Generic families of finite metric spaces with identical or trivial 1-dimensional persistence

Philip Smith[1] and Vitaliy Kurlin[1*]

[1*]Department of Computer Science, University of Liverpool, Ashton street, Liverpool, L69 3BX, United Kingdom.

*Corresponding author(s). E-mail(s): vitaliy.kurlin@gmail.com;
Contributing authors: philip.smith3@liverpool.ac.uk;

Abstract

Persistent homology is a popular and useful tool for analysing finite metric spaces, revealing features that can be used to distinguish sets of unlabeled points and as input into machine learning pipelines. The famous stability theorem of persistent homology provides an upper bound for the change of persistence in the bottleneck distance under perturbations of points, but without giving a lower bound.

This paper clarifies the possible limitations persistent homology may have in distinguishing finite metric spaces, which is evident for non-isometric point sets with identical persistence. We describe generic families of point sets in metric spaces that have identical or even trivial one-dimensional persistence. The results motivate stronger invariants to distinguish finite point sets up to isometry.

Keywords: persistent homology, isometry invariant, unlabeled point cloud

1 Motivations, problems, and outline of results

Topological Data Analysis (TDA) was pioneered by Serguei Barannikov [1], Claudia Landi [2], Vanessa Robins [3], and Herbert Edelsbrunner et al. [4]. The key papers of Gunnar Carlsson [5], Robert Ghrist [6], and Shmuel Weinberger [7] were followed by substantial developments of Fred Chazal [8] and others.

The main tool of TDA [9] is persistent homology, which is defined via filtration of complexes on a finite set A of unlabeled points in a Euclidean

or metric space. For standard filtrations of geometric complexes on A, the resulting persistence diagram or barcode is invariant up to isometry preserving inter-point distances, not up to more general continuous deformations.

Hence most persistence-based classifications distinguish point data only up to isometry, which is an important equivalence due to the rigidity of many real-life structures. Fig. 1 shows sets $A \subset \mathbb{R}^2$ whose points (in blue and red) form 1×2 'dominoes' that have identical persistence in dimensions 0 and 1.

The 0-dimensional persistence is determined by edge lengths of a Minimum Spanning Tree shown in green: one edge of length 2 and eight edges of length 1 for all sets in Fig. 1. The 1-dimensional persistence is trivial in these cases because all rectangular 'dominoes' do not create cycles in the filtrations of Vietoris-Rips, Cech, and Delaunay complexes, see details in Definition 2.2.

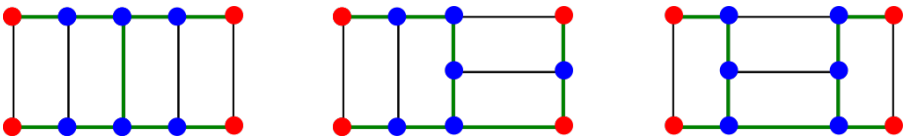

Fig. 1 Many non-isometric sets have the same 0D persistence and trivial 1D persistence.

To understand the strength of persistence as an isometry invariant, the following problem asks to fully describe the inverse of the persistence map.

Problem 1.1 (inverting persistence) *Give necessary and sufficient conditions for finite metric spaces to have identical persistence for a given filtration of complexes, e.g. describe all* 1D homologically trivial *point sets that have trivial 1D persistence.*

The analogue of Problem 1.1 was solved for 0-dimensional persistence of Morse functions on the interval [10] and the sphere S^2 [11]. Main Theorem 4.3 shows how a point sequence of arbitrary size can be added to any finite point set whilst leaving the 1-dimensional persistence unchanged, see Fig 2.

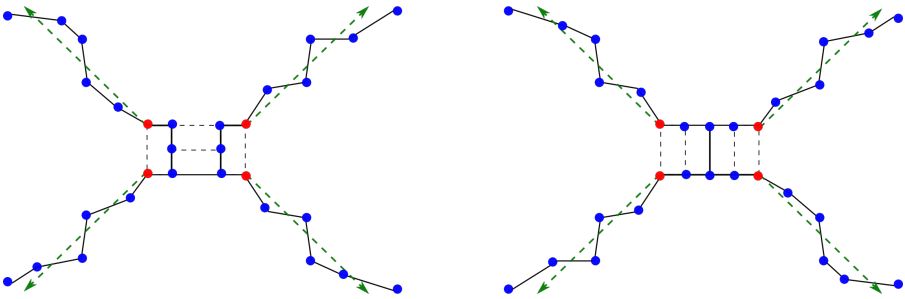

Fig. 2 The set A of 10 points in the centre is extended by four tails going out from red points. All such sets have trivial 1D persistence by Corollary 4.4, but all such sets in general position are not isometric to each other. The black edges form a Minimum Spanning Tree.

Corollary 4.4 describes generic families of finite metric spaces that have trivial 1-dimensional persistence computed using filtrations of simplicial complexes including Vietoris-Rips, Cech, and Delaunay complexes. These settings are standard for most applications, especially for high-dimensional data. Higher dimensional persistent homology will be included in future updates.

In the context of Problem 1.1, the resulting families of point sets in $\mathbb{R}^N$ form vast open subspaces mapped by persistence to a single value. This result complements the famous stability theorem [12] stating that under bounded noise, the bottleneck distance between persistence diagrams of a point set and its perturbation has an upper bound depending on the magnitude of the perturbation. However, there is no lower bound, so a perturbation of a point set can result in the corresponding persistent homology remaining unchanged.

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. No change in persistence under small perturbations motivates stronger isometry invariants discussed in section 6. Indeed, many applications [13] need 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 a more general continuous deformation of data changes persistence rather arbitrarily.

2 Edges that are 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 introduces Vietoris-Rips, Čech, and Delaunay complexes on a finite set A in any metric space M or for $A \subset \mathbb{R}^N$ for Delaunay complexes.

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

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

(b) The *dimension* of a simplex σ on $k+1$ 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 a family of simplicial complexes on the vertex set A, paremeterised by a *scale* α so that $C(A;\ \alpha') \subseteq C(A;\ \alpha)$ for $\alpha' \leq \alpha$. ■

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 $d(p, q)$, which will

be denoted by $|e|$ in the Euclidean space. 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 $\mathrm{VR}(A;\ \alpha)$ and $\check{\mathrm{C}}\mathrm{ech}(A;\ \alpha)$ 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 $\mathrm{Del}(A;\ \alpha) \subset \mathbb{R}^N$ will be defined for a finite set A only in $\mathbb{R}^N$ because of extra complications arising if a point set A lives in a more general metric space [14].

Definition 2.2 (Complexes $\mathrm{VR}(A;\ \alpha), \check{\mathrm{C}}\mathrm{ech}(A;\ \alpha), \mathrm{Del}(A;\ \alpha)$) 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 $\mathrm{VR}(A;\ \alpha)$ has all simplices on points $p_1, \ldots, p_k \in A$ whose pairwise distances are all at most 2α, so $d(p_i, p_j) \leq 2\alpha$ for all distinct $i, j \in \{1, \ldots, k\}$.

(b) The *Čech* complex $\check{\mathrm{C}}\mathrm{ech}(A;\ \alpha)$ has all simplices on points $p_1, \ldots, p_k \in A$ 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*

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

The Delaunay complex $\mathrm{Del}(A;\ \alpha)$ has all simplices on points $p_1, \ldots, p_k \in A$ such that the intersection $\cap_{i=1}^{k} (V(p_i) \cap \bar{B}(p_i;\ \alpha))$ is not empty [15]. Alternatively, a simplex σ on points $p_1, \ldots, p_k \in A$ is a *Delaunay* simplex if: (a) the smallest $(k-2)$-dimensional sphere S^{k-2} passing through $p_1, \ldots, p_k$ has a radius at most α; (b) there is also an $(N-1)$-dimensional sphere S^{N-1} passing through $p_1, \ldots, p_k$ 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 $\mathrm{Del}(A;\ \alpha)$ becomes a polyhedral *Delaunay mosaic* [16]. For simplicity, we choose any triangulation of H into Delaunay simplices. Then a Delaunay complex $\mathrm{Del}(A;\ \alpha) \subset \mathbb{R}^N$ is a subset of a Delaunay triangulation of the convex hull of A, which is unique in general position.

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

Both complexes $\mathrm{VR}(A;\ \alpha)$ and $\check{\mathrm{C}}\mathrm{ech}(A;\ \alpha)$ are abstract and so are not embedded in $\mathbb{R}^N$, even if $A \subset \mathbb{R}^N$. Though $\mathrm{Del}(A;\ \alpha)$ 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 $\mathrm{VR}(A;\ \alpha)$. Indeed, $\mathrm{VR}(A;\ \alpha)$ is a flag complex determined by its 1-dimensional skeleton $\mathrm{VR}^1(A;\ \alpha)$ so that any simplex of $\mathrm{VR}(A;\ \alpha)$ is built on a complete subgraph whose vertices are pairwise connected by edges in $\mathrm{VR}^1(A;\ \alpha)$.

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 $\check{\text{C}}\text{ech}(A;\ \alpha)$ 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 a 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 $\triangle pqv$ is 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 $\triangle pqv$ 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 $\alpha' < \alpha$.

Definition 2.3 defined classes of edges for any filtration of complexes. Lemma 2.4 interprets long edges in VR and Cech 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 A has a point v such that $e = [p, q]$ is a strictly longest edge in the triangle $\triangle pqv$.*

(b) *In the Čech filtration $\{\check{\text{C}}\text{ech}(A;\ \alpha)\}$, an edge e is long if and only if $\check{\text{C}}\text{ech}(A;\ \alpha)$ includes a triangle $\triangle pqv$ such that the edge $e = [p, q]$ is strictly longest in $\triangle pqv$ 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 $\triangle pqv$. Since the other two edges $[p, v], [v, q]$ entered the filtration at a smaller scale, the edge $e = [p, q]$ is longest in $\triangle pqv$. 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 $\check{\text{C}}\text{ech}(A;\ \alpha)$ includes the 2-simplex $\triangle pqv$, see Definition 2.2(b). □

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 $\{\mathrm{VR}(A;\ \alpha)\}$ 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. ■

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

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. □

Lemmas 2.7 and 2.8 interpret long edges in a Delaunay complex via angles.

Lemma 2.7 (Circumdisk of a triangle) *Let two triangles $\triangle pqu$, $\triangle pqv \subset \mathbb{R}^2$ 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 $\triangle pqu$ contains v, see Fig. 3 (left).*

Proof Let the infinite ray from p via v meet the circumcircle C of $\triangle pqu$ at a point w. The sine of $\angle puq = \angle pwq$ is $\dfrac{|p-q|}{2R}$, where R is the radius of C. Since $\angle puq = \angle pwq < \angle pvq$, the point v is inside $[p, w]$, hence enclosed by C, see Fig. 3. □

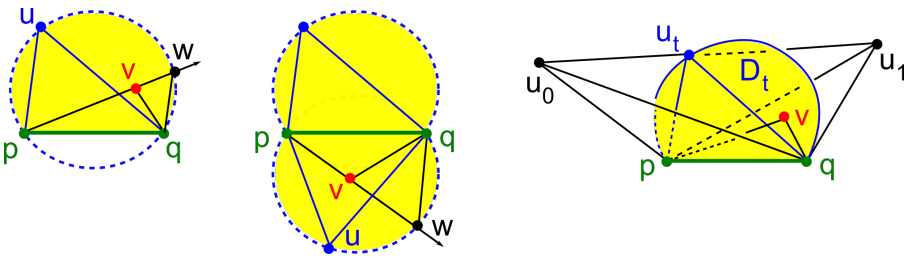

Fig. 3 **Left**: the circumdisk of $\triangle pqu$ contains $v \in A$, see Lemma 2.7 and Lemma 2.8 for $N = 2$ when $[p, q]$ is on the boundary of the convex hull of A. **Middle**: a proof of Lemma 2.8 for $N = 2$ when $[p, q]$ is inside the convex hull of A. **Right**: the circumdisk D_t of $\triangle pqu_t$ 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 Del(A; α)) *Let $\alpha \geq 0$ and $A \subset \mathbb{R}^N$ be a finite set. An edge $e = [p, q]$ in the complex* Del(A; α) *is long by Definition 2.3(b) if and only if*

(a) Del(A; α) *has a triangle $\triangle pqv$ whose angle at v is non-acute or, equivalently,*

(b) *the set A has a point v whose angle in the triangle $\triangle pqv$ is non-acute.* ■

Proof **(a)** By Definition 2.3(b) the edge e is long if the Delaunay complex Del(A; α) includes a triangle $\triangle pqv$ 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 $\alpha = \dfrac{|p-q|}{2}$. Since the intersection $\bar{B}(p;\ \alpha) \cap \bar{B}(q;\ \alpha)$ is the mid-point u of e, the triple intersection above is non-empty if and only if $d(u, v) \leq \alpha$. Equivalently, the circumcentre of $\triangle pqv$ is non-strictly outside $\triangle pqv$ or the angle at v in $\triangle pqv$ is non-acute.

(b) Due to **(a)**, it suffices to prove that if we have any triangle $\triangle pqv$ with a non-acute angle at v, we can find such a triangle within Del(A; α). Assume the contrary that all triangles in Del(A; α) 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 $\triangle pqu$ having an acute angle at the vertex u opposite to $[p, q]$. Then the triangle $\triangle pqu$ cannot be in Del(A; α) by Definition 2.2(c).

For $N \geq 3$, consider all $(N-1)$-dimensional Delaunay simplices containing the edge $[p, q]$. Then 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 $\sigma \in$ Del(A; α) 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 $\triangle pqu_0 = P_0 \cap \sigma \subset \sigma_0$ and $\triangle pqu_1 = P_1 \cap \sigma \subset \sigma_1$ are Delaunay triangles, while one intermediate plane P_t contains $\triangle pqv$ with a non-acute angle opposite to $[p, q]$, see Fig. 3 (right). By the assumption, both $\triangle pqu_0, \triangle pqu_1 \in$ Del(A; α) have acute angles opposite to $[p, q]$. The circumdisk D_t of each $\triangle pqu_t = P_t \cap \sigma$ 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 $\triangle pqu_t$, the angle opposite to $[p, q]$ has $\sin \angle pu_tq = \dfrac{|p-q|}{2R_t}$. Since both $\triangle pqu_0$ and $\triangle pqu_1$ have acute angles, the lower bound $R_t \geq \min\{R_0, R_1\}$ implies that $\angle pqu_t \leq \min\{\angle pqu_0, \angle pqu_1\}$ 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 $\triangle pqu_t$ should include 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 $\sigma \in$ Del(A; α) includes an extra point $v \in A$. □

3 Tails without medium edges in a metric space

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

Proposition 3.1 (No medium edges ⇒ trivial H_1) *For any filtration $\{C(A;\ \alpha)\}$ on a finite 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 complex $C(A;\ \alpha)$, if we add a short edge e, by Definition 2.3(a), the previously disjoint 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 $\triangle pqv$. 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.

Hence γ is homologically equivalent to a cycle in $C(A;\ \alpha')$ for some $\alpha' < \alpha$. So a long edge cannot create a 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 of points) For a fixed filtration $\{C(A;\ \alpha)\}$ on a finite set A from Definition 2.1, a *tail* T in a metric space M is any ordered sequence $T = \{p_1, \dots, p_n\}$, where p_1 is the *vertex* of T, any edge $[p_i, p_{i+1}]$ between successive points is short, and any edge $[p_i, p_j]$ between non-successive points 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 $\mathbb{R}^N$, a *ray* is any half-infinite line R going from a point v (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 R is the maximum angle $\angle(R, [p, q]) \in [0, \frac{\pi}{2}]$ over all distinct points $p, q \in T$. ■

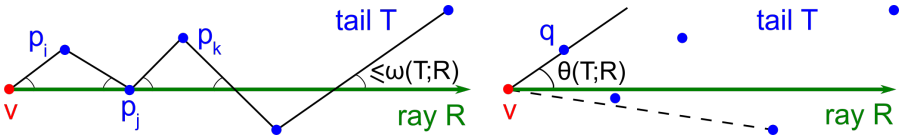

Fig. 4 A tail T around a ray R with vertex v in $\mathbb{R}^2$, see Definitions 3.4 and 3.6. **Left**: all angles are not greater than the angular deviation $\omega(T;\ R)$. **Right**: the angular thickness $\theta(T;\ R)$ can be smaller than $\omega(T;\ R)$.

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 $i < j < k$, 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 $\{C(T;\ \alpha)\}$ 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 $\triangle p_i p_j p_k$, 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 $[p_i, p_j]$ 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. □

Fig. 4 (right) illustrates the angular thickness below 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 $i = 2, \dots, n$. ■

4 Persistence for long wedges and with 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). ■

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

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:* $H_1(C(\cup_{i=1}^k A_i;\ \alpha)) = \oplus_{i=1}^k H_1(C(A_i;\ \alpha))$. *Hence the 1D persistence diagram* $\mathrm{PD}_1\{C(\cup_{i=1}^k A_i;\ \alpha)\}$ *is the union of the 1D persistence diagrams* $\mathrm{PD}_1\{C(A_i;\ \alpha)\}$ *for* $i = 1, \dots, k$. ■

Proof The inclusions $A_i \subset \cup_{i=1}^k A_i$ induce the homomorphism of the 1D homology groups $\oplus_{i=1}^k H_1(C(A_i;\ \alpha)) \to H_1(C(\cup_{i=1}^k A_i;\ \alpha))$ 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 $\triangle pqv$ 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 $C(A_i;\ \alpha)$, $i = 1, \dots, k$. □

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* $\mu(R;\ A) = \min\limits_{p \in A - \{v\}} \angle(\vec{R}, \overrightarrow{vp}) > \frac{\pi}{2}$. *Let* T *be any tail with vertex* v *for a filtration* $\{C(T;\ \alpha)\}$. *If* $\mu \geq \theta(T;\ R) + \frac{\pi}{2}$, *then* $\mathrm{PD}_1\{C(A \cup T;\ \alpha)\} = \mathrm{PD}_1\{C(T;\ \alpha)\}$. ■

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

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

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

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: $\mathrm{PD}_1\{C(A \cup T;\ \alpha)\} = \mathrm{PD}_1\{C(T;\ \alpha)\}$. □

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

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

5 Experiments on persistence of random sets

The experiments in this section use the Vietoris-Rips filtration whose 1-dimensional persistence is computed by the super-fast software Ripser[18].

The aim is to understand how often random point sets have trivial persistence or cycles with only low persistence, see more general conjectures [19]. The experiments depend on two parameters, the size n of a 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.

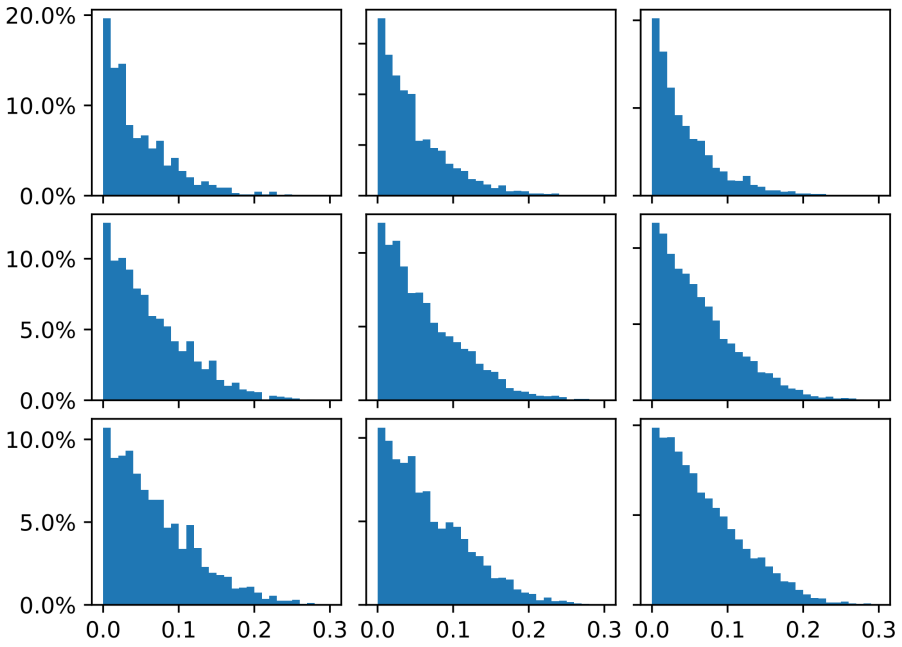

Fig. 5 Histograms of the persistence $p = \text{death} - \text{birth}$ in 1000 point sets in nine configurations of the parameters n and N. The x-axis is the persistence p, the y-axis is the percentage of pairs (birth,death) with the given persistence p. Top row: $N = 2$; middle row: $N = 5$; bottom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

Figure 5 shows histograms of the 1-dimensional persistence (death−birth) for nine configurations of the parameters: 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 [20]. If we order all pairs (birth,death) by their persistence $0 < p_1 \le \cdots \le p_k$, the widest gap has the largest difference $p_{i+1} - p_i$ over $i = 1, \ldots, k-1$. 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

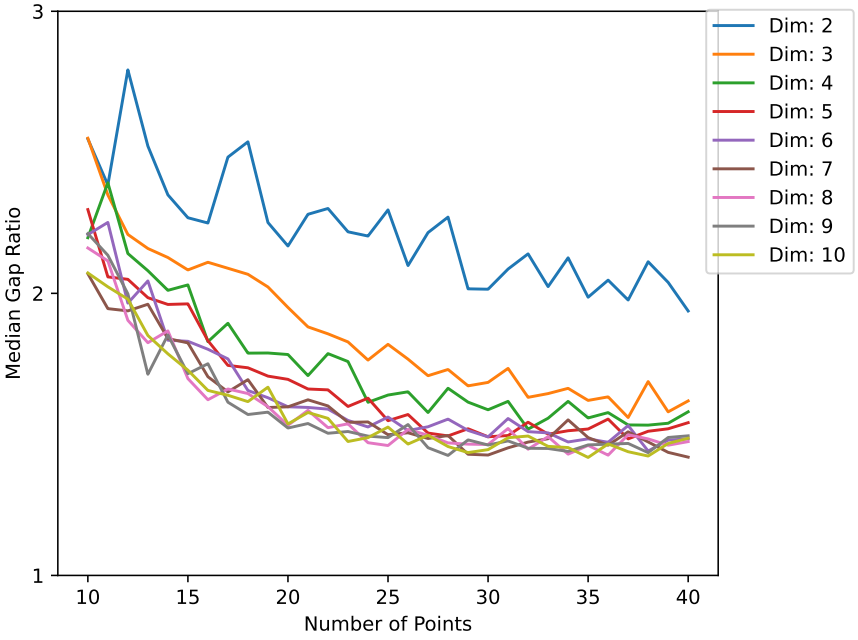

Fig. 6 The median gap ratio of a point set with at least two 1D persistent features, as the set size varies from $n = 10$ to $n = 40$ and the dimension N varies from $N = 2$ to $N = 10$.

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 $N = 2, \ldots, 10$ and point set sizes $n = 10, \ldots, 40$.

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 that are close to the diagonal.

6 Conclusions and discussion of other invariants

Main Theorem 4.3 showed how one can add an arbitrarily long tails to an existing point set without affecting the 1-dimensional persistent homology. Corollary 4.4 implies that families of sets with trivial 1D persistence form vast continuous subspaces in the space of isometry classes of finite sets. The bottleneck distance between persistence diagrams vanishes on these subspaces and cannot have a lower bound. We conjecture that any higher-dimensional persistence (not only 1D persistence as in Theorem 4.3) is preserved when adding tails under the same conditions. We plan further experiments to check how well the bottleneck distance separates point clouds from their perturbations.

The new results motivate a comparison of persistent homology with other isometry invariants of finite point sets. For finite sets of m labeled points, a complete isometry invariant is a classical distance $m \times m$ matrix whose brute-force adaptation to unlabelled points requires $m!$ permutations. The simpler collection of $\frac{m(m-1)}{2}$ pairwise distances (with repetitions) between m unlabeled points is complete for sets in general position [21] but do not distinguish infinitely many non-isometric m-point sets for $m \geq 4$.

The local distribution of distances [22] was recently studied under the name of the Pointwise Distance Distribution (PDD) for finite and periodic sets [23]. The completeness of PDD finite sets in general position is easy for finite sets in $\mathbb{R}^N$ [24, Theorem 16] and was recently extended to the much harder periodic case [23, Theorem 4.4]. The PDD is conjectured to be complete for $N = 2$ but cannot distinguish counter-examples [25] for $N = 3$, which were classified by higher order invariants in appendix C of the first version of [23] in 2021.

Another advantage of PDD is its new near-linear time based on recent algorithms for nearest neighbours [26], which also corrected gaps in past proofs for cover trees [27]. The actual speed is so fast that more than 200 billion pairwise comparisons of all 660K+ periodic crystals in the world's largest database of real materials were complete within two days on a modest desktop. This experiment detected physically impossible isometric duplicates whose underlying publications are investigated by five journals for data integrity [24, section 7].

More importantly, the above experiment justified the Crystal Isometry Principle saying that all real periodic crystals have unique locations determined by their DNA-style code or materials genome in a common Crystal Isometry Space (CRISP) continuously parameterised by complete isometry invariants. Even if examples of periodic sets with the same PDD emerge, the slower isoset invariant is provably complete [28] and has continuous metrics [29].

Returning to finite sets of m unlabeled points, [30, Theorems 4.6, 6.3, 6.6] justify polynomial times (in a fixed dimension) for complete isometry invariants and their continuous metrics, e.g. $O(m^{1.5}(2\log m)^N)$ in general position.

We thank all reviewers for their valuable time and helpful suggestions.

Acknowledgments. This work was supported by the £3.5M EPSRC grant 'Application-driven Topological Data Analysis' (2018-2023), the last author's Royal Academy of Engineering Fellowship 'Data Science for Next Generation Engineering of Solid Crystalline Materials' (2021-2023), and the EPSRC New Horizons grant 'Inverse design of periodic crystals' (2022-2024).

Declaration. The authors state that there is no conflict of interest.

References

[1] Barannikov, S.: The framed Morse complex and its invariants. Advances in Soviet Mathematics **21**, 93–116 (1994)

[2] Frosini, P., Landi, C.: Size theory as a topological tool for computer vision. Pattern Recognition and Image Analysis **9**(4), 596–603 (1999)

[3] Robins, V.: Towards computing homology from finite approximations. In: Topology Proceedings, vol. 24, pp. 503–532 (1999)

[4] Edelsbrunner, H., Letscher, D., Zomorodian, A.: Topological persistence and simplification. In: Proceedings 41st Annual Symposium on Foundations of Computer Science, pp. 454–463 (2000)

[5] Carlsson, G.: Topology and data. Bulletin of the American Mathematical Society **46**(2), 255–308 (2009)

[6] Ghrist, R.: Barcodes: the persistent topology of data. Bulletin of the American Mathematical Society **45**(1), 61–75 (2008)

[7] Weinberger, S.: What is... persistent homology? Notices of the AMS **58**(1), 36–39 (2011)

[8] Chazal, F., De Silva, V., Glisse, M., Oudot, S.: The structure and stability of persistence modules. Springer (2016)

[9] Edelsbrunner, H., Harer, J.: Persistent homology - a survey. Discrete & Computational Geometry **453** (2008)

[10] Curry, J.: The fiber of the persistence map for functions on the interval. Journal of Applied and Computational Topology **2**(3), 301–321 (2018)

[11] 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)

[12] Cohen-Steiner, D., Edelsbrunner, H., Harer, J.: Stability of persistence diagrams. Discrete & Computational Geometry **37**, 263–271 (2005)

[13] Edelsbrunner, H., Heiss, T., Kurlin, V., Smith, P., Wintraecken, M.: The density fingerprint of a periodic point set. In: Proceedings of SoCG, pp. 32–13216 (2021)

[14] 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)

[15] Delaunay, B.: Sur la sphere vide. Izv. Akad. Nauk USSR **7**, 793–800 (1934)

[16] Biswas, R., Cultrera di Montesano, S., Edelsbrunner, H., Saghafian, M.: Continuous and discrete radius functions on Voronoi tessellations and

Delaunay mosaics. Discrete & Computational Geometry, 1–32 (2022)

[17] Hatcher, A.: Algebraic topology. Cambridge University Press (2002)

[18] Bauer, U.: Ripser: efficient computation of vietoris-rips persistence barcodes. J Applied and Computational Geometry **5**(3), 391–423 (2021)

[19] Bobrowski, O., Skraba, P.: On the universality of random persistence diagrams. arXiv:2207.03926 (2022)

[20] Smith, P., Kurlin, V.: Skeletonisation algorithms with theoretical guarantees for unorganised point clouds with high levels of noise. Pattern Recognition **115**, 107902 (2021)

[21] Boutin, M., Kemper, G.: On reconstructing n-point configurations from the distribution of distances or areas. Advances in Applied Mathematics **32**(4), 709–735 (2004)

[22] Mémoli, F.: Gromov–Wasserstein distances and the metric approach to object matching. Found. Comp. Mathematics **11**, 417–487 (2011)

[23] Widdowson, D., Kurlin, V.: Resolving the data ambiguity for periodic crystals. In: Proceedings of NeurIPS (arXiv:2108.04798) (2022)

[24] 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)

[25] Pozdnyakov, S., Willatt, M., Bartók, A., Ortner, C., Csányi, G., Ceriotti, M.: Incompleteness of atomic structure representations. Phys. Rev. Lett. **125**, 166001 (2020)

[26] Elkin, Y.: New compressed cover tree for k-nearest neighbor search (phd thesis). arxiv:2205.10194 (2022)

[27] Elkin, Y., Kurlin, V.: Counterexamples expose gaps in the proof of time complexity for cover trees introduced in 2006. (2022)

[28] Anosova, O., Kurlin, V.: An isometry classification of periodic point sets. In: Discrete Geometry and Mathematical Morphology (2021)

[29] Anosova, O., Kurlin, V.: Algorithms for continuous metrics on periodic crystals. arxiv:2205.15298 (2022)

[30] Kurlin, V.: A computable and continuous metric on isometry classes of high-dimensional periodic sequences. arxiv:2205.04388 (2022)