1) as been prominently used and studied. It is a simple and concise topological representation, which encodes the main topological features of data.

With the increase of geometrical complexity discussed above, a new challenge has come forth as users are confronted to the emergence of *ensemble datasets*. With these representations, a given phenomenon is not only described by a single dataset, but by a *collection* of datasets, called *ensemble members*. This results in the analysis of an ensemble where each element is a topological representation, such as a persistence diagram.

However, this process moves the problem of analyzing an ensemble of scalar fields to analyzing an ensemble of persistence diagrams. To address this challenge, the question of finding a good *representative* of such an ensemble has emerged. An established answer to that question is the notion of *Wasserstein barycenter* [2], [78] of persistence diagrams, based on the so called  $q$ -Wasserstein distance [81]. To compute such barycenters, several algorithms have been proposed for  $q=2$  [48], [78], [80]. This barycenter emulates the behavior of an *average*, and as such, it is also sensitive to outliers.

Works on *Wasserstein medians* counter this issue [13], [77], showing their stability properties. In the context of probability measures, extensions to generic transportation costs have been introduced recently [12], [74].

This short paper proposes to leverage generic transportation costs for computing robust  $q$ -Wasserstein barycenters of ensembles of persistence diagrams, with  $q > 1$ . For this, we adapted a recent fixed-point method [74] from generic probability measures to persistence diagrams. As discussed above, the resulting barycenters have the advantage of being more *robust* to the presence of outlier diagrams in the ensemble. Experiments on synthetic and real-life data showcase this property for typical ensembles of persistence diagrams used previously in the literature. We show the utility of our barycenters by using them in a clustering problem on the Wasserstein metric space and by using them for a Wasserstein dictionary encoding problem [71]. We give details on this *robust barycentric* framework in Sec. III, then we show the utility of these robust barycenters with experiments in Sec. IV with two applications in Sec. IV-A and Sec. IV-B.

### A. Related work

The literature related to our work can be classified into two main categories, reviewed in the following: (i) ensemble visualization, and (ii) topological methods for ensemble of persistence diagrams.

**(i) Ensemble visualization:** A common approach to model data variability is using ensemble datasets. In this setting, the variability is modeled by a succession of empirical observations (i.e., the *members* of the ensemble). Existing approaches compute geometrical objects encoding the features of interest (such as level sets and streamlines for example) for each member of the ensemble. Then, a *representative* of this ensemble of geometrical objects can be estimated. In light of this, several techniques has been established. For example, spaghetti plots [23] are used to study level-set variability, especially for weather data [65], [68]. Box-plots [53], [82] are used to analyze the variability of contours and curves. In the case of flow ensemble, Hummel et al. [42] has proposed a Lagrangian framework for classification purposes.

More precisely, clustering techniques have been studied to highlight the main trends in ensemble of streamlines [30] and isocontours [31]. However, only few approaches applied those strategies to topological representations. Favelier et al. [28] and Athawale et al. [4] introduced techniques for analyzing the geometrical variability of critical points and gradient separatrices respectively. Ensemble layouts have been proposed for contour trees [47], [84]. However, the above techniques do not focus on the computation of a representative of an ensemble of topological objects.

**(ii) Topological methods:** To find a *representative* of an ensemble of persistence diagrams, notions from optimal transport [43], [54] were adapted to persistence diagrams. A central notion is the so-called *Wasserstein* distance [67]. The Wasserstein distance between persistence diagrams [25] (Sec. II-B) has been studied by the TDA community [18], [19]. This distance is computed by solving an assignment problem, for which exact [55] and approximate [5], [45] implementation can be found in open-source [8], [32]. Using this distance, the *Wasserstein* barycenter is used to find a *representative* diagram of an ensemble, [2]. Turner et al. [78] first introduced an algorithm for the computation of such a barycenter for persistence diagrams, along with convergence results and theoretical properties of the persistence diagram space. Lacombe et al. [48] proposed a method to compute a barycenter based on the entropic formulation of optimal transport [20], [21]. However, this method requires a vectorization of persistence diagrams, which is not only subject to parameters, but which also challenges visual analysis and inspection. Indeed, in this case the features of interest cannot be tracked by the users during the analysis. Vidal et al. [80] proposed an approach allowing the tracking of the features. This method is based on a progressive framework, which accelerates the computation time compared to Turner et al.'s method [78]. To take it further, several authors have proposed methods to find a representation of ensembles of topological descriptors by a basis of *representative* descriptors. Li et al. [50] leveraged sketching methods [83] for vectorized merge trees. Pont et al. [64] introduced a principal geodesic analysis method for merge trees, and Sisouk et al. [71] brought forth a Wasserstein dictionary encoding method for an ensemble of persistence diagrams. Both latter methods avoid the difficulties associated with vectorizations (e.g. quantization and linearization artifacts, inaccuracies in vectorization reversal). However, all of the above literature propose representatives that can be sensible to the presence of outliers. Turner et al. [77] studied the notion of median of a population of persistence diagrams, but no exact algorithm nor computation was proposed. In this work, we describe a general framework for computing a robust barycenter, by adapting a recent fixed point method [74] from generic probability measures to persistence diagrams. This robust barycenter is more stable to the presence of outliers, thereby enhancing other analysis frameworks such as clustering algorithms or dictionary-based encodings.

## B. Contributions

This paper makes the following contributions:

- 1) *A general framework for robust barycenters of persistence diagrams:* By adapting a recent approach [74] from

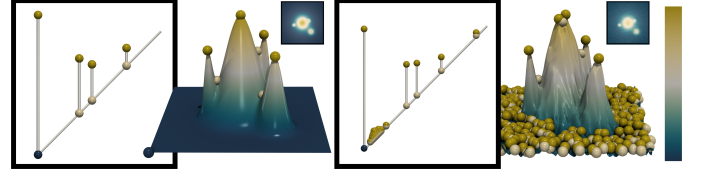


Fig. 1. Example of persistence diagrams of a smooth (left) and noisy scalar field (right). The four main features are represented with long bars in the persistence diagrams. In the noisy diagram, the noise in the scalar field is encoded by small bars near the diagonal.

probability measures to persistence diagrams, we show how barycenter diagrams can be reliably estimated, for generic  $W_q$  distances (Sec. III), despite outliers (Sec. IV).

- 2) *An application to clustering:* We present an application to clustering (Sec. IV-A), where our work yields an improved robustness to the outlier diagrams that are naturally present in ensembles used previously in the literature.
- 3) *An application to Wasserstein dictionary encoding:* We present an application to dictionary encoding (Sec. IV-B), where the added robustness of our generalized barycenters is demonstrated over standard barycenters.
- 4) *Implementation:* We provide a Python implementation of our work that can be used for reproducibility purposes.

## II. PRELIMINARIES

This section presents the required theoretical foundations to our work. We introduce persistence diagrams (Sec. II-A) and the usual metric used in topological data analysis (Sec. II-B).

### A. Persistence diagrams

Given a scalar field  $f$  on a  $(d_{\mathcal{M}})$ -manifold  $\mathcal{M}$ , with  $d_{\mathcal{M}} \leq 3$  in our applications, we denote  $f_{-\infty}^{-1}(w) = f^{-1}([-\infty, w])$  the sub-level set of  $f$  at value  $w \in \mathbb{R}$ . While sweeping  $w$  from  $-\infty$  to  $+\infty$ , the topology of the set  $f_{-\infty}^{-1}(w)$  changes at specific values  $w = f(c)$ , associated to the *critical points* of  $f$ . These points are classified by their index  $\mathcal{J}$ : 0 for minimal, 1 for 1-saddles, ...,  $d_{\mathcal{M}} - 1$  for  $(d_{\mathcal{M}} - 1)$ -saddles, and  $d_{\mathcal{M}}$  for maxima. The Elder rule [25] states that each topological feature is associated to a pair of critical points  $(b, d) = (f(c), f(c'))$ , with  $f(c) < f(c')$ , representing its *birth*  $b$  and *death*  $d$  in the sweep. These pairs can be visually represented as vertical bars in  $\mathbb{R}^2$ , with horizontal coordinate  $b$  and vertical coordinates  $b$  and  $d$ , where  $b - d$  encodes the lifespan of the associated topological feature. This representation is the so-called *Persistence Diagram*. Formally, a persistence diagram is the union of a finite set of 2D points  $X = \{x = (b, d) \in \mathbb{R}^2 \mid b < d\}$ , along with the diagonal  $\Delta = \{(b, b) \mid b \in \mathbb{R}\}$  of  $\mathbb{R}^2$  (the diagonal is not stored in any way in practice). In most cases, important topological features stand out from the diagonal while noise can be typically found near the diagonal (Fig. 1). In the following, we will simply denote a persistence diagram by  $X = \{x_1, \dots, x_K\}$  with  $K \in \mathbb{N}^*$ .

### B. Wasserstein distance and barycenter

The Wasserstein distance is widely used to compare persistence diagrams. In topological data analysis, in order to use this

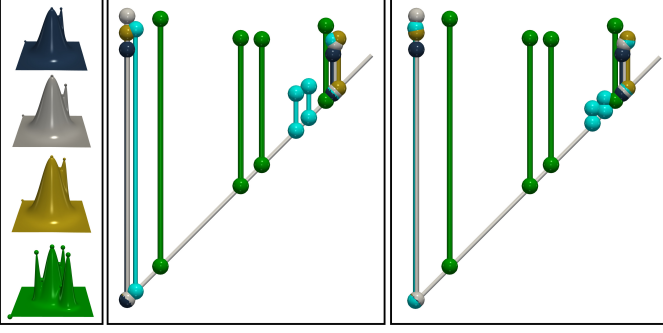


Fig. 2. Comparison of barycenters computed with different values of  $q$ . On the left we have terrain views of four scalar fields colored in blue, gray, yellow and green, the latter being an outlier (featuring more peaks). The corresponding persistence diagrams are represented with matching colors and the barycenters are represented in cyan. The barycenter with  $q = 2$  (center) is more sensitive to the presence of the green outlier, with two cyan bars of medium persistence, due to the outlier peaks in the green dataset. For  $q = 1.5$  (right), the persistence of these two bars is significantly reduced, and so will be their importance in distance computations.

distance, a typical pre-processing involves augmenting the persistence diagrams. Formally, let us consider  $X = \{x_1, \dots, x_{K_1}\}$  and  $Y = \{y_1, \dots, y_{K_2}\}$ , two persistence diagrams. Taking a point  $x = (b, d) \in \mathbb{R}^2$ , we denote by  $\pi_\Delta(x) = (\frac{b+d}{2}, \frac{b+d}{2})$  its projection on  $\Delta$ . We consider  $\Delta_X$  and  $\Delta_Y$  the diagonal projections of the points of  $X$  and  $Y$  respectively. We finish by taking  $X' = X \cup \Delta_Y$  and  $Y' = Y \cup \Delta_X$ . This results in  $|X'| = |Y'| = K$ . For the remainder of this paper, we consider that persistence diagrams are always augmented this way.

Then, for two persistence diagrams  $X$  and  $Y$ , the  $q$ -Wasserstein distance  $W_q$  is defined as:

$$W_q(X, Y) = \min_{\phi: X \rightarrow Y} \left( \sum_{\ell=1}^K c_q(x_\ell, \phi(x_\ell)) \right)^{1/q}, \quad (1)$$

where  $\phi: X \rightarrow Y$  is a bijection between  $X$  and  $Y$ . The *transportation cost*  $c_q$ , based on powered distances, is such that  $c_q(x, y) = \|x - y\|_2^q$  if  $x \notin \Delta$  or  $y \notin \Delta$ , and 0 otherwise. An optimal bijection  $\phi^*$  minimizing Eq. 1 is called an optimal transport plan. In practice,  $q$  is often set to 2, yielding the Wasserstein distance, noted  $W_2$ . A Wasserstein  $W_q$  barycenter of persistence diagrams  $X_1, \dots, X_m$  [78], is defined by minimizing the Fréchet energy:

$$\arg \min_B \sum_{i=1}^m \lambda_i W_q^q(B, X_i), \quad (2)$$

where  $\lambda_i \geq 0$  and  $\sum_i \lambda_i = 1$ . Practical algorithms have been proposed for the computation of Wasserstein barycenters [78], [80], but only for the specific case where  $q = 2$ . When  $q = 2$ , a solution  $B^*$  of Eq. 2 has the following property: a point  $x \in B^*$  is an arithmetic mean of  $m$  points each in  $X_1, \dots, X_m$ , which considerably eases the optimization of Eq. 2. Fig. 2 (center) presents an example of a  $W_2$  barycenter. A barycenter computed using the  $W_2$  distance emulates the behavior of a mean of an ensemble of scalars. As such, it is prone to the influence of outliers. This motivates the use of a more general framework for computing robust barycenters (detailed in Sec. III).

---

#### Algorithm 1: Barycenter computation algorithm.

---

**Input:** Set of persistence diagrams  $\{X_1, \dots, X_m\}$ , barycentric weights  $\lambda_1, \dots, \lambda_m$ , and iteration number  $T$ .

**Output :** Wasserstein barycenter  $B^{(T)} = \{x_1^{(T)}, \dots, x_K^{(T)}\}$ .

**Initialization:**  $B^{(0)} = X_1$ .

**for**  $0 \leq t \leq T - 1$  **do**

// 1. Assignment step

**for**  $i \in \{1, \dots, m\}$  **do**

Compute  $\phi_i \in \arg \min_{\phi: B^{(t)} \rightarrow X_i} \sum_{\ell=1}^K c_q(x_\ell^{(t)}, \phi(x_\ell^{(t)}))$ .

**end**

// 2. Update step

**for**  $\ell \in \{1, \dots, K\}$  **do**

// Ground barycenter computation

Find  $x_\ell^{(t+1)} = \mathbf{b}_q(\phi_1(x_\ell^{(t)}), \dots, \phi_m(x_\ell^{(t)}))$ .

**end**

**end**

---

### III. ROBUST BARYCENTER

This section presents a general framework for computing barycenters of persistence diagrams using  $W_q$  distances, for arbitrary  $q$  values such that  $q > 1$ .

#### A. Optimization

The optimization of Eq. 2 can be addressed by an iterative algorithm Alg. 1, where each iteration involves two steps. First, an *assignment step* computes the optimal assignment given the  $W_q$  metric between each input diagram and the current barycenter estimation. Second, in the *update step*, the Fréchet energy is minimized by computing, for each barycenter point, its *ground barycenter* in the birth/death plane. This is achieved by updating each barycenter point to its optimal location, given the assignments computed in the previous step.

For  $q = 2$ , the ground barycenter can be simply obtained by computing, for each barycenter point, the arithmetic means of its assigned points in the input diagrams [78], [80]. However, for  $q \neq 2$ , such a simple update procedure cannot be considered. As illustrated in Fig. 3, an update based on the arithmetic mean may increase the Fréchet energy for  $q \neq 2$ , hence potentially preventing Alg. 1 from converging toward a satisfying result.

Instead, to generalize the computation of ground barycenters for  $q \neq 2$ , we consider the following function, representing the ground barycenter in the birth/death plane:

$$\mathbf{b}_q: \begin{cases} \mathbb{R}^2 \times \dots \times \mathbb{R}^2 \rightarrow \mathbb{R}^2 \\ (y_1, \dots, y_m) \mapsto \arg \min_x \sum_{i=1}^m \lambda_i c_q(x, y_i) \end{cases} \quad (3)$$

To optimize Eq. 3, we use gradient descent, by leveraging the automatic differentiation capabilities of PyTorch. This optimization procedure is plugged into Alg. 1 in the *update step* (ground barycenter computation line). In practice we give a maximum number  $T$  of overall iterations. We noticed that

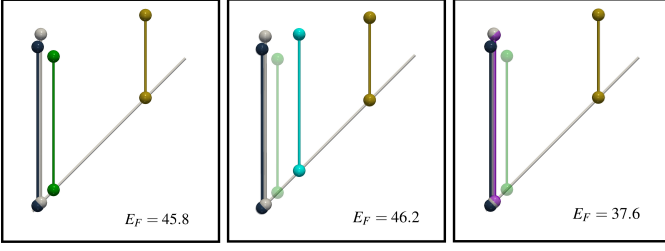


Fig. 3. Simple example where computing the arithmetic mean instead of optimizing  $b_q$  increases the Fréchet energy (noted  $E_F$ ) for  $q = 1$ . We have three simple persistence diagrams, in dark blue, gray and yellow, each having a single point. For this problem, the transport plans are fixed and the barycenter has only one point. On the left we initialized the barycenter as the diagram encoded in green. In the middle, we have the candidate of the barycenter encoded in cyan when computing an arithmetic mean after one iteration. We can see that the Fréchet energy (for  $q = 1$ ) increased. On the right, we have a candidate for the barycenter encoded in purple when optimizing  $b_q$  instead, this time displaying a decrease of the Fréchet energy at one iteration.

taking  $T < 10$  is sufficient for convergence. Assumptions for achieving convergence are discussed in the next section.

### B. Convergence

Tanguy et al. prove, in the setup of probability measures, that a fixed-point method for minimizing Eq. 3 converged under certain assumptions [74]. In this section, we review these assumptions in the setup of persistence diagrams to argue the convergence of our overall approach.

**Assumption 1 :** For all  $(y_1, \dots, y_m) \in \mathbb{R}^2 \times \dots \times \mathbb{R}^2$ , for all  $\lambda_1, \dots, \lambda_m$  barycentric coefficients,  $\arg \min_x \sum_{i=1}^m \lambda_i c_q(x, y_i)$  is reduced to a single element.

We prove that this assumption is satisfied in our case in Appendix A, when  $q > 1$ . When  $q = 1$ , this assumption does not hold in general (due to the presence of collinear points, in particular on the diagonal). In practice, this may lead to numerical instabilities when considering  $q = 1$ , especially when ground barycenters are not unique.

Under those assumptions, and denoting the Fréchet energy  $E_F(B) = \sum_{i=1}^m \lambda_i W_q^q(B, X_i)$ , Tanguy et al. [74] show that for two consecutive iterates  $B^{(t)}$  and  $B^{(t+1)}$ , we have  $E_F(B^{(t)}) \geq E_F(B^{(t+1)})$ . This means that fixed-point iterations  $(B^{(t)})$  decrease the energy optimized in Eq. 2.

For  $q \in (1, 2)$ , a resulting fixed point is a barycenter that is more robust to the presence of an outlier, in the initial set  $X_1, \dots, X_m$ , compared to a  $W_2$  barycenter. Fig. 2 illustrates this difference when computing a barycenter with an outlier.

## IV. RESULTS

This section presents two applications of our robust barycenters (Sec. III) along with detailed experiments. The experimental results are obtained on a computer with an NVIDIA GeForce RTX 2060 (Mobile Q) with 6 GB of dedicated VRAM. Our methods were implemented on Python, using Pytorch for computations on the GPU. We ran some experiments on two public ensembles [62] described in [63]. One is an acquired

2D ensemble and the other is a simulated 3D ensemble, both selected from past SciVis contests [59]. For the experiments, only the persistence pairs containing maxima were considered.

### A. Clustering on the persistence diagram metric space

The first natural application of our robust barycenters consists in using them for the problem of clustering an ensemble of persistence diagrams  $X_1, \dots, X_N$ . In particular, clustering methods on ensembles of persistence diagrams group together subset of members that have similar topological structures, highlighting trends of topological features in the ensemble.

For this, we consider the classic clustering method, the  $k$ -means algorithm. This is an iterative algorithm alternating between two phases: *computing  $k$  barycenters* and *labeling the elements into  $k$  clusters*. At first,  $k$  cluster barycenters  $B_j$  with  $j \in \{1, \dots, k\}$  are initialized as  $k$  diagrams in the initial input set  $X_1, \dots, X_N$ , typically using the  $k$ -means++ method [17]. Then, the labeling phase consists in assigning each diagrams  $X_n$  to the closest barycenter  $B_j$  by using the  $W_q$  distance. After the labeling phase, the barycenters are updated by computing new barycenters based on the new  $k$  clusters using Alg. 1. The algorithm stops when reaching a maximum number of iterations or when converging (i.e., when the labels do not change anymore). However, when using  $W_2$  as a distance, the presence of an outlier in an ensemble of persistence diagrams can incorrectly influence the barycenters in the  $k$ -means algorithm, and as a consequence the output labels of the clustering.

We leverage the robustness of the  $W_q$  barycenters for clustering problems when there are outliers in the ensemble to be clustered. We show a case in Fig. 4 where we artificially injected outlier pixels in an ensemble of synthetic scalar fields (a common degradation in real-life noisy data). This results in outlier diagrams in the input ensemble. We illustrate the clustering results with different  $q \in \{2, 1.8, 1.6, 1.4, 1.2, 1\}$ . This experiment shows that for  $q$  lower than 1.6, the resulting clustering put together the outliers into the correct groups, while for  $q = 2$  the outliers are incorrectly clustered. This can also be seen in Fig. 5 where an outlier is naturally present in the ensemble of scalar fields. In this example, our generic barycenters enable the computation of the correct clustering, as off  $q = 1.6$ . Moreover, as illustrated on the right of Fig. 5 for the last cluster, our generic barycenters are more representative, visually, of the input diagrams as the importance of outlier persistence pairs is decreased in our framework.

### B. Wasserstein dictionary encoding of persistence diagrams

Another application consists in using our robust barycenters as a core procedure for dictionary based encodings of ensembles of persistence diagrams [71]. Let  $X_1, \dots, X_N$  be an ensemble of persistence diagrams. A Wasserstein dictionary encoding aims at optimizing a set of persistence diagrams  $\mathcal{D}^* = \{a_1^*, \dots, a_m^*\}$  (called dictionary) and  $N$  vectors of barycentric coefficients  $\Lambda^* = \{\lambda_1^*, \dots, \lambda_N^*\}$  (i.e.,  $N$  vectors of size  $m$ , with positive elements summing to 1) by solving:

$$\arg \min_{\mathcal{D}, \Lambda} \sum_{\ell=1}^N W_2^2(B_2(\mathcal{D}, \lambda_\ell), X_\ell), \quad (4)$$



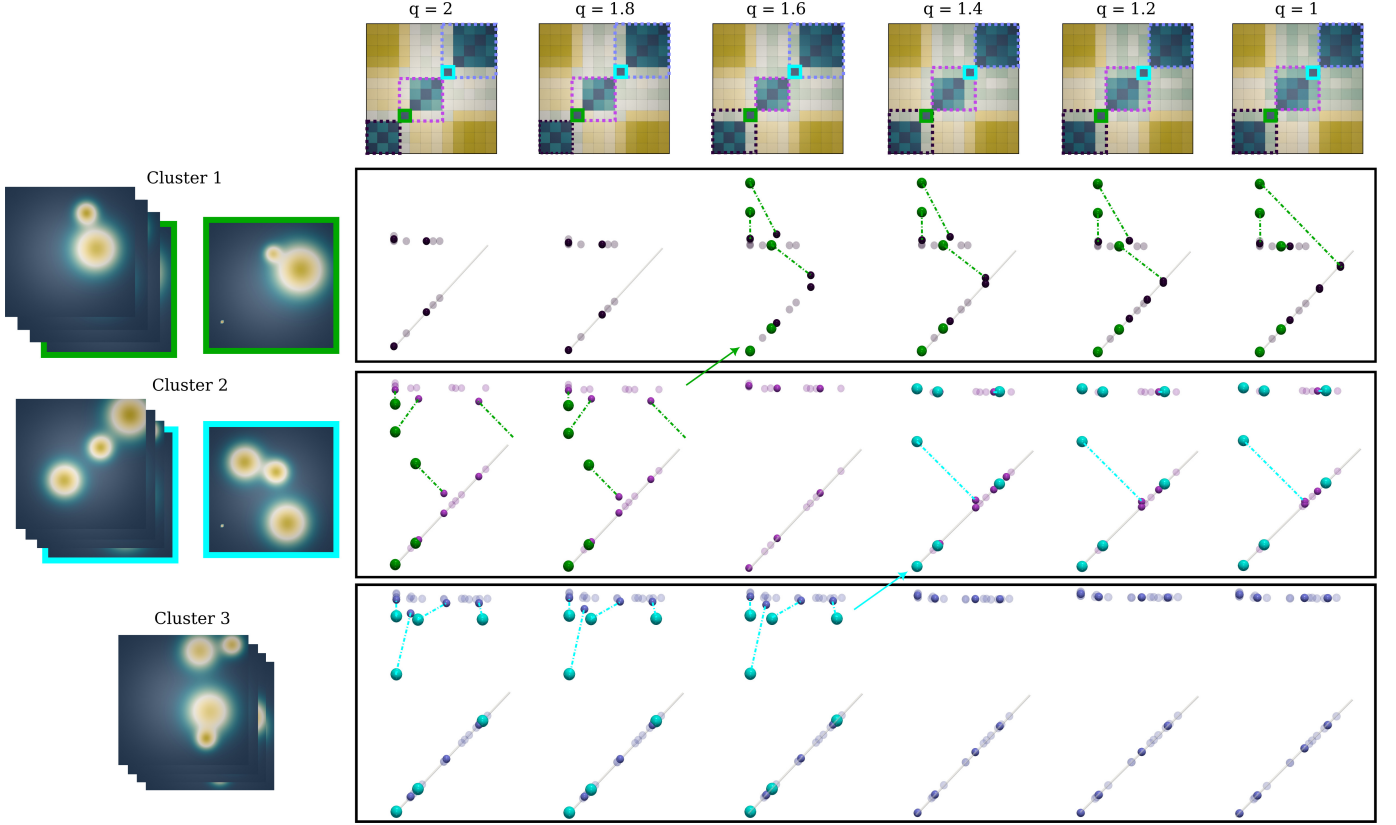


Fig. 4. Comparison of clustering results on an ensemble of diagrams of Gaussian mixtures. On the left we have the 3 clusters: one cluster of 2 Gaussians (top), one cluster of 3 Gaussians (middle) and one cluster of 4 Gaussians (bottom). In the first and second clusters, we inserted an outlier (highlighted in green and cyan respectively) by setting one isolated pixel to an arbitrarily high value. Those pixels result in persistent pairs in the corresponding diagrams. On top we have the distance matrices of  $W_q$  for  $q \in \{2, 1.8, 1.6, 1.4, 1.2, 1\}$ . In the distance matrices, the clustering results are shown with dashed squares (clusters are colored in dark purple, purple and pale purple) while the outlier diagrams are indicated with a plain square (green and cyan). In the three frames; we visualize the evolution of each cluster and their barycenters for each  $q$ . Each frame corresponds to a cluster (top: cluster 1, middle: cluster 2, bottom: cluster 3). The outlier diagrams are colored in green and cyan. The barycenters are shown in opaque while the diagrams of each cluster are shown in transparent. We observe that for  $q \in \{2, 1.8\}$  the green outlier is incorrectly assigned to the second cluster (as it exhibits the same number of persistence pairs, 3, as the entries of cluster 2). Similarly, given its number of persistence pairs, the cyan outlier is incorrectly assigned to the third cluster until  $q = 1.6$ . Beyond this value, the effect of the extra feature in these outlier diagrams is decreased, enabling their correct clustering.

where  $B_2(\mathcal{D}, \lambda_\ell)$  denotes a  $W_2$  barycenter of  $\mathcal{D}$  under barycentric coefficients  $\lambda_\ell$ . Informally, this framework works as a lossy compression for persistence diagrams. The goal is to optimize a smaller set of persistence diagrams ( $m \ll N$ ) and  $N$  vectors of barycentric weights such that the  $N$  Wasserstein barycenters defined by the barycentric weights are good approximations of the  $N$  input diagrams. This results in an encoding of much smaller size as only the dictionary and the  $N$  barycentric weights need to be stored to disk. This framework has two main applications: data reduction and dimensionality reduction. Naturally this framework can be extended to other Wasserstein distances. Then, Eq. 4 becomes:

$$\arg \min_{\mathcal{D}, \Lambda} \sum_{l=1}^N W_q^q(B_q(\mathcal{D}, \lambda_l), X_l), \quad (5)$$

where  $B_q(\mathcal{D}, \lambda_l)$  denotes a  $W_q$  barycenter returned by Alg. 1. For  $q = 2$ , Sisouk et al. [71] introduced the analytic expression of the gradient  $\nabla B_2(\mathcal{D}, \lambda_l)$  with respect to  $a_1, \dots, a_m$  and  $\lambda_l$ , enabling a simple gradient descent scheme for the optimization of Eq. 5. However, for  $q \neq 2$ , since ground barycenters are no longer obtained as arithmetic means, but by an iterative, fixed-point method (Sec. III-B), the gradient of the energy associated

with Eq. 5 cannot be derived analytically. Instead, we rely on automatic differentiation (implemented in PyTorch) and use Adam [46] to optimize both the dictionary  $\mathcal{D} = \{a_1, \dots, a_m\}$  and the vectors of barycentric coefficients  $\Lambda = \{\lambda_1, \dots, \lambda_N\}$ .

This extension results in a Wasserstein dictionary method that is more stable to the presence of outliers in the original input ensemble. Moreover, this extension lets us improve the initialization method for the dictionary. At first, the dictionary was chosen as the  $m$  elements that are the farthest to each other in the initial set [71]. But extending Eq. 4 to Eq. 5 allowed the dictionary to be initialized as barycenters issued from a  $k$ -means algorithm with  $k = m$ . From our experience, this initialization results in a better optimized energy (Eq. 5) and stability to the presence of outliers compared to the initial one [71]. We showcase this extension by applying a Wasserstein dictionary method using  $W_q$ . Fig. 6 shows a comparison of 2D planar layout of the barycenters, generated by the dictionaries and vectors optimized with Eq. 4 and Eq. 5, on the *Isabel* ensemble where we removed some entries to artificially inject an outlier (see the caption of Fig. 6 for more details). Specifically, this 2D planar layout is a direct application of the dictionary encoding when the dictionary has three atoms in it. After optimizing

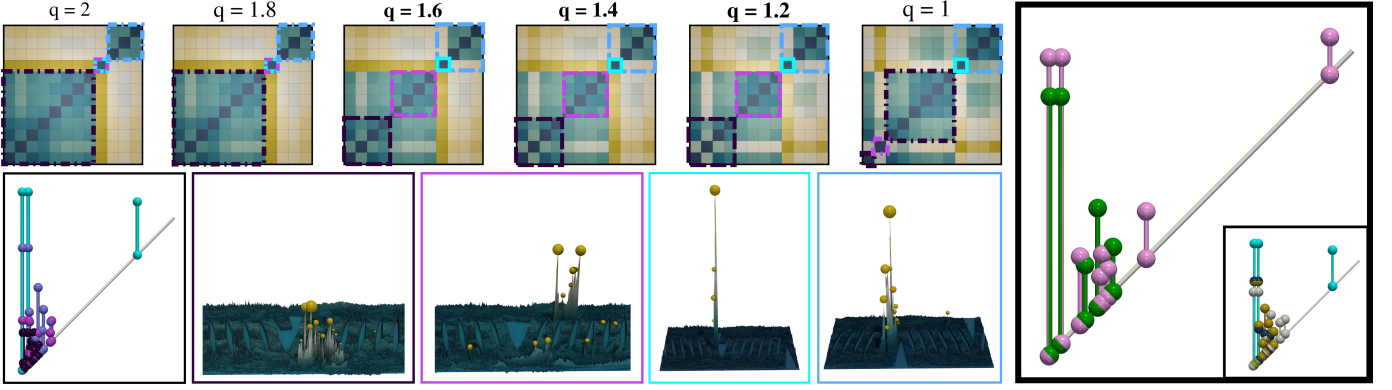


Fig. 5. Visual comparison of distance matrices using  $W_q$  for  $q \in \{2, 1.8, 1.6, 1.4, 1.2, 1\}$  on the *Volcanic Eruption* ensemble and the clustering results. Distance matrices are represented similarly to Fig. 4. This ensemble of 12 persistence diagrams has a natural outlier highlighted in cyan on the distance matrices. On the top, we can see that for  $q \in \{2, 1.8\}$ , the clustering algorithm keeps the outlier alone, groups the 8 first diagrams together and groups the last three together. Then starting from 1.6 to 1.2, the correct clusters are returned. But for  $q = 1$ , we can see that the clusters are not discriminated enough. On the bottom we have one representative scalar field for each cluster, and on the bottom left the corresponding diagrams, the cyan scalar field and diagram being the outlier. On the right, we have a visual comparison of two barycenters of the last cluster of four diagrams (represented on the bottom right of the square). The pink one encodes a  $W_2$  barycenter, while the green one encodes a  $W_{1.2}$  barycenter. We can see the influence of the outlier in the pink barycenter, as the two global pairs are higher than the green ones testifying the difference of scaling between the outlier (cyan) and the three other diagrams in the cluster. Also, we notice the presence of an isolated pair above the diagonal (pink diagram, rightmost pair) that is generated by the isolated persistent pair in the outlier diagram (cyan diagram, rightmost pair). This outlier pair is not present in the  $W_{1.2}$  barycenter (green diagram), hence making it more visually representative of the ensemble than the  $W_2$  barycenter (pink).

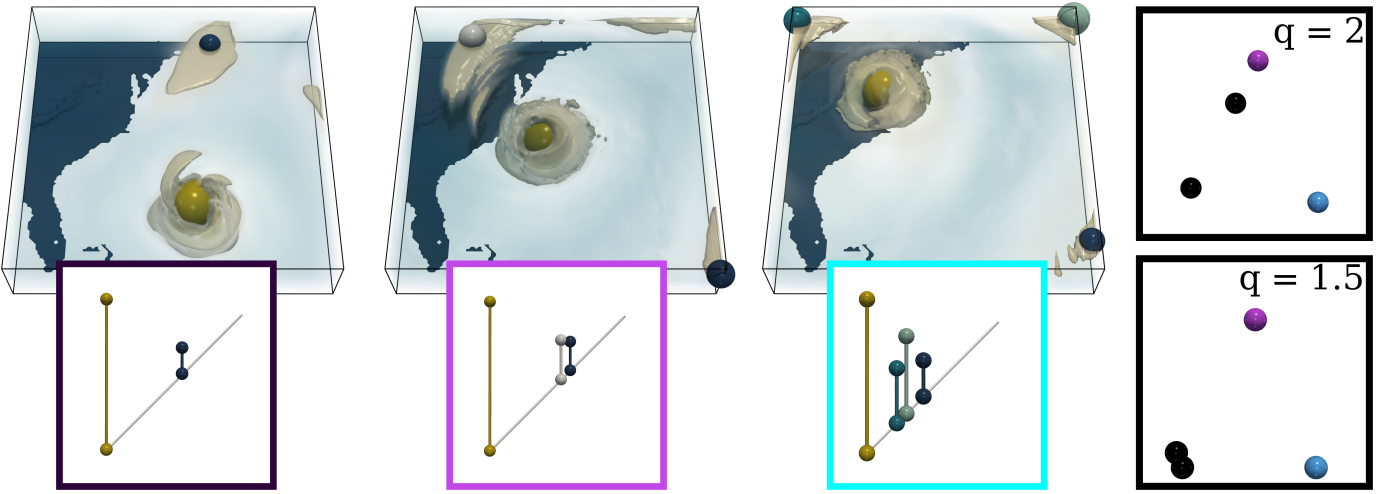


Fig. 6. Visual comparison of 2D planar layouts (on the right) of Wasserstein barycenters after solving Eq. 4 and Eq. 5 when taking a dictionary with 3 diagrams. The initial ensemble *Isabel* is composed of 12 diagrams divided in 3 classes of 4 diagrams each. We removed 2 diagrams from the first and second classes, yielding an imbalanced ensemble (in terms of class size) of 8 diagrams. We show representative scalar field for each cluster, along with their diagram. The clusters are colored in dark purple, purple and cyan respectively. In the planar layouts, the points, representing the barycenters, are colored by their ground truth classification. For  $q = 2$ , the barycenter approximating an element of the first cluster (dark purple) is misplaced (i.e., located near the second cluster, purple). For  $q = 1.5$ , this same barycenter is correctly placed, thus yielding a planar projection that is more faithful to the ground-truth classification.

the dictionary, with the three Wasserstein distances between them, the cosine law can be used to form a triangle in  $\mathbb{R}^2$  and then use the vectors as barycentric coefficients in  $\mathbb{R}^2$ . This experiment shows that this extension is more stable to the presence of outliers, resulting in planar projections that are more faithful to the ground truth classification.

### C. Computation time comparison

In this section, we compare the time needed to compute a barycenter, with regard to the  $W_2$  metric, using the arithmetic mean between points of  $\mathbb{R}^2$ , and the time when computing  $b_2$ . Our experiment consists in computing a barycenter of persistence diagrams for the *Isabel* ensemble (Sec. II-B), where

TABLE I  
RUNNING TIMES (IN SECONDS) FOR COMPUTING A  $W_2$  BARYCENTER.

Method	$m = 4$	$m = 6$
Arithmetic Mean	1.9	24.9
$b_2$ (Pytorch)	6.3	29.0

each diagram is threshold by persistence to feature only  $\sim 100$  pairs and where  $T$  is set to 5. We report the resulting running times in Tab. I. When considering  $m = 4$  input diagrams, the computation based on the arithmetic mean is 3 times faster than the one based on  $b_2$ . However, for a larger ensemble ( $m = 6$ ), the difference is reduced to 4 seconds, yielding compatible run

times between the two approaches.

## V. CONCLUSION

In this paper, we showcased the utility of a method for computing a robust Wasserstein barycenters of persistence diagrams. Specifically, we adapted a recent fixed-point method algorithm [74] to the case of persistence diagrams. We first gave a reminder on this fixed point method framework to compute this robust barycenter. We also gave a formal proof of the necessary hypothesis for the convergence of this method in the Appendix. Then we presented two applications of this robust barycenter to clustering and dictionary encoding of persistence diagrams in the presence of outliers. We believe that our work on the robustness of Wasserstein barycenters is a useful step toward improving their applicability in the analysis of real-life ensembles of persistence diagrams.

A legitimate direction for future work is the generalization of such robust barycenters to other topological descriptors such as merge trees, for which barycentric frameworks derived from the Wasserstein distance have been proposed [63]. Also, we will continue our investigation of the adaption of methods from Optimal Transport to ensembles of topological representations, as we believe it can become a key solution in the long term for the advanced analysis of large-scale ensemble datasets.

## ACKNOWLEDGMENTS

This work is partially supported by the European Commission grant ERC-2019-COG “TORI” (ref. 863464, <https://erc-tori.github.io/>). JD acknowledge the support of the “France 2030” funding ANR-23-PEIA-0004 (“PDE-AI”) and the support of the funding SOCOT - ANR-23-CE40-0017.

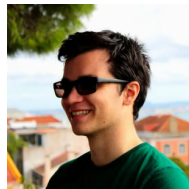
## REFERENCES

- [1] A. Acharya and V. Natarajan. A parallel and memory efficient algorithm for constructing the contour tree. In *IEEE PacificVis*, 2015.
- [2] M. Agueh and G. Carlier. Barycenters in the wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011.
- [3] 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*, 2018.
- [4] T. M. Athawale, D. Maljovec, C. R. Johnson, V. Pascucci, and B. Wang. Uncertainty Visualization of 2D Morse Complex Ensembles Using Statistical Summary Maps. *CoRR*, abs/1912.06341, 2019.
- [5] D. P. Bertsekas. A new algorithm for the assignment problem. *Mathematical Programming*, 21(1):152–171, 1981.
- [6] 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 Comp. Chem.*, 2018.
- [7] S. Biasotti, D. Giorgio, M. Spagnuolo, and B. Falcidieno. Reeb graphs for shape analysis and applications. *TCS*, 2008.
- [8] T. Bin Masood, J. Budin, M. Falk, G. Favelier, C. Garth, C. Gueunet, P. Guillou, L. Hofmann, P. Hristov, A. Kamakshidasan, C. Kappe, P. Klacansky, P. Laurin, J. Levine, J. Lukaszczuk, D. Sakurai, M. Soler, P. Steneteg, J. Tierny, W. Usher, J. Vidal, and M. Wozniak. An Overview of the Topology ToolKit. In *TopoInVis*, 2019.
- [9] A. Bock, H. Doraiswamy, A. Summers, and C. T. Silva. TopoAngler: Interactive Topology-Based Extraction of Fishes. *IEEE TVCG*, 2018.
- [10] P. Bremer, H. Edelsbrunner, B. Hamann, and V. Pascucci. A Multi-Resolution Data Structure for 2-Dimensional Morse Functions. In *Proc. of IEEE VIS*, 2003.
- [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 TVCG*, 2011.
- [12] C. Brizzi, G. Friesecke, and T. Ried.  $p$ -wasserstein barycenters, 2024.
- [13] G. Carlier, E. Chenchene, and K. Eichinger. Wasserstein medians: Robustness, pde characterization, and numerics. *SIAM Journal on Mathematical Analysis*, 56(5):6483–6520, Sept. 2024.
- [14] H. Carr, J. Snoeyink, and U. Axen. Computing contour trees in all dimensions. In *Symp. on Dis. Alg.*, 2000.
- [15] H. Carr, G. Weber, C. Sewell, and J. Ahrens. Parallel peak pruning for scalable SMP contour tree computation. In *IEEE LDAV*, 2016.
- [16] H. A. Carr, J. Snoeyink, and M. van de Panne. Simplifying Flexible Isosurfaces Using Local Geometric Measures. In *IEEE VIS*, 2004.
- [17] M. E. Celebi, H. A. Kingravi, and P. A. Vela. A comparative study of efficient initialization methods for the k-means clustering algorithm. *Expert Syst. Appl.*, 2013.
- [18] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. In *SoCG*, 2005.
- [19] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. *Discrete & Computational Geometry*, 2007.
- [20] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *NIPS*, 2013.
- [21] M. Cuturi and A. Doucet. Fast computation of wasserstein barycenters. In *ICML*, 2014.
- [22] L. De Floriani, U. Fugacci, F. Iuricich, and P. Magillo. Morse complexes for shape segmentation and homological analysis: discrete models and algorithms. *CGF*, 2015.
- [23] P. Diggle, P. Heagerty, K.-Y. Liang, and S. Zeger. *The Analysis of Longitudinal Data*. Oxford University Press, 2002.
- [24] H. Doraiswamy and V. Natarajan. Computing Reeb Graphs as a Union of Contour Trees. *IEEE TVCG*, 2013.
- [25] H. Edelsbrunner and J. Harer. *Computational Topology: An Introduction*. American Mathematical Society, 2009.
- [26] H. Edelsbrunner, J. Harer, V. Natarajan, and V. Pascucci. Morse-Smale complexes for piecewise linear 3-manifolds. In *SoCG*, 2003.
- [27] H. Edelsbrunner, J. Harer, and A. Zomorodian. Hierarchical morse complexes for piecewise linear 2-manifolds. In *SoCG*, 2001.
- [28] G. Favelier, N. Faraj, B. Summa, and J. Tierny. Persistence Atlas for Critical Point Variability in Ensembles. *IEEE TVCG*, 2018.
- [29] G. Favelier, C. Gueunet, and J. Tierny. Visualizing ensembles of viscous fingers. In *IEEE SciVis Contest*, 2016.
- [30] F. Ferstl, K. Bürger, and R. Westermann. Streamline variability plots for characterizing the uncertainty in vector field ensembles. *IEEE TVCG*, 2016.
- [31] F. Ferstl, M. Kanzler, M. Rautenhaus, and R. Westermann. Visual analysis of spatial variability and global correlations in ensembles of iso-contours. *CGF*, 2016.
- [32] R. Flamary, N. Courty, A. Gramfort, M. Z. Alaya, A. Boisbunon, S. Chambon, L. Chapel, A. Corenflos, K. Fatras, N. Fournier, L. Gautheron, N. T. Gayraud, H. Janati, A. Rakotomamonjy, I. Redko, A. Rolet, A. Schutz, V. Seguy, D. J. Sutherland, R. Tavenard, A. Tong, and T. Vayer. Pot: Python optimal transport. *Journal of Machine Learning Research*, 22(78):1–8, 2021.
- [33] R. Forman. A User’s Guide to Discrete Morse Theory. *AM*, 1998.
- [34] D. Guenther, R. Alvarez-Boto, J. Contreras-Garcia, J.-P. Piquemal, and J. Tierny. Characterizing Molecular Interactions in Chemical Systems. *IEEE TVCG*, 2014.
- [35] C. Gueunet, P. Fortin, J. Jomier, and J. Tierny. Task-Based Augmented Contour Trees with Fibonacci Heaps. *IEEE TPDS*, 2019.
- [36] C. Gueunet, P. Fortin, J. Jomier, and J. Tierny. Task-based Augmented Reeb Graphs with Dynamic ST-Trees. In *EGPGV*, 2019.
- [37] A. Gyulassy, P. Bremer, R. Grout, H. Kolla, J. Chen, and V. Pascucci. Stability of dissipation elements: A case study in combustion. *CGF*, 2014.
- [38] A. Gyulassy, P. Bremer, and V. Pascucci. Shared-Memory Parallel Computation of Morse-Smale Complexes with Improved Accuracy. *IEEE TVCG*, 2019.
- [39] A. Gyulassy, M. A. Duchaineau, V. Natarajan, V. Pascucci, E. Bringa, A. Higginbotham, and B. Hamann. Topologically Clean Distance Fields. *IEEE TVCG*, 2007.
- [40] 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 TVCG*, 2016.
- [41] 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. *CGF*, 2016.
- [42] M. Hummel, H. Obermaier, C. Garth, and K. I. Joy. Comparative visual analysis of lagrangian transport in CFD ensembles. *IEEE TVCG*, 2013.
- [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 TVCG*, 2011.
- [45] M. Kerber, D. Morozov, and A. Nigmatov. Geometry helps to compare persistence diagrams. *ACM J. of Experimental Algorithmics*, 22, 2017.
- [46] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization, 2017.
- [47] M. Kraus. Visualization of uncertain contour trees. In *IVTA*, 2010.
- [48] T. Lacombe, M. Cuturi, and S. Oudot. Large Scale computation of Means and Clusters for Persistence Diagrams using Optimal Transport. In *NIPS*, 2018.
- [49] D. E. Laney, P. Bremer, A. Mascarenhas, P. Miller, and V. Pascucci. Understanding the structure of the turbulent mixing layer in hydrodynamic instabilities. *IEEE TVCG*, 2006.
- [50] M. Li, S. Palande, L. Yan, and B. Wang. Sketching merge trees for scientific visualization. In *IEEE TopInVis*, 2023.
- [51] S. Maadasamy, H. Doraiswamy, and V. Natarajan. A hybrid parallel algorithm for computing and tracking level set topology. In *HiPC*, 2012.
- [52] 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*, 2016.
- [53] M. Mirzargar, R. Whitaker, and R. Kirby. Curve boxplot: Generalization of boxplot for ensembles of curves. *IEEE TVCG*, 20(12):2654–2663, 2014.
- [54] G. Monge. Mémoire sur la théorie des déblais et des remblais. *Académie Royale des Sciences de Paris*, 1781.
- [55] J. R. Munkres. *Elements of Algebraic Topology*. Addison-Wesley, 1984.
- [56] F. Nauleau, F. Vivodtzev, T. Bridel-Bertomeu, H. Beaugendre, and J. Tierny. Topological Analysis of Ensembles of Hydrodynamic Turbulent Flows – An Experimental Study. In *IEEE LDAV*, 2022.
- [57] M. Olejniczak, A. S. P. Gomes, and J. Tierny. A Topological Data Analysis Perspective on Non-Covalent Interactions in Relativistic Calculations. *International Journal of Quantum Chemistry*, 2019.
- [58] M. Olejniczak and J. Tierny. Topological Data Analysis of Vortices in the Magnetically-Induced Current Density in LiH Molecule. *Physical Chemistry Chemical Physics*, 2023.
- [59] Organizers. The IEEE SciVis Contest. <http://sciviscontest.ieeevis.org/>, 2004.
- [60] S. Parsa. A deterministic  $O(m \log m)$  time algorithm for the reeb graph. In *SoCG*, 2012.
- [61] V. Pascucci, G. Scorzelli, P. T. Bremer, and A. Mascarenhas. Robust on-line computation of Reeb graphs: simplicity and speed. *ACM ToG*, 2007.
- [62] M. Pont, J. Vidal, J. Delon, and J. Tierny. Wasserstein Distances, Geodesics and Barycenters of Merge Trees – Ensemble Benchmark. <https://github.com/MatPont/WassersteinMergeTreesData>, 2021.
- [63] M. Pont, J. Vidal, J. Delon, and J. Tierny. Wasserstein Distances, Geodesics and Barycenters of Merge Trees. *IEEE TVCG*, 2022.
- [64] M. Pont, J. Vidal, and J. Tierny. Principal geodesic analysis of merge trees (and persistence diagrams). *IEEE TVCG*, 2023.
- [65] K. Potter, A. Wilson, P. Bremer, D. Williams, C. Doutriaux, V. Pascucci, and C. R. Johnson. Ensemble-vis: A framework for the statistical visualization of ensemble data. In *2009 IEEE ICDM*, 2009.
- [66] V. Robins, P. J. Wood, and A. P. Sheppard. Theory and Algorithms for Constructing Discrete Morse Complexes from Grayscale Digital Images. *IEEE PAMI*, 2011.
- [67] F. Santambrogio. Optimal transport for applied mathematicians. *Birkhäuser, NY*, 55(58-63):94, 2015.
- [68] J. Sanyal, S. Zhang, J. Dyer, A. Mercer, P. Amburn, and R. Moorhead. Noodles: A tool for visualization of numerical weather model ensemble uncertainty. *IEEE TVCG*, 2010.
- [69] N. Shivashankar and V. Natarajan. Parallel Computation of 3D Morse-Smale Complexes. *CGF*, 2012.
- [70] 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 TVCG*, 2016.
- [71] K. Sisouk, J. Delon, and J. Tierny. Wasserstein Dictionaries of Persistence Diagrams. *IEEE Transactions on Visualization & Computer Graphics*, 30(02):1638–1651, Feb. 2024.
- [72] M. Soler, M. Petitfrère, G. Darche, M. Plainchault, B. Conche, and J. Tierny. Ranking Viscous Finger Simulations to an Acquired Ground Truth with Topology-Aware Matchings. In *IEEE LDAV*, 2019.
- [73] T. Sousbie. The Persistent Cosmic Web and its Filamentary Structure: Theory and Implementations. *Royal Astronomical Society*, 2011.
- [74] E. Tanguy, J. Delon, and N. Gozlan. Computing barycentres of measures for generic transport costs, 2025.
- [75] S. Tarasov and M. Vyalı. Construction of contour trees in 3D in  $O(n \log n)$  steps. In *SoCG*, 1998.
- [76] J. Tierny, A. Gyulassy, E. Simon, and V. Pascucci. Loop surgery for volumetric meshes: Reeb graphs reduced to contour trees. *IEEE TVCG*, 2009.
- [77] K. Turner. Medians of populations of persistence diagrams, 2019.
- [78] K. Turner, Y. Mileyko, S. Mukherjee, and J. Harer. Fréchet Means for Distributions of Persistence Diagrams. *DCG*, 2014.
- [79] J. Vidal, J. Budin, and J. Tierny. Progressive wasserstein barycenters of persistence diagrams. *IEEE TVCG*, 26(1):151–161, 2020.
- [80] J. Vidal, J. Budin, and J. Tierny. Progressive Wasserstein Barycenters of Persistence Diagrams. *IEEE TVCG*, 2020.
- [81] C. Villani. *Optimal transport: old and new*. Springer Verlag, 2008.
- [82] R. T. Whitaker, M. Mirzargar, and R. M. Kirby. Contour boxplots: A method for characterizing uncertainty in feature sets from simulation ensembles. *IEEE TVCG*, 2013.
- [83] D. P. Woodruff. *Sketching as a Tool for Numerical Linear Algebra*. Now Publishers, 2014.
- [84] K. Wu and S. Zhang. A contour tree based visualization for exploring data with uncertainty. *IJUQ*, 2013.



**Keanu Sisouk** is a Ph.D. student at Sorbonne University. He received his master's degree in Mathematics from Sorbonne University in 2021. His fields of interests lie on topological methods for data analysis, optimal transport, optimization methods, statistics and partial derivative equations.



**Eloi Tanguy** is a Ph.D. student at Université Paris-Cité. He received his master's degree in Mathematics from ENS Paris-Saclay in 2022 after his studies at Ecole Polytechnique. His research interests are primarily in optimal transport, in particular OT Barycentres, the Sliced Wasserstein Distance and GMM-OT.



**Julie Delon** received the Ph.D. degree in Mathematics from the Ecole Normale Supérieure Cachan in 2004. She is currently a professor at ENS Paris since 2025. She was a CNRS researcher affiliated with TELECOM ParisTech and obtained her professor tenure at Paris-Cité University in 2013. Her research interests lie in optimal transport, image processing, inverse problems and stochastic models for image restoration and editing.



**Julien Tierny** received the Ph.D. degree in Computer Science from the University of Lille in 2008. He is currently a CNRS research director, affiliated with Sorbonne University. His research expertise lies in topological methods for data analysis and visualization. He is the founder and lead developer of the Topology ToolKit (TTK), an open source library for topological data analysis.



APPENDIX A  
SUFFICIENT CONDITIONS FOR ASSUMPTION 1

**Conditions 1 :** Let  $y_1, \dots, y_m \in \mathbb{R}^d$  and  $(\lambda_1, \dots, \lambda_m) \in (0, 1)^m$  such that  $\sum_i \lambda_i = 1$ . Then for  $q \in (1, +\infty)$ , the function defined as:

$$V_q := x \mapsto \sum_{i=1}^m \lambda_i \|x - y_i\|_2^q,$$

has a unique minimiser in  $\mathbb{R}^d$ . If  $m \geq 3$  and there exists  $i_1 < i_2 < i_3 \in \{1, \dots, m\}$  such that the points  $y_{i_1}, y_{i_2}$  and  $y_{i_3}$  are not on a common affine line, then  $V_1$  also has a unique minimiser.

*Proof.* — *Step 1:* Case  $q > 1$ .

For  $q > 1$  and a fixed  $y \in \mathbb{R}^d$ , introduce the function  $h_q := x \mapsto \|x - y\|_2^q$ . We begin by showing that  $h_q$  is strictly convex. Take  $x_1, x_2 \in \mathbb{R}^d$  and  $t \in (0, 1)$ . We re-write:

$$h_q(tx_1 + (1-t)x_2) = \|t(x_1 - y) + (1-t)(x_2 - y)\|_2^q.$$

Introduce  $u := x_1 - y$  and  $v := x_2 - y$ . By convexity of  $\|\cdot\|_2$ , we have  $\|tu + (1-t)v\|_2 \leq t\|u\|_2 + (1-t)\|v\|_2$ . We consider the equality and strict inequality cases separately.

- 1) If  $\|tu + (1-t)v\|_2 < t\|u\|_2 + (1-t)\|v\|_2$ , we use consecutively the fact that  $a \mapsto a^q$  is increasing and convex on  $\mathbb{R}_+$ :

$$\begin{aligned} \|tu + (1-t)v\|_2^q &< (t\|u\|_2 + (1-t)\|v\|_2)^q \\ &\leq t\|u\|_2^q + (1-t)\|v\|_2^q. \end{aligned}$$

- 2) If  $\|tu + (1-t)v\|_2 = t\|u\|_2 + (1-t)\|v\|_2$ , equality in the triangle inequality for  $\|\cdot\|_2$  yields that  $u$  and  $v$  are positively co-linear, leading to the two following alternatives:

- a) If  $v = 0$  then we show the strict inequality as follows:

$$\|tu\|_2^q = t^q \|u\|_2^q < t\|u\|_2^q,$$

where the inequality comes from the fact that  $q > 1$  and  $t \in (0, 1)$ , with  $u \neq 0$  (indeed, if  $u = 0$  then we have  $u = v = 0$  yielding  $x_1 = x_2$  which is a contradiction).

- b) If  $v \neq 0$  then there exists  $\alpha \geq 0$  such that  $u = \alpha v$ . This implies that  $\|u\|_2 \neq \|v\|_2$ : if equality held, then since  $\alpha \geq 0$  we obtain  $\alpha = 1$ , then  $u = v$  yields  $x_1 = x_2$  which is a contradiction. Since  $\|u\|_2 \neq \|v\|_2$ , we can apply the strict convexity of  $a \mapsto a^q$  on  $\mathbb{R}_+$ , which shows:

$$\begin{aligned} \|tu + (1-t)v\|_2^q &= (t\|u\|_2 + (1-t)\|v\|_2)^q \\ &< t\|u\|_2^q + (1-t)\|v\|_2^q. \end{aligned}$$

In all cases, we obtain the inequality:

$$h_q(tx_1 + (1-t)x_2) < th_q(x_1) + (1-t)h_q(x_2),$$

showing strict convexity of  $h_q$ . As a convex combination of strictly convex functions,  $V_q$  is strictly convex. Since  $V_q$  is coercive, we conclude that it admits a unique minimiser.

— *Step 2:* Case  $q = 1$ .

We now assume that  $m \geq 3$  and that there exists  $i_1 < i_2 < i_3 \in \{1, \dots, m\}$  such that  $y_{j_1}, y_{j_2}$  and  $y_{j_3}$  are not on a common affine line. We prove that  $V_1$  is strictly convex: let  $x_1 \neq x_2 \in \mathbb{R}^d$

and  $t \in (0, 1)$ , by the triangle inequality (similarly to the case  $q > 1$ ):

$$V_1(tx_1 + (1-t)x_2) \leq \sum_{i=1}^m \lambda_i (\|t(x_1 - y_i)\|_2 + \|(1-t)(x_2 - y_i)\|_2),$$

using the triangle inequality. Our objective is to show that in this case, the inequality is strict. There is equality if and only if for each  $i \in \{1, \dots, m\}$ ,  $x_2 - y_i = 0$  or there exists  $\alpha_i \in \mathbb{R}_+$  such that  $x_1 - y_i = \alpha_i(x_2 - y_i)$ . We now reason by contradiction and assume that equality holds. We distinguish two cases concerning the points  $(y_{i_k})_{k=1}^3$  from the assumption.

- 1) If there exists  $k \in \{1, 2, 3\}$  such that  $x_2 - y_{i_k} = 0$ , without any loss of generality we take  $y_{i_1} = x_2$ , then  $y_{i_1} = x_2$ , furthermore the assumption on  $(y_{i_k})_{k=1}^3$  implies that:

$$\forall i \in \{i_2, i_3\}, x_2 - y_i \neq 0.$$

Using these properties, we deduce from the equality in the triangle inequality that there exists  $\alpha_{i_k} \geq 0$  such that  $x_1 - y_{i_k} = \alpha_{i_k}(x_2 - y_{i_k})$  for  $k \in \{2, 3\}$ . Note that the  $\alpha_{i_k} \neq 1$ , otherwise  $x_1 - y_{i_k} = x_2 - y_{i_k}$  which is impossible since  $x_1 \neq x_2$ . We can now re-write the equality as

$$y_{i_k} = x_1 + \frac{\alpha_{i_k}}{1 - \alpha_{i_k}}(x_1 - x_2),$$

concluding that  $y_{i_2}, y_{i_3}$  and  $x_2$  are on the common affine line  $x_1 + \mathbb{R}(x_1 - x_2)$ , which is a contradiction as  $x_2 = y_{i_1}$ . We conclude that the equality cannot hold in the triangle inequality in this case.

- 2) If  $\forall k \in \{1, 2, 3\}, x_2 - y_{i_k} \neq 0$ , then the triangle equality condition provides the existence of  $\alpha_{i_k} \geq 0$  such that  $x_1 - y_{i_k} = \alpha_{i_k}(x_2 - y_{i_k})$ . Again  $\alpha_{i_k} \neq 1$  for the same reasons discussed above. Similarly we re-write the equality as:

$$y_{i_k} = x_1 + \frac{\alpha_{i_k}}{1 - \alpha_{i_k}}(x_1 - x_2),$$

concluding that the three points  $(y_{i_k})_{k=1}^3$  are on the common affine line  $x_1 + \mathbb{R}(x_1 - x_2)$  which is a contradiction, hence equality cannot hold in the triangle inequality.

In both cases, we have proven that equality in the triangle inequality cannot happen, yielding:

$$V_1(tx + (1-t)x_2) < tV_1(x) + (1-t)V_1(x_2),$$

which shows the strict convexity of  $V_1$ . Since  $V_1$  is coercive, we can conclude that it admits a unique minimiser.  $\square$

**Remark 1 :** When  $q = 1$  it is crucial that at least three points are not on a common affine line. For instance if we have only two points, and take  $\lambda_1 = \lambda_2 = 1/2$ , then  $\arg \min_x \|x - y_1\|_2 + \|x - y_2\|_2 = \{ty_1 + (1-t)y_2 \mid t \in [0, 1]\}$ .