Decomposition of Zero-Dimensional Persistence Modules via Rooted Subsets

Ángel Javier Alonso ✉ ⓘ
Technische Universität Graz, Austria

Michael Kerber ✉ ⓘ
Technische Universität Graz, Austria

─── **Abstract** ───

We study the decomposition of zero-dimensional persistence modules, viewed as functors valued in the category of vector spaces factorizing through sets. Instead of working directly at the level of vector spaces, we take a step back and first study the decomposition problem at the level of sets.

This approach allows us to define the combinatorial notion of *rooted subsets*. In the case of a filtered metric space M, rooted subsets relate the clustering behavior of the points of M with the decomposition of the associated persistence module. In particular, we can identify intervals in such a decomposition quickly. In addition, rooted subsets can be understood as a generalization of the elder rule, and are also related to the notion of constant conqueror of Cai, Kim, Mémoli and Wang. As an application, we give a lower bound on the number of intervals that we can expect in the decomposition of zero-dimensional persistence modules of a density-Rips filtration in Euclidean space: in the limit, and under very general circumstances, we can expect that at least 25% of the indecomposable summands are interval modules.

1 Introduction

Multiparameter persistent homology is an active research area in topological data analysis. The motivation is that in many datasets there are multiple parameters that deserve attention in a multiscale analysis [11, 18, 31]. Concretely, when analyzing point clouds, we want to consider the distances between points, but also potentially remove points of low density.

A central object of persistent homology is the *persistence module*, which tracks algebraically how the topological features of the data change as we move through the parameter space. In the single-parameter case, every persistence module decomposes into a collection of intervals, called the *persistence barcode* [20], where each interval represents the lifetime of a topological feature in the data. In the multiparameter setting, there is a generalized notion of interval, which again represents the lifetime of a topological feature, but decomposing a multiparameter persistence module into intervals is not always possible, and one might be left with non-interval indecomposable persistence modules that lead to complications, both theoretically [12, 13, 18, 32] and computationally [1, 4, 23].
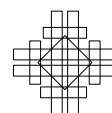
■ **Table 1** Number of intervals in the decomposition of zero-dimensional persistence modules for density-Rips filtrations. We tried both *clustered* samples where the points were sampled by a multivariate Gaussian distribution around 5 peaks, and *uniform* samples in the unit square. The density parameter was computed via a Gaussian kernel density estimate (*kde*) or a *random* density was assigned. The table shows the number of intervals for 5 independent test runs; for n points, the module is interval-decomposable if the number of intervals is n. This only happens for one run.

Sample	Densities	100 points					500 points				
		Run 1	2	3	4	5	Run 1	2	3	4	5
clustered	kde	100	98	95	98	98	474	487	478	479	479
uniform	kde	88	88	86	88	86	444	447	433	453	457
clustered	random	77	86	87	88	76	397	381	390	380	386
uniform	random	76	79	75	75	70	376	361	366	355	377

In fact, the classification of such indecomposable persistence modules is thought to be out of reach: certain involved posets are of *wild representation type*, even when accounting for certain simplifications [2]. Moreover, infinite families of complicated indecomposable persistence modules can be realized by simple geometric constructions [13], and, most recently, it has been shown in [3] that multiparameter persistence modules are, generically, close to being indecomposable, under the interleaving metric (we refer to [3] for a precise statement).

Still, the mentioned complications do not imply that the persistence modules that come up in practice are close to indecomposable, or that they are not decomposable into intervals. Indeed, is the decomposition of multiparameter persistence modules as badly behaved in practice as we can expect in theory? The authors of [2] and those of [3] state similar questions.

As an initial test, we computed the decomposition of persistence modules for a standard zero-dimensional construction, using a prototypical implementation of the algorithm by Dey and Xin [23] (this implementation will be discussed in another paper). As we see in Table 1, the assumption that persistence modules can be decomposed completely into intervals seems to be false most of the time, at least in this setting. However, Table 1 also shows that in all tested instances, *most* indecomposable summands are indeed intervals.

This begs the question whether we can provably expect many intervals in general. In addition, knowledge of the intervals can greatly simplify and speed up computational tasks for persistence modules: for instance, a popular way to analyze 2-parameter persistence modules is by considering 1-dimensional restrictions, so-called slices, resulting in a parameterized family of persistent barcodes [29, 31, 33, 34]. Every interval of the 2-dimensional persistence module gives one bar in the barcode of the slice, by intersecting the slice with the interval. Thus, by knowing the intervals, existing algorithms can focus on the non-interval "core" of the problem, which is typically of much smaller size.

The practical problem of the described approach is that decomposing a multiparameter persistence module is costly, despite ongoing efforts [23]. However, to leverage the knowledge of intervals there is no need to compute a total decomposition, or to even identify all intervals. It suffices to have a method to "peel off" intervals from a persistence module quickly. Thus, we pose the question whether there exist methods that work very fast in practice and still are capable of detecting many intervals.

Contributions. We focus on the case of zero-dimensional persistence modules. Already this case is of practical interest because of its connection to hierarchical clustering methods (see the Related work section below), and has received attention recently [2, 10, 14, 34]. In this context, we give some answers to the questions stated above:

For a point cloud M, a *nearest neighbor pair* is a pair $(x, y) \in M \times M$ such that y is the nearest neighbor of x and x is the nearest neighbor of y (breaking ties with a fixed total order). The theory we develop says that for a zero-dimensional persistence module of the density-Rips bifiltration (for any density estimation function), there are at least as many intervals as there are nearest neighbor pairs in M. These intervals are easily determined by the nearest neighbor pairs, and we refer to them as *NN-intervals*. Since all nearest neighbor pairs can be computed in $O(n \log n)$ time [19, 37], this yields a fast method to compute all NN-intervals of the decomposition. Moreover, we can expect many NN-intervals: using previous results on nearest neighbor graphs, we show that if M is sampled independently from an arbitrary, almost continuous density function, at least a quarter of the summands in the decomposition are intervals as $n \to \infty$. To our knowledge, this is the first result proving a non-constant lower bound on the number of intervals in a decomposition.

To arrive at this result, we use the following main idea: Instead of studying the decomposition of the persistence module directly in the category of (graded) vector spaces, we work in the category of *persistent sets*, whose objects can be interpreted as a two-parameter hierarchical clustering. The decomposition of a persistence module is governed by its idempotent endomorphisms, so we look for idempotent endomorphisms not of persistence modules, but of persistent sets, which are simpler. We show that such idempotent endomorphisms can be translated into *rooted subsets*, which are subsets of points that get consistently merged with a fixed point in the hierarchical clustering. Moreover, rooted subsets with a single element correspond to intervals in the associated persistence module.

Instead of peeling off intervals from the persistence module, we peel off rooted subsets from the persistent set. The advantage is that the remaining structure is still a hierarchical clustering, and the process can be iterated.

Related work.   Multiparameter persistent sets and zero-dimensional persistence modules, as we will study them here, are related to a multiparametric approach to the clustering problem first considered by Carlsson and Mémoli [16]. The need for multiple parameters, density and scale, is justified by an axiomatic approach to clustering [15, 17, 28]. The application of techniques from multiparameter persistence homology, like persistence modules and interleavings, to this setting has attracted attention recently [2, 14, 31, 33, 34].

Cai, Kim, Mémoli, and Wang [14] define a useful summary for zero-dimensional persistence modules coming from density-Rips, called the *elder-rule-staircode*, inspired by the elder rule [24]. They also introduce the related concept of constant conqueror, and they ask whether a constant conqueror induces an interval in the decomposition of the associated persistence module. We answer this question in the negative with Example 22, and, in contrast, we show that a *rooted generator*, as introduced here, does induce an interval in the decomposition (Corollary 13).

Brodzki, Burfitt, and Pirashvili [10] also study the decomposition of zero-dimensional persistence modules. They identify a class of persistence modules, called *semi-component modules*, that may appear as summands in the decomposition of zero-dimensional modules, but that are still hard to classify. Their methods have been of great inspiration, and in Theorem 24 we give another proof, within the theory we develop, of a theorem of theirs.

2   Preliminaries

Persistent sets and persistence modules.   In what follows, we let P be a finite poset, which we will view as a category. A **persistence module** (over P) is a functor from P to the category Vec of finite dimensional vector spaces, over a fixed field K. Such a functor

$F \colon P \to \mathsf{Vec}$ associates to each **grade** $p \in P$ a finite dimensional vector space F_p and to each morphism $p \leq q$ in P a linear map $F_{p \to q} \colon F_p \to F_q$, in such a way that $F_{p \to p} = \mathrm{id}$ and composition is preserved. We see persistence modules as the objects of the functor category Vec^P, where natural transformations are the morphisms. In this sense, a morphism $f \colon F \to G$ of persistence modules is a family of maps $\{f_p \colon F_p \to G_p\}_{p \in P}$ such that for every two $p \leq q$ the following diagram commutes

$$
\begin{array}{ccc}
F_p & \xrightarrow{F_{p \to q}} & F_q \\
\downarrow{\scriptstyle f_p} & & \downarrow{\scriptstyle f_q} \\
G_p & \xrightarrow{G_{p \to q}} & G_q.
\end{array}
$$

Similarly, a **persistent set** (over P) is a functor from P to Set, the category Set of finite sets, and morphisms of persistent sets are natural transformations as above.

We can obtain a persistence module from a persistent set by the application of the **linearization functor** $\mathsf{Set} \to \mathsf{Vec}$ that takes each set to the free vector space generated by it. This linearization functor induces a functor $\mathcal{L} \colon \mathsf{Set}^P \to \mathsf{Vec}^P$ by postcomposition.

From geometry to persistent sets.   Let (M, d) be a finite metric space, and consider a function $f \colon M \to \mathbb{R}$. We can understand f as an assignment of a *density* to each of the points of M; that is, a density estimation function [36]. We assume that f assigns lower values to points of *higher* density. Following [14], we call the triple (M, d, f) an **augmented metric space**. We construct a persistent set, the **density-Rips persistent set** of (M, d, f), that tracks how the clustering of points of M changes as we change the density and scale parameters, in a sense that we make precise shortly.

First, for a fixed scale parameter $\varepsilon \geq 0$, we define the **geometric graph of M at ε**, denoted by $\mathcal{G}_\varepsilon(M)$, as the undirected graph on the vertex set M and edges (x, y) where $d(x, y) \leq \varepsilon$. The connected components of $\mathcal{G}_\varepsilon(M)$, as ε goes from 0 to ∞, form the clusters of the dendrogram obtained via the single-linkage clustering method.

To introduce the density, for each $\sigma \in \mathbb{R}$ we let $M_\sigma \coloneqq \{x \in M \mid f(x) \leq \sigma\} \subseteq M$ be the metric subspace of points with (co)density below σ. For any two $\sigma \leq \sigma'$, $M_\sigma \subseteq M_{\sigma'}$ and by taking each (ε, σ) to the graph $\mathcal{G}_\varepsilon(M_\sigma)$, we obtain a functor $\mathcal{G}(M, f) \colon \mathbb{R}_{\geq 0} \times \mathbb{R} \to \mathsf{Graph}$, where the order in $\mathbb{R}_{\geq 0} \times \mathbb{R}$ is given by $(\varepsilon, \sigma) \leq (\varepsilon', \sigma')$ if and only if $\varepsilon \leq \varepsilon'$ and $\sigma \leq \sigma'$. We then consider the **connected components functor** $\pi_0 \colon \mathsf{Graph} \to \mathsf{Set}$, that takes each graph to its set of connected components. In this way, we obtain a functor $\pi_0 \circ \mathcal{G}(M, f) \colon \mathbb{R}_{\geq 0} \times \mathbb{R} \to \mathsf{Set}$.

▶ Remark 1. The linearized persistence module $\mathcal{L}(\pi_0 \circ \mathcal{G}(M, f)) \colon \mathbb{R}_{\geq 0} \times \mathbb{R} \to \mathsf{Vec}$ is isomorphic to the persistence module obtained by applying zero-dimensional homology at graph level, $H_0 \circ \mathcal{G}(M, f) \colon \mathbb{R}_{\geq 0} \times \mathbb{R} \to \mathsf{Vec}$. In this sense, the construction we have described is the zero-dimensional level of the density-Rips filtration, which is standard in multiparameter persistent homology (see [5, 18] and also [14]).

We can understand the functor $\pi_0 \circ \mathcal{G}(M, f) \colon \mathbb{R}_{\geq 0} \times \mathbb{R} \to \mathsf{Set}$ as a persistent set $S \colon P \to \mathsf{Set}$ indexed by a finite grid $P \subseteq \mathbb{R}_{\geq 0} \times \mathbb{R}$ in the following way. We consider the set of distances $D \coloneqq \{d(x, y) \mid x, y \in M\}$ and densities $T \coloneqq \{f(x) \mid x \in M\}$, and define a finite grid $P \coloneqq D \times T \subset \mathbb{R}_{\geq 0} \times \mathbb{R}$. Finally, we define the persistent set $S \colon P \to \mathsf{Set}$ by taking each $(\varepsilon, \sigma) \in P$ to $(\pi_0 \circ \mathcal{G}(M, f))_{(\varepsilon, \sigma)}$, and similarly for the morphisms.

▶ **Definition 2.** *Let (M, d, f) be an augmented metric space. We define its **density-Rips persistent set** as the functor $S \colon P \to \mathsf{Set}$, constructed as above.*

Decomposition of persistence modules. We can study persistence modules via their decomposition. For two persistence modules F and G their direct sum $F \oplus G$ is the persistence module given by taking direct sums pointwise, $(F \oplus G)_p = F_p \oplus G_p$. A persistence module is **indecomposable** if $F \cong F_1 \oplus F_2$ implies that either $F_1 = 0$ or $F_2 = 0$. Since persistence modules are actual modules (see, for instance, [8, Lemma 2.1]), by the Krull-Schmidt theorem, a decomposition of a persistence module $F = F_1 \oplus F_2 \oplus \ldots F_n$ into indecomposable summands is unique up to permutation and isomorphism of the summands.

Let I be a non-empty connected subposet of a poset P such that for any two $i, j \in I$ and any $l \in P$, if $i \le l \le j$ then $l \in I$. The **interval module supported on** I, $\mathcal{I}(I) \colon P \to \mathsf{Vec}$, is the indecomposable (by, e.g. [7, Proposition 2.2]) persistence module given by

$$\mathcal{I}(I)_p = \begin{cases} K, & \text{if } p \in I, \\ 0, & \text{otherwise,} \end{cases} \quad \text{with internal maps} \quad \mathcal{I}(I)_{p \to q} = \begin{cases} \text{id}, & \text{if } p, q \in I, \\ 0, & \text{otherwise.} \end{cases}$$

If P is a totally ordered set, every persistence module over P decomposes as a direct sum of interval modules [6], but such a nice decomposition does not exist in general for other posets.

Decomposition and endomorphisms. A direct sum $X = X_1 \oplus X_2$ of persistence modules is characterized up to isomorphism by morphisms $\iota_i \colon X_i \to X$ and $\pi_i \colon X \to X_i$ for $i = 1, 2$ such that $\pi_i \circ \iota_i = \text{id}_{X_i}$ and $\iota_1 \circ \pi_1 + \iota_2 \circ \pi_2 = \text{id}_X$ (see, for instance, [30]). In this case, for each $i = 1, 2$, the maps ι_i and π_i induce an endomorphism $\varphi_i := \iota_i \circ \pi_i$ of X. Such an endomorphism $\varphi_i \colon X \xrightarrow{\pi_i} X_i \xrightarrow{\iota_i} X$ is also split:

▶ **Definition 3.** *In any category, we say that an endomorphism $\varphi \colon X \to X$ is **split** if there exists an object Y and a factorization $\varphi \colon X \xrightarrow{\pi} Y \xrightarrow{\iota} X$ such that $\pi \circ \iota = \text{id}_Y$.*

We will use the following standard fact about split endomorphisms (proof in the full version):

▶ **Lemma 4.** *Let $\varphi \colon X \to X$ be a split endomorphism that has two factorizations $X \xrightarrow{\pi} Y \xrightarrow{\iota} X$ and $X \xrightarrow{\pi'} Y' \xrightarrow{\iota'} X$ with $\pi \circ \iota = \text{id}_Y$ and $\pi' \circ \iota' = \text{id}_{Y'}$. Then Y and Y' are isomorphic.*

Every split endomorphism $\varphi \colon X \xrightarrow{\pi} Y \xrightarrow{\iota} X$ is also **idempotent**, meaning that $\varphi \circ \varphi = \varphi$. Moreover, in our categories of interest, namely persistent sets Set^P and persistence modules Vec^P, every idempotent endomorphism splits through its **image**, see below. In these two categories, we define the image of a morphism f, $\text{img} f$, by taking the image pointwise, that is, $(\text{img} f)_p = f_p(S_p)$. The following two lemmas are standard (proof in the full version).

▶ **Lemma 5.** *Let $\varphi \colon X \to X$ be an idempotent endomorphism in Vec^P or Set^P. Then f splits through its image: there exists a factorization $f \colon X \xrightarrow{\pi} \text{img} \varphi \xrightarrow{\iota} X$ with $\pi \circ \iota = \text{id}_{\text{img} \varphi}$.*

▶ **Lemma 6.** *Let $F \colon P \to \mathsf{Vec}$ be a persistence module, and let $\varphi \colon F \to F$ be an idempotent endomorphism. Then F decomposes as $\text{img}(\text{id}_F - \varphi) \oplus \text{img} \varphi$.*

3 Endomorphisms of persistent sets and rooted subsets

As seen above, the decomposition of a persistence module is intimately related to its idempotent endomorphisms. Our main idea is that, when studying the decomposition of persistence modules of the form $\mathcal{L}S$, for a persistent set $S \colon P \to \mathsf{Set}$, we look for idempotent endomorphisms of S and study their image under the linearization functor $\mathcal{L}$.

▶ **Definition 7.** *Given a persistent set S, a **generator** is a pair (p_x, x) with $x \in S_{p_x}$ such that x is not in the image of any morphism $S_{q \to p_x}$ for any $q < p_x$. When it is clear, we will often suppress the grade p_x from the notation, and directly write that $x \in S_{p_x}$ is a generator.*

There is an induced preorder on the generators of S: for two generators $x \in S_{p_x}$ and $y \in S_{p_y}$ we say that $(p_x, x) \leq (p_y, y)$ if and only if $p_x \leq p_y$. This relation might not be antisymmetric, and so in general the preordered set of generators is not a poset.

Generators are useful because an endomorphism φ of a persistent set $S \colon P \to \mathsf{Set}$ is uniquely determined by the image of its generators: for each $z \in S_q$ we have $\varphi_q(z) = S_{p_x \to q} \circ \varphi_{p_x}(x)$ for some generator $x \in S_{p_x}$, by the commutativity property.

In linear algebra, an idempotent endomorphism can be thought as a projection onto its image, that is, onto its fixed points. This point of view and the concept of generators above motivates the following definition, which plays a fundamental role in our work.

▶ **Definition 8.** *A **rooted subset** A is a non-empty subset of the generators of S such that there exists an idempotent endomorphism φ of S whose set of generators that are not fixed is precisely A. If a rooted subset is a singleton, $A = \{x\}$, we say that x is a **rooted generator**.*

▶ Remark 9. In the case of an augmented metric space (M, d, f) and its density-Rips persistent set S of Definition 2 there exists a bijection between the points of M and the generators of S. A point $x \in M$ first appears in the graph $\mathcal{G}_0(M_{f(x)})$, where x is always its own connected component. In what follows, we will often identify a point $x \in M$ with its generator $x \in S_{p_x}$. In this sense, we can understand an endomorphism of S as an endomorphism of the set of points that is compatible with the connected components of all graphs $\mathcal{G}_\varepsilon(M_\sigma)$.

We are especially interested in persistent sets obtained from (augmented) metric spaces, and our objective is to relate rooted generators to the geometry of these objects. Considering an augmented metric space (M, d, f) and its density-Rips persistent set, Proposition 10 below characterizes rooted generators by the clustering behavior of the points of M.

▶ **Proposition 10.** *Let (M, d, f) be an augmented metric space and consider a point $x \in M$. If there exist some other point $y \in M$ such that*
1. *$f(y) \leq f(x)$ (i.e. y is "denser" than x), and*
2. *whenever x is in a cluster of more than one point, $y \in M$ is in the same cluster: for every $\mathcal{G}_\varepsilon(M_\sigma)$, if x is path-connected to some other point then x is path-connected to y,*
then the generator (p_x, x) of the density-Rips persistent set S of M is a rooted generator.

Conversely, if x is a rooted generator of S, then there exists a point $y \in M$ that satisfies conditions 1 and 2 above.

Proof. Before going into the proof, recall that, by the way we construct S and the inclusion $P \hookrightarrow \mathbb{R}_{\geq 0} \times \mathbb{R}$, for each $q \in P$ there is an associated graph $\mathcal{G}_\varepsilon(M_\sigma)$, for some $(\varepsilon, \sigma) \in \mathbb{R}_{\geq 0} \times \mathbb{R}$. Each element $z \in S_q$ is a connected component of this graph $\mathcal{G}_\varepsilon(M_\sigma)$, and the generators $x \in S_{p_x}$ such that $S_{p_x \to q}(x) = z$ are precisely the points in that connected component.

The first part follows from Proposition 11 below, which proves it in more generality.

For the converse, let φ be an idempotent of S whose only generator that is not fixed is $x \in S_{p_x}$. This means that there exists a generator $y \in S_{p_y}$, different from x, such that $\varphi_{p_x}(x) = S_{p_y \to p_x}(y)$. And clearly $\varphi_{p_z}(z) = z$ for any other generator $z \in S_{p_z}$. From the fact that $\varphi_{p_x}(x) = S_{p_y \to p_x}(y)$ we deduce that $f(y) \leq f(x)$, since $p_y \leq p_x$ in P. To see that the second condition holds, pick a $q \geq p_x$ and suppose that there exists a generator $w \in S_{p_w}$ such that $S_{p_x \to q}(x) = S_{p_w \to q}(w)$. This means that in the graph $\mathcal{G}_\varepsilon(M_\sigma)$ associated to q both x and w are in the same connected component, and we claim that y is also in this component. Indeed, by the definition of φ we have $S_{p_x \to q} \circ \varphi_{p_x}(x) = S_{p_y \to q}(y) = S_{p_w \to q}(w)$. ◀

▶ **Proposition 11.** *Let (M, d, f) be an augmented metric space. If for a set of points $A \subset M$ there exists a point $y \notin A$ such that for every $x \in A$, $f(y) \leq f(x)$ and for each $\mathcal{G}_\varepsilon(M_\sigma)$ either:*

- *x is path-connected to y, or*
- *the set of points that are path-connected to x is contained in A,*

then the set of generators $\{(p_x, x) \mid x \in A\}$ is a rooted subset in the density-Rips persistent set S of M.

Proof. To show that A is a rooted subset, we need to define an appropriate idempotent φ of S. Recalling that an endomorphism is uniquely determined by the image of its generators, we define φ by setting

$$\varphi_{p_x}(x) = \begin{cases} S_{p_y \to p_x}(y), & \text{if } x \in A, \\ x, & \text{otherwise,} \end{cases} \tag{1}$$

for every generator $x \in S_{p_x}$. We need to show that φ is indeed well-defined, which means that the image of $z \in S_q$, $\varphi_q(z) = S_{p_x \to q} \circ \varphi_{p_x}(x)$, is the same no matter the generator $x \in S_{p_x}$ we choose. Fix a $q \in P$ and a $z \in S_q$, and consider the set G of generators whose image in S_q is z, $G \coloneqq \{(p_x, x) \mid p_x \leq q, \ S_{p_x \to q}(x) = z\}$.

Then, to check that φ is well-defined, for every two $(p_x, x), (p_w, w) \in G$ it must hold that

$$S_{p_x \to q} \circ \varphi_{p_x}(x) = S_{p_w \to q} \circ \varphi_{p_w}(w). \tag{2}$$

If both (p_x, x) and (p_w, w) are not in A, or if both (p_x, x) and (p_w, w) are in A, then Equation (2) above trivially holds, by the way we have defined φ in Equation (1).

Thus, the only interesting case is that only one of (p_x, x) or (p_w, w) is in A. Say that $(p_x, x) \in A$ and $(p_w, w) \notin A$. Then, by assumption both x and w need to be path-connected to y at the graph $\mathcal{G}_\varepsilon(M_\sigma)$ associated to q, which means that, as desired,

$$S_{p_x \to q} \circ \varphi_{p_x}(x) = S_{p_y \to q}(y) = S_{p_w \to q}(w) = S_{p_w \to q} \circ \varphi_{p_w}(w).$$

Now, φ is idempotent, because for every $x \in A$ we have $\varphi_{p_x}^2(x) = \varphi_{p_x}(S_{p_y \to p_x}(y)) = S_{p_y \to p_x}(\varphi_{p_y}(y)) = S_{p_y \to p_x}(y)$. And it is clear that the only generators that are not fixed by φ are those in A. We conclude that, effectively, A is a rooted subset. ◀

Decomposition induced by rooted subsets.   As we have seen, rooted subsets are related to the clustering behavior of the points. They are also related to the decomposition of the linearized persistence module: they induce summands.

▶ **Theorem 12.** *Let φ be an idempotent endomorphism of a persistent set S. Then the persistence module $\mathcal{L}S$ decomposes into*

$$\text{img}(\text{id}_{\mathcal{L}S} - \mathcal{L}\varphi) \oplus \mathcal{L}(\text{img}\,\varphi).$$

Proof. Since φ is idempotent, $\mathcal{L}\varphi$ is idempotent. By Lemma 6, this induces a decomposition

$$\mathcal{L}S \cong \text{img}(\text{id} - \mathcal{L}\varphi) \oplus \text{img}\,\mathcal{L}\varphi.$$

It is left to show that $\text{img}\,\mathcal{L}\varphi \cong \mathcal{L}(\text{img}\,\varphi)$. Applying Lemma 5 to $\mathcal{L}\varphi$, we have a factorization $\mathcal{L}\varphi \colon \mathcal{L}S \xrightarrow{\pi} \text{img}\,\mathcal{L}\varphi \xrightarrow{\iota} \mathcal{L}S$ with $\pi \circ \iota = \text{id}$. Applying Lemma 5 again, this time to φ, we have a factorization $\varphi \colon S \xrightarrow{\pi'} \text{img}\,\varphi \xrightarrow{\iota'} S$ with $\pi' \circ \iota' = \text{id}$. Now, split endomorphisms are

preserved by every functor: in the diagram $\mathcal{L}S \xrightarrow{\mathcal{L}\pi'} \mathcal{L}(\text{img }\varphi) \xrightarrow{\mathcal{L}\iota'} \mathcal{L}S$ it holds $\mathcal{L}\iota' \circ \mathcal{L}\pi' = \mathcal{L}\varphi$ and $\mathcal{L}\pi' \circ \mathcal{L}\iota' = \text{id}$. Thus, the endomorphism $\mathcal{L}\varphi$ splits in two ways:

$$
\begin{array}{ccc}
 & \text{img }\mathcal{L}\varphi & \\
 \overset{\pi}{\nearrow} & \downarrow \cong & \overset{\iota}{\searrow} \\
\mathcal{L}S & & \mathcal{L}S \\
 \underset{\mathcal{L}\pi'}{\searrow} & & \underset{\mathcal{L}\iota'}{\nearrow} \\
 & \mathcal{L}(\text{img }\varphi), &
\end{array}
$$

where the middle arrow exists and is an isomorphism by Lemma 4, finishing the proof. ◄

Combining the above theorem with the Krull-Schmidt theorem, we obtain the following:

▶ **Corollary 13.** *A rooted subset of a persistent set S induces a summand in the decomposition of $\mathcal{L}S$. A rooted generator $x \in S_{p_x}$ induces an interval summand, and all other summands can be obtained by decomposing $\mathcal{L}(\text{img }\varphi)$, where φ is the endomorphism associated to x.*

This allows to iteratively *peel off* intervals of a persistence module of the form $\mathcal{L}S$: find a rooted generator of S, with associated idempotent φ, and continue considering $\text{img }\varphi$ instead of S. In the setting of an augmented metric space (M, d, f) and its density-Rips persistent set, the intervals that are peeled off are easily interpretable through the clustering behavior of the points M, by Proposition 10. Moreover, the conditions we describe actually happen in practice, as we see in Section 5.

Neighborly rooted points. In fact, certain points of an augmented metric space (M, d, f) can be seen to be rooted by looking at the nearest neighbors, which will be useful in Section 5. In what follows we fix a total order on M compatible with the order induced by f. Recall that the **nearest neighbor** of x is the element $x' \neq x$ of minimum distance to x, where ties have been broken by the fixed total order on M.

▶ **Definition 14.** *Let (M, d, f) be an augmented metric space. An element x is **neighborly rooted** if its nearest neighbor $y \in M$ satisfies $f(y) \leq f(x)$.*

▶ **Lemma 15.** *With the notation as above, if a point $x \in M$ is neighborly rooted then x is a rooted generator in the density-Rips persistent set of (M, d, f).*

Proof. It is clear that the nearest neighbor of x satisfies the conditions of Proposition 10. ◄

▶ Remark 16. We can identify all neighborly rooted points in the time it takes to solve the all-nearest-neighbor problem. Naturally, the all-nearest-neighbor problem can be solved in $O(n^2)$, where n is the number of points, by checking all possible pairs. When the points are in Euclidean space, the running time can be improved to $O(n \log n)$ time [19, 37].

Two notable intervals in the decomposition. The concept of rooted generators allows us to prove that, in certain cases, we can find at least two intervals in the decomposition of $\mathcal{L}S$, as in Theorem 18 below. We first prove Theorem 17, which has already appeared in [10, Theorem 5.3], where the proof method is to directly construct an endomorphism of the persistence module, as we also do after composing with the linearization functor.

▶ **Theorem 17.** *Let S be a persistent set. Suppose that the preordered set of generators of S has a bottom $\perp$ (that is, one has $\perp \leq x$ for any other generator x). Then the decomposition of $\mathcal{L}S$ consists of at least one interval, induced by $\perp$.*

Proof. Let $\bot \in S_{p_\bot}$ be a bottom and let $x \in S_{p_x}$ be a generator of S. Since $\bot \in S_{p_\bot}$ is a bottom, we have $p_\bot \leq p_x$. We can define an idempotent $\varphi \colon S \to S$ by $\varphi_{p_x}(x) = S_{p_\bot \to p_x}(\bot)$ for every generator $x \in S_{p_x}$ of S. This endomorphism is well-defined and its image has only one generator, namely $\bot$, and thus $\mathcal{L}(\operatorname{img} \varphi)$ is isomorphic to an interval module. ◀

▶ **Theorem 18.** *Let (M, d, f) be an augmented metric space, and let $S \colon P \to \mathsf{Set}$ be its density-Rips persistent set, as in Definition 2. If $|M| \geq 2$ then the decomposition of $\mathcal{L}S$ into indecomposable summands consists of at least two intervals.*

Proof. Consider a point $\top \in M$ of maximal function value, that is, $f(\top) \geq f(x)$ for any other $x \in M$. Let y be the nearest neighbor of $\top$. Since $M_{f(\top)} = M_\sigma$ for any $\sigma \geq f(\top)$, it is clear that $\top$ and its nearest neighbor y satisfy the conditions of Proposition 10, and thus $\top$ is a rooted generator, yielding the first interval. For the second interval, we note that there is at least one point $\bot \in M$ of minimal density value and apply Theorem 17. ◀

▶ **Example 19.** Not every summand of an indecomposable decomposition can be obtained by taking rooted subsets and applying Corollary 13. As an example, consider the augmented metric space given by six points $\{x_0, \dots, x_5\}$ in the plane as in Figure 1. Note that x_4 and x_5 are rooted in the associated density-Rips persistent set, and that they can be peeled off. After peeling, we obtain a persistent set $S \colon P \to \mathsf{Set}$ with $P := \{0, 2, 3, 4\} \times \{0, 1, 2, 3, 4, 5\} \subset \mathbb{R}^2$, which we describe in Figure 2. This example is an adaptation of [14, Example 4.12], which is introduced in the context of *conquerors* that we discuss in Section 4.
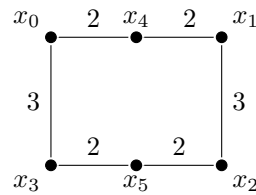
Figure 1 The augmented metric space (M, d, f) of Example 19, with $f(x_i) = i$. These are six points $\{x_0, \dots, x_5\}$ in the plane, where the distances are given by the numbers next to each line.

We claim that the persistence module $\mathcal{L}S \colon P \to \mathsf{Vec}$ decomposes into four summands, all of them interval modules. We denote these summands by I_0, I_1, I_2 and I_3, where each I_i is associated to the generator (p_i, x_i) of S, where $p_i = (0, i) \in P$. For each $i = 0, \dots, 3$, we set $(I_i)_p = 0$ for any $p < p_i$ and $(I_i)_{p_i} = K$, and we define $\iota_i \colon I_i \to \mathcal{L}S$ by

$$(\iota_0)_{p_0}(1) = [x_0], \qquad\qquad (\iota_1)_{p_1}(1) = [x_1] - [x_0],$$
$$(\iota_2)_{p_2}(1) = [x_2] - [x_1], \qquad\qquad (\iota_3)_{p_3}(1) = [x_3] - [x_0] + [x_1] - [x_2].$$

The support of each I_i are the grades $p \geq p_i$ such that $((\mathcal{L}S)_{p_i \to p} \circ (\iota_i)_{p_i})(1)$ is not zero. It can be seen that these maps induce a decomposition $\mathcal{L}S \cong I_0 \oplus I_1 \oplus I_2 \oplus I_3$.

However, no subset of the generators other than $\{x_1, x_2, x_3\}$ is rooted because each of the connected components given by $\{x_1, x_0\}$, $\{x_1, x_2\}$, $\{x_2, x_3\}$, and $\{x_0, x_3\}$ appear in S.

4 Rooted generators as a generalization of the elder rule

Single-parameter case. We now suppose that the poset P is a finite totally ordered poset. In this setting, the theory of rooted generators allows us to recover the *elder rule* [24] (see also [22] and [14]).
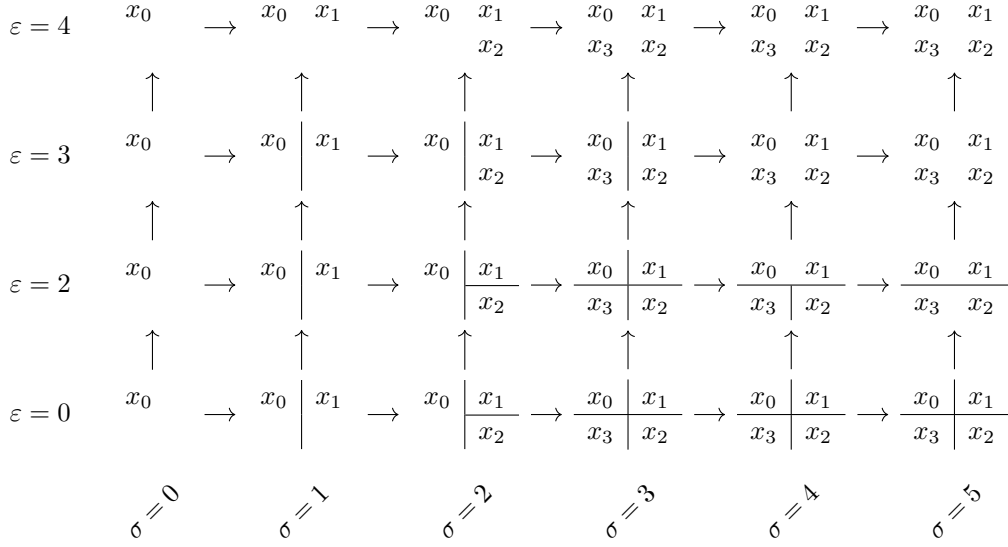
Figure 2 The persistent set $S\colon P \to \mathsf{Set}$ of Example 19 obtained by taking the density-Rips persistence set of Figure 1 and removing x_4 and x_5. Each node in the grid represents a partition of the x_i, where x_i and x_j are in the same partition if they are not separated by a line. The arrows are the functions that send the partition of x_i in one node to the partition of x_i in the other.

▶ **Proposition 20.** *Let P be a finite totally ordered poset and let $S\colon P \to \mathsf{Set}$ be a persistent set. Suppose that S has at least two generators and that $S_\top$ is a singleton, where $\top$ is the maximum element of P. Then every maximal generator (in the preorder of Definition 7) is rooted.*

Proof. Let $x \in S_{p_x}$ be a maximal generator, and define

$$I_x := \{q \in P \mid q \geq p_x \text{ and, for any other generator } w \in S_{p_w},\ S_{p_w \to q}(w) \neq S_{p_x \to q}(x)\}.$$

Since $p_x \in I_x$, I_x is not empty, and we can consider the set $U \subset P$ of upper bounds of I_x. Moreover, since $S_\top = \{*\}$ and there are at least two generators by assumption, the set $U \setminus I_x$ is not empty. Let α be the least element in $U \setminus I_x$. By construction of I_x and $U \setminus I_x$, there is a generator $y \in S_{p_y}$ such that $S_{p_y \to \alpha}(y) = S_{p_x \to \alpha}(x)$. Now, since x is maximal, it holds that $p_y \leq p_x$, and we can define an idempotent $\varphi\colon S \to S$ by $\varphi_{p_x}(x) = S_{p_y \to p_x}(y)$, and $\varphi_{p_z}(z) = z$ for any other generator $z \in S_{p_z}$. Such an idempotent is well-defined by the way we have defined α: if there is any other generator $w \in S_{p_w}$ such that $S_{p_w \to q}(w) = S_{p_x \to q}(x)$ then $\alpha \leq q$ and also $S_{p_y \to q}(y) = S_{p_x \to q}(x)$. We conclude that x is rooted, as desired. ◀

Thus, when P is a total order, we can decompose any persistence module $\mathcal{L}S$ by peeling off rooted generators, following Theorem 12 and by iteratively considering maximal generators.

Relation to constant conquerors. Let (M, d, f) be an augmented metric space. Cai, Kim, Mémoli and Wang [14] define the concept of a constant conqueror as follows. First, define an ultrametric on M: $u(x, x') := \min\{\varepsilon \in [0, \infty) \mid x \text{ and } x' \text{ are path-connected in } \mathcal{G}_\varepsilon(M)\}$.

Now fix a total order $\prec$ on M and let $x \in M$ be a non-minimal element with respect to this order. A **conqueror** of x in M is another point $x' \in M$ such that (1) $x' \prec x$, and (2) for

any x'' with $x'' \prec x$ one has $u(x, x') \leq u(x, x'')$. Given a function $f \colon M \to \mathbb{R}$, a **conqueror function** of a non-minimal $x \in M$, with respect to $\prec$, is a function $c_x \colon [f(x), \infty) \to M$ that sends each σ to a conqueror of x in M_σ. For the minimal element $\bot$ of M we define $c_\bot \colon [f(\bot), \infty) \to M$ to be the constant function at $\bot$.

Also, in the same paper [14], given a point $x \in M$, and assuming that $f \colon M \to \mathbb{R}$ is injective, the authors define the **staircode** of x as the set given by

$$I_x := \{(\varepsilon, \sigma) \in \mathbb{R}_{\geq 0} \times \mathbb{R} \mid x \in M_\sigma \text{ and } x \text{ is the oldest in } [x]_{(\varepsilon, \sigma)}\},$$

where $[x]_{(\varepsilon, \sigma)}$ is the set of points that are path-connected to x in $\mathcal{G}_\varepsilon(M_\sigma)$ and being the "oldest" means $f(x) < f(x')$ for any other $x' \in [x]_{(\varepsilon, \sigma)}$. The authors also define an analogous notion when f is not injective, which we do not reproduce here.

Finally, the authors ask the following question:

▶ **Question 21.** *Let (M, d, f) be an augmented metric space. If $x \in M$ has a constant conqueror function, is the interval module supported by I_x a summand of its density-Rips persistence module?*

If we replace constant conqueror by rooted generator then the answer is yes, by Corollary 13. The next example shows that the same cannot hold as originally stated in the question above.

▶ **Example 22.** Consider the subset M of $\mathbb{R}$ given by the points $x_0 = 0$, $x_1 = 7.5$, $x_2 = 3$ and $x_3 = 5$. Under the metric induced by the Euclidean distance on $\mathbb{R}$, M is a metric space, and can be made into an augmented metric space by defining $f(x_i) = i$, see Figure 3.
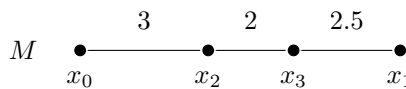
Figure 3 The augmented metric space (M, d, f) of Example 22, with $M \subset \mathbb{R}$ and $f(x_i) = i$.
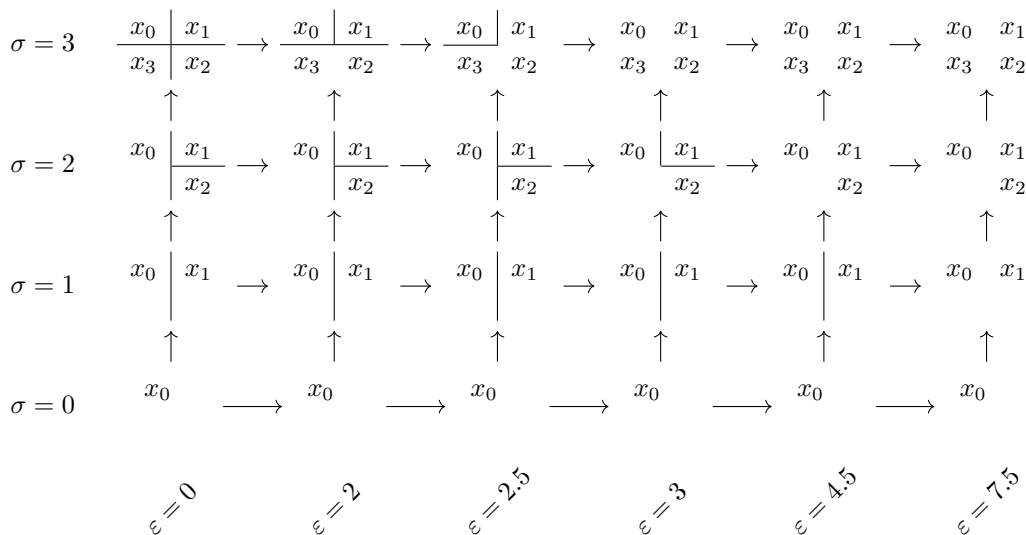
Figure 4 We picture the density-Rips persistent set of Figure 3.
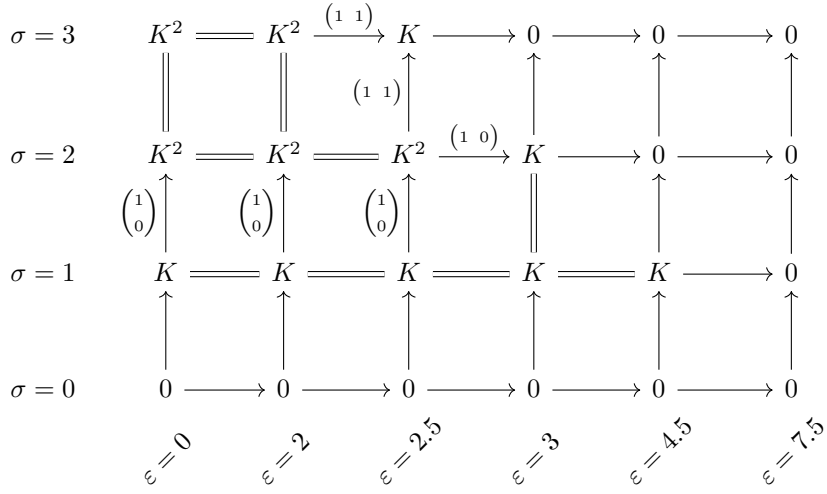
Figure 5 An indecomposable persistence module $F \colon P \to \mathsf{Vec}$, as referenced in Example 22.

Consider the only total order $\prec$ on M compatible with f, $x_0 \prec x_1 \prec x_2 \prec x_3$. The point x_1 has a constant conqueror: x_0 is the only candidate, and it is clear that, for every $i = 1, \ldots, 3$ and $x' \prec x_1$, $u_i(x_1, x_0) \leq u_i(x_1, x')$, where u_i is the ultrametric of M_i, precisely because x_0 is the only point that satisfies $x' \prec x_1$.

Let $S \colon P \to \mathsf{Set}$ be the density-Rips persistent set constructed from the augmented metric space (M, d, f). Here, P is the subposet of $\mathbb{R}^2$ given by $\{0, 2, 2.5, 3, 4.5, 7.5\} \times \{0, 1, 2, 3\}$, where the first coordinate represents the distances and the second coordinate the densities. We picture S in Figure 4. Now we proceed to decompose $\mathcal{L}S$. First, note that x_3 is a rooted generator, and consider an associated idempotent $\varphi \colon S \to S$. By Theorem 12, there is an interval $I \coloneqq \mathrm{img}(\mathrm{id}_{\mathcal{L}S} - \mathcal{L}\varphi)$ in the decomposition, and we can continue considering the persistent set $\mathrm{img}\,\varphi$. In $\mathrm{img}\,\varphi$, x_0 is a minimal generator. By Theorem 17 (and its proof) there is an idempotent $\psi \colon \mathrm{img}\,\varphi \to \mathrm{img}\,\varphi$ such that $I' \coloneqq \mathrm{img}\,\mathcal{L}\psi$ is an interval. Applying Theorem 12 again, we obtain a decomposition of $\mathcal{L}S$ of the form

$$I \oplus I' \oplus \mathrm{img}(\mathrm{id}_{\mathcal{L}\,\mathrm{img}\,\varphi} - \mathcal{L}\psi).$$

By direct computation, it can be seen that $\mathrm{img}(\mathrm{id}_{\mathcal{L}\,\mathrm{img}\,\varphi} - \mathcal{L}\psi)$ is isomorphic to the persistence module described in Figure 5. Moreover, this persistence module is indecomposable, which can be checked by looking at its endomorphisms: a persistence module F is indecomposable if and only if every endomorphism of F is either nilpotent or an isomorphism (see [9], and [10]).

Note that x_1 is not a rooted generator in $\mathcal{L}S$. In M_1, x_1 is its own connected component during $\varepsilon \in [0, 7.5)$, until x_0 joins the connected component. And in M_3 it is by itself during $\varepsilon \in [0, 2.5)$ and then joins the connected component of x_3, which is not connected to x_0 at that point. Similarly, x_2 is not rooted.

▶ **Remark 23.** Note that in Condition (2) of the definition of conqueror, we require that $x'' \prec x$. This requirement measures part of the difference between constant conqueror function and rooted generator for augmented metric spaces. If we drop this requirement, denoting the resulting concept by **conqueror**[*], we suppose that f is injective, and that $\prec$ is compatible with the order induced by f, then a non-minimal, with respect to $\prec$, point $x \in M$ has a constant **conqueror**[*] function if and only if x is a rooted generator, as in Proposition 10.

5 A lower bound on the number of expected intervals

We apply the theory we have developed to the study of how a typical decomposition of a persistence module coming from density-Rips might look like. In particular, suppose we sample independently n points from a common density function $f(x)$ in $\mathbb{R}^d$, obtaining a finite metric space $M \subset \mathbb{R}^d$. We can then consider the augmented metric space (M, d_M, f), where f, rather than being an estimated density, is the true underlying density function. This setting resembles actual practice, but is more suitable to theoretical study. Let S be the density-Rips persistent set of M. Then, how many intervals can we expect in the decomposition of $\mathcal{L}S$? The following theorem says that, under very general conditions on f, regardless of d, and as n goes to infinity, we can at least expect 25% of the summands to be intervals.

▶ **Theorem 24.** *Let $X_1, \ldots, X_n$ be i.i.d. points taking values in $\mathbb{R}^d$, sampled from a common density function $f(x)$ that is continuous almost everywhere with respect to the Lebesgue measure.*

Consider the finite augmented metric space $(M = \{X_1, \ldots, X_n\}, d_M, f)$, where d_M is induced by the Euclidean metric in $\mathbb{R}^d$, and let S be its density-Rips persistent set.

Let $\mathfrak{I}_n$ be the random variable that counts the number of intervals in the indecomposable decomposition of $\mathcal{L}S$, and let $\mathfrak{S}_n$ be the random variable that counts the total number of summands in the same decomposition. We have

$$\liminf_{n \to \infty} \mathrm{E}\left[\frac{\mathfrak{I}_n}{\mathfrak{S}_n}\right] \geq c(d), \tag{3}$$

where $c(d)$ is a constant that depends on d, and $c(1) = \frac{1}{3}$, $c(2) \approx 0.31$ and $c(d) \downarrow \frac{1}{4}$ as $d \to \infty$.

The rest of the section is dedicated to proving this theorem. The nearest neighbor graph of a metric space plays a fundamental role.

▶ **Definition 25.** *The **nearest neighbor graph** of M is the directed graph on M given by the directed edges of the form (x, x'), where x' is the nearest neighbor of x.*

Now, we are interested in estimating the number of neighborly rooted elements, as in Definition 14, as they induce an interval in the decomposition of $\mathcal{L}S$. However, in general being neighborly rooted depends on f. To do without the condition on f we have:

▶ **Lemma 26.** *Let (M, d_M, f) be an augmented metric space and let S be its density-Rips persistent set. There are at least as many intervals in the indecomposable decomposition of $\mathcal{L}S$ as 2-cycles in the nearest neighbor graph of M.*

Proof. We can assume without loss of generality that $|M| \geq 2$. Let G be the nearest neighbor graph of M. The only cycles in this graph are precisely the 2-cycles, and each weakly connected component of G contains exactly one 2-cycle (see [25]).

Let $C_1, \ldots, C_k$ be the weakly connected components of G. Fix $i \in \{1, \ldots, k\}$, and let $x, y \in M$ be such that (x, y) and (y, x) is the 2-cycle in C_i. Either $f(y) \leq f(x)$ or $f(x) \leq f(y)$, and either x is neighborly rooted, y is neighborly rooted, or both are neighborly rooted. Say x is neighborly rooted, and define an endomorphism $\varphi_i \colon S \to S$ by setting

$$(\varphi_i)_{p_z}(z) = \begin{cases} S_{p_y \to p_x}(y), & \text{if } x = z, \\ z, & \text{otherwise,} \end{cases}$$

for every generator $z \in S_{p_z}$. Such an endomorphism is well-defined as shown in Proposition 10.

Constructing, for each i, an idempotent φ_i as above, it is clear that we can iteratively peel off the associated intervals, yielding the desired conclusion. ◀

Naturally, the number of 2-cycles is half the number of points that are the nearest neighbor of its nearest neighbor. The problem of estimating the probability for a point to be the nearest neighbor of its nearest neighbor, assuming a random point process, has been studied by multiple authors (see [21, 25, 26, 27, 35]).

In our case, when we have $X_1, \ldots, X_n$ i.i.d. points in $\mathbb{R}^d$ sampled from a common density function f under the conditions of Theorem 24, by [27, Theorem 1.1], and letting $N_{i,n}$ denote the probability event that X_i is the nearest neighbor of its nearest neighbor, we have

$$\lim_{n \to \infty} \mathrm{P}(N_{i,n}) = b(d), \tag{4}$$

where $b(d)$ is the volume of a unit d-sphere divided by the volume of the union of two unit spheres with centers at distance 1. In fact, $b(1) = \frac{2}{3}$, $b(2) \approx 0.621$, and $b(d) \downarrow \frac{1}{2}$ as $d \to \infty$ (see [35, Table 2]), and we define $c(d) := \frac{b(d)}{2}$.

We are now ready to finish the proof of Theorem 24 at the start of the section. Applying Lemma 26 and the linearity of expectation, it holds

$$\mathrm{E}[\mathfrak{I}_n] \geq \mathrm{E}\left[\sum_{i=1}^{n} \frac{I(N_{i,n})}{2}\right] = \sum_{i=1}^{n} \frac{\mathrm{E}[I(N_{i,n})]}{2} = \sum_{i=1}^{n} \frac{\mathrm{P}(N_{i,n})}{2},$$

where $I(N_i)$ is the indicator random variable of $N_{i,n}$. By Equation (4) we have

$$\liminf_{n \to \infty} \mathrm{E}\left[\frac{\mathfrak{I}_n}{n}\right] \geq \frac{b(d)}{2} = c(d).$$

Finally, noting that the number of summands in the decomposition is bounded by the number of points, $\mathfrak{S}_n \leq n$ (see full version), Equation (3) of Theorem 24 follows, finishing the proof.

6   Discussion

Although we have focused our attention to augmented metric spaces and density-Rips, rooted subsets can be applied to other persistent sets. Of special interest for us is the degree-Rips filtration [5] of a metric space, where we filter by the degree of the vertices in the underlying geometric graphs. To accommodate this situation, one could modify condition 1 of Proposition 10 to take into account the evolution of the degrees, rather than the density. We leave an in-depth treatment of this case for future work.

We have seen, both in our lower bound of Section 5 and in preliminary experimental evaluation, that we can expect to find many intervals in the decomposition of those persistence modules coming from geometry, at least in the cases considered here. This is in contrast to the purely algebraic setting, where, in light of recent developments [2, 3], looking for a decomposition might fall short.

References

1   Hideto Asashiba, Mickaël Buchet, Emerson G. Escolar, Ken Nakashima, and Michio Yoshi-waki. On interval decomposability of 2D persistence modules. *Computational Geometry*, 105/106:Paper No. 101879, 33, 2022. `doi:10.1016/j.comgeo.2022.101879`.

2   Ulrich Bauer, Magnus B. Botnan, Steffen Oppermann, and Johan Steen. Cotorsion tor-sion triples and the representation theory of filtered hierarchical clustering. *Advances in Mathematics*, 369:107171, 51, 2020. `doi:10.1016/j.aim.2020.107171`.

3   Ulrich Bauer and Luis Scoccola. Generic two-parameter persistence modules are nearly indecomposable, November 2022. `arXiv:2211.15306`.

4   Håvard Bakke Bjerkevik, Magnus Bakke Botnan, and Michael Kerber. Computing the interleaving distance is NP-hard. *Foundations of Computational Mathematics*, 20(5):1237–1271, 2020. `doi:10.1007/s10208-019-09442-y`.

5   Andrew J. Blumberg and Michael Lesnick. Stability of 2-parameter persistent homology. *Foundations of Computational Mathematics*, 2022. `doi:10.1007/s10208-022-09576-6`.

6   Magnus Bakke Botnan and William Crawley-Boevey. Decomposition of persistence modules. *Proceedings of the American Mathematical Society*, 148(11):4581–4596, 2020. `doi:10.1090/proc/14790`.

7   Magnus Bakke Botnan and Michael Lesnick. Algebraic stability of zigzag persistence modules. *Algebraic & Geometric Topology*, 18(6):3133–3204, 2018. `doi:10.2140/agt.2018.18.3133`.

8   Magnus Bakke Botnan, Steffen Oppermann, Steve Oudot, and Luis Scoccola. On the bottleneck stability of rank decompositions of multi-parameter persistence modules, July 2022. `arXiv:2208.00300`.

9   Michel Brion. Representations of quivers. In *Geometric methods in representation theory. I*, volume 24 of *Sémin. Congr.*, pages 103–144. Soc. Math. France, Paris, 2012.

10   Jacek Brodzki, Matthew Burfitt, and Mariam Pirashvili. On the complexity of zero-dimensional multiparameter persistence, August 2020. `arXiv:2008.11532`.

11   Mickaël Buchet, Frédéric Chazal, Tamal K. Dey, Fengtao Fan, Steve Y. Oudot, and Yusu Wang. Topological analysis of scalar fields with outliers. In *31st International Symposium on Computational Geometry*, volume 34 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages 827–841, 2015. `doi:10.4230/LIPIcs.SOCG.2015.827`.

12   Mickaël Buchet and Emerson G. Escolar. Every 1D persistence module is a restriction of some indecomposable 2D persistence module. *Journal of Applied and Computational Topology*, 4(3):387–424, 2020. `doi:10.1007/s41468-020-00053-z`.

13   Mickaël Buchet and Emerson G. Escolar. Realizations of indecomposable persistence modules of arbitrarily large dimension. *Journal of Computational Geometry*, 13(1):298–326, 2022. `doi:10.20382/jocg.v13i1a12`.

14   Chen Cai, Woojin Kim, Facundo Mémoli, and Yusu Wang. Elder-rule-staircodes for augmented metric spaces. *SIAM Journal on Applied Algebra and Geometry*, 5(3):417–454, 2021. `doi:10.1137/20M1353605`.

15   Gunnar Carlsson and Facundo Mémoli. Characterization, stability and convergence of hierarchical clustering methods. *Journal of Machine Learning Research*, 11:1425–1470, 2010.

16   Gunnar Carlsson and Facundo Mémoli. Multiparameter hierarchical clustering methods. In *Classification as a tool for research*, Stud. Classification Data Anal. Knowledge Organ., pages 63–70. Springer, Berlin, 2010. `doi:10.1007/978-3-642-10745-0_6`.

17   Gunnar Carlsson and Facundo Mémoli. Classifying clustering schemes. *Foundations of Computational Mathematics*, 13(2):221–252, 2013. `doi:10.1007/s10208-012-9141-9`.

18   Gunnar Carlsson and Afra Zomorodian. The theory of multidimensional persistence. *Discrete & Computational Geometry*, 42(1):71–93, 2009. `doi:10.1007/s00454-009-9176-0`.

19   Kenneth L. Clarkson. Fast algorithms for the all nearest neighbors problem. In *24th Annual Symposium on Foundations of Computer Science*, pages 226–232. IEEE, 1983. `doi:10.1109/SFCS.1983.16`.

20   Anne Collins, Afra Zomorodian, Gunnar Carlsson, and Leonidas J. Guibas. A barcode shape descriptor for curve point cloud data. *Computers & Graphics*, 28(6):881–894, 2004. `doi:10.1016/j.cag.2004.08.015`.

21   Trevor F. Cox. Reflexive nearest neighbours. *Biometrics*, 37(2):367, 1981. `doi:10.2307/2530424`.

22   Justin Curry. The fiber of the persistence map for functions on the interval. *Journal of Applied and Computational Topology*, 2(3-4):301–321, 2018. `doi:10.1007/s41468-019-00024-z`.

23   Tamal K. Dey and Cheng Xin. Generalized persistence algorithm for decomposing multiparameter persistence modules. *Journal of Applied and Computational Topology*, 6(3):271–322, 2022. `doi:10.1007/s41468-022-00087-5`.

24 Herbert Edelsbrunner and John L. Harer. *Computational topology: an introduction.* American Mathematical Society, Providence, RI, 2010. `doi:10.1090/mbk/069`.

25 David Eppstein, Michael S. Paterson, and Frances F. Yao. On nearest-neighbor graphs. *Discrete & Computational Geometry*, 17(3):263–282, 1997. `doi:10.1007/PL00009293`.

26 Norbert Henze. On the probability that a random point is the j th nearest neighbour to its own k th nearest neighbour. *Journal of Applied Probability*, 23(1):221–226, 1986. `doi:10.2307/3214132`.

27 Norbert Henze. On the fraction of random points by specified nearest-neighbour interrelations and degree of attraction. *Advances in Applied Probability*, 19(4):873–895, 1987. `doi:10.2307/1427106`.

28 Jon Kleinberg. An impossibility theorem for clustering. In *Advances in Neural Information Processing Systems*, volume 15. MIT Press, 2002.

29 Michael Lesnick and Matthew Wright. Interactive visualization of 2-d persistence modules, December 2015. `arXiv:1512.00180`.

30 Saunders Mac Lane. *Categories for the Working Mathematician*, volume 5 of *Graduate Texts in Mathematics*. Springer New York, 1978. `doi:10.1007/978-1-4757-4721-8`.

31 Leland McInnes and John Healy. Accelerated hierarchical density clustering. In *2017 IEEE International Conference on Data Mining Workshops*, pages 33–42, 2017. `doi:10.1109/ICDMW.2017.12`.

32 Samantha Moore. Hyperplane restrictions of indecomposable n-dimensional persistence modules. *Homology, Homotopy and Applications*, 24(2):291–305, 2022. `doi:10.4310/HHA.2022.v24.n2.a14`.

33 Alexander Rolle and Luis Scoccola. Persistable: Persistent and stable clustering. URL: `https://github.com/LuisScoccola/persistable`.

34 Alexander Rolle and Luis Scoccola. Stable and consistent density-based clustering, July 2021. `arXiv:2005.09048`.

35 Mark F. Schilling. Mutual and shared neighbor probabilities: finite- and infinite-dimensional results. *Advances in Applied Probability*, 18(2):388–405, 1986. `doi:10.2307/1427305`.

36 Bernard W. Silverman. *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London, 1986.

37 Pravin M. Vaidya. An $O(n \log n)$ algorithm for the all-nearest-neighbors problem. *Discrete & Computational Geometry*, 4(2):101–115, 1989. `doi:10.1007/BF02187718`.