

Simplexwise Distance Distributions for finite spaces with metrics and measures

Vitaliy Kurlin

Computer Science department
University of Liverpool, UK

vitaliy.kurlin@gmail.com

Abstract

A finite set of unlabelled points in Euclidean space is the simplest representation of many real objects from mineral rocks to sculptures. Since most solid objects are rigid, their natural equivalence is rigid motion or isometry maintaining all inter-point distances. More generally, any finite metric space is an example of a metric-measure space that has a probability measure and a metric satisfying all axioms.

This paper develops Simplexwise Distance Distributions (SDDs) for any finite metric spaces and metric-measures spaces. These SDDs classify all known non-equivalent spaces that were impossible to distinguish by simpler invariants. We define metrics on SDDs that are Lipschitz continuous and allow exact computations whose parametrised complexities are polynomial in the number of given points.

1. Motivations for classifying metric spaces

The simplest representation of any rigid object such a car or a sculpture is a finite set (cloud) $S \subset \mathbb{R}^n$ of m unlabelled points, where $n = 2, 3$ are the most practical dimensions.

The rigidity of many solid objects motivates us to study them up to rigid motion, which is a composition of translations and rotations in Euclidean space \mathbb{R}^n . We can consider any finite set X with a metric that is a distance function $d_X : X \times X \rightarrow [0, +\infty)$ satisfying all metric axioms. The natural equivalence relation on metric spaces is an isometry that is any map $f : X \rightarrow Y$ maintaining all inter-point distances so that $d_X(p, q) = d_Y(f(p), f(q))$ for $p, q \in X$.

In \mathbb{R}^n , any isometry is a composition of a mirror reflection with some rigid motion. Any orientation-preserving isometry can be realised as a continuous rigid motion.

The shape of a rigid object is mathematically defined as its isometry class. Any non-rigid deformation defines a weaker equivalence (than isometry) with a smaller space of flexible shapes. Comparing shapes up to isometry requires finer invariants to distinguish many more isometry classes.

The mathematical approach to distinguish spaces up to isometry uses invariants that are properties preserved by any isometry. Any invariant I maps all isometric spaces to the same value, hence has no false negatives that are pairs of isometric spaces $S \cong Q$ with $I(S) \neq I(Q)$.

A complete invariant I should distinguish all non-isometric clouds, so if $S \not\cong Q$ then $I(S) \neq I(Q)$. Equivalently, I has no false positives that are pairs of non-isometric spaces $S \not\cong Q$ with $I(S) = I(Q)$. Then I is a DNA-style code that uniquely identifies any space S up to isometry.

Since real data are always noisy, a useful complete invariant must be also continuous under the movement of points. Satisfying both completeness and continuity is extremely challenging for sets of m unlabelled points because of $m!$ potential permutations that match all m points.

A complete and continuous invariant for $m = 3$ points consists of three pairwise distances a, b, c (sides of a triangle) and is known in school as the SSS theorem [69] about the congruence (isometry) of triangles. As a result, the isometry space of 3-point sets is continuously mapped as a quadrangular cone $\{0 < a \leq b \leq c \leq a + b\}$ parametrised by $a \leq b \leq c$ satisfying one triangle inequality $a + b \geq c$.

The full description above had no easy analogue for $m \geq 4$ points in \mathbb{R}^n . One obstacle was a family of 4-point sets in \mathbb{R}^2 that have the same six pairwise distances, see Fig. 1.

Problem 1.1 (complete isometry invariants with computable continuous metrics). *Design an invariant I of finite metric spaces satisfying the following properties:*

- (a) completeness : S, Q are isometric $\Leftrightarrow I(S) = I(Q)$;
- (b) Lipschitz continuity : if any point of C is perturbed within its ε -neighbourhood then $I(S)$ changes by at most $\lambda\varepsilon$ for a constant λ and a metric d satisfying all axioms:
 - (1) $d(I(S), I(Q)) = 0$ if and only if $S \cong Q$ are isometric,
 - (2) symmetry : $d(I(S), I(Q)) = d(I(Q), I(S))$,
 - (3) $d(I(S), I(Q)) + d(I(Q), I(T)) \geq d(I(S), I(T))$;
- (c) computability : $I(S)$ and d are computable in a polynomial time in the number m of points in given spaces.

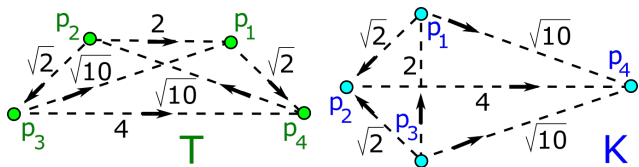


Figure 1. **Left:** trapezium $T = \{(1, 1), (-1, 1), (-2, 0), (2, 0)\}$. **Right:** kite $K = \{(0, 1), (-1, 0), (0, -1), (3, 0)\}$. Both T and K have the same six pairwise distances $\sqrt{2}, \sqrt{2}, 2, \sqrt{10}, \sqrt{10}, 4$.

Condition (1.1b) asking for a continuous metric is stronger than the completeness in (1.1a). Detecting an isometry $S \cong Q$ gives a discontinuous metric, say $d = 1$ for all non-isometric $S \not\cong Q$ even if S, Q are nearly identical. Any metric d satisfying the first axiom in (1.1b) detects an isometry $S \cong Q$ by checking if $d(I(S), I(Q)) = 0$.

Problem 1.1 was open at least since 1974 when Gilbert and Shepp [29] described a 4-parameter family of 4-point sets in \mathbb{R}^2 that have the same six pairwise distances. Problem 1.1 is also motivated by the weaknesses [26, 27, 63] of persistent homology in Topological Data Analysis.

Section 2 reviews the closely related work on invariants of point sets and more general metric spaces. Section 3 introduces the Simplexwise Distance Distributions (SDDs), which substantially generalise all past distance-based invariants of finite metric spaces. Section 4 shows that SDDs are simple enough for manual computations and classifying infinite families of clouds that cannot be distinguished by simpler distance distributions. Section 5 develops Lipschitz continuous metrics on SDDs that are computed in a parametrized polynomial time in the number m of points.

We consider Problem 1.1 a first important step toward understanding moduli spaces of any data objects. Metric spaces and isometry can be replaced by other data and equivalence, respectively, to get analogues of Problem 1.1. This paper extends section 3 of [73], whose 8-page version without proofs and big examples will appear soon. In the papers [71–73] the first author implemented all algorithms, the second author designed all theory, proofs, and examples.

2. Past work on isometries and metric spaces

This section reviews the related work starting from a simpler version of Problem 1.1 asking only to detect a potential isometry between clouds of m unlabelled points

Isometry detection refers to a simpler version of Problem 1.1 to algorithmically detect a potential isometry between given clouds of m points in \mathbb{R}^n . The best algorithm by Brass and Knauer [10] takes $O(m^{\lceil n/3 \rceil} \log m)$ time, so $O(m \log m)$ in \mathbb{R}^3 [11]. The latest advance is the $O(m \log m)$ algorithm in \mathbb{R}^4 [39]. These algorithms output a binary answer (yes/no) without quantifying similarity between non-isometric clouds by a continuous metric.

Multidimensional scaling (MDS). For a given $m \times m$ distance matrix of any m -point cloud A , MDS [61] finds an embedding $A \subset \mathbb{R}^k$ (if it exists) preserving all distances of M for a dimension $k \leq m$. A final embedding $A \subset \mathbb{R}^k$ uses eigenvectors whose ambiguity up to signs gives an exponential comparison time that can be close to $O(2^m)$.

The Heat Kernel Signature (HKS) is a complete isometry invariant of a manifold M whose the Laplace-Beltrami operator has distinct eigenvalues by [66, Theorem 1]. If M is sampled by points, HKS can be discretized and remains continuous [66, section 4] but the completeness is unclear.

The Hausdorff distance [35] can be defined for any subsets A, B in an ambient metric space as $d_H(A, B) = \max\{\vec{d}_H(A, B), \vec{d}_H(B, A)\}$, where the directed Hausdorff distance is $\vec{d}_H(A, B) = \sup_{p \in A} \inf_{q \in B} |p - q|$. To take into account isometries, one can minimize the Hausdorff distance over all isometries [17, 19, 37]. For $n = 2$, the Hausdorff distance minimized over isometries in \mathbb{R}^2 for sets of at most m point needs $O(m^5 \log m)$ time [18]. For a given $\varepsilon > 0$ and $n > 2$, the related problem to decide if $d_H \leq \varepsilon$ up to translations has the time complexity $O(m^{\lceil (n+1)/2 \rceil})$ [70, Chapter 4, Corollary 6]. For general isometry, only approximate algorithms tackled minimizations for infinitely many rotations initially in \mathbb{R}^3 [31] and in \mathbb{R}^n [4, Lemma 5.5].

The Gromov-Wasserstein distances can be defined for metric-measure spaces, not necessarily sitting in a common ambient space. The simplest Gromov-Hausdorff (GH) distance cannot be approximated with any factor less than 3 in polynomial time unless $P = NP$ [60, Corollary 3.8]. Polynomial-time algorithms for GH were designed for ultrametric spaces [50]. However, GH spaces are challenging even for finite point sets in the line \mathbb{R} , see [46] and [76].

The Gromov-Wasserstein distances can be defined for metric-measure spaces, not necessarily sitting in a common ambient space. The simplest Gromov-Hausdorff (GH) distance cannot be approximated with any factor less than 3 in polynomial time unless $P = NP$ [60, Corollary 3.8]. Polynomial-time algorithms for GH were designed for ultrametric spaces [50]. However, GH spaces are challenging even for finite point sets in the line \mathbb{R} , see [46] and [76].

Experimental approaches cover a wide variety of descriptors designed manually or optimised through machine learning, for example, Scale Invariant Feature Transform [57, 65, 67, 78]. Some of these descriptors are designed for invariance under permutations of points [56, 75], and also for invariance under isometry [16, 52, 62], for example, in Geometric Deep Learning [14, 15]. Among many obstacles [1, 21, 22, 34, 45], the hard one is to theoretically guarantee the completeness and Lipschitz continuity of such descriptors under perturbations as in Problem 1.1.

Local distributions of distances in Mémoli’s seminal work [48, 49] for metric-measure spaces, or shape distributions [8, 32, 47, 53, 54], are first-order versions of the new SDDs. Another approach to Problem 1.1 uses direction-based invariants [41], which inspired Complete Neural Networks [36]. The Lipschitz continuity was proved [41, Theorem 4.9] in general position but not for near-singular configurations, for example, when a triangle degenerates to a line. These degeneracies will be addressed in the forthcoming

ing work [44] extending SDD to a complete invariant of clouds in \mathbb{R}^n .

3. Simplexwise Distance Distribution (SDD)

This section introduces the isometry invariants SDD for a finite cloud C of unlabelled points in any metric space X . The *lexicographic* order $u < v$ on vectors $u = (u_1, \dots, u_h)$ and $v = (v_1, \dots, v_h)$ means that if the first i (possibly, $i = 0$) coordinates of u, v coincide then $u_{i+1} < v_{i+1}$. Let S_h denote the permutation group on indices $1, \dots, h$.

Definition 3.1 (RDD($C; A$)). *Let C be a cloud of m unlabelled points in a space with a metric d . A basis sequence $A = (p_1, \dots, p_h) \in C^h$ consists of $1 \leq h < m$ distinct points. Let $D(A)$ be the triangular distance matrix whose entry $D(A)_{i,j-1}$ is $d(p_i, p_j)$ for $1 \leq i < j \leq h$, all other entries are filled by zeros. Any permutation $\xi \in S_h$ acts on $D(A)$ by mapping $D(A)_{ij}$ to $D(A)_{kl}$, where $k \leq l$ is the pair of indices $\xi(i), \xi(j) - 1$ written in increasing order.*

For any other point $q \in C - A$, write distances from q to p_1, \dots, p_h as a column. The $h \times (m - h)$ -matrix $R(C; A)$ is formed by these $m - h$ lexicographically ordered columns. The action of ξ on $R(C; A)$ maps any i -th row to the $\xi(i)$ -th row, after which all columns can be written again in the lexicographic order. The Relative Distance Distribution RDD($C; A$) is the equivalence class of the pair $[D(A), R(C; A)]$ of matrices up to permutations $\xi \in S_h$.

For $h = 1$ and a basis sequence $A = (p_1)$, the matrix $D(A)$ is empty and $R(C; A)$ is a single row of distances (in the increasing order) from p_1 to all other points $q \in C$. For $h = 2$ and a basis sequence $A = (p_1, p_2)$, the matrix $D(A)$ is the single number $d(p_1, p_2)$ and $R(C; A)$ consists of two rows of distances from p_1, p_2 to all other points $q \in C$.

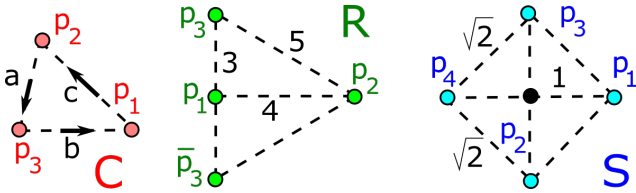


Figure 2. **Left:** a cloud $C = \{p_1, p_2, p_3\}$ with distances $a \leq b \leq c$. **Middle:** the triangular cloud $R = \{(0, 0), (4, 0), (0, 3)\}$. **Right:** the square cloud $S = \{(1, 0), (-1, 0), (0, 1), (-1, 0)\}$.

Example 3.2 (RDD for a 3-point cloud C). *Let $C \subset \mathbb{R}^2$ consist of p_1, p_2, p_3 with inter-point distances $a \leq b \leq c$ ordered counter-clockwise as in Fig. 2 (left). Then*

$$\text{RDD}(C; p_1) = [\emptyset; (b, c)], \text{RDD}(C; \begin{pmatrix} p_2 \\ p_3 \end{pmatrix}) = [a; \begin{pmatrix} c \\ b \end{pmatrix}],$$

$$\text{RDD}(C; p_2) = [\emptyset; (a, c)], \text{RDD}(C; \begin{pmatrix} p_3 \\ p_1 \end{pmatrix}) = [b; \begin{pmatrix} a \\ c \end{pmatrix}],$$

$$\text{RDD}(C; p_3) = [\emptyset; (a, b)], \text{RDD}(C; \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}) = [c; \begin{pmatrix} b \\ a \end{pmatrix}].$$

We have written $\text{RDD}(C; A)$ for a basis sequence $A = (p_i, p_j)$ of ordered points represented by a column. Swapping the points $p_1 \leftrightarrow p_2$ makes the last RDD above equivalent to $\text{RDD}(C; \begin{pmatrix} p_2 \\ p_1 \end{pmatrix}) = [c; \begin{pmatrix} a \\ b \end{pmatrix}]$.

Though $\text{RDD}(C; A)$ is defined up to a permutation $\xi \in S_h$ of h points in $A \in C^h$, comparisons of RDDs will be practical for $h = 2, 3$ with metrics independent of ξ .

Definition 3.3 (Simplexwise Distance Distribution SDD($C; h$)). *Let C be a cloud of m unlabelled points in a metric space. For an integer $1 \leq h < m$, the Simplexwise Distance Distribution SDD($C; h$) is the unordered set of RDD($C; A$) for all unordered h -point subsets $A \subset C$.*

For $h = 1$ and any m -point cloud C , the distribution SDD($C; 1$) can be considered as a matrix of m rows of ordered distances from every point $p \in C$ to all other $m - 1$ points. If we lexicographically order these m rows and collapse any $l > 1$ identical rows into a single one with the weight l/m , then we get the Pointwise Distance Distribution PDD($C; m - 1$) introduced in [72, Definition 3.1].

The PDD was simplified to the easier-to-compare vector of Average Minimum Distances [74]: $\text{AMD}_k(C) = \frac{1}{m} \sum_{i=1}^m d_{ik}$, where d_{ik} is the distance from a point $p_i \in C$ to its k -th nearest neighbor in C . These neighbor-based invariants can be computed in a near-linear time in m [25] and were pairwise compared for all 660K+ periodic crystals in the world's largest database of real materials [72]. Definition 3.4 similarly maps SDD to a smaller invariant.

Recall that the 1st moment of a set of numbers a_1, \dots, a_k is the average $\mu = \frac{1}{k} \sum_{i=1}^k a_i$. The 2nd moment is the standard deviation $\sigma = \sqrt{\frac{1}{k} \sum_{i=1}^k (a_i - \mu)^2}$. For $l \geq 3$, the l -th standardized moment [38, section 2.7] is $\frac{1}{k} \sum_{i=1}^k \left(\frac{a_i - \mu}{\sigma} \right)^l$.

Definition 3.4 (Simplexwise Distance Moments SDM). *For any m -point cloud C in a metric space, let $A \subset C$ be a subset of h unordered points. The Sorted Distance Vector SDV(A) is the list of all $\frac{h(h-1)}{2}$ pairwise distances between points of A written in increasing order. The vector $\vec{R}(C; A) \in \mathbb{R}^{m-h}$ is obtained from the $h \times (m - h)$ matrix $R(C; A)$ in Definition 3.1 by writing the vector of $m - h$ column averages in increasing order.*

The pair $[\text{SDV}(A); \vec{R}(C; A)]$ is the Average Distance Distribution ADD($C; A$) considered as a vector of length $\frac{h(h-3)}{2} + m$. The unordered collection of ADD($C; A$) for

RDD($T; A$) in SDD($T; 2$) $[\sqrt{2}, \begin{pmatrix} 2 & \sqrt{10} \\ \sqrt{10} & 4 \end{pmatrix}] \times 2$ $[2, \begin{pmatrix} \sqrt{2} & \sqrt{10} \\ \sqrt{10} & \sqrt{2} \end{pmatrix}]$ $[\sqrt{10}, \begin{pmatrix} \sqrt{2} & 4 \\ 2 & \sqrt{2} \end{pmatrix}] \times 2$ $[4, \begin{pmatrix} \sqrt{2} & \sqrt{10} \\ \sqrt{10} & \sqrt{2} \end{pmatrix}]$	RDD($K; A$) in SDD($K; 2$) $[\sqrt{2}, \begin{pmatrix} 2 & \sqrt{10} \\ \sqrt{2} & 4 \end{pmatrix}] \times 2$ $[2, \begin{pmatrix} \sqrt{2} & \sqrt{10} \\ \sqrt{2} & \sqrt{10} \end{pmatrix}]$ $[\sqrt{10}, \begin{pmatrix} \sqrt{2} & 2 \\ 4 & \sqrt{10} \end{pmatrix}] \times 2$ $[4, \begin{pmatrix} \sqrt{2} & \sqrt{2} \\ \sqrt{10} & \sqrt{10} \end{pmatrix}]$
ADD($T; A$) in ASD($T; 2$) $[\sqrt{2}, (\frac{2+\sqrt{10}}{2}, \frac{4+\sqrt{10}}{2})] \times 2$ $[2, (\frac{\sqrt{2}+\sqrt{10}}{2}, \frac{\sqrt{2}+\sqrt{10}}{2})]$ $[\sqrt{10}, (\frac{2+\sqrt{2}}{2}, \frac{4+\sqrt{2}}{2})] \times 2$ $[4, (\frac{\sqrt{2}+\sqrt{10}}{2}, \frac{\sqrt{2}+\sqrt{10}}{2})]$	ADD($K; A$) in ASD($K; 2$) $[\sqrt{2}, (\frac{2+\sqrt{2}}{2}, \frac{4+\sqrt{10}}{2})] \times 2$ $[2, (\sqrt{2}, \sqrt{10})]$ $[\sqrt{10}, (\frac{2+\sqrt{10}}{2}, \frac{4+\sqrt{2}}{2})] \times 2$ $[4, (\frac{\sqrt{2}+\sqrt{10}}{2}, \frac{\sqrt{2}+\sqrt{10}}{2})]$
SDM ₁ = $\frac{3 + \sqrt{2} + \sqrt{10}}{3}$ SDM ₂ = $\frac{6 + 2\sqrt{2} + 4\sqrt{10}}{12}$ SDM ₃ = $\frac{16 + 4\sqrt{2} + 4\sqrt{10}}{12}$	SDM ₁ = $\frac{3 + \sqrt{2} + \sqrt{10}}{3}$ SDM ₂ = $\frac{8 + 5\sqrt{2} + 3\sqrt{10}}{12}$ SDM ₃ = $\frac{16 + 3\sqrt{2} + 5\sqrt{10}}{12}$

Table 1. The Simplexwise Distance Distributions from Definition 3.3 for the 4-point clouds $T, K \subset \mathbb{R}^2$ in Fig. 1. The symbol $\times 2$ indicates a doubled RDD. The three bottom rows show coordinates of $\text{SDM}(C; 2, 1) \in \mathbb{R}^3$ from Definition 3.4 for $h = 2$, $l = 1$ and both $C = T, K$. Different elements are highlighted.

all $\binom{m}{h}$ unordered subsets $A \subset C$ is the Average Simplexwise Distribution $\text{ASD}(C; h)$. The Simplexwise Distance Moment $\text{SDM}(C; h, l)$ is the l -th (standardized for $l \geq 3$) moment of $\text{ASD}(C; h)$ considered as a probability distribution of $\binom{m}{h}$ vectors, separately for each coordinate.

Example 3.5 (SDD and SDM for T, K). Fig. 1 shows the non-isometric 4-point clouds T, K with the same Ordered Pairwise Distances: $\text{SDV} = \{\sqrt{2}, \sqrt{2}, 2, \sqrt{10}, \sqrt{10}, 4\}$, see infinitely many examples in [9]. The arrows on the edges of T, K show orders of points in each pair of vertices for RDDs. Then T, K are distinguished up to isometry by $\text{SDD}(T; 2) \neq \text{SDD}(K; 2)$ in Table 1. The 1st coordinate of $\text{SDM}(C; 2, 1) \in \mathbb{R}^3$ is the average of the six distances from SDV (the same for T, K) but the other two coordinates (column averages from $R(C; A)$ matrices) differ.

Some of the $\binom{m}{h}$ RDDs in $\text{SDD}(C; h)$ can coincide as in Example 3.5. If we collapse any $l > 1$ identical RDDs into

a single RDD with the weight $l/\binom{m}{h}$, SDD can be considered as a weighted probability distribution of RDDs.

All time complexities are proved for a random-access machine (RAM) model. In a general metric space, a point cloud C is usually given by a distance matrix on (arbitrarily ordered) points of C . Hence we assume that the distance between any points of C is accessible in a constant time.

Theorem 3.6 (invariance and time of SDDs). For $h \geq 1$ and any cloud C of m unlabelled points in a metric space, $\text{SDD}(C; h)$ is an isometry invariant, which can be computed in time $O(m^{h+1}/(h-1)!)$. For any $l \geq 1$, the invariant $\text{SDM}(C; h, l) \in \mathbb{R}^{m + \frac{h(h-3)}{2}}$ has the same time.

Proof. Any isometry $S \rightarrow Q$ preserves distances, hence induces a bijection $\text{SDD}(S; h) \rightarrow \text{SDD}(Q; h)$ for $h \geq 1$.

By Definition 3.3, for any $h \geq 1$ and a cloud C of m unlabelled points in a metric space, the Simplexwise Distance Distribution $\text{SDD}(C; h)$ consists of $\binom{m}{h} = \frac{m!}{h!(m-h)!} = O(m^h/h!)$ Relative Distance Distributions $\text{RDD}(C; A)$ for any unordered subset $A \subset C$ of h points.

For any order of points of A , every $\text{RDD}(C; A)$ consists of the distance matrix $D(A)$, which needs $O(h^2)$ time and $h \times (m-h)$ matrix $R(C; A)$, which needs $h(m-h)$ time. Since $h \leq m$, the extra factor $O(hm)$ gives the final time $O(m^{h+1}/(h-1)!)$ for $\text{SDD}(C; h)$.

For a fixed h -point subset $A \subset C$, the vector $\vec{R}(C; A)$ from Definition 3.4 needs $O(hm)$ time to average h distances in $m-h$ columns and $O(m \log m)$ time to order these averages. The list $\text{SDV}(A)$ of Ordered Pairwise Distances is obtained by sorting all pairwise distances from $D(A)$ in time $O(h^2 \log h)$. So the Average Distance Distribution $\text{ADD}(C; A)$ obtained by concatenating the ordered vectors $\text{SDV}(A) \in \mathbb{R}^{\frac{h(h-1)}{2}}$ and $\vec{R}(C; A) \in \mathbb{R}^{m-h}$ requires only $O((h^2 + m) \log m)$ extra time. Hence the Average Simplexwise Distribution $\text{ASD}(C; h)$ for all h -point subsets $A \subset C$ needs $O(m^{h+1}/(h-1)!)$ time including $O((h^2 + m) \log m)$, the same as the initial $\text{SDD}(C; h)$.

For $l = 1$, the first raw moment $\text{SDM}(C; h, 1)$ is the simple average of all $k = \binom{m}{h}$ vectors $\text{ADD}(C; A)$ of length $O(hm)$, hence needs $O(m^{h+1}/(h-1)!)$ time. For $l = 2$, the standard deviation σ of each coordinate in all vectors $\text{ADD}(C; A)$ requires the same time. Then, for any fixed $l \geq 3$, the l -th standardized moment $\frac{1}{k} \sum_{i=1}^k \left(\frac{a_i - \mu}{\sigma} \right)^l$ needs again the same time $O(m^{h+1}/(h-1)!)$. \square

We conjecture that $\text{SDD}(C; h)$ is a complete isometry invariant of a cloud $C \subset \mathbb{R}^n$ for some $h \leq n$. Section 4 shows that $\text{SDD}(C; 2)$ distinguishes all infinitely many known pairs [55, Fig. S4] of non-isometric m -point sets $S, Q \subset \mathbb{R}^3$ that have equal $\text{PDD}(S) = \text{PDD}(Q)$

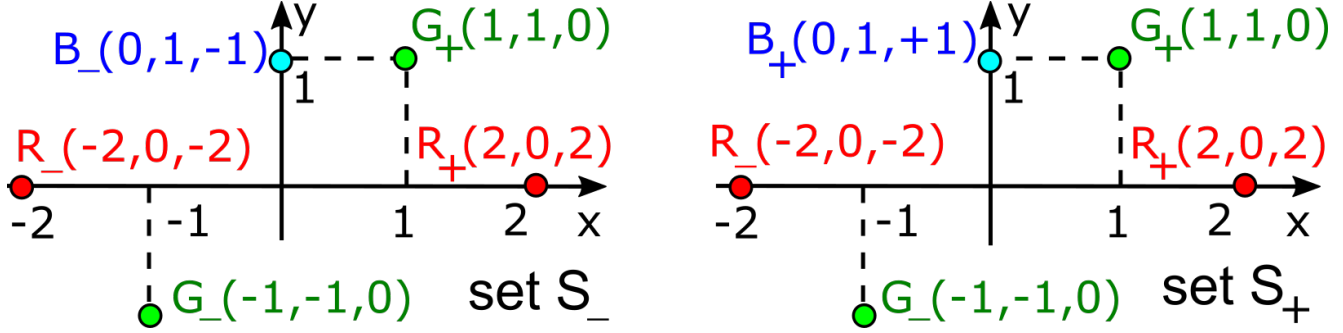


Figure 3. **Left:** (x, y) -projection of the 5-point set $S_- \subset \mathbb{R}^3$ consisting of the green points $G_- = (-1, -1, 0)$ and $G_+ = (1, 1, 0)$, the red points $R_- = (-2, 0, -2)$ and $R_+ = (2, 0, 2)$, and the blue point $B_- = (0, 1, -1)$. **Right:** to get the set $S_+ \subset \mathbb{R}^3$ from S_- , replace the point B_- with the new point $B_+ = (0, 1, 1)$.

4. The strength of the isometry invariant SDD

Examples 4.1 and 4.2 distinguish clouds of 5 points and 7 points, respectively, in \mathbb{R}^3 by comparing their SDDs of order 2. In Example 4.3, $\text{SDD}(C; 2)$ distinguishes 6-point clouds in a family of pairs depending on three parameters.

Example 4.1 (5-point clouds). *Fig. 3 shows the 5-point clouds $S_{\pm} \subset \mathbb{R}^3$ taken from [55, Figure S4(A)].*

distances of S_-	R_-	R_+	G_-	G_+	B_-
$R_-(-2, 0, -2)$	0	$\sqrt{32}$	$\sqrt{6}$	$\sqrt{14}$	$\sqrt{6}$
$R_+(+2, 0, +2)$	$\sqrt{32}$	0	$\sqrt{14}$	$\sqrt{6}$	$\sqrt{14}$
$G_-(-1, -1, 0)$	$\sqrt{6}$	$\sqrt{14}$	0	$\sqrt{8}$	$\sqrt{6}$
$G_+(+1, +1, 0)$	$\sqrt{14}$	$\sqrt{6}$	$\sqrt{8}$	0	$\sqrt{2}$
$B_-(0, +1, -1)$	$\sqrt{6}$	$\sqrt{14}$	$\sqrt{6}$	$\sqrt{2}$	0

distances of S_+	R_-	R_+	G_-	G_+	B_+
$R_-(-2, 0, -2)$	0	$\sqrt{32}$	$\sqrt{6}$	$\sqrt{14}$	$\sqrt{14}$
$R_+(+2, 0, +2)$	$\sqrt{32}$	0	$\sqrt{14}$	$\sqrt{6}$	$\sqrt{6}$
$G_-(-1, -1, 0)$	$\sqrt{6}$	$\sqrt{14}$	0	$\sqrt{8}$	$\sqrt{6}$
$G_+(+1, +1, 0)$	$\sqrt{14}$	$\sqrt{6}$	$\sqrt{8}$	0	$\sqrt{2}$
$B_+(0, +1, +1)$	$\sqrt{14}$	$\sqrt{6}$	$\sqrt{6}$	$\sqrt{2}$	0

Table 2. Distances between all points of the set S_{\mp} in Fig. 3.

The sets S_{\pm} are not isometric, because S_+ has the triple of points B_+, G_+, R_+ with pairwise distances $\sqrt{2}, \sqrt{6}, \sqrt{6}$, but S_- has no such a triple. Table 2 highlights differences between distance matrices. If we order distances to neighbors, the matrices in Table 3 differ only in one pair.

If we ignore labels of points, S_{\pm} have identical Pointwise Distance Distribution (PDD), which is the Simplexwise Distance Distributions (SDD) in Definition 3.3.

For easier visualization, the matrix below is obtained by

S_- distances to	1st neighbor	2nd	3rd	4th
$R_- = (-2, 0, -2)$	$\sqrt{6}$	$\sqrt{6}$	$\sqrt{14}$	$\sqrt{32}$
$R_+ = (+2, 0, +2)$	$\sqrt{6}$	$\sqrt{14}$	$\sqrt{14}$	$\sqrt{32}$
$G_- = (-1, -1, 0)$	$\sqrt{6}$	$\sqrt{6}$	$\sqrt{8}$	$\sqrt{14}$
$G_+ = (+1, +1, 0)$	$\sqrt{2}$	$\sqrt{6}$	$\sqrt{8}$	$\sqrt{14}$
$B_- = (0, +1, -1)$	$\sqrt{2}$	$\sqrt{6}$	$\sqrt{6}$	$\sqrt{14}$

S_+ distances to	1st neighbor	2nd	3rd	4th
$R_- = (-2, 0, -2)$	$\sqrt{6}$	$\sqrt{14}$	$\sqrt{14}$	$\sqrt{32}$
$R_+ = (+2, 0, +2)$	$\sqrt{6}$	$\sqrt{6}$	$\sqrt{14}$	$\sqrt{32}$
$G_- = (-1, -1, 0)$	$\sqrt{6}$	$\sqrt{6}$	$\sqrt{8}$	$\sqrt{14}$
$G_+ = (+1, +1, 0)$	$\sqrt{2}$	$\sqrt{6}$	$\sqrt{8}$	$\sqrt{14}$
$B_+ = (0, +1, -1)$	$\sqrt{2}$	$\sqrt{6}$	$\sqrt{6}$	$\sqrt{14}$

Table 3. For each point from the 5-point set S_+ in Fig. 3, the distances to neighbors from Table 2 are ordered in each row.

lexicographically ordering the rows in Table 3:

$$\text{PDD}(S_{\pm}) = \text{SDD}(S_{\pm}; 1) = \begin{pmatrix} \sqrt{2} & \sqrt{6} & \sqrt{6} & \sqrt{14} \\ \sqrt{2} & \sqrt{6} & \sqrt{8} & \sqrt{14} \\ \sqrt{6} & \sqrt{6} & \sqrt{8} & \sqrt{14} \\ \sqrt{6} & \sqrt{6} & \sqrt{14} & \sqrt{32} \\ \sqrt{6} & \sqrt{14} & \sqrt{14} & \sqrt{32} \end{pmatrix}$$

Now we show that $\text{SDD}(S_-; 2) \neq \text{SDD}(S_+; 2)$. For $h = 2$, the Simplexwise Distance Distribution $\text{SDD}(C; h)$ consists of $\text{RDD}(C; A)$ for 2-point subsets $A \subset C$. Both sets S_{\pm} have a single pair of points (G_+, B_-) and (G_+, B_+) at distance $\sqrt{2}$. Hence it suffices to show that the Relative Distance Distributions differ for this pair:

$$\text{RDD}(S_-, \begin{pmatrix} G_+ \\ B_- \end{pmatrix}) = \left[\sqrt{2}, \begin{pmatrix} \sqrt{8} & \sqrt{14} & \sqrt{6} \\ \sqrt{6} & \sqrt{6} & \sqrt{14} \\ G_- & R_- & R_+ \end{pmatrix} \right]$$

$$\text{RDD}(S_+, \begin{pmatrix} G_+ \\ B_+ \end{pmatrix}) = \left[\sqrt{2}, \begin{pmatrix} \sqrt{8} & \sqrt{14} & \sqrt{6} \\ \sqrt{6} & \sqrt{14} & \sqrt{6} \\ G_- & R_- & R_+ \end{pmatrix} \right]$$

The last rows in the above 3×3 matrices indicate a complementary point $q \in C - A$ for indexing columns of the 2×3 matrices $R(C; A)$ in Definition 3.1. The resulting RDDs differ because any permutation of rows or columns of $R(S_+; \{G_+, B_+\})$ keeps the pair $\sqrt{6}, \sqrt{6}$ in the same column but $R(S_-; \{G_+, B_+\})$ has no pair $\sqrt{6}, \sqrt{6}$ in one column. Hence $\text{SDD}(S_-; 2) \neq \text{SDD}(S_+; 2)$.

Example 4.2 (7-point sets). The sets Q_\pm in Fig. 4 taken from [55, Figure S4(B)] have distances in Table 4. Both sets have only two pairs of points at distance $\sqrt{6}$. Hence it suffices to compare RDDs for these pairs below.

$$R(Q_-; \begin{pmatrix} G \\ B_{+1} \end{pmatrix}) = \begin{pmatrix} \sqrt{32} & \sqrt{14} & \sqrt{17} & 3 & \sqrt{13} \\ \sqrt{6} & \sqrt{8} & \sqrt{5} & 1 & \sqrt{3} \\ R & B_{-1} & B_{-2} & B_{+2} & O_- \end{pmatrix},$$

$$R(Q_-; \begin{pmatrix} R \\ B_{-1} \end{pmatrix}) = \begin{pmatrix} \sqrt{32} & \sqrt{14} & 3 & \sqrt{17} & \sqrt{5} \\ \sqrt{14} & \sqrt{8} & 3 & \sqrt{13} & \sqrt{3} \\ G & B_{+1} & B_{-2} & B_{+2} & O_- \end{pmatrix}$$

The pair above has submatrices $\begin{pmatrix} 3 & \sqrt{13} \\ 1 & \sqrt{3} \end{pmatrix}$ and $\begin{pmatrix} 3 & \sqrt{5} \\ 3 & \sqrt{3} \end{pmatrix}$ but the pair below has no such submatrices.

$$R(Q_+; \begin{pmatrix} G \\ B_{+1} \end{pmatrix}) = \begin{pmatrix} \sqrt{32} & \sqrt{14} & \sqrt{17} & 3 & \sqrt{5} \\ \sqrt{6} & \sqrt{8} & \sqrt{5} & 1 & \sqrt{3} \\ R & B_{-1} & B_{-2} & B_{+2} & O_+ \end{pmatrix},$$

$$R(Q_+; \begin{pmatrix} R \\ B_{-1} \end{pmatrix}) = \begin{pmatrix} \sqrt{32} & \sqrt{14} & 3 & \sqrt{17} & \sqrt{13} \\ \sqrt{14} & \sqrt{8} & 3 & \sqrt{13} & \sqrt{3} \\ G & B_{+1} & B_{-2} & B_{+2} & O_+ \end{pmatrix}$$

The pair of $\text{RDD}(Q_-; \begin{pmatrix} G \\ B_{+1} \end{pmatrix})$ and $\text{RDD}(Q_-; \begin{pmatrix} R \\ B_{-1} \end{pmatrix})$ differs from the pair $\text{RDD}(Q_+; \begin{pmatrix} G \\ B_{+1} \end{pmatrix})$ and $\text{RDD}(Q_+; \begin{pmatrix} R \\ B_{-1} \end{pmatrix})$. Hence $\text{SDD}(Q_-; 2) \neq \text{SDD}(Q_+; 2)$.

Example 4.3 (6-point sets). The sets T_\pm in Fig. 5, which was motivated by [55, Figure S4(C)], have the points R, G, O_\pm from the sets Q_\pm in Example 4.2 and three new points $C_1(x_1, y_1, 0)$, $C_2(x_2, y_2, 0)$, $C_3(x_3, y_3, 0)$ such that $|RC_1| = |GC_2|$, $|RC_2| = |GC_3|$, $|RC_3| = |GC_1|$.

Denote by $2l_1, 2l_2, 2l_3$ the lengths of these three pairs of line segments after their projection to the xy -plane so that

$$(4.3.1) \begin{cases} (x_2 + 2)^2 + y_2^2 = |RC_2|^2 - 4 = (2l_1)^2, \\ (x_3 - 2)^2 + y_3^2 = |GC_3|^2 - 4 = (2l_1)^2; \end{cases}$$

$$(4.3.2) \begin{cases} (x_3 + 2)^2 + y_3^2 = |RC_3|^2 - 4 = (2l_2)^2, \\ (x_1 - 2)^2 + y_1^2 = |GC_1|^2 - 4 = (2l_2)^2; \end{cases}$$

$$(4.3.3) \begin{cases} (x_1 + 2)^2 + y_1^2 = |RC_1|^2 - 4 = (2l_3)^2, \\ (x_2 - 2)^2 + y_2^2 = |GC_2|^2 - 4 = (2l_3)^2. \end{cases}$$

Comparing the first part of (4.3.1) with the second part side of (4.3.3), we get $(2l_1)^2 - 4x_2 = (2l_3)^2 + 4x_2$, so $x_2 = \frac{l_1^2 - l_3^2}{2}$. Similarly, $x_3 = \frac{l_2^2 - l_1^2}{2}$, $x_1 = \frac{l_3^2 - l_2^2}{2}$ so that $x_1 + x_2 + x_3 = 0$. From the second part of (4.3.2), we get $x_1^2 - 4x_1 + 4 + y_1^2 = 4l_2^2$, so

$$|O_\pm C_1|^2 = x_1^2 + y_1^2 + 1 = 4l_2^2 + 4x_1 - 3 = 2l_2^2 + 2l_3^2 - 3,$$

$$\text{similarly } |O_\pm C_2|^2 = 2l_3^2 + 2l_1^2 - 3, \quad |O_\pm C_3|^2 = 2l_1^2 + 2l_2^2 - 3.$$

$$\text{Then } |C_1 C_2|^2 = (x_1 - x_2)^2 + (y_1 - y_2)^2 = x_1^2 + y_1^2.$$

Table 5 contains all pairwise distances between the points of T_\pm . We show that T_\pm differ by the simplified invariants $\widetilde{\text{SDD}}(T_\pm; 2)$ below. In each column of $R(C; A)$, we additionally allow any permutation of elements independent of other columns, so we could order each column (a pair of distances) lexicographically. Denote the resulting simplification of RDD by $\widetilde{\text{RDD}}$. Then $\widetilde{\text{SDD}}(T_\pm; 2)$ have identical $\widetilde{\text{RDD}}$ s for the 2-point subsets A from the list $\{R, G\}, \{O_\pm, C_i\}, \{C_i, C_j\}$ for distinct $i, j = 1, 2, 3$.

For example, both $\widetilde{\text{RDD}}(T_\pm; \{R, G\})$ start with the distance $|R - G| = \sqrt{32}$ and then include the same four pairs $(\sqrt{5}, \sqrt{13}), (2\sqrt{l_i^2 + 1}, 2\sqrt{l_{i-1}^2 + 1})$ for $i \in \{1, 2, 3\}$ modulo 3, which should be ordered and written lexicographically. Hence it makes sense to compare $\widetilde{\text{SDD}}(T_\pm; 2)$ only by the remaining $\widetilde{\text{RDD}}(T_\pm; A)$ for A from the list $\{R, O_\pm\}, \{G, O_\pm\}, \{R, C_i\}, \{G, C_j\}$ in Table 6.

Without loss of generality assume that $l_1 \geq l_2 \geq l_3$. If all the lengths are distinct, then $l_1 > l_2 > l_3$. Then the rows for $\{R, O_-\}$ and $\{G, O_+\}$ differ in Table 6 even after ordering each pair so that a smaller distance precedes a larger one, and after writing all pairs lexicographically. So $\widetilde{\text{SDD}}(T_-; 2) \neq \widetilde{\text{SDD}}(T_+; 2)$ unless two of l_i are equal.

If (say) $l_1 = l_2$, the lexicographically ordered rows of $\{R, O_-\}$ and $\{G, O_+\}$ coincide in $\widetilde{\text{SDD}}(T_\pm; 2)$, similarly for the rows of $\{G, O_-\}$ and $\{R, O_+\}$. Hence it suffices to compare only the six rows for the remaining pairs $\{R, C_i\}, \{G, C_j\}$ in $\widetilde{\text{SDD}}(T_\pm; 2)$.

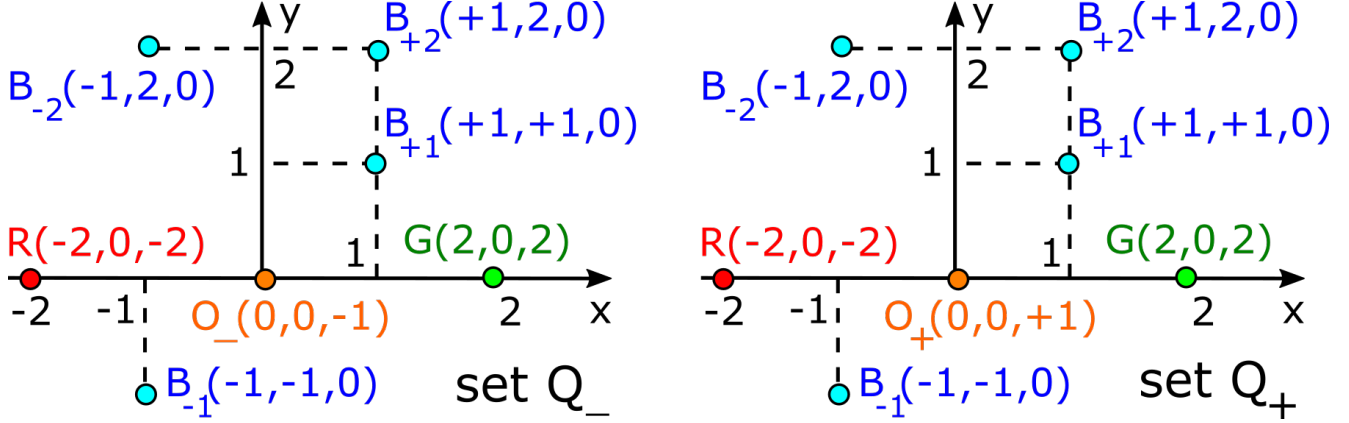


Figure 4. **Left:** (x, y) -projection of the 7-point set $Q_- \subset \mathbb{R}^3$ of the red point $R = (-2, 0, -2)$, green point $G = (2, 0, 2)$, four blue points $B_{\pm 1} = (\pm 1, \pm 1, 0)$, $B_{\pm 2} = (\pm 1, 2, 0)$, orange point $O_- = (0, 0, -1)$. **Right:** to get Q_+ from Q_- , replace O_- with $O_+ = (0, 0, +1)$.

distances of Q_-	R	G	B_{-1}	B_{+1}	B_{-2}	B_{+2}	O_-
$R = (-2, 0, -2)$	0	$\sqrt{32}$	$\sqrt{6}$	$\sqrt{14}$	3	$\sqrt{17}$	$\sqrt{5}$
$G = (+2, 0, +2)$	$\sqrt{32}$	0	$\sqrt{14}$	$\sqrt{6}$	$\sqrt{17}$	3	$\sqrt{13}$
$B_{-1} = (-1, -1, 0)$	$\sqrt{6}$	$\sqrt{14}$	0	$\sqrt{8}$	3	$\sqrt{13}$	$\sqrt{3}$
$B_{+1} = (+1, +1, 0)$	$\sqrt{14}$	$\sqrt{6}$	$\sqrt{8}$	0	$\sqrt{5}$	1	$\sqrt{3}$
$B_{-2} = (-1, 2, 0)$	3	$\sqrt{17}$	3	$\sqrt{5}$	0	2	$\sqrt{6}$
$B_{+2} = (+1, 2, 0)$	$\sqrt{17}$	3	$\sqrt{13}$	1	2	0	$\sqrt{6}$
$O_- = (0, 0, -1)$	$\sqrt{5}$	$\sqrt{13}$	$\sqrt{3}$	$\sqrt{3}$	$\sqrt{6}$	$\sqrt{6}$	0

distances of Q_+	R	G	B_{-1}	B_{+1}	B_{-2}	B_{+2}	O_+
$R = (-2, 0, -2)$	0	$\sqrt{32}$	$\sqrt{6}$	$\sqrt{14}$	3	$\sqrt{17}$	$\sqrt{13}$
$G = (+2, 0, +2)$	$\sqrt{32}$	0	$\sqrt{14}$	$\sqrt{6}$	$\sqrt{17}$	3	$\sqrt{5}$
$B_{-1} = (-1, -1, 0)$	$\sqrt{6}$	$\sqrt{14}$	0	$\sqrt{8}$	3	$\sqrt{13}$	$\sqrt{3}$
$B_{+1} = (+1, +1, 0)$	$\sqrt{14}$	$\sqrt{6}$	$\sqrt{8}$	0	$\sqrt{5}$	1	$\sqrt{3}$
$B_{-2} = (-1, 2, 0)$	3	$\sqrt{17}$	3	$\sqrt{5}$	0	2	$\sqrt{6}$
$B_{+2} = (+1, 2, 0)$	$\sqrt{17}$	3	$\sqrt{13}$	1	2	0	$\sqrt{6}$
$O_+ = (0, 0, +1)$	$\sqrt{13}$	$\sqrt{5}$	$\sqrt{3}$	$\sqrt{3}$	$\sqrt{6}$	$\sqrt{6}$	0

Table 4. The matrices of distances between all points of the 7-point set Q_{\mp} in Fig. 4 taken from [55, Figure S4(B)].

For $l_1 = l_2$, we get $x_3 = \frac{l_2^2 - l_1^2}{2} = 0$ and $x_1 = -x_2 = \frac{l_3^2 - l_2^2}{2}$. In equation (4.3.3) the equality $(x_1 + 2)^2 + y_1^2 = (x_2 - 2)^2 + y_2^2$ with $x_1 = -x_2$ implies that $y_1^2 = y_2^2$. The even more degenerate case $l_1 = l_2 = l_3$, means that $x_1 = x_2 = x_3 = 0$ and $y_1^2 = y_2^2 = y_3^2$, hence at least two of C_1, C_2, C_3 should coincide. The above contradiction means that it remains to consider only the case $l_1 = l_2 > l_3$ when $x_1 = -x_2 \neq 0 = x_3$ and $y_1 = \pm y_2$, see Fig. 5.

If $y_1 = y_2$, the sets T_{\pm} are isometric by $(x, y, z) \mapsto$

$(-x, y, -z)$. If $y_1 = -y_2$ and $y_3 = 0$, the sets T_{\pm} are isometric by $(x, y, z) \mapsto (-x, -y, -z)$. If $y_1 = -y_2$ and $y_3 \neq 0$, then $C_1 = (x_1, y_1, 0)$, $C_2 = (-x_1, -y_1, 0)$, $C_3 \neq (0, 0, 0)$. Then among the six remaining rows, only the rows of $\{R, C_1\}$, $\{G, C_2\}$ have points at the distance $2\sqrt{l_3^2 + 1}$, see Table 6 for $i = 3$ considered modulo 3. Then $i + 1 \equiv 1 \pmod{3}$, $i - 1 \equiv 2 \pmod{3}$, so $l_{i+1} = l_1 = l_2 = l_{i-1}$.

Looking at the rows of $\{R, C_1\}$, $\{G, C_2\}$, the three common pairs in each of $\text{SDD}(T_{\pm}; 2)$ include the same distance $2\sqrt{l_1^2 + 1} = 2\sqrt{l_2^2 + 1}$ but differ by $|C_{i-1}C_i| = |C_2C_3| \neq$

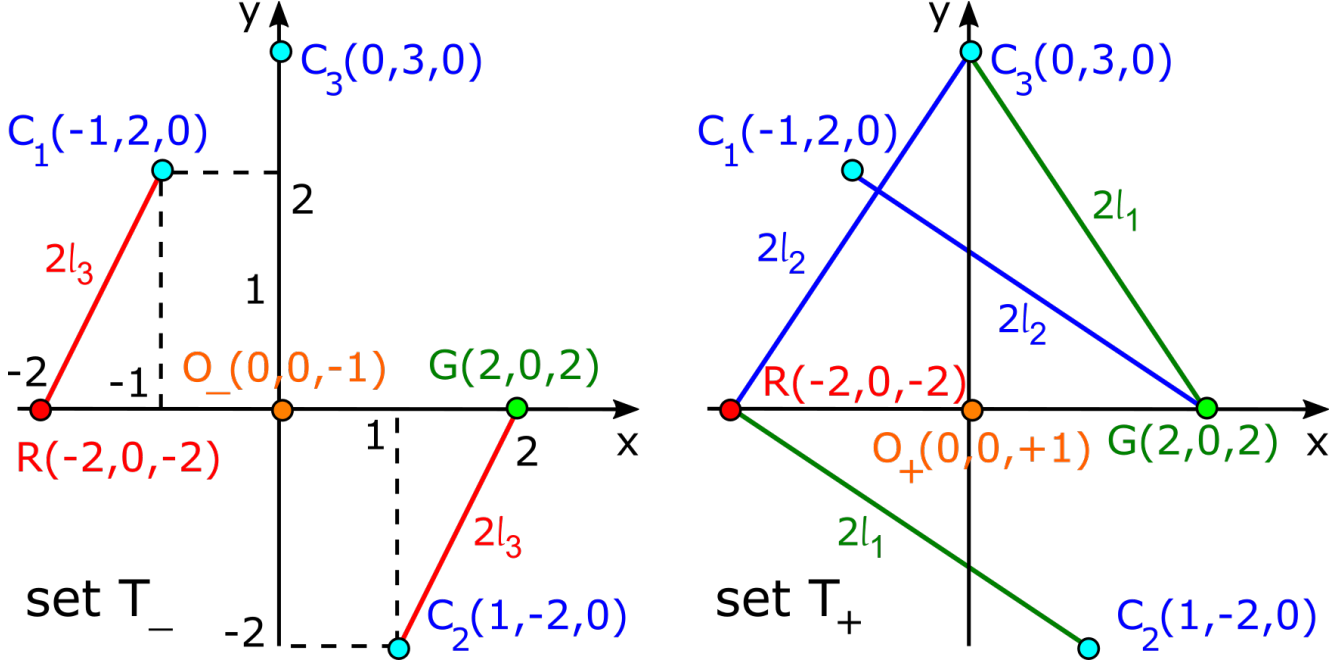


Figure 5. **Left:** (x, y) -projection of the 6-point set $T_- \subset \mathbb{R}^3$ consisting of the red point $R = (-2, 0, -2)$, green point $G = (2, 0, 2)$, three blue points $C_1 = (x_1, y_1, 0)$, $C_2 = (x_2, y_2, 0)$, $C_3 = (x_3, y_3, 0)$ and orange point $O_- = (0, 0, -1)$ so that $|RC_1| = 2l_3 = |GC_2|$, $|RC_2| = 2l_1 = |GC_3|$, $|RC_3| = 2l_2 = |GC_1|$. **Right:** to get the set $T_+ \subset \mathbb{R}^3$ from T_- , replace O_- with $O_+ = (0, 0, +1)$.

distances of T_-	R	G	C_1	C_2	C_3	O_-
$R = (-2, 0, -2)$	0	$\sqrt{32}$	$2\sqrt{l_3^2 + 1}$	$2\sqrt{l_1^2 + 1}$	$2\sqrt{l_2^2 + 1}$	$\sqrt{5}$
$G = (+2, 0, +2)$	$\sqrt{32}$	0	$2\sqrt{l_2^2 + 1}$	$2\sqrt{l_3^2 + 1}$	$2\sqrt{l_1^2 + 1}$	$\sqrt{13}$
$C_1 = (x_1, y_1, 0)$	$2\sqrt{l_3^2 + 1}$	$2\sqrt{l_2^2 + 1}$	0	$ C_1C_2 $	$ C_3C_1 $	$\sqrt{2l_2^2 + 2l_3^2 - 3}$
$C_2 = (x_2, y_2, 0)$	$2\sqrt{l_1^2 + 1}$	$2\sqrt{l_3^2 + 1}$	$ C_1C_2 $	0	$ C_2C_3 $	$\sqrt{2l_3^2 + 2l_1^2 - 3}$
$C_3 = (x_3, y_3, 0)$	$2\sqrt{l_2^2 + 1}$	$2\sqrt{l_1^2 + 1}$	$ C_3C_1 $	$ C_2C_3 $	0	$\sqrt{2l_1^2 + 2l_2^2 - 3}$
$O_- = (0, 0, -1)$	$\sqrt{5}$	$\sqrt{13}$	$\sqrt{2l_2^2 + 2l_3^2 - 3}$	$\sqrt{2l_3^2 + 2l_1^2 - 3}$	$\sqrt{2l_1^2 + 2l_2^2 - 3}$	0

distances of T_+	R	G	C_1	C_2	C_3	O_+
$R = (-2, 0, -2)$	0	$\sqrt{32}$	$2\sqrt{l_3^2 + 1}$	$2\sqrt{l_1^2 + 1}$	$2\sqrt{l_2^2 + 1}$	$\sqrt{13}$
$G = (+2, 0, +2)$	$\sqrt{32}$	0	$2\sqrt{l_2^2 + 1}$	$2\sqrt{l_3^2 + 1}$	$2\sqrt{l_1^2 + 1}$	$\sqrt{5}$
$C_1 = (x_1, y_1, 0)$	$2\sqrt{l_3^2 + 1}$	$2\sqrt{l_2^2 + 1}$	0	$ C_1C_2 $	$ C_3C_1 $	$\sqrt{2l_2^2 + 2l_3^2 - 3}$
$C_2 = (x_2, y_2, 0)$	$2\sqrt{l_1^2 + 1}$	$2\sqrt{l_3^2 + 1}$	$ C_1C_2 $	0	$ C_2C_3 $	$\sqrt{2l_3^2 + 2l_1^2 - 3}$
$C_3 = (x_3, y_3, 0)$	$2\sqrt{l_2^2 + 1}$	$2\sqrt{l_1^2 + 1}$	$ C_3C_1 $	$ C_2C_3 $	0	$\sqrt{2l_1^2 + 2l_2^2 - 3}$
$O_+ = (0, 0, +1)$	$\sqrt{13}$	$\sqrt{5}$	$\sqrt{2l_2^2 + 2l_3^2 - 3}$	$\sqrt{2l_3^2 + 2l_1^2 - 3}$	$\sqrt{2l_1^2 + 2l_2^2 - 3}$	0

Table 5. The matrices of distances between all points of the 6-point set T_{\mp} in Fig. 5 motivated by [55, Figure S4(C)].

$$|C_3C_1| = |C_iC_{i+1}| \text{ as } C_1 = \pm C_2, C_3 \neq (0, 0, 0).$$

$$l_1 = l_2 = \frac{\sqrt{13}}{2}, l_3 = \frac{\sqrt{5}}{2}.$$

This couple of different rows implies that $\text{SDD}(T_-; 2) \neq \text{SDD}(T_+; 2)$ due to the swapped distances $\sqrt{5}, \sqrt{13}$ in their remaining pairs, see Table 7 for the sets T_{\pm} in Fig. 5 with

T_- pair	distance	common pairs in $\widetilde{\text{SDD}}(T_{\pm}; 2)$	pairs that differ in $\widetilde{\text{SDD}}(T_{+}; 2)$
$\{R, O_{-}\}$	$\sqrt{5}$	$(\sqrt{13}, \sqrt{32})$ to G	$(2\sqrt{l_3^2 + 1}, \sqrt{2l_2^2 + 2l_3^2 - 3})$ to C_1 , $(2\sqrt{l_1^2 + 1}, \sqrt{2l_3^2 + 2l_1^2 - 3})$ to C_2 , $(2\sqrt{l_2^2 + 1}, \sqrt{2l_1^2 + 2l_2^2 - 3})$ to C_3
$\{G, O_{-}\}$	$\sqrt{13}$	$(\sqrt{5}, \sqrt{32})$ to R	$(2\sqrt{l_2^2 + 1}, \sqrt{2l_2^2 + 2l_3^2 - 3})$ to C_1 , $(2\sqrt{l_3^2 + 1}, \sqrt{2l_3^2 + 2l_1^2 - 3})$ to C_2 , $(2\sqrt{l_1^2 + 1}, \sqrt{2l_1^2 + 2l_2^2 - 3})$ to C_3
$\{R, C_{i+1}\}$	$2\sqrt{l_i^2 + 1}$	$(2\sqrt{l_{i-1}^2 + 1}, \sqrt{32})$ to G , $(2\sqrt{l_{i+1}^2 + 1}, C_{i+1}C_{i-1})$ to C_{i-1} , $(2\sqrt{l_{i-1}^2 + 1}, C_iC_{i+1})$ to C_i	$(\sqrt{5}, \sqrt{2l_{i-1}^2 + 2l_i^2 - 3})$ to O_{-}
$\{G, C_{i-1}\}$	$2\sqrt{l_i^2 + 1}$	$(2\sqrt{l_{i+1}^2 + 1}, \sqrt{32})$ to R , $(2\sqrt{l_{i-1}^2 + 1}, C_{i+1}C_{i-1})$ to C_{i+1} , $(2\sqrt{l_{i+1}^2 + 1}, C_{i-1}C_i)$ to C_i	$(\sqrt{13}, \sqrt{2l_i^2 + 2l_{i+1}^2 - 3})$ to O_{-}
T_+ pair	distance	common pairs in $\widetilde{\text{SDD}}(T_{\pm}; 2)$	pairs that differ in $\widetilde{\text{SDD}}(T_{-}; 2)$
$\{G, O_{+}\}$	$\sqrt{5}$	$(\sqrt{13}, \sqrt{32})$ to R	$(2\sqrt{l_2^2 + 1}, \sqrt{2l_2^2 + 2l_3^2 - 3})$ to C_1 , $(2\sqrt{l_3^2 + 1}, \sqrt{2l_3^2 + 2l_1^2 - 3})$ to C_2 , $(2\sqrt{l_1^2 + 1}, \sqrt{2l_1^2 + 2l_2^2 - 3})$ to C_3
$\{R, O_{+}\}$	$\sqrt{13}$	$(\sqrt{5}, \sqrt{32})$ to G	$(2\sqrt{l_3^2 + 1}, \sqrt{2l_2^2 + 2l_3^2 - 3})$ to C_1 , $(2\sqrt{l_1^2 + 1}, \sqrt{2l_3^2 + 2l_1^2 - 3})$ to C_2 , $(2\sqrt{l_2^2 + 1}, \sqrt{2l_1^2 + 2l_2^2 - 3})$ to C_3
$\{R, C_{i+1}\}$	$2\sqrt{l_i^2 + 1}$	$(2\sqrt{l_{i-1}^2 + 1}, \sqrt{32})$ to G , $(2\sqrt{l_{i+1}^2 + 1}, C_{i+1}C_{i-1})$ to C_{i-1} , $(2\sqrt{l_{i-1}^2 + 1}, C_iC_{i+1})$ to C_i	$(\sqrt{13}, \sqrt{2l_{i-1}^2 + 2l_i^2 - 3})$ to O_{+}
$\{G, C_{i-1}\}$	$2\sqrt{l_i^2 + 1}$	$(2\sqrt{l_{i+1}^2 + 1}, \sqrt{32})$ to R , $(2\sqrt{l_{i-1}^2 + 1}, C_{i+1}C_{i-1})$ to C_{i+1} , $(2\sqrt{l_{i+1}^2 + 1}, C_{i-1}C_i)$ to C_i	$(\sqrt{5}, \sqrt{2l_i^2 + 2l_{i+1}^2 - 3})$ to O_{+}

Table 6. For the sets T_{\pm} , the distributions $\widetilde{\text{SDD}}(T_{\pm}; 2)$ can differ only by $\widetilde{\text{RDD}}$ s of the pairs $\{R, O_{\pm}\}, \{G, O_{\pm}\}, \{R, C_i\}, \{G, C_i\}$ shown above, where $i \in \{1, 2, 3\}$ is considered modulo 3 so that $1 - 1 \equiv 3 \pmod{3}$. In rows of corresponding pairs of points, some pairs of distances are the same in both $\widetilde{\text{SDD}}(T_{\pm}; 2)$, but other pairs can differ. If parameters l_1, l_2, l_3 are pairwise distinct, the rows $\{R, O_{-}\}, \{G, O_{+}\}$ include three different pairs of distances, so $\widetilde{\text{SDD}}(T_{-}; 2) \neq \widetilde{\text{SDD}}(T_{+}; 2)$, see Example 4.3.

T_- pair	distance	distance to neighbor1	distance to neighbor 2	distance to neighbor 3	distance to neighbor 4
$\{R, C_1\}$	3	$(\sqrt{2}, \sqrt{17})$ to C_3	$(\sqrt{5}, \sqrt{6})$ to O_-	$(\sqrt{17}, \sqrt{20})$ to C_2	$(\sqrt{17}, \sqrt{32})$ to G
$\{G, C_2\}$	3	$(\sqrt{6}, \sqrt{13})$ to O_-	$(\sqrt{17}, \sqrt{20})$ to C_1	$(\sqrt{17}, \sqrt{26})$ to C_3	$(\sqrt{17}, \sqrt{32})$ to R
T_+ pair	distance	distance to neighbor1	distance to neighbor 2	distance to neighbor 3	distance to neighbor 4
$\{R, C_1\}$	3	$(\sqrt{2}, \sqrt{17})$ to C_3	$(\sqrt{6}, \sqrt{13})$ to O_+	$(\sqrt{17}, \sqrt{20})$ to C_2	$(\sqrt{17}, \sqrt{32})$ to G
$\{G, C_2\}$	3	$(\sqrt{5}, \sqrt{6})$ to O_+	$(\sqrt{17}, \sqrt{20})$ to C_1	$(\sqrt{17}, \sqrt{26})$ to C_3	$(\sqrt{17}, \sqrt{32})$ to R

Table 7. The above rows show that $\text{SDD}(T_-; 2) \neq \text{SDD}(T_+; 2)$ for the sets T_\pm with $C_1 = (-1, 2, 0)$, $C_2 = (1, -2, 0)$, $C_3 = (0, 3, 0)$ so that $l_1 = l_2 = \frac{\sqrt{5}}{2}$, $l_3 = \frac{\sqrt{13}}{2}$ in Table 6.

5. Continuous and computable metrics on SDD

The $m - h$ permutable columns of the matrix $R(C; A)$ in RDD from Definition 3.1 can be interpreted as $m - h$ unlabelled points in \mathbb{R}^h . Since any isometry is bijective, the simplest metric respecting bijections is the bottleneck distance (also called the Wasserstein distance W_∞).

Definition 5.1 (bottleneck distance W_∞). *For any vector $v = (v_1, \dots, v_n) \in \mathbb{R}^n$, the Minkowski norm is $\|v\|_\infty = \max_{i=1, \dots, n} |v_i|$. For any vectors or matrices N, N' of the same size, the Minkowski distance is $L_\infty(N, N') = \max_{i,j} |N_{ij} - N'_{ij}|$. For clouds $C, C' \subset \mathbb{R}^n$ of m unlabelled points, the bottleneck distance $W_\infty(C, C') = \inf_{g: C \rightarrow C'} \sup_{p \in C} \|g(p) - p\|_\infty$ is minimized over all bijections $g: C \rightarrow C'$.*

Lemma 5.2 (the max metric M_∞ on RDDs). *For any m -point clouds and ordered h -point sequences $A \subset C$ and $A' \subset C'$, set $d(\xi) = \max\{L_\infty(\xi(D(A)), D(A')), W_\infty(\xi(R(C; A)), R(C'; A'))\}$ for a permutation $\xi \in S_h$ on h points. Then the max metric $M_\infty(\text{RDD}(C; A), \text{RDD}(C'; A')) = \min_{\xi \in S_h} d(\xi)$ satisfies all metric axioms on RDDs from Definition 3.1 and can be computed in time $O(h!(h^2 + m^{1.5} \log^h m))$.*

Proof of Lemma 5.2. The first metric axiom says that $\text{RDD}(C; A), \text{RDD}(C'; A')$ are equivalent by Definition 3.1 if and only if $M_\infty(\text{RDD}(C; A), \text{RDD}(C'; A')) = 0$ or $d(\xi) = 0$ for some permutation $\xi \in S_h$. Then $d(\xi) = 0$ is equivalent to $\xi(D(A)) = D(A')$ and $\xi(R(C; A)) = R(C'; A')$ up to a permutation of columns due to the first axiom for W_∞ . The last two conclusions mean that the Relative Distance Distributions $\text{RDD}(C; A), \text{RDD}(C'; A')$ are equivalent by Definition 3.1. The symmetry axiom follows since any permutation ξ is invertible. To prove the triangle inequality $M_\infty(\text{RDD}(C; A), \text{RDD}(C''; A'')) + M_\infty(\text{RDD}(C''; A''), \text{RDD}(C'; A')) \geq M_\infty(\text{RDD}(C; A), \text{RDD}(C'; A'))$, let $\xi, \xi' \in S_h$ be optimal permutations for the M_∞ values in the left-hand side

above. The triangle inequality for L_∞ says that

$$\begin{aligned} &L_\infty(\xi(D(A)), D(A')) + \\ &L_\infty(\xi'(D(A'')), D(A')) \geq \\ &L_\infty(\xi(D(A)), \xi'(D(A''))) = \\ &L_\infty(\xi'^{-1}\xi(D(A)), D(A'')), \end{aligned}$$

similarly for the bottleneck distance W_∞ from Definition 5.1. Taking the maximum of L_∞, W_∞ preserves the triangle inequality. Then $M_\infty(\text{RDD}(C; A), \text{RDD}(C''; A'')) = \min_{\xi \in S_h} d(\xi)$ cannot be larger than $d(\xi'^{-1}\xi)$ for the composition of the permutations above, so the triangle inequality holds for M_∞ .

For a fixed permutation $\xi \in S_h$, the distance $L_\infty(\xi(D(A)), D(A'))$ requires $O(h^2)$ time. The bottleneck distance $W_\infty(\xi(R(C; A)), R(C'; A'))$ on the $h \times (m - h)$ matrices $\xi(R(C; A))$ and $R(C'; A')$ with permutable columns can be considered as the bottleneck distance on clouds of $(m - h)$ unlabelled points in \mathbb{R}^h , so $W_\infty(\xi(R(C; A)), R(C'; A'))$ needs only $O(m^{1.5} \log^h m)$ time by [24, Theorem 6.5]. The minimization over all permutations $\xi \in S_h$ gives the factor $h!$ in the final time. \square

For $h = 1$ and a 1-point subset $A \subset C$, the matrix $D(A)$ is empty, so $d(\xi) = W_\infty(\xi(R(C; A)), R(C'; A'))$. The metric M_∞ on RDDs will be used for intermediate costs to get metrics on unordered collections of RDDs (SDDs) by using the standard tools in Definitions 5.3 and 5.4 below.

Definition 5.3 (Linear Assignment Cost LAC [28]). *For any $k \times k$ matrix of costs $c(i, j) \geq 0$, $i, j \in \{1, \dots, k\}$,*

the Linear Assignment Cost LAC = $\frac{1}{k} \min_g \sum_{i=1}^k c(i, g(i))$ is minimized for all bijections g on the indices $1, \dots, k$.

The normalization factor $\frac{1}{k}$ in LAC makes this metric better comparable with EMD whose weights sum up to 1.

Definition 5.4 (Earth Mover's Distance on distributions). *Let $B = \{B_1, \dots, B_k\}$ be a finite unordered set of objects with weights $w(B_i)$, $i = 1, \dots, k$. Consider another set*

$D = \{D_1, \dots, D_l\}$ with weights $w(D_j)$, $j = 1, \dots, l$. Assume that a distance between B_i, D_j is measured by a metric $d(B_i, D_j)$. A flow from B to D is a $k \times l$ matrix whose entry $f_{ij} \in [0, 1]$ represents a partial flow from an object B_i to D_j . The Earth Mover's Distance [59] is the minimum of

$$\text{EMD}(B, D) = \sum_{i=1}^k \sum_{j=1}^l f_{ij} d(B_i, D_j) \text{ over } f_{ij} \in [0, 1] \text{ subject to } \sum_{j=1}^l f_{ij} \leq w(B_i) \text{ for } i = 1, \dots, k, \sum_{i=1}^k f_{ij} \leq w(D_j) \text{ for } j = 1, \dots, l, \text{ and } \sum_{i=1}^k \sum_{j=1}^l f_{ij} = 1.$$

The first condition $\sum_{j=1}^l f_{ij} \leq w(B_i)$ means that not more than the weight $w(B_i)$ of the object B_i ‘flows’ into all D_j via the flows f_{ij} , $j = 1, \dots, l$. The second condition $\sum_{i=1}^k f_{ij} \leq w(D_j)$ means that all flows f_{ij} from B_i for $i = 1, \dots, k$ ‘flow’ to D_j up to its weight $w(D_j)$. The last condition $\sum_{i=1}^k \sum_{j=1}^l f_{ij} = 1$ forces all B_i to collectively ‘flow’ into all D_j . LAC [28] and EMD [59] can be computed in a near cubic time in the sizes of given sets of objects.

Theorem 5.5(b) extends the $O(m^{1.5} \log^n m)$ algorithm for fixed clouds of m unlabelled points in [24, Theorem 6.5] to the harder case of isometry classes but keeps the polynomial time in m for a fixed dimension n .

Theorem 5.5 (time of metrics on SDDs). *For any m -point clouds C, C' in their own metric spaces and $h \geq 1$, let the Simplexwise Distance Distributions $\text{SDD}(C; h)$ and $\text{SDD}(C'; h)$ consist of $k = \binom{m}{h}$ RDDs with equal weights $\frac{1}{k}$ without collapsing identical RDDs.*

(a) *Using the $k \times k$ matrix of costs computed by the metric M_∞ between RDDs from $\text{SDD}(C; h)$ and $\text{SDD}(C'; h)$, the Linear Assignment Cost LAC from Definition 5.3 satisfies all metric axioms on SDDs and can be computed in time $O(h!(h^2 + m^{1.5} \log^h m)k^2 + k^3 \log k)$.*

(b) *Let $\text{SDD}(C; h)$ and $\text{SDD}(C'; h)$ have a maximum size $l \leq k$ after collapsing identical RDDs. Then EMD from Definition 5.4 satisfies all metric axioms on SDDs and is computed in time $O(h!(h^2 + m^{1.5} \log^h m)l^2 + l^3 \log l)$.*

Proof. The Linear Assignment Cost (LAC) from Definition 5.3 is symmetric because any bijective matching can be reversed. The triangle inequality for LAC follows from the triangle inequality for the metric M_∞ in Lemma 5.2 by using a composition of bijections $\text{SDD}(C; h) \rightarrow \text{SDD}(C'; h) \rightarrow \text{SDD}(C''; h)$ matching all RDDs similarly to the proof of Lemma 5.2. The first metric axiom for LAC means that LAC = 0 if and only if there is a bijection $g : \text{SDD}(C; h) \rightarrow \text{SDD}(C'; h)$ so that all matched RDDs

are at distance $M_\infty = 0$, so these RDDs are equivalent (hence SDDs are equal) due to the first axiom of $M_\infty = 0$, which was proved in Lemma 5.2.

The metric axioms for the Earth Mover's Distance (EMD) are proved in the appendix of [59] assuming the metric axioms for the underlying distance d , which is the metric M_∞ from Lemma 5.2 in our case.

The time complexities for LAC and EMD follow from the time $O((h^2 + m^{1.5} \log^h m)h!)$ for M_∞ in Lemma 5.2, after multiplying by a quadratic factor for the size of cost matrices and adding a near cubic time [28, 30]. \square

The Lipschitz continuity of SDD in Theorem 5.8 needs Lemma 5.7 follows from its partial case in Lemma 5.6.

Lemma 5.6. *For any $a, b, c, d \in \mathbb{R}$, if $a \leq b$ and $c \leq d$ then $\max\{|a - c|, |b - d|\} \leq \max\{|a - d|, |b - c|\}$.*

Proof. We consider several cases of the relative locations of the pairs $a \leq b$ and $c \leq d$ in the line \mathbb{R} .

Case $a \leq b \leq c \leq d$. The required inequality follows from $\max\{c - a, d - b\} \leq d - a = \max\{d - a, c - b\}$.

Case $a \leq c \leq b \leq d$. The required inequality follows from $\max\{c - a, d - b\} \leq d - a = \max\{d - a, b - c\}$.

Case $a \leq c \leq d \leq b$. The inequality $\max\{c - a, b - d\} \leq \max\{d - a, b - c\}$ holds as $c - a \leq d - a$, $b - d \leq b - c$.

All other cases reduce to the cases above by the transformation $(a, b) \mapsto (-b, -a)$, $(c, d) \mapsto (-d, -c)$, which preserves the given condition and required conclusion. \square

Lemma 5.7 (the metric L_∞ respects ordering). *For any vector $v = (v_1, \dots, v_k) \in \mathbb{R}^k$, the vector $\vec{v} \in \mathbb{R}^k$ is obtained from v by writing all coordinates in increasing order. Then $|\vec{u} - \vec{v}|_\infty \leq |u - v|_\infty$ for any vectors $u, v \in \mathbb{R}^k$.*

Proof. Since $|u - v|_\infty = \max_{i=1, \dots, k} |u_i - v_i|$, the metric L_∞ is preserved under any permutation $\xi \in S_k$ applied simultaneously to the coordinates of both u, v . Hence, without loss of generality, we can assume that all coordinates of one vector u are already in increasing order, so $u = \vec{u}$. For any pair of successive coordinates $u_i \leq u_{i+1}$, let the corresponding pair in v be in the opposite order $v_i > v_{i+1}$.

By Lemma 5.6 the swap $v_i \leftrightarrow v_{i+1}$ does not increase the L_∞ distance between (u_i, u_{i+1}) and (v_i, v_{i+1}) , hence between $u, v \in \mathbb{R}^k$. Applying such swaps puts all coordinates of v in increasing order without increasing L_∞ . \square

Theorem 5.8 substantially generalizes the fact that perturbing two points in their ε -neighborhoods changes the Euclidean distance between these points by at most 2ε .

Theorem 5.8 (Lipschitz continuity of SDDs). *In any metric space, let C' be obtained from a cloud C by perturbing each point within its ε -neighborhood. For*

any $h \geq 1$, $\text{SDD}(C; h)$ changes by at most 2ε in the LAC and EMD metrics. The lower bound holds: $\text{EMD}(\text{SDD}(C; h), \text{SDD}(C'; h)) \geq |\text{SDM}(C; h, 1) - \text{SDM}(C'; h, 1)|_\infty$.

Proof. Order all points of the given clouds C, C' so that every point $p_i \in C$ has the same index as its perturbation $p'_i \in C'$. In the given metric space, the distance $d(p_i, p_j)$ between any points in C changes under perturbation by at most 2ε so that $|d(p_i, p_j) - d(p'_i, p'_j)| \leq 2\varepsilon$. This upper bound 2ε remains for the max metric M_∞ from Lemma 5.2, also for the LAC and EMD metrics due to the total weight 1 of all costs in Definitions 5.3 and 5.4, respectively.

Lemma 5.7 implies that re-writing coordinates of a vector in increasing order cannot increase the metric L_∞ , hence $L_\infty(\xi(D(A)), D(A')) \geq |\text{SDV}(A) - \text{SDV}(A')|_\infty$ for any permutation $\xi \in S_h$ of indices $1, \dots, h$.

The bottleneck distance $W_\infty(\xi(R(C; A)), R(C'; A'))$ is the maximum of the metric L_∞ computed between corresponding column vectors of the $h \times (m - h)$ matrices $\xi(R(C; A))$ and $R(C'; A')$. Let $d = (d_1, \dots, d_h)$ and $d' = (d'_1, \dots, d'_h)$ be two such columns. The triangle inequalities imply that $|d - d'|_\infty = \max_{i=1, \dots, h} |d_i - d'_i| \geq \frac{1}{h} \sum_{i=1}^h |d_i - d'_i|$.

Hence taking averages of all vector coordinates cannot increase the metric L_∞ . Then $W_\infty(\xi(R(C; A)), R(C'; A'))$ has the lower bound equal to the metric L_∞ between the vectors of $m - h$ column averages in the matrices $\xi(R(C; A))$ and $R(C'; A')$. Applying Lemma 5.7 to these vectors in \mathbb{R}^{m-h} implies that $W_\infty(\xi(R(C; A)), R(C'; A')) \geq |\bar{R}(C; A) - \bar{R}(C'; A')|_\infty$.

Taking the maximum of the metrics L_∞ and W_∞ considered above, we get the lower bound in terms of the Average Distance Distribution from Definition 3.4: $d(\xi) = \max\{L_\infty(\xi(D(A)), D(A')), W_\infty(\xi(R(C; A)), R(C'; A'))\} \geq |\text{ADD}(C; A) - \text{ADD}(C'; A')|_\infty$.

Since the above argument holds for any permutation $\xi \in S_h$, we get $M_\infty(\text{RDD}(C; A), \text{RDD}(C'; A')) = \min d(\xi) \geq |\text{ADD}(C; A) - \text{ADD}(C'; A')|_\infty$.

Both $\text{SDD}(C; h)$ and $\text{ASD}(C; h)$ are unordered collections of $\binom{m}{h}$ RDD and vectors, respectively. If we use an optimal flow matrix f_{ij} for $\text{EMD}(\text{SDD}(C; h), \text{SDD}(C'; h))$ from Definition 5.4 to compute EMD on ASD vectors, we get an upper bound for $\text{EMD}(\text{ASD}(C; h), \text{ASD}(C'; h))$, which can be potentially smaller (for another flow matrix) but not larger, so $\text{EMD}(\text{SDD}(C; h), \text{SDD}(C'; h)) \geq \text{EMD}(\text{ASD}(C; h), \text{ASD}(C'; h))$. Considering $\text{ASD}(S; h)$ as a weighted distribution of vectors, $\text{SDM}(C; h, 1)$ is its centroid from section 3 in [20, section 3]. The lower bound $\text{EMD}(\text{ASD}(C; h), \text{ASD}(C'; h)) \geq |\text{SDM}(C; h, 1) - \text{SDM}(C'; h, 1)|_\infty$ follows from [20, Theorem 1]. \square

6. Measured Simplexwise Distribution (MSD)

This section adapts SDD for metric-measure spaces.

Definition 6.1 (metric-measure space). According to Gromov [33, section 3 $\frac{1}{2}$.1], a metric-measure space (X, d_X, μ_X) is a compact space X with a metric d_X and a Borel measure μ_X such that $\mu_X(X) < +\infty$. An isomorphism between metric-measure spaces is an isometry $f : X \rightarrow Y$ that respects the measures in the sense that $\mu_Y(U) = \mu_X(f^{-1}(U))$ for any subset $U \subset Y$.

Dividing $\mu_X(U)$ by the full measure $\mu_X(X) < +\infty$ for any $U \subset X$, we can assume that $\mu_X(X) = 1$, so μ_X is a probability measure. Any metric space X of m points can be considered a metric-measure space with the uniform measure $\mu_X(p) = \frac{1}{m}$ for all points $p \in X$.

On two points $0, 1$ in Euclidean line \mathbb{R} , the mm-spaces $X = (\{0, 1\}, 1, \{\frac{1}{2}, \frac{1}{2}\})$ and $Y = (\{0, 1\}, 1, \{\frac{1}{3}, \frac{2}{3}\})$ are isometric but not isomorphic because of different weights.

Definition 6.2 extends the local distribution of distances from [48, Definition 5.5] to higher orders $h > 1$.

Definition 6.2 (Measured Simplexwise Distribution MSD). Let (X, d_X, μ_X) be any metric-measure space. For any basis sequence $A = (p_1, \dots, p_h) \in X^h$ of $h \geq 1$ ordered points, write the triangular distance matrix $D(A)$ from Definition 3.1 row-by-row as the Vector of Interpoint Distances $\text{VID}(A) \in \mathbb{R}_+^{h(h-1)/2}$ so that $\text{VID}_k = d_X(p_i, p_j)$ for $k = h(i-1) + j - 1$, $1 \leq i < j \leq h$. For a vector $\vec{d} = (d_1, \dots, d_h) \in \mathbb{R}_+^h$ of distance thresholds, the Vector of Sequence-based Measures $\text{VSM}(A; \vec{d}) \in \mathbb{R}_+^h$ consists of h values $\mu_X(\{q \in X : d_X(q, p_i) \leq d_i\})$ for $i = 1, \dots, h$. The Measured Simplexwise Distribution of order $h \geq 1$ is the function $\text{MSD}[X; h] : X^h \times \mathbb{R}_+^h \rightarrow \mathbb{R}_+^{h(h+1)/2}$ mapping any $A \in X^h$ and $\vec{d} \in \mathbb{R}_+^h$ to the pair $[\text{VID}(A), \text{VSM}(A; \vec{d})]$ considered as a concatenated vector in $\mathbb{R}_+^{h(h+1)/2}$.

For $h = 1$, the vector $\text{VID}(A)$ is empty and the Measured Simplexwise Distribution of order $h = 1$ coincides with the local distribution of distances [48, Definition 5.5] $\text{MSD}[X; 1] : X \times \mathbb{R}_+ \rightarrow \mathbb{R}_+$ mapping any point $p \in X$ and a threshold $d \in \mathbb{R}_+$ to $\mu_X(\{q \in X : d_X(q, p) \leq d\})$.

Any permutation ξ on indices $1, \dots, h$ naturally permutes the components of $\text{MSD}[X; h]$. If X consists of m points, $\text{MSD}[X; h]$ reduces to the finite collection of $\binom{m}{h}$ vectors $\text{VID}(A)$ paired with fields $\text{VSM}(A; \vec{d}) : \mathbb{R}_+^h \rightarrow \mathbb{R}_+^h$ only for unordered h -point subsets $A \subset X$, which can be refined to a stronger invariant analog of SDD below.

Definition 6.3 (Weighted Simplexwise Distribution WSD). Let X be a finite mm-space whose any point p has a weight $w(p)$. For $h \geq 1$ and a sequence $A = (p_1, \dots, p_h) \in X^h$

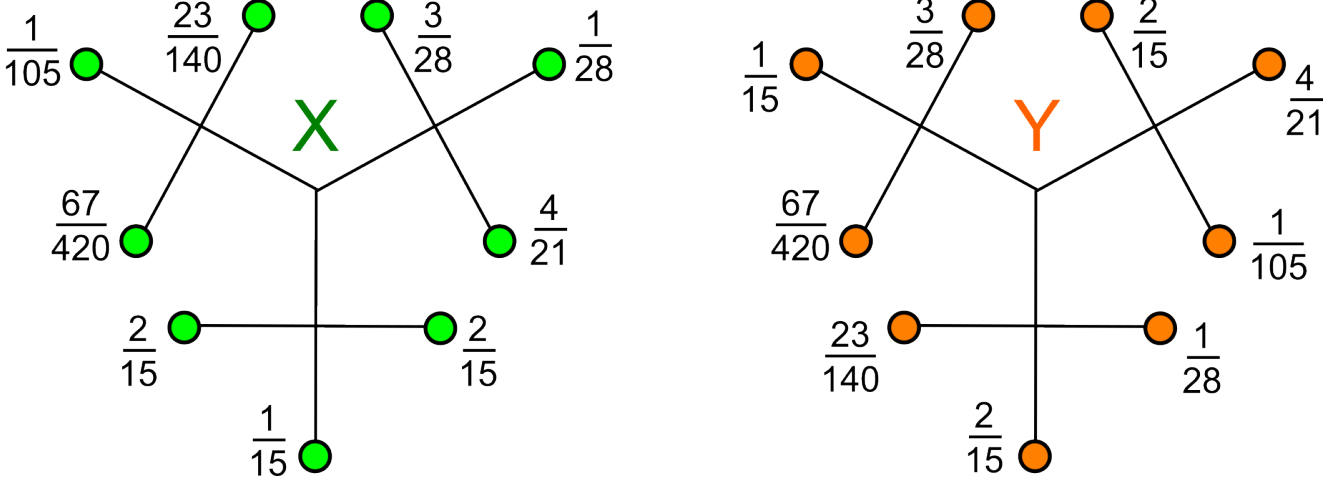


Figure 6. Non-isomorphic mm-spaces X, Y from [48, Fig. 8] have equal local distributions of distances but are distinguished by the new Weighted Simplexwise Distribution of order 1 and the Measured Simplexwise Distributions of order 2, see details in Example 6.4.

in Definition 3.1, endow any distance $d(p, q)$ in $D(A)$ with the unordered pair $w(p), w(q)$ of weights. For every point $q \in X - A$, put the weight $w(q)$ in the extra $(h + 1)$ -st row of the matrix $M(X; A)$ whose columns are indexed by unordered $q \in X - A$. If $h = 1$ and $A = p_1$, set $D(A) = w(p_1)$. The Weighted Distance Distribution $WDD(X; A)$ is the equivalence class of the pair $[D(A); M(X; A)]$ up to permutations $\xi \in S_h$ acting on A . The Weighted Simplexwise Distribution $WSD(X)$ is the unordered collection of $WDD(X; A)$ for all subsets $A \subset X$ of unordered h points.

For finite mm-spaces, a metric on WDDs can be defined similar to M_∞ from Lemma 5.2 by combining the weights and distances. Then LAC and EMD from Definitions 5.3 and 5.4 can be computed as in Theorem 5.5.

Example 6.4 (the strength of WSD). Fig. 6 shows mm-spaces X, Y on 9 points visualised as trees [48, Fig. 8]. All edges have length $\frac{1}{2}$ and induce the shortest-path metrics d_X, d_Y . The sum of weights in every small branch of 3 nodes is $\frac{1}{3}$. These mm-spaces X, Y have all inter-point distances only 1 and 2, and equal local distributions of distances $MSD[X; 1] = MSD[Y; 1]$ by [48, Example 5.6].

Indeed, both MSDs can be considered the same set of 9 piecewise constant functions $\mu(p)$ taking values $w(p), \frac{1}{3}$, and 1 on the intervals $[0, 1), [1, 2), [2, +\infty)$, respectively.

However, WSDs have more pointwise data: $WSD[X; 1]$ has $A(D) = w(p) = \frac{23}{140}$ and the following 2×8 matrix

$$M(X; p) = \begin{pmatrix} 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 \\ \frac{1}{105} & \frac{67}{420} & \frac{2}{15} & \frac{1}{15} & \frac{2}{15} & \frac{4}{21} & \frac{1}{28} & \frac{3}{28} \end{pmatrix},$$

but $WSD[Y; 1]$ has another matrix for $w(p) = \frac{23}{140}$.

$$M(Y; p) = \begin{pmatrix} 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 \\ \frac{2}{15} & \frac{1}{28} & \frac{1}{105} & \frac{4}{21} & \frac{2}{15} & \frac{3}{28} & \frac{1}{15} & \frac{67}{420} \end{pmatrix}.$$

The matrices with freely permutable columns are different, so X, Y are distinguished by WSD for $h = 1$

Also, $MSD[X; 2] \neq MSD[Y; 2]$ because, for any basis sequence $A = (p, q) \in X^2$, we have $VSM[X; 2](A; d_1, d_2) = (w(p), w(q))$ for $d_1, d_2 < 1$ since all other points have minimum distance 1 from p, q , similarly for Y . The unique points p, q of weights $w(p) = \frac{23}{140}$ and $w(q) = \frac{67}{420}$ have different distances $d_X(p, q) = 1$ and $d_Y(p, q) = 2$. Then $MSD[X; 2] \neq MSD[Y; 2]$ differ by the uniquely identifiable fields mapping $[0, 1)^2$ to the constant vector $(w(p), w(q))$ with $VID_X(A) = 1 \neq 2 = VID_Y(A)$.

We conjecture that any mm-spaces X, Y can be distinguished up to isomorphism by Measured Simplexwise Distributions for a high enough h depending on X, Y .

Future updates of this paper will include continuous metrics between Measured Simplexwise Distributions on mm-spaces. We are open to new ideas and collaboration.

New invariants in Definitions 3.3, 6.2 and main Theorems 3.6, 5.5, 5.8 essentially contribute to the new area of Geometric Data Science aiming to resolve all data challenges whose bottlenecks are analogues of Problem 1.1.

The earlier work has studied the following important cases of Problem 1.1: 1-periodic discrete series [5, 6, 41], 2D lattices [13, 43], 3D lattices [12, 40, 42, 51], periodic point sets in \mathbb{R}^3 [23, 64] and in higher dimensions [2–4].

The applications of Geometric Data Science to crystalline materials [7, 58, 68, 77] led to the Crystal Isometry Principle [71, 72, 74] extending Mendeleev's table of chemical elements to the Crystal Isometry Space of all periodic crystals continuously parametrised by complete invariants.

This work was supported by the Royal Academy of Engineering fellowship “Data science for next generation engineering of solid crystalline materials” (2021-2023, IF2122/186) and the EPSRC grants “Application-driven Topological Data Analysis” (2018-2023, EP/R018472/1) and “Inverse design of periodic crystals” (2022-2024, EP/X018474/1).

The author thanks all members of the Data Science Theory and Applications group in the Materials Innovation Factory (Liverpool, UK), especially Daniel Widdowson, Matthew Bright, Yury Elkin, Olga Anosova, also Justin Solomon (MIT), Steven Gortler (Harvard), Nadav Dym (Technion) for fruitful discussions, and any reviewers for their valuable time and helpful suggestions.

References

- [1] Naveed Akhtar and Ajmal Mian. Threat of adversarial attacks on deep learning in computer vision: A survey. *IEEE Access*, 6:14410–14430, 2018. [2](#)
- [2] Olga Anosova and Vitaliy Kurlin. Introduction to periodic geometry and topology. *arXiv:2103.02749*. [13](#)
- [3] Olga Anosova and Vitaliy Kurlin. An isometry classification of periodic point sets. In *Proceedings of Discrete Geometry and Mathematical Morphology*, 2021. [13](#)
- [4] Olga Anosova and Vitaliy Kurlin. Algorithms for continuous metrics on periodic crystals. *arxiv:2205.15298*, 2022. [2](#), [13](#)
- [5] Olga Anosova and Vitaliy Kurlin. Density functions of periodic sequences. *Lecture Notes in Computer Science (Proceedings of DGMM)*, 2022. [13](#)
- [6] O Anosova and V Kurlin. Density functions of periodic sequences of continuous events. *arXiv:2301.05137*, 2023. [13](#)
- [7] Jonathan Balasingham, Viktor Zamaraev, and Vitaliy Kurlin. Compact graph representation of crystals using Pointwise Distance Distributions. *arXiv:2212.11246*, 2022. [13](#)
- [8] Serge Belongie, Jitendra Malik, and Jan Puzicha. Shape matching and object recognition using shape contexts. *Transactions PAMI*, 24(4):509–522, 2002. [2](#)
- [9] Mireille Boutin and Gregor Kemper. On reconstructing n-point configurations from the distribution of distances or areas. *Adv. Appl. Math.*, 32(4):709–735, 2004. [4](#)
- [10] Peter Brass and Christian Knauer. Testing the congruence of d-dimensional point sets. In *SoCG*, pages 310–314, 2000. [2](#)
- [11] Peter Brass and Christian Knauer. Testing congruence and symmetry for general 3-dimensional objects. *Computational Geometry*, 27(1):3–11, 2004. [2](#)
- [12] Matthew J Bright, Andrew I Cooper, and Vitaliy A Kurlin. Welcome to a continuous world of 3-dimensional lattices. *arxiv:2109.11538*, 2021. [13](#)
- [13] Matthew J Bright, Andrew I Cooper, and Vitaliy A Kurlin. Geographic-style maps for 2-dimensional lattices. *Acta Crystallographica Section A*, 79(1):1–13, 2023. [13](#)
- [14] Michael M Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: grids, groups, graphs, geodesics, and gauges. *arXiv:2104.13478*, 2021. [2](#)
- [15] Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond Euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017. [2](#)
- [16] Haiwei Chen, Shichen Liu, Weikai Chen, Hao Li, and Randall Hill. Equivariant point network for 3D point cloud analysis. In *CVPR*, pages 14514–14523, 2021. [2](#)
- [17] Paul Chew, Dorit Dor, Alon Efrat, and Klara Kedem. Geometric pattern matching in d-dimensional space. *Discrete & Computational Geometry*, 21(2):257–274, 1999. [2](#)
- [18] Paul Chew, Michael Goodrich, Daniel Huttenlocher, Klara Kedem, Jon Kleinberg, and Dina Kravets. Geometric pattern matching under Euclidean motion. *Computational Geometry*, 7(1-2):113–124, 1997. [2](#)
- [19] Paul Chew and Klara Kedem. Improvements on geometric pattern matching problems. In *Scandinavian Workshop on Algorithm Theory*, pages 318–325, 1992. [2](#)
- [20] Scott Cohen and Leonidas Guibas. The Earth Mover’s Distance: lower bounds and invariance under translation. Technical report, Stanford University, 1997. [12](#)
- [21] Matthew J Colbrook, Vegard Antun, and Anders C Hansen. The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and Smale’s 18th problem. *PNAS*, 119(12):e2107151119, 2022. [2](#)
- [22] Yinpeng Dong, Fangzhou Liao, Tianyu Pang, Hang Su, Jun Zhu, Xiaolin Hu, and Jianguo Li. Boosting adversarial attacks with momentum. In *Computer vision and pattern recognition*, pages 9185–9193, 2018. [2](#)
- [23] H Edelsbrunner, T Heiss, V Kurlin, P Smith, and M Wintraecken. The density fingerprint of a periodic point set. In *Proceedings of SoCG*, pages 32:1–32:16, 2021. [13](#)
- [24] Alon Efrat, Alon Itai, and Matthew J Katz. Geometry helps in bottleneck matching and related problems. *Algorithmica*, 31(1):1–28, 2001. [10](#), [11](#)
- [25] Yury Elkin. New compressed cover tree for k-nearest neighbor search (phd thesis). *arxiv:2205.10194*, 2022. [3](#)
- [26] Y. Elkin and V. Kurlin. The mergegram of a dendrogram and its stability. In *Proceedings of MFCS*, 2020. [2](#)
- [27] Y. Elkin and V. Kurlin. Isometry invariant shape recognition of projectively perturbed point clouds by the mergegram extending 0d persistence. *Mathematics*, 9(17), 2021. [2](#)
- [28] Michael L Fredman and Robert Endre Tarjan. Fibonacci heaps and their uses in improved network optimization algorithms. *Journal ACM*, 34:596–615, 1987. [10](#), [11](#)
- [29] EN Gilbert and LA Shepp. Textures for discrimination experiments, 1974. [2](#)
- [30] Andrew Goldberg and Robert Tarjan. Solving minimum-cost flow problems by successive approximation. In *Proceedings of STOC*, pages 7–18, 1987. [11](#)
- [31] Michael T Goodrich, Joseph SB Mitchell, and Mark W Orletsky. Approximate geometric pattern matching under rigid motions. *Transactions PAMI*, 21:371–379, 1999. [2](#)
- [32] Cosmin Grigorescu and Nicolai Petkov. Distance sets for shape filters and shape recognition. *IEEE transactions on image processing*, 12(10):1274–1286, 2003. [2](#)
- [33] Mikhael Gromov, Misha Katz, Pierre Pansu, and Stephen Semmes. *Metric structures for Riemannian and non-Riemannian spaces*, volume 152. Springer, 1999. [12](#)

- [34] Chuan Guo, Jacob Gardner, Yurong You, Andrew Gordon Wilson, and Kilian Weinberger. Simple black-box adversarial attacks. In *ICML*, pages 2484–2493, 2019. [2](#)
- [35] Felix Hausdorff. Dimension und äußeres Maß. *Mathematische Annalen*, 79(2):157–179, 1919. [2](#)
- [36] Snir Hordan, Tal Amir, Steven J Gortler, and Nadav Dym. Complete neural networks for Euclidean graphs. *arXiv:2301.13821*, 2023. [2](#)
- [37] Daniel Huttenlocher, Gregory Klanderman, and William Rucklidge. Comparing images using the Hausdorff distance. *Transactions PAMI*, 15:850–863, 1993. [2](#)
- [38] Ernest Sydney Keeping. *Introduction to statistical inference*. Courier Corporation, 1995. [3](#)
- [39] Heuna Kim and Günter Rote. Congruence testing of point sets in 4 dimensions. *arXiv:1603.07269*, 2016. [2](#)
- [40] Vitaliy Kurlin. A complete isometry classification of 3-dimensional lattices. *arxiv:2201.10543*, 2022. [13](#)
- [41] Vitaliy Kurlin. Computable complete invariants for finite clouds of unlabeled points under Euclidean isometry. *arXiv:2207.08502*, 2022. [2](#), [13](#)
- [42] Vitaliy Kurlin. Exactly computable and continuous metrics on isometry classes of finite and 1-periodic sequences. *arxiv:2205.04388*, 2022. [13](#)
- [43] Vitaliy Kurlin. Mathematics of 2-dimensional lattices. *Found. Comp. Mathematics*, pages Dec 7: 1–59, 2022. [13](#)
- [44] Vitaliy Kurlin. The strength of a simplex is the key to a continuous isometry classification of Euclidean clouds of unlabelled points. *arXiv:2303.13486*, 2023. [3](#)
- [45] Cassidy Laidlaw and Soheil Feizi. Functional adversarial attacks. *Adv. Neural Inform. Proc. Systems*, 32, 2019. [2](#)
- [46] Sushovan Majhi, Jeffrey Vitter, and Carola Wenk. Approximating Gromov-Hausdorff distance in Euclidean space. *arXiv:1912.13008*, 2019. [2](#)
- [47] Siddharth Manay, Daniel Cremers, Byung-Woo Hong, Anthony Yezzi, and Stefano Soatto. Integral invariants for shape matching. *Trans. PAMI*, 28:1602–1618, 2006. [2](#)
- [48] Facundo Mémoli. Gromov–Wasserstein distances and the metric approach to object matching. *Foundations of Computational Mathematics*, 11(4):417–487, 2011. [2](#), [12](#), [13](#)
- [49] Facundo Mémoli and Tom Needham. Distance distributions and inverse problems for metric measure spaces. *Studies in Applied Mathematics*, 149(4):943–1001, 2022. [2](#)
- [50] Facundo Mémoli, Zane Smith, and Zhengchao Wan. The Gromov-Hausdorff distance between ultrametric spaces: its structure and computation. *arXiv:2110.03136*, 2021. [2](#)
- [51] Marco M Mosca and Vitaliy Kurlin. Voronoi-based similarity distances between arbitrary crystal lattices. *Crystal Research and Technology*, 55(5):1900197, 2020. [13](#)
- [52] Jigyasa Nigam, Michael J Willatt, and Michele Ceriotti. Equivariant representations for molecular hamiltonians and n-center atomic-scale properties. *Journal of Chemical Physics*, 156(1):014115, 2022. [2](#)
- [53] Robert et al Osada. Shape distributions. *Transactions on Graphics*, 21:807–832, 2002. [2](#)
- [54] Helmut Pottmann, Johannes Wallner, Qi-Xing Huang, and Yong-Liang Yang. Integral invariants for robust geometry processing. *Comp. Aided Geom. Design*, 26(1):37–60, 2009. [2](#)
- [55] Sergey N Pozdnyakov, Michael J Willatt, Albert P Bartók, Christoph Ortner, Gábor Csányi, and Michele Ceriotti. Incompleteness of atomic structure representations. *Phys. Rev. Lett.*, 125:166001, 2020. [4](#), [5](#), [6](#), [7](#), [8](#)
- [56] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in Neural Information Processing Systems*, 30, 2017. [2](#)
- [57] Blaine Rister, Mark A Horowitz, and Daniel L Rubin. Volumetric image registration from invariant keypoints. *Transactions on Image Processing*, 26(10):4900–4910, 2017. [2](#)
- [58] Jakob Ropers, Marco M Mosca, Olga D Anosova, Vitaliy A Kurlin, and Andrew I Cooper. Fast predictions of lattice energies by continuous isometry invariants of crystal structures. In *International Conference on Data Analytics and Management in Data Intensive Domains*, pages 178–192, 2022. [13](#)
- [59] Y. Rubner, C. Tomasi, and L. Guibas. The earth mover’s distance as a metric for image retrieval. *Intern. Journal of Computer Vision*, 40(2):99–121, 2000. [11](#)
- [60] Felix Schmiedl. Computational aspects of the Gromov–Hausdorff distance and its application in non-rigid shape matching. *Discrete Comp. Geometry*, 57:854–880, 2017. [2](#)
- [61] Isaac Schoenberg. Remarks to Maurice Frechet’s article “Sur la définition axiomatique d’une classe d’espace distances vectoriellement applicable sur l’espace de Hilbert. *Annals of Mathematics*, pages 724–732, 1935. [2](#)
- [62] Anthony Simeonov, Yilun Du, Andrea Tagliasacchi, Joshua B Tenenbaum, Alberto Rodriguez, Pulkit Agrawal, and Vincent Sitzmann. Neural descriptor fields: SE(3)-equivariant object representations for manipulation. In *ICRA*, pages 6394–6400, 2022. [2](#)
- [63] Philip Smith and Vitaliy Kurlin. Families of point sets with identical 1d persistence. *arxiv:2202.00577*, 2022. [2](#)
- [64] Phil Smith and Vitaliy Kurlin. A practical algorithm for degree-k voronoi domains of three-dimensional periodic point sets. In *Lecture Notes in Computer Science (Proceedings of ISVC)*, volume 13599, pages 377–391, 2022. [13](#)
- [65] Riccardo Spezialetti, Samuele Salti, and Luigi Di Stefano. Learning an effective equivariant 3d descriptor without supervision. In *ICCV*, pages 6401–6410, 2019. [2](#)
- [66] Jian Sun, Maks Ovsjanikov, and Leonidas Guibas. A concise and provably informative multi-scale signature based on heat diffusion. *Comp. Graph. Forum*, 28:1383–1392, 2009. [2](#)
- [67] Matthew Toews and William M Wells III. Efficient and robust model-to-image alignment using 3d scale-invariant features. *Medical image analysis*, 17(3):271–282, 2013. [2](#)
- [68] Aikaterini Vriza, Ioana Sovago, Daniel Widdowson, Peter Wood, Vitaliy Kurlin, and Matthew Dyer. Molecular set transformer: Attending to the co-crystals in the cambridge structural database. *Digital Discovery*, 1:834–850, 2022. [13](#)
- [69] E Weisstein. Triangle. <https://mathworld.wolfram.com>. [1](#)
- [70] Carola Wenk. Shape matching in higher dimensions. *PhD thesis, FU Berlin*, 2003. [2](#)
- [71] D Widdowson and V Kurlin. Pointwise distance distributions of periodic point sets. *arxiv:2108.04798*, 2021. [2](#), [13](#)
- [72] Daniel Widdowson and Vitaliy Kurlin. Resolving the data ambiguity for periodic crystals. *Advances in Neural Information Processing Systems (NeurIPS)*, 35, 2022. [2](#), [3](#), [13](#)

- [73] Daniel Widdowson and Vitaliy Kurlin. Recognizing rigid patterns of unlabeled point clouds by complete and continuous isometry invariants with no false negatives and no false positives. In *Proceedings of CVPR*, 2023. 2
- [74] Daniel Widdowson, Marco M Mosca, Angeles Pulido, Andrew I Cooper, and Vitaliy Kurlin. Average minimum distances of periodic point sets - foundational invariants for mapping all periodic crystals. *MATCH Comm. in Math. and in Computer Chemistry*, 87:529–559, 2022. 3, 13
- [75] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander J Smola. Deep sets. *Adv. Neural Inform. Proc. Systems*, 30, 2017. 2
- [76] N Zava. The Gromov-Hausdorff space isn't coarsely embeddable into any Hilbert space. *arXiv:2303.04730*, 2023. 2
- [77] Q Zhu, J Johal, D Widdowson, Z Pang, B Li, C Kane, V Kurlin, G Day, M Little, and A Cooper. Analogy powered by prediction and structural invariants: Computationally-led discovery of a mesoporous hydrogen-bonded organic cage crystal. *J Amer. Chem. Soc.*, 144:9893–9901, 2022. 13
- [78] Wen Zhu, Lingchao Chen, Beiping Hou, Weihan Li, Tianliang Chen, and Shixiong Liang. Point cloud registration of arvester based on scale-invariant points feature histogram. *Scientific Reports*, 12(1):1–13, 2022. 2