Approximate Čech Complexes in Low and High Dimensions

Aruni Choudhary* Michael Kerber† R. Sharathkumar‡

Abstract

Čech complexes reveal valuable topological information about point sets at a certain scale in arbitrary dimensions, but the sheer size of these complexes limits their practical impact. While recent work introduced approximation techniques for filtrations of (Vietoris-)Rips complexes, a coarser version of Čech complexes, we propose the approximation of Čech filtrations directly.

For fixed dimensional point set S, we present an approximation of the Čech filtration of S by a sequence of complexes of size linear in the number of points. We generalize well-separated pair decomposition (WSPD) to well-separated simplicial decomposition (WSSD) in which every simplex defined on S is covered by some element of WSSD. We give an efficient algorithm to compute a linear-sized WSSD in fixed dimensional spaces. Using a WSSD, we then present a linear-sized approximation of the filtration of Čech complex of S.

We also present a generalization of the known fact that the Rips complex approximates the Čech complex by a factor of $\sqrt{2}$. We define a class of complexes that interpolate between Čech and Rips complexes and that, given any parameter $\varepsilon > 0$, approximate the Čech complex by a factor $(1+\varepsilon)$. Our complex can be represented by roughly $O(n^{\lceil 1/2\varepsilon \rceil})$ simplices without any hidden dependence on the ambient dimension of the point set. Our results are based on an interesting link between Čech complex and coresets for minimum enclosing ball of high-dimensional point sets. As a consequence of our analysis, we show improved bounds on coresets that approximate the radius of the minimum enclosing ball.

1 Introduction

Motivation A common theme in topological data analysis is the analysis of point cloud data representing an unknown manifold. Although the ambient space can be high-dimensional, the manifold itself is usually of relatively low dimension. Manifold learning techniques try to infer properties of the manifold, like its dimension or its homological properties, from the point sample.

An early step in this pipeline is to construct a cell complex from the point sample which shares similarities with the hidden manifold. The *Čech complex at scale* α (with $\alpha \geq 0$) captures the intersection structure of balls of radius α centered at the input points. More precisely, it is the

*Max Planck Institute for Informatics, Saarbrücken, Germany (aruni.choudhary@mpi-inf.mpg.de)
†Max Planck Institute for Informatics, Saarbrücken, Germany (mkerber@mpi-inf.mpg.de)
‡Stanford University, Stanford, USA (sharathk@stanford.edu)

nerve of these balls, and is therefore homotopically equivalent to their union. Increasing α from 0 to ∞ yields a *filtration*, a sequence of nested Čech complexes, which can serve as the basis of multi-scale approaches for topological data analysis.

A notorious problem with Čech complexes is their representation: Its k-skeleton can consist of up to $O(n^{k+1})$ simplices, where n is the number of input points. Moreover, its construction requires the computation of minimum enclosing balls of point sets; we will make this relation explicit in Section 2. A common workaround is to replace the Čech complex by the *(Vietoris-)Rips complex* at the same scale α. Its definition only depends on the diameter of point sets and can therefore be computed by only looking at the pairwise distances. Although Rips complexes permit a sparser representation, they do not resolve the issue that the final complex can consist of a large number of simplices; Sheehy [25] and Dey et al. [11] have recently addressed this problem by defining an approximate Rips filtration whose size is only linear in the input size. On the other hand, efficient methods for approximating minimum enclosing balls have been established, even for high-dimensional problems, whereas the diameter of point sets appears to be a significantly harder problem in an approximate context. This suggests that Čech complexes might be more suitable objects than Rips complexes in an approximate context.

Contribution We give two different approaches to approximate filtrations of Čech complexes, both connecting the problem to well-known concepts in discrete geometry: The first approach yields, for a fixed constant dimension, a sequence of complexes, each of linear size in the number of input points, that approximate the Čech filtration. By approximate, we mean that the *persistence diagrams* of exact and approximate Čech filtration differ by an arbitrarily small multiplicative factor. To achieve this result, we generalize the famous *well-separated pair decomposition (WSPD)* to a higher-dimensional analogue, that we call the *well-separated simplicial decomposition (WSSD)*. A k-tuple in the WSSD can be viewed as k clusters of points of S with the property that whenever a ball contains at least one point of each cluster, a small expansion of the ball contains all points in all clusters. Furthermore, these tuples cover every simplex with vertices in S, i.e., given any k-simplex σ, there is a $(k+1)$-tuple of clusters such that each cluster contains one vertex of σ. We consider the introduction of WSSDs to be of independent interest: given the numerous applications of WSPD, we hope that its generalization will find further applications in approximate computational topology.

In the conference version of this paper [21], we presented a quadtree-based algorithm to construct WSSDs whose size depends on the ambient dimension d. In this version, we transfer the dependence from d to the *doubling dimension* Δ of the input point set. This requires us to re-express our constructions of WSSD and Čech complexes to use *hierarchical net-trees*, the generalization of quadtrees for spaces with small doubling dimension. This approach matches related work on the Rips filtration [25, 11] and makes our methods applicable even when the input set lies on a low-dimensional manifold in a high-dimensional space.

As our second contribution, we prove a generalized version of the well-known Vietoris-Rips lemma [13, p.62] which states that the Čech complex at scale α is contained in the Rips complex at scale $\sqrt{2}\alpha$. We define a family of complexes, called *completion complexes* such that for any ε, the Čech complex at scale α is contained in a completion complex at scale $(1+\varepsilon)\alpha$. These completions complexes are parametrized by an integer k; the k-completion is completely determined by its k-skeleton, consisting of up to $O(n^{k+1})$ complexes. To achieve $(1+\varepsilon)$-closeness to the Čech complex, we need to set $k \approx 1/(2\varepsilon)$ (see Theorem 29 for the precise statement); in particular, there is no dependence on the ambient dimension to approximate the Čech complex arbitrarily closely.

For proving this result, we use *coresets* for minimum enclosing ball (meb) [5]: the meb of a set of points can be approximated by selecting only a small subset of the input which is called a *coreset*; here approximation means that an ε-expansion of the meb of the coreset contains all input points. The size of the smallest coreset is at most $\lceil 1/\varepsilon \rceil$, independent of the number of points and the ambient dimension, and this bound is tight [5]. To obtain our result, we relax the definition of coreset for minimum enclosing balls. We only require the *radius* of the meb to be approximated, not the meb itself. We prove that even smaller coresets of size roughly $\lceil 1/(2\varepsilon) \rceil$ always exist for approximating the radius of the meb. Again, we consider this coreset result to be of independent interest.

Related work We use the hierarchical net-trees construction from Har-Peled and Mendel [18] in our algorithm; they compute a sequence of nested clusters called *nets*, representing the point set at different scales. Their algorithm takes $2^{O(\Delta)} n \log n$ time, where Δ denotes the doubling dimension of the point set of size n. Numerous applications, such as WSPD construction, approximate nearest neighbors, or metric spanners can be computed with a net-tree as the basic data structure; our work adds WSSDs as a new item to that list.

Sparse representation of complexes based on point cloud data are a popular subject in current research. Standard techniques are the *alpha complex* [14, 15] which contains all Delaunay simplices up to a certain circumradius (and their faces), *simplex collapses* which remove a pair of simplices from the complex without changing the homotopy type (see [2, 22, 27] for modern references), and *witness approaches* which construct the complex only on a small subset of landmark points and use the other points as witnesses [10, 3, 12]. A more extensive treatment of some of these techniques can be found in [13, Ch.III]. Another recent approach [24] constructs Rips complexes at several scales and connects them using *zigzag persistence* [6], an extension to standard persistence which allows insertions and deletions in the filtration. The aforementioned work by Sheehy [25] combines this theory with *net-trees* to get an approximate linear-size zigzag-filtration of the Rips complex in a first step and finally shows that the deletions in the zigzag can be ignored. Dey et al. [11] arrive at the same result more directly by constructing an hierarchy of ε-nets, defining a filtration from it where the elements are connected by simplicial maps instead of inclusions, and finally showing that this filtration is *interleaved* with the Rips-filtration in the sense of [7].

Outline We will introduce basic topological and geometric concepts in Section 2. Then we introduce WSSDs, our generalization of WSPDs and give an algorithm to compute them in Section 3. We show how to use WSSDs to approximates the persistence diagram of the Čech complex in Section 4. The existence of small coresets for approximating the radius of the meb is the subject of Section 5. k-completions and the generalized Vietoris-Rips Lemma are presented in Section 6. We conclude in Section 7.

2 Preliminaries

Simplicial complexes Let S denote a finite set of universal elements, called *vertices* [1] A *(simplicial) complex* C is a collection of subsets of S, called *simplices*, with the property that whenever a simplex σ is in C, all its (non-empty) subsets are in C as well. These non-empty subsets are called the *faces* of σ; a *proper face* is a face that is not equal to σ. Setting $k := |\sigma| - 1$, where $|\cdot|$ stands for the number of elements considered as a subset, we call σ a *k-simplex*. For a k-simplex $\sigma = \{v_0, \ldots, v_k\}$, we call $v_0, \ldots, v_k$ its *boundary vertices* of σ; we will also frequently write σ as a tuple of its boundary vertices, that is, $\sigma = (v_0, \ldots, v_k)$ with the convention that any permutation of the boundary vertices yields the same simplex. A *subcomplex* of C is a simplicial complex that is contained in S. One example of a subcomplex is the *k-skeleton* of a complex C, which is the set of all ℓ-simplices in C with $\ell \leq k$. Let K and K' be two simplicial complexes with vertex sets V and V' and consider a map $f : V \to V'$. If for any simplex $(v_0, \ldots, v_k)$ of K, $(f(v_0), \ldots f(v_k))$ yields a simplex in K', then f extends to a map from K to K' which we will also denote by f; in this case, f is called a *simplicial map*.

Let S be a set of arbitrary geometric objects, embedded in an ambient space $\mathbb{R}^d$. We call $|S| := \cup_{s \in S} s \subset \mathbb{R}^d$ the *union of* S. We define a simplicial complex C as follows: A k-simplex σ is in C if the corresponding $k+1$ objects have a common intersection in $\mathbb{R}^d$. It is easy to check that C is indeed closed under face relations and thus a simplicial complex with vertex set S, called the *nerve* of S. The famous *Nerve Theorem* [13, p.59] states that if all objects in S are convex, the union of S and its nerve are *homotopically equivalent*. This intuitively means that one can transform one into the other by bending, shrinking and expanding, but without gluing and cutting. A consequence of this theorem is that the *homology groups* of the union and the nerve are equal. We will give an intuitive meaning of homology groups later in this section; see [13, 23] for thorough introductions to homology.

For a finite point set P and $\alpha > 0$, the *Čech complex* $\mathscr{C}_\alpha(P)$ is the nerve of the set of (closed) balls of radius α centered at the points in P. Note that a k-simplex of the Čech complex can be identified with $(k+1)$ points $p_0, \ldots, p_k$ in P, the centers of the intersecting balls. Let $\mathrm{meb}(p_0, \ldots, p_k)$ denote the *minimum enclosing ball of* P, that is, the ball with minimal radius that contains each p_i.

Observation 1. *A k-simplex $\{p_0, \ldots, p_k\}$ is in $\mathscr{C}_\alpha(P)$ iff the radius of* $\mathrm{meb}(p_0, \ldots, p_k)$ *is at most α.*

A widely used approximation of Čech complexes is the *(Vietoris)-Rips complex* $\mathscr{R}_\alpha(P)$. It is defined as the maximal simplicial complex whose 1-skeleton equals the 1-skeleton of the Čech complex. Described as an iterative construction, starting with the edges of the Čech complex, a triangle is added to the Rips complex when its three boundary edges are present, a tetrahedron when its four boundary triangles are present, and so forth. The Rips complex is an example of a *clique complex* (also known as *flag complex* or *Whitney complex*). That means, it is completely determined by its 1-skeleton which in turn only depends on the pairwise distance between the input points. For $k+1$ points $p_0, \ldots, p_k$ in P, let the *diameter* $\mathrm{diam}(p_0, \ldots, p_k)$ denote the maximal pairwise distance between any two points p_i and p_j with $0 \leq i \leq j \leq k$.

Observation 2. *A k-simplex $\{p_0, \ldots, p_k\}$ is $\mathscr{R}_\alpha(P)$ iff* $\mathrm{diam}(p_0, \ldots, p_k)$ *is at most α.*

[1] Some of the defined concepts do not require that S is finite; however, since we will only deal with finite complexes in later sections, we decided to discuss this simpler setup.

For notational convenience, we will often omit the P from the notation and write $\mathscr{C}_\alpha$ and $\mathscr{R}_\alpha$ when P is clear from context.

Persistence modules For $A \subset \mathbb{R}$, a *persistent module* is a family $(F_\alpha)_{\alpha\in A}$ of vector spaces with homomorphisms $f_\alpha^{\alpha'}: F_\alpha \to F_{\alpha'}$ for any $\alpha \le \alpha'$ such that $f_{\alpha'}^{\alpha''} \circ f_\alpha^{\alpha'} = f_\alpha^{\alpha''}$ and f_α^α is the identity function.[2] The most common class are modules induced by a *filtration*, that is, a family of complexes $(C_\alpha)_{\alpha\in A}$ such that $C_\alpha \subseteq C_{\alpha'}$ for $\alpha \le \alpha'$. For some fixed dimension p, set $H_\alpha := H_p(C_\alpha)$, the p-th homology group of C_α. The inclusion map from C_α to $C_{\alpha'}$ induces an homomorphism $\hat{f}_\alpha^{\alpha'}: H_\alpha \to H_{\alpha'}$ and turns $(H_\alpha)_{\alpha\in\mathbb{R}}$ into a persistence module. Example of such filtrations and their induced modules are the *Čech filtration* $(\mathscr{C}_\alpha)_{\alpha\ge 0}$ and the *Rips filtration* $(\mathscr{R}_\alpha)_{\alpha\ge 0}$. However, we will also consider persistence modules which are not induced by filtrations. Generalizing the case of filtrations, given a sequence of simplicial complexes $(\mathscr{A}_\alpha)_{\alpha\in A}$ connected by simplicial maps $g_\alpha^{\alpha'}: \mathscr{A}_\alpha \to \mathscr{A}_{\alpha'}$ which satisfy $g_{\alpha'}^{\alpha''} \circ g_\alpha^{\alpha'} = g_\alpha^{\alpha''}$ and $g_\alpha^\alpha = \mathrm{id}$, the induced homology groups $H_\alpha := H_p(\mathscr{A}_\alpha)$ and induced homomorphisms $\hat{g}_\alpha^{\alpha'}: H_\alpha \to H_{\alpha'}$ also yield a persistence module. A persistence module $(F_\alpha)_{\alpha\in A}$ is *tame* if the rank of F_α is finite for all $\alpha \in A$. As our modules in this work will consist only of homology groups over finite simplicial complexes, all modules constructed in this paper will be tame, and we will ignore this technicality from now on when referring to previous results. We will frequently denote filtrations and modules by F_* instead of $(F_\alpha)_{\alpha\in A}$ for brevity if there is no confusion about A.

For a persistence module F_* with homomorphisms $f_\alpha^{\alpha'}$, we say that a generator (basis element) $\gamma \in F_\alpha$ is *born* at α if $\gamma \notin \mathrm{Im} f_{\alpha-\varepsilon}^{\alpha}$ for any $\varepsilon > 0$, where Im is the image of a map. If γ is born at α, we say that it *dies* at α' if α' is the smallest value such that $f_\alpha^{\alpha'}(\gamma) \in \mathrm{Im} f_{\alpha-\varepsilon}^{\alpha'}$ for some $\varepsilon > 0$. In other words, every generator can be represented by a point in the plane, determining its birth- and death-coordinate. F_* is completely characterized by this multiset of points, which is called the *persistence diagram* of the module and denote it as $\mathrm{Dgm} F_*$. Note that all points of the diagram lie on or above the diagonal in the birth-death-plane.

For the benefit of readers inexperienced with the concept of persistence, we explain the wealth of geometric-topological information contained in the persistence diagram, exemplified on a Čech filtration of a point set S in $\mathbb{R}^3$. As discussed, we can visualize the filtration as a sequence of growing balls centered at the points in S, and the union of these balls forms a sequence of growing shapes. During this process, the shape might create *voids*, that is, pockets of air completely enclosed by the shape. The rank of the second homology group $H_2(\mathscr{C}_\alpha)$ yields the number of voids present at a fixed scale α (this rank is also called the 2nd *Betti number*). The persistence diagram for $H_2(\mathscr{C}_*)$ provides multi-scale information about the voids in the process: every point (b,d) of the diagram represents a void that is formed for $\alpha = b$ and filled up for $\alpha = d$. The same information as for voids can be obtained for *connected components* and for *tunnels*, choosing the 0th and 1st homology groups, respectively.

Approximating persistence diagrams An important property of persistence diagrams is their stability under "small" perturbations of the underlying filtrations and modules; see Cohen-Steiner et al. [9] for the precise first statement of this type. We will use the more recent results by Chazal et al. [7] for this work, following Sheehy's notations and definitions [25]. For two modules F_*, G_*,

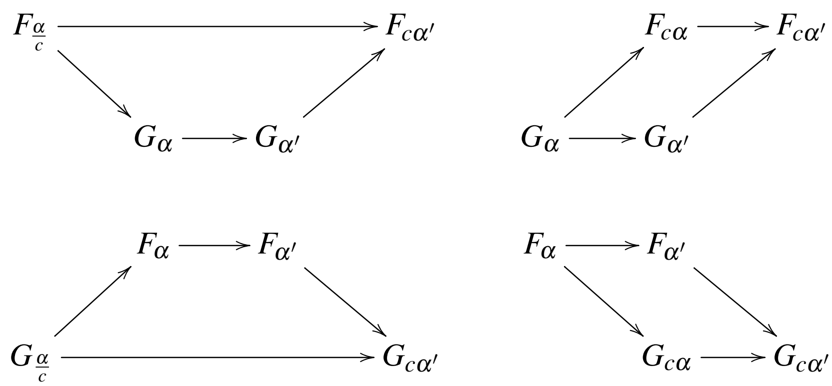
[2] This is not the most general definition of a persistent module; see [7].

we say that $\mathrm{Dgm}\, F_*$ is a *c-approximation* of $\mathrm{Dgm}\, G_*$ with $c \geq 1$ if there is a bijection $\pi : \mathrm{Dgm}\, F_* \to \mathrm{Dgm}\, G_*$ such that for any point (x,y) of $\mathrm{Dgm}\, F_*$, $\pi(x,y)$ lies in the axis-aligned box defined by $\frac{1}{c}(x,y)$ and $c(x,y)$. An equivalent statement is that the two diagrams have a bounded bottleneck distance on the log-scale.

We will use the following result which is a reformulation of [7, Def.4.2+Thm.4.4]:

Theorem 3. *Let $(F_\alpha)_{\alpha \geq 0}$ and $(G_\alpha)_{\alpha \geq 0}$ be two persistence module with two families of homomorphisms $\{\phi : F_\alpha \to G_{c\alpha}\}_{\alpha \geq 0}$ and $\{\psi : G_\alpha \to F_{c\alpha}\}_{\alpha \geq 0}$ such that all the following diagrams commute:*

(2.1)



Then, the persistence diagrams of F_α and G_α are c-approximations of each other.

In the case of modules induced by filtrations, there is a simple corollary, called the "Persistence Approximation Lemma" in [25]:

Lemma 4. *If two filtrations $(A_\alpha)_{\alpha \geq 0}$ and $(B_\alpha)_{\alpha \geq 0}$ satisfy $A_{\frac{\alpha}{c}} \subset B_\alpha \subset A_{c\alpha}$ for all $\alpha \geq 0$, then the persistence diagrams are c-approximations of each other.*

Doubling spaces A *discrete ball* centered at a point $q \in P$ with radius r is the set of points $Q \subseteq P$ which satisfy $\|p - q\| \leq r$ for all $q \in Q$. The *doubling constant* [1, 26] is the smallest integer λ such for all $p \in P$ and all $r > 0$, the discrete ball centered at p of radius r is covered by λ discrete balls of radius $r/2$. The *doubling dimension* Δ is $\lceil \log_2 \lambda \rceil$. Metric spaces with constant doubling dimension are called *doubling spaces*. A simple example of a doubling space is a point set in $\mathbb{R}^d$ (with arbitrary d) sampled from an affine space of constant dimension, using the Euclidean metric. In contrast, the boundary points of the standard $(n-1)$-simplex do not form a doubling space since their doubling dimension is $\lceil \log_2 n \rceil$. It is NP-hard to calculate the doubling dimension of a metric [17] but it can be approximated within a constant factor [18, Sec.9].

Net-trees A subset $Q \subseteq P$ is an (α, β)*-net*, denoted by $\mathcal{N}_{\alpha,\beta}$, if all points in P are in distance at most α from some point in Q and the distance between any two points in Q is at least β.

We can represent a nested sequence of nets for increasing scales α using a rooted tree structure, called the *net-tree* [18]. It has n leaves, each representing a point of P, and each internal node has at least two children. Every tree-node v represents the subsets of points given by the sub-tree rooted at v; we denote this set by P_v. Every v has a representative, $\mathrm{rep}_v \in P_v$ that equals the representative of one of its children if v is not a leaf. Moreover, v is associated with an integer $\ell(v)$ called the *level* of v which satisfies $\ell(v) < \ell(\mathrm{par}(v))$, where $\mathrm{par}(v)$ is the parent of v in the tree. Finally, the crucial properties of each node are given next; we let $\mathbb{B}(p,r)$ denote the ball centered at p with radius r, $\tau = 11$, and $\mathrm{scale}_v := \frac{2\tau^{\ell(v)+1}}{\tau-1}$

- *Covering property*: $P_v \subset \mathbb{B}(\mathrm{rep}_v, \mathrm{scale}_v)$
- *Packing property*: $P_v \supset P \bigcap \mathbb{B}(\mathrm{rep}_v, \frac{\tau-5}{4\tau^2}\mathrm{scale}_{(\mathrm{par}(v))})$

A simple consequence of the covering property is that

$$\mathrm{diam}_v := \mathrm{diam}(P_v) \leq 2\mathrm{scale}_v. \quad (2.2)$$

Moreover, the covering and packing properties ensure that each node v can admit at most $\lambda^{O(1)}$ children where λ is the doubling constant for P.

A net-tree can be constructed deterministically in time $2^{O(\Delta)} n \log(n \cdot \Phi)$ where Φ represents the *spread* of P, using the greedy clustering scheme of Gonzalez [16] as a precursor to the tree construction. The dependence on spread can be eliminated by constructing the tree in $2^{O(\Delta)}O(n \log n)$ time in expectation. Note that these bounds assume that distances between points can be computed in constant time in the underlying metric; since we restrict our attention to the Euclidean space, we do not need this simplifying assumption and multiply the bounds with $O(d)$, the time to compute the Euclidean distance of two points. The net-tree construction is oblivious to knowing the value of Δ. One can extract a net at scale by collecting the representatives of the highest nodes of level less than ℓ; see [18, Prop.2.2] for details.

The net-tree T can be augmented to maintain, for each node u, a list of close-by nodes with similar diameter. Specifically, for each node u the data structure maintains the set

$$\mathrm{Rel}(u) := \{v \in T \mid \ell(v) \leq \ell(u) < \ell(\mathrm{par}(v)) \text{ and } \|\mathrm{rep}_u - \mathrm{rep}_v\| \leq 14\tau^{\ell(u)}\}.$$

Computing these sets does not increase the running time of the construction.

3 Well-separated simplicial decomposition

In this section, we introduce the notion of Well-separated simplicial decomposition (WSSD) of point sets. WSSD can be seen as a generalization of well-separated pair decomposition of a point set. We first revisit the definition of WSPD and then generalize it to WSSD.

Notations Let $S \subset \mathbb{R}^d$ be a set of n points with doubling dimension Δ and let the corresponding doubling constant be λ. Let T denote a net tree for the point set. For a tuple $\gamma = (v_0, \ldots, v_k)$ where each v_i is a node of T, we denote by $\mathrm{meb}(\gamma)$ the minimum enclosing ball of the point set $\{\mathrm{rep}_{v_0}, \ldots, \mathrm{rep}_{v_k}\}$ where rep_{v_i} denotes the representative point of v_i. We denote the radius of $\mathrm{meb}(\gamma)$ by $\mathrm{rad}(\gamma)$.

For $e > 0$ and a ball $\mathbb{B}$ with center c and radius r, we let $e\mathbb{B}$ denote the ball with center c and radius $e \cdot r$. We state the following property, which follows directly by triangle inequality, but is used several times in our arguments:

Observation 5. *Let $\mathbb{B}$ be a ball with radius r that intersects a point set M whose diameter is at most λr for some $\lambda > 0$. Then, $M \subseteq \lambda\mathbb{B}$.*

Finally, whenever we make statements that depend on a parameter ε, it is implicitly assumed that $\varepsilon \in (0,1)$ from now on.

Well-Separated Pair Decomposition A pair of net-tree nodes $(u,v) \in T$ is said to be ε*-well-separated* if $max\{\text{diam}_u, \text{diam}_v\} \le \varepsilon \text{dist}(u,v)$, where $\text{dist}(u,v)$ denotes the minimal distance between a point in P_u and a point in P_v. Informally speaking, all pairs of points (p,q) with $p \in P_u$, $q \in P_v$ have a similar distance to each other if (u,v) is well-separated. We state a simple consequence which appears somewhat indirect, but allows a generalization to multivariate tuples:

Lemma 6. *If (q,q') with $q,q' \in T$ is ε-well separated, then for any ball $\mathbb{B}$ that contains at least one point of P_q and one point of $P_{q'}$, the ball $(1+2\varepsilon)\mathbb{B}$ contains all of P_q and all of $P_{q'}$.*

Proof. Let $\mathbb{B}$ be a ball with radius r intersecting both q and q', which means that $r \ge \text{dist}(q,q')/2$. Because (q,q') is well-separated,

$$\text{diam}_q \le \varepsilon \text{dist}(q,q') \le 2\varepsilon r,$$

implying that $(1+2\varepsilon)\mathbb{B}$ contains all of P_q by Observation 5. The same argument applies for q'. □

For a pair $(p,p') \in S \times S$ we say that a pair of net-tree nodes (q,q') *covers* (p,p') if $p \in q$ and $p' \in q'$, or $p \in q'$ and $p' \in q$. An ε*-well separated pair decomposition* (ε-WSPD) of S is a set of pairs $\Gamma = ((q_1,q'_1),(q_2,q'_2),\ldots,(q_m,q'_m))$ such that all pairs are ε-well separated and every edge in $S \times S$ is covered by some pair in Γ. We rely on the following property of WSPDs, adapted from [18, Sec.5]:

Theorem 7. *A ε-WSPD of size $n\varepsilon^{-O(\Delta)}$ can be computed in $O(n\log n + n/\varepsilon^{O(\Delta)})$ time, where Δ is the doubling dimension of the point set.*

Well-Separated Simplicial decomposition We generalize the construction of WSPD to higher dimensions: Let S and T be as above. A (k+1)-tuple $(v_0,v_1,\ldots,v_k)$ of net-tree nodes is called an ε*-well separated tuple (ε-WST)* if for any ball $\mathbb{B}$ which contains at least one point of each P_{v_i},

$$v_0 \cup v_1 \cup \ldots \cup v_k \subseteq (1+\varepsilon)\mathbb{B}$$

Moreover, we say that $(v_0,\ldots,v_k)$ *covers* a k-simplex $\sigma = (p_0,\ldots,p_k)$ with $p_0,\ldots,p_k \in S$ if there is a permutation π of $(0,\ldots,k)$ such that $p_{\pi(\ell)} \in P_{v_\ell}$ for all $0 \le \ell \le k$.

Definition 1. *An (ε,k)-WSSD is a set of $(k+1)$-tuples $\Gamma = \{\gamma_1,\gamma_2,\ldots,\gamma_l\}$ such that each γ_i is a ε-WST and each k-simplex is covered by some γ_i. An ε-WSSD is the union of (ε,k)-WSSDs for all $1 \le k \le \Delta$.*

It is easy to see with Lemma 6 that an $\frac{\varepsilon}{2}$-WSPD is an $(\varepsilon,1)$-WSSD.

Our algorithm We construct a WSSD recursively. If k=1, we use the WSPD algorithm from [18, Sec.5] to compute an $\frac{\varepsilon}{2}$-WSPD. For any $k>1$, we construct an (ε,k)-WSSD Γ_k recursively from an $(\varepsilon,k-1)$-WSSD Γ_{k-1} as follows: Each tuple $\gamma \in \Gamma_{k-1}$ is an ε-WST. Consider a tuple $\gamma = (v_0,v_1,v_2,\ldots,v_{k-1}) \in \Gamma_{k-1}$. Each v_i of this tuple represents a node in the net tree T. We compute an ε-approximation of $\text{meb}(\gamma)$ using the algorithm of [4] (recall that $\text{meb}(\gamma)$ is the meb of the representative points); more precisely, we compute a value r such that there is a ball of radius r covering all representatives and

$$\text{rad}(\gamma) \le r \le (1+\varepsilon)\text{rad}(\gamma) \qquad (3.1)$$

Recall that $\text{scale}_u = \frac{2\tau^{\ell(u)+1}}{\tau-1}$. We find the lowest ancestor u of v_0 in T that satisfies

$$\frac{\tau-5}{4\tau^2}\text{scale}_u \geq 8r \tag{3.2}$$

To find suitable nodes to be added to γ, we consider nodes $w \in \{u \cup \text{Rel}(u)\}$. For each such w, we traverse its sub-tree to find the highest descendants whose diameters are small enough for them to be valid candidates. Specifically, we find $w's$ highest descendants x satisfying

$$\frac{\varepsilon r}{4} \geq 2 \cdot \text{scale}_x \tag{3.3}$$

For each such node x, we add a $(k+1)$-tuple $(v_0, v_1, v_2, \ldots, v_{k-1}, x)$to Γ_k.

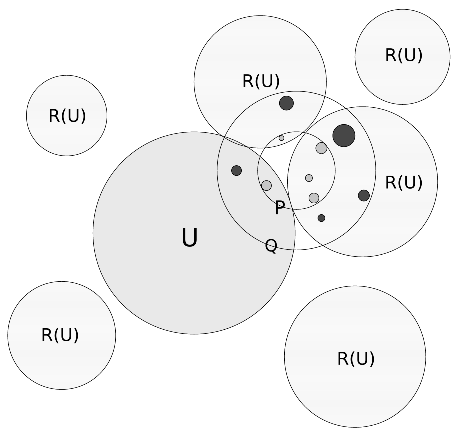


Figure 3.1: An example of constructing Γ_5 from Γ_4: The lightly shaded disks within ball P denote the nodes of a tuple $\gamma \in \Gamma_4$. Region U along with $\text{Rel}(U)$ depicted as $R(U)$ in the figure is examined for adding a node to γ. In this example, the dark disks in the region between P and Q are added to γ to form a tuple of Γ_5.

Correctness In order to prove the correctness of our construction procedure, we need to show that the generated tuples indeed form a (ε, k)-WSSD.

Lemma 8. *Each tuple added by the algorithm is an ε-WST.*

Proof. By induction: For $k = 1$, the statement is true as an $\frac{\varepsilon}{2}$-WSPD is an (ε,1)-WSSD, each tuple of which is an ε-WST. For $k > 1$, consider a k-tuple $\gamma = (v_0, v_1, v_2, \ldots, v_{k-1})$.Let x be added to this tuple to form the $(k+1)$-tuple $\gamma' = (v_0, v_1, v_2, \ldots, v_{k-1}, x)$. Now, consider a ball $\mathbb{B}$ that contains a point of each P_{v_i} and a point of P_x. By induction hypothesis, $(1+\varepsilon)\mathbb{B}$ contains all of P_{v_i}, so it suffices to show that it also contains all of P_x.

Let r' be the radius of $\mathbb{B}$. By Observation 5, it suffices to show that $\text{diam}_x \leq \varepsilon r'$. Using induction hypothesis again, the ball $(1+\varepsilon)\mathbb{B}$ contains all representatives rep_{v_i}. Therefore, it is an enclosing ball of the representatives and therefore, $(1+\varepsilon)r' \geq \text{rad}(\gamma)$. Since r as computed by the algorithm satisfies $r \le (1+\varepsilon)\mathrm{rad}(\gamma)$, we have that $r \le (1+\varepsilon)^2 r' \le 4r'$ since $\varepsilon \le 1$. Recall that $\mathrm{diam}_x \le 2\mathrm{scale}_x$ from 2.2. We can thus bound

$$\mathrm{diam}_x \le 2\mathrm{scale}_x \le \frac{\varepsilon r}{4} \le \varepsilon r'.$$

□

For showing that all k-simplices are covered, we need several preparatory results. The first one is taken from [5] – we note that the required bound also follows as a simple corollary of the main result of [5], but we decided to give a more low-level argument for clarity.

Lemma 9. *Let P be a point set with $|P| \ge 3$. Then, there exists a point $p \in P$ such that*

$$p \in \frac{1+1/d}{\sqrt{1-1/d^2}}\mathrm{meb}(P \setminus \{p\}).$$

In particular, $p \in 2\mathrm{meb}(P - \setminus\{p\})$ for $d \ge 2$.

Proof. Note that the statement is trivial if there exists a point $p \in P$ whose removal does not change the minimum enclosing ball. Therefore, assume wlog that $|P| \le d+1$, and all points of P are at the boundary of $\mathrm{meb}(P)$. Let c be the center and r be the radius of $\mathrm{meb}(P)$. The points in P span a polytope T; take the smallest ball $\mathbb{B}$ centered at c that is contained in T. By [5, Lem. 3.2], its radius is at most r/d. Moreover, $\mathbb{B}$ touches at least one facet of T. Let p be the point opposite of this facet, set $P' := P \setminus \{p\}$ and let c' and r' denote the center and radius of the meb of P'. Following the argumentation of [5, Lem. 3.3], it holds that

$$r' \ge r\sqrt{1-(1/d^2)}$$

and moreover, c' is the point where $\mathbb{B}$ touches the facet, so that $\|c-c'\| \le r/d$. Now, by triangle inequality

$$\|p-c'\| \le \|p-c\| + \|c-c'\| \le r + r/d \le (1+1/d)\frac{r'}{\sqrt{1-1/d^2}}$$

which implies the first claim. The second part follows easily by noting that

$$\frac{1+1/d}{\sqrt{1-1/d^2}} \le 2$$

for all $d \ge 5/3$. □

Lemma 10. *Let u be an internal node of the net tree T. Then*

$$\mathrm{diam}_u \ge \frac{\tau-5}{4\tau^2}\mathrm{scale}_u$$

Proof. Since u is internal, it has at least two children v_1, v_2. Let $p = \mathrm{rep}_{v_1}$. By the packing property of the net-tree, the ball around p with radius $\frac{\tau-5}{4\tau^2}\mathrm{scale}_{\mathrm{par}(v_1)}$ does only contain points of P_{v_1}. Since $\mathrm{par}(v_1) = u$ and v_2 contains at least one point not in P_{v_1}, the statement follows. □

Lemma 11. *Let u be a node of the net-tree T with* $\mathrm{diam}_u \geq 4R$*, where* $R > 0$*. Then, for any ball* $\mathbb{B}$ *of radius R which contains a point of P_u, the set* $\{u \cup \mathrm{Rel}(u)\}$ *covers* $S \cap 2\mathbb{B}$*.*

Proof. Consider a point $x \in S \cap 2\mathbb{B}$. The net tree T contains a path from the root to a leaf such that $x \in P_w$ for all nodes w on that path. In particular, there exists a node v with $x \in P_v$ and $\ell(v) \leq \ell(u) < \ell(\mathrm{par}(v))$. It can be observed from Figure 3.2 that

$$\|\mathrm{rep}_u - \mathrm{rep}_v\| \leq \mathrm{diam}_u + 3R + \mathrm{diam}_v < 3\mathrm{diam}_u < 14\tau^{\ell(u)}.$$

It follows that $v \in \mathrm{Rel}(u)$. □

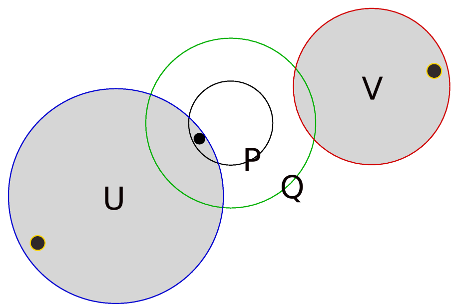


Figure 3.2: P represents a ball $\mathbb{B}$ of radius R and Q denotes $2\mathbb{B}$. Disk u denotes a node such that $diam(u) \geq 4R$ and it intersects $\mathbb{B}$. Disk v represents a node covering a point in $2\mathbb{B}$. The dark circles in u and v are their representatives.

Lemma 12. *Γ_k covers all k-simplices over S.*

Proof. By induction: For $k = 1$, the 1-simplices are pairs of points from S which are covered by the WSPD. Let Γ_{k-1} cover all $(k-1)$-simplices of S. Consider any k-simplex $\sigma = (p_0, p_1, \ldots, p_k)$. From Lemma 9, it follows that there exists a point, say p_k which belongs to $2\mathrm{meb}(\sigma')$, where $\sigma' := \sigma \setminus \{p_k\}$. σ' is a $(k-1)$-simplex and hence is covered by some k-tuple $\gamma \in \Gamma_{k-1}$. We show that when considering γ, the algorithm produces a $(k+1)$-tuple (γ, x) with $p_k \in P_x$, implying that σ is covered by that tuple.

Let r' be the radius of $\mathbb{B}' := \mathrm{meb}(\sigma')$, and let r be the approximate radius for $\mathrm{meb}(\gamma)$ as computed in the algorithm. Let $\mathbb{B}$ denote the corresponding enclosing ball of radius r. Since γ is an ε-WST, $(1+\varepsilon)\mathbb{B}$ contains $p_0, \ldots, p_{k-1}$. Therefore, we have that $r' \leq (1+\varepsilon)r \leq 2r$. Because of Lemma 10, the net-tree node u computed in the algorithm satisfies $\mathrm{diam}_u \geq 8r \geq 4r'$. Applying Lemma 11 on u, r', and $\mathbb{B}'$ yields that the set $\{u \cup \mathrm{Rel}(u)\}$ covers $S \cap 2\mathbb{B}'$. Thus, there is some node $w \in \{u \cup \mathrm{Rel}(u)\}$ that covers v_k. By construction, one of the descendants x of w for which a tuple (γ, x) is created satisfies $v_k \in P_x$. □

With Lemma 8 and Lemma 12, it follows that the constructed set Γ_k is an (ε, k)-WSSD.

Analysis. We bound the size of the (ε, k)-WSSD generated by our algorithm and the total time taken to compute it.

Lemma 13. *The size of (ε,k)-WSSD Γ_k is* $n(2/\varepsilon)^{O(\Delta \cdot k)}$

Proof. By induction: For $k = 1$, size of a $\frac{\varepsilon}{2}$-WSPD is $n(2/\varepsilon)^{O(\Delta)}$ which conforms with our hypothesis. Assume that the size of Γ_{k-1} is $n(2/\varepsilon)^{O(\Delta\cdot(k-1))}$. Consider any tuple $\gamma \in \Gamma_{k-1}$. Let r be the approximate radius of meb(γ) as computed in the algorithm. We bound the number of explored nodes. In the course of the algorithm, we find the lowest ancestor u of $v_0 \in \gamma$ which satisfies (3.2) and then explore the highest descendants of $\{u \bigcup \mathrm{Rel}(u)\}$ until we reach nodes x satisfying (3.3). Let us denote $\frac{\tau-5}{4\tau^2}$ by τ'. The child u' of u that is ancestor of v_0 satisfies $\tau' \cdot \mathrm{scale}_{u'} < 8r$ because the algorithm would not have chosen u otherwise. By a similar argument the parent x' of x satisfies $2 \cdot \mathrm{scale}_{x'} > \frac{\varepsilon r}{4}$. It follows that

$$\frac{\mathrm{scale}_{u'}}{\mathrm{scale}_{x'}} < \frac{\frac{8r}{\tau'}}{\frac{\varepsilon r}{8}} < \frac{64}{\varepsilon \tau'}.$$

Because $\frac{\mathrm{scale}_{\mathrm{par}(v)}}{\mathrm{scale}_v} \geq \tau$ for any node $v \in T$, u and x are at most $\ell := 2 + \log_\tau \frac{64}{\varepsilon\tau'}$ levels apart in the net-tree. Since any node has at most $\lambda^{O(1)}$ children, it immediately follows that the total number of nodes explored is $\lambda^{O(\ell)}$ which simplifies to $(2/\varepsilon)^{O(\Delta)}$, as λ=$2^{O(\Delta)}$. Because we add at most one tuple for each such node, the bound follows. □

Lemma 14. *Computing an ε-WSSD takes $2^{O(\Delta)} dn \log n + nd(2/\varepsilon)^{O(\Delta\cdot k)}$ time.*

Proof. For $k = 1$, Γ_1 is an $\frac{\varepsilon}{2}$-WSPD which requires $d\left(2^{O(\Delta)} n \log n + n(2/\varepsilon)^{O(\Delta)}\right)$ time to construct [18, Sec.5]; the additional factor of d is caused by the fact that distance computations require $O(d)$ time. For $k \geq 2$, consider the recursive construction of Γ_k from Γ_{k-1}. For each tuple $\gamma \in \Gamma_{k-1}$, we have to compute the approximate meb $\mathbb{B}$ and then explore ancestors and their descendants to find suitable nodes to be added.

Computing the approximate meb takes $O(kd/\varepsilon + \varepsilon^{-5})$ time [4]. From the arguments of Lemma 13, the number of nodes explored is $(2/\varepsilon)^{O(\Delta)}$, and only a constant amount of time is spent per node. Hence, for each γ time spent is:

$$O(kd/\varepsilon + \varepsilon^{-5}) + (2/\varepsilon)^{O(\Delta)}.$$

As the size of Γ_{k-1} is $n(2/\varepsilon)^{O(\Delta\cdot(k-1))}$, the additional time required to compute Γ_k from Γ_{k-1} is

$$n(2/\varepsilon)^{O(\Delta\cdot(k-1))} \left(O(kd/\varepsilon + \varepsilon^{-5}) + (2/\varepsilon)^{O(\Delta)}\right)$$

which simplifies to

$$n(2/\varepsilon)^{O(\Delta\cdot k)} k \cdot d.$$

The total time required to compute $\Gamma_1, \ldots, \Gamma_k$ is

$$d\left(2^{O(\Delta)} n \log n + n(2/\varepsilon)^{O(\Delta)}\right) + \sum_{i=2}^{k} n(2/\varepsilon)^{O(\Delta\cdot i)} i \cdot d$$

which simplifies to

$$d\left(2^{O(\Delta)} n \log n + n(2/\varepsilon)^{O(\Delta\cdot k)}\right).$$

□

We conclude the section with a property of our computed WSTs which will be useful in Section 4.

Lemma 15. *For any ε-WST t=$(v_0, v_1, \ldots, v_k)$ generated by our algorithm,* $\mathrm{scale}_{v_i} \leq \varepsilon \mathrm{rad}(t)$ *for all v_i.*

Proof. By induction: For $k = 1$, we construct a $\frac{\varepsilon}{2}$-WSPD by the algorithm in [18, Sec.5]. Their construction ensures that $max(\mathrm{scale}_a, \mathrm{scale}_b) \leq \frac{\varepsilon}{2}\mathrm{dist}(a,b)$. Since $\mathrm{dist}(a,b) \leq 2 \cdot \mathrm{rad}(a,b)$, it follows that $max(\mathrm{scale}_a, \mathrm{scale}_b) \leq \varepsilon \mathrm{rad}(a,b)$.

For $k \geq 2$, let the statement be true for Γ_{k-1}. Consider $\gamma = (v_0, \ldots, v_{k-1}) \in \Gamma_{k-1}$ and let r denote the radius of the ε-approximation of $\mathrm{meb}(\gamma)$ as computed by our algorithm. A node x is added to γ to form $\gamma' = (v_0, \ldots, v_{k-1}, x)$ only if $2 \cdot \mathrm{scale}_x \leq \frac{\varepsilon}{4} r \leq \frac{\varepsilon(1+\varepsilon)}{4}\mathrm{rad}(\gamma)$ from (3.1). As $\mathrm{rad}(\gamma) \leq \mathrm{rad}(\gamma')$, the statement follows for x, and it follows from induction hypothesis for $v_0, \ldots v_{k-1}$. □

4 Čech approximations of linear size

In this section, we will define a persistence module which is a $(1+\varepsilon)$-approximation of the Čech module in the sense of Section 2. We start with a summary of our construction: we first define a sequence of (non-nested) simplicial complexes $(\mathscr{A}_\alpha)_{\alpha \geq 0}$, which we define using a WSSD from Section 3. Then, we construct simplicial maps $g_\alpha^{\alpha'} : \mathscr{A}_\alpha \to \mathscr{A}_{\alpha'}$ such that $g_{\alpha'}^{\alpha''} \circ g_\alpha^{\alpha'} = g_\alpha^{\alpha''}$ and $g_\alpha^\alpha = \mathrm{id}$. As discussed in Section 2, applying the homology functor to that sequence yields a persistent module. To show that the constructed module approximates the Čech module, we define simplicial *cross-maps* $\phi : \mathscr{C}_{\frac{\alpha}{1+\varepsilon}} \to \mathscr{A}_\alpha$ and $\psi : \mathscr{A}_{\frac{\alpha}{1+\varepsilon}} \to \mathscr{C}_\alpha$ that connect the two sequences on a simplicial level. We then show that the induced maps on homology groups all commute and finally apply Theorem 3 to show that the constructed module $(1+\varepsilon)$-approximates the Čech module. We remark that this strategy follows the approach by Dey et al. [11] who get a similar result for the Rips module, simplifying the previous work of Sheehy [25].

More notations. We construct a $\frac{\varepsilon}{154}$-WSSD for the net tree T. For a level ℓ, we define

$$T_\ell := \{u \in T \mid \ell(u) \leq \ell < \ell(\mathrm{par}(u))\}.$$

For a net tree node v of level i or less we denote by $\mathrm{vcell}(v,i)$ its ancestor in T_ℓ, that is, its highest ancestor having level at most i.

We fix the following additional parameters: Set $\theta_\ell := (1 + \frac{2\varepsilon}{5})^\ell$ for any integer ℓ. Let Δ_α denote the integer such that

$$\theta_{\Delta_\alpha} \leq \alpha < \theta_{\Delta_\alpha + 1}.$$

We define h_α as the integer such that

$$\frac{2\tau}{\tau - 1}\tau^{h_\alpha} \leq \frac{\varepsilon \theta_{\Delta_\alpha}}{7} \leq \frac{2\tau}{\tau - 1}\tau^{h_\alpha + 1}.$$

When there is no ambiguity about α, we will skip the suffixes and write $\Delta := \Delta_\alpha$ and $h := h_\alpha$.

To give a rough intuition about the chosen terms, the approximate complex will only change at discrete values; more precisely, all $\alpha \in [\theta_\ell, \theta_{\ell+1})$ will result in the same approximation. This motivates the definition of Δ_α which determines the range in which α falls in. The second parameter h_α

determines the grid size on which the approximation is constructed. Note that h_α rather depends on Δ_α than on α itself. Consequently, for any $\alpha \in [\theta_k, \theta_{k+1})$, the same h_α is chosen. Before we formally describe our construction, we prove the following useful lemma:

Lemma 16. *Let $\alpha > 0$, $\Delta := \Delta_\alpha$ and $h := h_\alpha$. If an $\frac{\varepsilon}{154}$-WST $t = (v_0, v_1, \ldots, v_k)$ satisfies* $\mathrm{rad}(t) \leq \theta_{\Delta+1}$*, then the level of each v_i is h or smaller.*

Proof. By Lemma 15, each v_i of t satisfies $\mathrm{scale}_{v_i} \leq \frac{\varepsilon}{154}\mathrm{rad}(t) \leq \frac{\varepsilon}{154}\theta_{\Delta+1}$. As $\theta_{\Delta+1} = (1+\frac{2\varepsilon}{5})\theta_\Delta \leq 2\theta_\Delta$ and $\tau = 11$, we have

$$\frac{2\tau^{\ell(v_i)+1}}{\tau - 1} = \mathrm{scale}_{v_i} \leq \frac{\varepsilon\theta_\Delta}{77} \leq \frac{2\tau}{11(\tau-1)}\tau^{h+1} = \frac{2\tau^{h+1}}{\tau-1}.$$

It follows that $\ell(v_i) \leq h$. □

The approximation complex We construct a simplicial complex $\mathscr{A}_\alpha$ over the vertex set from net-tree T at scale h in the following way: For any WST $t = (v_0, v_1, \ldots, v_k)$ with all v_i at level h or less, we add $t' = (\mathrm{vcell}(v_0, h), \mathrm{vcell}(v_1, h), \ldots, \mathrm{vcell}(v_k, h))$ to $\mathscr{A}_\alpha$ if $\mathrm{rad}(t') \leq \theta_\Delta$. Note that some of the $\mathrm{vcell}(v_l, h)$ can be the same, so that the resulting simplex might be of dimension less than k. It is clear by construction and Lemma 13 that $\mathscr{A}_\alpha$ consists of at most $n(2/\varepsilon)^{O(\Delta^2)}$ simplices, but it requires a proof to show that it is well-defined:

Lemma 17. *$\mathscr{A}_\alpha$ is a simplical complex.*

Proof. Consider $\gamma = (v_0, v_1, \ldots, v_k) \in \mathscr{A}_\alpha$. We need to show that all faces of $(v_0, v_1, \ldots, v_k)$ lie in $\mathscr{A}_\alpha$. Consider $t = (v_0, v_1, \ldots, v_l)$ with $l < k$ and consider the simplex $\tau = (\mathrm{rep}_{v_0}, \ldots, \mathrm{rep}_{v_l})$. There exists a WST $t' = (v'_0, \ldots, v'_l)$ which covers τ. The ball $(1+\varepsilon/154)\mathrm{meb}(\tau)$ is an enclosing ball for all points of the $(\varepsilon/154)$-WST t'. Therefore,

$$\mathrm{rad}(t') \leq (1+\frac{\varepsilon}{154})\mathrm{rad}(\tau) = (1+\frac{\varepsilon}{154})\mathrm{rad}(t) \leq (1+\frac{\varepsilon}{154})\mathrm{rad}(\gamma) \leq (1+\frac{\varepsilon}{154})\theta_\Delta \leq \theta_{\Delta+1}$$

Lemma 16 implies that the level of each v'_i is at most h. In particular, $\mathrm{vcell}(v'_i) = v_i$ as v'_i and v_i share the point m_i and v_i has level at most h by construction. Hence t belongs to $\mathscr{A}_\alpha$. □

We define maps between the $\mathscr{A}_\alpha$ next: Consider two scales $\alpha_1 < \alpha_2$. We set $h_1 := h_{\alpha_1}$ and define h_2, Δ_1, and Δ_2 accordingly. Set Since $h_1 \leq h_2$, there is a natural map $g_{\alpha_1}^{\alpha_2} : T_{h_1} \to T_{h_2}$, mapping a net-tree node of level h_1 to its highest ancestor of level at most h_2. This naturally extends to a map

$$g_{\alpha_1}^{\alpha_2} : \mathscr{A}_{\alpha_1} \to \mathscr{A}_{\alpha_2},$$

by mapping a simplex $\sigma = (v_0, \ldots, v_k)$ to $g_{\alpha_1}^{\alpha_2}(\sigma) := (g_{\alpha_1}^{\alpha_2}(v_0), \ldots, g_{\alpha_1}^{\alpha_2}(v_k))$. It is easy to verify that $g_{\alpha'}^{\alpha''} \circ g_{\alpha}^{\alpha'} = g_{\alpha}^{\alpha''}$ and $g_\alpha^\alpha = \mathrm{id}$.

Lemma 18. *$g := g_{\alpha_1}^{\alpha_2} : \mathscr{A}_{\alpha_1} \to \mathscr{A}_{\alpha_2}$ is a simplical map.*

Proof. Let $t = (v_0, v_1, \dots, v_k)$ be a k-simplex of $\mathscr{A}_{\alpha_1}$. Let $g(v_i) = v'_i$ denote the ancestor of v_i at level h_2. We need to show that $t' = (v'_0, v'_1, \dots, v'_k)$ belongs to $\mathscr{A}_{\alpha_2}$. If $\Delta_1 = \Delta_2$, the statement is trivial since $h_1 = h_2$, $\mathscr{A}_{\alpha_1} = \mathscr{A}_{\alpha_2}$ and g is the identity map. So, we assume that $\Delta_1 < \Delta_2$.

Consider the meb of t. It contains the representatives of all nodes v_i and hence at least one point of each of v'_i. If we inflate the ball by the largest diameter of a node at level h_2, all v'_i will be covered completely. We need to show that the inflated radius is less than θ_{Δ_2}.

The diameter of a node u at level h_2 is at most $2\text{scale}_u \leq \frac{2\varepsilon\theta_{\Delta_2}}{7}$. Also, as $\Delta_1 < \Delta_2$, we have $\theta_{\Delta_1} < \frac{\theta_{\Delta_2}}{1+\frac{2\varepsilon}{5}}$. Therefore,

$$\text{rad}(t') \leq \text{rad}(t) + \frac{2\varepsilon\theta_{\Delta_2}}{7} \leq \theta_{\Delta_1} + \frac{2\varepsilon\theta_{\Delta_1}}{7(1+\frac{2\varepsilon}{5})} < \frac{1+\frac{2\varepsilon}{7}+\frac{4\varepsilon^2}{35}}{1+\frac{2\varepsilon}{5}}\theta_{\Delta_2} < \theta_{\Delta_2},$$

for $\varepsilon \leq 1$. Hence $t' \in \mathscr{A}_{\alpha_2}$. □

Cross maps Next, we investigate the *cross-map* $\phi : \mathscr{C}_{\frac{\alpha}{1+\varepsilon}} \to \mathscr{A}_\alpha$. To define it for a vertex $v \in \mathscr{C}_{\frac{\alpha}{1+\varepsilon}}$ (which is a point of S), set $\phi(v) = q$, where q is the net-tree node in T_h that contains v. For a simplex $(v_0, \dots, v_k)$, define $\phi(v_0, \dots, v_k) = (\phi(v_0), \dots, \phi(v_k))$.

Lemma 19. *ϕ is a simplical map.*

Proof. Let $\sigma = (m_0, \dots, m_k)$ be a simplex in $\mathscr{C}_{\frac{\alpha}{1+\varepsilon}}$. Let $t = (v_0, v_1, \dots, v_k)$ be a WST which covers σ. By the WSSD property, we have that

$$\text{rad}(t) \leq (1+\frac{\varepsilon}{154})\text{rad}(\sigma) \leq \frac{1+\frac{\varepsilon}{154}}{1+\varepsilon}\alpha \leq \theta_{\Delta+1}.$$

Hence all v_i are at level at most h. Let $t' = (v'_0, \dots, v'_k)$ where $v'_i = \text{vcell}(v_i, h)$. We need to show that $t' \in \mathscr{A}_\alpha$. As in the previous lemma, if we inflate meb(t) by $\frac{2\varepsilon\theta_\Delta}{7}$, we cover t'. Using $\alpha \leq \theta_{\Delta+1} = (1+\frac{2\varepsilon}{5})\theta_\Delta$,

$$\text{rad}(t') \leq \frac{1+\frac{\varepsilon}{154}}{1+\varepsilon}\alpha + \frac{2\varepsilon\theta_\Delta}{7} \leq \frac{1+\frac{107}{154}\varepsilon+\frac{111}{385}\varepsilon^2}{1+\varepsilon}\theta_\Delta \leq \theta_\Delta$$

for $\varepsilon \leq 1$. It follows that $t' \in \mathscr{A}_\alpha$.

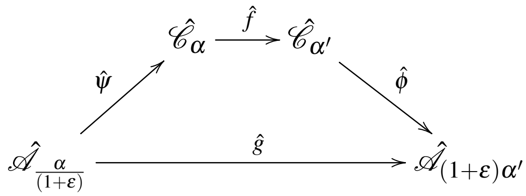
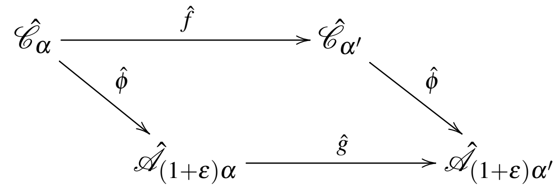
In the reverse direction, the map $\Psi : \mathscr{A}_{\frac{\alpha}{1+\varepsilon}} \to \mathscr{C}_\alpha$ maps a net tree node v of level h to its representative rep_v. For $t = (v_0, v_1, \dots, v_k) \in \mathscr{A}_{\frac{\alpha}{1+\varepsilon}}$, consider $\sigma = (\text{rep}_{v_0}, \dots, \text{rep}_{v_k})$. It follows that $\text{rad}(\sigma) \leq \text{rad}(t) \leq \frac{\alpha}{1+\varepsilon} \leq \alpha$. Hence $\sigma \in \mathscr{C}_\alpha$. □

Interleaving sequences We fix some integer $p \geq 0$ and consider the persistence modules

$$(\hat{\mathscr{C}}_\alpha)_{\alpha\geq 0} := (H_p(\mathscr{C}_\alpha))_{\alpha\geq 0}, \quad (\hat{\mathscr{A}}_\alpha)_{\alpha\geq 0} := (H_p(\mathscr{A}_\alpha))_{\alpha\geq 0},$$

where $H_p(\cdot)$ is the p-th homology group over an arbitrary base field, with the induced homomorphisms $\hat{f}_{\alpha_1}^{\alpha_2}$ (induced by inclusion) and $\hat{g}_{\alpha_1}^{\alpha_2}$, respectively. Moreover, since the cross-maps are simplicial, the induced homomorphisms $\hat{\phi} : \hat{\mathscr{C}}_{\frac{\alpha}{1+\varepsilon}} \to \hat{\mathscr{A}}_\alpha$ and $\hat{\psi} : \hat{\mathscr{A}}_{\frac{\alpha}{1+\varepsilon}} \to \hat{\mathscr{C}}_\alpha$ connect the two modules. We show that the cross-maps $\hat{\phi}$, $\hat{\psi}$ commute with the module maps $\hat{f}$, $\hat{g}$ in the next three lemmas.

Lemma 20. *The diagrams*

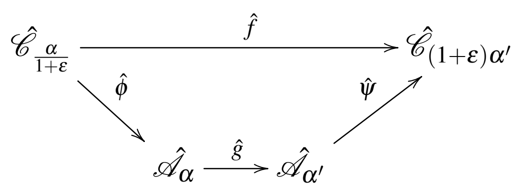



commute, that means, $\hat{\phi} \circ \hat{f} \circ \hat{\psi} = \hat{g}$ *for the left diagram and* $\hat{\phi} \circ \hat{f} = \hat{g} \circ \hat{\phi}$ *for the right diagram.*

Proof. The maps commute already on the simplicial level, that is, $\phi \circ f \circ \psi = g$ and $\phi \circ f = g \circ \phi$, respectively, as one can easily verify from the definition of the maps. $\square$

For the next two lemmas, we need the following definition: Two simplicial maps $h_1, h_2 : K \to L$ are *contiguous* if for any simplex $(v_0, \ldots, v_k) \in K$, the points $(h_1(v_0), \ldots, h_1(v_k), h_2(v_0), \ldots, h_2(v_k))$ form a simplex in L. In this case, the induced homomorphisms $\hat{h_1}, \hat{h_2}$ are equal [23, p.67].

Lemma 21. *For* $\alpha \leq \alpha'$*, the diagram*



commutes, that means, $\hat{\psi} \circ \hat{g} \circ \hat{\phi} = \hat{f}$.

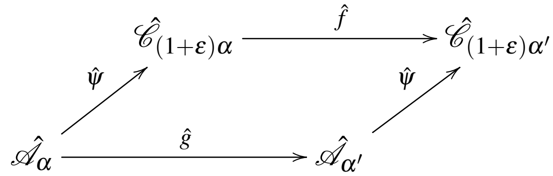
Proof. Note the simplicial maps do not commute here; we will show instead that they are contiguous. So, fix a simplex $\sigma = (v_0, \ldots, v_k)$ in $\mathscr{C}_{\frac{\alpha}{1+\varepsilon}}$. Let $q_i' := g(\phi(v_i))$; by definition, q_i' is a net-tree node that contains v_i, at the appropriate level for α'. By definition $w_i = \psi(q_i') = \psi(g(\phi(v_i))$ is the representative of q_i'. By the definition of $\mathscr{A}_{\alpha'}$, we have that $\mathrm{rad}(q_0', \ldots, q_k') \leq \theta_{\Delta'} \leq \alpha'$, where $\Delta' := \Delta_{\alpha'}$. It follows that $\mathrm{rad}(w_0, \ldots, w_k) \leq \alpha'$.

Since q_i' contains both v_i and w_i, if we inflate $\mathrm{meb}(w_0, \ldots, w_k)$ by the largest diameter of any node at the level h' corresponding to α', we will cover the simplex $\sigma = (v_0, \ldots, v_k, w_0, \ldots, w_k)$. The diameter of nodes at level h' is at most $\frac{2\varepsilon\theta_{\Delta'}}{7}$. The radius of a ball required to cover σ is at most

$$\mathrm{rad}(v_0, \ldots, v_k, w_0, \ldots, w_k) \leq \mathrm{rad}(w_0, \ldots, w_k) + \frac{2\varepsilon\theta_{\Delta'}}{7} \leq \alpha' + \frac{2\varepsilon}{7}\alpha' < (1+\varepsilon)\alpha'.$$

Therefore, the simplex $(v_0, \ldots, v_k, w_0, \ldots, w_k)$ is in $\mathscr{C}_{(1+\varepsilon)\alpha}$, so $\psi \circ g \circ \phi$ and f are contiguous. $\square$

Lemma 22. *For* $\alpha \leq \alpha'$*, the diagram*



commutes, that means, $\hat{\psi} \circ \hat{g} = \hat{f} \circ \hat{\psi}$.

Proof. Again, the corresponding simplicial maps do not commute in general (they do only if $h_\alpha = h_{\alpha'}$). We will show that the simplicial maps are contiguous. Fix some $t = (q_0, \dots, q_k) \in \mathscr{A}_\alpha$ and let v_ℓ be the representative of q_ℓ; in particular $f \circ \psi(q_\ell) = v_\ell$. Now, set $q'_\ell := g(q_\ell)$. It is clear that $q_\ell \subseteq q'_\ell$ and therefore, $v_\ell \in q'_\ell$. Set $w_\ell := \psi(g(q_\ell)) = \psi(q'_\ell)$ to be the representative of q'_ℓ. By the same argument as in Lemma 21, $\mathrm{rad}(v_0, \dots, v_k, w_0, \dots, w_k) \leq (1+\varepsilon)\alpha'$. This implies that the two maps are contiguous. □

Theorem 23. *The persistence module $\hat{\mathscr{A}}_*$ is a $(1+\varepsilon)$-approximation of the persistence module $\hat{\mathscr{C}}_*$.*

Proof. Lemmas 20-22 show that all diagrams in (2.1) commute for $c = 1+\varepsilon$. □

5 Coresets for minimal enclosing ball radii

Recall that for a point set $P = \{p_1, \dots, p_n\} \subset \mathbb{R}^d$, we denote by $\mathrm{meb}(P)$ the *minimum enclosing ball of P*. Let $\mathrm{center}(P) \in \mathbb{R}^d$ denote the center and $\mathrm{rad}(P) \geq 0$ the radius of $\mathrm{meb}(P)$. Fixing $\varepsilon > 0$, we call a subset $C \subseteq P$ a *meb-coreset for P* if the ball centered at $\mathrm{center}(C)$ and with radius $(1+\varepsilon)\mathrm{rad}(C)$ contains P. We call $C \subseteq P$ a *radius-coreset for P* if $\mathrm{rad}(P) \leq (1+\varepsilon)\mathrm{rad}(C)$. Informally, a radius-coreset approximates only the radius of the minimum enclosing ball, whereas the meb-coreset approximates the ball itself. A meb-coreset is also a radius-coreset by definition, but the opposite is not always the case; see Figure 5.1 for an example.

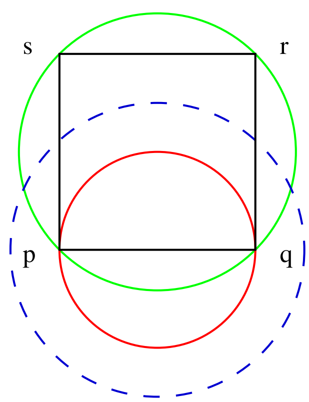


Figure 5.1: Consider the square formed by the points $p = (0,0), q = (2,0), r = (2,2)$ and $s = (0,2)$ on a plane. Let $P = \{p,q,r,s\}$ and $C = \{p,q\}$. Then, $\mathrm{center}(P) = (1,1)$, $\mathrm{rad}(P) = \sqrt{2}$, $\mathrm{center}(C) = (1,0)$ and $\mathrm{rad}(C) = 1$. For $\varepsilon = 0.5$, it is clear that C is a radius-coreset of P. However, C is not a meb-coreset as the ball centered at $\mathrm{center}(C) = (1,0)$ with radius 1.5 does not cover the points r and s.

Obviously, a point set is a coreset of itself, so coresets exist for any point set. We are interested in the coresets of small sizes. For the meb-coreset, this question is answered by Bădoiu and Clarkson [5]. We summarize their result in the following statement:

Theorem 24. *For $\varepsilon > 0$, and any (finite) point set, there exists a meb-coreset of size $\lceil \frac{1}{\varepsilon} \rceil$, and there exist point sets where any meb-coreset has size at least $\lceil \frac{1}{\varepsilon} \rceil$.*

Note that the size of the coreset is independent of both the number of points in P and the ambient dimension. However, since radius-coresets are a relaxed version of meb-coresets, we can hope for even smaller coresets. We start by showing a lower bound:

Lemma 25. *There is a point set such that any radius-coreset has size at least*

$$\delta := \lceil \frac{1}{2\varepsilon + \varepsilon^2} + 1 \rceil.$$

Proof. Consider the standard $(d-1)$-simplex in d dimensions, that is, P is the point set given by the d unit vectors in $\mathbb{R}^d$. By elementary calculations, it can be verified that $\text{center}(P) = (\frac{1}{d}, \dots, \frac{1}{d})$ and $\text{rad}(P) = \sqrt{\frac{d-1}{d}}$. Fixing a subset $C \subseteq P$ of size k, its points span a standard simplex in $\mathbb{R}^k$ and therefore, $\text{rad}(C) = \sqrt{\frac{k-1}{k}}$ by the same argument. Hence, C is a radius-coreset of P if and only if

$$\sqrt{\frac{d-1}{d}} \le (1+\varepsilon)\sqrt{\frac{k-1}{k}}.$$

Isolating k yields the equivalent condition that

$$k \ge \lceil \frac{(1+\varepsilon)^2}{(1+\varepsilon)^2 - \frac{d-1}{d}} \rceil = \lceil 1 + \frac{1}{\frac{d}{d-1}(2\varepsilon + \varepsilon^2 + \frac{1}{d})} \rceil.$$

The last expression is monotonously increasing in d, and converges to δ for $d \to \infty$. It follows that, for d large enough, any radius-coreset of a standard $(d-1)$-simplex has size at least δ. □

We will show next that any point set has a radius coreset of size δ. For a point set P in $\mathbb{R}^d$ and $1 \le k \le d$, let $r_k(P)$ denote the maximal radius of a meb among all subsets of P of cardinality k. We can assume that P contains at least $d+1$ points; otherwise it is contained in a lower-dimensional Euclidean space. On the other hand, if P contains at least $d+1$ points, there exists a subset P' of P containing exactly $d+1$ points such that the meb of P' equals the meb of P, which implies that $r_{d+1}(P) = \text{rad}(P)$. Moreover, $r_2(P) = \text{diam}(P)$ is the diameter of P. We use a result by Henk [19, Thm.1] (we adapt his notation to our context):

Theorem 26 (Generalized Jung's Theorem)**.** *Let $P \subset \mathbb{R}^d$ be a point set, and let i, j two integers with $2 \le j \le i \le d+1$. Then*

$$r_i(P) \le \sqrt{\frac{j(i-1)}{i(j-1)}} r_j(P)$$

The theorem generalizes an older result by Jung [20] which states the following relation between the circumradius and the diameter of P:

$$\text{rad}(P) = r_{d+1}(P) \le \sqrt{\frac{2d}{d+1}} r_2(P) = \sqrt{\frac{2d}{d+1}} \text{diam}(P). \tag{5.1}$$

We sketch the proof of Theorem 26 for completeness. It relies on the following property: Given a point set Q of $k+1$ linearly independent points in $\mathbb{R}^k$. Then,

$$\text{rad}(Q) \le \frac{k}{\sqrt{k^2-1}} r_k(Q), \tag{5.2}$$

in other words, there is a subset of k points whose circumradius is large in some sense; see also [5, Lemma 3.3]. We assume for simplicity that the i-subset of points of P that realizes $r_i(P)$ is linearly independent; otherwise, we can switch to an independent subset and a similar argument applies. Iteratively applying (5.2) yields that

$$r_i(P) \leq \prod_{t=j}^{i-1} \frac{t}{\sqrt{t^2-1}} r_j(P).$$

However, it is a straight-forward to prove by induction that

$$\prod_{t=j}^{i-1} \frac{t}{\sqrt{t^2-1}} = \sqrt{\frac{j(i-1)}{i(j-1)}}.$$

Theorem 27. *For $\varepsilon > 0$, any finite point set P has a radius-coreset of size δ.*

Proof. Applying Theorem 26 to the case that $i = d+1$ and $j = \delta$ yields

$$\mathrm{rad}(P) = r_{d+1}(P) \leq \sqrt{\frac{\delta \cdot d}{(d+1)(\delta-1)}} r_\delta(P) = \underbrace{\sqrt{\frac{d}{d+1}}}_{\leq 1} \sqrt{\frac{\delta}{\delta-1}} r_\delta(P).$$

Furthermore, since $\delta \geq \frac{1}{2\varepsilon+\varepsilon^2} + 1$, it follows that

$$\frac{\delta}{\delta-1} = 1 + \frac{1}{\delta-1} \leq (1+\varepsilon)^2.$$

So, letting C be a subset of cardinality δ with radius $r_\delta(P)$, we obtain that $\mathrm{rad}(P) \leq (1+\varepsilon)\mathrm{rad}(C)$, which means that C is a radius-coreset. □

We remark that our results immediately imply an algorithm for computing a radius-coreset of size δ: starting with the whole point set, iteratively remove points such that the remaining subset has the largest possible radius among all choices of removed points. When this process is stopped for a subset of size δ, the resulting subset is a radius-coreset. However, this algorithm is rather inefficient, because it is quadratic in n, and a natural question is how to compute radius coresets more efficiently. For meb-coresets of size $\lceil \frac{1}{\varepsilon} \rceil$, Bădoiu and Clarkson [5] prove existence algorithmically by defining an algorithm which starts with an arbitrary set of size $\lceil \frac{1}{\varepsilon} \rceil$ and alternatingly adds and removes points from the set until the set remains unchanged, and they prove that the resulting set is a meb-coreset. Their algorithm is an instance of a more general class of optimization problems as described in [8]; we were not able to find a reformulation of the radius-coreset problem in terms of this algorithmic framework.

6 A generalized Rips-Lemma

We define the following generalization of a flag-complex:

Definition 28 (*i*-completion)**.** *Let K denote a simplicial complex. The i-*completion *of K, $\mathscr{M}_i(K)$, is maximal complex whose i-skeleton equals the i-skeleton of K.*

With that notation, we have that $\mathscr{R}_\alpha = \mathscr{M}_1(\mathscr{C}_\alpha)$. Moreover, we have that $\mathscr{C}_\alpha = \mathscr{M}_d(\mathscr{C}_\alpha)$ as a consequence of Helly's Theorem [13, p.57].

We can show the following result as an application of Theorem 27.

Theorem 29. *For $\delta = \lceil 1/(2\varepsilon + \varepsilon^2) + 1 \rceil$,*

$$\mathscr{C}_\alpha \subseteq \mathscr{M}_{\delta-1}(\mathscr{C}_\alpha) \subseteq \mathscr{C}_{(1+\varepsilon)\alpha}$$

Proof. The first inclusion is clear. Now, consider a simplex σ in $\mathscr{M}_{\delta-1}(\mathscr{C}_\alpha)$. The second inclusion is trivial if $\dim \sigma \leq \delta - 1$, so let its dimension be at least δ. By Theorem 27, the boundary vertices of σ have a coreset of size at most δ. Let τ denote the simplex spanned by such a coreset. As τ is a face of σ, it is contained in $\mathscr{M}_{\delta-1}(\mathscr{C}_\alpha)$, and because it is of dimension at most $\delta - 1$, it is in particular contained in $C(\alpha)$. By the property of coresets, the minimal enclosing ball of σ has radius at most $(1+\varepsilon)\alpha$ which implies that $\sigma \in \mathscr{C}_{(1+\varepsilon)\alpha}$. □

As a special case, consider the choice $\varepsilon = \sqrt{2} - 1$, so that $\delta = 2$. The above result yields that

$$\mathscr{R}_\alpha = \mathscr{M}_1(\mathscr{C}_\alpha) \subseteq \mathscr{C}_{\sqrt{2}\alpha},$$

which is exactly the statement of the Vietoris-Rips Lemma as stated in [13, p.62].

Theorem 29 and Lemma 4 prove the closeness of the persistence diagrams of the Čech filtration and the completion complex:

Theorem 30. *The persistence diagram of $\mathscr{M}_{\delta-1}(\mathscr{C}_*)$ with $\delta := \lceil 1/(2\varepsilon + \varepsilon^2) + 1 \rceil$ is a $(1+\varepsilon)$-approximation of the persistence diagram of $\mathscr{C}_*$.*

Note that $\mathscr{M}_k(\mathscr{C}_\alpha)$ is determined by the k-skeleton of the Čech complex, which of size $O(n^{k+1})$. In this respect, the completion complex constitutes a trade-off between simplicity (i.e., its representation size) and approximation quality of the Čech complex. We emphasize that the approximation is solely determined by k and does not depend on the ambient dimension of the point set.

7 Conclusion and Outlook

We have presented two distinct ways to approximate Čech complexes; the fixed-dimensional result on approximating the Čech filtration to linear size is a technically challenging, but conceptually straight-forward extension of recent work on the Rips filtration; however, we believe that the concept of WSSDs to be interesting and hopefully applicable in different contexts, and we plan to identify application scenarios in the future. Our high-dimensional results are a first attempt to link the areas of computational topology, where data is often high-dimensional, and geometric approximation algorithms that try to overcome the curse of dimensionality. We want to achieve algorithmic results in that context in the future; one question is whether an optimal-size radius coreset can be computed efficiently. Moreover, the introduced concept of completions is not tied to start completing simplices at a fixed dimension; in fact, one can start with any complex C (not necessarily a skeleton) and define the completion as the largest complex containing C. With such *adaptive completions*, ε-close approximations of the Čech filtration might be possible with just a slightly larger representation size than the Rips filtration. The open question is, however, whether such a representation can be computed efficiently. Finally, we pose the question whether there are other applications, besides approximating Čech complexes, where the smaller size of radius-coresets in comparison to meb-coresets could be useful.

References

[1] P. Assouad. Plongements lipschitziens dans $\mathbb{R}^n$. *Bulletin de la Socit Mathmatique de France*, 111:429–448, 1983.

[2] D. Attali, A. Lieutier, and D. Salinas. Efficient data structures for representing and simplifying simplicial complexes in high dimensions. *Int. J. of Computational Geometry & Applications*, 22:279–303, 2012.

[3] J.-D. Boissonnat, L. Guibas, and S. Oudot. Manifold reconstruction in arbitrary dimensions using witness complexes. *Discr. Comput. Geom.*, 42:37–70, 2009.

[4] M. Bădoiu and K. Clarkson. Smaller core-sets for balls. In *Proc. 14th ACM-SIAM Symp. on Discr. Alg.*, pages 801–802, 2003.

[5] M. Bădoiu and K. Clarkson. Optimal core-sets for balls. *Computational Geometry: Theory and Applications*, 40:14–22, 2008.

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

[7] F. Chazal, D. Cohen-Steiner, M. Glisse, L. Guibas, and S. Oudot. Proximity of persistence modules and their diagrams. In *Proc. 25th ACM Symp. on Comp. Geom.*, pages 237–246, 2009.

[8] K. Clarkson. Coresets, sparse greedy approximation, and the Frank-Wolfe algorithm. In *Proc. 19th ACM-SIAM Symp. on Discr. Alg.*, pages 922–931, 2008.

[9] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. *Discrete and Computational Geometry*, 37:103–120, 2007.

[10] V. de Silva and G. Carlsson. Topological estimation using witness complexes. In *Symp. on Point-Based Graphics*, pages 157–166, 2004.

[11] T. Dey, F. Fan, and Y. Wang. Computing topological persistence for simplicial maps. *CoRR*, abs/1208.5018, 2012.

[12] T. Dey, F. Fan, and Y. Wang. Graph induced complex on point data. In *Proc. 29th ACM Symp. on Comp. Geom.*, 2013.

[13] H. Edelsbrunner and J. Harer. *Computational Topology, An Introduction*. American Mathematical Society, 2010.

[14] H. Edelsbrunner, D. Kirkpatrick, and R. Seidel. On the shape of a set of points in the plane. *IEEE Trans. Inform. Theory*, IT-29:551–559, 1983.

[15] H. Edelsbrunner and E. Mücke. Three-dimensional alpha shapes. *ACM Trans. Graphics*, 13:43–72, 1994.

[16] T. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theor. Comput. Sci.*, pages 293–306, 1985.

[17] L. Gottlieb and R. Krauthgamer. Proximity algorithms for nearly-doubling spaces. In *In APPROX-RANDOM*, 2010.

[18] S. Har-Peled and M. Mendel. Fast construction of nets in low dimensional metrics and their applications. *Siam J. on Computing*, 35:1148–1184, 2006.

[19] M. Henk. A generalization of Jung's theorem. *Geometriae Dedicata*, 42:235–240, 1992.

[20] H. Jung. Über die kleinste Kugel, die eine räumliche Figur einschliesst. *J. reine angewandte Mathematik*, 123:241–257, 1901.

[21] M. Kerber and R. Sharathkumar. Approximate Čech complex in low and high dimensions. In *Algorithms and Computation*, volume 8283 of *LNCS*, pages 666–676. Springer Berlin Heidelberg, 2013.

[22] M. Mrozek, P. Pilarczyk, and N. Zelazna. Homology algorithm based on acyclic subspace. *Computer and Mathematics with Applications*, 55:2395–2412, 2008.

[23] J. Munkres. *Elements of algebraic topology*. Westview Press, 1984.

[24] S. Oudot and D. Sheehy. Zigzag zoology: Rips zigzags for homology inference. In *Proc. 29th ACM Symp. on Comp. Geom.*, 2013.

[25] D. Sheehy. Linear-size approximation to the Vietoris-Rips filtration. In *Proc. 2012 ACM Symp. on Comp. Geom.*, pages 239–248, 2012.

[26] K. Talwar. Bypassing the embedding: Algorithms for low dimensional metrics. In *Proc. 36th ACM Symp. on Theory of Comput.*, pages 281–290, 2004.

[27] A. Zomorodian. The tidy set: A minimal simplicial set for computing homology of clique complexes. In *Proc. 26th ACM Symp. on Comp. Geom.*, pages 257–266, 2010.