Geometric Localization of Homology Cycles

Amritendu Dhar* Vijay Natarajan† Abhishek Rathod‡

Abstract

Computing an optimal cycle in a given homology class, also referred to as the homology localization problem, is known to be an NP-hard problem in general. Furthermore, there is currently no known optimality criterion that localizes classes geometrically and admits a stability property under the setting of persistent homology. We present a geometric optimization of the cycles that is computable in polynomial time and is stable in an approximate sense. Tailoring our search criterion to different settings, we obtain various optimization problems like optimal homologous cycle, minimum homology basis, and minimum persistent homology basis. In practice, the (trivial) exact algorithm is computationally expensive despite having a worst case polynomial runtime. Therefore, we design approximation algorithms for the above problems and study their performance experimentally. These algorithms have reasonable runtimes for moderate sized datasets and the cycles computed by these algorithms are consistently of high quality as demonstrated via experiments on multiple datasets.

1 Introduction

Homology groups and their persistent version called *persistent homology* play a central role in topological data analysis (TDA), a thriving research field of equal interest to computer scientists, mathematicians and data scientists [17, 18]. The ranks for homology groups and the barcodes for persistent homology groups have been extensively studied both from algorithmic and mathematical perspectives. With the growth of TDA in applications, there is an increasing need for computing homology cycles that localize given homology classes or constitute a basis for the homology group. Often applications require these cycles to be tightest possible or geometry-aware in some sense rather than being completely oblivious of the embedding space. This demand has led to studying homologous or basis cycles under various optimization criteria. A number of optimization results in this direction have now appeared in the literature both in persistent and non-persistent settings [4, 5, 7, 8, 10–13, 16, 27].

The quality of the optimal cycles depends on the choice of a weight function. For instance, one may choose a weight for each p-cycle ζ to be the sum of non-negative weights assigned to each p-simplex in ζ. Optimizing this measure over a class of a given cycle ζ localizes the class $[\zeta]$ in the sense that it selects a cycle in the class with the least weight. Unfortunately, this problem is known to be NP-hard in general [7, 10] except for some special cases [13, 15]. Polynomial time algorithms are known for certain optimization criteria [9, 15] or in lower dimensions [4, 7, 16, 20].

Outline and Contributions. Precisely, we achieve the following. Given a simplicial complex K with the vertices in a point set $P \subset \mathbb{R}^d$ and linearly embedded simplices, we define the weight of a cycle ζ as the radius of the smallest $(d-1)$-sphere that encloses ζ. This measure, in some sense, captures the locality of ζ with respect to its geometry. In Section 4, we study how homology localization serves as an archetype application. Then, we solve other versions of the optimal cycle problem including *minimum homology basis* in Section 5 and *minimum persistent homology basis* in Section 6. For the persistent version, in Appendix B, we show optimal persistent homology bases are stable in an approximate sense. For previous results on optimal persistent cycles [11,15] such stability is not known. The approximation algorithms described in this paper have been implemented. In Section 7, we report experimental results for the approximate algorithms. In our experiments, we found that even the approximate algorithms return cycles of consistently high quality confirming the value of ℓ_2–radius as an optimization criterion. We further compare experimental results on persistent homology with that of PersLoop [28], which is a state of the art software for computing optimal persistent 1-cycles. We visually infer that our cycles are "tighter" than those of PersLoop on multiple datasets of practical importance.

Related work. A criterion related to ours was considered by Chen and Freedman [9] who proposed to compute a minimum homology basis while optimizing

*Department of Computer Science and Automation, Indian Institute of Science, Bangalore, `amritendud@iisc.ac.in`

†Department of Computer Science and Automation, Indian Institute of Science, Bangalore and Zuse Institute, Berlin, `vijayn@iisc.ac.in`

‡Department of Computer Science, Ben Gurion University of the Negev, `arathod@post.bgu.ac.il`

the shortest path radius of the geodesic balls containing the basis cycles. With an embedding in the Euclidean space, the ℓ_2–radius of the geometric balls capture locality more concisely than the shortest path radius. Yet another measures of optimality for cycles that is tractable, namely lexicographically optimality [11], suffers from the drawback that it requires a parameter: a total order on simplices. In applications, it is sometimes desirable that the optimal cycles be stable with regard to the change in the input data [2]. An optimization criterion that is geometry-aware, polynomial time computable, and results in some kind of stability is *volume optimal cycles* by Obayashi [25, 26]. However, unlike our measure, the approach described in [25] works only for computing representatives of finite bars. In another related work, Li et al. [24] obtain minimal representatives using linear programming for a variety of optimization criteria with impressive runtimes. However, their software does not work for arbitrary filtrations yet [22]. In summary, our key contribution in this work is that we introduce a natural measure of optimality of cycles that has good theoretical properties and is well-behaved in practice.

2 Background and preliminaries

In this section, we recall some preliminaries on persistent homology. For the rest of this section, we work only with simplexwise filtrations: That is, we have a filtration $\mathcal{F}$ on $\mathbb{R}$ where the complexes change only at finite set of values $a_1 < a_2 < \ldots < a_n$ and every change involves addition of a unique simplex σ_{a_i} for $i \in [n]$.

$$\mathcal{F} : \emptyset = \mathsf{K}_{a_0} \overset{\sigma_{a_0}}{\hookrightarrow} \mathsf{K}_{a_1} \overset{\sigma_{a_1}}{\hookrightarrow} \mathsf{K}_{a_2} \overset{\sigma_{a_2}}{\hookrightarrow} \ldots \overset{\sigma_{a_{n-1}}}{\hookrightarrow} \mathsf{K}_n = \mathsf{K}$$

Using p-th homology groups of the complexes over the field $\mathbb{Z}_2$, we get a sequence of vector spaces connected by inclusion-induced linear maps:

$$H_p\mathcal{F} : H_p(\mathsf{K}_{a_0}) \to H_p(\mathsf{K}_{a_1}) \to H_p(\mathsf{K}_{a_2}) \to \ldots$$

The sequence $H_p\mathcal{F}$ with the linear maps is called a *persistence module*. There is a special persistence module called the *interval module* $\mathbb{I}^{[b,d)}$ associated to the interval $[b, d)$. Denoting the vector space indexed at $a \in \mathbb{R}$ as $\mathbb{I}_a$, this interval module is given by

$$\mathbb{I}^{[b,d)}_a = \begin{cases} \mathbb{Z}_2 & \text{if } a \in [b, d) \\ 0 & \text{otherwise} \end{cases}$$

together with identity maps $\mathrm{id}_{a,a'} : \mathbb{I}^{[b,d)}_a \to \mathbb{I}^{[b,d)}_{a'}$ for all $a, a' \in [b, d)$ with $a \le a'$.

It is known due to a result of Gabriel [21] that a persistence module defined with finite complexes admits a decomposition

$$H_p\mathcal{F} \cong \bigoplus_{\alpha} \mathbb{I}^{[b_\alpha,d_\alpha)}$$

which is unique up to isomorphism and permutation of the intervals. The intervals $[b_\alpha, d_\alpha)$ are called the *bars*. The multiset of bars forms the *barcode* of the persistence module $H_p\mathcal{F}$, denoted by $\mathcal{B}_p(\mathcal{F})$. The following two definitions are taken from [14].

Definition 1. For an interval $[b, d)$, we say that ζ is a *representative cycle* for $[b, d)$, or simply ζ *represents* $[b, d)$, if one of the following holds:

- $d \ne +\infty$, ζ is a cycle in K_b containing σ_b, and ζ is not a boundary in K_{d-1} but becomes one in K_d.
- $d = +\infty$, and ζ is a cycle in K_b containing σ_b.

Definition 2 (Persistent cycles)**.** A p-cycle ζ that represents an interval $[b, d) \in \mathcal{B}_p(\mathcal{F})$ is called a *persistent p-cycle* for $[b, d)$.

For a bar $[b, d)$, σ_b is said to be a *creator simplex* and σ_d is called a *destroyer simplex*.

It is easy to check that if ζ is a representative cycle for $[b_i, d_i)$ and ξ is a representative cycle for $[b_j, d_j)$, where $b_j < b_i$ and $d_j < d_i$, then $\zeta + \xi$ is also a representative cycle for $[b_i, d_i)$. The set of representative cycles for interval $[b_i, d_i)$ is denoted by $\mathcal{R}([b_i, d_i))$. Representatives of bars of the form $[b, \infty)$ are called *essential cycles*.

Definition 3 (Persistent basis)**.** Let J be the indexing set for the intervals in the barcode $\mathcal{B}_p(\mathcal{F})$ of filtration $\mathcal{F}$. That is, for every $j \in J$, $[b_j, d_j)$ is an interval in $\mathcal{B}_p(\mathcal{F})$. Then a set of p-cycles $\{\zeta_j \mid j \in J\}$ is called a *persistent p-basis* for $\mathcal{F}$ if

$$H_p\mathcal{F} = \bigoplus_{j\in J} \mathbb{I}^{\zeta_j} \text{ where } \mathbb{I}^{\zeta_j} \text{ is defined by}$$

$$\mathbb{I}^{\zeta_j}_a = \begin{cases} [\zeta_j] & \text{if } a \in [b_j, d_j) \\ 0 & \text{otherwise.} \end{cases}$$

Here, for every $j \in J$ and every $a, a' \in [b_j, d_j)$ with $a \le a'$ the maps $\mathbb{I}^{\zeta_j}_a \to \mathbb{I}^{\zeta_j}_{a'}$ are the induced maps on homology restricted to $[\zeta_j]$, respectively.

The following theorem by Dey et al. [14] relates persistent cycles to persistent bases.

Theorem 1 ([14, Theorem 1])**.** *Let J be the indexing set for the intervals in the barcode $\mathcal{B}_p(\mathcal{F})$ of filtration $\mathcal{F}$. Then, an indexed set of p-cycles $\{\zeta_j \mid j \in J\}$ is a persistent p-basis for a filtration $\mathcal{F}$ if and only if $\zeta_j \in \mathcal{R}([b_j, d_j))$ for every $j \in J$.*

3 The ℓ_2–radius metric

Given a complex K, let $Z_p(\mathsf{K})$ denote its p-th cycle group, $B_p(\mathsf{K})$ its p-th boundary group and $H_p(K)$ its p-th homology group with $\mathbb{Z}_2$ coefficients. Given a cycle ζ, our goal is to define a non-negative weight function $w : Z_p(\mathsf{K}) \to \mathbb{R}^+$ on the cycles in $Z_p(\mathsf{K})$ and compute a minimum-weight (optimal) cycle ζ^* in its homology class $[\zeta]$, that is,

$$\zeta^* \in \arg\min_{\hat{\zeta}\in[\zeta]} w(\hat{\zeta}). \tag{1}$$

We show that the ℓ_2–radius is an alternative natural geometric objective function defined on cycles that guarantees tractability. Let P be the vertex set of K. Then, K_V denotes the subcomplex of K induced by a subset $V \subseteq P$. Extending this notation, we say a complex is *induced* by a sphere $S_{c,r}$ if it is induced by the subset of vertices of P that are enclosed by $S_{c,r}$ (including on the sphere). Let the complex induced by $S_{c,r}$ be denoted as $\mathsf{K}_{c,r}$. We define a weight function $r : C_p(\mathsf{K}) \to \mathbb{R}^+$, $\xi \mapsto r(\xi)$ where

$$r(\xi) = \min_{c,\delta}\{\delta \mid \xi \in C_p(\mathsf{K}_{c,\delta})\}. \tag{2}$$

In words, $r(\xi)$ is the radius of the smallest Euclidean sphere whose induced complex in K contains ξ.

We now define an ℓ_2–radius measure for intervals in a barcode. For an interval $[b, d) \in \mathcal{B}_p(\mathcal{F})$, we define $r([b, d))$ as the radius of the smallest sphere that encloses a subset of vertices V of K_b that induces a subcomplex $\mathsf{K}_b^V \subset \mathsf{K}_b$, which supports a representative cycle for $[b, d)$. Equivalently,

$$r([b, d)) = \min_{\zeta\in\mathcal{R}([b,d))} r(\zeta), \tag{3}$$

where in Equation (3), the radius function r is restricted to the subcomplex $\mathsf{K}_b \subset \mathsf{K}$.

4 Computing optimal homologous cycle

Following Equation (1), we define an optimal cycle ζ^* in the class $[\zeta]$ by requiring $\zeta^* \in \arg\min_{\hat{\zeta}\in[\zeta]} r(\hat{\zeta})$. The cycle ζ^* represents an optimal localization of the class $[\zeta]$ with respect to the ℓ_2–radius. We consider the following Optimal Homologous Cycle problem:

> Given an p-cycle $\zeta \in Z_p(\mathsf{K})$, compute an optimal cycle ζ^* in $[\zeta]$ and $r(\zeta^*)$.

Remark 4.1. To compute the optimal homologous cycle, it is sufficient to look at the minimum circumspheres of all k-subsets of points $P = V(\mathsf{K})$, where $k \in \{2, \ldots, d + 1\}$, and check if the circumsphere encloses a cycle homologous to the input cycle. When the dimension d of the complex K is fixed, the search terminates in polynomial time. This describes a trivial exact algorithm which was found to be too expensive in our experiments in spite of polynomial time complexity.

Remark 4.2. By restricting the centers of the spheres in Equation (1) to the the sites $P = V(\mathsf{K})$ (vertices of K) yields a 2-approximation of ℓ_2–radius as follows: let S be a sphere that minimizes ℓ_2–radius of a chain ξ and let v be a vertex on S. Then, a sphere of twice the optimal radius centered at v encloses S, and therefore also encloses ξ. We define $r_c(\xi) = \min\{\delta \mid \xi \in C_p(\mathsf{K}_{c,\delta})\}$ and $r_P(\xi) = \min_{c\in P,\delta}\{\delta \mid \xi \in C_p(\mathsf{K}_{c,\delta})\}$.

Notations and Conventions. The notations and conventions described are common to all the problems in the paper. In our algorithms, a cycle (or a chain) ζ is represented by a 0–1 vector in the standard chain basis. That is, a p-cycle ζ is represented by a vector ζ where $\zeta[i] = 1$ ($\zeta[i] = 0$) if a p-simplex σ_i is (not) in the support of ζ. We often use cycle vectors of subcomplexes in computations involving cycles and boundaries of larger complexes. To ensure that we are working with vectors/matrices of the right dimensions, we make the following adjustment. For complexes $\mathsf{L} \subset \mathsf{K}$, the inclusion map $\mathsf{L} \hookrightarrow \mathsf{K}$ induces maps $Z_p(\mathsf{L}) \to Z_p(\mathsf{K})$ for every p. A cycle ξ in L is mapped to a cycle $\overline{\xi}$ in K with $\overline{\xi}[i] = \xi[i]$ for simplices $\sigma_i \in \mathsf{L}$, and $\overline{\xi}[i] = 0$ for simplices $\sigma_i \in \mathsf{K} \setminus \mathsf{L}$ (using standard chain basis). Likewise, a matrix $\mathbf{M}$ of cycle vectors of L can be treated as a matrix of cycle vectors $\overline{\mathbf{M}}$ of K by padding zeros in the rows corresponding to the simplices in $\mathsf{K} \setminus \mathsf{L}$. We call such cycle vectors $\overline{\xi}$ and matrices $\overline{\mathbf{M}}$, the *extensions* of ξ and $\mathbf{M}$ in K.

Let K be a simplicial complex, $\mathsf{K} \subset \mathbb{R}^d$. For any $v \in \mathbb{R}^d$ we can define a total ordering $\prec_v$ on the simplices of K as follows. If σ_1 is a face of σ_2 or $r_v(\sigma_1) < r_v(\sigma_2)$, then $\sigma_1 \prec_v \sigma_2$. Otherwise (when $r_v(\sigma_1) = r_v(\sigma_2)$ and σ_1 is neither a face or coface of σ_2), ties are arbitrarily broken. If $\zeta = \sigma_1 + \ldots + \sigma_s$ such that $\sigma_1 \prec_v \ldots \prec_v \sigma_s$, then we define $\kappa(\zeta) = \sigma_s$. Further, we extend this ordering to chains as follows: If $\zeta_1, \zeta_2 \in C_p(\mathsf{K})$ such that $\zeta_1 = \sigma_1 + \ldots + \sigma_s$, $\zeta_2 = \sigma'_1 + \ldots + \sigma'_{s'}$ with $\sigma_1 \prec_v \ldots \prec_v \sigma_s$ and $\sigma'_1 \prec_v \ldots \prec_v \sigma'_{s'}$, then $\zeta_1 \prec_v \zeta_2$ if $\sigma_s \prec_v \sigma'_{s'}$ i.e. $\kappa(\zeta_1) \prec_v \kappa(\zeta_2)$. Note that $r_v(\zeta) = r_v(\kappa(\zeta))$. The ordering $\prec_v$ induces a simplex-wise filtration on K which we denote by $\mathcal{D}_v(\mathsf{K})$.

The standard reduction algorithm [3] is used in many of our algorithms as subroutines. For completeness, we present an outline of the algorithm and recall some facts arising out of it in Appendix C. Algorithm 1 relies on the following proposition (Proof in Appendix C).

Proposition 4. *Let $\zeta_1 \prec_v \ldots \prec_v \zeta_s$ be the essential $p-$cycles of $\mathcal{D}_v(\mathsf{K})$ computed using standard reduction. Let* ζ be a p-cycle, $[\zeta] \neq 0 \in H_p(\mathsf{K})$, *such that* $\zeta = \zeta_{i_1} +$

$\ldots + \zeta_{i_m} + \partial c_{p+1}$ *where each* $\zeta_{i_k} \in \{\zeta_1, \ldots, \zeta_s\}$ *and* c_{p+1} *is a* $p+1$ *chain. If* $i_1 < \ldots < i_m$ *then* $r_v([\zeta]) = r_v(\zeta_{i_m})$. *In particular,* $\zeta_{i_1} + \ldots + \zeta_{i_m} \in \arg\min_{\xi \in [\zeta]} r_v(\xi)$.

We now describe a 2-approximation algorithm for Optimal Homologous Cycle for an input cycle ζ by optimizing with respect to r_P. For each site v, the algorithm invokes the subroutine optimal-hom-cycle-forsite (Line 15 of Algorithm 1) which computes $r_v([\zeta]) = \min_{\eta \in [\zeta]}\{r_v(\eta)\}$ and $\zeta_v^* \in \arg\min_{\eta \in [\zeta]}\{r_v(\eta)\}$. Finally it reports the minimum among all sites and the corresponding optimal homologous cycle. Procedure optimal-hom-cycle-forsite is motivated by Proposition 4. It first sorts the simplices of K based on distance from v. The ordering is monotonic, that is, faces gain precedence over cofaces.

In this way the ordering $\prec_v$ and hence the filtration $\mathcal{D}_v(\mathsf{K})$ is defined. Let $\zeta_1 \prec_v \ldots \prec_v \zeta_m$ be the essential $p-$cycles of $\mathcal{D}_v(\mathsf{K})$ computed using standard reduction. As noted before we consider cycle vectors to represent the cycles. To compute the linear combination of cycles $\{\zeta_i\}$ which is homologous to ζ, we solve for the system of equations $[\zeta_1 \ldots \zeta_s \,|\, B_p(\mathsf{K})].x = \zeta$. (We invoke subroutine SolveByReduction which solves $Ax = b$ over $\mathbb{Z}_2$, using standard reduction as a subroutine. Refer to Appendix C, Algorithm 5 for definition of this routine). If $i_1, \ldots i_s, j_1 \ldots, j_t$ is a solution where indices $i_1, \ldots i_s$ correspond to cycles in $\{\zeta_i\}_{i=1}^m$) and $j_1 \ldots j_t$ correspond to boundaries in B_p, then by Proposition 4 $\zeta_{i_1} + \ldots + \zeta_{i_s} \in \arg\min_{\eta \in [\zeta]\{r_v(\eta)\}}$.

Remark 4.3. Algorithm 1 runs in $O(|P|N^3)$, where N is the number of simplices in K.

Algorithm 1: Computing optimal homologous cycle for given set of sites

Input : $\mathsf{K}, \zeta \in C_p(\mathsf{K})$
Output: $r_P([\zeta]), \zeta^*$ (Optimal homologous cycle)

```
1  Procedure OptHomologousCycle
2      r_P([ζ]) ← ∞, ζ* ← ∅
3      for v ∈ P do
4          r_v([ζ]), ζ*_v ← OPTIMAL-HOM-CYCLE-FORSITE(v)
5          If r_v([ζ]) is less than the current value of r_P([ζ]), then update r_P([ζ]) with r_v([ζ]) and ζ* with ζ*_v
6  Procedure Optimal-Hom-Cycle-ForSite(v)
7      ▷ Description: Computes r_v([ζ]) = min_{η∈[ζ]} r_v(η), ζ*_v ∈ arg min_{η∈[ζ]} r_v(η)
8      Define ≺_v on K. Compute D_v(K)
9      Compute the essential cycles of D_v(K) by standard reduction. Let ζ_1, ..., ζ_m be essential cycles ordered with respect to ≺_v.
10     Compute the p^th boundary matrix of K, denote it by B_p.
11     Assemble matrix ∂ = [ζ_1, ..., ζ_m | B_p]
12     Solve ∂.x = ζ. Invokes( SOLVEBYREDUCTION(∂, ζ)). Let i_1, ..., i_s, j_1, ..., j_t be the solution where i_1, ..., i_s ≤ m (indices that correspond to cycles in {ζ_i}_{i=1}^m) and j_1 ... j_t > m ( indices correspond to boundaries in B_p.)
13     ζ*_v ← ζ_{i_1} + ... + ζ_{i_s}.
14     r_v([ζ]) ← r_v(ζ*_v)
15     Return r_v([ζ]), ζ*_v
```

5 Optimal homology basis

A set of p-cycles $\{\zeta_1, \ldots, \zeta_\beta\}$ ($\beta_p = \beta_p(\mathsf{K})$) is called a *homology cycle basis* if the set of classes $\{[\zeta_1], \ldots, [\zeta_\beta]\}$ forms a basis for $H_p(\mathsf{K})$. For simplicity, we use the term *homology basis* to refer to the set of cycles $\{\zeta_1, \ldots, \zeta_\beta\}$.

Definition 5. A $p-$homology basis $\{\zeta_1, \ldots, \zeta_m\}$ will be called a minimum $p-$homology basis ($p > 0$) with respect to a non-negative weight function $w : Z_p(K) \to \mathbb{R}$, if for all $p-$homology bases $\zeta_1' \ldots \zeta_m'$, $w([\zeta_1]) + \ldots + w([\zeta_m]) \leq w([\zeta_1']) + \ldots + w([\zeta_m'])$ and each $\zeta_i \in \arg\min_{\eta \in [\zeta_i]} w(\eta)$

We consider the following Optimal Homology Basis problem.

> For a given $p > 0$ compute a minimum $p-$homology basis with respect to the weight function r.

Algorithm 2 describes a $2\beta_p$-approximation algorithm for Optimal Homology Basis by computing a minimum homology basis with respect to r_P by restricting the centers of minimal spheres to sites. To compute the minimum homology basis $\mathscr{M}$ from Ω, (see Line 5) standard reduction is performed on $\partial = [B_p(\mathsf{K}) \,|\, \Omega]$. We examine the columns of the reduced matrix $\tilde{\partial}$ from left to right. For every non-zero column i that is an index from Ω, we add the corresponding cycle in Ω to $\mathscr{M}$. *Algorithm* 2 runs in $O(|P|N^3)$. See Appendix C.1 for a proof of correctness.

6 Optimal persistent homology basis

We now consider a filtration of a simplicial complex K with the aim of studying an extension of the problem to persistent homology [19]. We introduce the Minimum persistent homology Basis problem:

> Given a filtration $\mathcal{F}$ of complex K, compute a persistent p-basis $\Lambda_p = \{\zeta_i \mid i \in [|\mathcal{B}_p(\mathcal{F})|]\}$ that

Algorithm 2: Optimal homology basis for sites

Input : Complex $\mathsf{K} \subset \mathbb{R}^d$, $p > 0$
Output: A minimum homology basis with respect to r_P

```
Procedure Opt-hom-basis-for-sites(K, p)
    For each v ∈ P, define ≺_v. Using standard reduction compute the essential p−cycles of the filtration D_v(K), denote them by ζ_{v,1}, ..., ζ_{v,m}.
    Let Ω = {ζ_{v,i}}_{v∈P,1≤i≤m}. Sort the cycles in Ω so that if r_v(ζ_{v,i}) < r_{v'}(ζ_{v',i'}) then ζ_{v,i} precedes ζ_{v',i'} in Ω. If ζ_{v,i} ≺_v ζ_{v,i'}, then ζ_{v,i} precedes ζ_{v,i'} as well. Ties are broken arbitrarily. Denote this ordering on Ω as ≺_Ω.
    M ← ∅.
    for ζ in the ordered list Ω do
        Let η_1, ... η_k be the cycles currently in M.
        if [ζ] ∈ span{[η_1], ... [η_k]} then
            Discard ζ and continue.
        else
            Add ζ to M.
    Report M as a minimum homology basis.
```

minimizes $r(\Lambda_p) = \sum_{i=1}^{|\mathcal{B}_p(\mathcal{F})|} r(\zeta_i)$.

Theorem 1 states that for computing an optimal persistent homology basis it suffices to compute the minimum representative of each bar. Formally, an optimum representative of a bar $[b, d)$ is a cycle $\zeta^* \in \arg\min_{\eta \in \mathcal{R}([b,d))}\{r_P(\eta)\}$.

Algorithm 3 computes an minimum representative of a input bar $[b, d) \in \mathcal{B}_p(\mathcal{F})$ for a simplex-wise filtration $\mathcal{F}$ of K with $V(\mathsf{K}) = P$ with respect to r_P. For each site $v \in P$ the subroutine Opt-Pers-Cycle-Site is invoked which computes $\zeta_v^* \in \arg\min_{\eta\in\mathcal{R}([b,d))}\{r_v(\eta)\}$. Finally the minimum $\zeta_P^* = \arg\min_{v\in P}\{r_v(\zeta_v^*)\}$ among all sites is reported. Similar to Algorithm 1 a filtration $\mathcal{D}_v$ is defined on K_b. The essential p−cycles $Y = \{\zeta_1 \prec_v \ldots \prec_v \zeta_m\}$ of $\mathcal{D}_v(\mathsf{K}_b)$ are computed using standard reduction. We then compute the smallest $i > 0$ such that $\exists \xi \in span\{\zeta_1, \ldots, \zeta_i\}, \xi \in \mathcal{R}([b, d))$. If σ_b was added at index b of $\mathcal{F}$ and α is the index of the first cycle in Y containing σ_b, then update Y by adding Y_α to all other cycles containing σ_b. This ensures that only a single cycle now contains σ_b. Denoting these cycles of $Y \setminus \{\zeta_\alpha\}$ by Y' and the first i cycles of Y' by $Y'_{\leq i}$, it suffices to check if $[B_p(\mathsf{K}_d) \,|\, Y_{\leq i}].x = Y_\alpha$ has a solution. This is determined in Line 15 with a binary search over $i \in [1..m-1]$.

The proof of correctness of Algorithm 3 can be found in Appendix C.2. It runs in $O(|P|N^3 \log N)$.

Algorithm 3: Computing optimal representative of bar of persistence wrt r_P.

Input : $\mathsf{K}, \mathcal{F}$(simplex-wise filtration), $[b, d) \in \mathcal{B}_p(\mathcal{F})$
Output: $\zeta_P^*([b, d)) \in \arg\min_{v\in P,\eta\in\mathcal{R}([b,d))}\{r_v(\eta)\}$

```
1  Procedure Opt-PersHom-Rep(K, F, [b, d))
2      r_P([b, d)) ← ∞, ζ*_P([b, d)) ← ∅
3      for v ∈ P do
4          r_v(ζ*_v), ζ*_v ← Opt-Pers-Cycle-Site([b, d), v)
5          if r_v(ζ*_v) < r_P([b, d)]) then
6              r_P([b, d)) ← r_v(ζ*_v) ζ*_P([b, d)) ← ζ*_v
7      Return r_P([b, d)), ζ*_P([b, d))
8  Procedure Opt-Pers-Cycle-Site([b, d), v)
9      Define ≺_v on K_b. Compute D_v(K_b)
10     Compute the essential p-cycles ζ_1 ≺_v ... ≺_v ζ_m of D_v(K_b) using standard reduction
11     Y ← ζ_1, ..., ζ_m. (Y is a matrix of m columns, the column Y_i is the cycle-vector of (ζ_i). Let α be index of first cycle in Y containing σ_b
12     Add cycle Y_α to all other cycles in Y containing σ_b, resulting in matrix Ŷ.
13     Assemble matrix Y' by dropping the α^th column of Ŷ. Denote by Y'_{≤i} the first i columns of Y'.
14     ∂_d ← B_p(K_d)(∂_d is empty if d = ∞)
15     Compute the smallest i ∈ [1..m − 1] such that [∂_d|Y'_{≤i}].x = Y_α has a solution.
16     Let b_1, ..., b_t, i_1, ..., i_s be the solution computed by the previous step where b_1, ..., b_t are indices in ∂_d and i_1, ..., i_s are in Y'_{≤i}
17     ζ*_v ← Y'_{≤i,i_1} + ... + Y'_{≤i,i_s} + Y_α
18     Return ζ*_v, r_v(ζ*_v)
```

7 Experiments

We report results of experiments on real world datasets with a focus on computing cycle representatives of H_1. These results demonstrate the utility of the ℓ_2–radius towards the identification of meaningful representatives of homology classes. We consider three applications: localizing individual 1-cycles, optimal 1-homology basis computation, and minimum persistent 1-homology basis computation. Computing the exact ℓ_2–radius via an enumeration of all circumspheres is expensive. So, all experiments were conducted on implementations of the approximate algorithms described above.

We implement a heuristic to minimize the length of the cycle representative while preserving its radius. Essentially, we replace one of the paths P between two vertices in the cycle with the shortest length path between them if it is homologous to P. This heuristic results in smoother and shorter cycles for all datasets. This also addresses the issue with the non-unicity of the ℓ_2-metric in the sense that in our experiments we always find tighter cycles within a sphere when there are several homologous cycles of varying lengths within a sphere.

All experiments were performed on an Intel Xeon(R) Gold 6230 CPU @ 2.10 powered workstation with 20 cores and 384GB RAM running Ubuntu Linux. The algorithms were parallelized using Intel Thread Building Blocks (TBB). The PHAT library [3] was used in all routines that invoke the standard reduction algorithm.

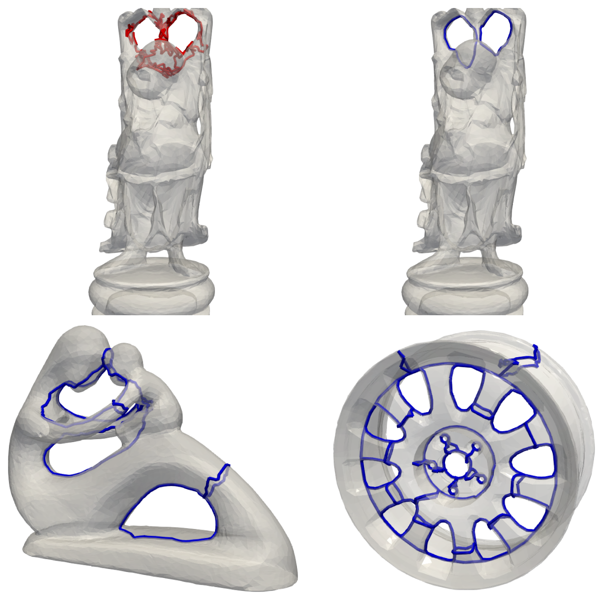
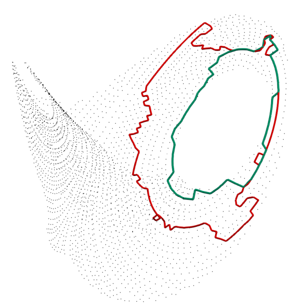


Figure 1: (**top**) The localization algorithm computes optimal (blue) 1-cycles that are homologous to input 1-cycles (red). (**bottom**) Minimum 1-homology basis.

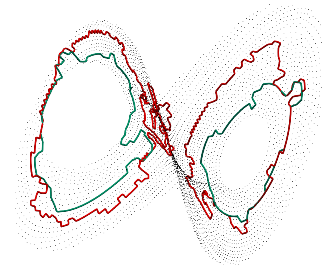
Homology localization. Given an input simplicial complex representing a surface and a 1-cycle, we compute a localized cycle that is homologous to the input cycle. Figure 1(top) shows two input cycles (red) and their localized versions (blue) for the Happy Buddha dataset. We can visually infer that the localized cycle computed by our approximate algorithm is close to the optimal cycle.

Optimal homology basis. Figure 1(bottom) shows the optimal homology basis computed for two 3D models. We observe that the cycles are tight and capture all tunnels and loops of the model. Results on additional datasets are available in Appendix D.

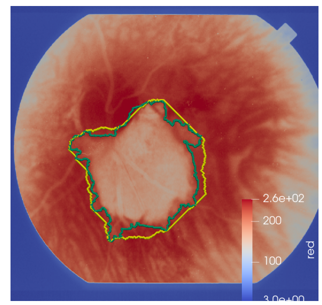
Optimal persistent homology cycle. We report our results on three classes of filtrations: Rips, lower star, and Delaunay. Figure 2 **(a,b)** shows the Lorenz'63 data set for 60% and 80% densities and the representatives of the longest and the top two longest bars, respectively, of a Rips filtration. These are of relevance in simulating weather phenomena [29]. Figure 2 **(c)** highlights the representative of the longest lived bar for a lower star filtration on a retinal image [23] with a retinal disorder. The cycle represents the region of the disorder. Figure 2 **(d)** highlights representatives of the top two bars of an alpha complex on a protein molecule (PDB-ID: 1OED). Cycles computed by our algorithm (green) appear "tighter" than those computed by PersLoop (red or yellow).



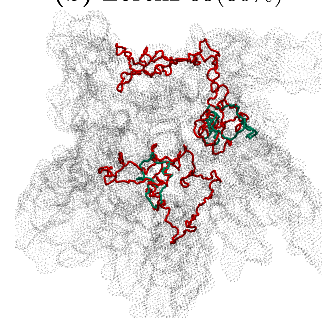
(a) Lorenz'63(60%)



(b) Lorenz'63(80%)



(c) Retina



(d) 1OED

Figure 2: The green persistent 1-homology cycles computed by our algorithm are tighter than the red (or yellow) cycles computed by PersLoop [15].

Execution time. Our algorithms are parallelizable if the optimization subroutines for each site can be executed independently. We obtain fast running times ($\sim$ few minutes) for moderate to large-sized datasets ($\sim$ millions of simplices) for all algorithms, see Table 1.

Data set	#Simplices	Execution time
Lorenz-63(60%)	$\sim 2 \times 10^6$	120s
Lorenz-63(80%)	$\sim 3 \times 10^6$	120s
Retina	$\sim 5 \times 10^6$	140s
1OED	$\sim 2.5 \times 10^6$	130s

Table 1: Mean time to compute optimal representative of each of the top-40 bars in the persistence barcode.

Acknowledgements

This work is partially supported by the PMRF, MoE Govt. of India, a SERB grant CRG/2021/005278, an NSF grant CCF 2049010 and ERC 101039913. VN acknowledges support from the Alexander von Humboldt Foundation, and Berlin MATH+ under the Visiting Scholar program. Part of this work was completed when VN was a guest Professor at the Zuse Institute Berlin. AR would like to thank Tamal K. Dey for several helpful discussions and for his invaluable support.

References

[1] H. Bakke Bjerkevik. On the stability of interval decomposable persistence modules. *Discrete & Computational Geometry*, 66(1):92–121, 2021.

[2] A. Barbensi, I. H. R. Yoon, C. D. Madsen, D. O. Ajayi, M. P. H. Stumpf, and H. A. Harrington. Hypergraphs for multiscale cycles in structured data, 2022.

[3] U. Bauer, M. Kerber, J. Reininghaus, and H. Wagner. PHAT–persistent homology algorithms toolbox. *Journal of symbolic computation*, 78:76–90, 2017.

[4] G. Borradaile, E. W. Chambers, K. Fox, and A. Nayyeriy. Minimum cycle and homology bases of surface-embedded graphs. *Journal of Computational Geometry*, 8(2), 2017.

[5] G. Borradaile, W. Maxwell, and A. Nayyeri. Minimum bounded chains and minimum homologous chains in embedded simplicial complexes. In *36th International Symposium on Computational Geometry, SoCG 2020*, volume 164 of *LIPIcs*, pages 21:1–21:15, 2020.

[6] M. Botnan. Mastermath course topological data analaysis. `http://www.few.vu.nl/~botnan/lecture_notes.pdf`, 2022.

[7] E. W. Chambers, J. Erickson, and A. Nayyeri. Minimum cuts and shortest homologous cycles. In *Proceedings of the twenty-fifth annual symposium on Computational geometry*, pages 377–385. ACM, 2009.

[8] E. W. Chambers, S. Parsa, and H. Schreiber. On complexity of computing bottleneck and lexicographic optimal cycles in a homology class. In *38th International Symposium on Computational Geometry, SoCG 2022*, volume 224 of *LIPIcs*, pages 25:1–25:15, 2022.

[9] C. Chen and D. Freedman. Measuring and computing natural generators for homology groups. *Computational Geometry*, 43(2):169–181, 2010.

[10] C. Chen and D. Freedman. Hardness results for homology localization. *Discrete & Computational Geometry*, 45(3):425–448, 2011.

[11] D. Cohen-Steiner, A. Lieutier, and J. Vuillamy. Lexicographic optimal homologous chains and applications to point cloud triangulations. In *36th International Symposium on Computational Geometry, SoCG 2020*, volume 164 of *LIPIcs*, pages 32:1–32:17, 2020.

[12] D. Cohen-Steiner, A. Lieutier, and J. Vuillamy. Lexicographic optimal homologous chains and applications to point cloud triangulations. *Discrete & Computational Geometry*, pages 1–20, 2022.

[13] T. K. Dey, A. N. Hirani, and B. Krishnamoorthy. Optimal homologous cycles, total unimodularity, and linear programming. *SIAM Journal on Computing*, 40(4):1026–1044, 2011.

[14] T. K. Dey, T. Hou, and S. Mandal. Persistent 1-cycles: Definition, computation, and its application. In *International Workshop on Computational Topology in Image Context*, pages 123–136. Springer, 2019.

[15] T. K. Dey, T. Hou, and S. Mandal. Computing minimal persistent cycles: Polynomial and hard cases. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2587–2606. SIAM, 2020.

[16] T. K. Dey, J. Sun, and Y. Wang. Approximating loops in a shortest homology basis from point data. In *Proceedings of the Twenty-Sixth Annual Symposium on Computational Geometry*, pages 166–175. ACM, 2010.

[17] T. K. Dey and Y. Wang. *Computational Topology for Data Analysis*. Cambridge University Press, 2022. `https://www.cs.purdue.edu/homes/tamaldey/book/CTDAbook/CTDAbook.pdf`.

[18] H. Edelsbrunner and J. Harer. *Computational Topology: An Introduction*. Applied Mathematics. American Mathematical Society, 2010.

[19] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discrete Comput. Geom.*, 28:511–533, 2002.

[20] J. Erickson and K. Whittlesey. Greedy optimal homotopy and homology generators. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1038–1046. Society for Industrial and Applied Mathematics, 2005.

[21] P. Gabriel. Unzerlegbare Darstellungen I. *Manuscripta Mathematica*, 6(1):71–103, 1972.

[22] G. Henselman-Petrusek. personal communication.

[23] A. Hoover and M. Goldbaum. Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels. *IEEE Transactions on Medical Imaging*, 22(8):951–958, 2003.

[24] L. Li, C. Thompson, G. Henselman-Petrusek, C. Giusti, and L. Ziegelmeier. Minimal cycle representatives in persistent homology using linear programming: An empirical study with user's guide. *Frontiers in artificial intelligence*, 4:681117, 2021.

[25] I. Obayashi. Volume-optimal cycle: Tightest representative cycle of a generator in persistent homology. *SIAM Journal on Applied Algebra and Geometry*, 2(4):508–534, 2018.

[26] I. Obayashi. Stable volumes for persistent homology. *Journal of Applied and Computational Topology*, 7(4):671–706, 2023.

[27] A. Rathod. Fast algorithms for minimum cycle basis and minimum homology basis. In *36th International Symposium on Computational Geometry, SoCG 2020*, volume 164 of *LIPIcs*, pages 64:1–64:11, 2020.

[28] Sayan Mandal, Tamal K Dey , Tao Hou. "persloop software for computing persistent 1-cycle. `https://github.com/Sayan-m90/Persloop-viewer`.

[29] K. Strommen, M. Chantry, J. Dorrington, and N. Otter. A topological perspective on weather regimes. *Climate Dynamics*, 60:1–31, 07 2022.

A Persistent homology

In this section, we provide the basic definitions for persistent homology, and the usual distances used in this context, namely, the interleaving distance and the matching distance.

Given a nested sequence of simplicial complexes indexed over $\mathbb{N}$, $\mathbb{Z}$, or $\mathbb{R}$, persistent homology of this sequence captures how a homology class evolves over the sequence. Formally, let $\mathcal{P}$ denote one of the poset categories $\mathbb{N}$, $\mathbb{Z}$, or $\mathbb{R}$. Given a finite complex K, let $\mathcal{K}$ denote the set of all possible subcomplexes of K including the empty one. A $\mathcal{P}$-indexed *filtration* of K is a map $\mathcal{F} : \mathcal{P} \to \mathcal{K}$ that satisfies $\mathcal{F}(a) \subseteq \mathcal{F}(a')$ for every pair of indices $a, a' \in \mathcal{P}$ with $a \leq a'$. In this case, $\mathcal{F}$ takes a sequence of real numbers

$$a_0 \leq a_1 \leq a_2 \leq \dots$$

to a nested sequence of simplicial complexes $\mathsf{K}_{a_i} \subseteq \mathsf{K}$

$$\mathcal{F} : \cdots \hookrightarrow \mathsf{K}_{a_0} \hookrightarrow \mathsf{K}_{a_1} \hookrightarrow \mathsf{K}_{a_2} \hookrightarrow \dots$$

Using p-th homology groups of the complexes over the field $\mathbb{Z}_2$, we get a sequence of vector spaces connected by inclusion-induced linear maps:

$$H_p\mathcal{F} : \cdots \to H_p(\mathsf{K}_{a_0}) \to H_p(\mathsf{K}_{a_1}) \to H_p(\mathsf{K}_{a_2}) \to \dots$$

For any pair a_i, a_j, let $\iota_{i,j} : H_p(\mathsf{K}_{a_i}) \to H_p(\mathsf{K}_{a_j})$ denote the linear map (internal morphism) induced by the inclusion $\mathsf{K}_{a_i} \hookrightarrow \mathsf{K}_{a_j}$. The sequence $H_p\mathcal{F}$ with the linear maps is called a *persistence module* which satisfies the following two properties : (i) for any triple $a_i \leq a_j \leq a_k$ in $\mathcal{P}$, one has $\iota_{i,k} = \iota_{j,k} \circ \iota_{i,j}$ and (ii) $\iota_{i,i}$ is identity for each $a_i \in \mathcal{P}$.

Let $b, d \in \mathcal{P}$. We define the *interval* $[b, d)$ as

$$[b, d) = \{a \in \mathcal{P} \mid b \leq a < d\} .$$

There is a special persistence module called the *interval module* $\mathbb{I}^{[b,d)}$ associated to the interval $[b, d)$. Denoting the vector space indexed at $a \in \mathcal{P}$ as $\mathbb{I}_a$, this interval module is given by

$$\mathbb{I}_a^{[b,d)} = \begin{cases} \mathbb{Z}_2 & \text{if } a \in [b, d) \\ 0 & \text{otherwise} \end{cases}$$

together with identity maps $\mathrm{id}_{a,a'} : \mathbb{I}_a^{[b,d)} \to \mathbb{I}_{a'}^{[b,d)}$ for all $a, a' \in [b, d)$ with $a \leq a'$.

It is known due to a result of Gabriel [21] that a persistence module defined with finite complexes admits a decomposition

$$H_p\mathcal{F} \cong \bigoplus_{\alpha} \mathbb{I}^{[b_\alpha, d_\alpha)} \tag{4}$$

which is unique up to isomorphism and permutation of the intervals. The intervals $[b_\alpha, d_\alpha)$ are called the *bars*. The multiset of bars forms the *barcode* of the persistence module $H_p\mathcal{F}$. Let $B_p(\mathcal{F})$ denote this barcode for the filtration $\mathcal{F}$.

A.1 Interleavings and matchings for persistence modules

We need the following concept of interleavings between persistence modules in order to establish a stability property for the so called ℓ_2–radius for Čech filtrations. To make the discussion accessible to a computer science audience, we strip the presentation off its original category-theoretic formulation, and instead provide a more concrete and accessible description.

Definition 6 (ϵ-shifts of persistence modules)**.** Let ϵ be a non-negative real number, and let $\mathbb{P}$ be a persistence module with internal linear maps given by $\phi_{a,a'}$. Then, the *ϵ-shift* of $\mathbb{P}$, denoted by $\mathbb{P}^\epsilon$, is given by $\mathbb{P}_a^\epsilon = \mathbb{P}_{a+\epsilon}$ with internal linear maps $\phi_{a,a'}^\epsilon = \phi_{a+\epsilon,a'+\epsilon}$.

Definition 7 (ε-interleaving)**.** Let $\mathbb{P}$ and $\mathbb{Q}$ be two persistence modules indexed over the real numbers and with the internal linear maps as $\phi_{a,a'}$ and $\psi_{a,a'}$ respectively. We say $\mathbb{P}$ and $\mathbb{Q}$ are ε-interleaved if there exist two families of maps $F_a : \mathbb{P}_a \to \mathbb{Q}_{a+\varepsilon}$ and $G_a : \mathbb{Q}_a \to \mathbb{P}_{a+\varepsilon}$ satisfying the following two conditions:

1. $\psi_{a+\varepsilon,a'+\varepsilon} \circ F_a = F_{a'} \circ \phi_{a,a'}$ and $\phi_{a+\varepsilon,a'+\varepsilon} \circ G_a = G_{a'} \circ \psi_{a,a'}$ [rectangular commutativity]

2. $G_{a+\varepsilon} \circ F_a = \phi_{a,a+2\varepsilon}$ and $F_{a+\varepsilon} \circ G_a = \psi_{a,a+2\varepsilon}$ [triangular commutativity]

$$\begin{tikzcd}
\mathbb{P}: \ldots \arrow[r] & \mathbb{P}_a \arrow[r] \arrow[dr] & \mathbb{P}_{a+\varepsilon} \arrow[r] \arrow[dr] & \mathbb{P}_{a+2\varepsilon} \arrow[r] \arrow[dr] & \ldots\ldots \\
\mathbb{Q}: \ldots \arrow[r] & \mathbb{Q}_a \arrow[r] \arrow[ur] & \mathbb{Q}_{a+\varepsilon} \arrow[r] \arrow[ur] & \mathbb{Q}_{a+2\varepsilon} \arrow[r] \arrow[ur] & \ldots\ldots
\end{tikzcd}$$

Some of the relevant maps for interleaving between two modules are shown above whereas the two parallelograms and the two triangles below depict the rectangular and the triangular commutativities respectively.

$$\begin{tikzcd}
\mathbb{P}_a \arrow[rr, "\phi_{a,a'}"] \arrow[dr, "F_a"'] & & \mathbb{P}_{a'} \arrow[dr, "F_{a'}"] & \\
& \mathbb{Q}_{a+\varepsilon} \arrow[rr, "\psi_{a+\varepsilon,a'+\varepsilon}"'] & & \mathbb{Q}_{a'+\varepsilon}
\end{tikzcd}
\qquad
\begin{tikzcd}
& \mathbb{P}_{a+\varepsilon} \arrow[rr, "\phi_{a+\varepsilon,a'+\varepsilon}"] & & \mathbb{Q}_{a'+\varepsilon} \\
\mathbb{Q}_a \arrow[ur, "G_a"] \arrow[rr, "\psi_{a,a'}"'] & & \mathbb{Q}_{a'} \arrow[ur, "G_{a'}"'] &
\end{tikzcd}$$

$$\begin{tikzcd}
\mathbb{P}_a \arrow[rr, "\phi_{a,a+2\varepsilon}"] \arrow[dr, "F_a"'] & & \mathbb{P}_{a+2\varepsilon} \\
& \mathbb{Q}_{a+\varepsilon} \arrow[ur, "G_{a+\varepsilon}"'] &
\end{tikzcd}
\qquad
\begin{tikzcd}
& \mathbb{P}_{a+\varepsilon} \arrow[dr, "F_{a+\varepsilon}"] & \\
\mathbb{Q}_a \arrow[ur, "G_a"] \arrow[rr, "\psi_{a,a+2\varepsilon}"] & & \mathbb{Q}_{a+2\varepsilon}
\end{tikzcd}$$

Definition 8 (Interleaving distance)**.** Given two persistence modules $\mathbb{P}$ and $\mathbb{Q}$, their interleaving distance is defined as

$$\mathrm{d}_I(\mathbb{P}, \mathbb{Q}) = \inf\{\varepsilon \,|\, \mathbb{P} \text{ and } \mathbb{Q} \text{ are } \varepsilon\text{-interleaved}\}.$$

Remark A.1. In Definition 7, the family of maps $\{F_a \mid a \in \mathbb{R}\}$ assemble to give a map $F : \mathbb{P} \to \mathbb{Q}^\epsilon$, and the family of maps $\{G_a \mid a \in \mathbb{R}\}$ assemble to give a map $G : \mathbb{Q} \to \mathbb{P}^\epsilon$.

Definition 9 (ϵ-shift of a family of maps)**.** Suppose there exists a family of maps $F : \mathbb{P} \to \mathbb{Q}$. Then, a shift of F by ϵ, denoted by F^ϵ, gives a new family maps $\{F^\epsilon_a = F_{a+\epsilon} \mid a \in \mathbb{R}\}$ from $\mathbb{P}^\epsilon$ to $\mathbb{Q}^\epsilon$.

We now define the notion of ϵ-matchings between two barcodes.

Definition 10 (ϵ-matching of barcodes)**.** Suppose that we are given two barcodes (which are multisets of intervals) X and Y. Then, a matching $\mathcal{M}$ between X and Y is a collection of pairs $\mathcal{M} = \{(\mathbf{i}, \mathbf{j}) \mid \mathbf{i} \in X, \mathbf{j} \in Y\}$ such that each $\mathbf{i}$ and $\mathbf{j}$ occur in at most one pair.

Let $\mathcal{U}(\mathcal{M})$ be the collection of intervals in $X \cup Y$ that do not appear in any of the pairings of $\mathcal{M}$.

For pairs $(\mathbf{i}, \mathbf{j}) \in \mathcal{M}$, where $\mathbf{i} = (b, d)$ and $\mathbf{j} = (b', d')$, define $c(\mathbf{i}, \mathbf{j}) = \max(|b - b'|, |d - d'|)$ and for intervals $\mathbf{i} = [b, d) \in \mathcal{U}$, define $c(\mathbf{i}) = \frac{d-b}{2}$. Then, the cost of the matching $\mathcal{M}$, denoted by $c(\mathcal{M})$, is defined as follows:

$$c(\mathcal{M}) = \max\left(\sup_{(\mathbf{i},\mathbf{j})\in\mathcal{M}} c((\mathbf{i}, \mathbf{j})), \sup_{\mathbf{i}\in\mathcal{U}(\mathcal{M})} c(\mathbf{i})\right).$$

Finally, we say that $\mathcal{M}$ is an ϵ-matching if $c(\mathcal{M}) \le \epsilon$.

B Approximate stability for ℓ_2-radius

We now study notions of stability for weight functions that serve as objective functions for intervals in a barcode. We limit the discussion to Čech filtrations, while noting that similar definitions can be derived for other commonly encountered filtrations like Rips or lower star.

B.1 Stability of weight functions on intervals

We recall the notion of Čech complexes before proceeding to introduce some new definitions of our own concerning stability of weight functions on intervals.

Definition 11 (Čech complexes). Let P be a finite point set in $\mathbb{R}^d$. Let $D_{r,x}$ denote a Euclidean ball of radius r centered at x. The Čech complex of P for radius r is the abstract simplicial complex given by

$$\check{\mathsf{C}}_r(P) = \{X \subset P \mid \bigcap_{x \in X} D_{r,x} \neq \emptyset\}.$$

The Čech filtration of P, denoted by $\check{\mathsf{C}}(P)$, is the nested sequence of complexes $\{\check{\mathsf{C}}_r(P)\}_{r \geq 0}$, where $\check{\mathsf{C}}_s(P) \subseteq \check{\mathsf{C}}_t(P)$ for $s \leq t$. We use the notation $B(\check{\mathsf{C}}(P))$ to denote the barcode of $\check{\mathsf{C}}(P)$.

Definition 12 (ϵ-perturbations of point sets). We say that Q is an *ϵ-perturbation* of P *realized through* a bijective map f, if f maps points in P to points in Q such that $\|p - f(p)\|_2 \leq \epsilon$.

Definition 13 (Weight functions on persistent homology bases). Given a filtration $\mathcal{F}$, and a p-th persistent homology basis $\Omega_p(\mathcal{F})$ of $\mathcal{F}$, a function that assigns a positive real number to every cycle in $\Omega_p(\mathcal{F})$ is called a *weight function on the p-th persistent homology basis* $\Omega_p(\mathcal{F})$ of $\mathcal{F}$. In this case, the (total) weight on $\Omega_p(\mathcal{F})$ is simply the sum of weights of basis cycles.

Definition 14 (Weight functions on persistent barcodes). For a filtration $\mathcal{F}$, its p-th barcode $B_p(\mathcal{F})$, a persistent p-homology basis $\Omega_p(\mathcal{F}) = \{\zeta_1 \dots \zeta_{|\Omega_p(\mathcal{F})|}\}$, and a weight function s on $\Omega_p(\mathcal{F})$ defined as $s(\Omega^p(\mathcal{F})) = \sum_{i=1}^{|\Omega_p(\mathcal{F})|} s(\zeta_i)$, a weight function w on $B_p(\mathcal{F})$ is defined as:

$$w(B_p(\mathcal{F})) = \min_{\Omega_p(\mathcal{F})} \; s(\Omega_p(\mathcal{F})) \quad \text{ where } \Omega_p(\mathcal{F}) \text{ is a persistent } p\text{-homology basis for } \mathcal{F}. \tag{5}$$

If $\Omega_p^\star(\mathcal{F})$ is a persistent p-homology basis of minimum weight in Equation (5), then the weight $w([b,d))$ for an interval $[b,d) \in B_p(\mathcal{F})$ is defined as $w([b,d)) = s(\zeta^\star)$ where $\zeta^\star$ is the representative for $[b,d)$ in $\Omega_p^\star(\mathcal{F})$.

Definition 15 (Stability of weight functions for filtration families). Suppose that we are given a weight function w on barcodes of a family of filtrations. Then, for a point set P, an associated filtration of complexes $\mathcal{F}_P$ and for real numbers $\epsilon, \delta \geq 0$, we say w is *(ϵ,δ)-stable* for an interval $[b,d) \in B(\mathcal{F}_P)$ with $(d-b) > 2\epsilon$, if for every ϵ-perturbation Q of P with an associated filtration of complexes $\mathcal{F}_Q$, there exists an ϵ-matching $\mathcal{M}$ between the intervals of $B(\mathcal{F}_P)$ and the intervals of $B(\mathcal{F}_Q)$ such that if the length of the interval $\mathcal{M}([b,d)) > 2\epsilon$ then

$$|w([b,d)) - w(\mathcal{M}([b,d)))| \leq 2\delta.$$

Furthermore, if w is (ϵ,δ)-stable for all intervals of $B(\mathcal{F}_P)$, then we say that w is *(ϵ,δ)-stable for the point set P*. Finally, if w is (ϵ,δ)-stable for every point set embedded in a Euclidean space, then we say that w is *(ϵ,δ)-stable for the filtration family.*

Definition 16 (Instability of weight functions). Given a family of filtrations $\mathfrak{F}$, and a weight function w defined on the persistent barcodes of filtrations belonging to $\mathfrak{F}$, we say that the pair $(\mathfrak{F}, w)$ is unstable, if for every $\epsilon, \alpha > 0$, there exists filtrations $\mathcal{F}_1, \mathcal{F}_2 \in \mathfrak{F}$ such that the barcodes of $\mathcal{F}_1, \mathcal{F}_2 \in \mathfrak{F}$ admit an ϵ-matching while satisfying $|w(\mathcal{F}_1) - w(\mathcal{F}_2)| > \alpha$.

In Figure 3, we provide an example depicting instability for the weight function given by length of 1-cycles. While this is not the same as the ℓ_2–radius, we maintain that similar counterexamples can be designed for ℓ_2–radius as well. We leave out the details here. In particular, it is easily seen that ℓ_2–radius is an unstable function for Čech filtrations.

B.2 Approximate representatives and bases

Given the inherent instability associated to the ℓ_2–radius, we formulate a slightly weaker notion of stability for the ℓ_2–radius. To state the result, we need a few definitions.

Definition 17 (ϵ-approximate representatives). For an interval $[b,d)$ with $(d-b) > 2\epsilon$, a cycle ζ is said to be an *ϵ-approximate representative* for $[b,d)$, if ζ is a nontrivial cycle that is born at $b' \in [b-\epsilon, b+\epsilon]$ and dies at $d' \in [d-\epsilon, d+\epsilon]$, and the class $[\zeta]$ is nontrivial in K_s for every $s \in [b', d')$.

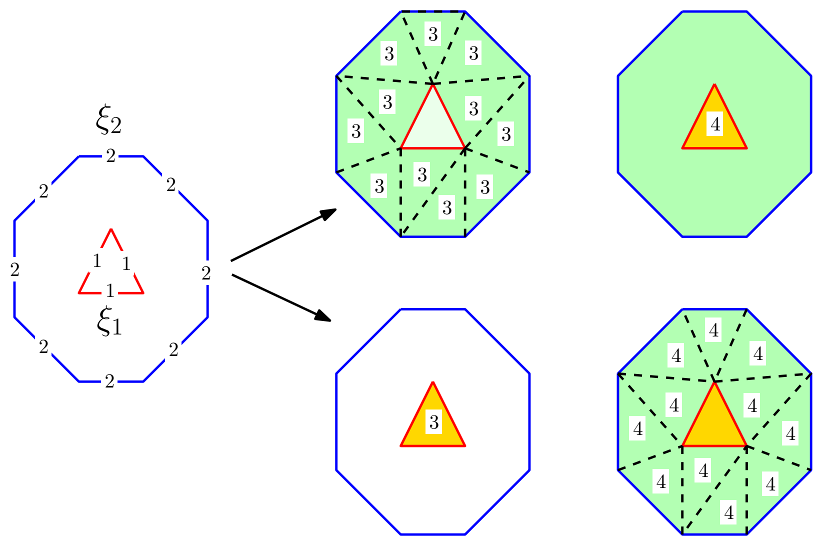


Figure 3: Two 1-cycles ξ_1 and ξ_2 appear at time (index) 1 and 2 respectively. In the upper filtration, the green triangles appear at time 3 filling up the annulus and the golden triangle appears at time 4 giving rise to the bars $[2, 3)$ and $[1, 4)$ with the minimal (lengthwise) persistent cycles $\xi_1 + \xi_2$ and ξ_1 resp. In the lower filtration the golden triangle comes first and green triangles next giving rise to slightly perturbed bars $[1, 3)$ and $[2, 4)$. Their minimal persistent cycles change to ξ_1 and ξ_2, resp. suggesting that the cycles may change considerably with respect to the length function.

Recall that in Section 4, we provided a definition for the ℓ_2–radius function r for intervals defined on chains and cycles of all subcomplexes of a complex K, whose vertex set is embedded in a Euclidean space. We now provide a definition for an approximate variant of this function.

Definition 18 (ϵ-approximate radii of intervals)**.** As an approximation of the radius function defined in Equation (3), we now define $r^\epsilon([b, d))$ for an interval $[b, d) \in B(\check{\mathsf{C}}(P))$ with $d-b > 2\epsilon$ to be the radius of the smallest Euclidean sphere that encloses all the vertices of some ϵ-approximate representative cycle of $[b, d)$. In other words, $r^\epsilon([b, d)) = \min_\zeta r(\zeta)$, where ζ is an ϵ-approximate representative cycle for $[b, d)$.

In Appendix B, we prove the following stability theorem (Theorem 2 restated as Theorem 4 with a proof) by building on techniques developed by Bjerkevik [1].

Theorem 2 (Approximate stability for ℓ_2–radius)**.** *Let P be a point set embedded in a Euclidean space and Q be an ϵ-perturbation of P. Then, there exists an ϵ-matching $\mathcal{M}$ between $B(\check{\mathsf{C}}(P))$ and $B(\check{\mathsf{C}}(Q))$ such that if $[b', d') = \mathcal{M}([b, d))$ and both $[b, d)$, $[b', d')$ have lengths greater than 2ϵ, then*

$$\begin{aligned} r^{2\epsilon}([b, d)) &\leq r([b', d')) + \epsilon, \\ r^{2\epsilon}([b', d')) &\leq r([b, d)) + \epsilon. \end{aligned} \tag{6}$$

Note that, as per Definition 15, we would have (ϵ, ϵ) stability for ℓ_2–radius, if there exists an ϵ-matching $\mathcal{M}$ that matches intervals $[b, d) \in B(\check{\mathsf{C}}(P))$ of length greater than 2ϵ to intervals $[b', d') \in B(\check{\mathsf{C}}(Q))$ of length greater than 2ϵ satisfying

$$\begin{aligned} r([b, d)) &\leq r([b', d')) + \epsilon, \\ r([b', d')) &\leq r([b, d)) + \epsilon. \end{aligned} \tag{7}$$

In this sense, Theorem 2 is an approximate version of ℓ_2–radius stability.

We now build towards a proof for Theorem 2. To begin with, let P be a point set embedded in a Euclidean space, and let Q be an ϵ-perturbation of P realized through a bijective map between point sets denoted by f. We denote the inverse of f by g. Let $\Omega^P = \{\zeta_1, \zeta_2, \ldots, \zeta_n\}$ be a persistent homology basis for $\check{\mathsf{C}}(P)$, and let $\Omega^Q = \{z_1, z_2, \ldots, z_m\}$ be a persistent homology basis for $\check{\mathsf{C}}(Q)$.

Then, it is easy to check that for every $s \in \mathbb{R}$, the map f induces a simplicial map $f_s : \check{\mathsf{C}}_s(P) \to \check{\mathsf{C}}_{s+\epsilon}(Q)$. Furthermore, the map f_s induces a map on the respective cycle groups $f_s^{\#} : Z_*(\check{\mathsf{C}}_s(P)) \to Z_*(\check{\mathsf{C}}_{s+\epsilon}(Q))$ as well as a map $\hat{f}_s : H_*(\check{\mathsf{C}}_s(P)) \to H_*(\check{\mathsf{C}}_{s+\epsilon}(Q))$ on homology groups.

Similarly, for every $s \in \mathbb{R}$, the map g induces a simplicial map $g_s : \check{\mathsf{C}}_s(Q) \to \check{\mathsf{C}}_{s+\epsilon}(P)$, which in turn, induces a map on the respective cycle groups $g_s^{\#} : Z_*(\check{\mathsf{C}}_s(Q)) \to Z_*(\check{\mathsf{C}}_{s+\epsilon}(P))$ as well as a map on homology groups $\hat{g}_s : H_*(\check{\mathsf{C}}_s(Q)) \to H_*(\check{\mathsf{C}}_{s+\epsilon}(P))$. Moreover, $g_{s+\epsilon} \circ f_s = \mathrm{id}$, and $f_{s+\epsilon} \circ g_s = \mathrm{id}$ for every $s \in \mathbb{R}$.

We define persistent modules $\mathbb{P}$ and $\mathbb{Q}$ as follows.

$$\begin{aligned} \mathbb{P} &= \bigoplus_{\substack{\mathbf{i} \in B(\check{\mathsf{C}}(P)) \\ \zeta_i \in \mathcal{R}(\mathbf{i})}} \mathbb{I}^{\zeta_i} \cong \bigoplus_{\mathbf{i} \in B(\check{\mathsf{C}}(P))} \mathbb{I}^{\mathbf{i}}, \\ \mathbb{Q} &= \bigoplus_{\substack{\mathbf{j} \in B(\check{\mathsf{C}}(Q)) \\ z_j \in \mathcal{R}(\mathbf{j})}} \mathbb{I}^{z_j} \cong \bigoplus_{\mathbf{j} \in B(\check{\mathsf{C}}(Q))} \mathbb{I}^{\mathbf{j}}. \end{aligned}$$

For some $\mathbf{j} = [b_j, d_j) \in B(\check{\mathsf{C}}(Q))$, let $z_j \in \mathcal{R}(\mathbf{j})$. Then, $\hat{g}_{b_j}([z_j])$ is a class in $\check{\mathsf{C}}_{b_j+\epsilon}(P)$. Since Ω^P is a persistent homology basis for $\check{\mathsf{C}}(P)$, $\hat{g}_{b_j}([z_j])$ can be written as follows.

$$\hat{g}_{b_j}([z_j]) = \sum_{i \in I} \kappa_{j,i} \cdot [\zeta_i]. \tag{8}$$

In Equation (8), $I \subset [n]$ indexes a subset of the representative cycles for the intervals in $B(\check{\mathsf{C}}(P))$, and the coefficients $\kappa_{j,i} \in \mathbb{Z}_2$.

We define $G|_{\mathbf{j}} : \mathbb{I}^{z_j} \to \mathbb{P}^{\epsilon}$ as follows.

$$G|_{\mathbf{j}}(\mathbb{I}^{z_j}_s) = \sum_{i \in I} \kappa_{j,i} \cdot \mathbb{I}^{\zeta_i}_{s+\epsilon} \quad \text{for } s \in [b_j, d_j).$$

By linear extension over all intervals $\mathbf{j} \in B(\check{\mathsf{C}}(Q))$, we obtain a map $G : \mathbb{Q} \to \mathbb{P}^{\epsilon}$.

Symmetrically, let $\zeta_i \in \Omega^P$ be a representative cycle for some $\mathbf{i} = [b_i, d_i) \in B(\check{\mathsf{C}}(P))$. Then, $\hat{f}_{b_i}([\zeta_i])$ is a class in $\check{\mathsf{C}}_{b_i+\epsilon}(Q)$ Thus, $\hat{f}_{b_i}([\zeta_i])$ can be written as follows.

$$\hat{f}_{b_i}([\zeta_i]) = \sum_{j \in J} \eta_{i,j} \cdot [z_j]. \tag{9}$$

In Equation (9), $J \subset [m]$ indexes a subset of the representative cycles for intervals in $B(\check{\mathsf{C}}(Q))$, and the coefficients $\eta_{i,j} \in \mathbb{Z}_2$.

We define $F|_{\mathbf{i}} : \mathbb{I}^{\zeta_i} \to \mathbb{Q}^{\epsilon}$ as follows.

$$F|_{\mathbf{i}}(\mathbb{I}^{\zeta_i}_s) = \sum_{j \in J} \eta_{i,j} \cdot \mathbb{I}^{z_j}_{s+\epsilon} \quad \text{for } s \in [b_i, d_i).$$

By linear extension over all intervals $\mathbf{i} \in B(\check{\mathsf{C}}(P))$, we obtain a map $F : \mathbb{P} \to \mathbb{Q}^{\epsilon}$.

It is easy to check that the maps F and G consitute an ϵ-interleaving between modules $\mathbb{P}$ and $\mathbb{Q}$. This is a simple consequence of the fact that $g_{s+\epsilon} \circ f_s = \mathrm{id}$, and $f_{s+\epsilon} \circ g_s = \mathrm{id}$ for every $s \in \mathbb{R}$.

Let $\imath^{\mathbb{P}}_{\mathbf{i}} : \mathbb{I}^{\zeta_i} \hookrightarrow \mathbb{P}$ and $\imath^{\mathbb{Q}}_{\mathbf{j}} : \mathbb{I}^{z_j} \hookrightarrow \mathbb{Q}$ denote the canonical inclusion maps, and let $\pi^{\mathbb{P}}_{\mathbf{i}} : \mathbb{P} \twoheadrightarrow \mathbb{I}^{\zeta_i}$ and $\pi^{\mathbb{Q}}_{\mathbf{j}} : \mathbb{Q} \twoheadrightarrow \mathbb{I}^{z_j}$ denote the canonical projection maps.

For the morphism $F : \mathbb{P} \to \mathbb{Q}^{\epsilon}$, we have $F|_{\mathbf{i}} = F \circ \imath^{\mathbb{P}}_{\mathbf{i}}$. Now, let $F_{\mathbf{i},\mathbf{j}} = (\pi^{\mathbb{Q}}_{\mathbf{j}})^{\epsilon} \circ F \circ \imath^{\mathbb{P}}_{\mathbf{i}}$. Likewise, for the morphism $G : \mathbb{Q} \to \mathbb{P}^{\epsilon}$, we have $G|_{\mathbf{j}} = G \circ \imath^{\mathbb{Q}}_{\mathbf{j}}$. Now, let $G_{\mathbf{j},\mathbf{i}} = (\pi^{\mathbb{P}}_{\mathbf{i}})^{\epsilon} \circ G \circ \imath^{\mathbb{Q}}_{\mathbf{j}}$.

Let $\phi^{\epsilon} : \mathbb{P} \to \mathbb{P}^{\epsilon}$ denote the collection of maps whose restriction to $\mathbb{P}_t$ gives the internal linear map $\phi_{t,t+\epsilon} : \mathbb{P}_t \to \mathbb{P}_{t+\epsilon}$. For an interval summand $\mathbb{I}^{\mathbf{i}}$ of $\mathbb{P}$, let $\phi^{\epsilon}_{\mathbf{i}}$ denote the collection of maps whose restriction to $\mathbb{I}^{\mathbf{i}}_t$ gives the internal linear map $\phi_{t,t+\epsilon}|_{\mathbb{I}^{\mathbf{i}}} : \mathbb{I}_t \to \mathbb{I}_{t+\epsilon}$.

Using the fact that $\phi^{2\epsilon}_{\mathbf{i}} = (\pi^{\mathbb{P}}_{\mathbf{i}})^{2\epsilon} \circ \phi^{2\epsilon} \circ \imath^{\mathbb{P}}_{\mathbf{i}}$, we can write

$$\phi^{2\epsilon}_{\mathbf{i}} = \sum_{\mathbf{j} \in B(\check{\mathsf{C}}(Q))} G^{\epsilon}_{\mathbf{j},\mathbf{i}} \circ F_{\mathbf{i},\mathbf{j}}. \tag{10}$$

Moreover, because $\mathbb{P}$ is a direct sum of interval modules, for $\mathbf{i} \neq \mathbf{i}'$, we obtain

$$\begin{aligned}(\pi^{\mathbb{P}}_{\mathbf{i}'})^{2\epsilon} \circ \phi^{2\epsilon} \circ \imath^{\mathbb{P}}_{\mathbf{i}} &= \sum_{\mathbf{j} \in B(\check{\mathsf{C}}(Q))} G^{\epsilon}_{\mathbf{j},\mathbf{i}'} \circ F_{\mathbf{i},\mathbf{j}} \\ &= 0. \end{aligned} \tag{11}$$

Definition 19 $(\Lambda(\mathbf{i}), \Lambda(A))$**.** For intervals $\mathbf{i} \in B(\check{\mathsf{C}}(P))$, we define

$$\widetilde{\Lambda}(\mathbf{i}) = \{\mathbf{j} \in B(\check{\mathsf{C}}(Q)) : \mathbb{I}^{\mathbf{i}} \text{ and } \mathbb{P}^{\mathbf{j}} \text{ are } \epsilon\text{-interleaved}\}, \text{ and}$$
$$\Lambda(\mathbf{i}) = \{\mathbf{j} \in B(\check{\mathsf{C}}(Q)) : \mathbb{I}^{\mathbf{i}} \text{ and } \mathbb{P}^{\mathbf{j}} \text{ are } \epsilon\text{-interleaved and either } \eta_{i,j} \neq 0 \text{ or } \kappa_{j,i} \neq 0\}.$$

For a collection of intervals $A \subset B(\check{\mathsf{C}}(P))$, we define

$$\widetilde{\Lambda}(A) = \bigcup_{\mathbf{i} \in A} \widetilde{\Lambda}(\mathbf{i}) \quad \text{and} \quad \Lambda(A) = \bigcup_{\mathbf{i} \in A} \Lambda(\mathbf{i}).$$

Definition 20 $(\Upsilon(\mathbf{j}), \Upsilon(A))$**.** For intervals $\mathbf{j} \in B(\check{\mathsf{C}}(Q))$, we define

$$\widetilde{\Upsilon}(\mathbf{j}) = \{\mathbf{i} \in B(\check{\mathsf{C}}(P)) : \mathbb{I}^{\mathbf{i}} \text{ and } \mathbb{P}^{\mathbf{j}} \text{ are } \epsilon\text{-interleaved}\}, \text{ and}$$
$$\Upsilon(\mathbf{j}) = \{\mathbf{i} \in B(\check{\mathsf{C}}(P)) : \mathbb{I}^{\mathbf{i}} \text{ and } \mathbb{P}^{\mathbf{j}} \text{ are } \epsilon\text{-interleaved and either } \eta_{i,j} \neq 0 \text{ or } \kappa_{j,i} \neq 0\}.$$

For a collection of intervals $A \subset B(\check{\mathsf{C}}(Q))$, we define

$$\widetilde{\Upsilon}(A) = \bigcup_{\mathbf{j} \in A} \widetilde{\Upsilon}(\mathbf{j}) \quad \text{and} \quad \Upsilon(A) = \bigcup_{\mathbf{j} \in A} \Upsilon(\mathbf{j}).$$

Remark B.1. From Definitions 19 and 20 we make the following observations.

$$\mathbf{j} \in \widetilde{\Lambda}(A) \setminus \Lambda(A) \quad \Longrightarrow \quad \eta_{i,j} = 0 \text{ and } \kappa_{j,i} = 0 \text{ for all } \mathbf{i} \in A.$$

$$\mathbf{i} \in \widetilde{\Upsilon}(A) \setminus \Upsilon(A) \quad \Longrightarrow \quad \eta_{i,j} = 0 \text{ and } \kappa_{j,i} = 0 \text{ for all } \mathbf{j} \in A.$$

Remark B.2. The sets $\widetilde{\Lambda}(A)$ is equivalent to $\mu(A)$ as defined in Botnan's lecture notes [6][Section 13.2]. The rationale behind defining $\Lambda(A)$ and $\Upsilon(A)$ is to ensure that the representative cycles of perturbed sets can be linearly related. Refer to Remark B.5,Theorem 4 for additional details.

We now recall some elementary propositions from Botnan's lecture notes [6]. The proofs for Propositions 21 to 23 can be found in Section 13.2 of [6].

Proposition 21. *For a morphism $\varrho : \mathbb{I}^{[b_1,d_1)} \to \mathbb{I}^{[b_2,d_2)}$, if $b_2 > b_1$ or $d_2 > d_1$, then ϱ is zero. On the other hand, if $b_2 \leq b_1$ and $d_2 \leq d_1$, then ϱ_t is determined by ϱ_{b_1} for $t \in [b_1, d_2)$, in the sense that, ϱ_t is nonzero if and only if ϱ_{b_1} is nonzero.*

Proposition 21 above says that the morphism $\varrho : \mathbb{I}^{[b_1,d_1)} \to \mathbb{I}^{[b_2,d_2)}$ between interval modules is completely determined by the function value at b_1.

For an interval $\mathbf{i} = [b, d)$, define $\alpha(\mathbf{i}) = b+d$. Specializing Bjerkevik's arguments [1] to single parameter persistence, Botnan [6] defines the following preorder on single parameter interval modules: $\mathbf{i} \leq_\alpha \mathbf{j}$ if and only if $\alpha_(\mathbf{i}) \leq \alpha_(\mathbf{j})$.

Proposition 22. *Let $\mathbf{i}_1 = [b_1, d_1)$, $\mathbf{i}_2 = [b_2, d_2)$ and $\mathbf{i}_3 = [b_3, d_3)$ be intervals such that $\mathbf{i}_1 \leq_\alpha \mathbf{i}_3$. If there exist nonzero maps $\varrho : \mathbb{I}^{\mathbf{i}_1} \to (\mathbb{I}^{\mathbf{i}_2})^\epsilon$ and $\chi : \mathbb{I}^{\mathbf{i}_2} \to (\mathbb{I}^{\mathbf{i}_3})^\epsilon$, then $\mathbb{I}^{\mathbf{i}_2}$ is ϵ-interleaved with either $\mathbb{I}^{\mathbf{i}_1}$ or $\mathbb{I}^{\mathbf{i}_3}$.*

Remark B.3. When $\mathbf{i}_3 = \mathbf{i}_1$, the contrapositive of Proposition 22 reads as follows: If $\mathbb{I}^{\mathbf{i}_2}$ is not ϵ-interleaved with $\mathbb{I}^{\mathbf{i}_1}$, then either $\varrho : \mathbb{I}^{\mathbf{i}_1} \to (\mathbb{I}^{\mathbf{i}_2})^\epsilon$ or $\chi : \mathbb{I}^{\mathbf{i}_2} \to (\mathbb{I}^{\mathbf{i}_3})^\epsilon$ is a zero map.

Proposition 23. *Let $\mathbf{i}_1 = [b_1, d_1)$, $\mathbf{i}_2 = [b_2, d_2)$ and $\mathbf{i}_3 = [b_3, d_3)$ be such that $(d_1 - b_1) > 2\epsilon$, $(d_3 - b_3) > 2\epsilon$ and $\mathbf{i}_1 \leq_\alpha \mathbf{i}_3$. If there exist nonzero maps $\varrho : \mathbb{I}^{\mathbf{i}_1} \to (\mathbb{I}^{\mathbf{i}_2})^\epsilon$ and $\chi : \mathbb{I}^{\mathbf{i}_2} \to (\mathbb{I}^{\mathbf{i}_3})^\epsilon$, then $\chi^\epsilon \circ \varrho \neq 0$.*

Remark B.4. It is worth noting that Proposition 23 applies only when $\mathbf{i}_1 \leq_\alpha \mathbf{i}_3$. When $\mathbf{i}_1 >_\alpha \mathbf{i}_3$, there is no guarantee that $\chi^\epsilon \circ \varrho$ is nonzero even when ϱ and χ are both nonzero. Later, in Proposition 24, this has implications in how Equation (16) is set up.

Proposition 24. *Suppose $\mathbb{P}$ and $\mathbb{Q}$ are ϵ-interleaved. Let A be a collection of intervals in $\mathbb{P}$ with length greater than 2ϵ. Then, $|A| \leq |\Lambda(A)|$.*

Proof. Let $\mu = |A|$ and $\nu = |\Lambda(A)|$. To begin with, order the elements of $A = \{\mathbf{i}_1, \mathbf{i}_2, \dots, \mathbf{i}_\mu\}$ in non-decreasing order of $\leq_\alpha$. Also, for $\mathbf{i} = [b_i, d_i)$ and $\mathbf{i}' = [b_{i'}, d_{i'})$, we write $i < i'$ if and only if $\mathbf{i} \leq_\alpha \mathbf{i}'$. Let $\Lambda(A) = \{\mathbf{j}_1, \mathbf{j}_2, \dots, \mathbf{j}_\nu\}$.

Next, we obtain an expression for $\phi_{\mathbf{i}}^{2\epsilon}$ as follows.

$$\begin{aligned}
\phi_{\mathbf{i}}^{2\epsilon} &= \sum_{\mathbf{j} \in B(\check{\mathbb{C}}(Q))} G^\epsilon_{\mathbf{j},\mathbf{i}} \circ F_{\mathbf{i},\mathbf{j}} && \text{using Equation (10)} \\
&= \sum_{\mathbf{j} \in \widetilde{\Lambda}(A)} G^\epsilon_{\mathbf{j},\mathbf{i}} \circ F_{\mathbf{i},\mathbf{j}} && \text{using Remark B.3} \\
&= \sum_{\mathbf{j} \in \Lambda(A)} G^\epsilon_{\mathbf{j},\mathbf{i}} \circ F_{\mathbf{i},\mathbf{j}} && \text{using Remark B.1.}
\end{aligned} \tag{12}$$

Using the same observation, for $\mathbf{i} \leq_\alpha \mathbf{i}'$, we obtain the following expression:

$$\begin{aligned}
\sum_{\mathbf{j} \in \Lambda(A)} G^\epsilon_{\mathbf{j},\mathbf{i}'} \circ F_{\mathbf{i},\mathbf{j}} &= \sum_{\mathbf{j} \in \widetilde{\Lambda}(A)} G^\epsilon_{\mathbf{j},\mathbf{i}'} \circ F_{\mathbf{i},\mathbf{j}} && \text{using Remark B.1} \\
&= \sum_{\mathbf{j} \in B(\check{\mathbb{C}}(Q))} G^\epsilon_{\mathbf{j},\mathbf{i}'} \circ F_{\mathbf{i},\mathbf{j}} && \text{using Proposition 22} \\
&= 0 && \text{using Equation (11).}
\end{aligned} \tag{13}$$

Using Propositions 21 and 23, Equations (12) and (13) can be written as

$$1 = \sum_{\mathbf{j} \in \Lambda(\mathbf{i})} \kappa_{j,i} \cdot \eta_{i,j} \quad \text{and} \tag{14}$$

$$0 = \sum_{\mathbf{j} \in \Lambda(\mathbf{i})} \kappa_{j,i'} \cdot \eta_{i,j} \text{ for } i < i', \text{ respectively.} \tag{15}$$

Putting Equations (14) and (15) in matrix form yields the following matrix equation, where the entries above the diagonal in the right hand side matrix are unknown.

$$\begin{bmatrix} \kappa_{j_1,i_1} & \dots & \kappa_{j_\nu,i_1} \\ \vdots & \ddots & \vdots \\ \kappa_{j_1,i_\mu} & \dots & \kappa_{j_\nu,i_\mu} \end{bmatrix} \begin{bmatrix} \eta_{i_1,j_1} & \dots & \eta_{i_\mu,j_1} \\ \vdots & \ddots & \vdots \\ \eta_{i_1,j_\nu} & \dots & \eta_{i_\mu,j_\nu} \end{bmatrix} = \begin{bmatrix} 1 & * & \dots & * \\ 0 & 1 & \dots & * \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}. \tag{16}$$

Since the matrix on the right hand side is upper triangular with ones on diagonal, it has rank $|A|$. On the other hand, the two matrices on the left have rank upper bounded by $\nu = |\Lambda(A)|$. It follows immediately that $|A| \leq |\Lambda(A)|$.

Proposition 25. *Suppose $\mathbb{P}$ and $\mathbb{Q}$ are ϵ-interleaved. Let A be a collection of intervals in $\mathbb{Q}$ with length greater than 2ϵ. Then, $|A| \leq |\Upsilon(A)|$.*

Proof. The statement of the theorem is symmetric to Proposition 25. Hence, we omit the details of the proof.

Remark B.5. Note that the proof of Proposition 24 closely follows Botnan's exposition of Bjerkevik's ideas for constructing an ϵ-matching given an ϵ-interleaving between persistence modules using Hall's theorem. However, there are three important differences.

- In our proof, it is vital to use the canonical ϵ-interleavings that are induced by the simplicial maps f_s and g_s for $s \in \mathbb{R}$ as described in Appendix B. In Bjerkevik's approach an arbitrary ϵ-interleaving can be used to derive an ϵ-matching. See Figure 4 for an example.

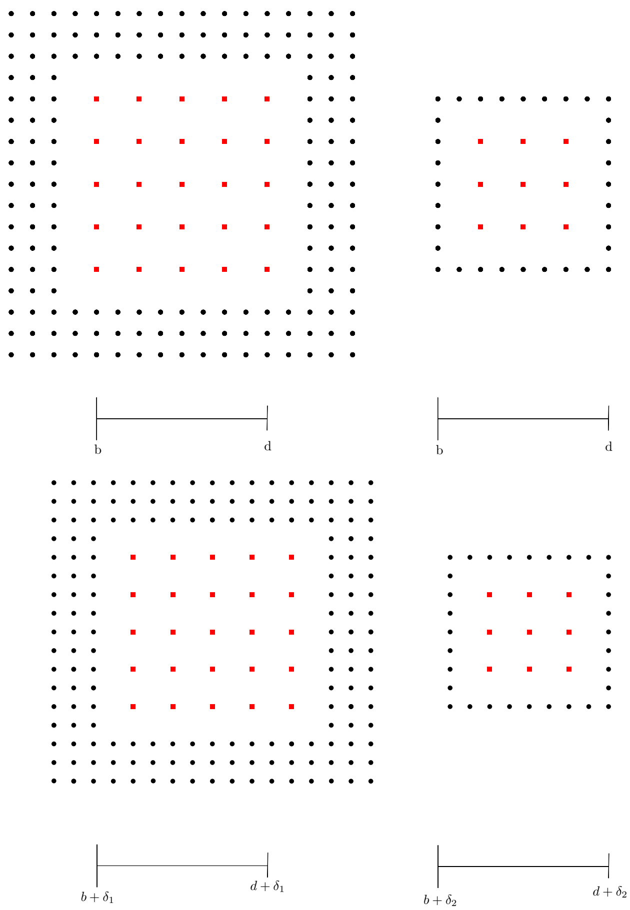


Figure 4: Consider the Čech filtrations on the point set shown in the top figure, denoted by P, which is δ-perturbed to obtain another point set Q shown in the bottom figure. In $\check{\mathsf{C}}(P)$, the cycle on the left supported by the inner rim of black grid points, and the cycle on the right supported by black points are born at b and die at d. Owing to a δ-perturbation, where $\delta_1, \delta_2 < \delta$, in $\check{\mathsf{C}}(Q)$, the first cycle is born at $b + \delta_1$ and dies at $d + \delta_1$, whereas the second cycle is born at $b + \delta_2$ and dies at $d + \delta_2$. If one uses Bjerkevik's approach to obtain a δ-matching, the bar $[b, d)$ (top left) may be matched to either the bar $[b+\delta_1, d+\delta_1)$ (bottom left) or $[b+\delta_2, d+\delta_2)$ (bottom right), whereas the bar $[b, d)$ (top right) may also be matched to either the bar $[b+\delta_1, d+\delta_1)$ (bottom left) or $[b+\delta_2, d+\delta_2)$ (bottom right). Thus, a naive approach to δ-matching is not approximately stable as the radius values of the enclosing spheres of matched representatives can be arbitrarily far apart. Our version of matching ensures that for this example, $[b, d)$ (top left) is matched to $[b+\delta_1, d+\delta_1)$ (bottom left) and $[b, d)$ (top right) is matched to the bar $[b+\delta_2, d+\delta_2)$ (bottom right).

- While for Bjerkevik's result it suffices to establish the inequality $|A| \leq |\widetilde{\Lambda}(A)|$, for our purposes it is necessary to establish the stricter inequality $|A| \leq |\Lambda(A)|$. In particular, we require that an interval $\mathbf{i}$ is matched only to one of the intervals in $\Lambda(\mathbf{i})$.
- While Bjerkevik uses arbitary interval decompositions of persistent modules, we are required to use the decompositions that come from fixed choices of persistence homology bases. This has the following consequence: even when the ϵ-interleaving maps F and G are canonical, the maps $F_{\mathbf{i},\mathbf{j}}$ and $G_{\mathbf{j},\mathbf{i}}$ depend on how we choose to represent the interval summands of $\mathbb{P}$ and $\mathbb{Q}$. Since $F_{\mathbf{i},\mathbf{j}}$ and $G_{\mathbf{j},\mathbf{i}}$ determine $\Lambda(\mathbf{i})$ for every $\mathbf{i}$, the underlying bipartite graph to be matched is determined by the choice of representative cycles. This in turn has a bearing on what kind of ϵ-interleaved interval summands of $\mathbb{P}$ and $\mathbb{Q}$ get matched.

Theorem 3 (Hall's theorem)**.** *Let G be a finite bipartite graph on sets U and V. For a subset of vertices $U' \subset U$, let $N_G(U')$ denote the subset of V adjacent to U'. Then, the following are equivalent:*

- *for all $U' \subset U$, $|U'| \leq |N_G(U')|$*
- *there exists an injective map $\imath : U \hookrightarrow V$ such that $\imath$ maps every vertex a of U to a vertex b of V only if there is*

an edge (a,b) *in* G.

Let $\Omega^P_{2\epsilon} \subseteq \Omega^P$ be the cycles of Ω^P that represent intervals of length greater than 2ϵ. Likewise, let $\Omega^Q_{2\epsilon} \subseteq \Omega^Q$ be the cycles of Ω^Q that represent intervals of length greater than 2ϵ. Since $\mathbb{P}$ and $\mathbb{Q}$ are ϵ-interleaved, combining Proposition 24 and Theorem 3, we obtain two injections $\imath : \Omega^P_{2\epsilon} \hookrightarrow \Omega^Q$ and $\jmath : \Omega^Q_{2\epsilon} \hookrightarrow \Omega^P$ such that

$$\imath(\zeta_i) = z_j \implies \mathbb{I}^{\zeta_i} \text{ and } \mathbb{I}^{z_j} \text{ are } \epsilon\text{-interleaved and either } F_{\mathbf{i},\mathbf{j}} \text{ or } G_{\mathbf{j},\mathbf{i}} \text{ is nonzero.} \tag{17}$$

$$\jmath(z_j) = \zeta_i \implies \mathbb{I}^{\zeta_i} \text{ and } \mathbb{I}^{z_j} \text{ are } \epsilon\text{-interleaved and either } F_{\mathbf{i},\mathbf{j}} \text{ or } G_{\mathbf{j},\mathbf{i}} \text{ is nonzero.} \tag{18}$$

Corollary 26. *There is a matching of representative cycles of* $\check{\mathsf{C}}(P)$ *with representative cycles of* $\check{\mathsf{C}}(Q)$ *such that*

- *All persistent cycles of* $\check{\mathsf{C}}(P)$ *and* $\check{\mathsf{C}}(Q)$ *representing intervals of length greater than* 2ϵ *are matched.*
- *If a representative* ζ_i *of* $\check{\mathsf{C}}(P)$ *is matched to a representative* z_j *of* $\check{\mathsf{C}}(Q)$*, then* $\mathbb{I}^{\mathbf{i}}$ *is* ϵ*-interleaved with* $\mathbb{I}^{\mathbf{j}}$*, and either* $F_{\mathbf{i},\mathbf{j}}$ *or* $G_{\mathbf{j},\mathbf{i}}$ *is nonzero.*

Proof. The proof is essentially a paraphrase of the proof of Theorem 13.14 in Botnan's notes [6].

Construct a bipartite graph $G = (V,E)$ with vertex set $V = \Omega^P \bigcup \Omega^Q$. The vertices of Ω^P are colored red and the vertices of Ω^Q are colored blue. The edges of G are built from the two injections $\imath : \Omega^P_{2\epsilon} \hookrightarrow \Omega^Q$ and $\jmath : \Omega^Q_{2\epsilon} \hookrightarrow \Omega^P$. In particular, we have a directed edge $\zeta \bullet\!\!\rightarrow z$ if and only if $\imath(\zeta) = z$, and a directed edge $z \bullet\!\!\rightarrow \zeta$ if and only if $\jmath(z) = \zeta$. It is easy to check that every connected component in G is either a directed cycle or a directed path.

The matching $\mathcal{M}$ is constructed as follows. For every cycle, pick alternate edges and include them in $\mathcal{M}$. For every directed path, pick the odd numbered edges and include them in $\mathcal{M}$. As a consequence, all vertices incident on some directed cycle are matched. Also, all vertices on directed paths of odd length are matched. The only vertices that are not matched are the terminal vertices of paths with even length, and these terminal vertices are representative cycles for intervals of length smaller than 2ϵ. This shows that $\mathcal{M}$ is an ϵ-matching.

Also, by construction of $\Lambda(A)$ and $\Upsilon(A)$ as described in Definition 19, we have a directed edge from interval $\mathbf{i}$ to $\mathbf{j}$ in G only if they are ϵ-interleaved and either $F_{\mathbf{i},\mathbf{j}}$ or $G_{\mathbf{j},\mathbf{i}}$ is nonzero, which proves the second claim.

Proposition 27. *For an interval* $[b',d') \in B(\check{\mathsf{C}}(Q))$ *represented by* z_j*, let* $\hat{g}_{b_j}([z_j]) = \sum_{i\in I} \kappa_{j,i} \cdot [\zeta_i]$*, where* I *indexes a subset of representative cycles of* $\check{\mathsf{C}}(P)$*. Let* $\Omega = \{\zeta_i \mid i \in I\}$*. Then, every cycle in the set* Ω *dies at or before* $d' + \epsilon$*.*

Proof. Targeting a contradiction, suppose that there exists a partition of cycles $\Omega = \Omega_1 \sqcup \Omega_2$, where the cycles in Ω_1 die at or before $d' + \epsilon$, whereas the cycles in Ω_2 die after $d' + \epsilon$. Then, Eqn. 8 can be written as

$$\sum_{\omega\in\Omega_2} [\omega] \;=\; \sum_{\gamma\in\Omega_1} [\gamma] \;+\; \hat{g}_{b_j}([z_j]). \tag{19}$$

First, note that $\hat{g}_{b_j}([z_j])$ is trivial at $d' + \epsilon$ because g is an ϵ-perturbation and z_j dies at d'. Then, at $d' + \epsilon$, the class on the left hand side, namely, $\sum_{\omega\in\Omega_2}[\omega]$ is nontrivial because the cycles in Ω_2 persist beyond $d' + \epsilon$, whereas the class on the right hand side, namely, $\sum_{\gamma\in\Omega_1}[\gamma] + \hat{g}_{b_j}([z_j])$ is trivial at $d' + \epsilon$ because it is a sum of trivial classes. This gives the required contradiction. Hence, every cycle in the set Ω dies at or before $d' + \epsilon$.

Theorem 4 (Approximate stability for ℓ_2–radius for Čech filtrations)**.** *Let* P *be a point set embedded in a Euclidean space and* Q *be an* ϵ*-perturbation of* P*. Then, the* ϵ*-matching described in Corollary 26 matches intervals of* $B(\check{\mathsf{C}}(P))$ *to intervals of* $B(\check{\mathsf{C}}(Q))$ *such that for every interval* $[b,d) \in B(\check{\mathsf{C}}(P))$ *with length greater than* 2ϵ*, if the interval* $[b',d') = \mathcal{M}([b,d))$ *has length greater than* 2ϵ*, then*

$$r^{2\epsilon}([b,d)) \le r([b',d')) + \epsilon, \tag{20}$$

$$r^{2\epsilon}([b',d')) \le r([b,d)) + \epsilon. \tag{21}$$

Proof. We will only prove Equation (20). Equation (21) follows from the symmetry of the argument. Let $\Omega^Q_\star = \{z_k \mid k \in [m]\}$ be an optimal persistent homology basis for $\check{\mathsf{C}}(Q)$ and let $\Omega^P = \{\zeta_\ell \mid \ell \in [n]\}$ be an arbitrary persistent homology basis for $\check{\mathsf{C}}(P)$. Suppose that z_j, a (optimal) representative cycle for the interval $[b_j,d_j) \in B(\check{\mathsf{C}}(Q))$, is matched to ζ_i, a representative cycle for the interval $[b_i,d_i) \in B(\check{\mathsf{C}}(P))$.

Recall that the matching in Corollary 26 guarantees that $\mathbb{I}^{z_j}$ and $\mathbb{I}^{\zeta_i}$ are ϵ-interleaved and either $\eta_{i,j} \neq 0$ or $\kappa_{j,i} \neq 0$.

If $\kappa_{j,i} \neq 0$, then we write $\hat{g}_{b_j}([z_j])$ as

$$\hat{g}_{b_j}([z_j]) = \sum_{\ell\in I} \kappa_{j,\ell} \cdot [\zeta_\ell].$$

where as before, I indexes a subset of $[n]$ and for $\ell = i$, $\kappa_{j,\ell} \neq 0$. Since $\mathbb{I}^{z_j}$ is ϵ-interleaved with $\mathbb{I}^{\zeta_i}$, $b_i \in [b_j - \epsilon, b_j + \epsilon]$ and $d_i \in [d_j - \epsilon, d_j + \epsilon]$. Then, using Proposition 27, every cycle ζ_ℓ for $\ell \in I$ dies at or before $d_i + 2\epsilon$. Furthermore, every cycle ζ_ℓ for $\ell \in I$ is born at or before $b_j + \epsilon$, and hence also at or before $b_i + 2\epsilon$. Therefore, $g^{\#}_{b_j}(z_j)$ is a valid (and possibly optimal) 2ϵ-approximate representative for ζ_i. Since g is an ϵ-perturbation, by triangle inequality, $r(g^{\#}_{b_j}(z_j)) \leq r(z_j) + \epsilon$. Since the ℓ_2–radius of an optimal choice of 2ϵ-representative for ζ_i is upper bounded by $r(g^{\#}_{b_j}(z_j))$, and is possibly even smaller than $r(g^{\#}_{b_j}(z_j))$, we have $r^{2\epsilon}([b,d)) \leq r(g^{\#}_{b_j}(z_j))$. Also, by definition, $r(z_j) = r([b',d'))$. This proves the claim when $\kappa_{j,i} \neq 0$.

On the other hand, if $\eta_{i,j} \neq 0$, we write $\hat{f}_{b_i}([\zeta_i])$ as

$$\hat{f}_{b_i}([\zeta_i]) = \sum_{k \in J} \eta_{i,k} \cdot [z_k]. \tag{22}$$

where as before, J indexes a subset of $[m]$ and for $k = j$, $\eta_{i,j} \neq 0$. Clearly, the cycles z_k for $k \in J$ are born before $b_i + \epsilon$. Furthermore, it is easy to prove along the lines of Proposition 27 that the cycles z_k for $k \in J$ also dies before $d_i + \epsilon$. Rewriting Equation (22) we obtain

$$[z_j] = \eta_{i,k} \cdot \eta_{i,j}^{-1} \sum_{k \in J'} [z_k] + \eta_{i,j}^{-1} \cdot \hat{f}_{b_i}([\zeta_i]) \quad \text{ where } J' = J \setminus \{j\}.$$

Applying $\hat{g}_{b_i+\epsilon}$ to Equation (22) gives

$$\begin{aligned} \hat{g}_{b_i+\epsilon}([z_j]) &= \eta_{i,k} \cdot \eta_{i,j}^{-1} \sum_{k \in J''} \hat{g}_{b_i+\epsilon}([z_k]) + \eta_{i,j}^{-1} \hat{g}_{b_i+\epsilon} \circ \hat{f}_{b_i}([\zeta_i]). \\ &= \eta_{i,k} \cdot \eta_{i,j}^{-1} \sum_{k \in J''} \hat{g}_{b_i+\epsilon}([z_k]) + \eta_{i,j}^{-1}([\zeta_i]). \end{aligned}$$

Here, $J'' \subset J'$ indexes cycles z_k with $k \in J'$ for which $\hat{g}_{b_i+\epsilon}([z_k])$ is not trivial. As before, $g^{\#}_{b_i+\epsilon}(z_j)$ is a valid (and possibly optimal) 2ϵ-approximate representative for ζ_i. Since g is an ϵ-perturbation, by triangle inequality, $r(g^{\#}_{b_i+\epsilon}(z_j)) \leq r(z_j) + \epsilon$. Since the ℓ_2–radius of an optimal choice of 2ϵ-representative for ζ_i is bounded from above by $r(g^{\#}_{b_i+\epsilon}(z_j))$, we have $r^{2\epsilon}([b,d)) \leq r(g^{\#}_{b_{i+\epsilon}}(z_j))$. Therefore, for $\eta_{i,j} \neq 0$, we have

$$\begin{aligned} r^{2\epsilon}([b,d)) &\leq r(g^{\#}_{b_{i+\epsilon}}(z_j)) \\ &\leq r(z_j) + \epsilon \\ &= r([b',d') + \epsilon. \end{aligned}$$

Approximate stability for ℓ_2–radius for Rips **complexes** As before, let P be a point set embedded in a Euclidean space, and let Q be an ϵ-perturbation of P realized through a bijective map between point sets, denoted by f. Also, let g be the inverse of f. Then, it is easy to check that for every $s \in \mathbb{R}$, the map f (resp. g) induces an inclusion induced simplicial map $f_s : \mathrm{V}_s(P) \to \mathrm{V}_{s+2\epsilon}(Q)$ (resp. $g_s : \mathrm{V}_s(Q) \to \mathrm{V}_{s+2\epsilon}(P)$). Moreover, the map f_s (resp. g_s) induces a map on the respective cycle groups $f^{\#}_s : Z_*(\mathrm{V}_s(P)) \to Z_*(\mathrm{V}_{s+2\epsilon}(Q))$ (resp. $g^{\#}_s : Z_*(\mathrm{V}_s(Q)) \to Z_*(\mathrm{V}_{s+2\epsilon}(P))$) as well as a map $\hat{f}_s : H_*(\mathrm{V}_s(P)) \to H_*(\mathrm{V}_{s+2\epsilon}(Q))$ (resp. $\hat{g}_s : H_*(\mathrm{V}_s(Q)) \to H_*(\mathrm{V}_{s+2\epsilon}(P))$) on homology groups.

The proof strategy in Appendix B can be repeated more or less vertbatim to obtain the following result for Rips complexes. We leave out the details.

Theorem 5 (Approximate stability for ℓ_2–radius for Rips complexes)**.** *Let P be a point set embedded in a Euclidean space and Q be an ϵ-perturbation of P. Then, there exists an 2ϵ-matching that matches intervals of $B(\mathrm{V}(P))$ to intervals of $B(\mathrm{V}(Q))$ such that for every interval $[b,d) \in B(\mathrm{V}(P))$ with length greater than 4ϵ, if the interval $[b',d') = \mathcal{M}([b,d))$ has length greater than 4ϵ, then*

$$\begin{aligned} r^{4\epsilon}([b,d)) &\leq r([b',d')) + \epsilon, \\ r^{4\epsilon}([b',d')) &\leq r([b,d)) + \epsilon. \end{aligned}$$

Remark B.6. The astute reader may have noticed that statements analogous to Theorem 4 can be obtained for many commonly encountered filtrations including Delaunay and lower star. To obtain the respective statements for each of these filtrations, the changes required to the proof of Theorem 4 are routine, and hence we do not discuss the topic of extensions of Theorem 4 any further.

C Correctness of Algorithms

In this section we give proofs for the correctness of algorithms stated and other associated results. The standard reduction algorithm [3] will be used in many of our algorithms as subroutines. We begin by outlining the standard reduction algorithm (Algorithm 4) and recalling some facts that arise out of it. For any matrix ∂ with entries in $\mathbb{Z}_2$ we define $low(j)$ to the row index of the lowest 1 in column j of ∂. It is undefined if column j is 0.

Remark C.1. For a $n \times m$ matrix standard reduction runs in $O(n^2 m)$ time.

Algorithm 4: Standard reduction algorithm for matrix reduction [3]

Input : Matrix ∂ with m columns and entries in $\mathbb{Z}_2$

Output: $\tilde{\partial}$(the reduced matrix), V (a collection of cycle-vectors). $\tilde{\partial}$ has the property that $low(j)$ is unique for every non-zero column of $\tilde{\partial}$. If ∂ is the boundary matrix of a filtration and if $i_1, \ldots i_\beta$ are the unpaired column indices then cycle vectors represented by $V_{i_1}, \ldots, V_{i_\beta}$ are the essential cycles of the filtration

Procedure `StandardReduction`(∂)
- $V \leftarrow I_m$
- **for** $j = 1$ *to* m **do**
  - **while** *there exists* $j_0 < j$ *with* $low(j_0) = low(j)$ **do**
    - Add column ∂_{j_0} to ∂_j
    - Add column V_{j_0} to V_j
- **for** $j = 1$ *to* m **do**
  - **if** $\partial_j \neq 0$ **then**
    - *pair* $(low(j), j)$
- Return V

For complex K and the filtration $\mathcal{D}_v$ let $\zeta_1, \ldots, \zeta_s$ be the essential $p-$cycles computed by the standard reduction algorithm with $\kappa(\zeta_1) \prec_v \ldots \prec_v \kappa(\zeta_s)$. The essential cycles form a basis of the p^{th} homology of K that is, $[\zeta_1], \ldots, [\zeta_s]$ form a basis of $H_p(\mathsf{K})$. Let ∂ be the boundary matrix corresponding to $\mathcal{D}_v$ and $\tilde{\partial}$ be the reduced matrix. If $\tilde{\partial}_i$ is 0 and i is a paired index then the cycle V_i is a $p-$boundary in K. If j is an index in the filtration where the simplex τ appears and K_j the subcomplex consisting of all simplices till index j of the filtration, then the cycles $Z_j = \{V_i \,|\, \text{i is an index that corresponds to a p-simplex, } \tilde{\partial}_i = 0,\, i \leq j\}$ form a basis of $Z_p(\mathsf{K}_j)$.

We now state a few propositions that follow from the standard reduction algorithm.

Proposition 28. *Let $V_i \in Z_j$ be a p-cycle. Assume that i is a paired index, that is, V_i is not an essential cycle. Let $V_{i_1}, \ldots, V_{i_s}$ be the essential $p-$cycles in Z_j. Then $[V_i] \in span\{[V_{i_1}], \ldots, [V_{i_s}]\}$ in $H_p(\mathsf{K})$.*

Proof. Let $j_1, \ldots, j_t = i$ be the paired indices $\leq i$. Assume $j_1 < \ldots < j_t$. We proceed by induction on t. Since j_1 is paired there exists a $p-$boundary $\tilde{\partial}_{j_1'}$ such that $low(\tilde{\partial}_{j_1'}) = j_1$. Therefore $low(\tilde{\partial}_{j_1'} + V_{j_1}) < j_1$ and so $\tilde{\partial}_{j_1'} + V_{j_1} \in Z_p(\mathsf{K}_{j_1 - 1})$. But $Z_p(\mathsf{K}_{j_1-1}) \subset span\{V_{i_1}, \ldots, V_{i_s}\}$ and so $[V_{j_1}] \in span\{[V_{i_1}], \ldots, [V_{i_s}]\}$. For $t > 1$ argueing as before we have $\tilde{\partial}_{j_t'} + V_{j_t} \in Z_p(\mathsf{K}_{j_t - 1})$. But $Z_p(\mathsf{K}_{j_t-1}) \subset span\{V_{i_1}, \ldots, V_{i_s}, V_{j_1}, \ldots, V_{j_{t-1}}\}$ and so by the inductive hypothesis $[V_{j_t}] \in span\{[V_{i_1}], \ldots, [V_{i_s}]\}$.

Proposition 29. *Let $\zeta_1 \prec_v \ldots \prec_v \zeta_s$ be the essential $p-$cycles computed by the standard reduction algorithm for $\mathcal{D}_v(\mathsf{K})$. If $\zeta \in Z_p(\mathsf{K})$ be such that $\zeta_i \prec_v \zeta \prec_v \zeta_{i+1}$ then $[\zeta] \in span\{[\zeta_1] \ldots [\zeta_i]\}$.*

Proof. $\zeta_i \prec_v \zeta \prec_v \zeta_{i+1} \implies \kappa(\zeta_i) \prec_v \kappa(\zeta) \prec_v \kappa(\zeta_{i+1})$. Let $\tau = \kappa(\zeta)$ and j be the index where τ appears. $Z_j \supset \{\zeta_1, \ldots \zeta_i\}$ forms a basis of $Z_p(\mathsf{K}_j)$. Write $\zeta = a_1\zeta_1 + \ldots + a_i\zeta_i + b$ where b is a cycle spanned by $Z_j \setminus \{\zeta_1, \ldots \zeta_i\}$. By proposition 28 for each $\xi \in Z_j \setminus \{\zeta_1, \ldots \zeta_i\}$, $[\xi] \in span\{[\zeta_1] \ldots [\zeta_i]\}$ in $H_p(\mathsf{K})$.

Proposition 30. *Let $\zeta_1 \prec_v \ldots \prec_v \zeta_s$ be as in Proposition 29. Let ζ* be a p-cycle, $[\zeta] \neq 0$ *in $H_p(\mathsf{K})$. such that $\zeta = \zeta_{i_1} + \ldots + \zeta_{i_m} + \partial c_{p+1}$ where each $\zeta_{i_k} \in \{\zeta_1, \ldots, \zeta_s\}$ and c_{p+1} is a $p+1$ chain. If $i_1 < \ldots < i_m$ then $r_v([\zeta]) = r_v(\zeta_{i_m})$. In particular, $\zeta_{i_1} + \ldots + \zeta_{i_m} \in \arg\min_{\xi \in [\zeta]} r_v(\xi)$.*

Proof. First note that $\kappa(\zeta_{i_1}) \prec_v \ldots \prec_v \kappa(\zeta_{i_m})$. So $r_v(\zeta_{i_1} + \ldots + \zeta_{i_m}) = r_v(\kappa(\zeta_{i_m})) = r_v(\zeta_{i_m})$. We prove by contradiction. Let $\hat{\zeta} \in \arg\min_{\eta \in [\zeta]} r_v \eta$. If $r_v(\hat{\zeta}) < r_v(\zeta_{i_m}) \implies \kappa(\hat{\zeta}) \prec_v \kappa(\zeta_{i_m}) \implies \hat{\zeta} \prec_v \zeta_{i_m}$. By the Proposition 29 we would have $[\hat{\zeta}] \in span\{[\zeta_1] \ldots [\zeta_{i_m - 1}]\}$. But by assumption $[\hat{\zeta}] = [\zeta_{i_1}] + \ldots + [\zeta_{i_m}]$.

Algorithm 5: Solve with Reduction

Input : $A(\mathbb{Z}_2 \text{ matrix}), b(\mathbb{Z}_2 \text{ vector})$
▷ computes a solution to $Ax = b$ if one exists
Output: If x is the solution computed, returns the indices of x equal to 1, returns *false* if no solution exists.

```
Procedure SolveByReduction(A, b)
    C ← [A|b]. Let s be the index of the last column of C
    V ← StandardReduction(C). Continue to denote the reduced matrix by C.
    if C_s = 0 then
        Let j_1, ..., j_q be the row indices of V_s that are 1. soln ← {j_1, ..., j_{q-1}}
        Return true, soln
    else
        Return false.
```

C.1 Optimal Homology Basis

Proposition 31. *Let $\zeta_1, \ldots, \zeta_m$ be the cycles in $\mathscr{M}$ such that $\zeta_1 \prec_\Omega \ldots \prec_\Omega \zeta_m$. Let $\zeta'_1, \ldots, \zeta'_m$ be any collection of cycles such that $span\{[\zeta'_1], \ldots, [\zeta'_m]\} = H_p(\mathsf{K})$. Assume further that if $r_P([\zeta'_i]) < r_P([\zeta'_j])$ then $i < j$. Then $r_P(\zeta_j) \leq r_P([\zeta'_j]), 1 \leq j \leq m$.*

Proof. Let ζ_j correspond to cycle $\zeta_{u,\tilde{j}}, u \in S$ in the list Ω. Let $\Omega_- = \{\eta \in \Omega | \, \eta \prec_\Omega \zeta_j\}$ and $\Omega_+ = \{\eta \in \Omega | \, \zeta_j \prec_\Omega \eta\} \cup \{\zeta_j\}$. Let $[\Omega_-] = span\{[\eta], \eta \in \Omega_-\}$. First note that $r_P([\zeta_j]) = r_u(\zeta_j)$, for if $u^* \in \arg\min_{x \in S} r_x([\zeta_j])$ and say $\zeta_j = \zeta_{u^*,k_1} + \ldots + \zeta_{u^*,k_q} + b', \zeta_{u^*,k_1} \prec_{u^*} \ldots \prec_{u^*} \zeta_{u^*,k_q}, b' \in B_p(\mathsf{K})$, then $\zeta_{u,\tilde{j}} \prec_\Omega \zeta_{u^*,k_q}$. (This is because if $\zeta_{u^*,k_q} \in \Omega_- \implies [\zeta_j] \in [\Omega_-]$. But $[\zeta_j] \notin [\Omega_-]$). Thus $r_u(\zeta_{u,\tilde{j}}) \leq r_{u^*}(\zeta_{u^*,k_q}) = r_P([\zeta_j])$.
Now assume the contrary, that is, $r_P([\zeta'_j]) < r_P([\zeta_j]$. Let $v \in \arg\min_{v \in S} r_v([\zeta'_j])$. Let $\zeta'_j = \zeta_{v,i_1} + \ldots + \zeta_{v,i_s} + b$, where $b \in B_p(\mathsf{K})$ and $\zeta_{v,i_1} \prec_v \ldots \prec_v \zeta_{v,i_s}$ By Proposition 30, $r_v([\zeta'_j]) = r_v(\zeta_{v,i_s}) = r_P([\zeta'_j])$. $r_P([\zeta'_j]) < r_P([\zeta_j] \implies r_v(\zeta_{v,i_l}) < r_P(\zeta_j)$, $1 \leq l \leq s$. We have $\zeta_{v,i_1}, \ldots, \zeta_{v,i_s} \in \Omega_-$. Clearly $dim([\Omega_-]) = j - 1$. The classes $\{[\zeta'_1], \ldots [\zeta'_{j-1}]\}$ cannot all be in $[\Omega_-]$ as $[\zeta'_j] \in [\Omega_-]$ and the classes $\{[\zeta'_1], \ldots [\zeta'_j]\}$ are assumed to be linearly independent. Let $\zeta'_k, k < j$ be such that $[\zeta'_k] \notin [\Omega_-]$. Let $w \in \arg\min_{x \in S} r_x([\zeta'_k])$. Let $\zeta'_k = \zeta_{w,j_1} + \ldots + \zeta_{w,j_t} + b'$ where $b' \in B_p(\mathsf{K})$ and $\zeta_{w,j_1} \prec_w \ldots \prec_w \zeta_{w,j_t}$. Then $[\zeta'_k] \notin \Omega_- \implies \zeta_{w,j_t} \in \Omega_+$. In particular, $r_P([\zeta'_k]) = r_w(\zeta_{w,j_t}) \geq r_u(\zeta_j) = r_P([\zeta_j]) > r_P([\zeta'_j])$. This violates the assumption that the cycles $\{\zeta'_i\}$ are ordered by $r_P([\zeta'_i])$.

C.2 Minimal Persistent Homology Basis

In the following propositions we assume a simplex-wise filtration $\mathcal{F}$ on K, $\partial_\mathcal{F}$ is the corresponding boundary matrix, $[b, d) \in \mathcal{B}_p(\mathcal{F})$ and σ_b is the simplex added at index b of $\mathcal{F}$.

Proposition 32. *If ω is a cycle in K_b such that σ_b is incident on ω then ω is not a boundary in K_{d-1}.*

Proof. Let $\partial_\mathcal{F}$ be the boundary matrix corresponding to $\mathcal{F}$, $\widetilde{\partial_\mathcal{F}}$ be the reduced boundary matrix after applying standard reduction on $\partial_\mathcal{F}$ and V be the output cycle vectors. For any j that corresponds to an index in $\mathcal{F}$ the non-zero columns of $\widetilde{\partial_\mathcal{F}}$ with index $\leq j$ form a basis for the boundaries of K_j. Let $\rho(\sigma)$ be the index in $\mathcal{F}$ where the simplex σ appears. For a cycle ζ, let $\rho(\zeta) = \max_{\sigma \in \zeta} \rho(\sigma)$. Since $\omega \in Z_p(\mathsf{K}_b)$ and $\sigma_b \in \omega \implies \rho(\omega) = b$. Since $\mathcal{F}$ is simplex-wise a death index d can occur in at most 1 bar. Therefore $\nexists$ a boundary $\widetilde{\partial_{\mathcal{F},i}}, i < d$ with $\rho(\widetilde{\partial_{\mathcal{F},i}}) = b$. Since all non-zero columns $\widetilde{\partial_{\mathcal{F},i}}$ have unique values of ρ, if η is a boundary that is a linear combination of cycles $\{\widetilde{\partial_{\mathcal{F},i}}\}, i < d$, then $\rho(\eta) \neq b$. It follows that ω is a not a boundary in K_{d-1}.

Proposition 33. *Let $\zeta_1 \prec_v \ldots \prec_v \zeta_m$ be the essential $p-$cycles of $\mathcal{D}_v(\mathsf{K}_b)$ computed using standard reduction. Let $\xi^* \in \arg\min_{\eta \in \mathcal{R}([b,d))}\{r_v(\eta)\}$. If $[\xi^*] \in span\{[\zeta_1], \ldots, [\zeta_i]\}$ in $H_p(\mathsf{K}_b)$, then $\exists \zeta^* \in span\{\zeta_1, \ldots, \zeta_i\}, \zeta^* \in$* $\arg\min_{\eta \in \mathcal{R}([b,d))}\{r_v(\eta)\}$.

Proof. Let $\xi^* = a_1\zeta_1 + \ldots + a_i\zeta_i + b$ where $a_i = 1$ and $b \in B_p(\mathsf{K}_b)$. Then $\kappa(\xi^*) \not\prec_v \kappa(\zeta_i)$ for otherwise by Proposition 29 we would have $[\xi^*] \in span\{[\zeta_1], \ldots, [\zeta_{i-1}]\}$ in $H_p(\mathsf{K}_b)$. It follows that $\zeta^* = \xi^* + b = a_1\zeta_1 + \ldots + a_i\zeta_i \in \arg\min_{\eta \in \mathcal{R}([b,d))}\{r_v(\eta)\}$.

Correctness of algorithm 3 Algorithm 3 correctly computes a representative of the input bar $[b, d)$ as the cycle contains σ_b and is a boundary at K_d. By appendix C.2 it is not a boundary in K_{d-1}. By proposition 33 the subroutine Opt-Pers-Cycle-Site correctly computes ζ_v^* for each site v.

D Further Experimental Results

D.1 Homology Localization

Our localization algorithm is robust to these inputs and produces cycles that correspond to the relevant geometric features in the 3D models. We can visually infer for the Wheel and Happy Buddha datasets that the localized cycle computed by our approximate algorithm is nearly equal to the optimal cycle. Note that even though the input cycle spans both tunnels in the Happy Buddha dataset, the localized cycle is a sum of the two disjoint cycles. These experimental results demonstrate how the localization algorithm may be utilized as a supplement to existing methods that compute cycle representatives.

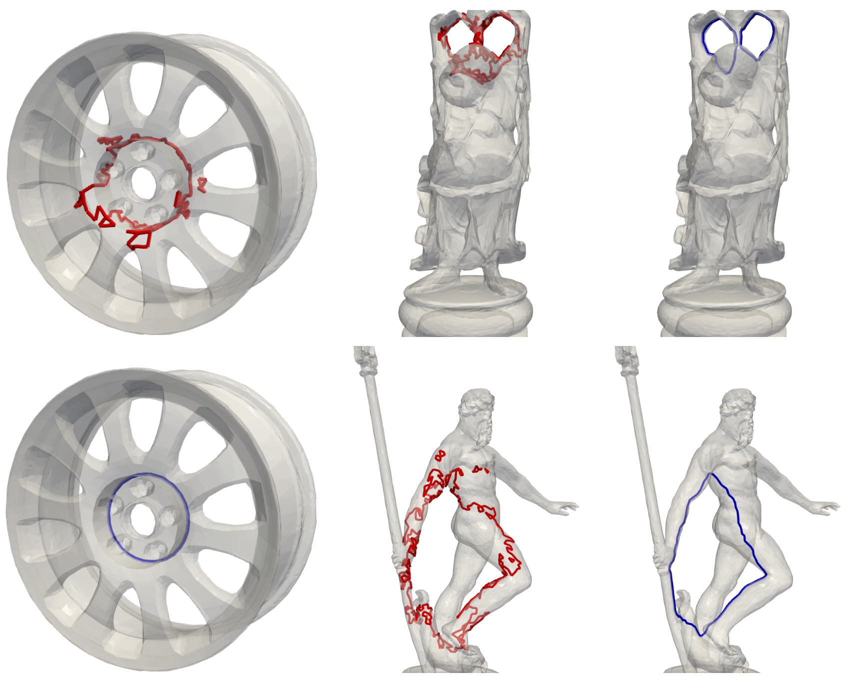


Figure 5: The localization algorithm computes optimal (blue) 1-cycles that are homologous to input 1-cycles (red).

D.2 Optimal homology basis

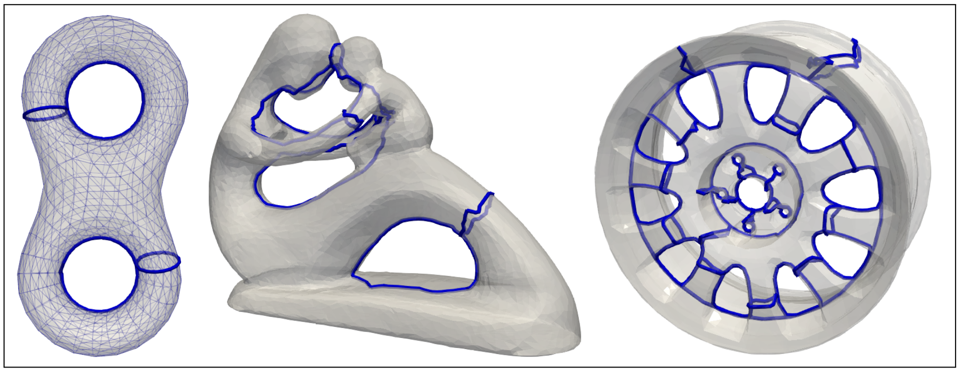
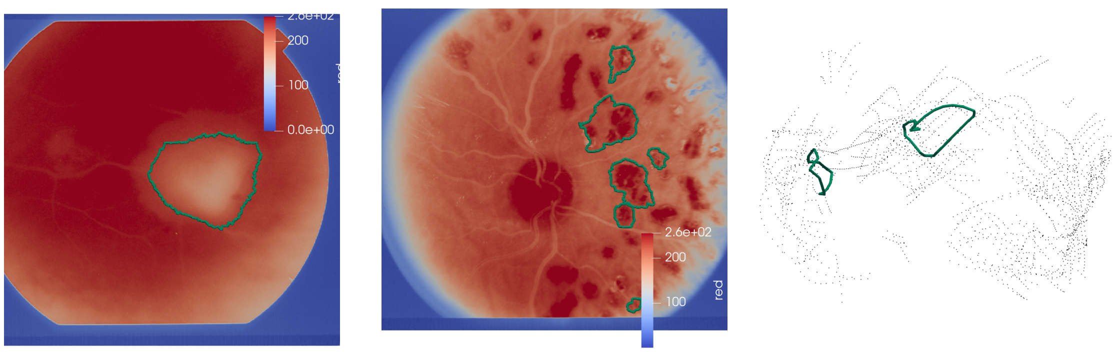


Figure 6: Optimal 1-homology basis computed using the localization algorithm.

D.3 Optimal Persistent Homology Representatives



(a) **(b)** **(c)**

Figure 7: **(a)**Persistent homology representative of longest bar bounding a region of disorder in the retinal image. **(b)** Representatives of few of the top-most bars indicating regions of disorder in the retinal image. **(c)** Top 2 representatives of the Lorenz'96 dataset(20%)