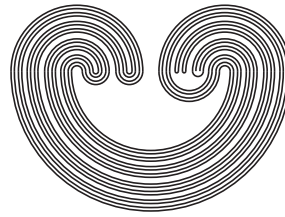


<http://topology.auburn.edu/tp/>

TOPOLOGY PROCEEDINGS



Volume 55, 2020

Pages 273–294

<http://topology.nipissingu.ca/tp/>

SINGULAR PERSISTENT HOMOLOGY WITH GEOMETRICALLY PARALLELIZABLE COMPUTATION

by

BORIS GOLDFARB

Electronically published on February 14, 2020

Topology Proceedings

Web: <http://topology.auburn.edu/tp/>

Mail: Topology Proceedings
Department of Mathematics & Statistics
Auburn University, Alabama 36849, USA

E-mail: topolog@auburn.edu

ISSN: (Online) 2331-1290, (Print) 0146-4124

COPYRIGHT © by Topology Proceedings. All rights reserved.

SINGULAR PERSISTENT HOMOLOGY WITH GEOMETRICALLY PARALLELIZABLE COMPUTATION

BORIS GOLDFARB

ABSTRACT. Persistent homology is a popular tool in Topological Data Analysis. It provides numerical characteristics of data sets which reflect global geometric properties. In order to be useful in practice, for example for feature generation in machine learning, it needs to be effectively computable. Classical homology is a computable topological invariant because of the Mayer-Vietoris exact and spectral sequences associated to coverings of a space. We state and prove versions of the Mayer-Vietoris theorem for persistent homology under mild and commonplace assumptions. This is done through the use of a new theory, the singular persistent homology, better suited for handling coverings of data sets. As an application, we create a distributed computational workflow where the advantage is not only or even primarily in speed improvement but in sheer feasibility for large data sets.

1. MOTIVATION AND STATEMENT OF RESULTS

In order to describe the problem addressed in this paper, we start by recalling the standard treatment of persistent homology. This construction is designed to leverage computations in algebraic topology in order to quantify geometric properties of finite data sets. It provides a multi-scale representation of geometric features of the data, including the relations between the scales. On a more sophisticated level, one can use filtrations

2010 *Mathematics Subject Classification.* Primary 55N99; Secondary 68W15.

Key words and phrases. Persistent homology, singular homology, Mayer-Vietoris theorem, distributed computation.

©2020 Topology Proceedings.

of a simplicial complex in place of scales to build a similar persistent homology signature of a filtration. Persistent homology has established itself as a robust feature included in machine learning applications in addition to other ad hoc uses in data science. We refer to the surveys Edelsbrunner/Harer [9] and Carlsson [6] for expositions.

Consider a filtration of a general topological space X with the indexing set $[0, +\infty)$. This filtration can be thought of as a covariant functor ϕ from $[0, +\infty)$ with the usual total ordering to the category of subspaces of X partially ordered by inclusion. Usually in the applications the topological space is a simplicial complex, and the filtration stages are simplicial subcomplexes. Let us use the notation X_ε for $\phi(\varepsilon)$. For a functor H_n from the appropriate category to modules over a ring R , one has the induced diagram with the nodes $H_n(X_\varepsilon)$ and the maps $H_n(X_\varepsilon \rightarrow X_{\varepsilon'})$ for any pair of non-negative reals $\varepsilon \leq \varepsilon'$. This whole diagram is sometimes referred to as the H_n -persistence module of the filtration and can be denoted as $pH_n(\phi)$.

The δ -persistent H_n -module of X_ε is the image of the induced map $H_n(X_\varepsilon) \rightarrow H_n(X_{\varepsilon+\delta})$. The δ -persistent n -th Betti number of X_ε can be defined as the rank of this submodule. This framework is then used to characterize the H_n -based features from X_ε that are still present in $X_{\varepsilon+\delta}$. For example, one can measure the *survival interval* for each of the elements in $H_n(X_\varepsilon)$ as the difference between the infimum of all ε' such that the element is not in the kernel of the map $H_n(X_\varepsilon) \rightarrow H_n(X_{\varepsilon'})$ and the value of ε itself.

In order to describe the appearance of persistence modules in data science, we specialize to a finite metric space M with t points. This is the usual framework for measuring dissimilarity between t data points. Associated to this metric space is the simplex X spanned by all t points as vertices.

Definition 1.1. The *Vietoris-Rips filtration* ρ of X is by subcomplexes X_ε defined by the condition that a subset S of M spans a simplex in X_ε if and only if $d(p, q) \leq \varepsilon$ for every pair of points from S .

Using the simplicial n -dimensional homology functor H_n gives a persistent module for each dimension n . In this case the δ -persistent homology module of X_ε characterizes the n -cycles in X_ε that are not the boundary of any $(n+1)$ -chain from the larger subcomplex $X_{\varepsilon+\delta}$. In order to make these invariants computable in practice and explainable to a broad population of data scientists, it is prudent to specialize to fields as coefficients. It becomes possible to use phrases such as “the δ -persistent Betti number of X_ε measures the number of n -dimensional holes in $X_{\varepsilon+\delta}$ created by the subcomplex X_ε .”

Notation 1.2. When M is a finite metric space, and the Vietoris-Rips filtration is understood, we will use the notation $H_n^\varepsilon(M)$ for $H_n(X_\varepsilon)$ and the notation $pH_n(M)$ for the resulting persistence module $pH_n(\rho)$.

Continuing with the setting in the last paragraph, there is a nice visual summary that can be constructed to describe the persistent module. This description requires one to use structural theorems for finitely generated graded $K[x]$ -modules, where K is a field, so we will assume that the coefficient ring is a field in the rest of this paper.

The *persistence barcode* of M is a collection of intervals indicating the birth of a new generator in some $H_n^\varepsilon(M)$ and its death in $H_n^{\varepsilon+\delta}(M)$, where δ is the survival interval for that generator. We refer to [6] for details and will include some illustrations shortly in this section.

Simplicial homology is one of the most computable algebraic invariants in topology. There are two reasons for that. One is that homology depends on a discrete representation of the space, its triangulation. This allows to use linear algebra to perform computations over fields whose complexity grows with the size of the triangulation but which use well understood and refined algorithms. The other reason is the excision property. This property is crucial for the axiomatic characterization of homology and distinguishes homology from other topological invariants.

This goal of this paper is to state and verify an analogue of the Mayer-Vietoris property for persistent homology. The goal of such a theorem should be to recover information about the persistent homology barcodes for the total data set from the barcodes associated to subsets that form a specific kind of covering. Of course, in practice this kind of approach to computation is very desirable because it allows to parallelize the local computations and therefore makes persistent homology computations feasible for larger data sets under the same memory constraints. It is also possible that this divide-and-concur strategy would be more efficient than direct homology computations with M .

We will use the rest of this section to demonstrate that this goal is not achievable in its entirety and offer a compromise applicable and sufficient in many practical situations. The fact that there is no perfect excision property for persistent homology shouldn't be surprising. Analogues of purely topological theorems usually fail without careful incorporation of local information or preservation of relevant information about the intersections as in excision results for schemes in algebraic geometry.

Example 1.3. Here we compare persistence barcodes for two data sets in $\mathbb{Z}^2$:

$$\begin{aligned} A &= \{(0, 1), (0, 2), (1, 0), (1, 3), (2, 0), (2, 3), (3, 1), (3, 2)\}, \\ B &= \{(-3, 1), (-2, 0), (-1, 0), (0, 1), (0, 2), (1, 0), (2, 0), (3, 1)\}. \end{aligned}$$

The sets are shown in Figure 1 together with coverings by three subsets each:

$$\begin{aligned} A_1 &= \{(0, 1), (0, 2), (1, 0), (1, 3), (2, 0)\}, \\ A_2 &= \{(1, 0), (2, 0), (2, 3), (3, 1), (3, 2)\}, \\ A_1 \cap A_2 &= \{(1, 0), (2, 0)\}, \\ B_1 &= \{(-3, 1), (-2, 0), (-1, 0), (0, 1), (0, 2)\}, \\ B_2 &= \{(0, 1), (0, 2), (1, 0), (2, 0), (3, 1)\}, \\ B_1 \cap B_2 &= \{(0, 1), (0, 2)\}. \end{aligned}$$

It is clear that the similarly indexed (and colored) subsets are pairwise isometric, and so pushouts of the local data should be the same in both cases. However, the two data sets are not isometric and, in fact, have distinct barcodes as also shown in Figure 1, in all dimensions.

In order to state our theorem as a partial resolution to this dilemma, we will need to define the following notion of a partial barcode.

Definition 1.4. For a chosen number ℓ , the ℓ -*prefix* of a persistent module is the restriction of the functor $H_n \circ \phi$ to the interval $[0, \ell]$. We will denote this functor by ℓ - pH_n . If we view the barcode as an unordered collection of intervals with specified times of birth and death of each homology generator, in the form of $\{(b_\alpha, d_\alpha)\}$, then the ℓ -*prefix* of the barcode is the collection of intervals $\{(b_\alpha, \min\{d_\alpha, \ell\})\}$. Visually, one simply erases the part of the given barcode that is beyond the threshold ℓ .

We will use the term *simplicial tree* for a simplicial complex that is connected and has no cycles. A covering of a metric space whose nerve is a simplicial tree and Lebesgue number at least $\ell \geq 0$ will be called a *tree-like decomposition of rank ℓ* . Suppose each of the covering sets with the subspace metric also has a tree-like decomposition of rank ℓ . Then we say that the resulting covering by smaller sets is a *hierarchical tree-like decomposition of rank ℓ and depth 2*.

Inductively, for a natural number D one defines a *hierarchical tree-like decomposition of rank ℓ and depth D* . We will refer to the sets that appear in such hierarchical decomposition and are not unions of other sets as *primary sets*.

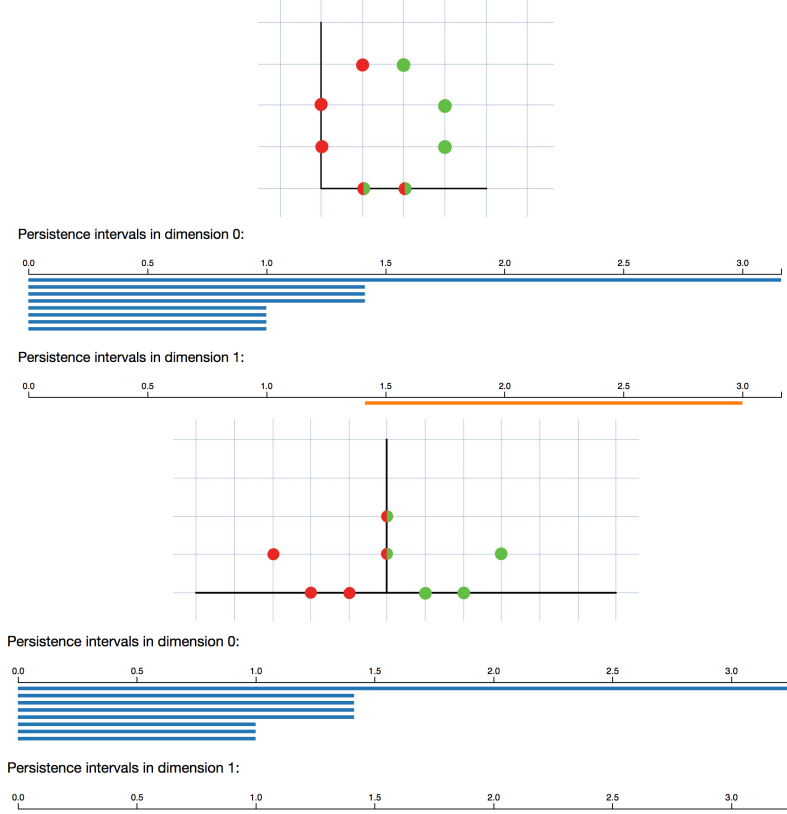


FIGURE 1. Reconstruction of the persistence barcode from a complete metric covering is not a well-posed problem.

Main Theorem. *Given a tree-like decomposition $\mathcal{U}$ of rank ℓ and depth D of a finite metric space M , the ℓ -prefix of the persistence module of M based on the Vietoris-Rips filtration can be reconstructed from the ℓ -prefixes of the persistence modules of each primary set U in $\mathcal{U}$, with the subspace metric, and the ℓ -prefixes of the persistence modules for a family of intersections $U \cap V$ for the subsets U and V in $\mathcal{U}$. This follows from the basic technical result which is a long exact Mayer-Vietoris sequence*

$$\begin{aligned} \dots \longrightarrow \bigoplus_{U,V \in \mathcal{U}} \ell\text{-p}H_n(U \cap V) &\xrightarrow{f_n} \bigoplus_{U \in \mathcal{U}} \ell\text{-p}H_n(U) \xrightarrow{g_n} \ell\text{-p}H_n(M) \\ &\longrightarrow \bigoplus_{U,V \in \mathcal{U}} \ell\text{-p}H_{n-1}(U \cap V) \longrightarrow \dots \end{aligned}$$

for a covering of M by two subsets U and V with Lebesgue number at least $\ell \geq 0$.

The easiest to construct and probably the most useful hierarchical tree-like decomposition of depth D can be obtained for any subset of the product of D simplicial or $\mathbb{R}$ -trees. It is explained in the following example. A further, simplest case of this example is for the trees obtained as triangulations of the real line worked out explicitly in section 4.

Example 1.5. Let T be a simplicial tree where each edge is given length 1 with the global metric induced as a path metric. We fix a vertex v_0 . Given another vertex $v \in T$, we define the subset $\text{Sh}(v) = \{t \in T \mid v \in [v_0, t]\}$. Let $B(v, r)$ stand for the open metric ball of radius r centered at v and $S(v, r)$ stand for its boundary sphere. We also define

$$\text{Sh}(v, l) = \text{Sh}(v) \cap B(v, l)$$

for a positive number l , and

$$\text{Sh}(v; l_1, l_2) = \text{Sh}(v, l_2) \setminus B(v, l_1) \setminus S(v, l_1),$$

for $l_2 > l_1 > 0$.

Given a number r greater than 1, consider the collection of open subsets of T consisting of the ball $B(v_0, 2r)$ and the differences $\text{Sh}(v; r-1, 3r)$ where the vertices v vary over $S(v_0, (2n-1)r)$ for arbitrary natural numbers n . It is easy to see that (1) this collection is a covering of T . Its nerve is itself a (union of) tree(s) where the vertices can be indexed by v_0 and the vertices $v \in S(v_0, (2n-1)r)$, the edges are the pairs (v, v') where $v' \in \text{Sh}(v, 2(n+1)r)$, (2) the diameter of each set in the covering is bounded by $6r$, and (3) the covering has a Lebesgue number at least r . Clearly, intersecting any subset V of T with the produced covering gives a covering of V with exactly the same three properties. This gives a hierarchical tree-like decomposition of V of rank r and depth 1.

For a product Π of D trees, one starts by performing this construction in each of the factors, then takes the product of the covering subsets to create a covering of Π . This is a hierarchical tree-like decomposition of rank ℓ and depth D . Again, intersecting any subset V of the product with the covering subsets is a hierarchical tree-like decomposition of V of rank ℓ and depth D .

An $\mathbb{R}$ -tree or *real tree* in geometry is a geodesic metric space where every triangle is a tripod. Divergence behavior of geodesics in an $\mathbb{R}$ -tree allows to generalize the above constructions in simplicial trees verbatim to $\mathbb{R}$ -trees and their products.

It will be clear from the main body of the paper that for any space with a hierarchical tree-like decomposition of rank ℓ and finite depth, there is a finite inductive procedure for reconstructing the ℓ -prefix of the persistence module of M through a finite number of extension problems from the ℓ -prefixes of the persistence modules of a selection of intersections between

the primary sets. The details and the implementation of this general algorithm will appear in a separate paper.

The proof of the Main Theorem will be presented in terms of a new theory, singular persistent homology, which we introduce in the next section. It is true that once the statement is made, the proof can be given entirely in terms of simplicial persistent homology. In fact the two theories are isomorphic when applied to a finite metric space as we will show. However, let us make three points. First, it is much easier to think in terms of coverings and related persistence diagrams within the new theory as will become apparent. This allows to quickly discover the correct statements. Second, the proofs also appear more natural and easy in the singular theory. Finally, when the complexity of related theories becomes greater, with more general sequences of maps than filtrations, such as in zig-zag or multidimensional persistent homology the transparent nature of the singular theory can be of greater benefit.

Acknowledgements. The author is grateful to the referee for corrections and insightful comments and suggestions.

2. SINGULAR PERSISTENT HOMOLOGY

In order to prove the Main Theorem we use an auxiliary theory that is conceptually better suited for decompositions of metric spaces via coverings. It is defined for metric spaces that are not necessarily finite or locally finite.

Definition 2.1. The symbol Δ^n will denote the metric space with $n + 1$ points where the distance between each pair of points is 1.

Let M be a metric space and R be a principal ideal domain.

Definition 2.2. An n -dimensional singular ε -simplex $\sigma: \Delta^n \rightarrow M$ is an arbitrary set function with the diameter of the image of σ bounded by the given real number $\varepsilon \geq 0$. Let us denote the set of all such simplices by $S_n^\varepsilon(M)$ and the R -module of finite chains they generate by $C_n^\varepsilon(M)$.

It is easy to see that the usual boundary operation restricts to $C_n^\varepsilon(M)$, so we get a chain complex $C_\bullet^\varepsilon(M)$ for each value of $\varepsilon \geq 0$. This gives the discrete singular homology modules $sH_n^\varepsilon(M)$. The ranks of these modules are the singular Betti numbers $s\beta_n^\varepsilon(M)$.

Clearly the maps induced by inclusions of chain complexes $C_\bullet^\varepsilon(M) \rightarrow C_\bullet^{\varepsilon'}(M)$ for $\varepsilon \leq \varepsilon'$ give homology homomorphisms which assemble into the familiar persistent module structure for the metric space M . We will refer to the resulting functor from $[0, +\infty)$ to the category of R -modules as the n -dimensional singular persistent homology module of M . It can be again visually represented by a persistence barcode where the number of bars present at each stage ε equals the corresponding singular Betti number.

Notation 2.3. We will use the notation $psH_n(M)$ for this functor. Of course, this new functor is no longer based on a geometric filtration of a space X . Instead, it exploits an algebraic filtration of the chain complex $C_\bullet^\infty(M)$.

Definition 2.4. It is clear how to do algebra with persistence modules as functors or diagrams modeled on the ordered ray $[0, +\infty)$. We will spell out just several operations that are needed in this paper.

A map of persistence modules $h: F \rightarrow G$ is a natural transformation between F and G as functors, so it is clear what the kernel K and cokernel C of h are in the abelian category of the persistence modules over the p.i.d. R . For example, each R -homomorphism $h_\varepsilon: F_\varepsilon \rightarrow G_\varepsilon$ has a kernel K_ε . There are R -homomorphisms $K_\varepsilon \rightarrow K_{\varepsilon'}$ induced from the structure maps of F , and this structure altogether gives a persistence module K which is the kernel of h .

There is also the evident direct sum construction generated by taking direct sums of individual modules at each stage ε and inducing the natural maps between these products as colimits in the module category.

Historical remark 1. The basic idea here is not new; it goes back to Vietoris's contribution to the development of homology theories for non-polyhedral spaces between 1910 and 1928, building on the pioneering work of L.E.J. Brouwer. The history can be found in Chapter 8 and specifically section 8-6 of [15]. At the time, for example, Čech used approximations of continua by nerves of coverings, and Borsuk used embeddings of spaces in geometric models that soon became the basic idea behind shape theory. The cycles that Brouwer and Vietoris considered were built out of ordered ε -simplices which were precisely the images of injective singular simplices from Definition 2.2. Of course all of this was happening before the comprehensive development of the singular theory by Eilenberg [11] in 1944.

Historical remark 2. One may also consider infinite chains of singular simplices. A useful additional condition would require that chains are *locally finite* in the sense that each metric ball in X intersects at most finitely many simplices in the chain. This gives the submodule $C_n^{\varepsilon,lf}(M)$. Of course, when M is a finite metric space, there is no distinction between $C_n^\varepsilon(M)$ and $C_n^{\varepsilon,lf}(M)$. When X is not compact, the colimit of $C_n^{\varepsilon,lf}(M)$ is essentially the module of uniformly bounded locally finite chains that were defined and used in the work on the Novikov conjecture in K -theory by Carlsson/Goldfarb [7].

Historical remark 3. Persistent homology for non-locally finite metric spaces and finite samples from such spaces in the context of topological

data analysis has been studied in several recent papers, for example by Chazal/De Silva/Oudot [8] and Adamaszek/Adams [1].

Remark 2.5. This comment is related to the historical remark 2 above. The version of locally finite homology in [7] is in fact based on a variant of the singular simplex. One can start with a standard simplex D^n in $\mathbb{R}^{n+1}$ and build a singular theory based on maps $\sigma: D^n \rightarrow M$ which are not necessarily continuous but have the diameter of the image $\sigma(D^n)$ bounded from above by some ε . We don't include details here but it is not hard to show that the resulting theory is equivalent to the discrete singular homology. In turn, using the standard subdivision tools, it is possible to prove the following. If M is connected then for any dimension n and any value $\varepsilon \geq 0$ we have $sH_n^\varepsilon(M) = H_n(M)$, where $H_n(M)$ is the usual n -dimensional singular homology of M .

We prove in this section a crucial for computation fact about singular persistent homology.

Theorem 2.6. *If M is a finite metric space then $psH_n(M) \cong pH_n(M)$.*

The isomorphism in this theorem is an isomorphism between persistent modules, so it is an isomorphism of functors from $[0, +\infty)$ to the category of finitely generated R -modules.

Sample computations quickly make one suspect that $sH_n^\varepsilon(M)$ is isomorphic to $H_n(R_\varepsilon M)$ for each value of the parameter $\varepsilon \geq 0$. Such comparisons are usually proved in algebraic topology by using refinements which are unavailable in finite metric spaces. Nevertheless, the suspicion is correct. This is seen by viewing the singular theory as homology of certain simplicial sets.

Let us emphasize that the useful direction in this theorem is certainly the interpretation of the singular persistent homology in terms of the simplicial persistent homology which produces fewer generators on the chain level. This fact is used to reformulate the theorem we prove in the next section using the singular theory as the Main Theorem, stated entirely in terms of the simplicial theory.

Proof. First we define two simplicial sets associated to M .

One is the standard nerve of the Vietoris-Rips complex $R_\varepsilon M$ which is denoted $N(R_\varepsilon M)$. This is the simplicial set associated to the poset of the simplices in $R_\varepsilon M$. To spell out what that means, $N(R_\varepsilon M)_k$ is viewed as all simplicial maps $\text{Hom}([n], R_\varepsilon M)$ from the clique $[n]$ with $n + 1$ vertices to $R_\varepsilon M$. The structure maps are the simplicial operators $f: [m] \rightarrow [n]$ which act by pre-composition $a \rightarrow a \circ f$ for an element $a: [n] \rightarrow R_\varepsilon M$ in $N(R_\varepsilon M)_k$. It is known that the simplicial homology of a complex is isomorphic to the homology of the geometric realization of its nerve.

This isomorphism is induced by the simplicial map from $|N(R_\varepsilon M)|$ to $R_\varepsilon M$. Notice right away that this is in fact a natural transformation between persistence modules and so is an isomorphism of persistence homology modules.

The second is the simplicial set $S^\varepsilon M$ given by $(S^\varepsilon M)_k = S_k^\varepsilon(M)$ and the usual for the singular simplicial set face and degeneracy formulas. It is easily seen to be a Kan complex applying the classical proof verbatim. The homology of the geometric realization of this simplicial set is the same as the usual homology of the simplicial set built through generating simplicial R -modules and further converting to a chain complex, see section III.2 of [13]. This last construction is precisely how the singular homology $sH_n(M)$ was defined. Each $a \in \text{Hom}([n], R_\varepsilon M)$ is uniquely determined by the values on the vertices. This clearly gives the isomorphism $t_\varepsilon: N(R_\varepsilon M) \rightarrow S^\varepsilon M$ which induces isomorphisms on homology. Moreover, the induced maps for all values of ε give a natural equivalence between $[0, +\infty)$ -diagrams which is an isomorphism of persistence homology modules.

Our conclusion is that there is an isomorphism between $pH_n(M)$ and $psH_n(M)$ for all n . $\square$

3. PROOF OF THE MAIN THEOREM

Our main theorem is a Mayer-Vietoris statement in singular persistent homology. We start with a couple of lemmas.

Lemma 3.1. *For any covering $\{M_\alpha\}$ of a metric space M with a Lebesgue number greater than some nonnegative number ε , the homomorphism*

$$g_n: \bigoplus_i C_n^\varepsilon(M_\alpha) \longrightarrow C_n^\varepsilon(M)$$

induced by inclusions $M_\alpha \rightarrow M$ is surjective for all n .

Proof. The Lebesgue number assumption guarantees that every singular ε -simplex, of any dimension, lands in some member of the covering. $\square$

Lemma 3.2. *Suppose the nerve of the covering $\{M_\alpha\}$ is a simplicial tree. Choose a total ordering of the indices, so for every edge $\{\alpha, \alpha'\}$ in the edge set E there is a well defined orientation $\alpha \leq \alpha'$. Then for each edge with the orientation $\alpha \leq \alpha'$ we have the inclusion $i_\alpha: M_\alpha \cap M_{\alpha'} \rightarrow M_\alpha$ and the inclusion $j_\alpha: M_\alpha \cap M_{\alpha'} \rightarrow M_{\alpha'}$. The homomorphism*

$$f_n: \bigoplus_{\{\alpha, \alpha'\} \in E} C_n^\varepsilon(M_\alpha \cap M_{\alpha'}) \longrightarrow \bigoplus_\alpha C_n^\varepsilon(M_\alpha),$$

with the components induced by i_α and $-j_\alpha$, is injective for all n .

Proof. Since all of the double intersections are disjoint from each other, an element of the kernel forces pairwise cancellations in the images of the homomorphisms induced from $-j_\alpha$ and $i_{\alpha'}$. Finiteness of a sum representing the kernel element guarantees that all summands are 0. $\square$

Theorem 3.3. *Given a covering $\{M_\alpha\}$ of M with a Lebesgue number greater than ε and nerve a simplicial tree, there is a long exact Mayer-Vietoris sequence*

$$\begin{aligned} \dots \longrightarrow \bigoplus_{\{\alpha, \alpha'\} \in E} sH_n^\varepsilon(M_\alpha \cap M_{\alpha'}) &\xrightarrow{f_n} \bigoplus_{\alpha} sH_n^\varepsilon(M_\alpha) \xrightarrow{g_n} sH_n^\varepsilon(M) \\ &\longrightarrow \bigoplus_{\{\alpha, \alpha'\} \in E} sH_{n-1}^\varepsilon(M_\alpha \cap M_{\alpha'}) \longrightarrow \dots \end{aligned}$$

Proof. This follows from the short exact sequence of chain complexes

$$0 \longrightarrow \bigoplus_{\{\alpha, \alpha'\} \in E} C_\bullet^\varepsilon(M_\alpha \cap M_{\alpha'}) \xrightarrow{f} \bigoplus_{\alpha} C_\bullet^\varepsilon(M_\alpha) \xrightarrow{g} C_\bullet^\varepsilon(M) \longrightarrow 0$$

The kernel of g is generated by the elements $x \oplus x'$, where $x \in C_\bullet^\varepsilon(M_\alpha)$, $x' \in C_\bullet^\varepsilon(M_{\alpha'})$, and $g(x) = g(-x')$. So both $x, x' \in C_\bullet^\varepsilon(M_\alpha \cap M_{\alpha'})$, and $f = \ker(g)$. Exactness at the other two terms is obtained from the two lemmas. $\square$

Proof of the Main Theorem. Theorem 3.3 can be applied inductively to a hierarchical tree-like decomposition of depth D starting with the computation of persistent homology of primary subsets. The assumption guarantees that the primary subsets form families that are tree-like decomposition of rank ℓ of their unions. Moreover, the unions of the families form a hierarchical tree-like decomposition of M of rank ℓ and depth $D - 1$. Since ε -psH of each union can be computed using Theorem 3.3, we achieve a reduction in depth. We inductively exploit hierarchical tree-like decompositions of M and apply Theorem 3.3 to finish the computation in D steps. $\square$

Example 3.4. Given a metric space M with a λ -Lipschitz function $\phi: M \rightarrow \mathbb{R}$, there is a covering indexed by the integers with a prescribed Lebesgue number ε constructed as follows. For any value $\ell > 0$ and any real number a , consider the covering of $\mathbb{R}$ by the family of intervals $R_{a, \ell, i} = [a + i\ell, a + (i + 1)\ell + \lambda\varepsilon]$ indexed by the integers $i \in \mathbb{Z}$. The overlaps between these closed intervals have size $\lambda\varepsilon$. Then the covering of M by subsets $\phi^{-1}[a + i\ell, a + (i + 1)\ell + \lambda\varepsilon]$ has Lebesgue number ε .

There is a generalization of this construction for a λ -Lipschitz function $\phi: M \rightarrow T$ to a simplicial tree or, indeed, an $\mathbb{R}$ -tree. Select a base point t_0 in T . Then consider concentric metric annuli $A_{a, \ell, i}$ in T centered at t_0

whose points have distance from t_0 fall within $R_{a,\ell,i}$ for positive values of a and i . The inverse images along ϕ of the connected components of all $A_{a,\ell,i}$ give the desired covering of M with Lebesgue number ε .

We restate the theorem in the case that is useful for our example in the next section. Suppose that the indexing set of the covering are the integers, and the intersections $M_k \cap M_l$ are empty unless $k = l \pm 1$.

Corollary 3.5. *If the covering $\{M_k\}$ has a Lebesgue number greater than ε , there is a long exact Mayer-Vietoris sequence*

$$\begin{aligned} \dots \longrightarrow \bigoplus_k sH_n^\varepsilon(M_k \cap M_{k+1}) &\xrightarrow{f_n} \bigoplus_k sH_n^\varepsilon(M_k) \xrightarrow{g_n} sH_n^\varepsilon(M) \\ &\longrightarrow \bigoplus_k sH_{n-1}^\varepsilon(M_k \cap M_{k+1}) \longrightarrow \dots \end{aligned}$$

The homomorphisms in the sequence have explicit classical descriptions, so the computation of $sH_n^\varepsilon(M)$ and specifically the Betti number $sB_n^\varepsilon(M)$ reduces to that of $sH_n^\varepsilon(M_k \cap M_{k+1})$ and $sH_n^\varepsilon(M_k)$ for a finite number of values of k .

Corollary 3.6. *Suppose the metric space M is a finite metric space, so we can use the Vietoris-Rips filtration from Definition 1.1. In view of Theorem 2.6, all the statements in this section still hold if we simply replace the singular homology sH with the simplicial homology H applied to each stage in the filtration.*

We want to state this more explicitly in the case when the coefficient ring R is a field. What is needed is the analysis of homomorphisms f_* because by splitting the long exact sequence into short exact sequences centered around $sH_*^\varepsilon(M)$ we get the description

$$(\dagger) \quad sH_n^\varepsilon(M) \cong \operatorname{coker}(f_{n,\varepsilon}) \oplus \ker(f_{n-1,\varepsilon}).$$

This leads to an algorithm for reconstructing the ε -prefix $\varepsilon\text{-ps}H_n(M)$ of the usual persistence module when M is a finite metric space and the coefficient ring R is a field. Before we present this algorithm, we want to make an important remark.

Remark 3.7 (Sufficiency of an ε -prefix). A necessary feature of the theorems above is that the top scale ε is fixed a priori and is related to the Lebesgue number of the covering. This allows to reconstruct only the ε -prefix of the persistence module or persistence barcode.

We want to argue that there are several common situations in practice where this is not a serious disadvantage. Suppose our data comes from a well-sampled manifold or another geometric shape with tame local behavior.

Then there is a value of the parameter that is known to give a guaranteed reconstruction of homology via homology of the Vietoris-Rips complex. For example, the main theorems in section 3 of Niyogi/Smale/Weinberger [19] are of this type. Another more intrinsic situation is the setting of totally bounded metric spaces due to Chazal/De Silva/Oudot [8]. The point is that in these situations the significant persistent homological features of the data set are guaranteed to appear early, likely before the filtration stage ε . So sufficiency of the ε -prefix of the barcode will increase with better quality and density of sampled data in conjunction with the greater importance of low dimensional computations.

This is illustrated in the example of coarser and then finer samples from a circle in $\mathbb{R}^2$ shown in Figure 2. The homological features such as connectedness and the single prominent generator in 1-dimensional homology are discernible earlier for the second, finer sample. It should be clear that the better the density of the sampling the earlier the significant homological features get detected in all dimensions.

Algorithm 3.8. The following is an explicit algorithm, leveraging Corollaries 3.5 and 3.6 and implemented in Figure 3, for computing the ε -prefix $\varepsilon\text{-}pH_n(M)$ when M is a finite metric space with a 1-Lipschitz function to the real line $\mathbb{R}$. The value of ε as well as the dimension n are arbitrary and are fixed from the outset.

Step 1. Cover the real line by intervals I_i of length T with overlaps of size $\varepsilon < T/2$. We see that at most two intervals overlap, so the nerve of the covering is 1-dimensional.

Step 2. The preimages of the intervals I_i along the given 1-Lipschitz function give subspaces M_i of M with a 1-dimensional nerve. This gives M a hierarchical tree-like decomposition of rank ε and depth 1 according to Definition 1.4. In this case the tree is simply the real line. We denote the lower bound of the index i such that M_i is nonempty by L and the upper bound by U . Let us also denote the double intersections by subsets indexed by consecutive integers as $M_{i,i+1}$.

Step 3. Compute the simplicial persistence vector spaces for all pieces $\varepsilon\text{-}pH_n(M_i)$ and $\varepsilon\text{-}pH_n(M_{i,i+1})$ in parallel using one of the available packages such as **Phat** [4], **Ripser** [2], **Eirene** [14].

We stress once again that it suffices to compute the usual simplicial persistent homology using the standard packages because of the identification between simplicial and singular persistent homologies in Theorem 2.6.

Step 4. Compute $\text{coker}(f_n)$ and $\ker(f_{n-1})$, then the direct sum $\text{coker}(f_{n,\varepsilon}) \oplus \ker(f_{n-1,\varepsilon})$. This is $\varepsilon\text{-}pH_n(M)$ from the equation (†).

Kernels, cokernels, and induced maps between them can be found implemented as methods in all open source software packages listed in step 3. The same open source libraries can be leveraged to compute $\varepsilon\text{-}pH_n(M)$.

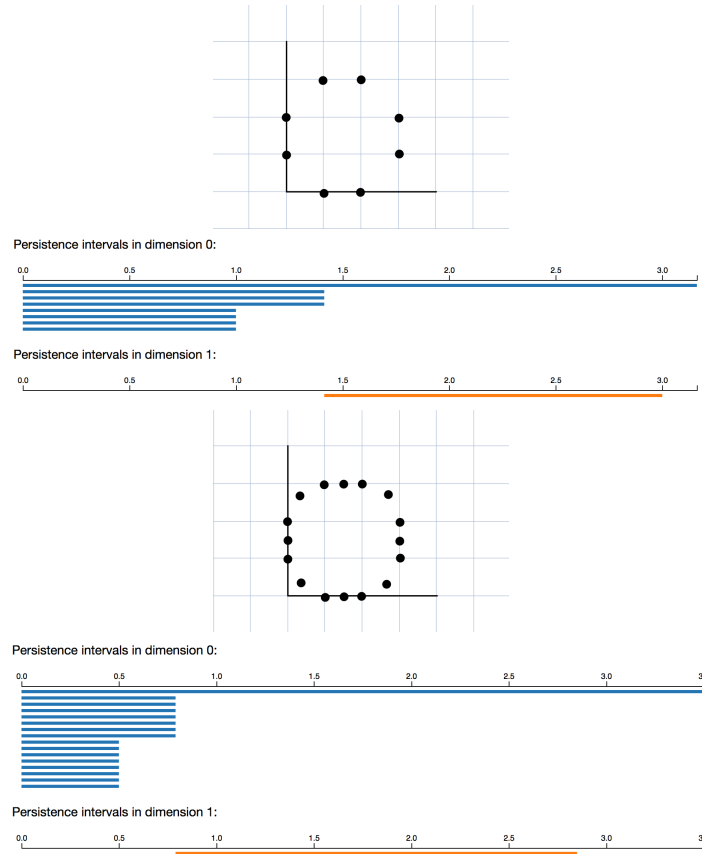


FIGURE 2. Comparison of coarser and finer samples from a circle.

The following remarks refer to some notation introduced in Figure 3.

Remark 3.9. When computing $\varepsilon\text{-}pH_n(M)$ for a range of dimensions, there is considerable redundancy between computations in adjacent dimensions. So the array $\mathbf{F}(n)$ would be exactly the same outcome as the array $\mathbf{F}((n+1)-1)$. Instead of repeating this routine for each dimension, such redundancy can be eliminated by processing the whole range of dimensions at each step and assembling the appropriate outcomes from pairs of consecutive dimensions in the last line.

```

1D-MAYER-VIETORIS( $M, n, \varepsilon$ )
1   $\mathbf{M} \leftarrow \text{1D-COVER}(M, \varepsilon)$ 
2  parallel for  $M(\bullet) \in \mathbf{M}$ 
3      do  $\varepsilon\text{-PH}(M(\bullet))$ 
4  parallel for  $i$  from  $L$  to  $U$ 
5      do  $\text{IND}(M(i, i+1) \subset M(i))$ 
6      do  $\text{IND}(M(i, i+1) \subset M(i+1))$ 
7   $\mathbf{O}\text{-double}(n) \leftarrow \text{PLUS } \{\varepsilon\text{-pH}(n)(M(i, i+1))\}$ 
8   $\mathbf{O}\text{-single}(n) \leftarrow \text{PLUS } \{\varepsilon\text{-pH}(n)(M(i))\}$ 
9   $\mathbf{O}\text{-double}(n-1) \leftarrow \text{PLUS } \{\varepsilon\text{-pH}(n-1)(M(i, i+1))\}$ 
10  $\mathbf{O}\text{-single}(n-1) \leftarrow \text{PLUS } \{\varepsilon\text{-pH}(n-1)(M(i))\}$ 
11  $\mathbf{F}(n) \leftarrow \text{BLOCK}(\mathbf{O}\text{-double}(n): \mathbf{O}\text{-single}(n))$ 
12  $\mathbf{F}(n-1) \leftarrow \text{BLOCK}(\mathbf{O}\text{-double}(n-1): \mathbf{O}\text{-single}(n-1))$ 
13  $\mathbf{K} \leftarrow \text{KER}(\mathbf{F}(n-1))$ 
14  $\mathbf{C} \leftarrow \text{COKER}(\mathbf{F}(n))$ 
15  $\mathbf{H} \leftarrow \text{PLUS}(\mathbf{K}, \mathbf{C})$ 

```

FIGURE 3. Pseudocode for computing the relevant segments of the persistent homology Mayer-Vietoris sequence from Corollary 3.5 and thus its global n -dimensional term $\varepsilon\text{-pH}_n(M)$ for a single value of n in parallel. 1D-COVER generates a cover of M by subspaces M_i and their double intersections $M_{i,i+1}$. The procedure $\varepsilon\text{-PH}$ can be any implementation generating the persistence module for $M_\bullet$ for a range of dimensions including $n-1$ and n but which is run only up to the value of the parameter ε . IND induces persistence diagram maps from inclusions. BLOCK assembles the blocks from lines 5 and 6 into the total matrix of the linear transformation f_n from $\mathbf{O}\text{-double}(n)$ to $\mathbf{O}\text{-single}(n)$. PLUS, KER, and COKER are the direct sum, kernel, and cokernel procedures described in Definition 2.4. KER, COKER and matrices for the induced maps are implemented as methods in open source software packages Phat, Ripser, Eirene, and others.

Remark 3.10 (Worst case analysis). This algorithm is most helpful for sparse and uniformly dense data. These features are common, for example, in graphics data or uniformly sampled data.

To see this point it is instructive to look at the worst case scenario which is all of the data set concentrated in one primary set M_i . In this case, there is only one instance of line 3 in Figure 3. Then line 5 is the identity map,

so eventually $\mathbf{H} = \varepsilon\text{-pH}(M(i))$. This shows that the algorithm brings no improvement to the computation of $\varepsilon\text{-pH}(M)$ in any sense. At the other extreme are the cases where the range of values in the selected feature of the data is much larger than R , and so the algorithm is able to generate many instances of parallel computations in line 3.

Remark 3.11 (Efficiency). The extension steps plus a number of reduced size parallel computations are faster and cheaper in terms of computing resources compared to the head-on computation of the full persistence barcode. We will analyze a common case in the next section.

Remark 3.12 (Feasibility). There exist geometric settings where the Mayer-Vietoris algorithm would not introduce efficiency—for example when the decomposition is modeled on a high-valence tree. Even in cases like that, the ability to parallelize the computation so that each of the barcode computations for covering subsets stays within memory bounds of the processor available for the task makes the whole computation feasible, even if not necessarily efficient.

4. AN EXAMPLE OF A HIERARCHICAL TREE-LIKE DECOMPOSITION

We include an example of a specific practical decomposition which can be applied to subsets of a Euclidean space. We aim to illustrate both efficiency and feasibility of computing persistent homology of the subset gained from the use of the Mayer-Vietoris algorithm.

It is most natural to divide multi-parameter data according to projections to subsets in one chosen parameter. The general kind of parameter for our purposes is tree-based, with partial order, but more commonly the parameter is a coordinate from the totally ordered real line. The set itself is usually a subset of a Euclidean space where the parameter is one of the Euclidean coordinates. Distributing the computation, one simplifies the computation by applying an algorithm to slices which are reduced in size compared to the total data set.

In our case this process will be performed inductively using isometric intervals in each real coordinate with overlaps that are at least ε -thick. This will guarantee that all slices used in the computation form coverings with a Lebesgue number at least ε .

Before presenting the algorithm, we need to set up a system of notations.

We assume that the given metric space M is a finite set of points embedded as a metric subspace of a Euclidean space of dimension D with the Euclidean metric. All points are stored as D -dimensional vectors. A simple search can identify the maximal difference R_j between the values of the j -th coordinate. For simplicity we will use $R = \max\{R_j\}$ but the algorithm can be fruitfully refined by adjusting to the difference in

sizes among R_j . Then M is contained in some D -dimensional hypercube $[a_1, a_1 + R] \times \dots \times [a_D, a_D + R]$.

Suppose we are given a constraint which restricts to $\leq p$ the number of computations we are able to run in parallel. The nature of this constraint can be the number of processors, or threads, or nodes available at the same time. Let $k = \lfloor p^{1/D} \rfloor$, the largest integer smaller than $p^{1/D}$. Since $p \geq 1$, this number is positive.

Of course, a larger number p will allow for a more effective parallelization scheme. The worst case analysis in Remark 3.10 corresponds to a single processor being used, with $p = 1$.

Now k , the floor of $p^{1/D}$, is the number of intervals we use in each Euclidean coordinate. Here is a prescription of the intervals in the i -th coordinate which achieves all lengths to be $T = R/k + \varepsilon$ and the only double overlaps between consecutive intervals of width ε :

$$\begin{aligned} A_{i,1} &= [a_i, a_i + R/k + \varepsilon], \\ A_{i,2} &= [a_i + R/k, a_i + 2R/k + \varepsilon], \\ &\dots \\ A_{i,k} &= [a_i + (k-1)R/k, a_i + R + \varepsilon]. \end{aligned}$$

If π_i is the projection onto the i -th coordinate, it is a 1-Lipschitz function and so produces the strips $S_{i,j} = \pi_i^{-1} A_{i,j}$ which form a covering of M with a Lebesgue number at most ε .

Given an integral vector x with coordinates from 1 to k , there are hypercubes

$$C_x = S_{1,x_1} \cap S_{2,x_2} \cap \dots \cap S_{D,x_D}.$$

We will generalize this notation as follows. Let x be an integral vector with coordinates from 0 to k and with the property that if 0 appears as a value then all subsequent coordinates must be 0. We will use $d(x)$ to denote the highest index for which the value is nonzero.

Notation 4.1. $X(s) = \{x \in \mathbb{Z}_{\geq 0}^D \mid D - d(x) = s\}$.

Now there is an extension to the previous geometric constructions. If the value of j is 0 in $S_{i,j}$, we interpret that set as the union

$$S_{i,0} = S_{i,1} \cup \dots \cup S_{i,k}.$$

Informally, index 0 indicates no constraint on the value of the i -th coordinate. So C_x has an interpretation as

$$C_x = S_{1,x_1} \cap \dots \cap S_{d,x_d}$$

where $d = d(x)$. There are also sets of the form

$$C'_x = S_{1,x_1} \cap \dots \cap S_{d,x_d} \cap S_{d,x_d+1}$$

when $x_d < k$.

Notation 4.2. $M_x = M \cap C_x$ and $M'_x = M \cap C'_x$.

What we have here is a hierarchical tree-like decomposition $\{M_x\}$ of M of rank ε and depth D . The intersections $\{M'_x\}$ correspond to edges that appear in this scheme. As part of the decomposition, this structure gives a system of 1-dimensional coverings of each M_x for $x \in X(s)$, $1 \leq s \leq D$.

Notation 4.3. For $1 \leq s \leq D$ and any $x \in X(s)$ there is a covering

$$\{M_x \cap S_{D-s+1,i} \mid 1 \leq i \leq k\}$$

of M_x which we denote by $\mathbf{M}(x)$. Another useful way to describe this covering is as $\{M_y\}$, where

$$y = (x_1, \dots, x_{D-s}, i, 0, \dots, 0)$$

for $1 \leq i \leq k$. When y is as above, we will use the notation

$$y + 1 = (x_1, \dots, x_{D-s}, i + 1, 0, \dots, 0).$$

Example 4.4. The coverings are illustrated in Figure 4. Altogether this is an example of a hierarchical tree-like decomposition of depth $D = 2$, rank ε equal to the width of double intersections between cubical sets, and the choice of $k = 3$. Corresponding to $s = 1$, we have three possible values of $x \in X(1)$ listed across the bottom of the picture. The sets C_x for these x are the blue shaded rectangles. For a data set M contained entirely inside the total region, we have the corresponding subsets $M_x = C_x \cap M$ which form a covering of M . This is a 1-dimensional covering with Lebesgue number ε . Now let's choose one of these values, say $x = (2, 0)$. Corresponding to $s = 2$, there are three possible values of y which restrict to x on the first $d - s$ coordinates, so $\mathbf{M}(2, 0)$ is the covering of $M_{(2,0)}$ consisting of sets $M_{(2,1)}$, $M_{(2,2)}$, $M_{(2,3)}$ as listed in Notation 4.3. These are the primary cubical sets colored in red. Again, this is a 1-dimensional covering with the Lebesgue number ε .

Algorithm 4.5. We start with a verbose description, then present a pseudocode.

Step 1. Compute the simplicial persistence vector spaces for all pieces $\varepsilon\text{-}pH_n(M_x)$, $\varepsilon\text{-}pH_n(M'_x)$, $\varepsilon\text{-}pH_{n-1}(M_x)$, and $\varepsilon\text{-}pH_{n-1}(M'_x)$ for x involving no zeros. The computations can be done in parallel and distributed among available CPUs. This is possible because there are at most p different options for x , so p is a sufficient number of processors to perform this step for M_x and then for M'_x .

Step 2. Consider all options for vectors x where the D -th coordinate is 0. Each M_x and M'_x is a union of appropriate k subsets from Step 1 and their $k - 1$ nonempty intersections. Clearly this covering of M_x and M'_x inherits the Lebesgue number bounded by ε . Applying essentially the procedure 1D-MAYER-VIETORIS we can build $\varepsilon\text{-}pH_n(M_x)$ and $\varepsilon\text{-}pH_n(M'_x)$.

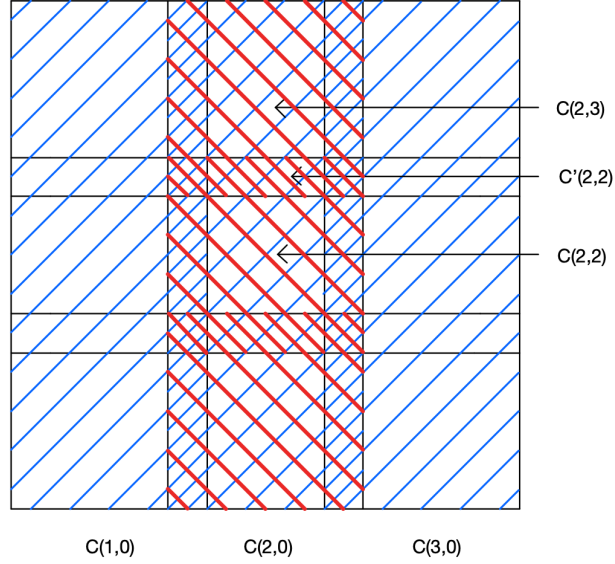


FIGURE 4. A hierarchical tree-like decomposition of depth 2.

Step S (for $3 \leq S \leq k$). We proceed inductively applying Step 2 verbatim except restricting to vectors x with $S - 1$ last coordinates 0. These M_x and M'_x have 1-dimensional coverings by sets from Step $S - 1$ with Lebesgue number ε , so the same procedure can be used to compute $\varepsilon\text{-}pH_n(M_x)$ and $\varepsilon\text{-}pH_n(M'_x)$. In the last step, that means computing $\varepsilon\text{-}pH_n(M_x)$ for the 0-vector x , that is $\varepsilon\text{-}pH_n(M)$.

The issue of time complexity is different from the issue that is often more important in persistence computations. Data sets of reasonable size make programs like MATLAB run out of resources even on high-end computers. Parallelizing the persistence algorithm and distributing the computation among a network of computers with average parameters make possible practical computations that are not even feasible directly, cf. Remark 3.12.

5. DISCUSSION

The major point in this paper is that it is possible to assemble parallel homology computations into partial but often crucial information about the persistence module of a metric space M . The quality of the answer is a decision to be made in the beginning of the computation, but the higher desired quality correlates with greater demands on the number and/or capacity of parallel processors.

```

D-DEPTH-HIERARCHICAL-ASSEMBLY( $M, n, \varepsilon, D, p$ )
1  build  $X(0)$  as in 4.1
2  parallel for  $x \in X(0)$ 
3      do  $\varepsilon\text{-PH}(M(x))$ 
4  parallel for  $x \in X(0)$ 
5      do  $\varepsilon\text{-PH}(M'(x))$ 
6  for counter = 1 to  $D$ 
7      do build  $X(\text{counter})$  as in 4.1
8      do parallel for  $x \in X(\text{counter})$ 
9          build  $M(x)$  as in 4.3
10          $M \leftarrow M(x)$ 
11         evaluate  $\varepsilon\text{-PH}(n)\text{-IND}(M'(y) \subset M(y))$ 
12         evaluate  $\varepsilon\text{-PH}(n)\text{-IND}(M'(y) \subset M(y+1))$ 
13         for dim = 0 to  $n$ 
14              $\mathbf{O}\text{-double}(n) \leftarrow \text{PLUS } \{\varepsilon\text{-pH}(\text{dim})(M'(y))\}$ 
15              $\mathbf{O}\text{-single}(n) \leftarrow \text{PLUS } \{\varepsilon\text{-pH}(\text{dim})(M(y))\}$ 
16              $\mathbf{F}(n) \leftarrow \text{BLOCK}(\mathbf{O}\text{-double}(\text{dim}); \mathbf{O}\text{-single}(\text{dim}))$ 
17              $\mathbf{K} \leftarrow \text{KER}(\mathbf{F}(\text{dim}-1))$  if  $\text{dim} \geq 1$ 
18              $\mathbf{K} \leftarrow 0$  if  $\text{dim} = 0$ 
19              $\mathbf{C} \leftarrow \text{COKER}(\mathbf{F}(\text{dim}))$ 
20              $\mathbf{H} \leftarrow \text{PLUS}(\mathbf{K}, \mathbf{C})$ 
21             save  $\varepsilon\text{-pH}(\text{dim})(M(x)) \leftarrow \mathbf{H}$ 
22 save  $\varepsilon\text{-pH}(n)(M) \leftarrow \varepsilon\text{-pH}(n)(M(x))$ 

```

FIGURE 5. Pseudocode for a parallel computation of the n -dimensional persistent homology term $\varepsilon\text{-pH}_n(M)$ for a finite data set M in $\mathbb{R}^D$ and a single value of n using p processors. The time complexity is that of the procedures in lines 3 and 5 which is known to be cubic in the number of simplices in the Vietoris-Rips complexes that get generated by primary subsets. The complexity of all subsequent processes is sub-cubic. The primary sets are the multi-cubes of volume $(R/k)^D$ for $k = \lfloor p^{1/D} \rfloor$. In other words, the reduction of complexity compared to direct persistence computation is proportional to the number p of available processors.

The stress here is meant to be on the term “homology”. This is distinct from virtually every practical attempt in the literature to parallelize the computation of persistent homology.

There are surely other attempts to parallelize the computation of persistent homology. They range from general descriptions of strategy to

methods that require special arrangements different from our Lebesgue number conditions. We will briefly survey them and compare.

Bauer, Kerber, and Reininghaus [4, 5] are concerned with distributing the matrix algebra involved in the homology computations and optimizations specific to persistence and so are very much transverse to the geometric decompositions in this paper. These methods are implemented and are part of the toolbox in modern software packages such as **Phat** [4], **Ripser** [2], **Eirene** [14].

Di Fabio and Landi [12] work on the level of Betti numbers and detect errors in a Mayer-Vietoris formula for ranks of homology. They observe that these estimates can be used in partial matching problems.

Lewis and Zomorodian [16] and Lewis and Morozov [17] use a very different, hierarchical approach to geometric decomposition of the metric space and use the blow-up complex construction for assembling the data. They build on the work of Carlsson and Zomorodian [21] which introduced the blow-up complex for an effective decomposition of the metric space. This is probably the closest in spirit to what is in this paper. Yet, just as in Bauer/Kerber/Reininghaus [4, 5], the decompositions are used for parallelizing matrix computations of homology rather than the homology computations themselves as we do in this paper. Another difference is a set of geometric constraints on the coverings that are left by these authors for the user to construct. The problem of finding balanced minimal blowups is NP-hard, and in all cases the resulting multicore algorithm uses a lot of preprocessing. In contrast, our scheme can be rather canonical and depends only on a specific pattern of overlaps. We illustrated the scheme with a covering that is always available for our purposes in subsets of a Euclidean space and useful in large well-sampled data sets.

Lipsky, Skraba, and Vejdemo-Johansson [18] is a very general discussion of how a spectral sequence computation would proceed in favorable situations, organized according to increasing dimension of the Vietoris-Rips complexes. The algorithmic issues are not addressed.

There is a paper of Zomorodian [20] that constructs a simplicial set called the “tidy set”. This set and its simplifications are used as preprocessing steps for homology computations. There is no direct relation of that simplicial set to singular homology, and it’s unlikely the reductions described by Zomorodian would be useful for singular simplicial sets.

REFERENCES

- [1] M. Adamaszek and H. Adams, *The Vietoris-Rips complexes of a circle*, Pacific J. Math. **290** (2017), 1–40.
- [2] U. Bauer, *Ripser*, 2018. Software available at <https://github.com/Ripser/ripser>.

- [3] U. Bauer, M. Kerber, and J. Reininghaus, *Distributed computation of persistent homology*, in Proceedings of the Meeting on Algorithm Engineering & Experiments, SIAM (2014), 31–38.
- [4] U. Bauer, M. Kerber, J. Reininghaus, and H. Wagner, *PHAT: Persistent homology algorithms toolbox*, in Mathematical Software, ICMS 2014, H. Hong and C. Yap, eds., Lecture Notes in Computer Science **8592** (2014), 137–143. Software available at <https://code.google.com/p/phat/>.
- [5] ———, *Clear and Compress: computing persistent homology in chunks*, in Topological Methods in Data Analysis and Visualization III (2014), 103–117.
- [6] G. Carlsson, *Topological pattern recognition for point cloud data*, Acta Numerica **23** (2014), 289–368.
- [7] G. Carlsson and B. Goldfarb, *The integral K -theoretic Novikov conjecture for groups with finite asymptotic dimension*, Invent. Math. **157** (2004), 405–418.
- [8] F. Chazal, V. De Silva, and S.Y. Oudot, *Persistence Stability for Geometric complexes*, Geometriae Dedicata **173** (2014), 193–214.
- [9] H. Edelsbrunner and J. Harer, *Persistent homology— a survey*, Contemp. Math. **453** (2008), 257–282.
- [10] H. Edelsbrunner and S. Parsa, *On the computational complexity of Betti numbers: reductions from matrix rank*, Proc. of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms (2014), 152–160.
- [11] S. Eilenberg, *Singular homology theory*, Annals of Math. **45** (1944), 407–447.
- [12] B. Di Fabio and C. Landi, *A Mayer-Vietoris formula for persistent homology with an application to shape recognition in the presence of occlusions*, Found. Comput. Math. **11** (2011), 499–527.
- [13] P. Goerss and J. Jardine, *Simplicial homotopy theory*, Birkhäuser Verlag, Basel, 2009.
- [14] G. Henselman, *Eirene*, 2018. Software available at <http://gregoryhenselman.org/eirene/>.
- [15] J.G. Hocking and G.S. Young, *Topology*, Addison-Wesley, 1961.
- [16] R. Lewis and A. Zomorodian, *Multicore homology via Mayer-Vietoris*, preprint, 2014; available as [arXiv:1407.2275](https://arxiv.org/abs/1407.2275)
- [17] R. Lewis and D. Morozov, *Parallel computation of persistent homology using the blowup complex*, Proc. of the Annual Symposium on Parallelism in Algorithms and Architectures (2015), 323–331.
- [18] D. Lipsky, P. Skraba, and M. Vejdemo-Johansson, *A spectral sequence for parallelized persistence*, preprint, 2011; available as [arXiv:1112.1245](https://arxiv.org/abs/1112.1245)
- [19] P. Niyogi, S. Smale, and S. Weinberger, *Finding the homology of submanifolds with high confidence from random samples*, Discrete Comput. Geometry **39** (2008), 419–441.
- [20] A. Zomorodian, *The Tidy Set: A Minimal Simplicial Set for Computing Homology of Clique Complexes*, Proceedings of the 26th ACM Symposium on Computational Geometry, Snowbird, Utah (2010), 257–266.
- [21] A. Zomorodian and G. Carlsson, *Localized homology*, Comput. Geom. **41** (2008), 126–148.

DEPARTMENT OF MATHEMATICS AND STATISTICS, STATE UNIVERSITY OF NEW YORK,
ALBANY, NY 12222, USA
Email address: bgoldfarb@albany.edu