Zigzag zoology: Rips zigzags for homology inference

Steve Y. Oudot, Donald R. Sheehy

Zigzag Zoology: Rips Zigzags for Homology Inference

Steve Y. Oudot[*], Donald R. Sheehy[†]

July 23, 2014

Abstract

For n points sampled near a compact set X, the persistence barcode of the Rips filtration built from the sample contains information about the homology of X as long as X satisfies some geometric assumptions. The Rips filtration is prohibitively large, however zigzag persistence can be used to keep the size linear in n, with a constant factor depending only (exponentially) on the intrinsic dimension of X. We present several species of Rips-like zigzags and compare them with respect to the signal-to-noise ratio, a measure of how well the underlying homology is represented in the persistence barcode relative to the noise in the barcode at the relevant scales. Some of these Rips-like zigzags have been available as part of the Dionysus library for several years while others are new. Interestingly, we show that some species of Rips zigzags will exhibit less noise than the (non-zigzag) Rips filtration itself. Thus, Rips zigzags can offer improvements in both size complexity and signal-to-noise ratio.

Along the way, we develop new techniques for manipulating and comparing persistence barcodes from zigzag modules. In particular, we give methods for reversing arrows and removing spaces from a zigzag while controlling the changes occurring in its barcode. These techniques were developed to provide our theoretical analysis of the signal-to-noise ratio of Rips-like zigzags, but they are of independent interest as they apply to zigzag modules generally.

Keywords Topological inference · Persistent homology · Vietoris-Rips complex · Quiver representations · Reflections

Mathematics Subject Classification 62-07 · 55U10 · 16G20 · 68U05

Communicated by Herbert Edelsbrunner

1 Introduction

The goal of homology inference is to extract the homology of a space from a finite sample. The problem is ill-posed in general, but under the right geometric assumptions about the input and the underlying space, one can compute an object called a *persistence barcode* which provably contains information about the underlying homology. Indeed, homology inference was and continues to be one of the main motivations for topological persistence theory.

The barcode is computed from a sequence of simplicial complexes, for which two main challenges arise. The first challenge is to guarantee that the simplicial complexes remain small. Commonly

[*]Geometrica group, Inria Saclay – Île de France, Alan Turing building, École Polytechnique, 91190 Palaiseau, France. Email: steve.oudot@inria.fr

[†]Computer Science Department, 371 Fairfield Way, Unit 4155, University of Connecticut, Storrs, CT 06269-4155, USA. Email: don.r.sheehy@gmail.com

used methods produce complexes that quickly become too large to fit in memory. The second challenge is to decrease noise in the barcode while preserving the signal, i.e. the information about the underlying space. We confront both challenges, analyze several approaches that give linear-size data structures, and provide guarantees on the signal-to-noise ratio in the barcodes.

Context. Persistent homology applies to nested, parameterized families of simplicial complexes called *filtrations*. The persistence algorithm takes a filtration and produces a barcode describing all the changes in homology as one goes from one complex to the next in the filtration [17, 29].

Persistent homology has an important connection with geometric inference results that describe conditions when homology inference is possible using a union of balls centered at the sample points—see the survey by Chazal and Cohen-Steiner [7]. The *(Vietoris-)Rips filtration* $\{\mathcal{R}_\alpha\}_{\alpha\geq 0}$ is useful when these conditions are met. It is defined to have a simplex in $\mathcal{R}_\alpha$ for every subset of points with diameter at most α. So, the filtration parameter is the geometric scale and the theory guarantees the existence of some range of scales for which the barcode encodes the homology of the underlying space. The barcode of this filtration thus has an elegant multi-scale interpretation as "the homology of the input point cloud across scales."

The immediate drawback to using the Rips filtration is its size. The scale at which it exceeds the available memory varies with the input data, the filtration, and the computer used. However, it is observed to happen early enough so that not all the interesting homological information hidden in the data can be discovered—see Section 6 for a compelling example. Recent research looks at how to reduce the size of the complexes in the filtration to postpone the breaking point. Among the most notable examples are the witness complex [11] and graph induced complex [14], which were introduced for this purpose. They are built on a sparse subset of the input data (called the landmarks set), and their construction is driven by the rest of the data points. The resulting filtrations have fewer vertices, but they do not degrade the quality of the inference guarantees as much as would a Rips filtration on the landmarks. Nevertheless, a minimum landmarks density is still required to ensure a correct inference, therefore the filtrations still incur a blowup at large scales.

A different approach was proposed by Chazal and Oudot [10], who used truncated Rips filtrations on a nested sequence of subsets of the input points corresponding to samplings at different scales. Their method computes the barcodes of the Rips filtration of each subset restricted to a range of scales near the sampling scale of the subset. This can prevent the size blowup in the Rips filtrations because every subset looks like a uniform sample at the relevant scale. The lingering challenge from this work is to relate the bars in the resulting barcodes for different scales.

Taking advantage of the recent introduction of zigzag persistence by Carlsson and de Silva [5], Morozov suggested a simple way to connect the truncated Rips filtrations of consecutive subsamples together, to obtain a single long sequence of simplicial complexes connected by inclusions—called the *Morozov zigzag* (M-ZZ) hereafter. Zigzag persistence relaxes the condition that the family of complexes be a filtration and instead allows consecutive spaces to be included in either direction, forwards or backwards, so the sequence is a zigzag diagram rather than a filtration. The M-ZZ has been integrated into the Dionysus library [16] since early 2009, and as reported by its author from preliminary experiments [26], it has given surprisingly good results in practice. However, to date it comes with no theoretical guarantees, so a primary motivation of our paper is to assess the theoretical quality of the results provided by this zigzag.

Existing methods for building sparse approximations to the Rips filtration have all focused on the size question, but have ignored the question of noise. For example, even the Rips filtration can have noise in the barcode at the scales where it represents the underlying topology. We show that

some variants of the M-ZZ not only recover the topological signal but also provably eliminate noise in the relevant range.

Contributions. We provide the following theoretical guarantees for the Morozov zigzag:

- When the input point cloud P is sufficiently close (in the Hausdorff distance) to a compact set X with positive weak feature size in $\mathbb{R}^d$, there is a *sweet range* of geometric scales over which the persistence barcode of the Morozov zigzag exhibits the homology of X (technically, the offsets X^λ for an arbitrarily small $\lambda > 0$). That is, the barcode has long intervals spanning the entire sweet range, and their number is at least the dimension of the homology group $\mathsf{H}X^\lambda$ (Theorem 4.3).
- If X has positive μ-reach, then there is a smaller (*sweeter*) range over which the number of spanning intervals is exactly $\dim \mathsf{H}X^\lambda$ and no other intervals are present.

This motivates the study of more elaborate variants of the Morozov zigzag that are less likely to carry topological noise in the sweet range, even when the underlying space X has zero μ-reach and positive weak feature size. We analyze three variants in the paper:

- The first one, called the *discretized Mozorov zigzag*, consists in considering only subsamples whose corresponding geometric scales are of the form ζ^i for a fixed constant ζ and an integer i. This discretization makes sure that the geometric scale drops significantly (by a factor of ζ) from one subsample to the next, so there is enough room in each connection between truncated filtrations to kill the noise.
- The second one, called the *oscillating Rips zigzag*, consists in somewhat relaxing the truncation parameter in the Rips filtrations before connecting them together. The effect is to leave enough room in every truncated filtration for the noise to be killed.
- The third one, called the *image Rips zigzag*, consists in taking a nested pair of Morozov zigzags with different filtration parameters, and in connecting them by canonical inclusions to obtain an image zigzag module at the homology level. Taking a pair of zigzags instead of a single zigzag kills the noise in the same way as taking a pair of Rips complexes instead of a single Rips complex did in [10].

Each of these variants comes with the desired guarantee that the sweet and sweeter ranges are equal, meaning that there is guaranteed to be only ephemeral noise in the sweet range even when the underlying space X merely has positive weak feature size. Thus, Rips zigzags offer improvements in both size complexity and signal-to-noise ratio compared to the Rips filtration. The price to pay compared to the basic Morozov zigzag is a somewhat increased time or space complexity (Theorems 5.1 and 5.2). The overhead depends on the variant considered but it always remains bounded, so the variants are tractable alternatives to the Morozov zigzag.

To prove the aforementioned results, we develop new techniques for manipulating zigzag modules and comparing their persistence barcodes:

- We show how arrows in a zigzag module can be reversed while preserving the persistence barcode (Theorem 3.1).
- We give a method for removing spaces from a zigzag module while tracking the intervals in its barcode (Theorem 3.2).

These low-level manipulations make it possible to transform one module into another while controlling the changes in its barcode, a strategy at the core of the proofs of our main theorems.

Related work. A different approach to the problem of building sparse filtrations for offsets of point clouds in Euclidean space was presented by Hudson et al. [22]. They used ideas from Delaunay refinement mesh generation to build linear-size filtrations that provide provably good

approximations to the persistence diagram of the offsets. However, that approach requires building a complex that covers the ambient space and includes simplices up to its dimension. Moreover, the construction requires the use of high degree predicates. In contrast, the new methods described here only depend on an intrinsic dimension of data and can be built using only distances comparisons.

Recently, Sheehy [27] proposed a method for building a sparse zigzag filtration whose barcode is provably close to that of the Rips filtration as well as a non-zigzagging variant achieving similar guarantees. Also, Dey et al. gave an alternative persistence algorithm for simplicial maps rather than inclusions, which is closely related to zigzag persistence [15]. Their approach, when applied to Rips filtrations, similarly gives barcodes that are provably close to that of the Rips filtration. We obtain comparable space/time bounds to these results but get stronger guarantees regarding noise. Methods that approximate the Rips filtration directly can, in principle, have noise that is as large as the noise in the Rips filtration itself (or worse).

2 Background

We assume the reader is familiar with homology theory—see e.g. [20] for an introduction. Throughout the paper, we use singular homology with coefficients in a fixed field, omitted in our notations. The homology functor is denoted by H.

The definitions of Rips and Čech filtrations are given in Section 2.1. Then, Section 2.2 introduces some aspects of the sampling theory for compact sets in Euclidean spaces, which will be instrumental in the geometric part of our analysis. We refer the reader to [7] for a comprehensive survey on this topic. Section 2.3 gives a brief overview of the concepts and results from zigzag persistence theory that will be used in the algebraic part of our analysis. Our terminology is the same as in [5] up to a few minor variants, and we refer the reader to that paper for an in-depth treatment. See also [4] for a more general and high-level introduction to the topological analysis of point cloud data.

2.1 Rips and Čech Filtrations

Let P be a finite set in $\mathbb{R}^d$. The Rips complex of P at scale α, denoted $\mathcal{R}_\alpha(P)$, is the abstract simplicial complex consisting of all subsets of P with diameter (maximum pairwise distance) at most α. The *Rips filtration* is the collection of Rips complexes at all nonnegative scales.

The Čech complex of P at scale α, denoted $\mathcal{C}_\alpha(P)$ is the abstract simplicial complex consisting of all subsets of P with minimum enclosing ball radius at most α. It is isomorphic to the *nerve* of the collection of balls of radius α centered at the points of P, i.e. there is a simplex for every subset of balls that have a common intersection. The *Čech filtration* is the collection of Čech complexes at all nonnegative scales.

The Rips and Čech filtrations are *interleaved* as follows[1] in $\mathbb{R}^d$, where $\vartheta_d = \sqrt{\frac{d}{2(d+1)}} \in [\frac{1}{2}, \frac{1}{\sqrt{2}})$.

$$\forall \alpha \geq 0,\ \mathcal{C}_{\frac{\alpha}{2}}(P) \subseteq \mathcal{R}_\alpha(P) \subseteq \mathcal{C}_{\vartheta_d \alpha}(P). \tag{1}$$

2.2 Critical Point Theory for Distance Functions

The geometric part of our analysis takes place in Euclidean space $\mathbb{R}^d$, where $\|\cdot\|$ denotes the Euclidean norm. The distance from a point y to a set $X \subset \mathbb{R}^d$ is $\mathrm{d}(y, X) = \inf_{x \in X} \|x - y\|$. When

[1] See e.g. [13] for a proof. Our definition of the Rips complex differs from the one in [13] by a factor of 2 in the parameter value. This explains the slight discrepancy between our chain of inclusions and the one in [13].

X is compact, the infimum becomes a minimum, and we let d_X denote the function *distance to* X.

$$\forall y \in \mathbb{R}^d,\ \mathrm{d}_X(y) \stackrel{\text{def}}{=} \mathrm{d}(y, X) = \min_{x \in X} \|x - y\|.$$

The α-*offset* of X is the locus of the points of $\mathbb{R}^d$ whose distance to X is at most α:

$$X^\alpha \stackrel{\text{def}}{=} \mathrm{d}_X^{-1}([0, \alpha]).$$

Although d_X may not be differentiable everywhere in $\mathbb{R}^d$, its gradient can be extended to be well-defined over all of $\mathbb{R}^d$ [8]. The extended gradient is denoted ∇_X in the following. When working with offsets, it is useful to observe that for $y \notin X^\alpha$, $\mathrm{d}_{X^\alpha}(y) = \mathrm{d}_X(y) - \alpha$, so $\nabla_{X^\alpha}(y) = \nabla_X(y)$.

Definition 2.1. *A* critical point *of the distance function* d_X *to a compact set* $X \subset \mathbb{R}^d$ *is a point* p *of* $\mathbb{R}^d$ *such that* $\nabla_X(p) = 0$. *Equivalently, a* critical point *is a point of* $\mathbb{R}^d \setminus X$ *that is in the convex hull of its nearest points in* X. *A number* $r \in \mathbb{R}$ *is a* critical value *if there exists a critical point* p *such that* $\mathrm{d}_X(p) = r$.

Definition 2.2. *The* weak feature size *of a compact set* X, *denoted* $\mathrm{wfs}(X)$, *is the smallest positive critical value of its distance function* d_X.

Given $X \subseteq \mathbb{R}^d$ and $\beta \geq \alpha \geq 0$, we let $\mathsf{H}X^\beta_\alpha$ denote the image of the homomorphism $\mathsf{H}X^\alpha \to \mathsf{H}X^\beta$ induced at the homology level by the canonical inclusion $X^\alpha \hookrightarrow X^\beta$.

Lemma 2.3 ([10])**.** *Let* X *be a compact set and* P *a finite set in* $\mathbb{R}^d$, *such that* $\mathrm{d}_H(X, P) < \varepsilon$ *for some* $\varepsilon < \frac{1}{4}\mathrm{wfs}(X)$. *Then,* $\mathsf{H}P^\beta_\alpha \cong \mathsf{H}X^\lambda$ *for any* $\alpha, \beta \in [\varepsilon, \mathrm{wfs}(X) - \varepsilon]$ *such that* $\beta - \alpha \geq 2\varepsilon$, *and for any* $\lambda \in (0, \mathrm{wfs}(X))$.

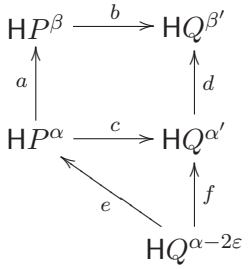
For any finite sets $P \subseteq Q \subset \mathbb{R}^d$ and any non-negative parameters $\alpha, \alpha', \beta, \beta'$ such that $\alpha \leq \alpha'$, $\beta \leq \beta'$, $\alpha \leq \beta$ and $\alpha' \leq \beta'$, we have the following commutative diagram where all linear maps are induced by inclusions of offsets:

$$\begin{array}{ccc} \mathsf{H}P^\beta & \to & \mathsf{H}Q^{\beta'} \\ \uparrow & & \uparrow \\ \mathsf{H}P^\alpha & \to & \mathsf{H}Q^{\alpha'} \end{array} \tag{2}$$

This commutative diagram induces a homomorphism $\mathsf{H}P^\beta_\alpha \to \mathsf{H}Q^{\beta'}_{\alpha'}$.

Lemma 2.4. *Let* X *be a compact set and* $P \subseteq Q$ *be finite sets in* $\mathbb{R}^d$, *such that* $\mathrm{d}_H(X, P) < \varepsilon$ *and* $\mathrm{d}_H(Q, X) < \varepsilon$ *for some* $\varepsilon < \frac{1}{6}\mathrm{wfs}(X)$. *Then, for any* $\alpha, \alpha', \beta, \beta' \in [3\varepsilon, \mathrm{wfs}(X) - \varepsilon]$ *such that* $\beta - \alpha \geq 2\varepsilon$, $\beta' - \alpha' \geq 2\varepsilon$, $\alpha' \geq \alpha$ *and* $\beta' \geq \beta$, *the linear map* $\mathsf{H}P^\beta_\alpha \to \mathsf{H}Q^{\beta'}_{\alpha'}$ *induced by the diagram (2) is an isomorphism.*

Proof. According to Lemma 2.3, $\mathsf{H}P^\beta_\alpha$ and $\mathsf{H}Q^{\beta'}_{\alpha'}$ are isomorphic vector spaces, and they are finite-dimensional because P and Q are finite. Therefore, it suffices to show that $\operatorname{rank} \mathsf{H}P^\beta_\alpha \to \mathsf{H}Q^{\beta'}_{\alpha'} = \dim \mathsf{H}P^\beta_\alpha$. We have the following commutative diagram where all the maps are induced by inclusions (note that $Q^{\alpha-2\varepsilon} \subseteq P^\alpha$ since $\mathrm{d}_H(P, Q) \leq \mathrm{d}_H(P, X) + \mathrm{d}_H(Q, X) \leq 2\varepsilon$):



The homomorphism $\mathsf{H}P_\alpha^\beta \to \mathsf{H}Q_{\alpha'}^{\beta'}$ we are interested in is the restriction of b to $\operatorname{im} a$, whose rank is the same as the one of $b \circ a$. By composition and commutativity, we have

$$\operatorname{rank} d \circ f = \operatorname{rank} b \circ a \circ e \leq \operatorname{rank} b \circ a \leq \operatorname{rank} a,$$

and by Lemma 2.3 we have $\operatorname{rank} d \circ f = \operatorname{rank} a = \dim \mathsf{H}P_\alpha^\beta$ since $\alpha - 2\varepsilon \geq \varepsilon$ and $\beta' \leq \operatorname{wfs}(X) - \varepsilon$. Hence, $\operatorname{rank} b \circ a = \dim \mathsf{H}P_\alpha^\beta$. □

Combining the above analysis with the Persistent Nerve Lemma[2] [10, Lemma 3.4], we obtain the following result where the notation $\mathsf{H}\mathcal{C}_\alpha^\beta(P)$ stands for the image of the homomorphism $\mathsf{H}\mathcal{C}_\alpha(P) \to \mathsf{H}\mathcal{C}_\beta(P)$ induced at the homology level by the inclusion $\mathcal{C}_\alpha(P) \hookrightarrow \mathcal{C}_\beta(P)$.

Theorem 2.5. *Let X be a compact set and P and Q be finite sets in $\mathbb{R}^d$ with $P \subseteq Q$, such that $\mathrm{d}_H(P, X) < \varepsilon$ and $\mathrm{d}_H(Q, X) < \varepsilon$.*

(i) *If $\varepsilon < \frac{1}{4}\operatorname{wfs}(X)$, then for any $\alpha, \beta \in [\varepsilon, \operatorname{wfs}(X) - \varepsilon]$ such that $\beta - \alpha \geq 2\varepsilon$, for any $\lambda \in (0, \operatorname{wfs}(X))$, the spaces $\mathsf{H}\mathcal{C}_\alpha^\beta(P)$ and $\mathsf{H}X^\lambda$ are isomorphic.*

(ii) *If $\varepsilon < \frac{1}{6}\operatorname{wfs}(X)$, then for any $\alpha, \alpha', \beta, \beta' \in [3\varepsilon, \operatorname{wfs}(X) - \varepsilon]$ such that $\beta - \alpha \geq 2\varepsilon$, $\beta' - \alpha' \geq 2\varepsilon$, $\alpha' \geq \alpha$ and $\beta' \geq \beta$, the homomorphism $\mathsf{H}\mathcal{C}_\alpha^\beta(P) \to \mathsf{H}\mathcal{C}_{\alpha'}^{\beta'}(Q)$ induced by the following commutative diagram (where the maps are induced by inclusions) is an isomorphism.*

$$\begin{array}{ccc} \mathsf{H}\mathcal{C}_\beta(P) & \to & \mathsf{H}\mathcal{C}_{\beta'}(Q) \\ \uparrow & & \uparrow \\ \mathsf{H}\mathcal{C}_\alpha(P) & \to & \mathsf{H}\mathcal{C}_{\alpha'}(Q) \end{array}$$

(iii) *If $\varepsilon < \frac{1}{4}\operatorname{wfs}(X)$, then for any $\alpha, \beta \in [2\vartheta_d\varepsilon, \frac{1}{2\vartheta_d}(\operatorname{wfs}(X) - \varepsilon)]$ such that $\beta - \alpha \geq 2\varepsilon$, the map $\mathsf{H}\mathcal{C}_\alpha^\beta(P) \to \mathsf{H}\mathcal{R}_{2\beta}(P)$ is injective, and the map $\mathsf{H}\mathcal{R}_{\frac{\alpha}{\vartheta_d}}(P) \to \mathsf{H}\mathcal{C}_\alpha^\beta(P)$ is surjective, where both maps are induced by inclusions. The constant ϑ_d is the same as in (1).*

Proof. Parts (i) and (ii) follow from Lemmas 2.3 and 2.4 respectively, when combined with the persistent Nerve Lemma. To prove part (iii), first observe that the hypotheses and (i) imply that $\operatorname{rank} \mathsf{H}\mathcal{C}_\alpha^\beta = \operatorname{rank} \mathsf{H}\mathcal{C}_\alpha^{2\vartheta_d\beta}$. So, in the following sequence of maps induced by inclusions,

$$\mathsf{H}\mathcal{C}_\alpha(P) \to \mathsf{H}\mathcal{C}_\beta(P) \to \mathsf{H}\mathcal{R}_{2\beta}(P) \to \mathsf{H}\mathcal{C}_{2\vartheta_d\beta}(P)$$

the map $\mathsf{H}\mathcal{C}_\beta(P) \to \mathsf{H}\mathcal{R}_{2\beta}(P)$ restricted to the image of $\mathsf{H}\mathcal{C}_\alpha(P) \to \mathsf{H}\mathcal{C}_\beta(P)$ is injective and therefore $\mathsf{H}\mathcal{C}_\alpha^\beta(P) \to \mathsf{H}\mathcal{R}_{2\beta}(P)$ is injective as well. Similarly, we have that $\operatorname{rank} \mathsf{H}\mathcal{C}_{\frac{\alpha}{2\vartheta_d}}^\beta = \operatorname{rank} \mathsf{H}\mathcal{C}_\alpha^\beta$. If we decompose these maps as

$$\mathsf{H}\mathcal{C}_{\frac{\alpha}{2\vartheta_d}}(P) \to \mathsf{H}\mathcal{R}_{\frac{\alpha}{\vartheta_d}}(P) \to \mathsf{H}\mathcal{C}_\alpha(P) \to \mathsf{H}\mathcal{C}_\beta(P)$$

we see that $\mathsf{H}\mathcal{R}_{\frac{\alpha}{\vartheta_d}}(P) \to \mathsf{H}\mathcal{C}_\alpha^\beta(P)$ is surjective as desired. □

[2]We are in fact using an extended version of the Persistent Nerve Lemma, stated in [9], where the index sets of the open covers may differ.

2.3 Zigzag persistence

The algebraic part of our analysis relies on zigzag persistence theory. We use the terminology introduced by Carlsson and de Silva [5]. A *zigzag module* $\mathbb{V}$ is a finite diagram of finite-dimensional vector spaces over a fixed field $\boldsymbol{k}$ where the underlying graph is a path:

$$\mathbb{V} = V_1 \overset{v_1}{\longleftrightarrow} V_2 \overset{v_2}{\longleftrightarrow} \cdots \overset{v_{n-1}}{\longleftrightarrow} V_n,$$

where the notation $V_i \overset{v_i}{\longleftrightarrow} V_{i+1}$ indicates that the linear map v_i can be oriented either forwards ($v_i : V_i \to V_{i+1}$) or backwards ($v_i : V_i \leftarrow V_{i+1}$). An equivalent notation is $v_i : V_i \leftrightarrow V_{i+1}$. The sequence of map orientations defines the *type* of the module $\mathbb{V}$. A *persistence module*, as defined in the standard (non-zigzag) persistence literature [29], is a zigzag module in which all the maps are oriented forwards. Thus, all persistence modules of the same length have the same type.

A *submodule* $\mathbb{X}$ of a zigzag module $\mathbb{V}$ is defined by subspaces $X_i \subseteq V_i$ such that for all i we have $v_i(X_i) \subseteq X_{i+1}$ if $v_i : V_i \leftrightarrow V_{i+1}$ is a forward map and $v_i(X_{i+1}) \subseteq X_i$ if v_i is a backward map. The maps in $\mathbb{X}$ are the restrictions of the maps in $\mathbb{V}$ to the X_is, so $\mathbb{V}$ and $\mathbb{X}$ have the same type.

A *homomorphism* Φ between two zigzag modules $\mathbb{V}$ and $\mathbb{W}$ of the same type, denoted $\Phi : \mathbb{V} \to \mathbb{W}$, is a collection of linear maps $\phi_i : V_i \to W_i$ such that the following diagram commutes for all $i = 1, \cdots, n-1$:

$$\begin{array}{ccc} V_i & \overset{v_i}{\longleftrightarrow} & V_{i+1} \\ \downarrow{\scriptstyle \phi_i} & & \downarrow{\scriptstyle \phi_{i+1}} \\ W_i & \overset{w_i}{\longleftrightarrow} & W_{i+1} \end{array} \tag{3}$$

The image of Φ, denoted $\operatorname{im}\Phi$, is the submodule of $\mathbb{W}$ composed of the spaces $\operatorname{im}\phi_i$. The kernel of Φ, denoted $\ker\Phi$, is the submodule of $\mathbb{V}$ composed of the spaces $\ker\phi_i$. Φ is called an *isomorphism* if every map $\phi_i : V_i \to W_i$ is an isomorphism.

A submodule $\mathbb{X}$ of $\mathbb{V}$ is a *summand* if there exists another submodule $\mathbb{Y}$ of $\mathbb{V}$ where $V_i = X_i \oplus Y_i$ for all i. In that case, we say that $\mathbb{V}$ is the *direct sum* of $\mathbb{X}$ and $\mathbb{Y}$, written $\mathbb{V} = \mathbb{X} \oplus \mathbb{Y}$. As pointed out in [5], all summands are submodules, but not all submodules are summands. A zigzag module $\mathbb{V}$ is called *indecomposable* if it admits no other summand than the zero module and itself. It has been known since Gabriel [18] that the indecomposable zigzag modules are the so-called *interval modules*. Given a module type τ and an integer interval $[b, d]$, the interval τ-module with birth time b and death time d is written $\mathbb{I}_\tau[b, d]$ and defined with spaces I_i such that $I_i = \boldsymbol{k}$ if $i \in [b, d]$ and $I_i = 0$ otherwise, and with identity maps between adjacent copies of the base field $\boldsymbol{k}$ and zero maps elsewhere (the maps are oriented according to τ). A module $\mathbb{I}_\tau[b, d]$ is said to be an *ephemeral module* when the interval $[b, d]$ itself is *ephemeral*, i.e. when it has length zero ($b = d$).

A consequence of Gabriel's result is that every τ-module is isomorphic to a direct sum of finitely many τ-intervals. Moreover, the Krull-Schmidt principle guarantees that this decomposition is unique up to a reordering of the terms.

Theorem 2.6 (Interval Decomposition)**.** *For every τ-module $\mathbb{V}$ there exists a unique finite multiset of interval modules $\{\mathbb{I}_\tau[b_i, d_i]\}$ and an isomorphism*

$$\Phi : \mathbb{V} \to \bigoplus_i \mathbb{I}_\tau[b_i, d_i].$$

Thus, as in standard (non-zigzag) persistence theory, the structure of $\mathbb{V}$ is fully and uniquely described by the multiset of integer intervals[3] $\{[b_i, d_i]\}$, called the *persistence barcode* of $\mathbb{V}$ and denoted $\mathsf{Pers}(\mathbb{V})$.

[3]Note that we follow [5] and depart from the traditional persistence barcode representation by using closed intervals instead of half-open intervals.

Carlsson and de Silva gave a constructive proof of Theorem 2.6—see [5, Thm. 4.1], which led to an algorithm for computing the decomposition of a zigzag module. Among the concepts and results presented in their paper, the following one plays an important part here.

Given a zigzag module $\mathbb{V} = V_1 \overset{v_1}{\longleftrightarrow} \cdots \overset{v_{n-1}}{\longleftrightarrow} V_n$ and two integers $p \leq q \in [1, n]$, let $\mathbb{V}[p, q]$ denote the restriction of $\mathbb{V}$ to the index set $[p, q]$. That is, $\mathbb{V}[p, q] = V_p \overset{v_p}{\longleftrightarrow} \cdots \overset{v_{q-1}}{\longleftrightarrow} V_q$. Let $\mathsf{Pers}(\mathbb{V})|_{[p,q]}$ denote the restriction of $\mathsf{Pers}(\mathbb{V})$ to $[p, q]$:

$$\mathsf{Pers}(\mathbb{V})|_{[p,q]} = \{[b, d] \cap [p, q] \mid [b, d] \in \mathsf{Pers}(\mathbb{V})\}.$$

The Restriction Principle [5, Prop. 2.12] connects the two types of restrictions together:

Theorem 2.7 (Restriction)**.**

$$\mathsf{Pers}(\mathbb{V}[p, q]) = \mathsf{Pers}(\mathbb{V})|_{[p,q]}.$$

3 Manipulating Zigzag Modules

3.1 Arrow reversal

Suppose we have a zigzag module $\mathbb{V} = V_1 \leftrightarrow \cdots \leftrightarrow V_n$ and we want to reverse the map $V_k \leftrightarrow V_{k+1}$ for some arbitrary index k in the range $[1, n-1]$, while preserving the persistence barcode of $\mathbb{V}$. The following theorem states that this is always possible, moreover with a reverse map that is closely related to the original map.

Theorem 3.1 (Arrow Reversal)**.**
Let $\mathbb{V} = V_1 \leftrightarrow \cdots \leftrightarrow V_k \overset{f}{\longleftrightarrow} V_{k+1} \leftrightarrow \cdots \leftrightarrow V_n$ be a zigzag module. Then, there is a map $g : V_k \leftrightarrow V_{k+1}$ oriented opposite to f, such that $f \circ g|_{\operatorname{im} f} = \mathbb{1}_{\operatorname{im} f}$ and $g \circ f|_{\operatorname{im} g} = \mathbb{1}_{\operatorname{im} g}$, and the zigzag module $\mathbb{V}^*$ obtained from $\mathbb{V}$ by replacing the submodule $V_k \overset{f}{\longleftrightarrow} V_{k+1}$ by $V_k \overset{g}{\longleftrightarrow} V_{k+1}$ has the same persistence barcode as $\mathbb{V}$.

When f is injective, g is surjective and $g \circ f$ is the identity over the domain of f. Conversely, when f is surjective, g is injective and $f \circ g$ is the identity over the codomain of f. These properties are useful when $\mathbb{V}$ is part of a commutative diagram, as they sometimes imply commutativity is preserved after the arrow reversal—as in the proof of Theorem 4.2.

Proof. Let $\Phi : \mathbb{V} \to \bigoplus_i \mathbb{I}_\tau[b_i, d_i]$ be the decomposition of $\mathbb{V}$ given by the Interval Decomposition Theorem 2.6. Denote the spaces of $\mathbb{I}_\tau[b_i, d_i]$ by $I_1^i \ldots, I_n^i$, and let $\Phi = (\phi_1, \ldots, \phi_n)$ where each $\phi_j : V_j \to \bigoplus_i I_j^i$ is an isomorphism.

We assume without loss of generality that f is oriented forwards; the case when f is oriented backwards is symmetric. The map $f' = \bigoplus_i (I_k^i \to I_{k+1}^i)$ makes the following diagram commute.

$$\begin{array}{ccc} V_k & \xrightarrow{\quad f \quad} & V_{k+1} \\ \downarrow{\scriptstyle \phi_k} & & \downarrow{\scriptstyle \phi_{k+1}} \\ \bigoplus_i I_k^i & \xrightarrow{\quad f' \quad} & \bigoplus_i I_{k+1}^i \end{array} \tag{4}$$

To reverse f, we first reverse each map $I_k^i \to I_{k+1}^i$ separately. Recall that this map is either the identity or zero. The reversal of an identity map is an identity map and the reversal of a zero map is a zero map, so we take $I_k^i \overset{\mathbb{1}}{\leftarrow} I_{k+1}^i$ if $I_k^i \overset{\mathbb{1}}{\to} I_{k+1}^i$ and $I_k^i \overset{0}{\leftarrow} I_{k+1}^i$ if $I_k^i \overset{0}{\to} I_{k+1}^i$. We thus get a new

interval module $\mathbb{I}_{\tau^*}[b_i, d_i]$, where τ^* is the zigzag type of $\mathbb{V}^*$. Let now $g' = \bigoplus_i (I^i_k \leftarrow I^i_{k+1})$ be the reverse of f', and let $g = \phi_k^{-1} \circ g' \circ \phi_{k+1}$ be the reverse of f, which gives the following commutative diagram.

$$\begin{array}{ccc} V_k & \xleftarrow{\quad g \quad} & V_{k+1} \\ \downarrow{\scriptstyle \phi_k} & & \downarrow{\scriptstyle \phi_{k+1}} \\ \bigoplus_i I^i_k & \xleftarrow{\quad g' \quad} & \bigoplus_i I^i_{k+1} \end{array} \tag{5}$$

The commutativity of this diagram and the definition of Φ imply that the the same isomorphisms ϕ_j induce an isomorphism $\Phi^* : \mathbb{V}^* \to \bigoplus_i \mathbb{I}_{\tau^*}[b_i, d_i]$. Therefore,

$$\mathsf{Pers}(\mathbb{V}^*) = \mathsf{Pers}(\bigoplus_i \mathbb{I}_{\tau^*}[b_i, d_i]) = \mathsf{Pers}(\bigoplus_i \mathbb{I}_{\tau}[b_i, d_i]) = \mathsf{Pers}(\mathbb{V}).$$

It only remains to prove that $f \circ g|_{\operatorname{im} f} = \mathbb{1}_{\operatorname{im} f}$ and $g \circ f|_{\operatorname{im} g} = \mathbb{1}_{\operatorname{im} g}$. By symmetry, it suffices to prove just one of these, say $f \circ g|_{\operatorname{im} f} = \mathbb{1}_{\operatorname{im} f}$. Using the isomorphism Φ, it suffices to prove the equivalent statement for f' and g'. The statement now follows from the observation that

$$f' \circ g' = \bigoplus_i ((I^i_k \to I^i_{k+1}) \circ (I^i_k \leftarrow I^i_{k+1})) = \mathbb{1}_{\operatorname{im} f'} \oplus 0_{\ker g'},$$

where the final step separates the identity maps and the zero maps into two groups. □

3.2 Space Removal

Suppose we want to remove the space V_k from a zigzag module $\mathbb{V} = V_1 \leftrightarrow \cdots \leftrightarrow V_n$ while preserving most of the persistence barcode. The following theorem shows that this is always possible, and that every interval $[b, d]$ in $\mathsf{Pers}(\mathbb{V})$ becomes $[b, d] \setminus \{k\}$ in the new index set $\{1, \cdots, k-1, k+1, \cdots, n\}$. Note that this is still a single interval, denoted by $[b, d]_{\hat{k}}$ for clarity.

Theorem 3.2 (Space Removal)**.**
Let $\mathbb{V}$ be a zigzag module containing $V_{k-1} \xleftrightarrow{f} V_k \xleftrightarrow{g} V_{k+1}$. There exists a map $h : V_{k-1} \leftrightarrow V_{k+1}$ such that the zigzag module $\mathbb{V}^*$ formed by removing V_k, f, and g from $\mathbb{V}$ and replacing them with h has barcode

$$\mathsf{Pers}(\mathbb{V}^*) = \left\{ [b, d]_{\hat{k}} \mid [b, d] \in \mathsf{Pers}(\mathbb{V}) \right\}.$$

Furthermore, if any map $V_{k-1} \leftrightarrow V_{k+1}$ commutes with f and g, then so does h.

The first step towards the proof of this theorem is the following special case for removing a space between maps oriented in the same direction.

Lemma 3.3 (Composition)**.**
Given $\mathbb{V} = V_1 \leftrightarrow \cdots \leftrightarrow V_{k-1} \xrightarrow{f} V_k \xrightarrow{g} V_{k+1} \leftrightarrow \cdots \leftrightarrow V_n$, let $\mathbb{V}^* = V_1 \leftrightarrow \cdots \leftrightarrow V_{k-1} \xrightarrow{g \circ f} V_{k+1} \leftrightarrow \cdots \leftrightarrow V_n$. Then, $\mathsf{Pers}(\mathbb{V}^*) = \{[b, d]_{\hat{k}} \mid [b, d] \in \mathsf{Pers}(\mathbb{V})\}$.

Proof. Let $\Phi : \mathbb{V} \to \bigoplus_i \mathbb{I}_\tau[b_i, d_i]$ be the decomposition of $\mathbb{V}$ given by the Interval Decomposition Theorem 2.6. Denoting the spaces of $\mathbb{I}_\tau[b_i, d_i]$ by $I^i_1 \ldots, I^i_n$, and letting $f' = \bigoplus_i (I^i_{k-1} \to I^i_k)$,

$g' = \bigoplus_i (I^i_k \to I^i_{k+1})$, and $\Phi = (\phi_1, \ldots, \phi_n)$ where each $\phi_j : V_j \to \bigoplus_i I^i_j$ is an isomorphism, we have the following commutative diagram.

$$\begin{array}{ccccc} V_{k-1} & \xrightarrow{f} & V_k & \xrightarrow{g} & V_{k+1} \\ \downarrow{\scriptstyle \phi_{k-1}} & & \downarrow{\scriptstyle \phi_k} & & \downarrow{\scriptstyle \phi_{k+1}} \\ \bigoplus_i I^i_{k-1} & \xrightarrow{f'} & \bigoplus_i I^i_k & \xrightarrow{g'} & \bigoplus_i I^i_{k+1} \end{array} \tag{6}$$

By removing the kth space from each interval module $\mathbb{I}_\tau[b_i, d_i]$, we get a new interval decomposition $\bigoplus_i \mathbb{I}_{\tau^*}[b_i, d_i]_{\hat{k}}$, where τ^* is the type of $\mathbb{V}^*$. Observe that the commutativity of (6) implies that the following diagram commutes as well.

$$\begin{array}{ccc} V_{k-1} & \xrightarrow{g \circ f} & V_{k+1} \\ \downarrow{\scriptstyle \phi_{k-1}} & & \downarrow{\scriptstyle \phi_{k+1}} \\ \bigoplus_i I^i_{k-1} & \xrightarrow{g' \circ f'} & \bigoplus_i I^i_{k+1} \end{array} \tag{7}$$

Thus, $\Phi^* = (\phi_1, \ldots, \phi_{k-1}, \phi_{k+1}, \ldots, \phi_n)$ is an isomorphism from $\mathbb{V}^*$ to $\bigoplus_i \mathbb{I}_{\tau^*}[b_i, d_i]_{\hat{k}}$. It follows that

$$\mathsf{Pers}(\mathbb{V}^*) = \mathsf{Pers}(\bigoplus_i \mathbb{I}_{\tau^*}[b_i, d_i]_{\hat{k}}) = \{[b, d]_{\hat{k}} \mid [b, d] \in \mathsf{Pers}(\mathbb{V})\}$$

as claimed in the lemma. □

We now combine this special case with the Arrow Reversal Theorem 3.1 to prove the general Space Removal Theorem.

Proof of Theorem 3.2. There are four cases to consider depending on the orientations of the maps f and g. Cases $(\xrightarrow{f}, \xrightarrow{g})$ and $(\xleftarrow{f}, \xleftarrow{g})$ are covered by the Composition Lemma 3.3. The remaining cases $(\xrightarrow{f}, \xleftarrow{g})$ and $(\xleftarrow{f}, \xrightarrow{g})$ are handled by reversing one of the maps using the Arrow Reversal Theorem 3.1 and then applying the Composition Lemma 3.3. The arrow reversal preserves the persistence diagram and the composition removes the kth index from each interval as desired. It only remains to show that the commutativity property holds.

When there exists a map $h' : V_{k-1} \leftrightarrow V_{k+1}$ that commutes with f and g, then either

$$f = g \circ h' \text{ or } f = h' \circ g \text{ or } g = f \circ h' \text{ or } g = h' \circ f.$$

The proofs of these four cases are similar, so we only give the proof of the first case here, leaving the others as an exercise.

Assuming that $f = g \circ h'$, we have $\operatorname{im} f \subseteq \operatorname{im} g$. So, applying the Arrow Reversal Theorem 3.1 on the map g results in a map g' such that $g \circ g'|_{\operatorname{im} f} = \mathbb{1}_{\operatorname{im} f}$. Then, the Composition Lemma 3.3 lets $h = g' \circ f$, so we obtain

$$g \circ h = g \circ g' \circ f = g \circ g'|_{\operatorname{im} f} \circ f = \mathbb{1}_{\operatorname{im} f} \circ f = f$$

as claimed. □

4 Rips Zigzags

Let P be a finite point cloud in some metric space, and suppose that the matrix of pairwise distances between the points of P is known. Given any ordering $(p_1, \cdots, p_n)$ on the points of P, let $P_i := \{p_1, \ldots, p_i\}$ denote the ith prefix, and define the *ith geometric scale* as $\varepsilon_i = \mathrm{d}_H(P_i, P)$. Since P_i grows as i increases, we have $\varepsilon_1 \geq \varepsilon_2 \geq \cdots \geq \varepsilon_n = 0$.

Given a choice of multipliers $\eta \leq \rho$, Chazal and Oudot [10] proposed to do homological inference from P using the sequence of short filtrations $\mathcal{R}_{\eta\varepsilon_i}(P_i) \hookrightarrow \mathcal{R}_{\rho\varepsilon_i}(P_i)$. Zigzag persistence makes it possible to replace this sequence of short filtrations by a single zigzag filtration, a representative portion of which is depicted below.

$$\begin{array}{ccccccc} & \mathcal{R}_{\rho\varepsilon_{i-1}}(P_i) & & \mathcal{R}_{\rho\varepsilon_i}(P_{i+1}) & & \\ \nearrow & & \nwarrow & & \nearrow & & \nwarrow \\ \mathcal{R}_{\eta\varepsilon_{i-1}}(P_{i-1}) & & \mathcal{R}_{\eta\varepsilon_i}(P_i) & & \mathcal{R}_{\eta\varepsilon_{i+1}}(P_{i+1}) \end{array} \tag{8}$$

The zigzag module induced at the homology level by this diagram is referred to as the *oscillating Rips zigzag* (*oR-ZZ* for short) hereafter. Note that from a computational point of view, the smaller ρ the smaller the maximum complex size in the zigzag. In addition, the closer η to ρ the fewer simplex additions and deletions during the zigzag calculation. Therefore, as a rule of thumb, one should try to make ρ as small as possible while η as close to ρ as possible.

Before proceeding with the analysis of the oscillating Rips zigzag in Section 4.2 and of its variants in the subsequent sections, we first make a short detour and study another zigzag over the sequence of vertex sets $P_1, \cdots, P_n$ that will play a central role in our analysis.

4.1 The image Čech zigzag

Canonical inclusions between Čech complexes give the following pair of horizontal zigzags connected by vertical arrows, where each zigzag alternately adds one point to the vertex set and reduces the geometric scale.

$$\begin{array}{ccccccccccc} \cdots & \leftarrow & \mathsf{H}\mathcal{C}_{\rho\varepsilon_i}(P_i) & \rightarrow & \mathsf{H}\mathcal{C}_{\rho\varepsilon_i}(P_{i+1}) & \leftarrow & \mathsf{H}\mathcal{C}_{\rho\varepsilon_{i+1}}(P_{i+1}) & \rightarrow & \cdots \\ & & \uparrow & & \uparrow & & \uparrow & & \\ \cdots & \leftarrow & \mathsf{H}\mathcal{C}_{\eta\varepsilon_i}(P_i) & \rightarrow & \mathsf{H}\mathcal{C}_{\eta\varepsilon_i}(P_{i+1}) & \leftarrow & \mathsf{H}\mathcal{C}_{\eta\varepsilon_{i+1}}(P_{i+1}) & \rightarrow & \cdots \end{array}$$

This commutative diagram induces the following zigzag of images, referred to as the *image Čech zigzag* hereafter.

$$\cdots \leftarrow \mathsf{H}\mathcal{C}^{\rho\varepsilon_i}_{\eta\varepsilon_i}(P_i) \rightarrow \mathsf{H}\mathcal{C}^{\rho\varepsilon_i}_{\eta\varepsilon_i}(P_{i+1}) \leftarrow \mathsf{H}\mathcal{C}^{\rho\varepsilon_{i+1}}_{\eta\varepsilon_{i+1}}(P_{i+1}) \rightarrow \cdots \tag{9}$$

Theorem 4.1. *Let ρ and η be multipliers such that $3 < \eta < \rho - 2$. Let $X \subset \mathbb{R}^d$ be a compact set, let $P \subset \mathbb{R}^d$ be a finite set, and let $\varepsilon = \mathrm{d}_H(P, X)$. Then, for any $l > k$ such that*

$$\max\left\{\frac{3\varepsilon}{\eta - 3},\ \frac{2\varepsilon}{\rho - \eta - 2}\right\} \leq \varepsilon_l \leq \varepsilon_k < \min\left\{\frac{1}{6}\mathrm{wfs}(X) - \varepsilon,\ \frac{1}{\rho + 1}\left(\mathrm{wfs}(X) - \varepsilon\right)\right\},$$

the image Čech zigzag restricted to $\mathsf{H}\mathcal{C}^{\rho\varepsilon_k}_{\eta\varepsilon_k}(P_k) \rightarrow \cdots \leftarrow \mathsf{H}\mathcal{C}^{\rho\varepsilon_l}_{\eta\varepsilon_l}(P_l)$ contains only isomorphisms, and its spaces are isomorphic to $\mathsf{H}X^\lambda$ for any $\lambda \in (0, \mathrm{wfs}(X))$. Therefore, its persistence barcode is made only of full-length intervals, whose number equals the dimension of $\mathsf{H}X^\lambda$.

Proof. By the triangle inequality, for any $i \in [1,n]$ we have $\mathrm{d}_H(P_i, X) \leq \mathrm{d}_H(P_i, P) + \mathrm{d}_H(P, X) < \varepsilon_i + \varepsilon$. Since the geometric scale ε_i decreases with i, we have $\varepsilon_k \geq \varepsilon_i \geq \varepsilon_l$ for all $i \in [k,l]$, and therefore the bounds on ε_l and ε_k imply that $\varepsilon_i + \varepsilon < \frac{1}{6}\mathrm{wfs}(X)$, that $\eta\varepsilon_i$ and $\rho\varepsilon_i$ belong to the interval $[3(\varepsilon_i+\varepsilon), \mathrm{wfs}(X) - (\varepsilon_i+\varepsilon)]$, and that $\rho\varepsilon_i - \eta\varepsilon_i \geq 2(\varepsilon_i+\varepsilon)$. Thus, the hypotheses of Theorem 2.5 (ii) are satisfied within the range $[k,l]$, and so the result follows from that theorem. □

Theorem 4.1 guarantees a range of scales for which the zigzag has the correct persistent homology (*the sweet range*) only when the Hausdorff distance ε between the sample P and the set X is sufficiently small with respect to the weak feature size of X. Since ε_l must be less than ε_k, the bounds in the hypothesis of the theorem imply that

$$\varepsilon < \min\left\{\frac{\eta-3}{6\eta},\ \frac{\eta-3}{3\rho+\eta},\ \frac{\rho-\eta-2}{6(\rho-\eta)},\ \frac{\rho-\eta-2}{3\rho-\eta}\right\}\mathrm{wfs}(X) = O(\mathrm{wfs}(X)).$$

4.2 Analysis of the oscillating Rips zigzag

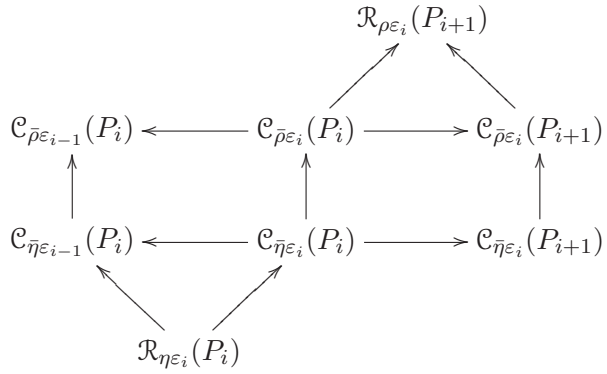
The following result gives conditions on η and ρ for the persistence barcode of the oR-ZZ to exhibit the homology of the shape underlying an input point cloud $P \subset \mathbb{R}^d$. The main idea will be to relate the oR-ZZ to the image Čech zigzag.

Theorem 4.2. *Let ρ and η be multipliers such that $\rho > 10$ and $\frac{3}{\vartheta_d} < \eta < \frac{\rho-4}{2\vartheta_d}$. Let $X \subset \mathbb{R}^d$ be a compact set, let $P \subset \mathbb{R}^d$ be a finite set, and let $\varepsilon = \mathrm{d}_H(P,X)$. Then, for any $l > k$ such that*

$$\max\left\{\frac{3\varepsilon}{\vartheta_d\eta - 3},\ \frac{4\varepsilon}{\rho - 2\vartheta_d\eta - 4}\right\} \leq \varepsilon_l \leq \varepsilon_k < \min\left\{\frac{1}{6}\mathrm{wfs}(X) - \varepsilon,\ \frac{1}{\vartheta_d\rho + 1}\left(\mathrm{wfs}(X) - \varepsilon\right)\right\},$$

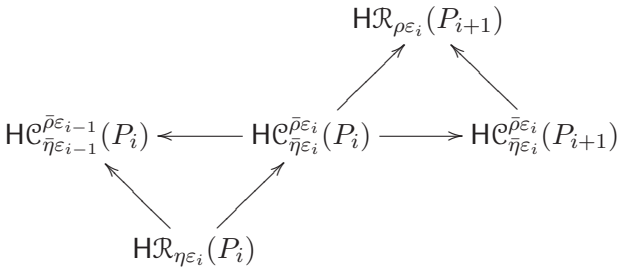
the oR-ZZ restricted to $\mathsf{H}\mathcal{R}_{\rho\varepsilon_k}(P_{k+1}) \leftarrow \cdots \leftarrow \mathsf{H}\mathcal{R}_{\eta\varepsilon_l}(P_l)$ has a persistence barcode made only of full-length intervals and ephemeral (length zero) intervals, the number of full-length intervals being equal to the dimension of $\mathsf{H}X^\lambda$ for any $\lambda \in (0, \mathrm{wfs}(X))$.

Proof. Let $\bar{\rho} = \frac{\rho}{2}$ and $\bar{\eta} = \vartheta_d\eta$. Our hypotheses imply $\frac{\rho}{2} \geq \vartheta_d\eta$, so using (1) we can factor the inclusion maps in (8) through Čech complexes with multipliers $\bar{\eta}$ and $\bar{\rho}$ as follows.



This commutative diagram induces the following interleaving between the oscillating Rips zigzag

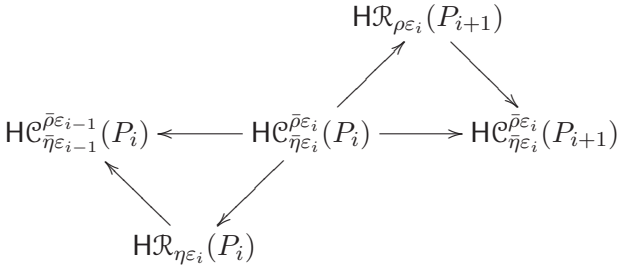
and the image Čech zigzag.



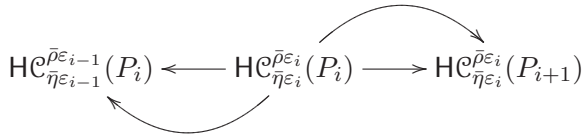
Let $\mathbb{U}$ be the restriction of the oR-ZZ to $\mathsf{H}\mathcal{R}_{\rho\varepsilon_k}(P_{k+1}) \leftarrow \cdots \leftarrow \mathsf{H}\mathcal{R}_{\eta\varepsilon_l}(P_l)$, let $\mathbb{W}$ be the restriction of the image Čech zigzag to $\mathsf{H}\mathcal{C}^{\bar{\rho}\varepsilon_k}_{\bar{\eta}\varepsilon_k}(P_k) \rightarrow \cdots \leftarrow \mathsf{H}\mathcal{C}^{\bar{\rho}\varepsilon_l}_{\bar{\eta}\varepsilon_l}(P_l)$, and let $\mathbb{V}$ be the mixed zigzag $\mathsf{H}\mathcal{C}^{\bar{\rho}\varepsilon_k}_{\bar{\eta}\varepsilon_k}(P_k) \rightarrow \mathsf{H}\mathcal{R}_{\rho\varepsilon_k}(P_{k+1}) \leftarrow \mathsf{H}\mathcal{C}^{\bar{\rho}\varepsilon_k}_{\bar{\eta}\varepsilon_k}(P_{k+1}) \leftarrow \mathsf{H}\mathcal{R}_{\eta\varepsilon_{k+1}}(P_{k+1}) \rightarrow \mathsf{H}\mathcal{C}^{\bar{\rho}\varepsilon_{k+1}}_{\bar{\eta}\varepsilon_{k+1}}(P_{k+1}) \rightarrow \cdots \rightarrow \mathsf{H}\mathcal{C}^{\bar{\rho}\varepsilon_{l-1}}_{\bar{\eta}\varepsilon_{l-1}}(P_{l-1}) \rightarrow \mathsf{H}\mathcal{R}_{\rho\varepsilon_{l-1}}(P_l) \leftarrow \mathsf{H}\mathcal{C}^{\bar{\rho}\varepsilon_{l-1}}_{\bar{\eta}\varepsilon_{l-1}}(P_l) \leftarrow \mathsf{H}\mathcal{R}_{\eta\varepsilon_l}(P_l) \rightarrow \mathsf{H}\mathcal{C}^{\bar{\rho}\varepsilon_l}_{\bar{\eta}\varepsilon_l}(P_l)$. Our goal is to relate $\mathsf{Pers}(\mathbb{V})$ to both $\mathsf{Pers}(\mathbb{U})$ and $\mathsf{Pers}(\mathbb{W})$, which we will do by turning $\mathbb{V}$ successively into $\mathbb{U}$ and $\mathbb{W}$ via arrow reversals and space removals while tracking the changes in its persistence barcode. It is an elementary exercise to show that the conditions bounding ε_l and ε_k satisfy the hypotheses of Theorem 4.1 with multipliers $\bar{\eta}$ and $\bar{\rho}$. These conditions also satisfy the hypotheses of Theorem 2.5(iii) for $\alpha = \bar{\eta}\varepsilon_i$ and $\beta = \bar{\rho}\varepsilon_i$ and the sets P_i and P_{i+1} for all $i \in [k, l-1]$. Thus, these Theorems imply the following facts:

(i) all spaces in $\mathbb{W}$ are isomorphic to $\mathsf{H}X^\lambda$, and all maps in $\mathbb{W}$ are isomorphisms,
(ii) the map $\mathsf{H}\mathcal{R}_{\rho\varepsilon_i}(P_{i+1}) \leftarrow \mathsf{H}\mathcal{C}^{\bar{\rho}\varepsilon_i}_{\bar{\eta}\varepsilon_i}(P_{i+1})$ is injective for any $i \in [k, l-1]$, and
(iii) the map $\mathsf{H}\mathcal{R}_{\eta\varepsilon_i}(P_i) \rightarrow \mathsf{H}\mathcal{C}^{\bar{\rho}\varepsilon_i}_{\bar{\eta}\varepsilon_i}(P_i)$ is surjective for any $i \in [k, l]$.

To turn $\mathbb{V}$ into $\mathbb{W}$, we first use Theorem 3.1 to reverse every injective map $\mathsf{H}\mathcal{R}_{\rho\varepsilon_i}(P_{i+1}) \leftarrow \mathsf{H}\mathcal{C}^{\bar{\rho}\varepsilon_i}_{\bar{\eta}\varepsilon_i}(P_{i+1})$ and every surjective map $\mathsf{H}\mathcal{R}_{\eta\varepsilon_i}(P_i) \rightarrow \mathsf{H}\mathcal{C}^{\bar{\rho}\varepsilon_i}_{\bar{\eta}\varepsilon_i}(P_i)$, to get a new zigzag $\mathbb{V}^* = \mathsf{H}\mathcal{C}^{\bar{\rho}\varepsilon_k}_{\bar{\eta}\varepsilon_k}(P_k) \rightarrow \mathsf{H}\mathcal{R}_{\rho\varepsilon_k}(P_{k+1}) \rightarrow \mathsf{H}\mathcal{C}^{\bar{\rho}\varepsilon_k}_{\bar{\eta}\varepsilon_k}(P_{k+1}) \leftarrow \mathsf{H}\mathcal{R}_{\eta\varepsilon_{k+1}}(P_{k+1}) \leftarrow \mathsf{H}\mathcal{C}^{\bar{\rho}\varepsilon_{k+1}}_{\bar{\eta}\varepsilon_{k+1}}(P_{k+1}) \rightarrow \cdots \leftarrow \mathsf{H}\mathcal{C}^{\bar{\rho}\varepsilon_{l-1}}_{\bar{\eta}\varepsilon_{l-1}}(P_{l-1}) \rightarrow \mathsf{H}\mathcal{R}_{\rho\varepsilon_{l-1}}(P_l) \rightarrow \mathsf{H}\mathcal{C}^{\bar{\rho}\varepsilon_{l-1}}_{\bar{\eta}\varepsilon_{l-1}}(P_l) \leftarrow \mathsf{H}\mathcal{R}_{\eta\varepsilon_l}(P_l) \leftarrow \mathsf{H}\mathcal{C}^{\bar{\rho}\varepsilon_l}_{\bar{\eta}\varepsilon_l}(P_l)$ that has the same persistence barcode as $\mathbb{V}$. Moreover, the reverse maps provided by Theorem 3.1 make the triangles commute in the resulting diagram interleaving $\mathbb{V}^*$ and $\mathbb{W}$.



Now, we remove the Rips complexes from $\mathbb{V}^*$ by composing all the adjacent maps with same orientation. Since composition preserves commutativity of the subdiagrams, the following diagram involving $\mathbb{W}$ (straight path) and the newly obtained zigzag $\mathbb{W}^*$ (curved path) commutes.



Hence, the zigzags $\mathbb{W}$ and $\mathbb{W}^*$ are identical. It suffices now to compare $\mathsf{Pers}(\mathbb{V}^*)$ to $\mathsf{Pers}(\mathbb{W}^*)$. Recall that $\mathbb{W}^*$ is obtained from $\mathbb{V}^*$ by removing the Rips complexes. In the process, every interval of $\mathsf{Pers}(\mathbb{V}^*)$ either vanishes or turns into some interval of $\mathsf{Pers}(\mathbb{W}^*)$. By the Space Removal Theorem 3.2, only the ephemeral intervals of $\mathsf{Pers}(\mathbb{V}^*)$ may vanish because there are no consecutive Rips

complexes in $\mathbb{V}^*$. Moreover, the full-length intervals of $\mathsf{Pers}(\mathbb{V}^*)$ are mapped bijectively to those of $\mathsf{Pers}(\mathbb{W}^*)$ because $\mathbb{V}^*$ and $\mathbb{W}^*$ have the same endpoints. It follows then from (i) that $\mathsf{Pers}(\mathbb{V}^*)$—and thus $\mathsf{Pers}(\mathbb{V})$—contains only full-length intervals of multiplicity $\dim \mathsf{H}X^\lambda$ and possibly some ephemeral intervals.

We turn $\mathbb{V}$ into $\mathbb{U}$ by removing the Čech complexes. First, we restrict $\mathbb{V}$ to $\mathsf{H}\mathcal{R}_{\rho\varepsilon_k}(P_{k+1}) \leftarrow \mathsf{H}\mathcal{C}^{\bar{\rho}\varepsilon_k}_{\bar{\eta}\varepsilon_k}(P_{k+1}) \leftarrow \cdots \leftarrow \mathsf{H}\mathcal{C}^{\bar{\rho}\varepsilon_{l-1}}_{\bar{\eta}\varepsilon_{l-1}}(P_l) \leftarrow \mathsf{H}\mathcal{R}_{\eta\varepsilon_l}(P_l)$, thus removing the Čech complexes at either ends of the zigzag. Since $k < l$, the Restriction Theorem 2.7 tells us that the full-length intervals in the barcode of the shortened zigzag $\mathbb{V}^*$ are in bijection with the ones in the barcode of $\mathbb{V}$, while the other intervals in $\mathsf{Pers}(\mathbb{V}^*)$ come from length-zero intervals in $\mathsf{Pers}(\mathbb{V})$ and cannot be longer. We then compose the incoming and outgoing maps at Čech complexes in the sequence to obtain $\mathbb{U}$. By the Space Removal Theorem 3.2, only the intervals starting or ending at a Čech complex can be affected by this operation, and these can only be shortened. Therefore, the number of full-length intervals remains the same as in the barcode of $\mathbb{V}^*$, and the other intervals remain ephemeral. □

4.3 Morozov zigzag

The limiting case of the oscillating Rips zigzag occurs when the multipliers η and ρ are equal. This case has been integrated into the Dionysus library [16] since early 2009.

$$\cdots \leftarrow \mathcal{R}_{\rho\varepsilon_i}(P_i) \rightarrow \mathcal{R}_{\rho\varepsilon_i}(P_{i+1}) \leftarrow \mathcal{R}_{\rho\varepsilon_{i+1}}(P_{i+1}) \rightarrow \cdots \tag{10}$$

We call the zigzag module induced at the homology level by this diagram the *Morozov zigzag* (*M-ZZ* for short). The motivation for letting $\eta = \rho$ is obvious from a computational point of view; as η gets closer to ρ, fewer simplex additions and deletions are required during the zigzag persistence calculation. Moreover, as reported by its author [26], the M-ZZ has given surprisingly good results in preliminary experiments, despite the fact that $\eta = \rho$ clearly violates the hypotheses of Theorem 4.2. We are able to provide the following weaker guarantee.

Theorem 4.3. *Let $\rho > 10$ be a multiplier. Let $X \subset \mathbb{R}^d$ be a compact set and let $P \subset \mathbb{R}^d$ be such that $\mathrm{d}_H(P, X) < \varepsilon$ with $\varepsilon < \frac{\rho-10}{(3+10\vartheta_d)\rho}\,\mathrm{wfs}(X)$. Then, for any $l > k$ such that*

$$\frac{10\varepsilon}{\rho-10} \leq \varepsilon_l \leq \varepsilon_k < \min\left\{\frac{1}{6}\mathrm{wfs}(X) - \varepsilon,\ \frac{5}{(1+5\vartheta_d)\rho+5}(\mathrm{wfs}(X)-\varepsilon)\right\},$$

the M-ZZ restricted to $\mathsf{H}\mathcal{R}_{\rho\varepsilon_k}(P_{k+1}) \leftarrow \cdots \leftarrow \mathsf{H}\mathcal{R}_{\rho\varepsilon_l}(P_l)$ has a number of full-length intervals that is at least the dimension of $\mathsf{H}X^\lambda$ for any $\lambda \in (0, \mathrm{wfs}(X))$.

Proof. Let $\mathbb{V}$ denote the restriction of the M-ZZ to

$$\mathsf{H}\mathcal{R}_{\rho\varepsilon_k}(P_{k+1}) \leftarrow \cdots \leftarrow \mathsf{H}\mathcal{R}_{\rho\varepsilon_l}(P_l).$$

Let $\mathbb{U}$ and $\mathbb{W}$ respectively be the image Čech zigzags with multipliers (η_1, ρ_1) and (η_2, ρ_2) restricted to the index set of $\mathbb{V}$, where

$$\eta_1 = \frac{\rho}{2} - 2\left(1 + \frac{\varepsilon}{\varepsilon_l}\right),\ \rho_1 = \frac{\rho}{2},\ \eta_2 = \vartheta_d\rho,\ \text{ and } \rho_2 = \vartheta_d\rho + 2\left(1 + \frac{\varepsilon}{\varepsilon_l}\right).$$

The canonical inclusions between Čech complexes induce homomorphisms between the spaces of $\mathbb{U}$ and $\mathbb{W}$ with the same index. This family of homomorphisms forms a module homomorphism $\mathbb{U} \rightarrow \mathbb{W}$. By (1), every inclusion $\mathcal{C}_{\rho_1\varepsilon_i}(Q) \hookrightarrow \mathcal{C}_{\eta_2\varepsilon_i}(Q)$ factors through $\mathcal{R}_{\rho\varepsilon_i}(Q)$, so $\mathbb{U} \rightarrow \mathbb{W}$ itself factors through $\mathbb{V}$. Let $\mathbb{U} \xrightarrow{\Phi} \mathbb{V} \xrightarrow{\Psi} \mathbb{W}$ be the factorization.

The multipliers and the bounds on ε_l and ε_k have been carefully chosen to satisfy the hypotheses of all three parts of Theorem 2.5 so that the following three statements are true for all i and j such that $k \leq i \leq j \leq l$.

(i) $\mathsf{H}\mathcal{C}^{\rho_1\varepsilon_i}_{\eta_1\varepsilon_i}(P_j)$ is isomorphic to $\mathsf{H}X^\lambda$ for all $\lambda \in (0, \mathrm{wfs}(X))$.

(ii) $\mathsf{H}\mathcal{C}^{\rho_1\varepsilon_i}_{\eta_1\varepsilon_i}(P_j) \to \mathsf{H}\mathcal{C}^{\rho_2\varepsilon_i}_{\eta_2\varepsilon_i}(P_j)$ is an isomorphism.

(iii) $\mathsf{H}\mathcal{C}^{\rho_1\varepsilon_i}_{\eta_1\varepsilon_i}(P_j) \to \mathsf{H}\mathcal{R}_{\rho\varepsilon_i}(P_j)$ is injective.

Items (i) and (ii) imply that the spaces in $\mathbb{U}$ are all isomorphic to $\mathsf{H}X^\lambda$ and the maps in $\mathbb{U}$ are all isomorphisms. Thus, $\mathsf{Pers}(\mathbb{U})$ contains at least $\dim \mathsf{H}X^\lambda$ full-length intervals. Item (ii) also implies that $\Psi \circ \Phi$ is an isomorphism. So, we may write $\mathbb{V} = \operatorname{im}\Phi \oplus \ker\Psi$. The uniqueness clause of the Interval Decomposition Theorem (Thm. 2.6) then implies that $\mathsf{Pers}(\operatorname{im}\Phi) \subseteq \mathsf{Pers}(\mathbb{V})$. Item (iii) implies that Φ is injective, and thus that $\mathsf{Pers}(\operatorname{im}\Phi) = \mathsf{Pers}(\mathbb{U})$. This completes the proof as we have shown that $\mathsf{Pers}(\mathbb{V})$ contains $\mathsf{Pers}(\mathbb{U})$, which contains at least $\dim \mathsf{H}X^\lambda$ full-length intervals. □

Thus, the topological signal of X persists in the M-ZZ throughout a sweet range of the form $[O(\varepsilon),\ \Omega(\mathrm{wfs}(X))]$. The question of the resilience of the topological noise within this range remains open, and there is currently no theoretical evidence that the noise should not persist under the mere assumption that X has positive weak feature size. This will be our motivation for introducing a discretized variant of the Morozov zigzag that provably kills the residual noise in Section 4.4.

For now we will tackle the problem differently and add further restrictions to our setting: on the one hand, the dimension of the homology groups will be restricted to be 0 or 1 (Section 4.3.1); on the other hand, the shape X underlying the data points will be assumed to have positive μ-reach for some large enough value μ (Section 4.3.2).

4.3.1 0-th or 1-st homology

If we are only concerned with the 0-th or 1-st homology groups, then we can take advantage of the following simple observation.

Lemma 4.4. *For any $Q \subset \mathbb{R}^d$ and any $\beta \geq \alpha \geq 0$,*

$$\begin{aligned}
\operatorname{rank}\mathsf{H}_0(\mathcal{R}_\alpha(Q)) \to \mathsf{H}_0(\mathcal{R}_\beta(Q)) &= \operatorname{rank}\mathsf{H}_0(\mathcal{C}_{\frac{\alpha}{2}}(Q)) \to \mathsf{H}_0(\mathcal{C}_{\frac{\beta}{2}}(Q)),\\
\operatorname{rank}\mathsf{H}_1(\mathcal{R}_\alpha(Q)) \to \mathsf{H}_1(\mathcal{R}_\beta(Q)) &\leq \operatorname{rank}\mathsf{H}_1(\mathcal{C}_{\frac{\alpha}{2}}(Q)) \to \mathsf{H}_1(\mathcal{C}_{\frac{\beta}{2}}(Q)),
\end{aligned}$$

where the homomorphisms are induced at the homology level by canonical inclusions between the complexes.

Proof. Recall from [17] that for any finite simplicial complexes $X \subseteq Y$, the rank of the homomorphism induced at the r-th homology level by the canonical inclusion $X \subseteq Y$ is given by

$$\operatorname{rank}\mathsf{H}_r(X) \to \mathsf{H}_r(Y) = \dim \frac{Z_r(X)}{Z_r(X) \cap B_r(Y)}, \tag{11}$$

where $Z_r(X)$ denotes the space of r-cycles in X and $B_r(Y)$ denotes the space of r-boundaries in Y (both are subgroups of the space of r-chains in Y).

When $Q \subset \mathbb{R}^d$, it follows from the definitions of Čech and Rips complexes that $\mathcal{C}_{\frac{\gamma}{2}}(Q)$ and $\mathcal{R}_\gamma(Q)$ have the same 1-skeleton, given any $\gamma \geq 0$. Hence, $Z_0(\mathcal{C}_{\frac{\alpha}{2}}(Q)) = Z_0(\mathcal{R}_\alpha(Q))$ and $B_0(\mathcal{C}_{\frac{\beta}{2}}(Q)) =$

$B_0(\mathcal{R}_\beta(Q))$, which implies by (11) that the maps $\mathsf{H}_0(\mathcal{R}_\alpha(Q)) \to \mathsf{H}_0(\mathcal{R}_\beta(Q))$ and $\mathsf{H}_0(\mathcal{C}_{\frac{\alpha}{2}}(Q)) \to \mathsf{H}_0(\mathcal{C}_{\frac{\beta}{2}}(Q))$ have same rank.

The definitions of Čech and Rips complexes also imply that the 2-skeleton of $\mathcal{R}_\gamma(Q)$ contains the one of $\mathcal{C}_{\frac{\gamma}{2}}(Q)$, while their 1-skeleton is the same as mentioned previously. Hence, $Z_1(\mathcal{C}_{\frac{\alpha}{2}}(Q)) = Z_1(\mathcal{R}_\alpha(Q))$ and $B_1(\mathcal{C}_{\frac{\beta}{2}}(Q)) \subseteq B_1(\mathcal{R}_\beta(Q))$, which implies by (11) that $\operatorname{rank}\mathsf{H}_1(\mathcal{R}_\alpha(Q)) \to \mathsf{H}_1(\mathcal{R}_\beta(Q)) \leq \operatorname{rank}\mathsf{H}_1(\mathcal{C}_{\frac{\alpha}{2}}(Q)) \to \mathsf{H}_1(\mathcal{C}_{\frac{\beta}{2}}(Q))$. □

Letting now $P, \rho, \varepsilon, \varepsilon_k, \varepsilon_l$ follow the hypotheses of Theorem 4.3, we have by Lemma 4.4 and for every index $i \in [k, l]$

$$\operatorname{rank}\mathsf{H}_0(\mathcal{R}_{\rho\varepsilon_{i+1}}(P_{i+1})) \to \mathsf{H}_0(\mathcal{R}_{\rho\varepsilon_i}(P_{i+1})) = \operatorname{rank}\mathsf{H}_0(\mathcal{C}_{\frac{\rho}{2}\varepsilon_{i+1}}(P_{i+1})) \to \mathsf{H}_0(\mathcal{C}_{\frac{\rho}{2}\varepsilon_i}(P_{i+1})) = \dim(\mathsf{H}_0(X^\lambda)),$$
$$\operatorname{rank}\mathsf{H}_1(\mathcal{R}_{\rho\varepsilon_{i+1}}(P_{i+1})) \to \mathsf{H}_1(\mathcal{R}_{\rho\varepsilon_i}(P_{i+1})) \leq \operatorname{rank}\mathsf{H}_1(\mathcal{C}_{\frac{\rho}{2}\varepsilon_{i+1}}(P_{i+1})) \to \mathsf{H}_1(\mathcal{C}_{\frac{\rho}{2}\varepsilon_i}(P_{i+1})) = \dim(\mathsf{H}_1(X^\lambda)).$$

Hence, in 0-th or 1-st homology, the noise in the Morozov zigzag is killed when going through the link $\mathcal{R}_{\rho\varepsilon_i}(P_{i+1}) \leftarrow \mathcal{R}_{\rho\varepsilon_{i+1}}(P_{i+1})$. More precisely, given $r \in \{0, 1\}$, call $\mathbb{V}$ the restriction of the Morozov zigzag to $\mathsf{H}_r(\mathcal{R}_{\rho\varepsilon_k}(P_{k+1})) \leftarrow \cdots \leftarrow \mathsf{H}_r(\mathcal{R}_{\rho\varepsilon_l}(P_l))$. On the one hand, the Restriction Theorem 2.7 implies that the total multiplicity of the intervals including $[\mathsf{H}_r(\mathcal{R}_{\rho\varepsilon_i}(P_{i+1})),\ \mathsf{H}_r(\mathcal{R}_{\rho\varepsilon_{i+1}}(P_{i+1}))]$ in $\mathbb{V}$ is at most $\dim(\mathsf{H}_r(X^\lambda))$. On the other hand, Theorem 4.3 implies that the multiplicity of the full-length interval in $\mathbb{V}$ is precisely $\dim(\mathsf{H}_r(X^\lambda))$. It follows that among the intervals containing $[\mathsf{H}_r(\mathcal{R}_{\rho\varepsilon_i}(P_{i+1})),\ \mathsf{H}_r(\mathcal{R}_{\rho\varepsilon_{i+1}}(P_{i+1}))]$, only the full-length one has non-zero multiplicity. Thus, $\mathsf{Pers}_r(\mathbb{V})$ contains only full-length intervals and intervals of type $[\mathsf{H}_r(\mathcal{R}_{\rho\varepsilon_i}(P_i)),\ \mathsf{H}_r(\mathcal{R}_{\rho\varepsilon_i}(P_{i+1}))]$. These are not ephemeral in the index scale of $\mathbb{V}$, however they become so once represented on the scale of the geometric scales. Hence,

Theorem 4.5. *Given a choice of multiplier $\rho > 10$, suppose $P \subset \mathbb{R}^d$ and there is some compact set $X \subset \mathbb{R}^d$ such that $\mathrm{d}_H(P, X) < \varepsilon$ with $\varepsilon < \frac{\rho-10}{(3+10\vartheta_d)\rho}\operatorname{wfs}(X)$. Then, for any $k < l$ such that*

$$\frac{10\varepsilon}{\rho - 10} \leq \varepsilon_k, \varepsilon_l < \min\left\{\frac{1}{6}\operatorname{wfs}(X) - \varepsilon,\ \frac{5}{(1 + 5\vartheta_d)\rho + 5}(\operatorname{wfs}(X) - \varepsilon)\right\},$$

the zigzag module induced by (10) at the r-th homology level ($r \in \{0, 1\}$), once restricted to $\mathsf{H}_r(\mathcal{R}_{\rho\varepsilon_k}(P_{k+1})) \leftarrow \cdots \leftarrow \mathsf{H}_r(\mathcal{R}_{\rho\varepsilon_l}(P_l))$, has a persistence barcode made only of two types of intervals:

- *full-length intervals (the signal), whose number is equal to the dimension of $\mathsf{H}_r(X^\lambda)$ for any $\lambda \in (0, \operatorname{wfs}(X))$,*
- *intervals of type $[\mathsf{H}_r(\mathcal{R}_{\rho\varepsilon_i}(P_i)),\ \mathsf{H}_r(\mathcal{R}_{\rho\varepsilon_i}(P_{i+1}))]$ (the noise), which are ephemeral (length zero) on the geometric scale.*

4.3.2 Sampled compact sets of positive μ-reach

Let us assume that X has positive μ-reach, denoted $\operatorname{rch}_\mu(X) > 0$, for some sufficiently large μ. Recall that $\operatorname{rch}_\mu(X)$ is the infimum of distances from X to points outside of X where the gradient of the distance to X is less than μ [7]. This quantity is smaller than $\operatorname{wfs}(X)$, so having positive μ-reach (for some $\mu > 0$) implies having positive weak feature size. Moreover it is a stricly stronger hypothesis, under which the set X and its offsets enjoy further properties. Relevant to our problem is the fact that, if $\mathrm{d}_H(P, X) < \varepsilon$ with ε sufficiently small compared to $\operatorname{rch}_\mu(X)$ and μ sufficiently large, then for some values of α the Rips complex $\mathcal{R}_\alpha(P)$ is homotopy equivalent to X^λ for $\lambda \in (0, \operatorname{rch}_\mu(X))$—see [1, Theorem 14]. An immediate consequence of this result is that for a multiplier

ρ and an index i, $\mathcal{R}_{\rho\varepsilon_i}(P_i)$ and $\mathcal{R}_{\rho\varepsilon_i}(P_{i+1})$ are both homotopy equivalent to X^λ for $\lambda \in (0, \mathrm{rch}_\mu(X))$ whenever

$$\frac{\mu(2-\mu)(2\rho\varepsilon_i - 2\vartheta_d\varepsilon_i - 2(\varepsilon_i+\varepsilon))}{1+\mu(1-\mu) - \sqrt{1-\mu(2-\mu)\left(\frac{(2\vartheta_d\rho+1)\varepsilon_i+\varepsilon}{\mathrm{rch}_\mu(X)}\right)^2}} > \mathrm{rch}_\mu(X).$$

This condition depends on $\mathrm{rch}_\mu(X)$, its parameter μ, the multiplier ρ, the Hausdorff distance of the sample ε, and ε_i. Attali et al. [1] showed that there do exist values for which the condition is satisfied. We do not derive the space of valid assignment of constants here, but merely note that this result implies that there is a multiplier ρ and a range of scales of the form $[O(\varepsilon),\ \Omega(\mathrm{rch}_\mu(X))]$ for which the M-ZZ exhibits no noise at all. This holds because Theorem 4.3 implies that the signal is present in the sweet range and the Attali et al. result shows that every space in the strictly smaller range has the same homology as X^λ. We call this the *sweeter range*. Note that the quantities hidden in the big-O and big-Ω notations depend on μ, so the sweeter range requires a sufficiently large μ to be non-empty. Moreover, since the upper bound depends on $\mathrm{rch}_\mu(X)$ rather than $\mathrm{wfs}(X)$, the sweeter range can be arbitrarily smaller than the sweet range. However, when the input permits a sweeter range, the guarantees regarding the signal-to-noise ratio are correspondingly stronger.

4.4 Discretized Morozov Zigzag

It is possible to construct the M-ZZ over just a subsequence of the indices $1 \dots n$. Given such a subsequence $n_1, \dots, n_r$ with $n_1 = 1$ and $n_r = n$, the *discretized Morozov zigzag* (*dM-ZZ* for short) is defined as

$$\cdots \leftarrow \mathcal{R}_{\rho\varepsilon_{n_i}}(P_{n_i}) \rightarrow \mathcal{R}_{\rho\varepsilon_{n_i}}(P_{n_{i+1}}) \leftarrow \mathcal{R}_{\rho\varepsilon_{n_{i+1}}}(P_{n_{i+1}}) \rightarrow \cdots .$$

The dM-ZZ has the same guarantee as the M-ZZ with regards to preserving the signal in the sweet range, but it also has the added benefit that it can kill all of the noise in that range.

Theorem 4.6. *Let $\rho > 10$ be a multiplier. Let $X \subset \mathbb{R}^d$ be a compact set, let $P \subset \mathbb{R}^d$ be a finite set, and let $\varepsilon = \mathrm{d}_H(P, X)$ where $\varepsilon < \frac{\rho-10}{(3+10\vartheta_d)\rho}\mathrm{wfs}(X)$. Let $n_1, \dots, n_r$ be a subsequence of the indices $[1, n]$ such that*

$$\varepsilon_{n_i} \geq 2\vartheta_d\varepsilon_{n_{i+1}} + \frac{4}{\rho}\left(\varepsilon + \varepsilon_{n_{i+1}}\right)$$

for all $i \in [1, r-1]$. Then, for any $n_l > n_k$ such that

$$\frac{10\varepsilon}{\rho-10} \leq \varepsilon_{n_l} \leq \varepsilon_{n_k} < \min\left\{\frac{1}{6}\mathrm{wfs}(X) - \varepsilon,\ \frac{5}{(1+5\vartheta_d)\rho+5}(\mathrm{wfs}(X)-\varepsilon)\right\},$$

the dM-ZZ restricted to $\mathsf{H}\mathcal{R}_{\rho\varepsilon_{n_k}}(P_{n_{k+1}}) \leftarrow \cdots \leftarrow \mathsf{H}\mathcal{R}_{\rho\varepsilon_{n_l}}(P_{n_l})$ has a barcode with only three classes of intervals:

- *full-length intervals, whose number is equal to the dimension of $\mathsf{H}X^\lambda$ for any $\lambda \in (0, \mathrm{wfs}(X))$,*
- *ephemeral (length zero) intervals,*
- *intervals of the form $[\mathsf{H}\mathcal{R}_{\rho\varepsilon_{n_i}}(P_{n_i}),\ \mathsf{H}\mathcal{R}_{\rho\varepsilon_{n_i}}(P_{n_{i+1}})]$, which are ephemeral on the scale of the geometric scales.*

Proof. The proof of Theorem 4.3 applies verbatim to the dM-ZZ, with indices $1, 2, \cdots, n$ replaced by $n_1, n_2, \cdots, n_r$. Thus, the restriction $\mathbb{V}$ of the dM-ZZ to $\mathsf{H}\mathcal{R}_{\rho\varepsilon_{n_k}}(P_{n_{k+1}}) \leftarrow \cdots \leftarrow \mathsf{H}\mathcal{R}_{\rho\varepsilon_{n_l}}(P_{n_l})$ has a persistence barcode with at least $\dim(\mathsf{H}X^\lambda)$ full-length intervals.

In the hypothesis, $\varepsilon_{n_i} \geq 2\vartheta_d\varepsilon_{n_{i+1}} + \frac{4}{\rho}\left(\varepsilon + \varepsilon_{n_{i+1}}\right)$, the first term in the sum guarantees that the inclusion $\mathcal{R}_{\rho\varepsilon_{n_i}}(P_{n_{i+1}}) \leftarrow \mathcal{R}_{\rho\varepsilon_{n_{i+1}}}(P_{n_{i+1}})$ factors through $\mathcal{C}_{\frac{\rho}{2}\varepsilon_{n_i}}(P_{n_{i+1}}) \leftarrow \mathcal{C}_{\vartheta_d\rho\varepsilon_{n_{i+1}}}(P_{n_{i+1}})$. The

second term in the sum guarantees that the scale change is sufficiently large compared to the Hausdorff distance from $P_{n_{i+1}}$ to X, which is at most $\varepsilon + \varepsilon_{n_{i+1}}$. This allows us to apply Theorem 2.5(i) to guarantee that

$$\operatorname{rank}\left(\mathsf{H}\mathcal{C}_{\frac{\rho}{2}\varepsilon_{n_i}}(P_{n_{i+1}}) \leftarrow \mathsf{H}\mathcal{C}_{\vartheta_d\rho\varepsilon_{n_{i+1}}}(P_{n_{i+1}})\right) = \dim(\mathsf{H}X^\lambda).$$

Therefore, by composition we have

$$\operatorname{rank}\left(\mathsf{H}\mathcal{R}_{\rho\varepsilon_{n_i}}(P_{n_{i+1}}) \leftarrow \mathsf{H}\mathcal{R}_{\rho\varepsilon_{n_{i+1}}}(P_{n_{i+1}})\right) \leq \dim(\mathsf{H}X^\lambda).$$

Intuitively, this means that only the signal can go through the link $\mathsf{H}\mathcal{R}_{\rho\varepsilon_{n_i}}(P_{n_{i+1}}) \leftarrow \mathsf{H}\mathcal{R}_{\rho\varepsilon_{n_{i+1}}}(P_{n_{i+1}})$, and that the noise gets killed. More formally, by the Restriction Theorem 2.7 the total multiplicity of the intervals spanning $[\mathsf{H}\mathcal{R}_{\rho\varepsilon_{n_i}}(P_{n_{i+1}}),\ \mathsf{H}\mathcal{R}_{\rho\varepsilon_{n_{i+1}}}(P_{n_{i+1}})]$ in $\mathsf{Pers}(\mathbb{V})$ is at most $\dim(\mathsf{H}X^\lambda)$. It follows that among these intervals only the full-length interval has non-zero multiplicity. Thus, $\mathsf{Pers}(\mathbb{V})$ contains three types of intervals: full-length intervals, ephemeral intervals, and intervals of the form $[\mathsf{H}\mathcal{R}_{\rho\varepsilon_{n_i}}(P_{n_i}),\ \mathsf{H}\mathcal{R}_{\rho\varepsilon_{n_i}}(P_{n_{i+1}})]$. These last intervals are not ephemeral on the index scale, but they are ephemeral on the geometric scale because the birth and death times are both $\rho\varepsilon_{n_i}$. □

Note that ε usually remains unknown in practice, so the user cannot directly choose the maximal index set that guarantees $\varepsilon_{n_i} \geq 2\vartheta_d\varepsilon_{n_{i+1}} + \frac{4}{\rho}\left(\varepsilon + \varepsilon_{n_{i+1}}\right)$. The bounds given in Theorem 4.6 suggest to choose the indices

$$n_1 = 1 \text{ and } n_{i+1} = \min\{j > n_i \mid \varepsilon_j \leq \zeta\varepsilon_{n_i}\} \text{ for } i \geq 1,$$

where

$$\zeta = \frac{3\rho + 20}{10(\vartheta_d\rho + 2)}. \tag{12}$$

Thus,the conclusion of the theorem continues to hold within the same sweet range as long as $\varepsilon_{n_i} \geq \frac{10\varepsilon}{\rho - 10}$. Any smaller value could be chosen for ζ without affecting the sweet range. Nevertheless, a larger value of ζ has advantages, since the smaller the gaps on the geometric scale the more chances there are that the set of discretization values $\{\varepsilon_{n_i}\}$ intersects the sweet range, and furthermore, smaller gaps mean smaller complexes.

4.5 Image Rips zigzag

We end this section with another Rips zigzag construction similar to the image Čech zigzag. It consists of a nested pair of Morozov zigzags with multipliers $\rho \geq \eta \geq 0$. Canonical inclusions between Rips complexes give the following commutative diagram at the homology level.

$$\begin{array}{ccccccc}
\cdots \leftarrow \mathsf{H}\mathcal{R}_{\rho\varepsilon_i}(P_i) & \rightarrow & \mathsf{H}\mathcal{R}_{\rho\varepsilon_i}(P_{i+1}) & \leftarrow & \mathsf{H}\mathcal{R}_{\rho\varepsilon_{i+1}}(P_{i+1}) & \rightarrow \cdots \\
\uparrow & & \uparrow & & \uparrow & \\
\cdots \leftarrow \mathsf{H}\mathcal{R}_{\eta\varepsilon_i}(P_i) & \rightarrow & \mathsf{H}\mathcal{R}_{\eta\varepsilon_i}(P_{i+1}) & \leftarrow & \mathsf{H}\mathcal{R}_{\eta\varepsilon_{i+1}}(P_{i+1}) & \rightarrow \cdots
\end{array}$$

The vertical arrows in this diagram form a module homomorphism from the M-ZZ with multiplier η to the M-ZZ with multiplier ρ. The *image Rips zigzag* (iR-ZZ) is defined as the image of this homomorphism. It is written as follows, where each $\mathsf{H}\mathcal{R}^{\rho\varepsilon_r}_{\eta\varepsilon_r}(P_s)$ denotes the image of the map $\mathsf{H}\mathcal{R}_{\eta\varepsilon_r}(P_s) \rightarrow \mathsf{H}\mathcal{R}_{\rho\varepsilon_r}(P_s)$.

$$\cdots \leftarrow \mathcal{R}^{\rho\varepsilon_i}_{\eta\varepsilon_i}(P_i) \rightarrow \mathcal{R}^{\rho\varepsilon_i}_{\eta\varepsilon_i}(P_{i+1}) \leftarrow \mathcal{R}^{\rho\varepsilon_{i+1}}_{\eta\varepsilon_{i+1}}(P_{i+1}) \rightarrow \cdots$$

Image Rips zigzags have been available in the Dionysus library [16] since early 2009, with no theoretical guarantee on their behavior. Here we provide a guarantee on the output that is similar to the one obtained in Theorem 4.2 for the oscillating Rips zigzag.

Theorem 4.7. *Let ρ and η be multipliers such that $\rho > 10$ and $\frac{3}{\vartheta_d} < \eta < \frac{\rho-4}{2\vartheta_d}$. Let $X \subset \mathbb{R}^d$ be a compact set, let $P \subset \mathbb{R}^d$ be a finite set, and let $\varepsilon = \mathrm{d}_H(P, X)$. Then, for any $l > k$ such that*

$$\max\left\{\frac{3\varepsilon}{\vartheta_d\eta - 3},\ \frac{4\varepsilon}{\rho - 2\vartheta_d\eta - 4}\right\} \leq \varepsilon_l \leq \varepsilon_k < \min\left\{\frac{1}{6}\mathrm{wfs}(X) - \varepsilon,\ \frac{1}{\vartheta_d\rho + 1}\left(\mathrm{wfs}(X) - \varepsilon\right)\right\},$$

the iR-ZZ restricted to $\mathsf{H}\mathcal{R}^{\rho\varepsilon_k}_{\eta\varepsilon_k}(P_{k+1}) \leftarrow \cdots \leftarrow \mathsf{H}\mathcal{R}^{\rho\varepsilon_l}_{\eta\varepsilon_l}(P_l)$ contains only isomorphisms, and its spaces are isomorphic to $\mathsf{H}X^\lambda$ for any $\lambda \in (0, \mathrm{wfs}(X))$. Therefore, its persistence barcode is made only of full-length intervals, whose number equals the dimension of $\mathsf{H}X^\lambda$.

Proof. Our hypotheses imply $\frac{\rho}{2} \geq \vartheta_d\eta$, so we can use (1) to obtain the following diagram where all arrows are canonical inclusions.

$$\begin{array}{ccccccc}
\cdots \leftarrow & \mathcal{C}_{\vartheta_d\rho\varepsilon_i}(P_i) & \rightarrow & \mathcal{C}_{\vartheta_d\rho\varepsilon_i}(P_{i+1}) & \leftarrow & \mathcal{C}_{\vartheta_d\rho\varepsilon_{i+1}}(P_{i+1}) & \rightarrow \cdots \\
& \uparrow & & \uparrow & & \uparrow & \\
\cdots \leftarrow & \mathcal{R}_{\rho\varepsilon_i}(P_i) & \rightarrow & \mathcal{R}_{\rho\varepsilon_i}(P_{i+1}) & \leftarrow & \mathcal{R}_{\rho\varepsilon_{i+1}}(P_{i+1}) & \rightarrow \cdots \\
& \uparrow & & \uparrow & & \uparrow & \\
\cdots \leftarrow & \mathcal{C}_{\frac{\rho}{2}\varepsilon_i}(P_i) & \rightarrow & \mathcal{C}_{\frac{\rho}{2}\varepsilon_i}(P_{i+1}) & \leftarrow & \mathcal{C}_{\frac{\rho}{2}\varepsilon_{i+1}}(P_{i+1}) & \rightarrow \cdots \\
& \uparrow & & \uparrow & & \uparrow & \\
\cdots \leftarrow & \mathcal{C}_{\vartheta_d\eta\varepsilon_i}(P_i) & \rightarrow & \mathcal{C}_{\vartheta_d\eta\varepsilon_i}(P_{i+1}) & \leftarrow & \mathcal{C}_{\vartheta_d\eta\varepsilon_{i+1}}(P_{i+1}) & \rightarrow \cdots \\
& \uparrow & & \uparrow & & \uparrow & \\
\cdots \leftarrow & \mathcal{R}_{\eta\varepsilon_i}(P_i) & \rightarrow & \mathcal{R}_{\eta\varepsilon_i}(P_{i+1}) & \leftarrow & \mathcal{R}_{\eta\varepsilon_{i+1}}(P_{i+1}) & \rightarrow \cdots \\
& \uparrow & & \uparrow & & \uparrow & \\
\cdots \leftarrow & \mathcal{C}_{\frac{\eta}{2}\varepsilon_i}(P_i) & \rightarrow & \mathcal{C}_{\frac{\eta}{2}\varepsilon_i}(P_{i+1}) & \leftarrow & \mathcal{C}_{\frac{\eta}{2}\varepsilon_{i+1}}(P_{i+1}) & \rightarrow \cdots
\end{array}$$

This diagram induces a homomorphism Φ from the image Rips zigzag $\mathbb{U}$ of parameters η, ρ to the image Čech zigzag $\mathbb{V}$ of parameters $\vartheta_d\eta, \vartheta_d\rho$. Call $\mathbb{U}^*$ (resp. $\mathbb{V}^*$) the restriction of $\mathbb{U}$ (resp. $\mathbb{V}$) to the scale range $[(\varepsilon_k, P_{k+1}),\ (\varepsilon_l, P_l)]$. The restriction of Φ to $\mathbb{U}^*$ is an isomorphism onto $\mathbb{V}^*$. To see this, pick an arbitrary index i in the range $[k+1, l]$ and consider the following sequence of inclusions between Čech complexes:

$$\mathcal{C}_{\frac{\eta}{2}\varepsilon_i}(P_i) \xrightarrow{a} \mathcal{C}_{\vartheta_d\eta\varepsilon_i}(P_i) \xrightarrow{b} \mathcal{C}_{\frac{\rho}{2}\varepsilon_i}(P_i) \xrightarrow{c} \mathcal{C}_{\vartheta_d\rho\varepsilon_i}(P_{i+1}).$$

Under our geometric hypotheses, it follows from Theorem 2.5(i) that the maps $c \circ b \circ a$ and b induce homomorphisms of same rank at the homology level. Since a and c factor through Rips complexes as shown above, rank inequalities induced by composition imply that the map $\mathsf{H}\mathcal{R}^{\rho\varepsilon_i}_{\eta\varepsilon_i}(P_i) \rightarrow \mathsf{H}\mathcal{C}^{\vartheta_d\rho\varepsilon_i}_{\vartheta_d\eta\varepsilon_i}(P_i)$ is an isomorphism. The same argument holds for the map $\mathsf{H}\mathcal{R}^{\rho\varepsilon_i}_{\eta\varepsilon_i}(P_{i+1}) \rightarrow \mathsf{H}\mathcal{C}^{\vartheta_d\rho\varepsilon_i}_{\vartheta_d\eta\varepsilon_i}(P_{i+1})$, with $i \in [k, l-1]$. Thus, the restriction of Φ to $\mathbb{U}^*$ is indeed an isomorphism onto $\mathbb{V}^*$. The conclusion of the theorem follows because the spaces in $\mathbb{V}^*$ are isomorphic to $\mathsf{H}X^\lambda$ and the maps in $\mathbb{V}^*$ are isomorphisms, as guaranteed by Theorem 4.1. □

5 Complexity bounds

We assume that the ordering of the points of P is by furthest point sampling[4]. Then, every prefix P_i is an ε_i-sparse ε_i-sample of P, and so standard ball packing arguments like the one used in [19, Lemma 4.1] can be applied to bound the memory usage and running times of our Rips zigzags. The corresponding bounds are given below. Note that they are stated in terms of the doubling dimension of the input point set, because it may be smaller than the ambient Euclidean dimension and thus gives sharper bounds.

[4]Arbitrary orderings may lead to local oversampling and thus to an uncontrolled local growth of the complex, regardless of the zigzag considered.

Memory usage. The relevant parameter here is the multiplier ρ, which conditions the size of the biggest complex in a Rips zigzag.

Theorem 5.1. *Suppose P is sitting in some metric space of doubling dimension d. Then, for any $k \geq 0$, the number of k-simplices in the current complex at any time of the construction of the M-ZZ of parameter ρ is at most $2^{O(kd \log \rho)}n$, where n is the cardinality of P. The same bound applies to the oR-ZZ and iR-ZZ of parameter ρ, regardless of the value of parameter $\eta \leq \rho$. Finally, given a constant $\zeta \in (0, 1]$ and indices $n_1 = 1$ and $n_{i+1} = \min\{j > n_i \mid \varepsilon_j \leq \zeta\varepsilon_{n_i}\}$, the number of k-simplices in any complex of the dM-ZZ is at most $2^{O(kd \log \frac{\rho}{\zeta})}n$.*

Proof. We prove the result in the case of the dM-ZZ; the other cases following by letting $\zeta = 1$. Let $1 = n_1 < n_2 < \cdots < n_{r-1} < n_r$ be the discretization indices. Since each prefix P_{n_i} is ε_{n_i}-sparse, a standard ball packing argument shows that every point $p \in P_{n_i}$ is connected to at most $2^{O(d \log \rho)}$ neighbors in the Rips complex $\mathcal{R}_{\rho\varepsilon_{n_i}}(P_{n_i})$. These neighbors can form at most $\binom{2^{O(d \log \rho)}}{k} = 2^{O(kd \log \rho)}$ k-simplices with p. Thus, the total number of k-simplices in $\mathcal{R}_{\rho\varepsilon_{n_i}}(P_{n_i})$ is at most $2^{O(kd \log \rho)}n_i \leq 2^{O(kd \log \rho)}n$.

Let us now bound the size of $\mathcal{R}_{\rho\varepsilon_{n_i}}(P_{n_{i+1}})$. It follows from the definition of n_{i+1} that $P_{n_{i+1}-1}$ is $\zeta\varepsilon_{n_i}$-sparse. Then, the same ball packing argument as above shows that every point $p \in P$ has at most $2^{O(d \log \frac{\rho}{\zeta})}$ points of $P_{n_{i+1}-1}$ within distance $\rho\varepsilon_{n_i}$. Applying this result to every $p \in P_{n_{i+1}}$, and observing that the set difference $P_{n_{i+1}} \setminus P_{n_{i+1}-1}$ consists only of the point $p_{n_{i+1}}$, we deduce that every vertex of $\mathcal{R}_{\rho\varepsilon_{n_i}}(P_{n_{i+1}})$ has at most $2^{O(d \log \frac{\rho}{\zeta})}$ neighbors, as previously. The rest of the analysis follows. □

Since the theoretical lower bounds on ρ derived in Section 4 are constant ($\rho > 10$), one is allowed to set ρ to some constant value in practice and benefit from our guarantees on the quality of the output. Meanwhile, Theorem 5.1 ensures that the number of k-simplices in the current complex remains at most $2^{O(kd)}n$ throughout the construction of the M-ZZ, oR-ZZ or iR-ZZ. When using the dM-ZZ, one can also set ζ to be a constant as in (12), thus benefiting from the theoretical guarantees on the quality of the output while maintaining the number of k-simplices in the current complex below $2^{O(kd)}n$ throughout the construction of the zigzag. Note however that the exact complex size is bigger than the one achieved with the other types of Rips zigzags when $\zeta < 1$. These asymptotic bounds are as good in order of magnitude as the ones achieved with previous lightweight structures [15, 27].

Runtime. In experiments, we observed a significant slowdown in the running time of the oR-ZZ compared to the M-ZZ, iR-ZZ and dM-ZZ. This is a result of the oR-ZZ inserting and removing the same simplices many times. Thus, the total number of insertions is a relevant indicator of running time, and it is driven by the value of η: the closer it is to ρ, the fewer simplex insertions and deletions occur during the zigzag calculation.

Theorem 5.2. *Suppose P is sitting in some metric space of doubling dimension d. Then, for any $k \geq 0$, the total number of k-simplices inserted in the current complex throughout the construction of the M-ZZ of parameter ρ is at most $2^{O(kd \log \rho)}n$, where n is the cardinality of P. The same bound applies to the iR-ZZ of parameter ρ, regardless of the value of parameter $\eta \leq \rho$. For the oR-ZZ of parameters η, ρ, the bound becomes $2^{O(kd \log \rho)}n^2$. Finally, given a constant $\zeta \in (0, 1]$ and indices $n_1 = 1$ and $n_{i+1} = \min\{j > n_i \mid \varepsilon_j \leq \zeta\varepsilon_{n_i}\}$, the bound for the dM-ZZ is $2^{O(kd \log \frac{\rho}{\zeta})}n$.*

Proof. We begin with the M-ZZ, for which we will use a simple charging argument. Observe that simplex insertions occur only when a forward arrow is encountered in the zigzag. For any such arrow, the current complex is enlarged by adding a new vertex p_{i+1} and connecting it to the rest of the complex. By the same packing argument as in the proof of Theorem 5.1, p_{i+1} forms at most $2^{O(d \log \rho)}$ edges with the points of P_i, therefore the number of k-simplices in its star in $\mathcal{R}_{\rho\varepsilon_i}(P_{i+1})$ is at most $2^{O(kd \log \rho)}$. This is also the number of k-simplices created at this stage of the algorithm. Hence, the total number of k-simplices inserted throughout the process is at most $2^{O(kd \log \rho)}n$. This bound applies also to the iR-ZZ, which maintains two Morozov zigzags: one of parameter ρ, the other of parameter $\eta \leq \rho$.

The case of the dM-ZZ is similar, with the additional twist that more than one vertex is added to the current complex when going from $\mathcal{R}_{\rho\varepsilon_{n_i}}(P_{n_i})$ to $\mathcal{R}_{\rho\varepsilon_{n_i}}(P_{n_{i+1}})$. Nevertheless, as observed in the proof of Theorem 5.1, the points of $P_{n_{i+1}-1}$ are $\zeta\varepsilon_{n_i}$-sparse, so the number of edges in the star of any point of $P_{n_{i+1}}$ in $\mathcal{R}_{\rho\varepsilon_{n_i}}(P_{n_{i+1}})$ is at most $2^{O(d \log \frac{\rho}{\zeta})}$, and the number of k-simplices is bounded by $2^{O(kd \log \frac{\rho}{\zeta})}$. Hence, the total number of k-simplices inserted at this stage of the algorithm is at most $2^{O(d \log \frac{\rho}{\zeta})}(n_{i+1} - n_i)$. The result follows.

Finally, the case of the oR-ZZ with parameters $\eta < \rho$ is trickier. Due to the fact that both the vertex set and the Rips parameter increase when a forward arrow is encountered in the zigzag, we cannot simply charge the new simplex insertions to the newly added vertices: former vertices also form new simplices together. In fact our bound is obtained by a cruder argument: for any forward arrow, the current complex contains at most $2^{O(kd \log \rho)}n$ k-simplices before and after following the arrow. Hence, in the worst case the complex is rebuilt entirely, which means inserting at most $2^{O(kd \log \rho)}n$ k-simplices in total. Since this is true for any one of the $n-1$ forward arrows, the claimed quadratic bound follows. □

6 Experiments

Our implementation of the Rips zigzags of Section 4 is available as a package (called *homology-zigzags*) of the C++ library Dionysus [16].

Manufactured data. As a proof of concept, we ran our code on the so-called *Clifford data set* from [22], which was obtained by evenly spacing $2,000$ points along the line $l : y = 20x \mod 2\pi$ in the 2-d flat torus $(\mathbb{R} \mod 2\pi)^2$, then mapping the points onto the Clifford torus in $\mathbb{R}^4$ via the embedding $f : (u, v) \mapsto (\cos u, \sin u, \cos v, \sin v)$. This data set admits three non-trivial candidate underlying spaces: at small scales, the image of l through f, which is a closed helicoidal curve on the torus; at larger scales, the torus itself; at even larger scales, the 3-sphere of radius $\sqrt{2}$ on which the torus is sitting.

zigzag	parameters	max. complex size	total # insertions
M-ZZ	$\rho = 3$	107927	398107
dM-ZZ	$\rho = 3$, $\zeta = 0.9$	162919	604084
iR-ZZ	$\eta = 3$, $\rho = 3.2$	174436	1003215
oR-ZZ	$\eta = 3$, $\rho = 3.2$	174436	7252772

Table 1: Maximum complex size (in number of simplices) and total number of simplex insertions.

We ran the M-ZZ, dM-ZZ, iR-ZZ and oR-ZZ using the parameter values given in Table 1. Although these values lie outside the intervals prescribed by the theory, they were sufficient to

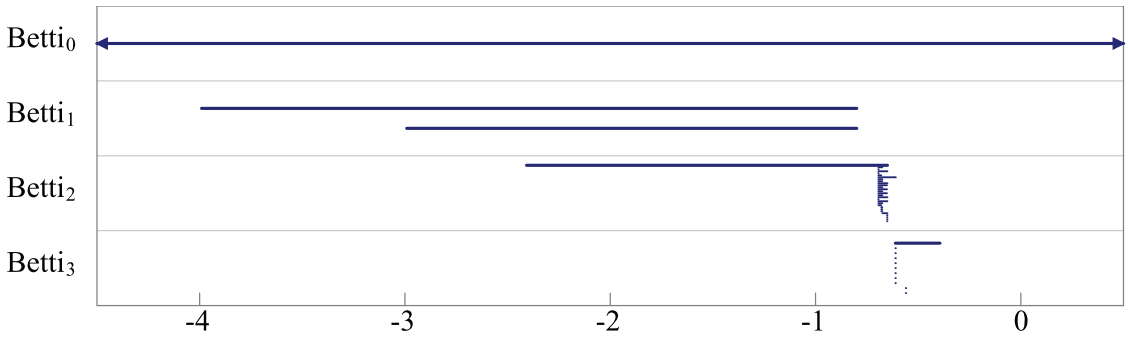


Figure 1: Barcode of the M-ZZ.

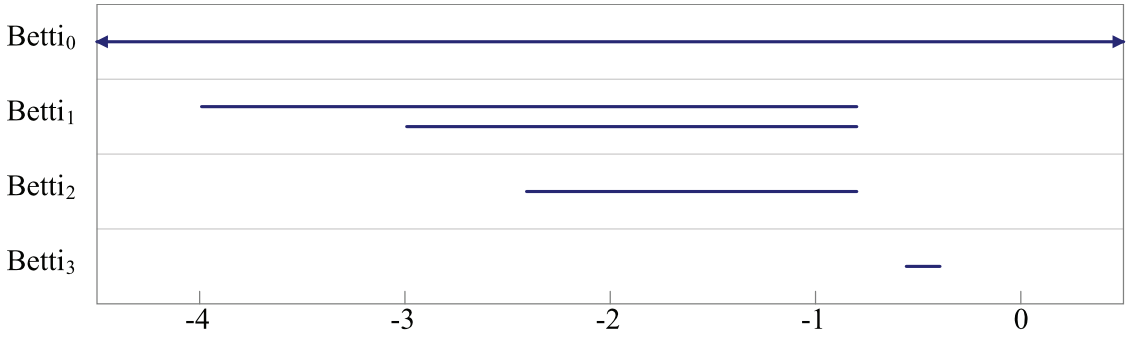


Figure 2: Barcode of the dM-ZZ.

obtain good results in practice. The corresponding results for the M-ZZ and dM-ZZ are reported in Figures 1 and 2 respectively. The barcodes are represented on a logarithmic geometric scale (i.e. the horizontal axis shows the value of $\log_2 \varepsilon_i$), with ephemeral (length zero) intervals removed for clarity. The results obtained with the iR-ZZ and oR-ZZ are similar to Figure 2 and therefore omitted.

The three spaces underlying the input data (curve, torus, 3-sphere) appear in all these barcodes, meanwhile the topological noise remains small (M-ZZ) or even ephemeral (dM-ZZ, iR-ZZ, oR-ZZ). Of particular interest is the 3-sphere, whose corresponding 3-homology cycle appears only at large scales and for a short while due to the fact that the 3-sphere is not densely sampled by the point cloud. This delicate 3-cycle certainly exists in the barcode of the standard Rips filtration, however it cannot be observed in practice. As mentioned in [22] the union of balls of radius α around the data points covers the entire sphere only for $\alpha \geq \sqrt{4-2\sqrt{2}}$, so the corresponding 3-homology cycle can appear in the Rips filtration only at a parameter value $\alpha \geq \frac{\sqrt{4-2\sqrt{2}}}{\vartheta_d} > \sqrt{2}$. Now, a simulation reveals that the 4-skeleton of the Rips complex at such parameter values contains more than 31 billion simplices, a size that lies at least 2 orders of magnitude beyond the sizes currently handled by existing implementations. On a 24-GB machine we were able to store the 4-skeleton of the Rips filtration and compute its persistent homology within the main memory up to $\alpha \approx 0.625$ using Dionysus. The corresponding truncated barcode is represented on a $\log_2$ scale in Figure 3. As expected, it shows only the curve and the torus, not the 3-sphere.

For comparison we recall in Figure 4 the barcode obtained by Hudson et al. using their mesh-based filtration [22]. Although the scale has been adapted to be the same as in the previous figures, any direct comparison of the barcodes should be made with the caveat that the intervals are half-

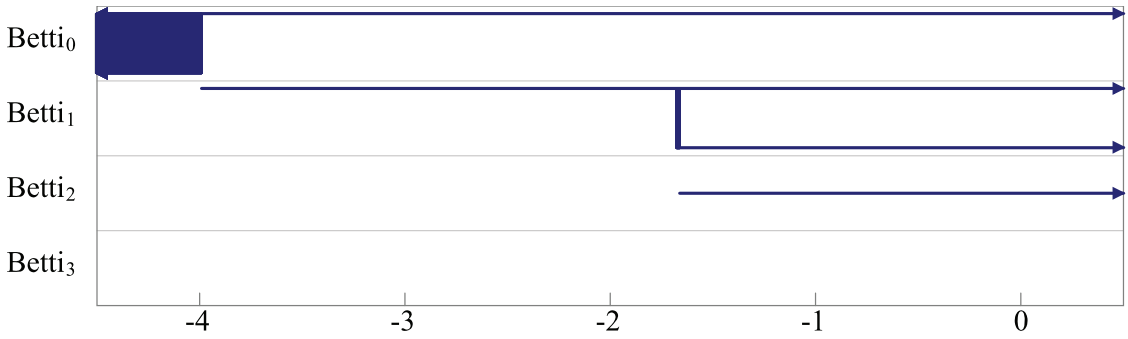


Figure 3: Barcode of the standard Rips filtration.

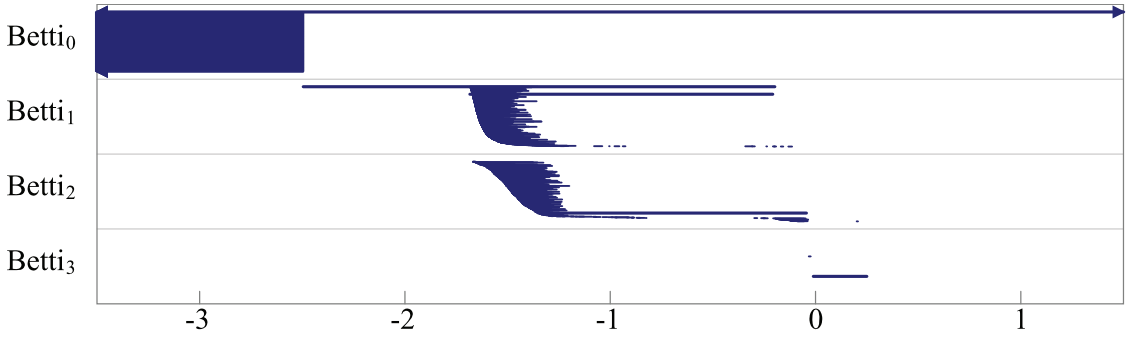
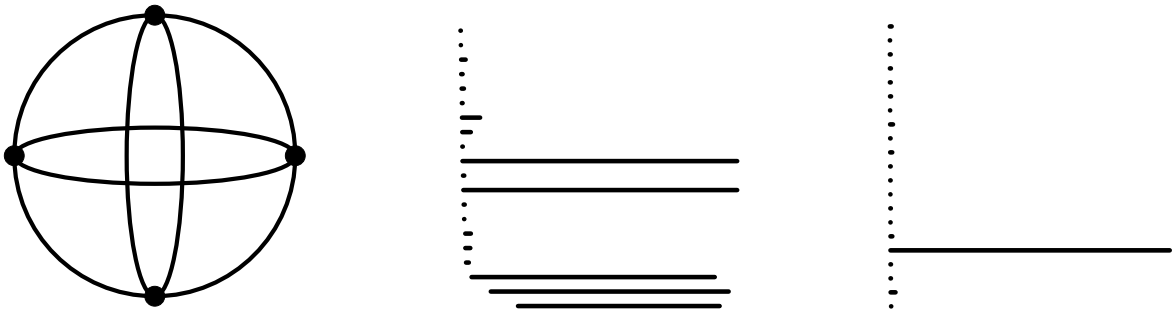


Figure 4: Barcode of the mesh-based filtration.

open as in [22], rather than closed as in the present paper. Nevertheless, the general trend here is that although the three spaces underlying the input data do appear in the barcode, the amount of noise is significant and its structure is quite irregular, despite the regularity of the data. Most notably, the amplitude of the noise in 1- or 2-homology is larger than the amplitude of the signal in 3-homology, which tends to obscure the information carried within the barcode. The superiority of Rips zigzags over mesh-based filtrations is also in terms of efficiency: as reported in [22] the mesh-based filtration contains 12 million simplices, whereas our Rips zigzags contain less than 200,000 simplices at any given time—see Table 1. The oR-ZZ did perform over 7 million simplex insertions and deletions in total, however the other zigzags performed much fewer such operations as predicted by Theorem 5.2. In practice, computing any of our Rips zigzags together with its barcode took only a few minutes on a single Intel Xeon CPU core running at 2.40 GHz, whereas computing the mesh-based filtration and its barcode took hours on a similar architecture.

Real data. To demonstrate the practicality of our approach, we applied it to a real-life example derived from the statistics of natural images. The data set, described by Lee et al. in [25], consists of 4.2 million high-contrast log-intensity 3×3 patches extracted from van Hateren's collection of still images [28]. Each patch is normalized in such a way that it can be represented as a point on the unit sphere $\mathbb{S}^7$ in the Euclidean space $\mathbb{R}^8$.

This 8-dimensional data set was studied in depth by Carlsson and co-authors [6, 12], who used the witness complex filtration to analyze its topological structure. Their findings led them to conjecture that the data are concentrated mostly around three circles with four points of intersection, as depicted in Figure 5(a). To uncover this structure, they applied several statistical filters to the point cloud (called P hereafter):



(a) The 3-circles model

(b) Barcode obtained with $k = 1,200$ and $x = 30\%$

(c) Barcode obtained with $k = 24,000$ and $x = 30\%$

Figure 5: Experimental results from [12] (images courtesy of G. Carlsson and V. de Silva).

1. Since the focus was primarily on high-density areas of the data, they thresholded P by density using the k-th nearest neighbor density estimator, keeping only the fraction x of the data points with lowest k-th nearest neighbor distance. Varying k from $1,200$ to $24,000$ and x from 10% to 30%, they obtained a collection $P_{k,x}$ of high-density subsets of P.

2. Considering each set $P_{k,x}$ independently, they selected a subset $W_{k,x}$ of $5,000$ random points sampled uniformly, which they took as their witness set.

3. Among the points of $W_{k,x}$ they selected a subset $L_{k,x}$ of 50 landmarks by a furthest-point sampling strategy.

After this preprocessing phase, they computed the barcode of the witness complex filtration for each pair $(L_{k,x}, W_{k,x})$ independently. Their results in 1-homology show two trends :

- For smaller values of k, the barcodes reveal 5 long bars—see e.g. Figure 5(b)—corresponding to the homology of the 3-circle model of Figure 5(a).
- For larger values of k, the barcodes reveal only 1 long bar—see e.g. Figure 5(c), suggesting that one of the circles is prevailing over the other two, which once again corroborates the 3-circles model as depicted in Figure 5(a).

In [6], the authors also conjectured that the 3 circles might actually lie on a Klein bottle. Their intuition came from the observation that the 3 circles can be combined into a natural parametrization of the space of high-contrast 3×3 patches that has the topology of a Klein bottle. Unfortunately, as intellectually pleasing as this hypothesis may be, they were not able to verify it beyond any doubt in practice, as it took them a fair amount of tweaking to make the Klein bottle *appear* in their barcodes. One possible explanation is that they were using too few landmarks in an attempt to keep the size of the witness complex filtration from becoming intractable. The question of the size and choice of the landmarks set has an impact on the validity of their experimental results as a whole, since after all, there are not so many shapes out there that can be faithfully sampled with only 50 points. Thus, even their results in 1-homology call for further experimental validation.

Our Rips zigzags are particularly relevant for this validation. Using the same preprocessed data, we took the whole set $W_{k,x}$ as landmarks and computed the persistence (up to 7 dimensions) of their associated Morozov zigzags. We subsampled the data down to 500 points to compute the diagrams for dimensions 4 and up. This took less than an hour per data set and used less than 2 GB of RAM. The barcodes obtained for $W_{k,x}$ with $(k = 1,200,\ x = 30)$ and $(k = 24,000,\ x = 30)$ are shown in Figures 6 and 7 respectively. While they do corroborate the results of Figure 5, they show a clear absence of any higher-dimensional topological structure. This further strengthens the

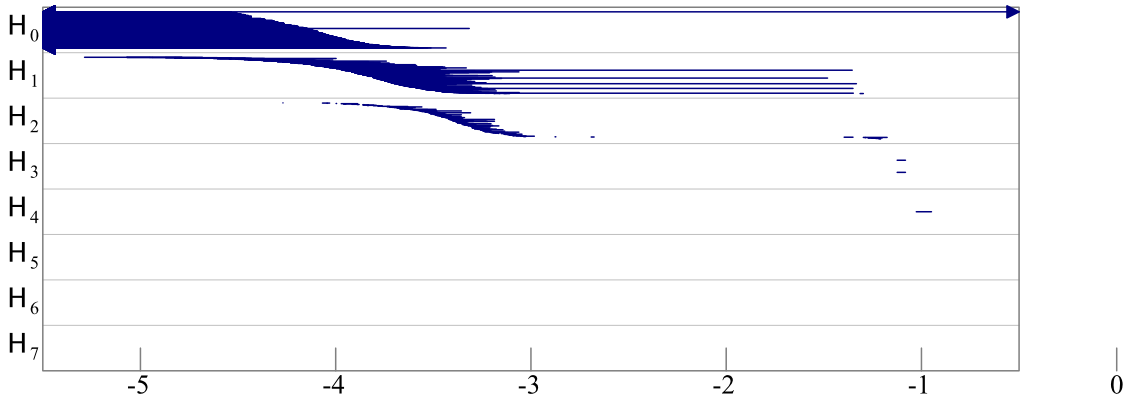


Figure 6: Barcode of the M-ZZ of $W_{k,x}$ with $k = 1,200$ and $x = 30$.

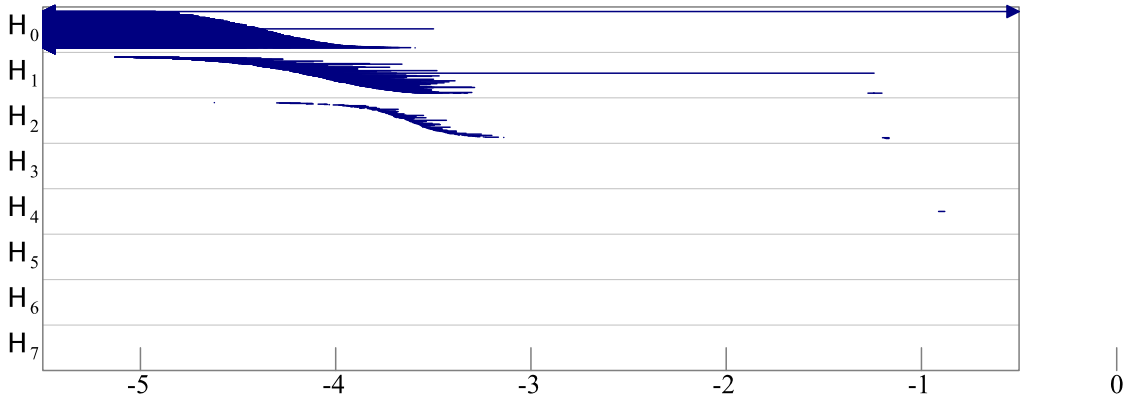


Figure 7: Barcode of the M-ZZ of $W_{k,x}$ with $k = 24,000$ and $x = 30$.

3-circles model hypothesis, while weakening the Klein bottle model hypothesis. It does not discard the latter entirely though, since our results are subject to the initial density filtering step, which might be too rough in its current form. A more subtle statistical preprocessing of the data might change the situation[5], and so the question of the presence (or not) of the Klein bottle in this data set is left open.

7 Discussion

In this paper, we explored several Rips zigzags that achieve both small size and bounded topological noise for homology inference. While our size bounds are comparable to those obtained in previous work, our bounds on the noise are clearly better since they guarantee the noise is either totally absent or just ephemeral. Our proofs rely on general new techniques for manipulating and comparing zigzag modules. We hope that these techniques will find further use and stimulate new research and applications of zigzag persistence. For now we provide an in-depth discussion of the potential and limitations of our results.

[5]This idea was suggested in [6], where it led to further tweaking of the data.

Choosing among the Rips zigzags in practice. Each one of our zigzags has its own strengths and weaknesses, whether in terms of output quality or in terms of computing efficiency. Table 2 summarizes their theoretical behavior with respect to the key aspects of our analysis: length of the sweet range, amplitude of the topological noise, memory usage and running time versus the number of input points.

zigzag	sweet range	topological noise	size vs input	runtime vs input
iR-ZZ	widest	none	linear	linear
oR-ZZ	widest	ephemeral	linear	quadratic
dM-ZZ	close to widest	ephemeral on geometric scale	linear	linear
M-ZZ	close to widest	potentially large	linear	linear

Table 2: Theoretical behaviour of the Rips zigzags of Section 4.

This table gives a clear advantage to the iR-ZZ and should encourage the user to prefer it over the other zigzags in practice. However, it is somewhat misleading because only the asymptotic memory usage and running time bounds are given. In fact, maintaining two Morozov zigzags requires up to twice the amount of memory required to maintain just one, which makes a huge difference on some datasets, including in terms of actual running time. Moreover, the iR-ZZ requires an implementation of the zigzag persistence algorithm for images, a version that is likely to be less efficient than its non-image counterpart. For instance, in our experiments on the Clifford data set the iR-ZZ was ten times slower to compute than the M-ZZ.

Among the other zigzags, the oR-ZZ clearly gives the best results in terms of quality, both in theory and in practice. Unfortunately, its computation turns out to be much slower than those of the other zigzags. In our experiments the slow-down factor compared to the M-ZZ went up to thousands.

The discretized Morozov zigzag achieves a good trade-off between running time and output quality. In our experiments the running times were comparable to that of the M-ZZ (only slower by a factor of 2 at most), while the output barcodes contained much less topological noise. A shortcoming though is that large discretization steps tend to over-simplify the barcodes, sometimes treating some relevant yet delicate topological features as noise and removing the corresponding bars in the output. Such issues occur less often with the other zigzags, which insert the data points one by one instead of in batches.

Thus, as a general rule of thumb, we suggest to use the Morozov zigzag first in practice when dealing with a new data set. This is especially true if the data are supposed to be sampled from "simple" shapes (e.g. ones with positive reach or μ-reach), in which case the quality of the output should be comparable to that achieved with the other zigzags (albeit with a smaller sweet range). Then, in cases where the quality of the result is not sufficient and one needs to obtain cleaner (less noisy) barcodes, either one of the three variants may be considered: dM-ZZ, oR-ZZ, or iR-ZZ. Although we did prefer the dM-ZZ for its good trade-off in general, in some specific scenarios the iR-ZZ or oR-ZZ was definitely the better choice.

Beyond Euclidean spaces. The geometric part of our analysis of the Rips zigzags of Section 4 assumes the vertex set is sitting in some Euclidean space, Hausdorff-close to some compact set. This hypothesis is made for the sake of convenience as it allows us to refer to the sampling theory for compact sets developped by Chazal and co-authors [7]. However, it does not reflect the variety of scenarios in which Rips zigzags can be used.

Since their construction only requires a matrix of pairwise distances as input, Rips zigzags

are applicable in any metric space. Although they come with no theoretical guarantees in such a generality, there are many contexts in which some things can be said about their persistence barcodes. Finite samples of a compact subset of $\mathbb{R}^d$ is but one example. Another important example is when P is sampled from a compact Riemannian manifold, or more generally from a compact Alexandrov space with curvature bounded from above, or even more generally from any compact length space X with positive convexity radius — see e.g. [3] for an introduction to these spaces. It is beyond the scope of this paper to redo our analysis in this context, however for completeness we provide high-level directions on how to adapt it:

- Compact spaces with positive convexity radius admit finite covers with (small enough) convex metric balls, so the Nerve Lemma and its persistent variant hold. One can then reproduce the results of Section 2.2, and in particular Theorem 2.5, with X being the whole space instead of some compact subset, and with the convexity radius of X playing the role formerly played by the weak feature size[6]. This settles most of the geometric aspects of the analysis of Section 4.
- The one part of the analysis that still remains to be adapted is when the results by Attali, Lieutier and Salinas [1] are invoked to bound the amplitude of the noise in the barcode of the Morozov zigzag. These results require P to be sitting close to some compact set with positive μ-reach in $\mathbb{R}^d$. However, it turns out that they follow previous work by Hausmann [21] and Latschev [24], which focuses precisely on cases where P lies on Riemannian manifolds or more general length spaces, and whose results apply directly to the present context.
- Once these geometric aspects have been adapted, the rest of the analysis can be reproduced and theorems similar to the ones of Section 4 can be derived for the oR-ZZ, M-ZZ, dM-ZZ, and iR-ZZ. While the lower bounds of the sweet ranges still depend on the sampling density parameter ε, the upper bounds now depend on the convexity radius rather than on the weak feature size.

Zigzags manipulations and functoriality. The topological part of our analysis of Rips zigzags relies on manipulating zigzags via arrow reversals or space removals, and of paramount importance is the fact that such manipulations can be performed while preserving (most of) the internal structure of the zigzags as well as the commutativity of the diagrams. In a sense, one can see the Arrow Reversal and Space Removal theorems as providing operators from the class of zigzags of a certain type τ to the class of zigzags of another type τ^*. In the arrow reversal case, τ and τ^* differ by a single arrow orientation, while in the space removal case they differ by a single space. An important question is to understand when and why these operators would preserve commutativity, which can be cast mathematically into the question of whether these operators can be made functorial.

To formalize this question, we use the language of quiver theory, in which a module type τ is called a *quiver* and is represented as a directed graph whose undirected version is a simple path, as in the following example.

$$\bullet \longrightarrow \bullet \longleftarrow \bullet \longrightarrow \bullet \longrightarrow \bullet$$

A zigzag module $\mathbb{V}$ of type τ is called a *representation* of τ. The class of all representations of τ, together with the module homomorphisms connecting them, forms a category denoted Rep_τ. In the case of arrow reversals, we are given another module type τ^* that differs from τ by a single arrow orientation, and we ask ourselves whether there exist functors $\mathrm{F} : \mathrm{Rep}_\tau \to \mathrm{Rep}_{\tau^*}$ that preserve the interval decompositions of τ's representations, i.e.,

$$\forall \mathbb{V} \in \mathrm{Rep}_\tau,\ \mathsf{Pers}(\mathbb{V}) = \mathsf{Pers}(\mathrm{F}(\mathbb{V})). \tag{13}$$

[6]The analysis is even simpler in this case since Čech complexes of a suitable parameter carry the same homological information as X, as opposed to the images of inclusions between Čech complexes.

Beyond the question of the existence of such functors (which remains a mystery to us at present) lies that of their intrinsic properties. Considering that the arrow reversal operator is naturally *reversible* (applying it twice to a module $\mathbb{V}$ gives either $\mathbb{V}$ itself or some module isomorphic to it, depending on the choice of isomorphism in the Interval Decomposition Theorem 2.6), it is desireable that the functor F be also *reversible* in some sense. There exist many notions of reversibility for functors in the theory of categories, and here we check F against the following classical notions: (a) isomorphism of categories, (b) equivalence of categories, (c) left or right adjoint. There is a natural gradation between these concepts: isomorphisms are equivalences and equivalences are adjoint.

We show in Appendix A that, given τ, τ^* as above, there is no covariant functor $\mathrm{F} : \mathrm{Rep}_\tau \to \mathrm{Rep}_{\tau^*}$ of type (a), (b) or (c) that satisfies (13). In other words, there is no hope to find *reversible* covariant arrow-reversing functors that preserve persistence diagrams, at least for the aforementioned classical notions of functor reversibility. We also show that a similar conclusion holds for contravariant functors, under the extra hypothesis that the length of τ, τ^* (i.e. the number of nodes in the graph) is at least 3 and not just 2. Indeed, for module types of length 2, plain vector space duality induces a contravariant functor that is an isomorphism of categories.

These negative results regarding the functoriality of arrow reversals are a pretty bad piece of news. And unfortunately, claims of a similar flavor can be made for space removals as well[7]. This being said, we have only scratched the surface of the problem so far. For instance, we still do not know if functors satisfying (13) exist. Further investigations should be conducted to assess the real potential of the arrow reversal and space removal operators, to derive more canonical variants, and to make connections with other operations on quivers such as for instance reflection and Coxeter functors [2].

A Functoriality

To prove that no functor of type (c) exists—and thus no functor of type (a) or (b) either, we use the fact that left (resp. right) adjoint functors are right (resp. left) exact, i.e., they preserve right (resp. left) exact sequences [23, §3.3.6].

Lemma A.1. *Let τ, τ^* be module types of length at least 2 that differ by a single arrow orientation. Then, there is no covariant functor $\mathrm{F} : \mathrm{Rep}_\tau \to \mathrm{Rep}_{\tau^*}$ satisfying (13) that is either left or right exact.*

Proof. By the restriction principle we only need to focus on module types of length 2. Let then τ be the quiver $\bullet \longrightarrow \bullet$ and τ^* the quiver $\bullet \longleftarrow \bullet$. Suppose that there exists a covariant functor $\mathrm{F} : \mathrm{Rep}_\tau \to \mathrm{Rep}_{\tau^*}$ that satisfies (13). Then, consider the following sequence of module homomorphisms (drawn horizontally) connecting various representations of τ (drawn vertically upwards), where $\boldsymbol{k}$ denotes the field of coefficients:

$$\begin{array}{ccccccccc} 0 & \xrightarrow{0} & \boldsymbol{k} & \xrightarrow{\mathbb{1}} & \boldsymbol{k} & \xrightarrow{0} & 0 & \xrightarrow{0} & 0 \\ \uparrow 0 & & \uparrow 0 & & \uparrow \mathbb{1} & & \uparrow 0 & & \uparrow 0 \\ 0 & \xrightarrow{0} & 0 & \xrightarrow{0} & \boldsymbol{k} & \xrightarrow{\mathbb{1}} & \boldsymbol{k} & \xrightarrow{0} & 0 \end{array}$$

It is easily seen that this sequence is exact. By (13) it is sent to a sequence of the following form

[7] With the notable exception of compositions as in Lemma 3.3, which canonically induce a functor.

through F, where the spaces A, B, C, D are isomorphic to $\boldsymbol{k}$:

$$\begin{array}{ccccccccc} 0 & \xrightarrow{0} & A & \xrightarrow{0} & C & \xrightarrow{0} & 0 & \xrightarrow{0} & 0 \\ \downarrow 0 & & \downarrow 0 & & \downarrow \cong & & \downarrow 0 & & \downarrow 0 \\ 0 & \xrightarrow{0} & 0 & \xrightarrow{0} & B & \xrightarrow{0} & D & \xrightarrow{0} & 0 \end{array}$$

Note that the map $A \longrightarrow C$ has to be zero for the quadrant $A, C, B, 0$ to commute. Idem, the map $B \longrightarrow D$ has to be zero for the quadrant $C, 0, D, B$ to commute. Thus, this new sequence is neither left nor right exact. □

Lemma A.2. *Let τ, τ^* be module types of length at least 3 that differ by a single arrow orientation. Then, there is no contravariant functor $\mathrm{F} : \mathrm{Rep}_\tau \to \mathrm{Rep}_{\tau^*}$ satisfying (13) that is either left or right exact.*

Proof. By the restriction principle we only need to focus on module types of length 3. Let then τ be the quiver $\bullet \longrightarrow \bullet \longleftarrow \bullet$ and τ^* the quiver $\bullet \longrightarrow \bullet \longrightarrow \bullet$, the other option ($\tau = \bullet \longleftarrow \bullet \longrightarrow \bullet$ and $\tau^* = \bullet \longrightarrow \bullet \longrightarrow \bullet$) being similar and left as an exercise. Suppose that there exists a contravariant functor $\mathrm{F} : \mathrm{Rep}_\tau \to \mathrm{Rep}_{\tau^*}$ that satisfies (13). Then, consider the following sequence of module homomorphisms (drawn horizontally) connecting various representations of τ (drawn vertically upwards), where $\boldsymbol{k}$ denotes the field of coefficients:

$$\begin{array}{ccccccccc} 0 & \xrightarrow{0} & 0 & \xrightarrow{0} & \boldsymbol{k} & \xrightarrow{\mathbb{1}} & \boldsymbol{k} & \xrightarrow{0} & 0 \\ \downarrow 0 & & \downarrow 0 & & \downarrow \mathbb{1} & & \downarrow 0 & & \downarrow 0 \\ 0 & \xrightarrow{0} & \boldsymbol{k} & \xrightarrow{\mathbb{1}} & \boldsymbol{k} & \xrightarrow{0} & 0 & \xrightarrow{0} & 0 \\ \uparrow 0 & & \uparrow 0 & & \uparrow \mathbb{1} & & \uparrow 0 & & \uparrow 0 \\ 0 & \xrightarrow{0} & 0 & \xrightarrow{0} & \boldsymbol{k} & \xrightarrow{\mathbb{1}} & \boldsymbol{k} & \xrightarrow{0} & 0 \end{array}$$

It is easily seen that this sequence is exact. By (13) it is sent to a sequence of the following form through F, where the spaces A, B, C, D, E, F are isomorphic to $\boldsymbol{k}$:

$$\begin{array}{ccccccccc} 0 & \xleftarrow{0} & 0 & \xleftarrow{0} & D & \longleftarrow & F & \xleftarrow{0} & 0 \\ \uparrow 0 & & \uparrow 0 & & \uparrow \cong & & \uparrow 0 & & \uparrow 0 \\ 0 & \xleftarrow{0} & A & \xleftarrow{0} & C & \xleftarrow{0} & 0 & \xleftarrow{0} & 0 \\ \uparrow 0 & & \uparrow 0 & & \uparrow \cong & & \uparrow 0 & & \uparrow 0 \\ 0 & \xleftarrow{0} & 0 & \xleftarrow{0} & B & \xleftarrow{0} & E & \xleftarrow{0} & 0 \end{array}$$

Note that the map $A \longleftarrow C$ has to be zero for the quadrant $A, C, B, 0$ to commute. Idem, the map $B \longleftarrow E$ has to be zero for the quadrant $C, 0, E, B$ to commute. Thus, this new sequence is neither left nor right exact. □

Acknowledgements

The authors wish to thank the referees for their careful reading of the paper and their insightful comments. They also thank Marc Glisse for suggesting the approach used in the proof of Theorem 4.3, Frédéric Chazal and Vin de Silva for helpful discussions, and Dmitriy Morozov for accepting to integrate their code into the Dionysus library. This work was supported by the European project CG-Learning No. 255827 and by the ANR project TopData.

References

[1] D. Attali, A. Lieutier, and D. Salinas. Vietoris-Rips complexes also provide topologically correct reconstructions of sampled shapes. *Computational Geometry: Theory and Applications*, 46(4):448–465, 2012.

[2] I. N. Bernstein, I. M. Gelfand, and V. A. Ponomarev. Coxeter functors and Gabriel's theorem. *Russian Mathematical Surveys*, 28(2):17–32, 1973.

[3] D. Burago, Y. Burago, and S. Ivanov. *A Course in Metric Geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001.

[4] G. Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009.

[5] G. Carlsson and V. de Silva. Zigzag persistence. *Foundations of Computational Mathematics*, 10(4):367–405, 2010.

[6] G. Carlsson, T. Ishkhanov, V. de Silva, and A. Zomorodian. On the local behavior of spaces of natural images. *Int. J. Comput. Vision*, 76(1):1–12, January 2008.

[7] F. Chazal and D. Cohen-Steiner. Geometric inference. In *Tesselations in the Sciences*. Springer-Verlag, 2013.

[8] F. Chazal, D. Cohen-Steiner, and A. Lieutier. A sampling theory for compact sets in Euclidean space. *Discrete & Computational Geometry*, 41:461–479, 2009.

[9] F. Chazal, L. J. Guibas, S. Y. Oudot, and P. Skraba. Analysis of scalar fields over point cloud data. *Discrete and Computational Geometry*, 46(4):743–775, December 2011.

[10] F. Chazal and S. Y. Oudot. Towards persistence-based reconstruction in Euclidean spaces. In *Proc. 24th ACM Symposium on Computational Geometry*, pages 232–241, 2008.

[11] V. de Silva. A weak characterisation of the Delaunay triangulation. *Geometriae Dedicata*, 135(1):39–64, August 2008.

[12] V. de Silva and G. Carlsson. Topological estimation using witness complexes. *IEEE Symposium on Point-based Graphic*, pages 157–166, 2004.

[13] V. de Silva and R. Ghrist. Coverage in sensor networks via persistent homology. *Algebraic and Geometric Topology*, 7:339–358, 2007.

[14] T. K. Dey, F. Fan, and Y. Wang. Graph induced complex on point data. In *Proc. 29th Annual Symposium on Computational Geometry*, pages 107–116, New York, NY, USA, 2013. ACM.

[15] Tamal K. Dey, Fengtao Fan, and Yusu Wang. Computing topological persistence for simplicial maps. In *Proc. 30th Annual Symposium on Computational Geometry*, 2014.

[16] Dionysus. By D. Morozov (`http://www.mrzv.org/software/dionysus/`).

[17] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discrete and Computational Geometry*, 28:511–533, 2002.

[18] P. Gabriel. Unzerlegbare Darstellungen I. *Manuscripta Mathematica*, 6:71–103, 1972.

[19] L. J. Guibas and S. Y. Oudot. Reconstruction using witness complexes. *Discrete and Computational Geometry*, 40(3):325–356, 2008.

[20] A. Hatcher. *Algebraic Topology*. Cambridge Univ. Press, 2001.

[21] J.-C. Hausmann. On the Vietoris-Rips complexes and a cohomology theory for metric spaces. *Prospects in topology, Ann. of Math. Stud.*, 138:175–188, 1995.

[22] B. Hudson, G. Miller, S. Y. Oudot, and D. Sheehy. Topological inference via meshing. In *Proc. ACM Symposium on Computational geometry*, pages 277–286. ACM, 2010.

[23] M. Kashiwara and P. Schapira. *Categories and Sheaves*, volume 332 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, 2006.

[24] J. Latschev. Vietoris-Rips complexes of metric spaces near a closed Riemannian manifold. *Archiv der Mathematik*, 77(6):522–528, 2001.

[25] A. B. Lee, K. S. Pedersen, and D. Mumford. The nonlinear statistics of high-contrast patches in natural images. *Int. J. Comput. Vision*, 54(1-3):83–103, August 2003.

[26] D. Morozov. Personal communication.

[27] D. R. Sheehy. Linear-size approximations to the vietoris-rips filtration. In *Proc. ACM Symposium on Computational Geometry*, pages 239–248. ACM, 2012.

[28] J. H. van Hateren and A. van der Schaaff. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society, London, B*, 265:359–366, 1997.

[29] A. Zomorodian and G. Carlsson. Computing persistent homology. *Discrete Comput. Geom.*, 33(2):249–274, 2005.