Journal of Applied and Computational Topology (2021) 5:99–140
https://doi.org/10.1007/s41468-020-00063-x

Probabilistic convergence and stability of random mapper graphs

Adam Brown[1] · **Omer Bobrowski**[2] · **Elizabeth Munch**[3] · **Bei Wang**[4]

Received: 21 October 2019 / Accepted: 23 November 2020 / Published online: 17 December 2020

Abstract

We study the probabilistic convergence between the mapper graph and the Reeb graph of a topological space $\mathbb{X}$ equipped with a continuous function $f : \mathbb{X} \to \mathbb{R}$. We first give a categorification of the mapper graph and the Reeb graph by interpreting them in terms of cosheaves and stratified covers of the real line $\mathbb{R}$. We then introduce a variant of the classic mapper graph of Singh et al. (in: Eurographics symposium on point-based graphics, 2007), referred to as the enhanced mapper graph, and demonstrate that such a construction approximates the Reeb graph of $(\mathbb{X}, f)$ when it is applied to points randomly sampled from a probability density function concentrated on $(\mathbb{X}, f)$. Our techniques are based on the interleaving distance of constructible cosheaves and topological estimation via kernel density estimates. Following Munch and Wang (In: 32nd international symposium on computational geometry, volume 51 of Leibniz international proceedings in informatics (LIPIcs), Dagstuhl, Germany, pp 53:1–53:16, 2016), we first show that the mapper graph of $(\mathbb{X}, f)$, a constructible $\mathbb{R}$-space (with a fixed open cover), approximates the Reeb graph of the same space. We then construct an isomorphism between the mapper of $(\mathbb{X}, f)$ to the mapper of a super-level set of a probability density function concentrated on $(\mathbb{X}, f)$. Finally, building on the approach of Bobrowski et al. (Bernoulli **23**(1):288–328, 2017b), we show that, with high probability, we can recover the mapper of the super-level set given a sufficiently large sample. Our work is the first to consider the mapper construction using the theory of cosheaves in a probabilistic setting. It is part of an ongoing effort to combine sheaf theory, probability, and statistics, to support topological data analysis with random data.

Keywords Topological data analysis · Mapper · Computational topology · Constructible cosheaves

Mathematics Subjects Classification: 55N31 · 62R40

Extended author information available on the last page of the article

1 Introduction

In recent years, topological data analysis has been gaining momentum in aiding knowledge discovery of large and complex data. A great deal of work has been focused on data modeled as scalar fields. For instance, scientific simulations and imaging tools produce data in the form of point cloud samples equipped with scalar values, such as temperature, pressure and grayscale intensity. One way to understand and characterize the structure of a scalar field $f : \mathbb{X} \to \mathbb{R}$ is through various forms of topological descriptors, which provide meaningful and compact abstraction of the data. Popular topological descriptors can be classified into vector-based ones such as persistence diagrams (Edelsbrunner et al. 2002) and barcodes (Ghrist 2008; Carlsson et al. 2004), graph-based ones such as Reeb graphs (Reeb 1946) and their variants merge trees (Beketayev et al. 2014) and contour trees (Carr et al. 2003), and complex-based ones such as Morse complexes, Morse–Smale complexes (Gerber and Potter 2012; Edelsbrunner et al. 2003a, b), and the mapper construction (Singh et al. 2007).

For a topological space $\mathbb{X}$ equipped with a function $f : \mathbb{X} \to \mathbb{R}$, the *Reeb graph*, denoted as $\mathcal{R}(\mathbb{X}, f)$, encodes the connected components of the level sets $f^{-1}(a)$ for a ranging over $\mathbb{R}$. It summarizes the structure of the data, represented as a pair $(\mathbb{X}, f)$, by capturing the evolution of the topology of its level sets. Research surrounding Reeb graphs and their variants has been very active in recent years, from theoretical, computational and applications aspects, see Biasotti et al. (2008) for a survey. In the multivariate setting, Reeb spaces (Edelsbrunner et al. 2008) generalize Reeb graphs and serve as topological descriptors of multivariate functions $f : \mathbb{X} \to \mathbb{R}^d$. The Reeb graph is then a special case of a Reeb space for $d = 1$.

One issue with Reeb spaces are their limited applicability to point cloud data. To facilitate their practical usage, a closely related construction called *mapper* (Singh et al. 2007) was introduced to capture the topological structure of a pair $(\mathbb{X}, f)$ (where $f : \mathbb{X} \to \mathbb{R}^d$). Given a topological space $\mathbb{X}$ equipped with a $\mathbb{R}^d$-valued function f, for the classic mapper construction, we work with a finite good cover $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$ of $f(\mathbb{X})$ for some indexing set A, such that $f(\mathbb{X}) \subseteq \bigcup U_\alpha$. Let $f^*(\mathcal{U})$ denote the cover of $\mathbb{X}$ obtained by considering the path-connected components of $f^{-1}(U_\alpha)$ for each α. The mapper construction of $(\mathbb{X}, f)$ is defined to be the nerve of $f^*(\mathcal{U})$, denoted as $\mathcal{N}_{f^*(\mathcal{U})}$, see Fig. 1h for an example. By definition, the *mapper* is an abstract simplicial complex; and its 1-dimensional skeleton is referred to as the classic *mapper graph* in this paper.

As a computable alternative to the Reeb space, the mapper has enjoyed tremendous success in data science, including cancer research (Nicolau et al. 2011) and sports analytics (Alagappan 2012); it is also a cornerstone of several data analytics companies such as Ayasdi and Alpine Data Labs. Many variants have been studied in recent years. The *α-Reeb graph* (Chazal and Sun 2014) redefines the equivalence relation between points using open intervals of length at most α. The *multiscale mapper* (Dey et al. 2016) studies a sequence of mapper constructions by varying the granularity of the cover. The *multinerve mapper* (Carriére and Oudot 2018) computes the multinerve (de Verdiére et al. 2012) of the connected cover. The *Joint Contour Net* (JCN) (Carr and Duke 2013, 2014) introduces quantizations to the cover elements by rounding

the function values. The *extended Reeb graph* (Barral and Biasotti 2014) uses cover elements from a partition of the domain without overlaps.

Although the mapper construction has been widely appreciated by the practitioners, our understanding of its theoretical properties remains fragmentary. Some questions important in theory and in practice center around its structure and its relation to the Reeb graph.

Q1. *Information content* What information is encoded by the mapper? How much information can we recover about the original data from the mapper by solving an inverse problem?
Q2. *Stability* What is the structural stability of the mapper with respect to perturbations of its function, domain and cover?
Q3. *Convergence* What is an appropriate metric under which the mapper converges to the Reeb graph as the number of sampled points goes to infinity and the granularity of the cover goes to zero?

To the best of our knowledge, our work is the first to address convergence in **a probabilistic setting**. Given a mapper construction applied to points randomly sampled from a probability density function, we prove an asymptotic result: as the number of points $n \to \infty$, the mapper graph construction approximates that of the Reeb graph up to the granularity of the cover with high probability.

Information, stability and convergence We discuss our work in the context of related literature in topological data analysis. As many topological descriptors, the mapper summarizes the information from the original data through a *lossy* process. To quantify its information content, Dey et al. (2017) studied the topological information encoded by Reeb spaces, mappers and multi-scale mappers, where 1-dimensional homology of the mapper was shown to be no richer than the domain $\mathbb{X}$ itself. Carriére and Oudot (2018) characterized the information encoded in the mapper using the extended persistence diagram of its corresponding Reeb graph. Gasparovic et al. (2018) provided full descriptions of persistent homology information of a metric graph via its intrinsic Čech complex, a special type of nerve complex. In this paper, we study the information content of the mapper via a (co)sheaf-theoretic approach; in particular, through the notion of *display locale*, we introduce an intermediate object called the *enhanced mapper graph*, that is, a CW complex with weighted 0-cells. We show that the enhanced mapper graph reduces the information loss during summarization and may be of independent interest.

In terms of stability, Carriére and Oudot (2018) derived stability for the mapper graph using the stability of extended persistence diagrams equipped with the bottleneck distance under Hausdorff or Wasserstein perturbations of the data (Cohen-Steiner et al. 2009). Our work is similar to Carriére and Oudot (2018) in a sense that we study the stability of the enhanced mapper graph with respect to perturbation of the data $(\mathbb{X}, f)$, where the local stability depends on how the cover $\mathcal{U}$ is positioned in relation to the critical values of f. However, we formalize the structural stability of the enhanced mapper graph using a categorification of the mapper algorithm and the interleaving distance of constructible cosheaves.

When f is a scalar field and the connected cover of its domain $\mathbb{R}$ consists of a collection of open intervals, the mapper construction is conjectured to recover the

Reeb graph precisely as the granularity of the cover goes to zero (Singh et al. 2007). Babu (2013) studied the above convergence using levelset zigzag persistence modules and showed that the mapper converges to the Reeb graph in the bottleneck distance. Munch and Wang (2016) characterized the mapper using constructible cosheaves and proved the convergence between the (classic) mapper and the Reeb space (for $d \geq 1$) in interleaving distance. The enhanced mapper graph defined in this paper is similar to the geometric mapper graph introduced in Munch and Wang (2016). The differences between the enhanced mapper graph and geometric mapper consist of technical changes in the geometric realization of each space as a quotient of a disjoint union of closed intervals. Proposition 1 implies that the enhanced mapper graph is isomorphic to the display locale of the mapper cosheaf, giving theoretic significance to the geometrically realizable enhanced mapper graph.

Dey et al. (2017) established a convergence result between the mapper and the domain under a Gromov–Hausdorff metric. Carriére and Oudot (2018) showed convergence between the (multinerve) mapper and the Reeb graph using the functional distortion distance (Bauer et al. 2014). The enhanced mapper graph we define plays a role roughly analogous to the multinerve mapper in Carriére and Oudot (2018), although with several important distinctions. Most significantly is the fact that the enhanced mapper graph is an $\mathbb{R}$-space, and as such is not a purely combinatorial object, in contrast to the multinerve mapper, which is a simplicial poset. Carriére et al. (2018) proved convergence and provided a confidence set for the mapper using a bottleneck distance on certain extended persistence diagrams. They showed that the mapper is an optimal estimator of the Reeb graph and provided a statistical method for automatic parameter tuning using the rate of convergence. Like Carriére et al. (2018), this paper studies a notion of consistency (detailed below) for the mapper algorithm. In contrast to Carriére et al. (2018), the results provided here use the Reeb distance on constructible $\mathbb{R}$-graphs (defined in Sect. 2) rather than bottleneck distances on extended persistence diagrams, and are applicable to more general topological spaces (i.e., we do not require $\mathbb{X}$ to be a smooth manifold).

Probabilistic mapper inference This work is part of an effort to harness the theory of probability and statistics to support and analyze the use of topological methods with random data. To date, most of this effort has been put into problems related to the homology and persistent homology of random point clouds. The problem of *homological inference* relates to the ability to recover the homology (or persistent homology) of an unknown space or function given random observations. In a noiseless setup this problem was studied in Niyogi et al. (2008), Bobrowski (2019), Chazal et al. (2015), de Kergorlay et al. (2019), Wang and Wang (2018). The noisy setup was studied in Niyogi et al. (2011), Bobrowski et al. (2017b), Chazal et al. (2017), Fasy et al. (2014). Briefly, these works provide methods to recover the homology, together with assumptions that guarantee correct recovery with high probability. In many of these, the results are asymptotic, taking the number of points $n \to \infty$. The main reason for taking limits, is that the mathematics become more tractable, and provide simpler and more intuitive statements. Such asymptotic results can be considered as proofs of *consistency* for such homology estimation procedures. In Sect. 3, we apply results of Bobrowski et al. (2017b) to study consistency of the enhanced mapper construction

introduced in Sect. 2. The statistical techniques we use are similar to those developed in Chazal et al. (2011). For further discussion of the differences between the techniques used in Sect. 3 and the results of Chazal et al. (2011), see Bobrowski et al. (2017b).

In a way, the work here uses similar ideas to perform "mapper inference", a type of *structural inference*, and proves consistency. Other probabilistic studies related to applied topology mainly include limiting theorems (laws of large numbers, and central limit theorems), and extreme value analysis for the homology and persistent homology of random data (see e.g. Yogeshwaran et al. 2016; Hiraoka et al. 2018; Owada and Adler 2017; Bobrowski et al. 2017a; Kahle and Meckes 2013). However, these are much more detailed quantitative statements than what we are looking for when working with the mapper construction.

Contributions We highlight four contributions of this paper.

- First, in Sect. 2.3, we introduce and construct an *enhanced mapper graph*. This graph retains more geometric information about the underlying space than the combinatorially defined classic mapper graph, multinerve mapper graph, and geometric mapper graph (defined in Munch and Wang (2016)). Moreover, we show that the enhanced mapper graph construction provides a concrete realization of the display locale of a constructible cosheaf.
- Second, in Sect. 2.5, we give a categorical interpretation of the mapper construction. This categorification allows us to view mapper construction as a functor from the category of cosheaves to the category of constructible cosheaves. We can recover a geometric realization of the mapper construction from the categorical realization by taking enhanced mapper graphs, i.e., the display locales, of the corresponding constructible cosheaves.
- Third, we prove convergence (Theorem 1) and stability (Theorem 3) for the mapper cosheaf in the interleaving distance.
- Finally, we obtain results on the approximation quality of random mapper graphs obtained from noisy data on spaces which are not assumed to be manifolds (Theorem 2).

Moreover, using the results of de Silva et al. (2016), each of our theorems are reinterpreted in terms of the geometrically-defined enhanced mapper graph and Reeb distance on $\mathbb{R}$-graphs. This reinterpretation allows us to state our main result below without referring to the machinery of cosheaf theory.

Theorem (Corollary 3) *Let $\mathcal{R}(\mathbb{X}, f)$ be the Reeb graph of a constructible $\mathbb{R}$-space $(\mathbb{X}, f)$, $\hat{\mathfrak{D}}_n^{\pi}$ be the enhanced mapper graph associated to the cosheaf $\hat{\mathscr{D}}_n^{\pi}$ defined in Sect. 4, and $d_R(\cdot,\cdot)$ be the Reeb distance defined in Sect. 2. Using the notation defined in Sect. 3, if there exists $\varepsilon < \delta_{\mathcal{U}}$ such that p is ε-concentrated on $\mathbb{X}$, then*

$$\lim_{n\to\infty} \mathbb{P}\left(d_R\big(\hat{\mathfrak{D}}_n^{\pi}, \mathcal{R}(\mathbb{X}, f)\big) \leq res_f \mathcal{U}\right) = 1.$$

Intuitively speaking, the above theorem states that we can recover (a variant of) the mapper graph using the theory of cosheaves in a probabilistic setting. In particular, with high probability, the distance between an enhanced mapper graph and the Reeb graph

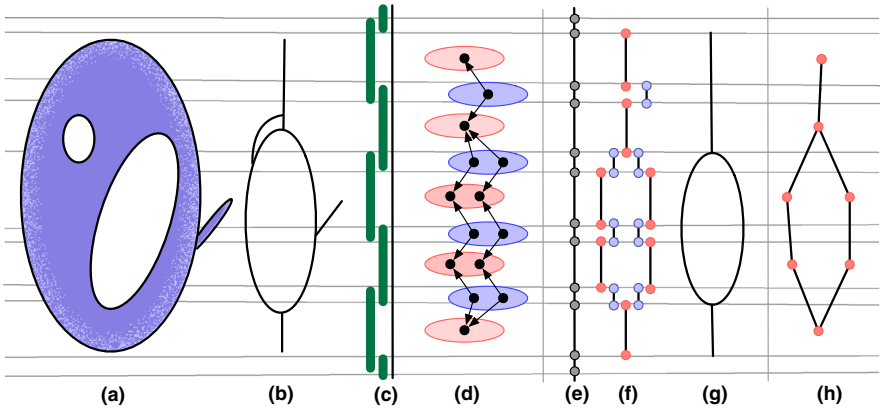


Fig. 1 An example of an enhanced mapper graph. **a** An $\mathbb{R}$-space $(\mathbb{X}, f)$ given by a topological space $\mathbb{X}$ (in blue) equipped with a height function $f : \mathbb{X} \to \mathbb{R}$. **b** Reeb graph of $(\mathbb{X}, f)$. **c** Nice cover of $\mathbb{R}$ with open intervals. **d** Visualization of the mapper cosheaf. **e** Stratification of $\mathbb{R}$. **f** Disjoint union of closed intervals ($\widetilde{\mathfrak{D}}$, in the notation of Sect. 2.3), with quotient isomorphic to the enhanced mapper graph. **g** Enhanced mapper graph ($\mathfrak{D}$, in the notation of Sect. 2.3). **h** Classic mapper graph of $(\mathbb{X}, f)$ (color figure online)

is upper bounded by the resolution of the cover (denoted as $\text{res}_f \mathcal{U}$, see Definition 15) as the number of samples goes to infinity. The proof of the theorem relies on two preliminary results. First, in Theorem 1, we construct an interleaving between the Reeb cosheaf and mapper cosheaf. Proposition 8 is the second key ingredient of the proof, giving a probabilistic recovery of the mapper cosheaf from random points. By interpreting the enhanced mapper graph in terms of cosheaf theory, we are able to simplify many of the proofs for convergence and stability. Generally, this paper illustrates the utility of combining sheaf theory with statistics in order to study robust topological and geometric properties of data.

Pictorial overview To better illustrate our key constructions, we give an example of an enhanced mapper graph. As illustrated in Fig. 1, given a topological space equipped with a height function $(\mathbb{X}, f)$, we are interested in studying how well its classic mapper graph (h) (with a fixed cover) approximates its Reeb graph (b). In order to study this problem, we construct a categorification of the mapper graph, through the theory of constructible cosheaves (d). The display locale functor is used to recover a geometric object from these category-theoretic constructible cosheaves. The geometric realization of the display locale of the mapper cosheaf is referred to as the enhanced mapper graph (g). We outline an explicit geometric realization of the enhanced mapper graph as a quotient of a disjoint union of closed intervals (f).

The main result of the paper, Theorem 2, gives (with high probability) a bound on the interleaving distance between the Reeb cosheaf and the enhanced mapper cosheaf. In order to interpret this result in terms of probabilistic convergence (Corollary 3), we apply the display locale functor to obtain the Reeb graph and the enhanced mapper graph from their cosheaf-theoretic analogues. This procedure results (with high probability) in a bound on the Reeb distance between an enhanced mapper graph and the Reeb graph of a constructible $\mathbb{R}$-space with random data.

2 Background

In this section, we review the results of de Silva et al. (2016) together with Munch and Wang (2016), showing that the interleaving distance between the mapper of the constructible $\mathbb{R}$-space $(\mathbb{X}, f)$ relative to the open cover $\mathcal{U}$ of $\mathbb{R}$ and the Reeb graph of $(\mathbb{X}, f)$ is bounded by the resolution of the open cover. Motivated by the categorification of Reeb graphs in de Silva et al. (2016), we introduce a categorified mapper algorithm, and restate the main results of Munch and Wang (2016) in this framework.

Categorification, in this context, means that we are interested in using the theory of constructible cosheaves to study Reeb graphs and mapper graphs. We can accomplish this by defining a cosheaf (the Reeb cosheaf) whose display locale is isomorphic to a given Reeb graph. One goal (completed in de Silva et al. 2016) of this approach is to use cosheaf theory to define an extended metric on the category of Reeb graphs. A natural candidate from the perspective of cosheaf theory is the interleaving distance. Suppose we want to use the interleaving distance of cosheaves to determine if two Reeb graphs are homeomorphic. We can first think of each Reeb graph as the display locale of a cosheaf, $\mathscr{F}$ and $\mathscr{G}$, respectively. This allows us to rephrase our problem as that of determining if the cosheaves, $\mathscr{F}$ and $\mathscr{G}$, are isomorphic. In general, interleaving distances cannot answer this question, since the interleaving distance is an extended *pseudo*-metric on the category of all cosheaves. In other words, having interleaving distance equal to 0 is not enough to guarantee that $\mathscr{F}$ and $\mathscr{G}$ are isomorphic as cosheaves. This seems to suggest that the interleaving distance is insufficient for the study of Reeb graphs. However (due to results of de Silva et al. 2016), if we restrict our study to the category of constructible cosheaves (over $\mathbb{R}$), we can avoid this subtlety. The interleaving distance is in fact an extended *metric* on the category of constructible cosheaves. If two constructible cosheaves have interleaving distance equal to 0, then they are isomorphic as cosheaves. Therefore, the display locales of constructible cosheaves (over $\mathbb{R}$) are homeomorphic if the interleaving distance between the cosheaves is equal to 0. In other words, if we want to know if two Reeb graphs are homeomorphic, it is sufficient to consider the interleaving distance between constructible cosheaves $\mathscr{F}$ and $\mathscr{G}$, provided that the display locales of the constructible cosheaves recover the Reeb graphs. Therefore, in the remainder of this section, we define a mapper cosheaf, and show that the Reeb cosheaf of a constructible $\mathbb{R}$-space is a constructible cosheaf, and that the mapper cosheaves are constructible. This allows us to use the commutativity of diagrams and the interleaving distance to prove convergence of the corresponding display locales, that is, the Reeb graphs and the enhanced mapper graphs. We use the example in Fig. 1 as a reference for various notions.

2.1 Constructible $\mathbb{R}$-spaces

We begin by defining constructible $\mathbb{R}$-spaces, which we consider to be the underlying spaces for estimating the Reeb graphs, see Fig. 1. Constructible $\mathbb{R}$-spaces can be considered as a class of topological spaces which provide a natural setting for generalizing aspects of classical Morse theory to the study of singular spaces. Like smooth manifolds equipped with a Morse function, constructible $\mathbb{R}$-spaces are topological

spaces equipped with a real valued function f, whose fibers, $f^{-1}(x)$, satisfy certain regularity conditions. Specifically, the topological structure of the fibers of the real valued function are required to only change at a finite set of function values. The function values which mark changes in the topological structure of fibers are referred to as critical values.

Definition 1 (de Silva et al. 2016) An $\mathbb{R}$-*space* is a pair $(\mathbb{X}, f)$, where $\mathbb{X}$ is a topological space and $f : \mathbb{X} \to \mathbb{R}$ is a continuous map. A *constructible* $\mathbb{R}$-*space* is an $\mathbb{R}$-space $(\mathbb{X}, f)$ satisfying the following conditions:

1. There exists a finite increasing sequence of points $S = \{a_0, \ldots, a_n\} \subset \mathbb{R}$, two finite sets of locally path-connected spaces $\{\mathbb{V}_0, \ldots, \mathbb{V}_n\}$ and $\{\mathbb{E}_0, \ldots, \mathbb{E}_{n-1}\}$, and two sets of continuous maps $\{\ell_i : \mathbb{E}_i \to \mathbb{V}_i\}$ and $\{r_i : \mathbb{E}_i \to \mathbb{V}_{i+1}\}$, such that $\mathbb{X}$ is the quotient space of the disjoint union

$$\coprod_{i=0}^{n} \mathbb{V}_i \times \{a_i\} \sqcup \coprod_{i=0}^{n-1} \mathbb{E}_i \times [a_i, a_{i+1}]$$

   by the relations

$$(\ell_i(x), a_i) \sim (x, a_i) \text{ and } (r_i(x), a_{i+1}) \sim (x, a_{i+1})$$

   for all i and $x \in \mathbb{E}_i$.
2. The continuous function $f : \mathbb{X} \to \mathbb{R}$ is given by projection onto the second factor of $\mathbb{X}$.

These are the objects of categories $\mathbb{R}$-**space** and $\mathbb{R}$-**space**$^{\mathbf{c}}$, consisting of $\mathbb{R}$-spaces and constructible $\mathbb{R}$-spaces, respectively. Morphisms in these categories are function-preserving maps; that is, $\varphi : (\mathbb{X}, f) \to (\mathbb{Y}, g)$ is given by a continuous map $\varphi : \mathbb{X} \to \mathbb{Y}$ such that $g \circ \varphi(x) = f(x)$.

Example 1 A smooth compact manifold $\mathbb{X}$ with a Morse function f constitutes a constructible $\mathbb{R}$-space. For instance, Fig. 1a illustrates a topological space $\mathbb{X}$ equipped with a height function f; the pair $(\mathbb{X}, f)$ is an $\mathbb{R}$-space. Similarly, a height function f on a torus $\mathbb{X}$ gives rise to an $\mathbb{R}$-space $(\mathbb{X}, f)$ in Fig. 6a.

In fact, $\mathbb{X}$ is not required to be a manifold for $(\mathbb{X}, f)$ to be an $\mathbb{R}$-space. Throughout the remainder of this paper, we assume that $(\mathbb{X}, f)$ is a constructible $\mathbb{R}$-space.

Definition 2 (de Silva et al. 2016) An $\mathbb{R}$-*graph* is a constructible $\mathbb{R}$-space such that the sets $\mathbb{V}_i$ and $\mathbb{E}_i$ are finite sets (with the discrete topology) for all i.

Example 2 The Reeb graph of a constructible $\mathbb{R}$-space is an $\mathbb{R}$-graph. For instance, the Reeb graph of $(\mathbb{X}, f)$ in Fig. 1b is an $\mathbb{R}$-graph. Similarly, the Reeb graph of a Morse function on a torus is an $\mathbb{R}$-graph, see Fig. 6b.

2.2 Constructible cosheaves

Sheaves and cosheaves are category-theoretic structures, called functors, which provide a framework for associating data to open sets in a topological space. These associations are required to preserve certain properties inherent to the topology of the space. In this way, one can study the topological structure of the space by studying the data associated to each open set by a given sheaf or cosheaf. In the following sections, we will use cosheaves to encode information about a constructible $\mathbb{R}$-space by associating open intervals in the real line to sets of (path-)connected components of fibers of the real valued function corresponding to the constructible $\mathbb{R}$-space.

Let **Int** be the category of connected open sets in $\mathbb{R}$ with inclusions which we refer to as intervals, and **Set** the category of abelian groups with group homomorphism maps. We first define a cosheaf over $\mathbb{R}$, which we propose to be the natural objects for categorifying the mapper algorithm.

Definition 3 A *pre-cosheaf* $\mathscr{F}$ on $\mathbb{R}$ is a covariant functor $\mathscr{F} : \mathbf{Int} \to \mathbf{Set}$. The category of precosheaves on $\mathbb{R}$ is denoted $\mathbf{Set}^{\mathbf{Int}}$ with morphisms given by natural transformations.

A pre-cosheaf $\mathscr{F}$ is a *cosheaf* if

$$\varinjlim_{V \in \mathcal{V}} \mathscr{F}(V) = \mathscr{F}(U)$$

for each open interval $U \in \mathbf{Int}$ and each open interval cover $\mathcal{V} \subset \mathbf{Int}$ of U, which is closed under finite intersections. The full subcategory of $\mathbf{Set}^{\mathbf{Int}}$ consisting of cosheaves is denoted **Csh**.

Remark 1 We note that usually, cosheaves are defined over the category of arbitrary open sets rather than the category of connected open sets. However, the category of cosheaves defined over connected open sets is equivalent to the category of cosheaves defined over arbitrary open sets, by the colimit property of cosheaves. When we define smoothing operations on cosheaves in Sect. 2.4, there are important distinctions that will make clear the need for the definition with respect to **Int**, as set-thickening operations do not preserve the cosheaf property otherwise.

Since we are interested in working with cosheaves which can be described with a finite amount of data, we will restrict our attention to a well-behaved subcategory of **Csh**, consisting of constructible cosheaves (defined below). Constructibility can be thought of as a type of "tameness" assumption for sheaves and cosheaves.

Definition 4 A cosheaf $\mathscr{F}$ is *constructible* if there exists a finite set $S \subset \mathbb{R}$ of *critical values* such that $\mathscr{F}[U \subset V]$ is an isomorphism whenever $S \cap U = S \cap V$. The full subcategory of **Csh** consisting of constructible cosheaves is denoted $\mathbf{Csh}^{\mathbf{c}}$.

2.3 The Reeb cosheaf and display locale functors

We introduce the Reeb cosheaf and display locale functors. These functors relate the category of constructible cosheaves to the category of $\mathbb{R}$-graphs, and provide a natural

categorification of the Reeb graph (de Silva et al. 2016). In other words, via both Reeb cosheaf functor and display locale functors, one could consider the translation between the data and their corresponding categorical interpretations.

Let $\mathscr{R}_f$ be the *Reeb cosheaf* of $(\mathbb{X}, f)$ on $\mathbb{R}$, defined by

$$\mathscr{R}_f(U) = \pi_0(\mathbb{X}^U),$$

where $\mathbb{X}^U := f^{-1}(U)$ and $\pi_0(\mathbb{X}^U)$ denotes the set of path components of $\mathbb{X}^U$.

Definition 5 The *Reeb cosheaf functor* $\mathcal{C}$ from the category of constructible $\mathbb{R}$-spaces to the category of constructible cosheaves

$$\mathbb{R}\textbf{-space}^{\mathbf{c}} \xrightarrow{\mathcal{C}} \textbf{Csh}^{\mathbf{c}}$$

is defined by $\mathcal{C}((\mathbb{X}, f)) = \mathscr{R}_f$. For a function-preserving map $\varphi : (\mathbb{X}, f) \to (\mathbb{Y}, g)$, the resulting morphism $\mathcal{C}[\varphi]$ is given by $\mathcal{C}[\varphi] : \mathscr{R}_f(U) = \pi_0 \circ f^{-1}(U) \to \pi_0 \circ g^{-1}(U) = \mathscr{R}_g(U)$ induced by $\varphi \circ f^{-1}(U) \subseteq g^{-1}(U)$.

Definition 6 The *costalk* of a (pre-)cosheaf $\mathscr{F}$ at $x \in \mathbb{R}$ is

$$\mathscr{F}_x = \varprojlim_{I \ni x} \mathscr{F}(I).$$

For each costalk $\mathscr{F}_x$, there is a natural map $\mathscr{F}_x \to \mathscr{F}(I)$ (given by the universal property of limits) for each open interval I containing x.

In order to related the Reeb and mapper cosheaves to geometric objects, we make use of the notion of *display locale*, introduced in Funk (1995).

Definition 7 The *display locale* of a cosheaf $\mathscr{F}$ (as a set) is defined as

$$\mathcal{D}(\mathscr{F}) = \coprod_{x \in \mathbb{R}} \mathscr{F}_x.$$

A topology on $\mathcal{D}(\mathscr{F})$ is generated by open sets of the form

$$U_{I,a} = \{s \in \mathscr{F}_x : x \in I \text{ and } s \mapsto a \in \mathscr{F}(I)\},$$

for each open interval $I \in \textbf{Int}$ and each section $a \in \mathscr{F}(I)$.

The display locale gives a functor from the category of cosheaves to the category of $\mathbb{R}$-graphs,

$$\textbf{Csh}^{\mathbf{c}} \xrightarrow{\mathcal{D}} \mathbb{R}\textbf{-graph}.$$

We proceed by giving an explicit geometric realization of the display locale of a constructible cosheaf. Let $\mathscr{F}$ be a constructible cosheaf with set of critical values $\mathbb{R}_0 \subset \mathbb{R}$. Let $\mathbb{R}_1 = \mathbb{R} \setminus \mathbb{R}_0$ be the complement of $\mathbb{R}_0$, so that we form a stratification

$$\mathbb{R} = \mathbb{R}_0 \sqcup \mathbb{R}_1,$$

See Fig. 1e for an example (black points are in $\mathbb{R}_0$, their complements are in $\mathbb{R}_1$). Let S_1 be the set of connected components of $\mathbb{R}_1$, i.e., the 1-dimensional stratum pieces. For $x \in \mathbb{R}_0$, let I_x denote the largest open interval containing x such that $I_x \cap \mathbb{R}_0 = \{x\}$. Let

$$\tilde{\mathfrak{D}}(\mathscr{F}) := \coprod_{V \in S_1} \overline{V} \times \mathscr{F}(V) \sqcup \coprod_{x \in \mathbb{R}_0} \{x\} \times \mathscr{F}(I_x),$$

where $\overline{V}$ is the closure of V and the product $C \times \emptyset$ of a set C with the empty set is understood to be empty. Geometrically, $\tilde{\mathfrak{D}}(\mathscr{F})$ is a disjoint union of connected closed subsets of $\mathbb{R}$; if the support of $\mathscr{F}$ is compact, then $\tilde{\mathfrak{D}}(\mathscr{F})$ is a disjoint union of closed intervals and points. Let π denote the projection map

$$\begin{aligned} \pi : \tilde{\mathfrak{D}}(\mathscr{F}) &\to \mathbb{R} \\ (x, a) &\mapsto x. \end{aligned}$$

Suppose $(x, a) \in \overline{V} \times \mathscr{F}(V) \subset \tilde{\mathfrak{D}}(\mathscr{F})$ and $x \in \mathbb{R}_0$. We have that $V \cap \mathbb{R}_0 = \emptyset$ and $I_x \cap V \neq \emptyset$ (because x lies on the boundary of V). By maximality of I_x, we have the inclusion $V \subset I_x$. Let $\varphi_{(x,a)}$ be the map

$$\varphi_{(x,a)} : \mathscr{F}(V) \to \mathscr{F}(I_x)$$

induced by the inclusion $V \subset I_x$. We can extend this map to the fiber of π over x,

$$\psi_x : \pi^{-1}(x) \to \mathscr{F}(I_x),$$

where $\psi_x((x, a)) := \varphi_{(x,a)}(a)$ if $(x, a) \in \overline{V} \times \mathscr{F}(V)$ and $\psi_x((x, a)) := a$ if $(x, a) \in \{x\} \times \mathscr{F}(I_x)$. Finally, we define an equivalence relation of points in $\tilde{\mathfrak{D}}(\mathscr{F})$. Suppose $(x, a), (y, b) \in \tilde{\mathfrak{D}}(\mathscr{F})$. Then $(x, a) \sim (y, b)$ if

1. $x = y \in \mathbb{R}_0$, and
2. $\psi_x(a) = \psi_x(b) \in \mathscr{F}(I_x)$.

Finally, let

$$\mathfrak{D}(\mathscr{F}) := \tilde{\mathfrak{D}}(\mathscr{F}) / \sim$$

be the quotient of $\tilde{\mathfrak{D}}(\mathscr{F})$ by the equivalence relation. The projection π factors through the quotient, giving a map $\bar{\pi} : \mathfrak{D}(\mathscr{F}) \to \mathbb{R}$.

Proposition 1 *If $\mathscr{F}$ is a constructible cosheaf with set of critical values S, then $\mathfrak{D}(\mathscr{F})$ is a 1-dimensional CW-complex which is isomorphic (as an $\mathbb{R}$-space) to the display locale, $\mathcal{D}(\mathscr{F})$, of $\mathscr{F}$.*

Proof We will construct a homeomorphism $\gamma : \mathfrak{D}(\mathscr{F}) \to \mathcal{D}(\mathscr{F})$ which preserves the natural quotient maps $\bar{f} : \mathcal{D}(\mathscr{F}) \to \mathbb{R}$ and $\bar{\pi} : \mathfrak{D}(\mathscr{F}) \to \mathbb{R}$. Given $x \in \mathbb{R}_1$, we have that $\bar{\pi}^{-1}(x) = \{x\} \times \mathscr{F}(V)$, where V is the connected component of $\mathbb{R}_1$ which contains x. Since $\mathscr{F}$ is constructible with respect to the chosen stratification, we have that $\mathscr{F}(V) \cong \mathscr{F}_x$. This gives a bijection from $\bar{\pi}^{-1}(x)$ to $\bar{f}^{-1}(x)$. For $x \in \mathbb{R}_0$, the fiber $\bar{\pi}^{-1}(x)$ is by construction in bijection with $\mathscr{F}(I_x)$. Again, since $\mathscr{F}$ is constructible and $I_x \cap \mathbb{R}_0 = B(x) \cap \mathbb{R}_0$ for each sufficiently small neighborhood $B(x)$ of x, we have that $\mathscr{F}(I_x) \cong \mathscr{F}_x$. These bijections define a map $\gamma : \mathfrak{D}(\mathscr{F}) \to \mathcal{D}(\mathscr{F})$, which preserves the quotient maps by construction. All that remains is to show that γ is continuous.

Suppose $x \in \mathbb{R}_1$, and let V be the connected component of $\mathbb{R}_1$ which contains x, and $B(x)$ be an open neighborhood of x such that $B(x) \subset V$. Then $\mathscr{F}_y \cong \mathscr{F}(V)$ for each $y \in B(x)$, and $\mathscr{F}(B(x)) \cong \mathscr{F}(V)$. Recall the definition of the basic open sets $U_{I,a}$ in the definition of display locale (with notation adjusted to better align with the current proof),

$$U_{I,a} = \left\{ s \in \mathscr{F}_y \subset \coprod_{x \in \mathbb{R}} \mathscr{F}_x : y \in I \text{ and } s \mapsto a \in \mathscr{F}(I) \right\}.$$

Using the above isomorphisms to simplify the definition according to the current set-up, we get

$$U_{B(x),a} \cong \left\{ a \in \coprod_{y \in B(x)} \mathscr{F}(V) \right\}.$$

Therefore, $\gamma^{-1}(U_{B(x),a}) = B(x) \times \{a\}$, which is open in the quotient topology on $\mathfrak{D}(\mathscr{F})$.

Suppose $x \in \mathbb{R}_0$, and let $B(x)$ be a neighborhood of x such that $B(x) \subset I_x$. Let V_1 and V_2 denote the two connected components of $\mathbb{R}_1$ which are contained in I_x. If $y \in B(x)$, then $\mathscr{F}_y$ is isomorphic to either $\mathscr{F}(V_1)$, $\mathscr{F}(V_2)$, or $\mathscr{F}(I_x)$. Moreover, since $\mathscr{F}$ is constructible, we have that $\mathscr{F}(B(x)) \cong \mathscr{F}(I_x)$. Let $a' \in \mathscr{F}(I_x)$ correspond to $a \in \mathscr{F}(B(x))$ under the isomorphism $\mathscr{F}(I_x) \cong \mathscr{F}(B(x))$. Following the definitions, we have that

$$\begin{aligned} \pi^{-1}\left(\gamma^{-1}(U_{B(x),a})\right) = &\left(\overline{V_1} \cap B(x)\right) \times \mathscr{F}[V_1 \subset I_x]^{-1}(a') \\ &\sqcup \left(\overline{V_2} \cap B(x)\right) \times \mathscr{F}[V_2 \subset I_x]^{-1}(a') \\ &\sqcup \{x\} \times \{a'\}, \end{aligned}$$

where $\mathscr{F}[V_i \subset I_x]^{-1}(a')$ is understood to be a (possibly empty) subset of $\mathscr{F}(V_i)$. It follows that $\gamma^{-1}(U_{B(x),a})$ is open in the quotient topology on $\mathfrak{D}(\mathscr{F})$. Therefore, γ^{-1}

maps open sets to open sets, and we have shown that γ is a homeomorphism which preserves the quotient maps $\bar{f}$ and $\bar{\pi}$, i.e., $\bar{f}(\gamma((x,a))) = \bar{\pi}((x,a)) = x$. □

It follows from the proposition that $\mathfrak{D}(\mathscr{F})$ is independent (up to isomorphism) of choice of critical values $\mathbb{R}_0$. Additionally, we now note that we can freely use the notation $\mathfrak{D}(\mathscr{F})$ or $\mathcal{D}(\mathscr{F})$ to refer to the display locale of a constructible cosheaf over $\mathbb{R}$. We will continue to use both symbols, reserving $\mathcal{D}$ for the display locale of an arbitrary cosheaf, and using $\mathfrak{D}$ when we want to emphasize the above equivalence for constructible cosheaves.

In de Silva et al. (2016), it is shown that the Reeb graph $\mathcal{R}(\mathbb{X}, f)$ of $(\mathbb{X}, f)$ is naturally isomorphic to the display locale of $\mathscr{R}_f$. Moreover, the display locale functor $\mathcal{D}$ and the Reeb functor $\mathcal{C}$ are inverse functors and define an equivalence of categories between the category of Reeb graphs and the category of constructible cosheaves on $\mathbb{R}$. This equivalence is closely connected to the more general relationships between constructible cosheaves and stratified coverings studied in Woolf (2009). The result allows us to define a distance between Reeb graphs by taking the interleaving distance between the associated constructible cosheaves as shown in the following section.

2.4 Interleavings

We start by defining the interleavings on the categorical objects. Interleaving is a typical tool in topological data analysis for quantifying proximity between objects such as persistence modules and cosheaves. For $U \subseteq \mathbb{R}$, let $U \mapsto U_\varepsilon := \{y \in \mathbb{R} \mid \|y - U\| \leq \varepsilon\}$. If $U = (a,b) \in \mathbf{Int}$, then $U_\varepsilon = (a-\varepsilon, b+\varepsilon)$.

Definition 8 Let $\mathscr{F}$ and $\mathscr{G}$ be two cosheaves on $\mathbb{R}$. An *ε-interleaving* between $\mathscr{F}$ and $\mathscr{G}$ is given by two families of maps

$$\varphi_U : \mathscr{F}(U) \to \mathscr{G}(U_\varepsilon), \quad \psi_U : \mathscr{G}(U) \to \mathscr{F}(U_\varepsilon)$$

which are natural with respect to the inclusion $U \subset U_\varepsilon$, and such that

$$\psi_{U_\varepsilon} \circ \varphi_U = \mathscr{F}[U \subset U_{2\varepsilon}], \quad \varphi_{U_\varepsilon} \circ \psi_U = \mathscr{G}[U \subset U_{2\varepsilon}]$$

for all open intervals $U \subset \mathbb{R}$. Equivalently, we require that the diagram

$$\begin{array}{ccccc} \mathscr{F}(U) & \longrightarrow & \mathscr{F}(U_\varepsilon) & \longrightarrow & \mathscr{F}(U_{2\varepsilon}) \\ & {\scriptstyle\varphi_U}\searrow \quad \nearrow{\scriptstyle\psi_U} & & {\scriptstyle\varphi_{U_\varepsilon}}\searrow \quad \nearrow{\scriptstyle\psi_{U_\varepsilon}} & \\ \mathscr{G}(U) & \longrightarrow & \mathscr{G}(U_\varepsilon) & \longrightarrow & \mathscr{G}(U_{2\varepsilon}) \end{array}$$

commutes, where the horizontal arrows are induced by $U \subseteq U_\varepsilon \subseteq U_{2\varepsilon}$.

The *interleaving distance* between two cosheaves $\mathscr{F}$ and $\mathscr{G}$ is given by

$$d_I(\mathscr{F}, \mathscr{G}) := \inf\{\varepsilon \mid \text{ there exists an } \varepsilon\text{-interleaving between } \mathscr{F} \text{ and } \mathscr{G}\}.$$

Now that we have an interleaving for elements of $\mathbf{Csh^c}$ along with an equivalence of categories between $\mathbf{Csh^c}$ and $\mathbb{R}$-**graph**, we can develop this into an interleaving distance for the Reeb graphs themselves. The interleaving distance for Reeb graphs will be defined using a smoothing functor, which we construct below.

Definition 9 Let $(\mathbb{X}, f)$ be a constructible $\mathbb{R}$-space. For $\varepsilon \geq 0$, define the *thickening functor* $\mathcal{T}_\varepsilon$ to be

$$\mathcal{T}_\varepsilon(\mathbb{X}, f) = (\mathbb{X} \times [-\varepsilon, \varepsilon], f_\varepsilon),$$

where $f_\varepsilon(x, t) = f(x) + t$. Given a morphism $\alpha : \mathbb{X} \to \mathbb{Y}$,

$$\begin{aligned} \mathcal{T}_\varepsilon(\alpha) : \mathbb{X} \times [-\varepsilon, \varepsilon] &\to \mathbb{Y} \times [-\varepsilon, \varepsilon] \\ (x, t) &\mapsto (\alpha(x), t). \end{aligned}$$

The *zero section map* is the morphism $(\mathbb{X}, f) \to \mathcal{T}_\varepsilon(\mathbb{X}, f)$ induced by

$$\begin{aligned} \mathbb{X} &\to \mathbb{X} \times [-\varepsilon, \varepsilon] \\ x &\mapsto (x, 0). \end{aligned}$$

Proposition 2 (de Silva et al. 2016, Proposition 4.23) *The thickening functor $\mathcal{T}_\varepsilon$ maps $\mathbb{R}$-graphs to constructible $\mathbb{R}$-spaces, i.e., if $(\mathbb{G}, g) \in \mathbb{R}-$**graphs** then $\mathcal{T}_\varepsilon(\mathbb{G}, g) \in \mathbb{R}-$**spaces**$^{\mathbf{c}}$.*

In general, the thickening functor $\mathcal{T}_\varepsilon$ will output a constructible $\mathbb{R}$-space, and not an $\mathbb{R}$-graph. In order to define a 'smoothing' functor for $\mathbb{R}$-graphs (following de Silva et al. 2016), we need to introduce a Reeb functor, which maps a constructible $\mathbb{R}$-space to an $\mathbb{R}$-graph.

Definition 10 The *Reeb graph functor* $\mathcal{R}$ maps a constructible $\mathbb{R}$-space $(\mathbb{X}, f)$ to an $\mathbb{R}$-graph $(\mathbb{X}_f, \bar{f})$, where $\mathbb{X}_f$ is the Reeb graph of $(\mathbb{X}, f)$ and $\bar{f}$ is the function induced by f on the quotient space $\mathbb{X}_f$. The *Reeb quotient map* is the morphism $(\mathbb{X}, f) \to \mathcal{R}(\mathbb{X}, f)$ induced by the quotient map $\mathbb{X} \to \mathbb{X}_f$.

Now we can define a smoothing functor on the category of $\mathbb{R}$-graphs.

Definition 11 Let $(\mathbb{G}, f) \in \mathbb{R}$-**graph**. The *Reeb smoothing functor* $\mathcal{S}_\varepsilon : \mathbb{R}$-**graph** $\to$ $\mathbb{R}$-**graph** is defined to be the Reeb graph of an ε-thickened $\mathbb{R}$-graph

$$\mathcal{S}_\varepsilon(\mathbb{G}, f) = \mathcal{R}\left(\mathcal{T}_\varepsilon(\mathbb{G}, f)\right).$$

The Reeb smoothing functor $\mathcal{S}_\varepsilon$ defined above is used to define an interleaving distance for Reeb graphs, called the Reeb interleaving distance. The Reeb interleaving distance, defined below, can be thought of as a geometric analogue of the interleaving distance of constructible cosheaves. Let $\zeta^\varepsilon_{\mathbb{F}}$ be the map from $(\mathbb{F}, f)$ to $\mathcal{S}_\varepsilon(\mathbb{F}, f)$ given by the composition of the zero section map $(\mathbb{F}, f) \to \mathcal{T}_\varepsilon(\mathbb{F}, f)$ with the Reeb quotient map $\mathcal{T}_\varepsilon(\mathbb{F}, f) \to \mathcal{R}(\mathcal{T}_\varepsilon(\mathbb{F}, f))$. To ease notation, we will denote the composition of $\zeta^\varepsilon_{\mathbb{F}} : (\mathbb{F}, f) \to \mathcal{S}_\varepsilon(\mathbb{F}, f)$ with $\zeta_{\mathcal{S}_\varepsilon(\mathbb{F}, f)} : \mathcal{S}_\varepsilon(\mathbb{F}, f) \to \mathcal{S}_\varepsilon(\mathcal{S}_\varepsilon(\mathbb{F}, f))$ by $\zeta^\varepsilon_{\mathbb{F}}(\zeta^\varepsilon_{\mathbb{F}}(\mathbb{F}, f))$.

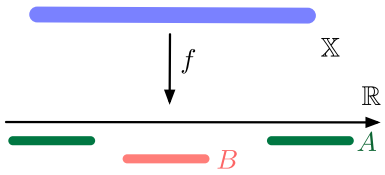


Fig. 2 A counter example showing why we use **Int** rathar than **Open**($\mathbb{R}$) for the definition of cosheaves in our context. See Remark 3

Definition 12 Let $(\mathbb{F}, f)$ and $(\mathbb{G}, g)$ be $\mathbb{R}$-graphs. We say that $(\mathbb{F}, f)$ and $(\mathbb{G}, g)$ are *ε-interleaved* if there exists a pair of function-preserving maps

$$\alpha : (\mathbb{F}, f) \to \mathcal{S}_\varepsilon(\mathbb{G}, g) \qquad \text{and} \qquad \beta : (\mathbb{G}, g) \to \mathcal{S}_\varepsilon(\mathbb{F}, f)$$

such that

$$\mathcal{S}_\varepsilon(\beta)\,(\alpha(\mathbb{F}, f)) = \zeta^\varepsilon_{\mathbb{F}}\left(\zeta^\varepsilon_{\mathbb{F}}(\mathbb{F}, f)\right) \quad \text{and} \quad \mathcal{S}_\varepsilon(\alpha)\,(\beta(\mathbb{G}, g)) = \zeta^\varepsilon_{\mathbb{G}}\left(\zeta^\varepsilon_{\mathbb{G}}(\mathbb{G}, g)\right).$$

That is, the diagram

$$\begin{array}{ccccc}
(\mathbb{F}, f) & \longrightarrow & \zeta^\varepsilon_{\mathbb{F}}(\mathbb{F}, f) & \longrightarrow & \zeta^\varepsilon_{\mathbb{F}}\left(\zeta^\varepsilon_{\mathbb{F}}(\mathbb{F}, f)\right) \\
 & \overset{\alpha}{\searrow}\ \overset{\beta}{\nearrow} & & \overset{\mathcal{S}_\varepsilon(\alpha)}{\searrow}\ \overset{\mathcal{S}_\varepsilon(\beta)}{\nearrow} & \\
(\mathbb{G}, g) & \longrightarrow & \zeta^\varepsilon_{\mathbb{G}}(\mathbb{G}, g) & \longrightarrow & \zeta^\varepsilon_{\mathbb{G}}\left(\zeta^\varepsilon_{\mathbb{G}}(\mathbb{G}, g)\right)
\end{array}$$

commutes.

The *Reeb interleaving distance*, $d_R\,((\mathbb{F}, f), (\mathbb{G}, g))$, is defined to be the infimum over all ε such that there exists an ε-interleaving of $(\mathbb{F}, f)$ and $(\mathbb{G}, g)$:

$$d_R\,((\mathbb{F}, f), (\mathbb{G}, g)) := \inf\{\varepsilon : \text{ there exists an } \varepsilon\text{-interleaving of } (\mathbb{F}, f) \text{ and } (\mathbb{G}, g)\}.$$

Remark 2 We should remark on a technical aspect of the above definition. The composition $\zeta^\varepsilon_{\mathbb{F}} \circ \zeta^\varepsilon_{\mathbb{F}}(\mathbb{F}, f)$ is naturally isomorphic to $\zeta^{2\varepsilon}_{\mathbb{F}}(\mathbb{F}, f)$. However, since the definition of the Reeb interleaving distance requires certain diagrams to commute, it is necessary to specify *an* isomorphism between $\zeta^\varepsilon_{\mathbb{F}} \circ \zeta^\varepsilon_{F}(\mathbb{F}, f)$ and $\zeta^{2\varepsilon}_{\mathbb{F}}(\mathbb{F}, f)$ if one would like to replace $\zeta^\varepsilon_{\mathbb{F}} \circ \zeta^\varepsilon_{F}(\mathbb{F}, f)$ with $\zeta^{2\varepsilon}_{\mathbb{F}}(\mathbb{F}, f)$ in the commutative diagrams. Therefore, we choose to work exclusively with the composition of zero section maps, rather than working with diagrams which commute up to natural isomorphism.

The remaining proposition of this section gives a geometric realization of the interleaving distance of constructible cosheaves.

Proposition 3 (de Silva et al. 2016) *$\mathcal{D}(\mathscr{F})$ and $\mathcal{D}(\mathscr{G})$ are ε-interleaved as $\mathbb{R}$-graphs if and only if $\mathscr{F}$ and $\mathscr{G}$ are ε-interleaved as constructible cosheaves.*

Remark 3 Cosheaves are usually defined as functors on the category of open sets instead of functors on the connected open sets. We choose to use **Int** instead of

Open($\mathbb{R}$) due to technical issues that arise when we begin smoothing the functors. Basically, smoothing the functor does not produce a cosheaf when the intervals are replaced by arbitrary open sets in $\mathbb{R}$. Consider the example of Fig. 2, where $\mathbb{X}$ is a line with map f projection onto $\mathbb{R}$. Say U^ε is the thickening of a set, $U^\varepsilon = \{x \in \mathbb{R} \mid |x - U| < \varepsilon\}$. Then we can pick an ε so that A^ε is two disjoint intervals, and $(A \cup B)^\varepsilon$ is one interval. Let F be the functor $U \mapsto \pi_0 f^{-1}(U)$ which is a cosheaf representing the Reeb graph. Then the functor $F \circ (\cdot)^\varepsilon$ is not a cosheaf since by the diagram,

$$\begin{array}{ccc} \emptyset = F((A \cap B)^\varepsilon) & \longrightarrow & F(A^\varepsilon) = \{\bullet\bullet\} \\ \downarrow & & \vdots \\ \{\bullet\} = F(B^\varepsilon) & \dashrightarrow & \text{colim} = \{\bullet\,\bullet\,\bullet\} \end{array}$$

$F(A \cup B)^\varepsilon = \{\bullet\}$ is not the colimit of $F(A^\varepsilon)$ and $F(B^\varepsilon)$.[1]

2.5 Categorified mapper

In this section, we interpret classic mapper (for scalar functions), a topological descriptor, as a category theoretic object. This interpretation, in terms of cosheaves and category theory, simplifies many of the arguments used to prove convergence results in Sect. 4. We first review the classic mapper and then discuss the categorified mapper. The main ingredient needed to define the mapper construction is a choice of cover. We say a cover of $\mathbb{R}$ is *good* if all intersections are contractible. A cover $\mathcal{U}$ is *locally finite* if for every $x \in \mathbb{R}$, $\mathcal{U}_x = \{V \in \mathcal{U} : x \in V\}$ is a finite set. In particular, locally finiteness implies that the cover restricted to a compact set is finite. For the remainder of the paper, we work with *nice covers* which are good, locally finite, and consist only of connected intervals, see Fig. 1c for an example.

We will now introduce a categorification of mapper. Let $\mathcal{U}$ be a nice cover of $\mathbb{R}$. Let $\mathcal{N}_\mathcal{U}$ be the nerve of $\mathcal{U}$, endowed with the Alexandroff topology. Consider the continuous map

$$\begin{aligned} \eta : \mathbb{R} &\to \mathcal{N}_\mathcal{U} \\ x &\mapsto \bigcap_{V \in \mathcal{U}_x} V, \end{aligned}$$

where the intersection $\bigcap_{V \in \mathcal{U}_x} V$ is viewed as an open simplex of $\mathcal{N}_\mathcal{U}$. The *mapper functor* $\mathcal{M}_\mathcal{U} : \mathbf{Set}^{\mathbf{Int}} \to \mathbf{Set}^{\mathbf{Int}}$ can be defined as

$$\mathcal{M}_\mathcal{U}(\mathscr{C}) = \eta^*(\eta_*(\mathscr{C})),$$

where η^* and η_* are the (pre)-cosheaf-theoretic pull-back and push-forward operations respectively. However, rather than defining η^* and η_* in generality, we choose to work

[1] We thank Vin de Silva for this counterexample.

with an explicit description of $\mathcal{M}_{\mathcal{U}}(\mathscr{C})$ given below. For notational convenience, define

$$\begin{aligned}\mathcal{I}_{\mathcal{U}} : \mathbf{Int} &\to \mathbf{Int}\\ U &\mapsto \eta^{-1}(\mathrm{St}(\eta(U))),\end{aligned}$$

where $\mathrm{St}(\eta(U))$ denotes the minimal open set in $\mathcal{N}_{\mathcal{U}}$ containing $\eta(U) := \cup_{x\in U}\eta(x)$ (the open star of $\eta(U)$ in $\mathcal{N}_{\mathcal{U}}$). It is often convenient to identify $\mathcal{I}_{\mathcal{U}}(U)$ with a union of open intervals in $\mathbb{R}$.

Lemma 1 *Using the notation defined above, we have the equality*

$$\mathcal{I}_{\mathcal{U}}(U) = \bigcup_{x\in U}\bigcap_{V\in\mathcal{U}_x} V,$$

where $\bigcap_{V\in\mathcal{U}_x} V$ *is viewed as a subset of* $\mathbb{R}$ *(not as a simplex of* $\mathcal{N}_{\mathcal{U}}$*).*

Proof If $y \in \bigcup_{x\in U}\bigcap_{V\in\mathcal{U}_x} V$, then there exists an $x \in U$ such that $y \in V$ for all $V \in \mathcal{U}_x$. In other words, $\mathcal{U}_x \subseteq \mathcal{U}_y$. Therefore, $\eta(y) \geq \eta(x)$ in the partial order of $\mathcal{N}_{\mathcal{U}}$. Therefore, $\eta(y) \in \mathrm{St}(\eta(U))$. This implies that $\bigcup_{x\in U}\bigcap_{V\in\mathcal{U}_x} V \subseteq \mathcal{I}_{\mathcal{U}}(U)$. For the reverse inclusion, assume that $u \in \mathcal{I}_{\mathcal{U}}(U)$, i.e., $\eta(u) \in \mathrm{St}(\eta(U))$. This implies that there exists $v \in U$ such that $\eta(u) \geq \eta(v)$. In other words, $\mathcal{U}_v \subseteq \mathcal{U}_u$. Therefore $u \in \cap_{V\in\mathcal{U}_v} V$, and $u \in \bigcup_{v\in U}\bigcap_{V\in\mathcal{U}_v} V$. □

Under this identification, it is clear that $\mathcal{I}_{\mathcal{U}}(U)$ is an open set in $\mathbb{R}$ (since the open cover $\mathcal{U}$ is locally finite), and if $U \subset V$ then $\mathcal{I}_{\mathcal{U}}(U) \subset \mathcal{I}_{\mathcal{U}}(V)$. Moreover, since $\bigcap_{V\in\mathcal{U}_x} V$ is an interval open neighborhood of x and U is an open interval, then $\mathcal{I}_{\mathcal{U}}(U)$ is an open interval. Therefore, $\mathcal{I}_{\mathcal{U}}$ can be viewed as a functor from **Int** to **Int**.

Finally, we can give $\mathcal{M}_{\mathcal{U}}(\mathscr{C})$ an explicit description in terms of the functor $\mathcal{I}_{\mathcal{U}}$.

Definition 13 The *mapper functor* $\mathcal{M}_{\mathcal{U}} : \mathbf{Set}^{\mathbf{Int}} \to \mathbf{Set}^{\mathbf{Int}}$ is defined by

$$\mathcal{M}_{\mathcal{U}}(\mathscr{C})(U) := \mathscr{C}(\mathcal{I}_{\mathcal{U}}(U)),$$

for each open interval $U \in \mathbf{Int}$.

Since $\mathcal{I}_{\mathcal{U}}$ is a functor from **Int** to **Int**, it follows that $\mathcal{M}_{\mathcal{U}}$ is a functor from $\mathbf{Set}^{\mathbf{Int}}$ to $\mathbf{Set}^{\mathbf{Int}}$. Hence, $\mathcal{M}_{\mathcal{U}}(\mathscr{C})$ is a functor from the category of pre-cosheaves to the category of pre-cosheaves. In the following proposition, we show that if $\mathscr{C}$ is a cosheaf, then $\mathcal{M}_{\mathcal{U}}(\mathscr{C})$ is in fact a constructible cosheaf.

Proposition 4 *Let* $\mathcal{U}$ *be a finite nice open cover of* $\mathbb{R}$*. The mapper functor* $\mathcal{M}_{\mathcal{U}}$ *is a functor from the category of cosheaves on* $\mathbb{R}$ *to the category of constructible cosheaves on* $\mathbb{R}$*:*

$$\mathcal{M}_{\mathcal{U}} : \mathbf{CSh} \to \mathbf{CSh}^{\mathbf{c}}.$$

Moreover, the set of critical points of $\mathcal{M}_{\mathcal{U}}(\mathscr{F})$ *is a subset of the set of boundary points of open sets in* $\mathcal{U}$*.*

Proof We will first show that if $\mathscr{C}$ is a cosheaf on $\mathbb{R}$, then $\mathcal{M}_{\mathcal{U}}(\mathscr{C})$ is a cosheaf on $\mathbb{R}$. We have already shown that $\mathcal{M}_{\mathcal{U}}(\mathscr{C})$ is a pre-cosheaf. So all that remains is to prove the colimit property of cosheaves. Let $U \in \mathbf{Int}$ and $\mathcal{V} \subset \mathbf{Int}$ be a cover of U by open intervals which is closed under intersections. By definition of $\mathcal{M}_{\mathcal{U}}(\mathscr{C})$, we have

$$\varinjlim_{V \in \mathcal{V}} \mathcal{M}_{\mathcal{U}}(\mathscr{C})(V) = \varinjlim_{V \in \mathcal{V}} \mathscr{C}(\mathcal{I}_{\mathcal{U}}(V)).$$

Notice that $\mathcal{I}_{\mathcal{U}}(\mathcal{V}) := \{\mathcal{I}_{\mathcal{U}}(V) : V \in \mathcal{V}\}$ forms an open cover of $\mathcal{I}_{\mathcal{U}}(U)$. However, in general this cover is no longer closed under intersections. We will proceed by showing that passing from $\mathcal{V}$ to $\mathcal{I}_{\mathcal{U}}(\mathcal{V})' := \{\bigcap_{i \in I} W_i : \{W_i\}_{i \in I} \subset \mathcal{I}_{\mathcal{U}}(\mathcal{V})\}$ does not change the colimit

$$\varinjlim_{V \in \mathcal{V}} \mathscr{C}(\mathcal{I}_{\mathcal{U}}(V)).$$

Suppose I_1 and I_2 are two open intervals in $\mathcal{V}$ such that $I_1 \cap I_2 = \emptyset$ and $\mathcal{I}_{\mathcal{U}}(I_1) \cap \mathcal{I}_{\mathcal{U}}(I_2) \neq \emptyset$. Recall that $\mathcal{U}'$ is the union of $\mathcal{U}$ with all intersections of cover elements in $\mathcal{U}$, i.e., the closure of $\mathcal{U}$ under intersections. By the identification

$$\mathcal{I}_{\mathcal{U}}(I_i) = \bigcup_{x \in I_i} \bigcap_{V \in \mathcal{U}_x} V,$$

there exists a subset $\{W_j\}_{j \in J} \subset \mathcal{U}'$ such that

$$\mathcal{I}_{\mathcal{U}}(I_1) \cap \mathcal{I}_{\mathcal{U}}(I_2) = \bigcup_{j \in J} W_j.$$

Suppose there exist $V_1, V_2 \in \mathcal{U}'$ such that $V_i \subsetneq V_1 \cup V_2$ (i.e., one set is not a subset of the other), and $V_1 \cup V_2 \subset \mathcal{I}_{\mathcal{U}}(I_1) \cap \mathcal{I}_{\mathcal{U}}(I_2)$. In other words, suppose that the cardinality of J, for any suitable choice of indexing set, is strictly greater than 1. Then there exists $x_1, x_2 \in I_1$ such that $x_1 \in V_1 \setminus V_2$ and $x_2 \in V_2 \setminus V_1$. Let w either be a point contained in $V_1 \cap V_2$ (if $V_1 \cap V_2 \neq \emptyset$) or a point which lies between V_1 and V_2. Since I_1 is connected, we have that $w \in I_1$. A similar argument shows that $w \in I_2$, which implies the contradiction $I_1 \cap I_2 \neq \emptyset$. Therefore,

$$\mathcal{I}_{\mathcal{U}}(I_1) \cap \mathcal{I}_{\mathcal{U}}(I_2) = W,$$

for some $W \in \mathcal{U}'$. Suppose $W = \bigcap_{k \in K} W_k$ for some $\{W_k\}_{k \in K} \subset \mathcal{U}$, and let $I_1 = J_1, J_2, \ldots, J_n = I_2$ be a chain of open intervals in $\mathcal{V}$, such that $J_j \cap J_{j+1} \neq \emptyset$. We have that

$$I_1 \cup \bigcup_{k \in K} W_k \cup I_2$$

is connected, because I_1, I_2, and $\bigcup_{k\in K} W_k$ are intervals with $\bigcup_{k\in K} W_k \cap I_1$ and $\bigcup_{k\in K} W_k \cap I_2$ nonempty. Therefore, for each j, $J_j \cap W_k \neq \emptyset$ for some k, i.e., $W \subset \mathcal{I}_{\mathcal{U}}(J_j)$. In conclusion, we have shown that

$$\mathcal{I}_{\mathcal{U}}(I_1) \cap \mathcal{I}_{\mathcal{U}}(I_2) \subseteq \mathcal{I}_{\mathcal{U}}(J_j) \text{ for each } j.$$

Following the arguments in the proof of Proposition 4.17 of de Silva et al. (2016), it can be shown that

$$\varinjlim_{V\in\mathcal{V}} \mathscr{C}(\mathcal{I}_{\mathcal{U}}(V)) = \varinjlim_{U\in\mathcal{I}_{\mathcal{U}}(\mathcal{V})} \mathscr{C}(U) = \varinjlim_{U\in\mathcal{I}_{\mathcal{U}}(\mathcal{V})'} \mathscr{C}(U).$$

Since $\mathscr{C}$ is a cosheaf, we can use the colimit property of cosheaves to get

$$\varinjlim_{V\in\mathcal{V}} \mathcal{M}_{\mathcal{U}}(\mathscr{C})(V) = \mathscr{C}(\mathcal{I}_{\mathcal{U}}(U)).$$

Therefore $\mathcal{M}_{\mathcal{U}}(\mathscr{C})$ is cosheaf. We will proceed to show that $\mathcal{M}_{\mathcal{U}}(\mathscr{C})$ is constructible.

Let S be the set of boundary points for open sets in $\mathcal{U}$. Since $\mathcal{U}$ is a finite, good cover of $\mathbb{R}$, S is a finite set. If $U \subset V$ are two open sets in $\mathbb{R}$ such that $U \cap S = V \cap S$, then $\mathcal{I}_{\mathcal{U}}(U) = \mathcal{I}_{\mathcal{U}}(V)$. Therefore $\mathcal{M}_{\mathcal{U}}(\mathscr{F})(U) \to \mathcal{M}_{\mathcal{U}}(\mathscr{F})(V)$ is an isomorphism. □

We use the mapper functor to relate Reeb graphs (the display locale of the Reeb cosheaf $\mathscr{R}_f$) to the enhanced mapper graph (the display locale of $\mathcal{M}_{\mathcal{U}}(\mathscr{R}_f)$). In particular, the error is controlled by the resolution of the cover, as defined below.

Definition 14 Let $\mathcal{U}$ be a nice cover of $\mathbb{R}$ and $\mathscr{F}$ a cosheaf on $\mathbb{R}$. The *resolution* of $\mathcal{U}$ relative to $\mathscr{F}$, denoted $\operatorname{res}_{\mathscr{F}} \mathcal{U}$, is defined to be the maximum of the set of diameters of $\mathcal{U}_{\mathscr{F}} := \{V \in \mathcal{U} : \mathscr{F}(V) \neq \emptyset\}$:

$$\operatorname{res}_{\mathscr{F}} \mathcal{U} := \max\{\operatorname{diam}(V) : V \in \mathcal{U}_{\mathscr{F}}\}.$$

Here we understand the diameter of open sets of the form $(a, +\infty)$ or $(-\infty, b)$ to be infinite. Therefore, the resolution $\operatorname{res}_{\mathscr{F}} \mathcal{U}$ can take values in the extended non-negative numbers $\mathbb{R}_{\geq 0} \sqcup \{+\infty\}$.

Remark 4 If $\mathscr{R}_f$ is a Reeb cosheaf of a constructible $\mathbb{R}$-space $(\mathbb{X}, f)$, then $\mathscr{R}_f(V) \neq \emptyset$ if and only if $V \cap f(\mathbb{X}) \neq \emptyset$.

Definition 15 Define $\operatorname{res}_f \mathcal{U}$ by

$$\operatorname{res}_f \mathcal{U} := \max\{\operatorname{diam}(V) : V \in \mathcal{U}_f\},$$

where $\mathcal{U}_f := \{V \in \mathcal{U} : V \cap f(\mathbb{X}) \neq \emptyset\}$.

The following theorem is analogous to Munch and Wang (2016, Theorem 1), adapted to the current setting. Specifically, our definition of the mapper functor $\mathcal{M}_{\mathcal{U}}$ differs from the functor $\mathcal{P}_K$ of Munch and Wang (2016), and the convergence result of Munch and Wang (2016) is proved for multiparameter mapper (whereas the following result is only proved for the one-dimensional case).

Theorem 1 (cf. Munch and Wang 2016, Theorem 1) *Let $\mathcal{U}$ be a nice cover of $\mathbb{R}$, and $\mathscr{F}$ a cosheaf on $\mathbb{R}$. Then*

$$d_I(\mathscr{F}, \mathcal{M}_{\mathcal{U}}(\mathscr{F})) \leq \operatorname{res}_{\mathscr{F}} \mathcal{U}.$$

Proof If $\operatorname{res}_{\mathscr{F}} \mathcal{U} = +\infty$, then the inequality is automatically satisfied. Therefore, we will work with the assumption that $\operatorname{res}_{\mathscr{F}} \mathcal{U} < +\infty$. Let $\delta_{\mathcal{U}} = \operatorname{res}_{\mathscr{F}} \mathcal{U} < +\infty$. We will prove the theorem by constructing a $\delta_{\mathcal{U}}$-interleaving of the sheaves $\mathscr{F}$ and $\mathcal{M}_{\mathcal{U}}(\mathscr{F})$. Suppose $I \in \mathbf{Int}$. For each $x \in I$, let $W_x = \bigcap_{V \in \mathcal{U}_x} V$. Recall that

$$\mathcal{I}_{\mathcal{U}}(I) = \bigcup_{x \in I} W_x.$$

Ideally, we would construct an interleaving based on an inclusion of the form $\mathcal{I}_{\mathcal{U}}(I) \subset I_{\delta_{\mathcal{U}}}$. However, this inclusion will not always hold. For example, if $\mathcal{U}$ is a finite cover, then it is possible for I to be a bounded open interval, and for $\mathcal{I}_{\mathcal{U}}(I)$ to be unbounded.

We will include a simple example to illustrate this behavior. Suppose $\mathcal{U} = \{(-\infty, -1), (-2, 2), (1, +\infty)\}$ and let $\mathscr{F}$ be the constant cosheaf supported at 0, i.e. $\mathscr{F}(U) = \emptyset$ if $0 \notin U$ and $\mathscr{F}(V) = \{*\}$ if $0 \in V$. Consider the interval $I = (0, 3)$. For each $x \in (0, 1] \subset I$, we have that $W_x = (-2, 2)$. If $x \in (1, 2) \subset I$, then $W_x = (-2, 2) \cap (1, +\infty)$. Finally, if $x \in [2, 3) \subset I$, then $W_x = (1, +\infty)$. Therefore, $\mathcal{I}_{\mathcal{U}}(I) = (-2, +\infty)$, which is unbounded. However, we observe that $\mathscr{F}((-\infty, -1)) = \emptyset$, $\mathscr{F}((-2, 2)) = \{*\}$, and $\mathscr{F}((1, +\infty)) = \emptyset$. Therefore, (in the notation of Definition 14) $\mathcal{U}_{\mathscr{F}} = \{(-2, 2)\}$, and $\operatorname{res}_{\mathscr{F}} \mathcal{U} = \operatorname{diam}((-2, 2)) = 4$.

Although $\mathcal{I}_{\mathcal{U}}(I)$ may be unbounded, we can construct an interval I' which is contained in $I_{\delta_{\mathcal{U}}}$ and satisfies the equality $\mathscr{F}(\mathcal{I}_{\mathcal{U}}(I)) = \mathscr{F}(I')$. The remainder of the proof will be dedicated to constructing such an interval.

Let $\mathcal{W} := \{U : U = \cap_{a \in A} W_a \text{ for some } A \subset I\}$ be an open cover of $\mathcal{I}_{\mathcal{U}}(I)$ which is closed under intersections and generated by the open sets W_x. Then the colimit property of cosheaves gives us the equality

$$\mathscr{F}(\mathcal{I}_{\mathcal{U}}(I)) = \varinjlim_{U \in \mathcal{W}} \mathscr{F}(U).$$

Let $E := \{e \in I : \mathscr{F}(W_e) = \emptyset\}$. If $U = \cap_{a \in A} W_a$ and $A \cap E \neq \emptyset$, then $\mathscr{F}(U) = \emptyset$. Let $\mathcal{W}_{I \setminus E} = \{U \in \mathcal{W} : U = \cap_{a \in A} W_a \text{ for some } A \subset I \setminus E\}$. We should remark on a small technical matter concerning $I \setminus E$. In general, this set is not necessarily connected. If that is the case, we should replace $I \setminus E$ with the minimal interval which covers $I \setminus E$. Going forward, we will assume that $I \setminus E$ is connected. Altogether we have

$$\mathcal{M}_{\mathcal{U}}(\mathscr{F})(I) = \mathscr{F}(\mathcal{I}_{\mathcal{U}}(I)) = \varinjlim_{U \in \mathcal{W}} \mathscr{F}(U) = \varinjlim_{U \in \mathcal{W}_{I \setminus E}} \mathscr{F}(U) = \mathscr{F}\left(\bigcup_{x \in I \setminus E} W_x \right).$$

If $x \in I \setminus E$, then $W_x \cap I \neq \emptyset$ and $\mathscr{F}(W_x) \neq \emptyset$. Therefore, $W_x \subseteq I_{\delta_{\mathcal{U}}}$, since $\operatorname{diam}(W_x) \leq \delta_{\mathcal{U}}$. Moreover,

$$\bigcup_{x \in I \setminus E} W_x \subseteq I_{\delta_{\mathcal{U}}}.$$

The above inclusion induces the following map of sets

$$\varphi_I : \mathcal{M}_{\mathcal{U}}(\mathscr{F})(I) \to \mathscr{F}(I_{\delta_{\mathcal{U}}}),$$

which gives the first family of maps of the $\delta_{\mathcal{U}}$-interleaving. The second family of maps

$$\psi_I : \mathscr{F}(I) \to \mathcal{M}_{\mathcal{U}}(\mathscr{F})(I_{\delta_{\mathcal{U}}}),$$

follows from the more obvious inclusion $I \subset \mathcal{I}_{\mathcal{U}}(I_{\delta_{\mathcal{U}}})$. Since the interleaving maps are defined by inclusions of intervals, it is clear that the composition formulae are satisfied:

$$\psi_{I_{\delta_{\mathcal{U}}}} \circ \varphi_I = \mathscr{F}[I \subset I_{2\delta_{\mathcal{U}}}], \qquad \varphi_{I_{\delta_{\mathcal{U}}}} \circ \psi_I = \mathcal{M}_{\mathcal{U}}(\mathscr{F})\left[I \subset I_{2\delta_{\mathcal{U}}}\right].$$

□

Remark 5 One might think that Theorem 1 can be used to obtain a convergence result for the mapper graph of a general $\mathbb{R}$-space. However, we should emphasize that the interleaving distance is only an extended *pseudo*-metric on the category of all cosheaves. Therefore, even if the interleaving distance between $\mathscr{F}$ and $\mathcal{M}_{\mathcal{U}}(\mathscr{F})$ goes to 0, this does not imply that the cosheaves are isomorphic. We only obtain a convergence result when restricting to the subcategory of constructible cosheaves, where the interleaving distance gives an extended metric.

The display locale $\mathfrak{D}(\mathcal{M}_{\mathcal{U}}(\mathscr{R}_f))$ of the mapper cosheaf is a 1-dimensional CW-complex obtained by gluing the boundary points of a finite disjoint union of closed intervals, see Fig. 1h. We will refer to this CW-complex as the *enhanced mapper graph* of $(\mathbb{X}, f)$ relative to $\mathcal{U}$, see Fig. 1g. There is a natural surjection from $\mathfrak{D}(\mathcal{M}_{\mathcal{U}}(\mathscr{R}_f))$ to the nerve of the connected cover pull-back of $\mathcal{U}$, $\mathcal{N}_{f^*(\mathcal{U})}$, i.e., from the enhanced mapper graph to the mapper graph, when the cover $\mathcal{U}$ contains open sets with empty triple intersections.

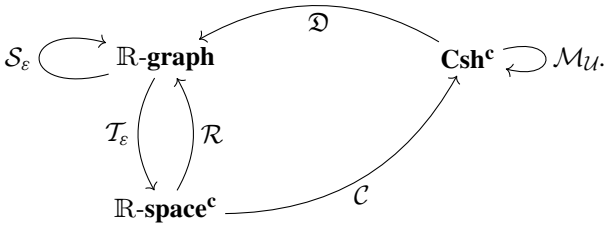
Using the Reeb interleaving distance and the enhanced mapper graph, we obtain and reinterpret the main result of Munch and Wang (2016) in the following corollary.

Corollary 1 (cf. Munch and Wang 2016, Corollary 6) *Let $\mathcal{U}$ be a nice cover of $\mathbb{R}$, and $(\mathbb{X}, f) \in \mathbb{R}$-**space**$^{\mathbf{c}}$. Then*

$$d_R(\mathcal{R}(\mathbb{X}, f), \mathfrak{D}(\mathcal{M}_{\mathcal{U}}(\mathscr{R}_f))) \leq \operatorname{res}_f \mathcal{U}.$$

Throughout this section we introduce several categories and functors which we will now summarize. Let $\mathbb{R}$-**graph** be the category of $\mathbb{R}$-graphs (i.e., Reeb graphs),

$\mathbb{R}$-**space**$^{\mathbf{c}}$ the category of constructible $\mathbb{R}$-spaces, **Csh**$^{\mathbf{c}}$ be the category of constructible cosheaves on $\mathbb{R}$, $\mathcal{S}_\varepsilon$ and $\mathcal{T}_\varepsilon$ the smoothing and thickening functors, $\mathfrak{D}$ the display locale functor, and $\mathcal{M}_\mathcal{U}$ the mapper functor. Altogether, we have the following diagram of functors and categories,



Enhanced mapper graph algorithm Finally, we briefly describe an algorithm for constructing the enhanced mapper graph, following the example in Fig. 1. Let $(\mathbb{X}, f)$ be a constructible $\mathbb{R}$-space (see Sect. 2.1). For simplicity, suppose that the cover $\mathcal{U}$ consists of open intervals, and contains no nonempty triple intersections ($U \cap V \cap W = \emptyset$ for all $U, V, W \in \mathcal{U}$). Let $\mathbb{R}_0$ be the union of boundary points of cover elements in the open cover $\mathcal{U}$. Let $\mathbb{R}_1$ be the complement of $\mathbb{R}_0$ in $\mathbb{R}$. The set $\mathbb{R}_0$ is illustrated with gray dots in Fig. 1e. We begin by forming the disjoint union of closed intervals,

$$\coprod_I \overline{I} \times \pi_0(f^{-1}(U_I)),$$

where the disjoint union is taken over all connected components I of $\mathbb{R}_1$, $\overline{I}$ denotes the closure of the open interval I, and U_I denotes the smallest open set in $\mathcal{U} \cup \{U \cap V \mid U, V \in \mathcal{U}\}$ which contains I. In other words, U_I is either the intersection of two cover elements in $\mathcal{U}$ or U_I is equal to a cover element in $\mathcal{U}$. The sets $\pi_0(f^{-1}(U_I))$ are illustrated in Fig. 1d. Notice that there is a natural projection map from the disjoint union to $\mathbb{R}$, given by projecting each point (y, a) in the disjoint union onto the first factor, $y \in \mathbb{R}$. The enhanced mapper graph is a quotient of the above disjoint union by an equivalence relation on endpoints of intervals. This equivalence relation is defined as follows. Let $(y, a) \in \overline{I} \times \pi_0(f^{-1}(U_I))$ and $(z, b) \in \overline{J} \times \pi_0(f^{-1}(U_J))$ be two elements of the above disjoint union. If $y \in \mathbb{R}_0$, then y is contained in exactly one cover element in $\mathcal{U}$, denoted by U_y. Moreover,if $y \in \mathbb{R}_0$, then there is a map $\pi_0(f^{-1}(U_I)) \to \pi_0(f^{-1}(U_y))$ induced by the inclusion $U_I \subseteq U_y$. Denote this map by $\psi_{(y,I)}$. An analogous map can be constructed for (z, b), if $z \in \mathbb{R}_0$. We say that $(y, a) \sim (z, b)$ if two conditions hold: $y = z$ is contained in $\mathbb{R}_0$, and $\psi_{(y,I)}(a) = \psi_{(z,J)}(b)$. The *enhanced mapper graph* is the quotient of the disjoint union by the equivalence relation described above.

For example, as illustrated in Fig. 1, seven cover elements of $\mathcal{U}$ in (c) give rise to a stratification of $\mathbb{R}$ into a set of points $\mathbb{R}_0$ and a set of intervals $\mathbb{R}_1$ in (e). For each interval I in $\mathbb{R}_1$, we look at the set of connected components in $f^{-1}(U_I)$. We then construct disjoint unions of closed intervals based on the cardinality of $\pi_0(f^{-1}(U_I))$ for each $I \in \mathbb{R}_1$. For adjacent intervals I_1 and I_2 in $\mathbb{R}_1$, suppose that I_1 is contained in the cover element V and I_2 is equal to the intersection of cover elements V and

W in $\mathcal{U}$. We consider the mapping from $\pi_0(f^{-1}(U_{I_2}))$ to $\pi_0(f^{-1}(U_{I_1}))$ (d). Here, we have that $U_{I_2} = V \cap W$ and $U_{I_1} = V$. We then glue these closed intervals following the above mapping, which gives rise to the enhanced mapper graph (g). Appendix A outlines these algorithmic details in the form of pseudocode.

3 Model

Let $\mathbb{X}$ be a compact locally path connected subset of $\mathbb{R}^d$. As stated in the introduction, study related to topological inference usually splits between noiseless and noisy settings. In the former, we assume that a given sample is drawn from $\mathbb{X}$ directly, while in the latter we allow random perturbations that produce samples in $\mathbb{R}^d$ that need not be on $\mathbb{X}$, but rather in its vicinity. In this paper, we address the noisy setting directly, using the machinery for super-level sets estimation developed in Bobrowski et al. (2017b). The basic inputs are a continuous function $f : \mathbb{R}^d \to \mathbb{R}$, and a probability density function $p : \mathbb{R}^d \to \mathbb{R}$. Our $\mathbb{R}$-space of interest will be $(\mathbb{X}, f|_{\mathbb{X}})$, and we will assume we are provided samples of $\mathbb{X}$ via p. Then, given a nice cover $\mathcal{U}$, we can compare the Reeb graph of $(\mathbb{X}, f|_{\mathbb{X}})$ to the mapper graph computed from the samples.

3.1 Setup

In this section, we give some basic assumptions on f, p, $\mathcal{U}$, and their interactions. We start with some notation for the various sets of interest. Let $\mathbb{X}_\delta = \{y \in \mathbb{R}^d : \inf_{x\in\mathbb{X}} d(x, y) \leq \delta\}$ be the δ-thickening of $\mathbb{X}$, and let $D_L = p^{-1}([L, +\infty))$ be a super-level set of p. Given an open set $V \subset \mathbb{R}$, define $\mathbb{X}^V := \mathbb{X} \cap f^{-1}(V)$. Let $\mathbb{X}_\delta^V := \mathbb{X}_\delta \cap f^{-1}(V)$ be the elements of $\mathbb{X}_\delta$ which map to V, and $D_L^V := D_L \cap f^{-1}(V)$. See Fig. 3 for an example of this notation. It is important to note that $\mathbb{X}_\delta^V$ is not necessarily equal to the δ-thickening of $\mathbb{X}^V$.

With this notation, we will assume that p is ε-concentrated on $\mathbb{X}$ as defined next with an example given in Fig. 4.

Definition 16 A probability density function p is *ε-concentrated on* $\mathbb{X}$ if there exists open intervals I_1, I_2, and a real number $\delta > 0$ such that

$$\mathbb{X} \subset D_{L_1} \subset \mathbb{X}_\delta \subset D_{L_2} \subset \mathbb{X}_\varepsilon,$$

for any $L_1 \in I_1$ and $L_2 \in I_2$.

Definition 17 A probability density function p is *concentrated on* $\mathbb{X}$ if p is ε-concentrated on $\mathbb{X}$ for all $\varepsilon > 0$.

We now turn our attention to $\mathcal{U}$, a nice cover of $\mathbb{R}$.

Definition 18 The *local H_0-critical value over V* is defined as

$$\delta_V = \sup\{\delta \mid H_0(\mathbb{X}^V) \xrightarrow{\cong} H_0(\mathbb{X}_\delta^V)\}.$$

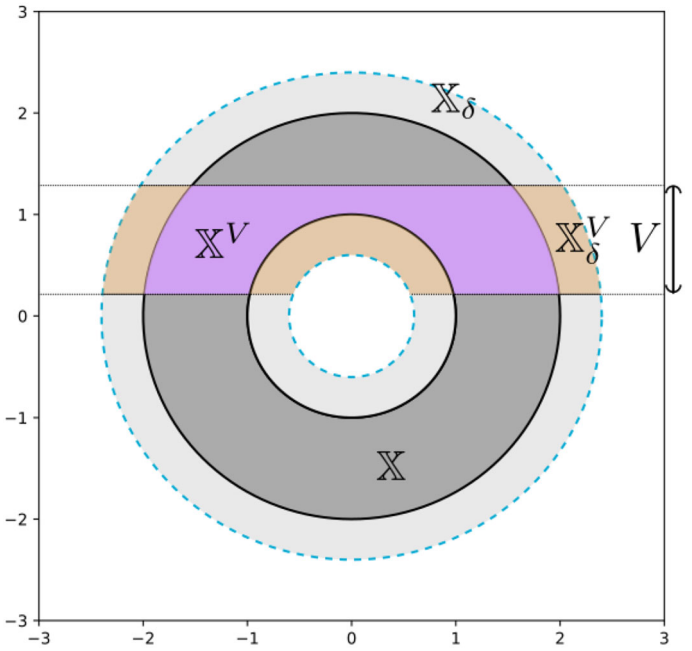


Fig. 3 This figure illustrates the inverse images $\mathbb{X}^V$ (in purple) and $\mathbb{X}_\delta^V$ (union of tan and purple) for an annulus with height function and open interval V. Notice that in this example the δ-thickening of $\mathbb{X}^V$ would include $\mathbb{X}_\delta^V$ as a proper subset, hence $(\mathbb{X}^V)_\delta \neq \mathbb{X}_\delta^V$

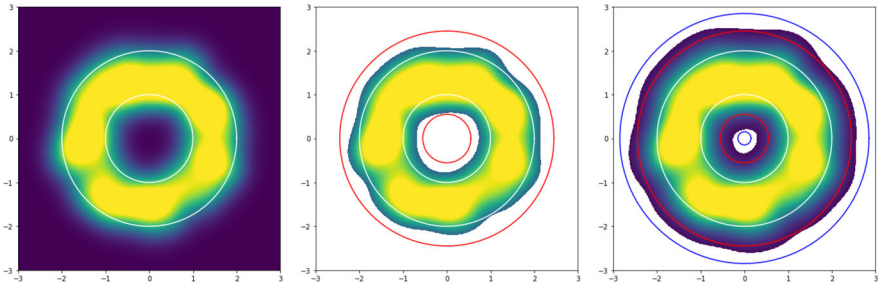


Fig. 4 An example of the concentrated definition. The left side of the figure illustrates a probability density function (PDF) p which is ε-concentrated on an annulus $\mathbb{X}$. The center image illustrates the thickened space $\mathbb{X}_\delta$, bounded by the red curves, and the super-level set D_{L_1}. The right side of the figure illustrates the thickened space $\mathbb{X}_\varepsilon$, bounded by the blue curves, and the super-level set D_{L_2}. Together, we see that $\mathbb{X} \subset D_{L_1} \subset \mathbb{X}_\delta \subset D_{L_2} \subset \mathbb{X}_\varepsilon$ (color figure online)

Let $\mathcal{U}' := \{V \subset \mathbb{R} : V = \cap_{\alpha \in A} U_\alpha \text{ for some } \{U_\alpha\}_{\alpha \in A} \subset \mathcal{U}\}$. The *global* H_0-*critical value over* $\mathcal{U}$ is defined as

$$\delta_{\mathcal{U}} := \min_{V \in \mathcal{U}'} \delta_V .$$

Throughout the paper, we assume that the global H_0-critical value over $\mathcal{U}$ is positive, i.e.

$$\delta_{\mathcal{U}} = \min_{V \in \mathcal{U}'} \delta_V > 0.$$

The positivity of local H_0-critical values is nontrivial and often fails for constructible $\mathbb{R}$-spaces which have singularities which lie over the boundary of one of the open sets in the open cover $\mathcal{U}$. In future work, it would be interesting to relax this assumption, and study convergence when the diameter of the union of open sets V for which $\delta_V = 0$, is small.

3.2 Approximation by super-level sets

In this section, we study how super-level sets of probability density functions (PDFs) can model the topology of constructible $\mathbb{R}$-spaces.

We need some further control over the relationship between the PDF p and the cover elements via the following definition.

Definition 19 Given an open set V, we say that L is H_0-*regular over* V if there exists $\nu > 0$ such that for all $\varepsilon_1 < \varepsilon_2 \in (L - \nu, L + \nu)$, the inclusion $D_{\varepsilon_2} \subset D_{\varepsilon_1}$ induces an isomorphism $H_0(D^V_{\varepsilon_2}) \xrightarrow{\cong} H_0(D^V_{\varepsilon_1})$.

Throughout the paper we will assume that the PDF p is *tame*, in the sense that the set of points which are H_0-regular over V is dense in $\mathbb{R}$, for any given open set V.

Assume the global H_0-critical value $\delta_{\mathcal{U}}$ is positive, and p is δ_2-concentrated on $\mathbb{X}$ for some δ_2 such that $0 < \delta_2 < \delta_{\mathcal{U}}$. By definition, there exist L_1, L_2 and δ_1 such that

1. $\mathbb{X} \subset D_{L_1} \subset \mathbb{X}_{\delta_1} \subset D_{L_2} \subset \mathbb{X}_{\delta_2}$
2. $0 < \delta_1 < \delta_2 < \delta_{\mathcal{U}}$
3. L_1 and L_2 are H_0-regular over V for each $V \in \mathcal{U}'$.

The set of points which are H_0-regular over V for each $V \in \mathcal{U}'$ is dense in $\mathbb{R}$. If L_1 is not H_0-regular over V for some $V \in \mathcal{U}'$, then L_1 can be turned into a regular value with an arbitrarily small perturbation. Moreover, by the Definition 16, this perturbation can be done without breaking the chain of inclusions $\mathbb{X} \subset D_{L_1} \subset \mathbb{X}_{\delta_1} \subset D_{L_2} \subset \mathbb{X}_{\delta_2}$. We therefore continue under the assumption that L_1 is H_0-regular over V for each $V \in \mathcal{U}'$.

Proposition 5 *Assume that p is ε-concentrated on $\mathbb{X}$ for some $\varepsilon < \delta_{\mathcal{U}}$. Let*

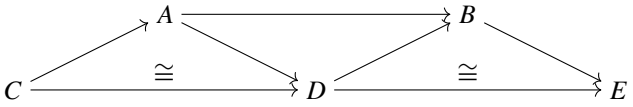
$$\mathscr{D}(V) := \mathrm{Im}\left(H_0(D^V_{L_1}) \to H_0(D^V_{L_2})\right).$$

Then for each $V \in \mathcal{U}'$, we have $H_0(\mathbb{X}^V) \cong \mathscr{D}(V)$ and further for each $V \subset W \in \mathcal{U}'$ the following diagram commutes,

$$\begin{array}{ccc} H_0(\mathbb{X}^V) & \xrightarrow{\cong} & \mathscr{D}(V) \\ \downarrow & & \downarrow \\ H_0(\mathbb{X}^U) & \xrightarrow{\cong} & \mathscr{D}(U). \end{array}$$

The proof will require the following technical lemma.

Lemma 2 *Suppose we have the following commutative diagram of vector spaces*



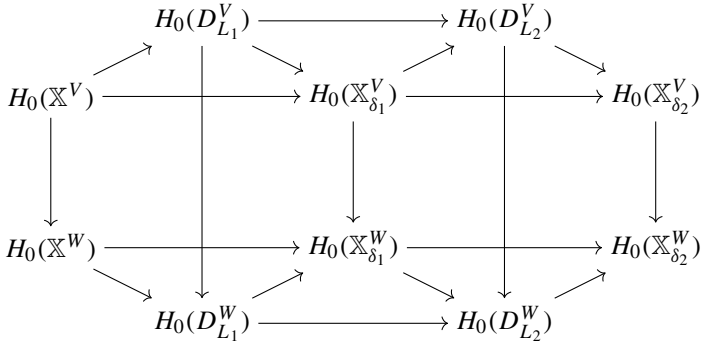
with $C \cong D \cong E$. Then $\mathrm{Im}(D \to B) = \mathrm{Im}(A \to B)$ and the map

$$D \xrightarrow{\cong} \mathrm{Im}(A \to B)$$

is an isomorphism of vector spaces.

Proof The map $D \to B$ is injective since $D \to E$ is an isomorphism and the diagram commutes. Therefore, $\mathrm{Im}(D \to B) \cong D$. Moreover, since the diagram commutes, $\mathrm{Im}(A \to B) \subset \mathrm{Im}(D \to B)$. Suppose $b \in \mathrm{Im}(D \to B)$, i.e., there exists $d \in D$ which maps to b. Since $C \to D$ is an isomorphism, there exists $c \in C$ which maps to $d \in D$. Let $a \in A$ be the image of $c \in C$ under the map $C \to A$. Since the diagram commutes, we have that $a \in A$ maps to $b \in B$ under the map $A \to B$. Therefore, $b \in \mathrm{Im}(A \to B)$. We have therefore shown that $\mathrm{Im}(A \to B) = \mathrm{Im}(D \to B) \cong D$. □

Proof of 5 Choose $\delta_2 > 0$ such that $\delta_2 < \delta_{\mathcal{U}}$ and p is δ_2-concentrated on $\mathbb{X}$. Applying the definition of δ_2-concentrated, we have $\mathbb{X} \subset D_{L_1} \subset \mathbb{X}_{\delta_1} \subset D_{L_2} \subset \mathbb{X}_{\delta_2}$. For $V \subset W$ we have the following commutative diagram of vector spaces



Since $\delta_1 < \delta_2 < \delta_{\mathcal{U}}$, by definition of global H_0-critical value over $\mathcal{U}$, all four horizontal maps

$$H_0(\mathbb{X}^V) \longrightarrow H_0(\mathbb{X}^V_{\delta_1}) \longrightarrow H_0(\mathbb{X}^V_{\delta_2}) \quad \text{and} \quad H_0(\mathbb{X}^W) \longrightarrow H_0(\mathbb{X}^W_{\delta_1}) \longrightarrow H_0(\mathbb{X}^W_{\delta_2})$$

are isomorphisms. Applying 2, we can conclude that

$$H_0(\mathbb{X}^V) \longrightarrow \mathscr{D}(V) \quad \text{and} \quad H_0(\mathbb{X}^W) \longrightarrow \mathscr{D}(W)$$

are isomorphisms of vector spaces. Since the diagram commutes, the image of $\mathscr{D}(V)$ under the map $H_0(D^V_{L_2}) \to H_0(D^W_{L_2})$ is contained in $\mathscr{D}(W)$. Therefore, $H_0(D^V_{L_2}) \to H_0(D^W_{L_2})$ induces a map $\mathscr{D}(V) \to \mathscr{D}(W)$, which completes the commutative diagram of the theorem. □

3.3 Point-cloud mapper algorithm

Given data $\{X_1, \ldots, X_n\} \overset{\text{i.i.d}}{\sim} p$, where p is a PDF, we can estimate p using a kernel density estimator (KDE) of the form,

$$\hat{p}(x) := \frac{1}{C_K n r^d} \sum_{i=1}^{n} K_r(x - X_i),$$

where $K(x)$ is a given kernel function, $K_r := K(x/r)$, and C_K is a constant defined below. The kernel function should satisfy the following:

1. $\text{supp}(K) \subset B_1(0)$, and $K(x)$ is smooth in $B_1(0)$.
2. $K(x) \in [0, 1]$, and $\max_x K(x) = K(0) = 1$,
3. $\int_{\mathbb{R}^d} K(x)dx = C_K$ with $C_K \in (0, 1)$.

Using $\hat{p}$ we can estimate the super-level sets D_L using

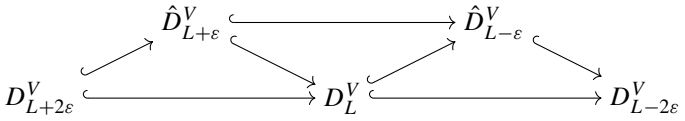
$$\hat{D}_L(n, r) := \bigcup_{i:\hat{p}(X_i)\geq L} B_r(X_i), \tag{1}$$

and the sets D^V_L using

$$\hat{D}^V_L(n, r) := \hat{D}_L(n, r) \cap f^{-1}(V). \tag{2}$$

Choose ε_i such that $L_i + 2\varepsilon_i$, $L_i - 2\varepsilon_i$ are within the H_0-regularity range of L_i over V for each $V \in \mathcal{U}$ and $L_1 - 2\varepsilon_1 > L_2 + 2\varepsilon_2$. In the following, we will use the term "with high probability" (w.h.p.) to mean that the probability of an event to occur converges to 1 as $n \to \infty$.

Proposition 6 *Fix L and V, and set $\hat{D}^V_L := \hat{D}^V_L(n, r)$. Fixing $\varepsilon > 0$, there exists a constant $C_\varepsilon > 0$ (independent of L and V) such that if $nr^d \geq C_\varepsilon \log n$, then the following diagram of inclusion relations holds w.h.p.,*



Proof The proof appears as part of the proof of Theorem 3.3 in Bobrowski et al. (2017b). □

Next, define the random vector space

$$\hat{\mathscr{D}}_i(V) := \mathrm{Im}\left(H_0(\hat{D}^V_{L_i+\varepsilon_i}) \to H_0(\hat{D}^V_{L_i-\varepsilon_i})\right).$$

Corollary 2 *If $nr^d \geq C_{\varepsilon_i} \log n$, then w.h.p. the random map*

$$H_0(D^V_{L_i}) \to \hat{\mathscr{D}}_i(V)$$

is an isomorphism.

Proof The corollary follows from applying 2–6. □

From here on, unless otherwise stated, we will assume that r is chosen so that $nr^d \geq \max(C_{\varepsilon_1}, C_{\varepsilon_2}) \log n$, so that 2 holds for both $\varepsilon_1, \varepsilon_2$.

Proposition 7 *For every $V \subset W \in \mathcal{U}'$, we have the following commutative diagram w.h.p.,*

$$\begin{array}{ccc} H_0(D^V_{L_i}) & \xrightarrow{\cong} & \hat{\mathscr{D}}_i(V) \\ \downarrow & \circlearrowright & \downarrow \\ H_0(D^W_{L_i}) & \xrightarrow{\cong} & \hat{\mathscr{D}}_i(W). \end{array}$$

Proof The proof follows the same arguments as the proof of Proposition 5, and using Corollary 2. □

Finally, we define the following random vector space,

$$\hat{\mathscr{D}}(V) := \mathrm{Im}\left(H_0(\hat{D}^V_{L_1+\varepsilon_1}) \to H_0(\hat{D}^V_{L_2-\varepsilon_2})\right).$$

Proposition 8 *Assume that p is ε-concentrated on $\mathbb{X}$ for some $\varepsilon < \delta_{\mathcal{U}}$. For every $V \subset W \in \mathcal{U}'$, we have the following commutative diagram w.h.p.,*

$$\begin{array}{ccc} H_0(\mathbb{X}^V) & \xrightarrow{\cong} & \hat{\mathscr{D}}(V) \\ \downarrow & \circlearrowright & \downarrow \\ H_0(\mathbb{X}^W) & \xrightarrow{\cong} & \hat{\mathscr{D}}(W), \end{array}$$

where the constants L_1 and L_2 (defining $\hat{\mathscr{D}}$) are given by the definition of ε-concentrated, and the constants ε_1 and ε_2 are given by the H_0-regularity of L_1 and L_2, respectively.

Proof We will use the assumption that $L_1 - 2\varepsilon_1 > L_2 + 2\varepsilon_2$ repeatedly for each of the super-level set inclusions in the proof. The inclusion of spaces $\hat{D}^V_{L_1-\varepsilon_1} \subset \hat{D}^V_{L_2-\varepsilon_2}$ induces a homomorphism $H_0(\hat{D}^V_{L_1-\varepsilon_1}) \to H_0(\hat{D}^V_{L_2-\varepsilon_2})$. Restricting the domain of this map, we get a homomorphism $\hat{\mathscr{D}}_1(V) \to H_0(\hat{D}^V_{L_2-\varepsilon_2})$. Since $L_1 - \varepsilon_1 > L_2 + \varepsilon_2 > L_2 - \varepsilon_2$, the map $\hat{\mathscr{D}}_1(V) \to H_0(\hat{D}^V_{L_2-\varepsilon_2})$ factors through $H_0(\hat{D}^V_{L_2+\varepsilon_2}) \to H_0(\hat{D}^V_{L_2-\varepsilon_2})$, forming the commutative diagram

$$\begin{array}{ccc} \hat{\mathscr{D}}_1(V) & \longrightarrow & H_0(\hat{D}^V_{L_2-\varepsilon_2}) \\ \downarrow & \circlearrowleft & \nearrow \\ H_0(\hat{D}^V_{L_2+\varepsilon_2}) & & \end{array}$$

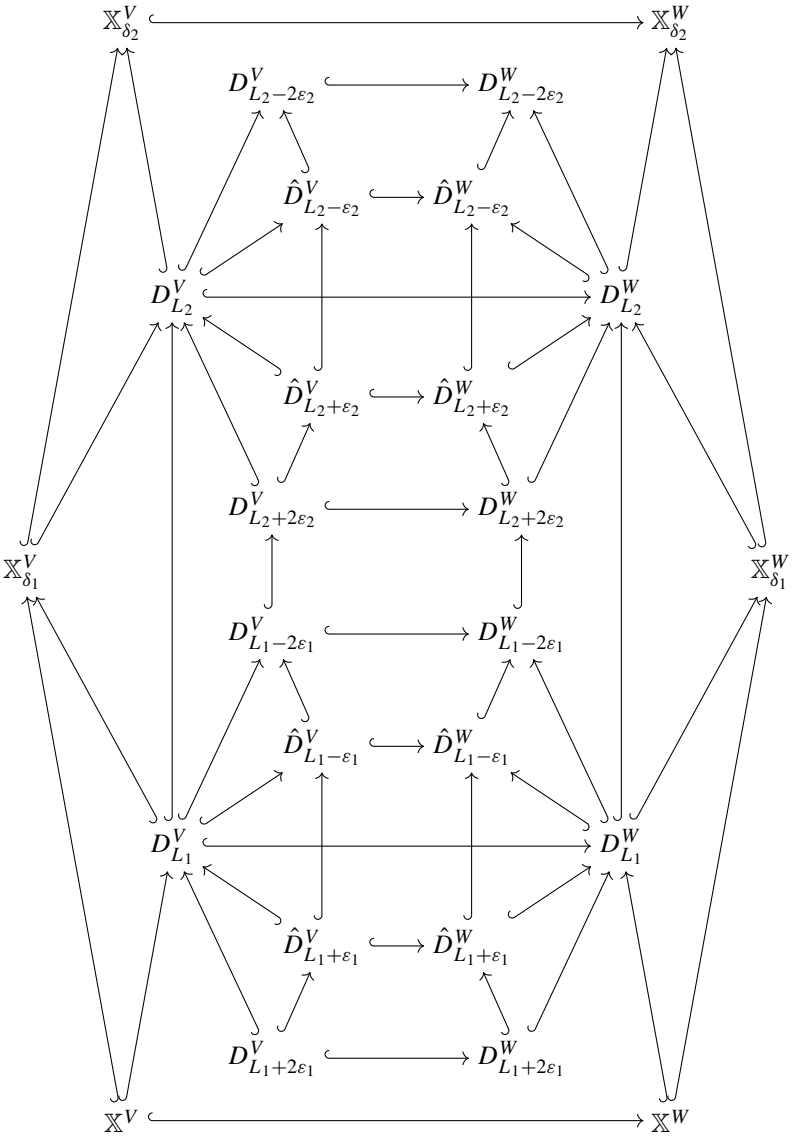
This implies that $\mathrm{Im}(\hat{\mathscr{D}}_1(V) \to H_0(\hat{D}^V_{L_2-\varepsilon_2})) \subset \hat{\mathscr{D}}_2(V)$, and gives a map $\hat{\mathscr{D}}_1(V) \to \hat{\mathscr{D}}_2(V)$, which w.h.p. completes the following commutative diagram,

$$\begin{array}{ccc} H_0(D^V_{L_1}) & \xrightarrow{\cong} & \hat{\mathscr{D}}_1(V) \\ \downarrow & \circlearrowleft & \downarrow \\ H_0(D^V_{L_2}) & \xrightarrow{\cong} & \hat{\mathscr{D}}_2(V) \end{array}$$

where the horizontal maps are given by Corollary 2. Therefore, applying Proposition 5, we have

$$H_0(\mathbb{X}^V) \xrightarrow{\cong} \mathscr{D}(V) \overset{\text{w.h.p.}}{\cong} \mathrm{Im}\left(\hat{\mathscr{D}}_1(V) \to \hat{\mathscr{D}}_2(V)\right).$$

Since $\hat{D}^V_{L_1+\varepsilon_1} \subset \hat{D}^V_{L_1-\varepsilon_1} \subset \hat{D}^V_{L_2+\varepsilon_2} \subset \hat{D}^V_{L_2-\varepsilon_2}$, we have that $\mathrm{Im}\left(\hat{\mathscr{D}}_1(V) \to \hat{\mathscr{D}}_2(V)\right) = \hat{\mathscr{D}}(V)$. The map $\hat{\mathscr{D}}(V) \to \hat{\mathscr{D}}(W)$ in the statement of the proposition (and the commutativity of the resulting diagram) is induced by the inclusion $\hat{D}^V_{L_2-\varepsilon_2} \hookrightarrow \hat{D}^W_{L_2-\varepsilon_2}$ in the following commutative diagram.



□

4 Main results

In this section, we prove convergence of the random enhanced mapper graph to the Reeb graph, as well as stability of the enhanced mapper graph under certain perturbations of the corresponding real valued function. Using the model described in Sect. 3, we generate random data, which is used to define a cosheaf which estimates

the connected components of fibers of the real valued function associated to a given constructible $\mathbb{R}$-space. In the proof of Theorem 2, we show that, with high probability, the cosheaf constructed using random data is isomorphic to the mapper functor applied to the Reeb cosheaf defined in Sect. 2. We then use the results established in Sect. 2 to translate the cosheaf theoretic statement into a geometric statement (Corollary 3) for the corresponding $\mathbb{R}$-graphs.

To begin, we identify a sufficient condition for determining when a morphism of constructible cosheaves is an isomorphism. A morphism $\mathscr{F} \to \mathscr{G}$ of cosheaves is a family of maps $\mathscr{F}(V) \to \mathscr{G}(V)$, for each open set V, which form a commutative diagram

$$\begin{array}{ccc} \mathscr{F}(V) & \longrightarrow & \mathscr{G}(V) \\ \downarrow & \circlearrowleft & \downarrow \\ \mathscr{F}(W) & \longrightarrow & \mathscr{G}(W) \end{array}$$

for each pair of open sets $V \subset W$. The morphism $\mathscr{F} \to \mathscr{G}$ is an isomorphism if each of the maps $\mathscr{F}(V) \to \mathscr{G}(V)$ is an isomorphism. Our first result shows that for cosheaves of the form $\mathcal{M}_{\mathcal{U}}(\mathscr{F})$, it is sufficient to consider only the maps $\mathcal{M}_{\mathcal{U}}(\mathscr{F})(V) \to \mathcal{M}_{\mathcal{U}}(\mathscr{G})(V)$ for open sets $V \in \mathcal{U}'$.

Proposition 9 *Let $\mathscr{C}$ and $\mathscr{D}$ be cosheaves on $\mathbb{R}$. An isomorphism $\mathcal{M}_{\mathcal{U}}(\mathscr{C}) \to \mathcal{M}_{\mathcal{U}}(\mathscr{D})$ of cosheaves is uniquely determined by a family of isomorphisms $\mathcal{M}_{\mathcal{U}}(\mathscr{C})(V) \to \mathcal{M}_{\mathcal{U}}(\mathscr{D})(V)$ for each $V \in \mathcal{U}'$, which form a commutative diagram*

$$\begin{array}{ccc} \mathcal{M}_{\mathcal{U}}(\mathscr{C})(V) & \xrightarrow{\cong} & \mathcal{M}_{\mathcal{U}}(\mathscr{D})(V) \\ \downarrow & \circlearrowleft & \downarrow \\ \mathcal{M}_{\mathcal{U}}(\mathscr{C})(W) & \xrightarrow{\cong} & \mathcal{M}_{\mathcal{U}}(\mathscr{D})(W) \end{array}$$

for each pair $V \subset W \in \mathcal{U}'$.

Proof Proposition 4 shows that $\mathcal{M}_{\mathcal{U}}(\mathscr{C})$ and $\mathcal{M}_{\mathcal{U}}(\mathscr{D})$ are constructible cosheaves over $\mathbb{R}$. The proof then follows from de Silva et al. (2016, Proposition 3.10). □

Recalling the notation of Sects. 2 and 3, for a super-level set D_L of p, let $\mathscr{R}_{D_L}$ be the Reeb cosheaf of (D_L, f) on $\mathbb{R}$, defined by

$$\mathscr{R}_{D_L}(U) = \pi_0(D_L^U)$$

for each open set $U \subset \mathbb{R}$. Let $\mathscr{R}_{\hat{D}_L}$ be the Reeb cosheaf of $(\hat{D}_L, f)$ on $\mathbb{R}$, defined by

$$\mathscr{R}_{\hat{D}_L}(U) = \pi_0(\hat{D}_L^U)$$

where $\hat{D}_L$, $\hat{D}_L^U$ are defined in (1), (2), respectively, and $U \subset \mathbb{R}$ is an open set. We should note that (D_L, f) and $(\hat{D}_L, f)$ are not apriori constructible spaces, so the cosheaves $\mathscr{R}_{D_L}$ and $\mathscr{R}_{\hat{D}_L}$ are not necessarily constructible. However, in what follows we will work exclusively with $\mathcal{M}_{\mathcal{U}}(\mathscr{R}_{D_L})$ and $\mathcal{M}_{\mathcal{U}}(\mathscr{R}_{\hat{D}_L})$, which are constructible cosheaves by Proposition 4.

Let $\hat{\mathscr{D}}_n^{\pi}$ be the cosheaf defined by

$$\hat{\mathscr{D}}_n^{\pi} := \mathcal{M}_{\mathcal{U}}\left(\mathrm{Im}\left(\mathscr{R}_{\hat{D}_{L_1+\varepsilon_1}} \to \mathscr{R}_{\hat{D}_{L_2-\varepsilon_2}}\right)\right),$$

with constants n, L_1, L_2, ε_1, and ε_2 chosen in Sect. 3. More explicitly, $\hat{\mathscr{D}}_n^{\pi}$ maps an open interval U to elements of the set $\mathscr{R}_{\hat{D}_{L_2-\varepsilon_2}}(\mathcal{I}_{\mathcal{U}}(U))$ which lie in the image of the set $\mathscr{R}_{\hat{D}_{L_1+\varepsilon_1}}(\mathcal{I}_{\mathcal{U}}(U))$ under the map induced by the inclusion $\hat{D}_{L_1+\varepsilon_1} \subseteq \hat{D}_{L_2-\varepsilon_2}$. By Proposition 4, $\hat{\mathscr{D}}_n^{\pi}$ is a constructible cosheaf.

Theorem 2 *Assume there exists $\varepsilon < \delta_{\mathcal{U}}$ such that p is ε-concentrated on $\mathbb{X}$, then*

$$\lim_{n\to\infty} \mathbb{P}\left(d_I(\hat{\mathscr{D}}_n^{\pi}, \mathscr{R}_{\mathbb{X}}) \leq res_f \mathcal{U}\right) = 1.$$

Proof An inclusion of open sets $Y \subset Z$ induces a map

$$\pi_0(Y) \to \pi_0(Z),$$

of the corresponding sets of path-connected components of Y and Z respectively. Each set of path-connected components forms a basis for the homology group in degree 0. Therefore, the map from $\pi_0(Y)$ to $\pi_0(Z)$ extends to a map between homology groups, resulting in the following commutative diagram

$$\begin{array}{ccc} \pi_0(Y) & \longrightarrow & H_0(Y) \\ \downarrow & \circlearrowleft & \downarrow \\ \pi_0(Z) & \longrightarrow & H_0(Z). \end{array}$$

By combining the preceding commutative diagram with Proposition 8, we see that for every $V \subset W \in \mathcal{U}$, the following diagram commutes w.h.p.

$$\begin{array}{ccc} \pi_0(\mathbb{X}^V) & \xrightarrow{\cong} & \hat{\mathscr{D}}_n^{\pi}(V) \\ \downarrow & \circlearrowleft & \downarrow \\ \pi_0(\mathbb{X}^W) & \xrightarrow{\cong} & \hat{\mathscr{D}}_n^{\pi}(W). \end{array}$$

Notice that if $V \in \mathcal{U}'$, then $\mathcal{I}_{\mathcal{U}}(V) = V$. By Proposition 9 we have that

$$\hat{\mathscr{D}}_n^{\pi} \overset{\text{w.h.p.}}{\cong} \mathcal{M}_{\mathcal{U}}(\mathscr{R}_{\mathbb{X}}).$$

Therefore, w.h.p.

$$d_I(\hat{\mathscr{D}}_n^{\pi}, \mathcal{M}_{\mathcal{U}}(\mathscr{R}_{\mathbb{X}})) = 0.$$

Theorem 1, combined with the triangle inequality, implies the theorem. □

Corollary 3 *Let $\mathcal{R}(\mathbb{X}, f)$ be the Reeb graph of a constructible $\mathbb{R}$-space $(\mathbb{X}, f)$, and $\mathcal{D}(\hat{\mathscr{D}}_n^{\pi})$ be the display locale of the mapper cosheaf defined above. If there exists $\varepsilon < \delta_{\mathcal{U}}$ such that p is ε-concentrated on $\mathbb{X}$, then*

$$\lim_{n\to\infty} \mathbb{P}\left(d_R\big(\mathcal{D}(\hat{\mathscr{D}}_n^{\pi}), \mathcal{R}(\mathbb{X}, f)\big) \leq res_f \mathcal{U}\right) = 1.$$

If p is concentrated on $\mathbb{X}$, then the above corollary will hold for nice open covers with arbitrarily small resolution, as long as $\delta_{\mathcal{U}}$ remains positive. Therefore, Corollary 3 implies that we can use random point samples from p to construct mapper graphs that are (w.h.p.) arbitrarily close (in the Reeb distance) to the Reeb graph of $\mathbb{X}$.

To conclude, we will turn our attention to the stability of mapper cosheaves corresponding to a constructible space $(\mathbb{X}, f)$ under perturbations of the function f. The following theorem uses the machinery of cosheaf theory to prove that the mapper cosheaf is stable as long as the singular points of the constructible $\mathbb{R}$-space $\mathbb{X}$ are sufficiently "far away" from the set of boundary points of our open cover $\mathcal{U}$.

Theorem 3 *Suppose $\mathscr{F}$ and $\mathscr{G}$ are constructible cosheaves over $\mathbb{R}$, with a common set of critical values S. Let $\mathcal{U}$ be a nice open cover of $\mathbb{R}$, with set of boundary points B. Assume that*

$$d_I(\mathscr{F}, \mathscr{G}) < \min\{|s - b| : s \in S, b \in B\}.$$

Then

$$d_I(\mathcal{M}_{\mathcal{U}}(\mathscr{F}), \mathcal{M}_{\mathcal{U}}(\mathscr{G})) < d_I(\mathscr{F}, \mathscr{G}).$$

Moreover, if $\mathscr{F}$ is the Reeb cosheaf of $(\mathbb{X}, f)$ and $\mathscr{G}$ is the Reeb cosheaf of $(\mathbb{X}, g)$, then

$$d_I(\mathcal{M}_{\mathcal{U}}(\mathscr{F}), \mathcal{M}_{\mathcal{U}}(\mathscr{G})) < ||f - g||_{\infty}.$$

Proof Suppose $\varphi_U : \mathscr{F}(U) \to \mathscr{G}(U_\varepsilon)$ and $\psi_U : \mathscr{G}(U) \to \mathscr{F}(U_\varepsilon)$ give an ε-interleaving of $\mathscr{F}$ and $\mathscr{G}$. Recall that

$$\mathcal{M}_{\mathcal{U}}(\mathscr{F})(U) = \mathscr{F}(\mathcal{I}_{\mathcal{U}}(U)).$$

Then

$$\varphi_{\mathcal{I}_{\mathcal{U}}(U)} : \mathcal{M}_{\mathcal{U}}(\mathscr{F})(U) \to \mathscr{G}(\mathcal{I}_{\mathcal{U}}(U)_\varepsilon).$$

In general, this does not give us an ε-interleaving of $\mathcal{M}_{\mathcal{U}}(\mathscr{F})$ and $\mathcal{M}_{\mathcal{U}}(\mathscr{G})$, because $\mathcal{I}_{\mathcal{U}}(U)_\varepsilon \neq \mathcal{I}_{\mathcal{U}}(U_\varepsilon)$. However, we will proceed by showing that each of these sets contain the same set of critical values.

Following the definition of $\mathcal{I}_{\mathcal{U}}$, we see that for each $U \in \mathbf{Int}$, $\mathcal{I}_{\mathcal{U}}(U)$ is an open interval in $\mathbb{R}$, with boundary points contained in B. Therefore $\mathcal{I}_{\mathcal{U}}(U) \cap S \subset \mathcal{I}_{\mathcal{U}}(U)_\varepsilon \cap S$. If the inclusion is not an equality, then there must exist $s \in S$ such that $s \in \mathcal{I}_{\mathcal{U}}(U)_\varepsilon$ and $s \notin \mathcal{I}_{\mathcal{U}}(U)$. In other words, if $\mathcal{I}_{\mathcal{U}}(U) \cap S \subsetneq \mathcal{I}_{\mathcal{U}}(U)_\varepsilon \cap S$, then there exists $s \in S$ and $b \in B$ such that $|s - b| < \varepsilon$.

Define

$$N_{\mathcal{U},\varepsilon}(U) := \mathcal{I}_{\mathcal{U}}(U_\varepsilon) \cap \mathcal{I}_{\mathcal{U}}(U)_\varepsilon.$$

By the arguments above, if $\varepsilon < \min\{|s - b| : s \in S, b \in B\}$, then $\mathcal{I}_{\mathcal{U}}(U) \cap S = \mathcal{I}_{\mathcal{U}}(U)_\varepsilon \cap S$. It follows that $N_{\mathcal{U},\varepsilon}(U) \cap S = \mathcal{I}_{\mathcal{U}}(U) \cap S = \mathcal{I}_{\mathcal{U}}(U)_\varepsilon \cap S$, because $\mathcal{I}_{\mathcal{U}}(U) \subset \mathcal{I}_{\mathcal{U}}(U_\varepsilon)$. By the definition of constructibility, this implies that the natural extension map $\mathscr{G}[N_{\mathcal{U},\varepsilon}(U) \subset \mathcal{I}_{\mathcal{U}}(U)_\varepsilon]$ (denoted by e for notational brevity)

$$\mathscr{G}(N_{\mathcal{U},\varepsilon}(U)) \xrightarrow{e} \mathscr{G}(\mathcal{I}_{\mathcal{U}}(U)_\varepsilon)$$

is an isomorphism, and therefore is invertible. The composition

$$\mathcal{M}_{\mathcal{U}}(\mathscr{F})(U) \xrightarrow{\varphi} \mathscr{G}(\mathcal{I}_{\mathcal{U}}(U)_\varepsilon) \xrightarrow{e^{-1}} \mathscr{G}(N_{\mathcal{U},\varepsilon}(U)) \to \mathscr{G}(\mathcal{I}_{\mathcal{U}}(U_\varepsilon)) = \mathcal{M}_{\mathcal{U}}(\mathscr{G})(U_\varepsilon)$$

gives an ε-interleaving of $\mathcal{M}_{\mathcal{U}}(\mathscr{F})$ and $\mathcal{M}_{\mathcal{U}}(\mathscr{G})$, because each map in the composition is natural with respect to inclusions. Therefore

$$d_I(\mathcal{M}_{\mathcal{U}}(\mathscr{F}), \mathcal{M}_{\mathcal{U}}(\mathscr{G})) < d_I(\mathscr{F}, \mathscr{G}).$$

When $\mathscr{F}$ is the Reeb cosheaf of $(\mathbb{X}, f)$ and $\mathscr{G}$ is the Reeb cosheaf of $(\mathbb{X}, g)$, the second statement of the theorem is a direct consequence of the above inequality and de Silva et al. (2016, Theorem 4.4). □

5 Discussion

In this paper, we work with a categorification of the Reeb graph (de Silva et al. 2016) and introduce a categorified version of the mapper construction. This categorification provides the framework for using cosheaf theory and interleaving distances to study convergence and stability for mapper constructions applied to point cloud data. In this setting, the Reeb graph of a constructible $\mathbb{R}$-space is realized as the display locale of a constructible cosheaf (which we refer to as the Reeb cosheaf, following de Silva et al.

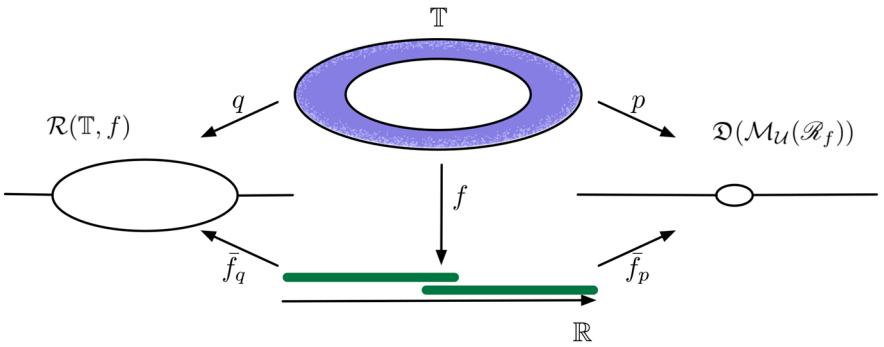


Fig. 5 An example of the enhanced mapper graph $\mathfrak{D}(\mathcal{M}_{\mathcal{U}}(\mathscr{R}_f))$ and the Reeb graph $\mathcal{R}(\mathbb{T}, f)$ of the height function f on the torus $\mathbb{T}$, with an open cover $\mathcal{U}$ of $f(\mathbb{T})$ consisting of two open intervals. The maps q and $\overline{f}_q$ are the natural quotient factorization of f obtained from the definition of the Reeb graph. Similarly, p and $\overline{f}_p$ are the quotient map and factorization of f obtained from the definition of the enhanced mapper graph

2016). In Sect. 2.5, we define a mapper functor from the category of cosheaves to the category of constructible cosheaves, giving a category theoretic interpretation of the mapper construction. We then define the *enhanced mapper graph* to be the display locale of the mapper functor applied to the Reeb cosheaf. We give an explicit geometric realization of the display locale as the quotient of a disjoint union of closed intervals, as illustrated in Fig. 5. In Sect. 3, we give a model for randomly sampling points from a probability density function concentrated on a constructible $\mathbb{R}$-space. After applying kernel density estimates, we consider an enhanced mapper graph generated by the random data. The main result of the paper, Theorem 2, then gives (with high probability) a bound on the Reeb distance between the Reeb graph and the enhanced mapper graph generated by a random sample of points.

Refinement to classic mapper graph The enhanced mapper graph suggests a few refinements to the classic mapper construction. Firstly, rather than an open cover $\mathcal{U}$ of $f(\mathbb{X})$ (the image of the constructible $\mathbb{R}$-space $\mathbb{X}$ in $\mathbb{R}$), it is more natural from the enhanced mapper perspective to start with a finite subset $\mathbb{R}_0$ of $\mathbb{R}$. From this finite subset, the enhanced mapper graph can be computed by first producing a finite disjoint union of closed intervals, with each interval associated to a connected component of the complement of $\mathbb{R}_0$. Then, by prescribing attaching maps on boundary points of the disjoint union of closed intervals, one can obtain a combinatorial description of the enhanced mapper graph as a graph with vertices labeled with real numbers. The enhanced mapper graph then has the structure of a stratified cover of $f(\mathbb{X})$, the image of the constructible $\mathbb{R}$-space $\mathbb{X}$ in $\mathbb{R}$. As such, the enhanced mapper graph contains more information than the classic mapper graph. Specifically, edges of the enhanced mapper graph have a naturally defined length which captures geometric information about the underlying constructible $\mathbb{R}$-space. Therefore, the enhanced mapper graph is naturally geometric, meaning that it comes equipped with a map to $\mathbb{R}$.

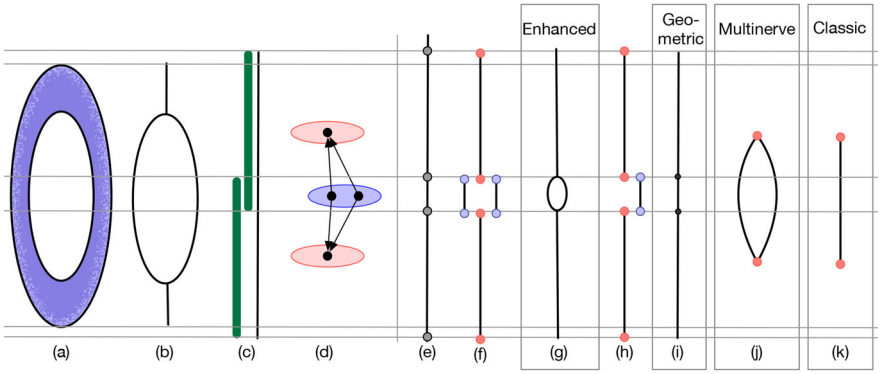


Fig. 6 Variations of mapper graphs for the height function on a torus. **a** Torus with a height function. **b** Reeb graph. **c** Nice cover. **d** Visualization of the mapper cosheaf. **e** Stratification of $\mathbb{R}$. **f** Disjoint union of closed intervals, $\widetilde{\mathfrak{D}}(\mathcal{M}_{\mathcal{U}}(\mathscr{R}_f))$, with quotient isomorphic to the enhanced mapper graph. **g** Enhanced mapper graph, $\mathfrak{D}(\mathcal{M}_{\mathcal{U}}(\mathscr{R}_f))$. **h** Disjoint union of closed intervals used to construct geometric mapper graph (Munch and Wang 2016). **i** Geometric mapper graph. **j** Multinerve mapper graph. **k** Classic mapper graph

Variations of mapper graphs We return to an in-depth discussion among variations of classic mapper graphs. As illustrated in Fig. 6 for the $\mathbb{R}$-**space** $(\mathbb{T}, f)$, that is, a torus equipped with a height function, the enhanced mapper graphs (g), geometric mapper graphs (i) studied by Munch and Wang (2016), and multinerve mapper graphs (j), have all been shown to be interleaved with Reeb graphs (b) (Munch and Wang 2016; Carriére and Oudot 2018). To further illustrate the subtle differences among the enhanced, geometric, mutinerve and classic mapper graphs, we give additional examples in Figs. 7 and 8. In certain scenarios, some of these constructions appear to be identical or very similar to each other. We would like to understand the information content associated with the above variants of mapper graphs, all of which are used as approximations of the Reeb graph of a constructible $\mathbb{R}$-space. As illustrated in Fig. 6, given an enhanced mapper graph (g) and an open cover (c), one can recover the the multinerve mapper graph (j), the geometric mapper graph (i), and the classic mapper graph (k). In future work, it would be interesting to quantify precisely the reconstruction ordering of these variants with and without any knowledge of the open cover.

In order to study convergence and stability of each variation of the mapper graph, it is necessary to assign function values to vertices of the graph. For the classic mapper graph or multinerve mapper graph, each vertex can be assigned, for instance, the value of the midpoint of a corresponding interval in $\mathbb{R}$. However, the display locale of a cosheaf over $\mathbb{R}$ admits a natural projection onto the real line, making a choice of function values unnecessary for the enhanced mapper graph. For this reason, we view the enhanced mapper graph as a natural variation of the mapper graph, well-suited for studying stability and convergence, with a natural interpretation in terms of cosheaf theory.

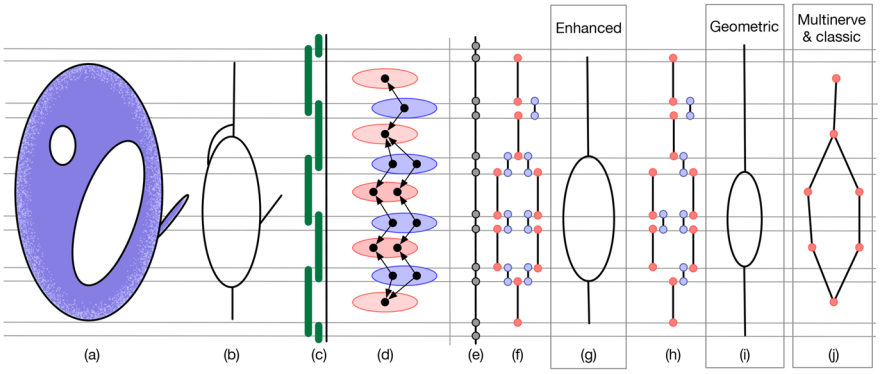


Fig. 7 A return to the example illustrated in Fig. 1. Variations of mapper graphs of a height function on a topological space. **a** A topological space with a height function. **b** Reeb graph. **c** Nice cover. **d** Visualization of the mapper cosheaf. **e** Stratification of $\mathbb{R}$. **f** Disjoint union of closed intervals with quotient isomorphic to the enhanced mapper graph. **g** Enhanced mapper graph. **h** Disjoint union of closed intervals used to construct geometric mapper graph (Munch and Wang 2016). **i** Geometric mapper graph. **j** Multinerve and classic mapper graph

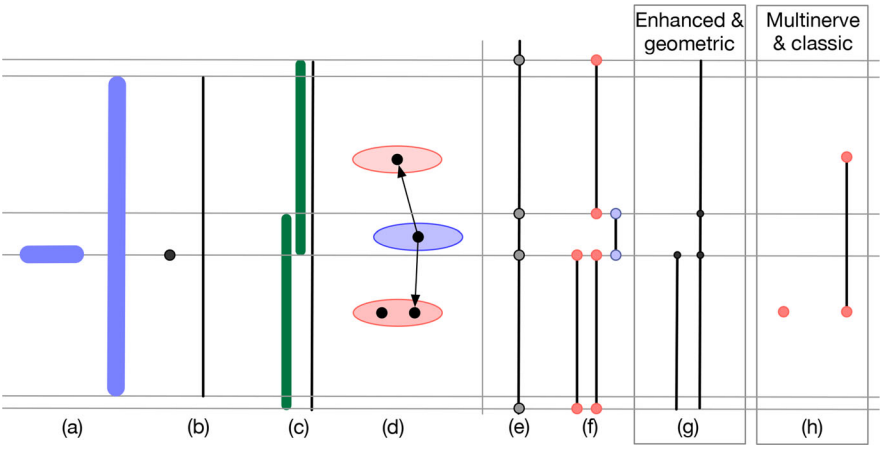


Fig. 8 Variations of mapper graphs of a height function on a topological space consisting of two line segments. **a** A topological space consisting of two line segments. **b** Reeb graph. **c** Nice cover. **d** Visualization of the mapper cosheaf. **e** Stratification of $\mathbb{R}$. **f** Disjoint union of closed intervals with quotient isomorphic to the enhanced mapper graph. **g** Enhanced and geometric mapper graph. **i** Multinerve and classic mapper graph

Multidimensional setting and parameter tuning It is natural to extend the enhanced mapper graph (and more generally the categorification of mapper graphs) to multidimensional Reeb spaces and multi-parameter mapper through studying constructible cosheaves and stratified covers of $\mathbb{R}^N$, for $N > 1$. We would also like to study the behavior of the parameter $\delta_{\mathcal{U}}$ for various constructible spaces and open covers. In general, this parameter can vanish for "bad" choices of open cover $\mathcal{U}$. It would be worthwhile to extend the results of this paper to obtain bounds on the interleaving distance when $\delta_{\mathcal{U}}$ vanishes. In conclusion, we hope for the results of this paper to pro-

mote the utility of combining methods from statistics and sheaf theory for the purpose of analyzing algorithms in computational topology.

Acknowledgements AB was supported in part by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Grant Agreement No. 754411 and NSF IIS-1513616. OB was supported in part by the Israel Science Foundation, Grant 1965/19. BW was supported in part by NSF IIS-1513616 and DBI-1661375. EM was supported in part by NSF CMMI-1800466, DMS-1800446, and CCF-1907591. We would like to thank the Institute for Mathematics and its Applications for hosting a workshop titled *Bridging Statistics and Sheaves* in May 2018, where this work was conceived.

Funding Open Access funding provided by Institute of Science and Technology (IST Austria)

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

A Pseudocode for the enhanced mapper graph algorithm

The following pseudocode (Algorithm 1) outlines an algorithm for computing the enhanced mapper graph, which is stored as a graph $G = (F, E)$ with a vertex set F and an edge set E, together with a real-valued function $f : F \to \mathbb{R}$.

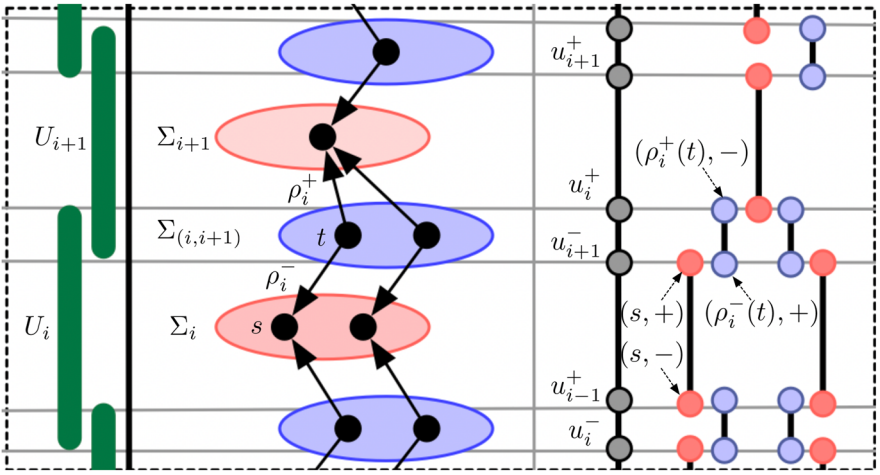


Fig. 9 An illustration of notations used in the pseudocode of Algorithm 1

Algorithm 1: Compute an enhanced mapper graph with oracles.

Input:

- A finite set of pairwise intersecting open intervals: $\{U_i := (u_i^-, u_i^+)\}_{i\in A}$
- For each interval, a set returned by a set oracle:
  - For each U_i, a set $\Sigma_i := \pi_0(f^{-1}(U_i))$
  - For each (U_i, U_{i+1}), a set $\Sigma_{(i,i+1)} := \pi_0(f^{-1}(U_i \cap U_{i+1}))$
- For each pair of intervals, a set map returned by a set-map oracle:
  For each (U_i, U_{i+1}), set maps

$$\rho_i^- : \Sigma_{(i,i+1)} \to \Sigma_i, \quad \rho_i^+ : \Sigma_{(i,i+1)} \to \Sigma_{i+1}.$$

Output:

- A graph $G = (F, E)$ with a vertex set F and an edge set $E \subseteq F \times F$
- A function $f : F \to \mathbb{R}$

Initialize $F = \emptyset$ and $E = \emptyset$
for $i \in A$ **do**
  Set $\Sigma_i^+ := \Sigma_i \times \{+\}$
  Set $\Sigma_i^- := \Sigma_i \times \{-\}$
  $F \leftarrow F \sqcup \Sigma_i^- \sqcup \Sigma_i^+$
  for $s \in \Sigma_i$ **do**
    $E \leftarrow E \sqcup ((s,-),(s,+))$

$$f((s,+)) := \begin{cases} u_{i+1}^- & \text{if } i+1 \in A \\ u_i^+ & \text{if } i+1 \notin A, \end{cases}$$

$$f((s,-)) := \begin{cases} u_{i-1}^+ & \text{if } i-1 \in A \\ u_i^- & \text{if } i-1 \notin A, \end{cases}$$

  end
end
for $(i, i+1) \in A \times A$ **do**
  for $t \in \Sigma_{(i,i+1)}$ **do**
    $E \leftarrow E \sqcup ((\rho_i^-(t),+),(\rho_i^+(t),-))$
  end
end

The algorithm assumes that we are given sets $\pi_0(f^{-1}(U))$ (denoted by Σ in the pseudocode) and set maps $\pi_0(f^{-1}(U)) \to \pi_0(f^{-1}(V))$ (denoted by ρ in the pseudocode) for various $U \subset V \subset \mathbb{R}$. In other words, the algorithm assumes that there is an oracle (referred to as a *set oracle*) that takes as input an inverse mapping of an interval and returns its corresponding set of path-connected components. It also assumes that there is a *set-map oracle* that keeps tracks of set maps between a pair of path-connected components (each component is denoted by s in the pseudocode). In Section 3, we give a statistical approach for computing such sets and set maps through kernel density estimates.

In Algorithm 1, let $\mathcal{U} = \{U_i\}_{i\in A}$ denote a finite set of pairwise intersecting open intervals. For simplicity, suppose the index set $A \subset \mathbb{Z}$ contains consecutive integers. That is, for each interval $U_i := (u_i^-, u_i^+)$ (for some $i \in A$), we have $u_i^- < u_{i-1}^+ <$

$u_{i+1}^- < u_i^+ < u_{i+1}^+$ (assuming $i-1, i+1 \in A$). For each interval U_i, $\Sigma_i := \pi_0(f^{-1}(U_i))$ denotes the set of path-connected components. For each path-connected component $s \in \Sigma_i$, the pairs $(s,+) \in \Sigma_i \times \{+,-\}$ and $(s,-) \in \Sigma_i \times \{+,-\}$ represent the two vertices associated to the edge in the enhanced mapper graph which corresponds to s. Similarly, for each path-connected component $t \in \Sigma_{(i,i+1)}$, the pairs $(\rho_i^-(t),+)$ and $(\rho_i^+(t),-)$ represent the two vertices associated to the edge in the enhanced mapper graph which corresponds to t.

For clarity, Fig. 9 illustrates notations used in the pseudocode of Algorithm 1. It is based on a zoomed view of Fig. 1c–f. The maps ρ_i^- and ρ_i^+ define how the red vertices and blue vertices (as end points of intervals) are glued together to form an enhanced mapper graph. In this particular example, $(\rho_i^-(t),+)$ (a blue vertex) matches with $(s,+)$ (a red vertex), due to the fact that $\rho_i^-(t) = s$.

References

Alagappan, M.: From 5 to 13: redefining the positions in basketball. In: MIT Sloan Sports Analytics Conference (2012)

Babu, A.: Zigzag Coarsenings, Mapper Stability and Gene-Network Analyses. Ph.D. Thesis, Stanford University (2013)

Barral, V., Biasotti, S.: 3D shape retrieval and classification using multiple kernel learning on extended reeb graphs. Vis. Comput. **30**(11), 1247–1259 (2014)

Bauer, U., Ge, X., Wang, Y.: Measuring distance between reeb graphs. In: Proceedings of the 30th Annual Symposium on Computational Geometry, pp. 464–473 (2014)

Beketayev, K., Yeliussizov, D., Morozov, D., Weber, G., Hamann, B.: Measuring the distance between merge trees. In: Topological Methods in Data Analysis and Visualization III: Theory, Algorithms, and Applications, Mathematics and Visualization, pp. 151–166 (2014)

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

Bobrowski, O.: Homological connectivity in random Čech complexes (2019). arXiv:1906.04861

Bobrowski, O., Kahle, M., Skraba, P.: Maximally persistent cycles in random geometric complexes. Ann. Appl. Probab. **27**(4), 2032–2060 (2017a)

Bobrowski, O., Mukherjee, S., Taylor, J.E.: Topological consistency via kernel estimation. Bernoulli **23**(1), 288–328 (2017b)

Carlsson, G., Zomorodian, A.J., Collins, A., Guibas, L.J.: Persistence barcodes for shapes. In: Proceedings Eurographs/ACM SIGGRAPH Symposium on Geometry Processing, pp. 124–135 (2004)

Carr, H., Duke, D.: Joint contour nets: computation and properties. In: IEEE Pacific Visualization Symposium, pp. 161–168 (2013)

Carr, H., Duke, D.: Joint contour nets. IEEE Trans. Vis. Comput Graph. **20**(8), 1100–1113 (2014)

Carr, H., Snoeyink, J., Axen, U.: Computing contour trees in all dimensions. Comput. Geom. **24**(2), 75–94 (2003)

Carriére, M., Oudot, S.: Structure and stability of the one-dimensional mapper. Found. Comput. Math. **18**(6), 1333–1396 (2018)

Carriére, M., Michel, B., Oudot, S.: Statistical analysis and parameter selection for mapper. J. Mach. Learn. Res. **19**, 1–39 (2018)

Chazal, F., Sun, J.: Gromov–Hausdorff approximation of filament structure using reeb-type graph. In: Proceedings 13th Annual Symposium on Computational Geometry, pp. 491–500 (2014)

Chazal, F., Guibas, L.J., Oudot, S., Skraba, P.: Scalar field analysis over point cloud data. Discrete Comput. Geom. **46**, 743–775 (2011)

Chazal, F., Glisse, M., Labruére, C., Michel, B.: Convergence rates for persistence diagram estimation in topological data analysis. J. Mach. Learn. Res. **16**, 3603–3635 (2015)

Chazal, F., Fasy, B., Lecci, F., Michel, B., Rinaldo, A., Rinaldo, A., Wasserman, L.: Robust topological inference: distance to a measure and kernel distance. J. Mach. Learn. Res. **18**(1), 5845–5884 (2017)

Cohen-Steiner, D., Edelsbrunner, H., Harer, J.: Extending persistence using poincaré and lefschetz duality. Found. Comput. Math. **9**(1), 79–103 (2009)
de Kergorlay, H.-L., Tillmann, U., Vipond, O.: Random Čech complexes on manifolds with boundary (2019). arXiv:1906.07626
de Silva, V., Munch, E., Patel, A.: Categorified reeb graphs. Discrete Comput. Geom. **55**, 854–906 (2016)
de Verdiére, Éric Colin, Ginot, G., Goaoc, X.: Multinerves and helly numbers of acyclic families. In: Proceedings of the 28th Annual Symposium on Computational Geometry, pp. 209–218 (2012)
Dey, T.K., Mémoli, F., Wang, Y.: Mutiscale mapper: a framework for topological summarization of data and maps. In: Proceedings of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 997–1013 (2016)
Dey, T.K., Mémoli, F., Wang, Y.: Topological analysis of nerves, reeb spaces, mappers, and multiscale mappers. In: Aronov, B. Katz, M.J. (eds) 33rd International Symposium on Computational Geometry, volume 77 of Leibniz International Proceedings in Informatics (LIPIcs), pp. 36:1–36:16. Dagstuhl, Germany (2017). Schloss Dagstuhl–Leibniz–Zentrum fuer Informatik
Edelsbrunner, H., Letscher, D., Zomorodian, A.J.: Topological persistence and simplification. Discrete Comput. Geom. **28**, 511–533 (2002)
Edelsbrunner, H., Harer, J., Natarajan, V., Pascucci, V.: Morse–Smale complexes for piece-wise linear 3-manifolds. In: Proceedings of the 19th Annual Symposium on Computational Geometry, pp. 361–370 (2003a)
Edelsbrunner, H., Harer, J., Zomorodian, A.J.: Hierarchical Morse–Smale complexes for piecewise linear 2-manifolds. Discrete Comput. Geom. **30**, 87–107 (2003b)
Edelsbrunner, H., Harer, J., Patel, A.K.: Reeb spaces of piecewise linear mappings. In: Proceedings of the 24th Annual Symposium on Computational Geometry, pp. 242–250 (2008)
Fasy, B.T., Lecci, F., Rinaldo, A., Wasserman, L., Balakrishnan, S., Singh, A.: Confidence sets for persistence diagrams. Ann. Stat. **42**(6), 2301–2339 (2014)
Funk, J.: The display locale of a cosheaf. Cahiers Topologie Géom. Différentielle Catég. **36**(1), 53–93 (1995)
Gasparovic, E., Gommel, M., Purvine, E., Sazdanovic, R., Wang, B., Wang, Y., Ziegelmeier, L.: A complete characterization of the one-dimensional intrinsic Čech persistence diagrams for metric graphs. In: Research in Computational Topology (2018)
Gerber, S., Potter, K.: Data analysis with the Morse–Smale complex: the MSR package for R. J. Stat. Softw. **50**(2) (2012)
Ghrist, R.: Barcodes: the persistent topology of data. Bull. Am. Math. Soc. **45**, 61–75 (2008)
Hiraoka, Y., Shirai, T., Trinh, K.D.: Limit theorems for persistence diagrams. Ann. Appl. Probab. **28**(5), 2740–2780 (2018)
Kahle, M., Meckes, E.: Limit the theorems for Betti numbers of random simplicial complexes. Homology, Homotopy Appl. **15**(1), 343–374 (2013)
Munch, E., Wang, B.: Convergence between categorical representations of Reeb space and mapper. In: Fekete, S., Lubiw, A. (eds) 32nd International Symposium on Computational Geometry, volume 51 of Leibniz International Proceedings in Informatics (LIPIcs), pp. 53:1–53:16, Dagstuhl, Germany (2016). Schloss Dagstuhl–Leibniz–Zentrum fuer Informatik
Nicolau, M., Levine, A.J., Carlsson, G.: Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. Proc. Natl. Acad. Sci. **108**(17), 7265–7270 (2011)
Niyogi, P., Smale, S., Weinberger, S.: Finding the homology of submanifolds with high confidence from random samples. Discrete Comput. Geom. **39**(1–3), 419–441 (2008)
Niyogi, P., Smale, S., Weinberger, S.: A topological view of unsupervised learning from noisy data. SIAM J. Comput. **40**(3), 646–663 (2011)
Owada, T., Adler, R.J.: Limit theorems for point processes under geometric constraints (and topological crackle). Ann. Probab. **45**(3), 2004–2055 (2017)
Reeb, G.: Sur les points singuliers d'une forme de pfaff completement intergrable ou d'une fonction numerique (on the singular points of a complete integral pfaff form or of a numerical function). Comptes Rendus Acad. Sci. Paris **222**, 847–849 (1946)
Singh, G., Mémoli, F., Carlsson, G.: Topological methods for the analysis of high dimensional data sets and 3D object recognition. In: Eurographics Symposium on Point-Based Graphics, pp. 91–100 (2007)
Wang, Y., Wang, B.: Topological Inference of Manifolds with Boundary (2018). arXiv:1810.05759
Woolf, J.: The fundamental category of a stratified space. J. Homotopy Relat. Struct. **4**(1), 359–387 (2009)

Yogeshwaran, D., Subag, E., Adler, R.J.: Random geometric complexes in the thermodynamic regime. Probab. Theory Relat. Fields **167**, 107–142 (2016)

Affiliations

Adam Brown[1] · Omer Bobrowski[2] · Elizabeth Munch[3] · Bei Wang[4]

✉ Adam Brown
adam.brown@ist.ac.at

Omer Bobrowski
omer@ee.technion.ac.il

Elizabeth Munch
muncheli@msu.edu

Bei Wang
beiwang@sci.utah.edu

1 IST Austria, Klosterneuburg, Austria

2 Technion - Israel Institute of Technology, Haifa, Israel

3 Michigan State University, Michigan, USA

4 University of Utah, Salt Lake City, USA