

Families of point sets with identical 1D persistence

Philip Smith¹ and Vitaliy Kurlin (corresponding author)¹

¹Computer Science, University of Liverpool, Liverpool, L69 3BX, UK,
philip.smith3@liverpool.ac.uk, vitaliy.kurlin@liverpool.ac.uk

April 1, 2022

Abstract

Persistent homology is a popular and useful tool for analysing point sets, revealing features of a point set that can be used to highlight key information, distinguish point sets and as an input into machine learning pipelines. The famous stability theorem of persistent homology provides an upper bound for the change in persistence under perturbations, but it does not provide a lower bound. This paper clarifies the possible limitations persistent homology may have in distinguishing point sets, which is clearly evident for point sets that have trivial persistence. We describe large families of point sets that have identical or even trivial one-dimensional persistence. The results motivate stronger invariants to distinguish point sets up to isometry.

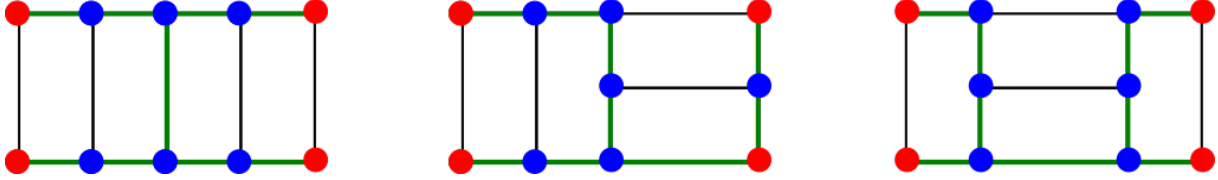
1 Introduction: persistence is an isometry invariant

Topological Data Analysis (TDA) was pioneered by Claudia Landi [15] (initially called size theory) and Herbert Edelsbrunner et al. [12]. The important papers by Gunnar Carlsson [4], Robert Ghrist [16] and Shmuel Weinberger [20] were followed by substantial developments by Fred Chazal [6] and many others. Persistent homology is a key tool of TDA [10] and is invariant up to isometry (transformations which maintain inter-point distances).

The famous stability theorem [7] states that under bounded noise, the bottleneck distance between persistence diagrams of a point set and its perturbation has an upper bound dependent on the magnitude of the perturbation. As such, a small perturbation of a point set results in at most a small change in its corresponding persistent homology.

However, there is no lower bound, which means that a perturbation of a point set can result in the corresponding persistent homology remaining unchanged. This is an issue for applications that require invariants to reliably distinguish point sets up to isometry or similar equivalence relations such as rigid motion or uniform scaling. A uniform scaling also scales persistence, but any non-uniform scaling or a more general continuous deformation of data changes persistence rather arbitrarily. Hence any persistence-based approach analyses data only up to isometry, which is an important equivalence due to the rigidity of many real-life structures, see Fig. 1.

Figure 1. Many non-isometric sets whose points form 1×2 dominoes have the same 0D persistence (determined by edge-lengths of a Minimum Spanning Tree in green) and trivial 1D persistence for the filtrations of Vietoris-Rips, Čech and Delaunay complexes. Theorem 4.3 extends these examples to a large open subspace of point sets by adding ‘tails’ at red corners.

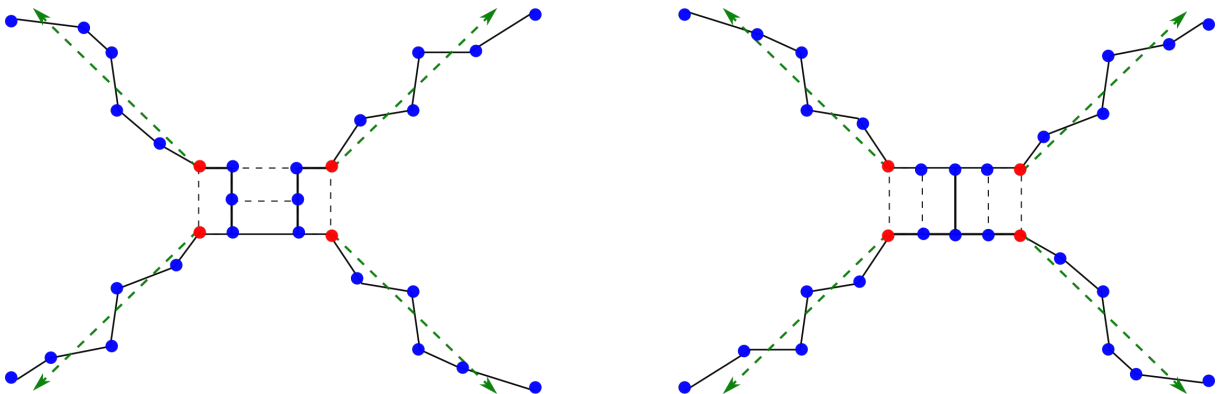


The above facts motivate a comparison of persistent homology with other isometry invariants of point sets. One key application is for rigid periodic crystals whose isometry classification was recently initiated by Herbert Edelsbrunner et al. [11]. A periodic crystal is naturally modeled by a periodic set of points representing atomic centres.

This paper describes how a point sequence of arbitrary size can be added to any finite point set whilst leaving the one-dimensional persistent homology unchanged, and goes on to define a large continuous family of point sets that have trivial persistence. We focus on one-dimensional persistent homology of point sets in any dimension, computed using filtrations of simplicial complexes including Vietoris-Rips, Čech and Delaunay complexes.

Our main Theorem 4.3 and Corollary 4.4 build large open spaces of point sets that all have identical 1D persistence. With similar aims, Curry et al. described spaces of Morse functions on the interval [8] and the sphere S^2 [5] that have identical persistence. The experiments use the Vietoris-Rips filtration whose persistence is implemented by Ripser [1]. Higher dimensional persistent homology will be included in future updates.

Figure 2. The initial set A of 10 points in the centre is extended by four tails going out from red points. All such sets have trivial 1D persistence by Corollary 4.4, but all such sets in general position are not isometric to each other. The black edges form a Minimum Spanning Tree of the set.



Section 4 introduces definitions and proves auxiliary lemmas needed for our main Theorem 4.3, which describes how, given a finite point set, we can add an arbitrarily large point set without affecting the one-dimensional persistent homology. Section 5 summarises large-scale experiments that reveal interesting information on the prevalence, or more likely lack, of significant persistent features occurring in randomly generated point sets.

2 Three classes of edges important for 1D persistence

This section introduces three classes of edges (short, medium and long) that will help build point sets with identical 1D persistence. Since persistent homology can be defined for any filtration of simplicial complexes on an abstract finite set A , the most general settings are recalled in Definition 2.1. Definition 2.2 describes the more explicit filtrations of Vietoris-Rips, Čech and Delaunay complexes on a finite set A in any metric space M (or just in \mathbb{R}^N for Delaunay complexes).

Definition 2.1 (Filtration of complexes $\{C(A; \alpha)\}$). *Let A be any abstract finite set.*

(a) *A (simplicial) complex C on A is a finite collection of subsets $\sigma \subset A$ (called simplices) such that all subsets of σ and any intersection of simplices are also simplices of C .*

(b) *The dimension of a simplex σ consisting of $k + 1$ points is k . We assume that all points of A are 0-dimensional simplices, sometimes called vertices of C . A 1-dimensional simplex (or edge) e between points $p, q \in A$ is the unordered pair denoted as $[p, q]$.*

(c) *An (ascending) filtration $\{C(A; \alpha)\}$ is any family of simplicial complexes on the vertex set A , parameterised by a scale $\alpha \geq 0$ such that $C(A; \alpha') \subseteq C(A; \alpha)$ for all $\alpha' \leq \alpha$. ■*

Let M be any metric space with a distance d satisfying all metric axioms. For any points $p, q \in A \subset M$, the edge $e = [p, q]$ has length $|e| = d(p, q)$. An example of a metric space is \mathbb{R}^N with the Euclidean metric. If $A \subset \mathbb{R}^N$, the edge $e = [p, q]$ can be geometrically interpreted as the straight-line segment connecting the points $p, q \in A \subset \mathbb{R}^N$.

Definition 2.2 introduces the simplicial complexes $\text{VR}(A; \alpha)$ and $\check{\text{Cech}}(A; \alpha)$ on any finite set A inside an ambient metric space M , although $A = M$ is possible. For a point $p \in A$ and $\alpha \geq 0$, let $\bar{B}(p; \alpha) \subset M$ denote the closed ball with centre p and radius α .

A Delaunay complex $\text{Del}(A; \alpha) \subset \mathbb{R}^N$ will be defined for a finite set $A \subset \mathbb{R}^N$ because of extra complications arising if a point set A lives in a more general metric space [3].

Definition 2.2 (Complexes $\text{VR}(A; \alpha)$, $\check{\text{Cech}}(A; \alpha)$, $\text{Del}(A; \alpha)$). *Let $A \subset M$ be any finite set of points. Fix a scale $\alpha \geq 0$. Each complex $C(A; \alpha)$ below has vertex set A .*

(a) *The Vietoris-Rips complex $\text{VR}(A; \alpha)$ has all simplices on points $p_1, \dots, p_k \in A$ whose pairwise distances are all at most 2α , so $d(p_i, p_j) \leq 2\alpha$ for all distinct $i, j \in \{1, \dots, k\}$.*

(b) *The Čech complex $\check{\text{Cech}}(A; \alpha)$ has all simplices on points $p_1, \dots, p_k \in A$ such that the full intersection $\cap_{i=1}^k \bar{B}(p_i; \alpha)$ is not empty.*

(c) *For any finite set of points $A \subset \mathbb{R}^N$, the convex hull of A is the intersection of all closed half-spaces of \mathbb{R}^N containing A . Each point $p_i \in A$ has the Voronoi domain*

$$V(p_i) = \{q \in \mathbb{R}^n \mid |q - p_i| \leq |q - p_j| \text{ for any point } p_j \in A - \{p_i\}\}.$$

The Delaunay complex $\text{Del}(A; \alpha)$ has all simplices on points $p_1, \dots, p_k \in A$ such that the intersection $\cap_{i=1}^k (V(p_i) \cap \bar{B}(p_i; \alpha))$ is not empty [9]. Alternatively, a simplex σ on points $p_1, \dots, p_k \in A$ is a Delaunay simplex if: (a) the smallest $(k - 2)$ -dimensional sphere S^{k-2} passing through p_1, \dots, p_k has a radius at most α ; (b) there is also an $(N - 1)$ -dimensional sphere S^{N-1} passing through p_1, \dots, p_k that does not enclose any points of A .

In a degenerate case, the smallest $(k - 2)$ -dimensional sphere S^{k-2} above can contain more than k points of A . If σ is enlarged to the convex hull H of all points $A \cap S^{k-2}$,

then $\text{Del}(A; \alpha)$ becomes a polyhedral Delaunay mosaic [2]. For simplicity, we choose any triangulation of H into Delaunay simplices. Then a Delaunay complex $\text{Del}(A; \alpha) \subset \mathbb{R}^N$ is a subset of a Delaunay triangulation of the convex hull of A , which is unique in general position.

The complexes of the three types above will be called geometric complexes for brevity. ■

Both complexes $\text{VR}(A; \alpha)$ and $\check{\text{Cech}}(A; \alpha)$ are abstract and so are not embedded in \mathbb{R}^N , even if $A \subset \mathbb{R}^N$. Though $\text{Del}(A; \alpha)$ is embedded into \mathbb{R}^N , its construction is fast enough only in dimensions $N = 2, 3$. For high dimensions $N > 3$ or any metric space M , the simplest complex to build and store is $\text{VR}(A; \alpha)$. Indeed, $\text{VR}(A; \alpha)$ is a flag complex determined by its 1-dimensional skeleton $\text{VR}^1(A; \alpha)$ so that any simplex of $\text{VR}(A; \alpha)$ is built on a complete subgraph whose vertices are pairwise connected by edges in $\text{VR}^1(A; \alpha)$.

The key idea of TDA is to view any finite set $A \subset \mathbb{R}^N$ through lenses of a variable scale $\alpha \geq 0$. When α is increasing from the initial value 0, the points of A become blurred to balls of radius α and may start forming topological shapes that ‘persist’ over long intervals of α . More formally, for any fixed $\alpha \geq 0$, the union $\cup_{p \in A} \bar{B}(p; \alpha)$ of closed balls is homotopy equivalent to the Čech complex $\check{\text{Cech}}(A; \alpha)$ and also to the Delaunay complex $\text{Del}(A; \alpha) \subset \mathbb{R}^N$ by the Nerve Lemma [17, Corollary 4G.3].

For any geometric complex $C(A; \alpha)$ from Definition 2.2, all connected components of $C(A; \alpha)$ are in a 1-1 correspondence with all connected components of the union $\cup_{p \in A} \bar{B}(p; \alpha)$. Any edge e enters $C(A; \alpha)$ when α equals the edge’s half-length $|e|/2$.

Definition 2.3 (Short, medium, long edges in a filtration). *Let $\{C(A; \alpha)\}$ be any filtration of complexes on an abstract finite vertex set A , see Definition 2.1. Let an edge $e = [p, q]$ between points $p, q \in A$ enter the simplicial complex $C(A; \alpha)$ at scale α .*

(a) *Consider the 1-dimensional graph $C'(A; \alpha)$ with vertex set A and all edges from $C(A; \alpha)$ except the edge e . If the endpoints of e are in different connected components of $C'(A; \alpha)$, then the edge e is called short in the filtration $\{C(A; \alpha)\}$.*

(b) *The edge e is called long in $\{C(A; \alpha)\}$ if A has a vertex v such that the 2-simplex Δpqv is contained in $C(A; \alpha)$ and both edges $[p, v], [v, q]$ are in $C(A; \alpha')$ for some $\alpha' < \alpha$.*

(c) *If e is neither short nor long, then the edge e is called medium in $\{C(A; \alpha)\}$. ■*

Definition 2.3(b) implies that any long edge enters $C(A; \alpha)$ with a 2-simplex Δpqv at the same scale α and the boundary of this 2-simplex is homologically trivial in $C(A; \alpha)$ due to the other two edges $[p, v], [v, q]$ that entered the filtration at a smaller scale $\alpha' < \alpha$.

Classes of edges in Definition 2.3 were introduced only in terms of an abstract filtration of complexes. Lemma 2.4 interprets long edges in VR and Čech filtrations via distances.

Lemma 2.4 (Long edges in VR and Čech). *Let A be a finite set in a metric space.*

(a) *In the Vietoris-Rips filtration $\{\text{VR}(A; \alpha)\}$, an edge $e = [p, q]$ is long if and only if the set A has a point v such that $e = [p, q]$ is a strictly longest edge in the triangle Δpqv .*

(b) *In the Čech filtration $\{\check{\text{Cech}}(A; \alpha)\}$, an edge e is long if and only if $\check{\text{Cech}}(A; \alpha)$ includes a triangle Δpqv such that the edge $e = [p, q]$ is strictly longest in Δpqv and the triple intersection $\bar{B}(p; \alpha) \cap \bar{B}(q; \alpha) \cap \bar{B}(v; \alpha)$ is not empty for $\alpha = d(p, q)/2$. ■*

Proof. For both filtrations, an edge e enters $C(A; \alpha)$ when $\alpha = |e|/2$. By Definition 2.3(b), a long edge enters $C(A; \alpha)$ together with a 2-simplex Δpqv . Since the other two edges $[p, v], [v, q]$ entered the filtration at a smaller scale, the edge $e = [p, q]$ is longest in Δpqv . For the Čech filtration, the triple intersection $\bar{B}(p; \alpha) \cap \bar{B}(q; \alpha) \cap \bar{B}(v; \alpha)$ should be non-empty to guarantee that $\check{C}ech(A; \alpha)$ includes the 2-simplex Δpqv , see Definition 2.2(b). \square

Example 2.5. For any 3-point set $A \subset \mathbb{R}^N$, let the edges of A have lengths $|e_1| \leq |e_2| < |e_3|$. By Definition 2.3, in $\{VR(A; \alpha)\}$ the edge e_3 is long whilst the edges e_1, e_2 are short. If $|e_1| < |e_2| = |e_3|$, then the edge e_1 is short but both edges e_2, e_3 are medium, not long. If $|e_1| = |e_2| = |e_3|$, then all three edges are medium.

Let $C(A; \alpha)$ be any geometric complex from Definition 2.2 on a finite set $A \subset \mathbb{R}^2$. If A consists of four vertices of the unit square, all square sides are medium whilst both diagonals are long. If A consists of four vertices of a rectangle that is not a square, the two shorter sides are short, the longer sides are medium and both diagonals are long. \blacksquare

Proposition 2.6 (Three classes of edges). For any finite set A and a filtration $\{C(A; \alpha)\}$ from Definition 2.2, all edges are split into three disjoint classes: short, medium, long. \blacksquare

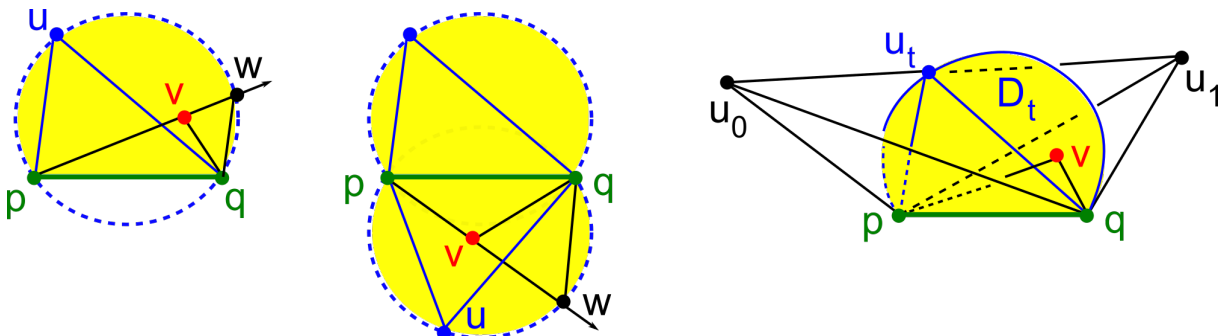
Proof. By Definition 2.3(b), the endpoints p, q of any long edge $e = [p, q] \subset C(A; \alpha)$ are connected by a chain of two edges $[p, v] \cup [v, q]$ that entered the filtration at a smaller scale $\alpha' < \alpha$. Hence the long edge e cannot be short by Definition 2.3(a). Hence the three classes of edges in Definition 2.3 are disjoint. \square

Lemmas 2.7 and 2.8 will interpret long edges in a Delaunay complex in terms of angles.

Lemma 2.7 (Circumdisk of a triangle). Let two triangles $\Delta pqv, \Delta pqw \subset \mathbb{R}^2$ lie on the same side of a common edge $[p, q]$. If $\angle puq < \angle pvq$, for example, if $\angle puq$ is acute and $\angle pvq$ is non-acute, then the open circumdisk of Δpqv contains w , see Fig. 3 (left).

Proof. Let the infinite ray from p via v meet the circumcircle C of Δpqv at a point w . We have equal angles $\angle puq = \angle pwq$ whose sine is $\frac{d(p, q)}{2R}$, where R is the radius of C . Since $\angle puq = \angle pwq < \angle pvq$, the point v is inside the edge $[p, w]$, hence enclosed by C . \square

Figure 3. **Left:** the circumdisk of Δpqv contains $w \in A$, see Lemma 2.7 and Lemma 2.8 for $N = 2$ when $[p, q]$ is on the boundary of the convex hull of A . **Middle:** a proof of Lemma 2.8 for $N = 2$ when $[p, q]$ is inside the convex hull of A . **Right:** the circumdisk D_t of Δpqv_t contains $v \in A$, leading to a contradiction in the proof of Lemma 2.8 for $N \geq 3$.



Lemma 2.8 (Long edges in $\text{Del}(A; \alpha)$). *Let $\alpha \geq 0$ and $A \subset \mathbb{R}^N$ be a finite set. An edge $e = [p, q]$ in the Delaunay complex $\text{Del}(A; \alpha)$ is long by Definition 2.3(b) if and only if*

- (a) $\text{Del}(A; \alpha)$ includes a triangle Δpqv whose angle at v is non-acute or, equivalently,
- (b) the set A has a point v whose angle in the triangle Δpqv is non-acute. ■

Proof. (a) By Definition 2.3(b) the edge e is long if the Delaunay complex $\text{Del}(A; \alpha)$ includes a triangle Δpqv whose edge $e = [p, q]$ is strictly longest (hence the opposite angle at v is strictly largest) and $\bar{B}(p; \alpha) \cap \bar{B}(q; \alpha) \cap \bar{B}(v; \alpha) \neq \emptyset$ for $\alpha = d(p, q)/2$. Since the intersection $\bar{B}(p; \alpha) \cap \bar{B}(q; \alpha)$ is the mid-point u of the edge e , the triple intersection above is non-empty if and only if $d(u, v) \leq \alpha$. Equivalently, the circumcentre of Δpqv lies non-strictly outside the triangle Δpqv or the angle at v in Δpqv is non-acute.

(b) Due to part (a), it suffices to prove that if we have any triangle Δpqv with a non-acute angle at v , we can find such a triangle within $\text{Del}(A; \alpha)$. Assume the contrary that all triangles in $\text{Del}(A; \alpha)$ containing $[p, q]$ have only acute angles opposite to $[p, q]$.

For $N = 2$, the edge $[p, q]$ can have one or two Delaunay triangles whose edge $[p, q]$ is on the boundary or inside the convex hull of A , see the first two pictures of Fig. 3. In both cases by Lemma 2.7, the above point v with a non-acute angle opposite to $[p, q]$ should be inside the circumdisk of one of these triangles Δpqv having an acute angle at the vertex u opposite to $[p, q]$. Then the triangle Δpqv cannot be in $\text{Del}(A; \alpha)$ by Definition 2.2(c).

For $N \geq 3$, consider all $(N - 1)$ -dimensional Delaunay simplices containing the edge $[p, q]$. The above point v lies (non-strictly) between a pair of successive $(N - 1)$ -dimensional subspaces spanned by two such simplices σ_0, σ_1 with common edge $[p, q]$. Let D^N be the circumball of the N -dimensional simplex $\sigma \in \text{Del}(A; \alpha)$ with faces σ_0, σ_1 .

Choose a 1-parameter family of 2-dimensional planes P_t , $t \in [0, 1]$, rotating around $[p, q]$ from σ_0 to σ_1 so that $\Delta pqv_0 = P_0 \cap \sigma \subset \sigma_0$ and $\Delta pqv_1 = P_1 \cap \sigma \subset \sigma_1$ are Delaunay triangles, while one intermediate plane P_t contains Δpqv with a non-acute angle opposite to $[p, q]$, see Fig. 3 (right). By the assumption, both $\Delta pqv_0, \Delta pqv_1 \in \text{Del}(A; \alpha)$ have acute angles opposite to $[p, q]$. The circumdisk D_t of each $\Delta pqv_t = P_t \cap \sigma$ has radius $R_t = \sqrt{R^2 - d_t^2}$, where R is the radius of D^N and d_t is the distance from the centre O of D^N to P_t . Then R_t varies from R_0 to R_1 over $t \in [0, 1]$, possibly with a maximum corresponding to the plane P_t passing through O , so $R_t \geq \min\{R_0, R_1\}$ for $t \in [0, 1]$.

By the sine theorem in each Δpqv_t , the angle opposite to $[p, q]$ has $\sin \angle pv_tq = \frac{d(p, q)}{2R_t}$. Since both Δpqv_0 and Δpqv_1 have acute angles, the lower bound $R_t \geq \min\{R_0, R_1\}$ implies that $\angle pv_tq \leq \min\{\angle pv_0q, \angle pv_1q\}$ is acute for all $t \in [0, 1]$. For the intermediate plane P_t containing the vertex v , by Lemma 2.7 the circumdisk $D_t \subset D^N$ of Δpqv_t should include the point v because $\angle pvq$ is non-acute and $\angle pu_tq$ is acute.

We get a contradiction with Definition 2.2(c) because the open circumball D^N of the Delaunay simplex $\sigma \in \text{Del}(A; \alpha)$ includes an extra point $v \in A$. □

3 Trivial persistence and tails without medium edges

As usual in TDA, we consider homology groups with coefficients in a field, say \mathbb{Z}_2 .

Proposition 3.1 (No medium edges \Rightarrow trivial H_1). *For any filtration $\{C(A; \alpha)\}$ on a finite abstract set A from Definition 2.1, when a scale $\alpha \geq 0$ is increasing, a new homology*

cycle in $H_1(C(A; \alpha))$ can be created only due to a medium edge in $C(A; \alpha)$. Hence, if $\{C(A; \alpha)\}$ has no medium edges, then $H_1(C(A; \alpha))$ is trivial for $\alpha \geq 0$. ■

Proof. When building the given complex $C(A; \alpha)$, if we add a short edge e , by Definition 2.3(a), the two previously disconnected components of $C^1(A; \alpha)$ containing the endpoints p, q of e become connected. Hence no 1-dimensional cycle in $C^1(A; \alpha)$ is created.

By Definition 2.3(b) any long edge $e = [p, q]$ enters $C(A; \alpha)$ strictly after two edges $[p, v], [v, q]$, and at the same time as the 2-simplex Δpqv . Any closed cycle γ including the new edge $[p, q]$ is homologically equivalent to the cycle with $[p, q]$ replaced by the 2-chain $[p, v] \cup [v, q]$. If γ has other edges e_1, \dots, e_k of the same length as e , the endpoints of e_i are connected by the complementary path $\gamma - e_i$, so each e_i cannot be short by Definition 2.3(a), $i = 1, \dots, k$. If any edge e_i , $i = 1, \dots, k$ is long by Definition 2.3(b), then e_i can be similarly replaced by a 2-chain of earlier edges.

In all cases, the cycle γ is homologically equivalent to a cycle in $C(A; \alpha')$ for a smaller $\alpha' < \alpha$. So a long edge cannot create a new class in $H_1(C(A; \alpha))$. Since only medium edges lead to non-trivial cycles, if A has no medium edges, then $H_1(C(A; \alpha))$ is trivial. □

Definition 3.2 (Tail T of points). *For a fixed filtration $\{C(A; \alpha)\}$ on a finite abstract set A from Definition 2.1, a tail $T \subset M$ is any ordered sequence $T = \{p_1, \dots, p_n\}$, where p_1 is the vertex of T , any edge between successive points $[p_i, p_{i+1}]$ is short, and any edge between non-successive points $[p_i, p_j]$ is long for any $1 \leq i < j \leq n$. ■*

Proposition 3.3 (Tails have trivial PD_1). *Any tail T from Definition 3.2 for a filtration $\{C(T; \alpha)\}$ of complexes from Definition 2.1 has trivial 1D persistence.*

Proof. Since any tail T has no medium edges by Proposition 2.6, the tail T has trivial $H_1(C(T; \alpha))$ for any $\alpha \geq 0$ by Proposition 3.1, hence trivial 1D persistence. □

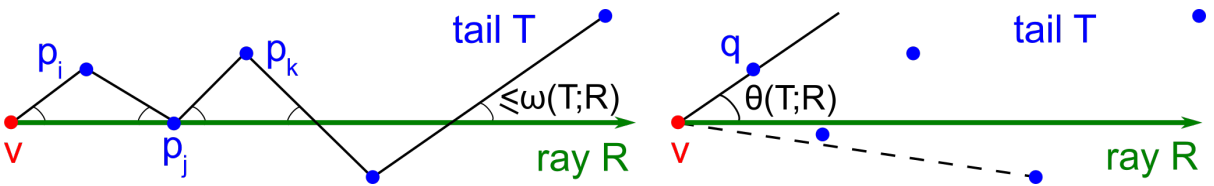
If vectors are not explicitly specified, all edges and straight lines are unoriented. We measure the angle between unoriented straight lines as their minimum angle within $[0, \frac{\pi}{2}]$.

Definition 3.4 (Angular deviation $\omega(T; R)$ from a ray R in \mathbb{R}^N). *In \mathbb{R}^N , a ray is any half-infinite line R going from a point v called the vertex of R . For any sequence $T = \{p_1, \dots, p_n\}$ of ordered points in \mathbb{R}^N , the angular deviation $\omega(T; R)$ of T relative to the ray R is the maximum angle $\angle(R, [p, q]) \in [0, \frac{\pi}{2}]$ over all distinct points $p, q \in T$. ■*

Figure 4. A tail T around a ray R with vertex v in \mathbb{R}^2 , see Definitions 3.4 and 3.6.

Left: all angles are not greater than the angular deviation $\omega(T; R)$.

Right: the angular thickness $\theta(T; R)$ can be smaller than $\omega(T; R)$.



Lemma 3.5 (Tails in \mathbb{R}^N). *Let R be a ray with vertex $v = p_1$ and $T = \{p_1, \dots, p_n\} \subset \mathbb{R}^N$ be any sequence of points with angular deviation $\omega(T; R) < \frac{\pi}{4}$.*

(a) For any p_i, p_j, p_k with $i < j < k$, the angle $\angle p_i p_j p_k$ is non-acute. The edge between the non-successive points p_i, p_k is long in any filtration $\{C(T; \alpha)\}$ in Definition 2.2.

(b) Any edge between successive points p_{i-1}, p_i , $i = 2, \dots, n$, is short in $\{C(T; \alpha)\}$.

Hence T has no medium edges in the filtration $\{C(T; \alpha)\}$ and is a tail by Definition 3.2.

Proof. (a) The condition $\omega(T; R) < \frac{\pi}{4}$ implies that all points of T are ordered by their distance from the vertex $v = p_1$ to their orthogonal projections to the line through R .

Apply a parallel shift to the points p_i, p_j, p_k so that $p_j \in R$. In the triangle $\Delta p_i p_j p_k$, the angle $\angle(\overrightarrow{p_j p_i}, \overrightarrow{p_j p_k}) = \pi - \angle(R, [p_j p_i]) - \angle(R, [p_j p_k]) \geq \pi - 2\omega(T; R) > \frac{\pi}{2}$ is non-acute, hence strictly largest, due to $\omega(T; R) < \frac{\pi}{4}$. The edge $[p_i, p_k]$ is long in any filtration $\{C(T; \alpha)\}$ by Definition 2.3(b) and (for the Delaunay filtration) Lemma 2.8(b). In particular, the edge $[p_i, p_k]$ is longer than both edges $[p_i, p_j]$ and $[p_j, p_k]$.

(b) The points p_{i-1}, p_i remain in disjoint components of $C^1(T; \alpha)$ after adding all other edges of length $d(p_{i-1}, p_i)$. Indeed, we proved above that any other edge connecting points p_j, p_k for $j \leq i-1 < i \leq k$ is longer than the edge $[p_{i-1}, p_i]$ between successive points. \square

The angular thickness below is illustrated in Fig. 4 (right) for Theorem 4.3 later.

Definition 3.6 (Angular thickness $\theta(T; R)$). Let $R \subset \mathbb{R}^N$ be a ray with vertex v and $T = \{p_1 = v, \dots, p_n\}$ be any finite sequence of points. The angular thickness $\theta(T; R)$ of T with respect to R is the maximum angle $\angle(\vec{R}, \overrightarrow{p_1 p_i})$ over $i = 2, \dots, n$. \blacksquare

4 Persistence for long wedges and with added tails

Definition 4.1 (Long wedges). Let A_1, \dots, A_k be finite sets sharing a common point v . For any filtration of complexes $\{C(\cup_{i=1}^k A_i; \alpha)\}$ from Definition 2.1, the wedge $\cup_{i=1}^k A_i$ is called long if any edge $[q_i, q_j]$ between distinct points $q_i \in A_i, q_j \in A_j, i \neq j$, is long in the filtration $\{C(\cup_{i=1}^k A_i; \alpha)\}$ in the sense of Definition 2.3(b). \blacksquare

For any filtration of simplicial complexes $\{C(A; \alpha)\}$ from Definition 2.1, the 1D persistence diagram of this filtration is denoted by $\text{PD}_1\{C(A; \alpha)\}$.

Theorem 4.2 (Persistence of long wedges). For any filtration $\{C(\cup_{i=1}^k A_i; \alpha)\}$ of a long wedge from Definition 4.1, the 1D homology group of the filtration at a given α is the direct sum: $H_1(C(\cup_{i=1}^k A_i; \alpha)) = \bigoplus_{i=1}^k H_1(C(A_i; \alpha))$. Hence the 1D persistence diagram $\text{PD}_1\{C(\cup_{i=1}^k A_i; \alpha)\}$ is the union of the 1D persistence diagrams $\text{PD}_1\{C(A_i; \alpha)\}$ for $i = 1, \dots, k$. \blacksquare

Proof. The inclusions $A_i \subset \cup_{i=1}^k A_i$ induce the homomorphism of the 1D homology groups $\bigoplus_{i=1}^k H_1(C(A_i; \alpha)) \rightarrow H_1(C(\cup_{i=1}^k A_i; \alpha))$ whose bijectivity follows below. Any long edge $e = [p, q]$ in a complex $C(\cup_{i=1}^k A_i; \alpha)$ can be replaced by a chain of two edges $[p, v] \cup [v, q]$ in $C(\cup_{i=1}^k A_i; \alpha')$ for some $\alpha' < \alpha$ due to a 2-simplex Δpqv included into $C(\cup_{i=1}^k A_i; \alpha)$ by Definition 2.3(b). Continue applying these replacements until any cycle of edges in $C(\cup_{i=1}^k A_i; \alpha)$ becomes homologous to a sum of cycles in $C(A_i; \alpha)$, $i = 1, \dots, k$. \square

Theorem 4.3 (A long wedge with a tail). Let $A \subset \mathbb{R}^N$ be any finite set, $v \in A$ be a point on the boundary of the convex hull of A , and R be a ray with vertex v so that $\mu(R; A) = \min_{p \in A - \{v\}} \angle(\vec{R}, \overrightarrow{vp}) > \frac{\pi}{2}$. Let T be any tail with vertex v for a filtration $\{C(T; \alpha)\}$, see Definition 3.2. If $\mu \geq \theta(T; R) + \frac{\pi}{2}$, then $\text{PD}_1\{C(A \cup T; \alpha)\} = \text{PD}_1\{C(T; \alpha)\}$, \blacksquare

Proof. For any points $p \in A$ and $q \in T$, we get the non-acute angle

$$\angle(\vec{vp}, \vec{vq}) \geq \angle(\vec{R}, \vec{vp}) - \angle(\vec{R}, \vec{vq}) \geq \mu - \angle(\vec{R}, \vec{vq}) \geq \mu - \theta(T; R) \geq \frac{\pi}{2}.$$

If p, q are in the same half-plane bounded by the line through R , the first inequality above becomes equality. Otherwise, $\angle(\vec{vp}, \vec{vq}) = \angle(\vec{R}, \vec{vp}) + \angle(\vec{R}, \vec{vq}) \geq \angle(\vec{R}, \vec{vp}) - \angle(\vec{R}, \vec{vq})$.

Then the edge $[p, q]$ is long in the filtration $\{C(A \cup T; \alpha)\}$, not medium, by Definition 2.3(b). Hence $A \cup T$ is a long wedge by Definition 2.3(b). Since the tail T has trivial 1D persistence by Proposition 3.3, Theorem 4.2 implies that the persistence diagrams are identical: $\text{PD}_1\{C(A \cup T; \alpha)\} = \text{PD}_1\{C(T; \alpha)\}$. \square

Corollary 4.4 (Trivial 1D persistence). *If a set A in a metric space M has $\text{PD}_1\{C(A; \alpha)\} = \emptyset$, then any long wedge $A \cup T$ with a tail $T \subset M$ also has $\text{PD}_1\{C(A \cup T; \alpha)\} = \emptyset$.* \blacksquare

Proof. Since the tail T has trivial 1D persistence by Proposition 3.3, Theorem 4.2 implies that $\text{PD}_1\{C(A \cup T; \alpha)\} = \text{PD}_1\{C(A; \alpha)\} = \emptyset$. \square

5 Experiments, discussion and future work

The experiments in this section increase the understanding of how regularly persistent homology reveals persistent features separated from noise. The experiments depend on two parameters, the size n of a point set, and the dimension N that the point set lies in. For each n, N in the ranges chosen, we generate 1000 point sets of n points uniformly sampled in a unit N -dimensional cube.

Figure 5 shows histograms of persistence (death–birth) of the one-dimensional features for nine configurations of the parameters: point set sizes $n = 10, 15, 20$ and dimensions $N = 2, 5, 8$. Each histogram highlights that the overwhelming majority of one-dimensional persistent features are skewed towards a low persistence, namely less than 10% of the unit cube size. Geometrically, the corresponding dots (birth,death) would be close to the diagonal in a persistence diagram.

Recall that highly persistent features (birth,death) are naturally separated from others with lower persistence $p = \text{death} - \text{birth}$ by the widest diagonal gap in the persistence diagram, see [19]. If we order all pairs (birth,death) by their persistence $0 < p_1 \leq \dots \leq p_k$, the widest gap has the largest difference $p_{i+1} - p_i$ over $i = 1, \dots, k - 1$. This widest gap can separate several pairs (birth,death) from the rest, not necessarily just a single feature. However, the first widest gap is significant only if it can be easily distinguished from the second widest gap.

So the significance of persistence can be measured as the ratio of the first widest gap over the second widest gap. This invariant up to uniform scaling of given data is called the *gap ratio*. Figure 6 shows the median gap ratio calculated over 1000 random point clouds in a unit cube for many dimensions $N = 2, \dots, 10$ and point set sizes $n = 10, \dots, 40$.

Figure 6 implies that for higher dimensions N , the median gap ratio quickly decreases to within the range $[1, 2]$ as the number n of points is increasing. Hence, when a persistence diagram contains at least two pairs (birth,death) above the diagonal, it is becoming harder to separate highly persistent features from noisy artefacts close to the diagonal.

The future updates will include similar experiments for filtrations of Čech complexes and Delaunay complexes. In conclusion, our main Theorem 4.3 describes how we can add

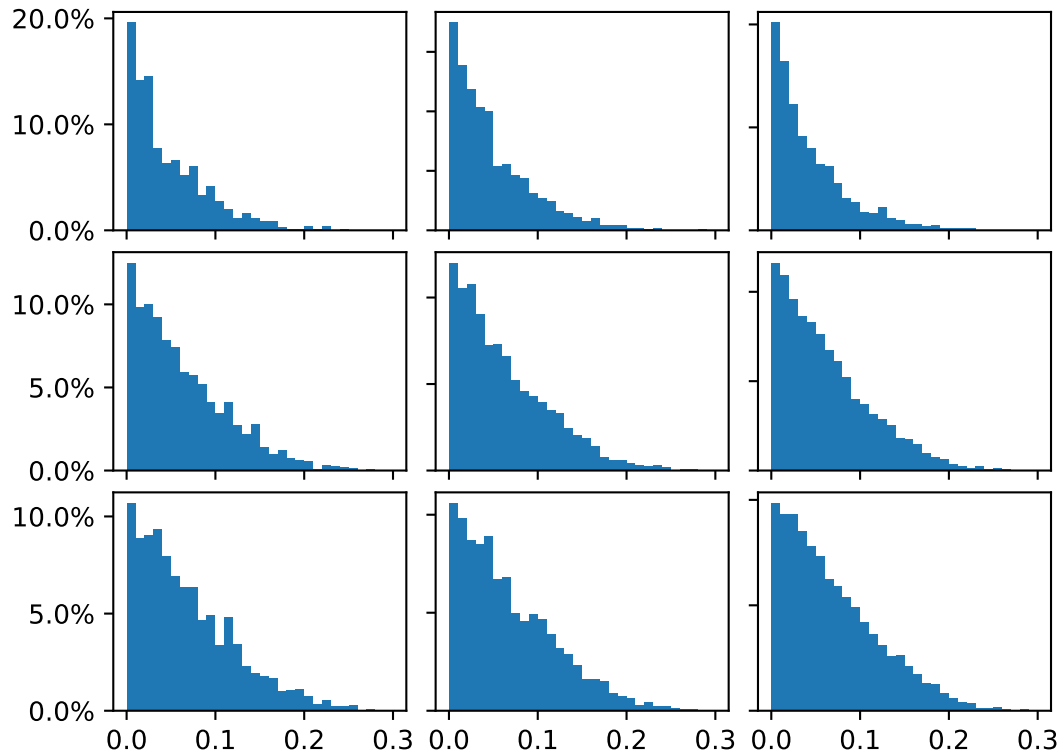


Figure 5. Histograms of the persistence $p = \text{death} - \text{birth}$ in 1000 point sets in nine configurations of the parameters n and N . The x -axis is the persistence p , the y -axis is the percentage of pairs (birth,death) with the given persistence p . Top row: $N = 2$; middle row: $N = 5$; bottom row: $N = 8$. Left column: $n = 10$; middle column: $n = 15$; right column $n = 20$.

an arbitrarily large point set to an existing point set without affecting the one-dimensional persistent homology, whilst Corollary 4.4 states how we can form a large continuous family of sets with trivial 1D persistence, implying that the bottleneck distance between persistence diagrams has no lower bound. We plan further experiments to check how well the bottleneck distance separates point clouds from their perturbations.

Other continuous isometry invariants [21,22] of finite and periodic point sets are complete in general position, hence distinguish almost all sets in \mathbb{R}^N . All counter-examples [18] to the completeness of past invariants were distinguished in [21, appendix C]. These latest invariants are based on the k -nearest neighbour search, a classical problem in Computer Science, which has near-linear time algorithms in the number of points [13,14].

This research was supported by the £3.5M EPSRC grant ‘Application-driven Topological Data Analysis’ (2018-2023, EP/R018472/1), the £10M Leverhulme Research Centre for Functional Materials Design (2016-2026) and the last author’s Royal Academy of Engineering Fellowship ‘Data Science for Next Generation Engineering of Solid Crystalline Materials’ (2021-2023, IF2122/186).

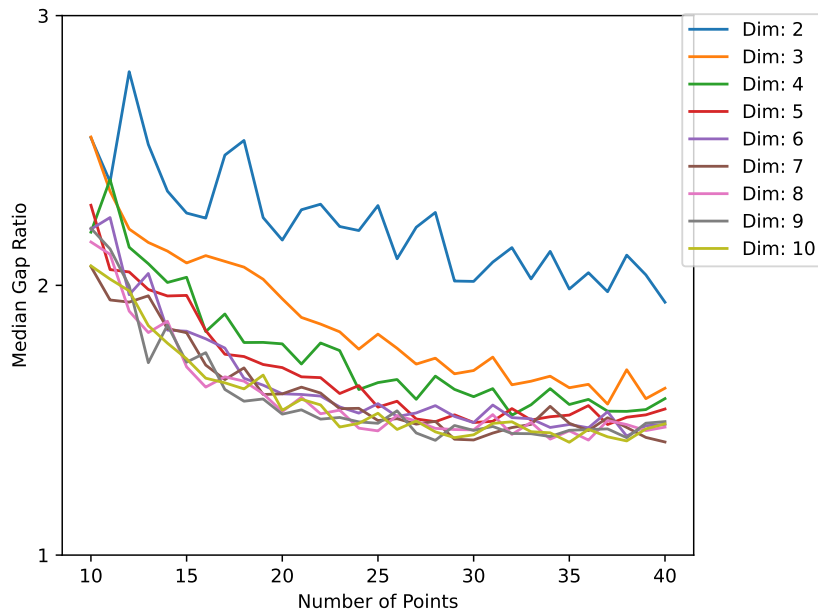


Figure 6. The median gap ratio of a point set with at least two 1D persistent features, as the size of the point set varies from $n = 10$ to $n = 40$ and the dimension N varies from $N = 2$ to $N = 10$.

References

- [1] Bauer, U.: Ripser: efficient computation of vietoris-rips persistence barcodes. *Journal of Applied and Computational Geometry* **5**(3), 391–423 (2021). DOI 10.1007/s41468-021-00071-5
- [2] Biswas, R., Cultrera di Montesano, S., Edelsbrunner, H., Saghafian, M.: Continuous and discrete radius functions on voronoi tessellations and delaunay mosaics. *Discrete & Computational Geometry* pp. 1–32 (2022)
- [3] Boissonnat, J.D., Dyer, R., Ghosh, A., Martynchuk, N.: An obstruction to delaunay triangulations in riemannian manifolds. *Discrete & Comp. Geometry* **59**(1), 226–237 (2018)
- [4] Carlsson, G.: Topology and data. *Bulletin of the American Mathematical Society* **46**(2), 255–308 (2009)
- [5] Catanzaro, M.J., Curry, J.M., Fasy, B.T., Lazovskis, J., Malen, G., Riess, H., Wang, B., Zabka, M.: Moduli spaces of morse functions for persistence. *Journal of Applied and Computational Topology* **4**(3), 353–385 (2020)
- [6] Chazal, F., De Silva, V., Glisse, M., Oudot, S.: *The structure and stability of persistence modules*. Springer (2016)
- [7] Cohen-Steiner, D., Edelsbrunner, H., Harer, J.: Stability of persistence diagrams. *Discrete & Computational Geometry - DCG* **37**, 263–271 (2005). DOI 10.1007/s00454-006-1276-5
- [8] Curry, J.: The fiber of the persistence map for functions on the interval. *Journal of Applied and Computational Topology* **2**(3), 301–321 (2018)

- [9] Delaunay, B.: Sur la sphere vide. *Izv. Akad. Nauk USSR* **7**, 793–800 (1934)
- [10] Edelsbrunner, H., Harer, J.: Persistent homology - a survey. *Discrete & Computational Geometry - DCG* **453** (2008). DOI 10.1090/conm/453/08802
- [11] Edelsbrunner, H., Heiss, T., Kurlin, V., Smith, P., Wintraecken, M.: The density fingerprint of a periodic point set. In: *Proceedings of SoCG*, pp. 32:1–32:16 (2021)
- [12] Edelsbrunner, H., Letscher, D., Zomorodian, A.: Topological persistence and simplification. In: *Proceedings 41st annual symposium on foundations of computer science*, pp. 454–463 (2000)
- [13] Elkin, Y., Kurlin, V.: A new compressed cover tree guarantees a near linear parameterized complexity for all k -nearest neighbors search in metric spaces. arXiv:2111.15478 (2021)
- [14] Elkin, Y., Kurlin, V.: Paired compressed cover trees guarantee a near linear parametrized complexity for all k -nearest neighbors search in an arbitrary metric space. arXiv:2201.06553 (2022)
- [15] Frosini, P., Landi, C.: Size theory as a topological tool for computer vision. *Pattern Recognition and Image Analysis* **9**(4), 596–603 (1999)
- [16] Ghrist, R.: Barcodes: the persistent topology of data. *Bulletin of the American Mathematical Society* **45**(1), 61–75 (2008)
- [17] Hatcher, A.: *Algebraic Topology*. Cambridge University Press (2001)
- [18] Pozdnyakov, S., Willatt, M., Bartók, A., Ortner, C., Csányi, G., Ceriotti, M.: Incompleteness of atomic structure representations. *Phys. Rev. Lett.* **125**, 166001 (2020). URL arXiv:2001.11696
- [19] Smith, P., Kurlin, V.: Skeletonisation algorithms with theoretical guarantees for unorganised point clouds with high levels of noise. *Pattern Recognition* **115**, 107902 (2021)
- [20] Weinberger, S.: What is... persistent homology? *Notices of the AMS* **58**(1), 36–39 (2011)
- [21] Widdowson, D., Kurlin, V.: Pointwise distance distributions of periodic sets. arXiv:2108.04798 (early draft) (2021). URL <http://kurlin.org/projects/periodic-geometry-topology/PDD.pdf>
- [22] Widdowson, D., Mosca, M., Pulido, A., Kurlin, V., Cooper, A.: Average minimum distances of periodic point sets. *MATCH Communications in Mathematical and in Computer Chemistry* **87**, 529–559 (2022). URL <http://kurlin.org/projects/periodic-geometry-topology/AMD.pdf>