A GEOMETRIC PERSPECTIVE ON SPARSE FILTRATIONS

NICHOLAS J. CAVANNA, MAHMOODREZA JAHANSEIR, AND DONALD R. SHEEHY

1. Introduction

Nerve complexes are discrete structures used for computing topological information about a continuous space, but they often suffer from exponential blowup in size. Sparsifying these complexes allows one to retain most topological information while maintaining a significantly smaller complex. We present a geometric perspective on sparse filtrations, viewing them as a nerve of geometric cones, rather than a union of nerves. This new perspective leads to simpler algorithms, simpler proofs with more general results. We also provide the runtime for our construction and prove that vertex removal can be implemented as a sequence of edge collapses. A video illustrating this approach is available [2].

2. Background

2.1. **Distances and Metrics.** The input is a finite set $P \subset \mathbb{R}^d$ with a convex metric $\mathbf{d}$, e.g. the Euclidean or L_p metric (for $p \geq 1$). Let $\mathrm{ball}(c, r) = \{x \in \mathbb{R}^d \mid \mathbf{d}(x, c) \leq r\}$ denote the closed metric ball centered at c with radius r. For all $\alpha \geq 0$, the α-*offsets* of P are defined as $P^\alpha = \bigcup_{p \in P} \mathrm{ball}(p, \alpha)$. The family of offsets parameterized by α is called the *offset filtration*, denoted $\{P^\alpha\}_{\alpha \geq 0}$. It is a *filtration* because $P^\alpha \subseteq P^\beta$ if $\alpha \leq \beta$.

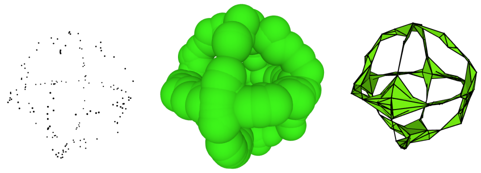

Figure 1. Left to right: point set, offsets and nerve complex

2.2. **Persistent Homology.** The homology of a space tells one the number of holes in each dimension in the space. Homology is useful for characterizing spaces topologically and is invariant under homeomorphisms. Given an offset filtration $\{P^\alpha\}_{\alpha \geq 0}$, each set in the filtration has its own homology. The differences between the homology of P^α and P^β for $\alpha < \beta$ indicate at what scales features are "born" and "die". A feature's lifespan can be represented by an interval $[b, d]$ where b is the *birth time* and d is the *death time*. Collectively, the lifespan intervals of all features is the *persistence barcode* and the information provided is the *persistent homology* of the filtration. We say the persistent homology of two spaces are c-*approximate* if we can pair each's intervals such that the birth and death times respectively are within a multiplicative factor c and those for which $d/b < c$ can be ignored.

2.3. **Simplicial Complexes and Nerves.** A *simplicial complex* is a collection of subsets of a vertex set that is closed under taking subsets. The *nerve* of a collection of closed, convex sets $\mathcal{U} = \{U_1, \ldots, U_n\}$ is a simplicial complex defined as $\mathrm{Nrv}(\mathcal{U}) := \{I \subseteq [n] \mid \bigcap_{j \in I} U_j \neq \emptyset\}$. The nerve has as a vertex for each set, an edge for each two-way intersection, a triangle for each three way intersection etc. In the case of an offset filtration, P^α is covered by $\{\mathrm{ball}(p, \alpha) \mid p \in P\}$. Each ball grows throughout the filtration, and by taking the nerve of the cover of each P^α, we get a filtration of nerves. Given a collection of filtrations of closed, convex sets $\mathcal{U} = \{\{U_1^\alpha\}, \ldots, \{U_n^\alpha\}\}$, , where $\mathcal{U}^\alpha := \{U_1^\alpha, \ldots, U_n^\alpha\}$, the Persistent Nerve Lemma [3] implies the filtrations $\{\bigcup U_i^\alpha\}_{\alpha \geq 0}$ and $\{\mathrm{Nrv}(\mathcal{U}^\alpha)\}_{\alpha \geq 0}$ have identical persistent homology i.e. their barcodes are the same.

2.4. **Greedy Permutations.** Given a set P of n points in a metric space with metric $\mathbf{d}$, we say that $P = \{p_1, \ldots, p_n\}$ is a *greedy permutation* of P if for all $i \in \{1, \ldots, n\}$, $\mathbf{d}(p_i, P_{i-1}) = \max_{p \in P} \mathbf{d}(p, P_{i-1})$, where $P_i = \{p_1, \ldots, p_i\}$. The insertion radius of a point p_i is the value $\lambda_i := \mathbf{d}(p_i, P_{i-1})$. By convention, $\lambda_1 = \infty$. Greedy permutations have the nice property that P_i is

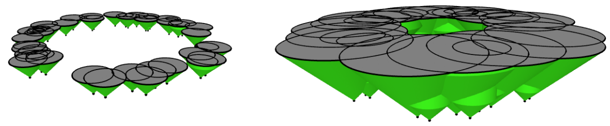

FIGURE 2. The cones of the offsets

a λ_i-net—for all $p, q \in P_i$, $\mathbf{d}(p, q) \geq \lambda_i$ [packing] and $P \subseteq P_i^{\lambda_i}$ [covering].

3. Results

3.1. **Sparsification.** We consider a point set $P = \{p_1, \ldots, p_n\}$ ordered by a greedy permutation. We fix a sparsification constant $\varepsilon \in (0, 1)$. The radius of the ball at p_i is limited to $\lambda_i(1 + \varepsilon)/\varepsilon$ and so the radius at scale α is defined as $r_i(\alpha) = \min\{\alpha, \lambda_i(1+\varepsilon)/\varepsilon\}$. The *perturbed α-offsets* are defined as $\tilde{P}^\alpha := \bigcup_{i\in[n]} \operatorname{ball}(p_i, r_i(\alpha))$. The sparsification process is induced by the *α-balls*, which are defined as $b_i(\alpha) = \operatorname{ball}(p_i, r_i(\alpha))$ if $\alpha \leq \lambda_i(1+\varepsilon)^2/\varepsilon$, and empty otherwise. There are far fewer intersections between α-balls and so the resulting nerve will be much smaller.

Proposition 1. *The persistence barcode of the perturbed offsets $\{\tilde{P}^\alpha\}_{\alpha\geq 0}$ $(1 + \varepsilon)$-approximates the persistence barcode of the offsets $\{P^\alpha\}_{\alpha\geq 0}$.*

3.2. **Sparse filtration.** In previous work, sparse filtrations were defined as a union of nerves at different scales. We provide a simpler geometric view of a sparse filtration as a nerve of cones. We add another spatial dimension to the α-balls, viewing α as the height of a cone. Formally, the perturbed cone for p_i is the set $U_i^\alpha := \bigcup_{\delta\leq\alpha}(b_i(\delta)\times\{\delta\})$ (see Fig. 2). This leads to an equivalent definition of the sparse nerve filtration, $\{S^\alpha\} = \{\operatorname{Nrv}\{U_i^\alpha\}\}_{\alpha\geq 0}$.

Theorem 2. *The persistence barcode of the sparse nerve filtration $\{S^\alpha\}_{\alpha\leq 0}$ is a $(1+\varepsilon)$-approximation to the persistence barcode of the offsets $\{P^\alpha\}_{\alpha\geq 0}$.*

3.3. **Algorithmic construction.** In our full paper [1], we create a data structure that allows insertion of points given a greedy permutation. We define the predecessor of p_i as the point $\operatorname{pred}(p_i)$ such that $\lambda_i = \mathbf{d}(p_i, \operatorname{pred}(p_i))$. With this data structure one can compute the edges in linear time. From the edges, one can find the k-simplices in the standard way: For each point p, check if the cones of each k-tuple of adjacent points along with the cone of p intersect at some $\alpha < \infty$, and if so the simplex defined by this $(k+1)$-tuple is in the filtration. With the simplex birth times and the maximal complex, we know the sparse nerve filtration. Theorem 3 summarizes the runtime of our algorithm.

Theorem 3. *Given a finite metric $(P, \boldsymbol{d})$ and a greedy permutation of P with $\operatorname{pred}(p_i)$ for each $p_i \in P$, one can find the k-simplices of $\{S^\alpha\}$ in $\kappa^{O(k\rho)}n$ time, where ρ is the doubling dimension of $\boldsymbol{d}$, $\kappa = (\varepsilon^2 + 3\varepsilon + 2)/\varepsilon$, and $\varepsilon \in (0, 1)$.*

3.4. **Removing vertices.** Since the sparse filtration is a filtration, we do not remove the vertices when we delete the α-balls at some scale. In practice, one may wish to remove the vertices and we prove this is possible using an operation called an edge contraction. Dey et al.[4] prove that if an edge satisfies the so-called *Link Condition*, then the topology of a simplicial complex is unchanged by contracting that edge. Theorem 4 uses this result and implies vertices can be removed by edge contractions as the scale α increases.

Theorem 4. *If $(P, \boldsymbol{d})$ is a finite subset of a convex metric space and $\{S^\alpha\}$ is its sparse filtration, then the last vertex p_n inserted has a neighbor p_i such that the edge $[p_np_i] \in S^\alpha$ satisfies the link condition, where $\alpha \geq \lambda_i(1 + \varepsilon)^2/\varepsilon$ and λ_n is the insertion radius of p_n.*

References

[1] Nicholas J. Cavanna, Mahmoodreza Jahanseir, and Donald R. Sheehy. A geometric perspective on sparse filtrations. *arXiv*, abs/1506.03797, 2015.

[2] Nicholas J. Cavanna, Mahmoodreza Jahanseir, and Donald R. Sheehy. Visualizing sparse filtrations. In *Proceedings of the 31st Symposium on Computational Geometry (Multimedia Session)*, 2015.

[3] Frédéric Chazal and Steve Y. Oudot. Towards persistence-based reconstruction in Euclidean spaces. In *Proceedings of the 24th ACM Symposium on Computational Geometry*, pages 232–241, 2008.

[4] Tamal K. Dey, Herbert Edelsbrunner, Sumanta Guha, and Dmitry V. Nekhayev. Topology preserving edge contraction. *Publications de l'Institut Mathematique (Beograd)*, 60:23–45, 1999.