Recognizing Rigid Patterns of Unlabeled Point Clouds by Complete and Continuous Isometry Invariants with no False Negatives and no False Positives

Daniel Widdowson
Computer Science department
University of Liverpool, UK
d.e.widdowson@liverpool.ac.uk

Vitaliy Kurlin
Computer Science department
University of Liverpool, UK
vitaliy.kurlin@gmail.com

Abstract

Rigid structures such as cars or any other solid objects are often represented by finite clouds of unlabeled points. The most natural equivalence on these point clouds is rigid motion or isometry maintaining all inter-point distances.

Rigid patterns of point clouds can be reliably compared only by complete isometry invariants that can also be called equivariant descriptors without false negatives (isometric clouds having different descriptions) and without false positives (non-isometric clouds with the same description).

Noise and motion in data motivate a search for invariants that are continuous under perturbations of points in a suitable metric. We propose the first continuous and complete invariant of unlabeled clouds in any Euclidean space. For a fixed dimension, the new metric for this invariant is computable in a polynomial time in the number of points.

1. Strong motivations for complete invariants

In Computer Vision, real objects such as cars and solid obstacles are considered rigid and often represented by a finite set $C \subset \mathbb{R}^n$ (called a *cloud*) of m *unlabeled* (or unordered) points, usually in low dimensions $n = 2, 3, 4$.

The rigidity of many real objects motivates the most fundamental equivalence of *rigid motion* [67], a composition of translations and rotations in $\mathbb{R}^n$. In a general metric space M, the most relevant equivalence is *isometry*: any map $M \to M$ maintaining all inter-point distances in M.

Any isometry in $\mathbb{R}^n$ is a composition of a mirror reflection with some rigid motion. Any orientation-preserving isometry can be realized as a continuous rigid motion.

There is no sense in distinguishing rigid objects that are related by isometry or having the same shape. Formally, the *shape* of a cloud C is its isometry class [48] defined as a collection of all infinitely many clouds isometric to C.

The only reliable tool for distinguishing clouds up to isometry is an *invariant* defined as a function or property preserved by any isometry. Since any isometry is bijective, the number of points is an isometry invariant, but the coordinates of points are not invariants even under translation. This simple invariant is *incomplete* (non-injective) because non-isometric clouds can have different numbers of points.

Any invariant I maps all isometric clouds to the same value. There are no isometric clouds $C \cong C'$ with $I(C) \neq I(C')$, meaning that I has *no false negatives*. Isometry invariants are also called *equivariant descriptors* [55].

A *complete* invariant I should distinguish all non-isometric clouds, so if $C \ncong C'$ then $I(C) \neq I(C')$. Equivalently, if $I(C) = I(C')$ then $C \cong C'$, so I has *no false positives*. Then I can be considered as a DNA-style code or genome that identifies any cloud uniquely up to isometry.

Since real data is always noisy and motions of rigid objects are important to track, a useful complete invariant must be also continuous under the movement of points.

A complete and continuous invariant for $m = 3$ points consists of three pairwise distances (sides of a triangle) and is known in school as the SSS theorem [68]. But all pairwise distances are incomplete for $m \geq 4$ [9], see Fig. 1.

Problem 1.1 (complete isometry invariants with computable continuous metrics)**.** *For any cloud of m unlabeled points in $\mathbb{R}^n$, find an invariant I satisfying the properties*

(a) completeness *: C, C' are isometric $\Leftrightarrow I(C) = I(C')$;*

(b) Lipschitz continuity *: if any point of C is perturbed within its ε-neighborhood then $I(C)$ changes by at most $\lambda\varepsilon$ for a constant λ and a metric d satisfying these axioms:*

1) $d(I(C), I(C')) = 0$ if and only if $C \cong C'$ are isometric,

2) symmetry *: $d(I(C), I(C')) = d(I(C'), I(C))$,*

3) $d(I(C), I(C')) + d(I(C'), I(C'')) \geq d(I(C), I(C''))$;

(c) computability *: I and d are computed in a polynomial time in the number m of points for a fixed dimension n.*

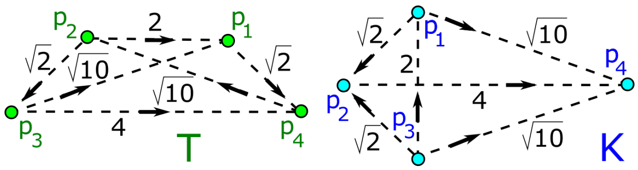

Figure 1. **Left**: the cloud $T = \{(1,1),(-1,1),(-2,0),(2,0)\}$. **Right**: the kite $K = \{(0,1),(-1,0),(0,-1),(3,0)\}$. T and K have the same 6 pairwise distances $\sqrt{2},\sqrt{2},2,\sqrt{10},\sqrt{10},4$.

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

Theorem 4.7 will solve Problem 1.1 for any m in $\mathbb{R}^n$. Continuous invariants in Theorem 3.10 are conjectured to be complete (no known counter-examples) in any metric space. The first author implemented all algorithms, the second author wrote all theory, proofs, examples in [38, 39].

2. Past work on cloud recognition/classification

Labeled clouds $C \subset \mathbb{R}^n$ are easy for isometry classification because the matrix of distances d_{ij} between indexed points p_i, p_j allows us to reconstruct C by using the known distances to the previously constructed points [28, Theorem 9]. For any clouds of the same number m of labeled points, the difference between $m \times m$ matrices of distances (or Gram matrices of $p_i \cdot p_j$) can be converted into a continuous metric by taking a matrix norm. If the given points are unlabeled, comparing $m \times m$ matrices requires $m!$ permutations, which makes this approach impractical.

Multidimensional scaling (MDS). For a given $m \times m$ distance matrix of any m-point cloud A, MDS [57] 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 $O(2^m)$.

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 $\mathbb{R}^3$. These algorithms output a binary answer (yes/no) without quantifying similarity between non-isometric clouds by a continuous metric.

The Hausdorff distance [30] 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 $\vec{d}_H(A,B) = \sup\limits_{p \in A} \inf\limits_{q \in B} |p - q|$. To take into account isometries, one can minimize the Hausdorff distance over all isometries [15, 17, 32]. 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 [16]. 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 $O(m^{\lceil (n+1)/2 \rceil})$ [69, Chapter 4, Corollary 6]. For general isometry, only approximate algorithms tackled minimizations for infinitely many rotations initially in $\mathbb{R}^3$ [26] 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 [56, Corollary 3.8]. Polynomial-time algorithms for GH were designed for ultrametric spaces [45]. However, GH spaces are challenging even for point clouds sets in $\mathbb{R}$, see [41] and [74].

The Heat Kernel Signature (HKS) is a complete isometry invariant of a manifold M whose the Laplace-Beltrami operator has distinct eigenvalues by [64, Theorem 1]. If M is sampled by points, HKS can be discretized and remains continuous [64, section 4] but the completeness is unclear.

Equivariant descriptors can be experimentally optimized [47, 59] on big datasets of clouds that are split into predefined clusters. Using more hidden parameters can improve accuracy on any finite dataset at a higher cost but will require more work for any new data. Point cloud registration filters outliers [58], samples rotations for Scale Invariant Feature Transform or uses a basis [52, 63, 65, 76], which can be unstable under perturbations of a cloud. The PCA-based complete invariant of unlabelled clouds [35] can discontinuously change when a basis degenerates to a lower dimensional subspace but inspired Complete Neural Networks [31] though without the Lipschitz continuity.

Geometric Deep Learning produces descriptors that are equivariant by design [13] and go beyond Euclidean space $\mathbb{R}^n$ [14], hence aiming to experimentally solve Problem 1.1. Motivated by obstacles in [1, 18, 19, 29, 40], Problem 1.1 needs a justified solution without relying on finite data.

Geometric Data Science solves analogs of Problem 1.1 for any real data objects considered up to practical equivalences instead of rigid motion on clouds [23, 24, 61]: 1-periodic discrete series [5, 6, 35], 2D lattices [12, 37], 3D lattices [11, 34, 36, 46], periodic point sets in $\mathbb{R}^3$ [20, 62] and in higher dimensions [2–4]. The applications of to crystalline materials [7, 53, 66, 75] led to the *Crystal Isometry Principle* [70, 71, 73] extending Mendeleev's table of elements to the *Crystal Isometry Space* of all periodic crystals parametrised by complete invariants like a geographic map of a planet.

Local distributions of distances in Mémoli's seminal work [43, 44] for metric-measure spaces, or shape distributions [8,27,42,49], are first-order versions of the new SDD below.

3. Simplexwise Distance Distribution (SDD)

We will refine Sorted Distance Vector in any metric space to get a complete invariant in $\mathbb{R}^n$ as shown in Fig. 2. All proofs from sections 3 and 4 are in [38,39], respectively.

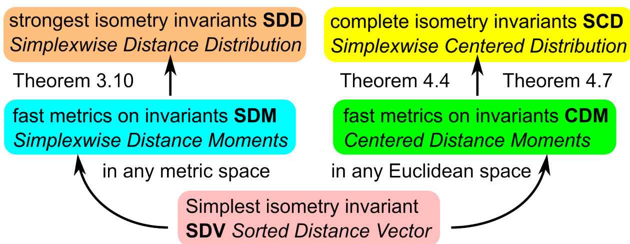

Figure 2. Hierarchy of new invariants on top of the classical SDV.

The *lexicographic* order $u < v$ on vectors $u = (u_1, \dots, u_h)$ and $v = (v_1, \dots, v_h)$ in $\mathbb{R}^h$ 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 ($\mathrm{RDD}(C; A)$)**.** *Let C be a cloud of m unlabeled points in a space with a metric d. Let $A = (p_1, \dots, p_h) \subset C$ be an ordered subset of $1 \leq h < m$ 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 $\mathrm{RDD}(C; A)$ *is the equivalence class of the pair $[D(A), R(C; A)]$ of matrices up to permutations $\xi \in S_h$.*

For a 1-point subset $A = \{p_1\}$ with $h = 1$, 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 a 2-point subset $A = (p_1, p_2)$ with $h = 2$, 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 $q \in C$.

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. 3 (left). Then*

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

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

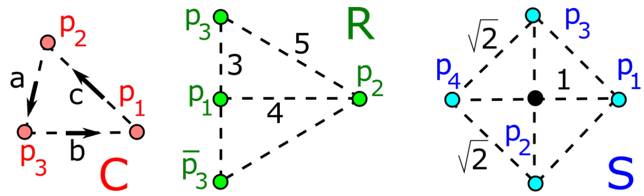

Figure 3. **Left**: a cloud $C = \{p_1, p_2, p_3\}$ with distances $a \leq b \leq c$. **Middle**: the triangular cloud $R = \{(0,0), (4,0), (0,3)\}$. **Right**: the square cloud $S = \{(1,0), (-1,0), (0,1), (-1,0)\}$.

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

We will always represent RDD *for a specified order $A = (p_i, p_j)$ of points that are written as a column. Swapping the points $p_1 \leftrightarrow p_2$ makes the last* RDD *above equivalent to another form:* $\mathrm{RDD}(C; \begin{pmatrix} p_2 \\ p_1 \end{pmatrix}) = [c; \begin{pmatrix} a \\ b \end{pmatrix}]$.

Though $\mathrm{RDD}(C; A)$ is defined up to a permutation ξ of h points in $A \subset C$, we later use only $h = n$, which makes comparisons of RDDs practical in dimensions $n = 2, 3$. Metrics on isometry classes of C will be independent of ξ.

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

For $h = 1$ and any m-point cloud C, the distribution $\mathrm{SDD}(C; 1)$ 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 $\mathrm{PDD}(C; m - 1)$ introduced in [71, Definition 3.1].

The PDD was simplified to the easier-to-compare vector of Average Minimum Distances [73]: $\mathrm{AMD}_k(C) = \dfrac{1}{m} \sum\limits_{i=1}^{m} d_{ik}$, 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 [22] and were pairwise compared for all all 660K+ periodic crystals in the world's largest database of real materials [71]. 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* $\mu = \dfrac{1}{k} \sum\limits_{i=1}^{k} a_i$. The 2nd moment is the *standard deviation* $\sigma = \sqrt{\dfrac{1}{k} \sum\limits_{i=1}^{k} (a_i - \mu)^2}$. For $l \geq 3$, the l-th *standardized moment* [33, section 2.7] is $\dfrac{1}{k} \sum\limits_{i=1}^{k} \left(\dfrac{a_i - \mu}{\sigma} \right)^l$.

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 $\mathrm{SDV}(A)$ *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 $[\mathrm{SDV}(A); \vec{R}(C;A)]$ is the Average Distance Distribution $\mathrm{ADD}(C;A)$ *considered as a vector of length $\frac{h(h-3)}{2} + m$. The unordered collection of $\mathrm{ADD}(C;A)$ for all $\binom{m}{h}$ unordered subsets $A \subset C$ is the Average Simplexwise Distribution $\mathrm{ASD}(C;h)$. The* Simplexwise Distance Moment $\mathrm{SDM}(C;h,l)$ *is the l-th (standardized for $l \geq 3$) moment of $\mathrm{ASD}(C;h)$ 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 Sorted Distance Vector* $\mathrm{SDV} = \{\sqrt{2}, \sqrt{2}, 2, \sqrt{10}, \sqrt{10}, 4\}$*, see infinitely many examples in [9]. The arrows on the edges of T,K show orders of points in each pair of vertices for* RDD*s. Then T,K are distinguished up to isometry by* $\mathrm{SDD}(T;2) \neq \mathrm{SDD}(K;2)$ *in Table 1. The 1st coordinate of* $\mathrm{SDM}(C;2,1) \in \mathbb{R}^3$ *is the average of 6 distances from* $\mathrm{SDV}(T) = \mathrm{SDV}(K)$ *but the other two coordinates (column averages from $R(C;A)$ matrices) differ.*

$\mathrm{RDD}(T;A)$ in $\mathrm{SDD}(T;2)$	$\mathrm{RDD}(K;A)$ in $\mathrm{SDD}(K;2)$
$[\sqrt{2}, \begin{pmatrix} 2 & \sqrt{10} \\ \sqrt{10} & 4 \end{pmatrix}] \times 2$	$[\sqrt{2}, \begin{pmatrix} 2 & \sqrt{10} \\ \sqrt{2} & 4 \end{pmatrix}] \times 2$
$[2, \begin{pmatrix} \sqrt{2} & \sqrt{10} \\ \sqrt{10} & \sqrt{2} \end{pmatrix}]$	$[2, \begin{pmatrix} \sqrt{2} & \sqrt{10} \\ \sqrt{2} & \sqrt{10} \end{pmatrix}]$
$[\sqrt{10}, \begin{pmatrix} \sqrt{2} & 4 \\ 2 & \sqrt{2} \end{pmatrix}] \times 2$	$[\sqrt{10}, \begin{pmatrix} \sqrt{2} & 2 \\ 4 & \sqrt{10} \end{pmatrix}] \times 2$
$[4, \begin{pmatrix} \sqrt{2} & \sqrt{10} \\ \sqrt{10} & \sqrt{2} \end{pmatrix}]$	$[4, \begin{pmatrix} \sqrt{2} & \sqrt{2} \\ \sqrt{10} & \sqrt{10} \end{pmatrix}]$
$\mathrm{ADD}(T;A)$ in $\mathrm{ASD}(T;2)$	$\mathrm{ADD}(K;A)$ in $\mathrm{ASD}(K;2)$
$[\sqrt{2}, (\frac{2+\sqrt{10}}{2}, \frac{4+\sqrt{10}}{2})] \times 2$	$[\sqrt{2}, (\frac{2+\sqrt{2}}{2}, \frac{4+\sqrt{10}}{2})] \times 2$
$[2, (\frac{\sqrt{2}+\sqrt{10}}{2}, \frac{\sqrt{2}+\sqrt{10}}{2})]$	$[2, (\sqrt{2}, \sqrt{10})]$
$[\sqrt{10}, (\frac{2+\sqrt{2}}{2}, \frac{4+\sqrt{2}}{2})] \times 2$	$[\sqrt{10}, (\frac{2+\sqrt{10}}{2}, \frac{4+\sqrt{2}}{2})] \times 2$
$[4, (\frac{\sqrt{2}+\sqrt{10}}{2}, \frac{\sqrt{2}+\sqrt{10}}{2})]$	$[4, (\frac{\sqrt{2}+\sqrt{10}}{2}, \frac{\sqrt{2}+\sqrt{10}}{2})]$
$\mathrm{SDM}_1 = \dfrac{3+\sqrt{2}+\sqrt{10}}{3}$	$\mathrm{SDM}_1 = \dfrac{3+\sqrt{2}+\sqrt{10}}{3}$
$\mathrm{SDM}_2 = \dfrac{6+2\sqrt{2}+4\sqrt{10}}{12}$	$\mathrm{SDM}_2 = \dfrac{8+5\sqrt{2}+3\sqrt{10}}{12}$
$\mathrm{SDM}_3 = \dfrac{16+4\sqrt{2}+4\sqrt{10}}{12}$	$\mathrm{SDM}_3 = \dfrac{16+3\sqrt{2}+5\sqrt{10}}{12}$

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 $\mathrm{SDM}(C;2,1) \in \mathbb{R}^3$ from Definition 3.4 for $h=2$, $l=1$ and both $C = T, K$. Different elements are highlighted.

Some of the $\binom{m}{h}$ RDDs in $\mathrm{SDD}(C;h)$ can be identical 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.

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

Definition 3.6 (bottleneck distance W_∞)**.** *For any vector $v = (v_1, \dots, v_n) \in \mathbb{R}^n$, the* Minkowski *norm is* $||v||_\infty = \max\limits_{i=1,\dots,n} |v_i|$*. For any vectors or matrices N, N' of the same size, the* Minkowski *distance is* $L_\infty(N,N') = \max\limits_{i,j} |N_{ij} - N'_{ij}|$*. For clouds $C, C' \subset \mathbb{R}^n$ of m unlabeled points, the* bottleneck distance $W_\infty(C,C') = \inf\limits_{g:C \to C'} \sup\limits_{p \in C} ||p - g(p)||_\infty$ *is minimized over all bijections $g: C \to C'$.*

Lemma 3.7 (the max metric M_∞ on RDDs)**.** *For any m-point clouds and ordered h-point subsets $A \subset C$ 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* $M_\infty(\mathrm{RDD}(C;A), \mathrm{RDD}(C';A')) = \min\limits_{\xi \in S_h} d(\xi)$ *satisfies all metric axioms on* RDD*s from Definition 3.1 and can be computed in time $O(h!(h^2 + m^{1.5} \log^h m))$.*

We will use only $h = n$ for Euclidean space $\mathbb{R}^n$, so the factor $h!$ in Lemma 3.7 is practically small for $n = 2, 3$.

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 between two unordered collections of RDDs by using standard Definitions 3.8 and 3.9 below.

Definition 3.8 (Linear Assignment Cost LAC [25])**.** *For any $k \times k$ matrix of costs $c(i,j) \geq 0$, $i,j \in \{1,\dots,k\}$, the* Linear Assignment Cost $\mathrm{LAC} = \frac{1}{k} \min\limits_{g} \sum\limits_{i=1}^{k} c(i, g(i))$ *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 3.9 (Earth Mover's Distance on distributions). *Let $B = \{B_1, \ldots, B_k\}$ be a finite unordered set of objects with weights $w(B_i)$, $i = 1, \ldots, k$. Consider another set $D = \{D_1, \ldots, D_l\}$ with weights $w(D_j)$, $j = 1, \ldots, l$. 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 *[54] is the minimum of $\mathrm{EMD}(B, D) = \sum\limits_{i=1}^{k}\sum\limits_{j=1}^{l} f_{ij} d(B_i, D_j)$ over $f_{ij} \in [0,1]$ subject to $\sum\limits_{j=1}^{l} f_{ij} \le w(B_i)$ for $i = 1, \ldots, k$, $\sum\limits_{i=1}^{k} f_{ij} \le w(D_j)$ for $j = 1, \ldots, l$, and $\sum\limits_{i=1}^{k}\sum\limits_{j=1}^{l} f_{ij} = 1$.*

The first condition $\sum\limits_{j=1}^{l} f_{ij} \le w(B_i)$ 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}, $j = 1, \ldots, l$. The second condition $\sum\limits_{i=1}^{k} f_{ij} \le w(D_j)$ means that all flows f_{ij} from B_i for $i = 1, \ldots, k$ 'flow' to D_j up to its weight $w(D_j)$. The last condition $\sum\limits_{i=1}^{k}\sum\limits_{j=1}^{l} f_{ij} = 1$ forces all B_i to collectively 'flow' into all D_j. LAC [25] and EMD [54] can be computed in a near cubic time in the sizes of given sets of objects.

Theorems 3.10(c) and 4.7 will extend $O(m^{1.5} \log^n m)$ algorithms for fixed clouds of m unlabeled points in [21, Theorem 6.5] to the harder case of isometry classes but keep the polynomial time in m for a fixed dimension n. All complexities are for a random-access machine (RAM) model.

Theorem 3.10 (invariance and continuity of SDDs). ***(a)*** *For $h \ge 1$ and any cloud C of m unlabeled points in a metric space, $\mathrm{SDD}(C; h)$ is an isometry invariant, which can be computed in time $O(m^{h+1}/(h-1)!)$. For any $l \ge 1$, the invariant $\mathrm{SDM}(C; h, l) \in \mathbb{R}^{m + \frac{h(h-3)}{2}}$ has the same time.*

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

(b) *Using the $k \times k$ matrix of costs computed by the metric M_∞ between RDDs from $\mathrm{SDD}(C; h)$ and $\mathrm{SDD}(C'; h)$, the Linear Assignment Cost LAC from Definition 3.8 satisfies all metric axioms on SDDs and can be computed in time $O(h!(h^2 + m^{1.5} \log^h m)k^2 + k^3 \log k)$.*

(c) *Let $\mathrm{SDD}(C; h)$ and $\mathrm{SDD}(C'; h)$ have a maximum size $l \le k$ after collapsing identical RDDs. Then EMD from Definition 3.9 satisfies all metric axioms on SDDs and is computed in time $O(h!(h^2 + m^{1.5} \log^h m)l^2 + l^3 \log l)$.*

(d) *Let C' be obtained from C by perturbing each point within its ε-neighborhood. For any $h \ge 1$, $\mathrm{SDD}(C; h)$ changes by at most 2ε in the LAC and EMD metrics. The lower bound holds: $\mathrm{EMD}\big(\mathrm{SDD}(C; h), \mathrm{SDD}(C'; h)\big) \ge |\mathrm{SDM}(C; h, 1) - \mathrm{SDM}(C'; h, 1)|_\infty$.*

Theorem 3.10(d) substantially generalizes the fact that perturbing two points in their ε-neighborhoods changes the Euclidean distance between these points by at most 2ε.

We conjecture that $\mathrm{SDD}(C; h)$ is a complete isometry invariant of a cloud $C \subset \mathbb{R}^n$ for some $h \ge n - 1$. [38, section 4] shows that $\mathrm{SDD}(C; 2)$ distinguished all infinitely many known pairs [50, Fig. S4] of non-isometric m-point clouds $C, C' \subset \mathbb{R}^3$ with identical $\mathrm{PDD}(C) = \mathrm{SDD}(C; 1)$.

4. Simplexwise Centered Distribution (SCD)

While all constructions of section 3 hold in any metric space, this section develops faster continuous metrics for complete isometry invariants of unlabeled clouds in $\mathbb{R}^n$.

The Euclidean structure of $\mathbb{R}^n$ allows us to translate the *center of mass* $\frac{1}{m}\sum\limits_{p \in C} p$ of a given m-point cloud $C \subset \mathbb{R}^n$ to the origin $0 \in \mathbb{R}^n$. Then Problem 1.1 reduces to only rotations around 0 from the orthogonal group $\mathrm{O}(\mathbb{R}^n)$.

Though the center of mass is uniquely determined by any cloud $C \subset \mathbb{R}^n$ of unlabeled points, real applications may offer one or several labeled points of C that substantially speed up metrics on invariants. For example, an atomic neighborhood in a solid material is a cloud $C \subset \mathbb{R}^3$ of atoms around a central atom, which may not be the center of mass of C, but is suitable for all methods below.

This section studies metrics on complete invariants of $C \subset \mathbb{R}^n$ up to rotations around the origin $0 \in \mathbb{R}^n$, which may or may not belong to C or be its center of mass.

For any subset $A = \{p_1, \ldots, p_{n-1}\} \subset C$, the distance matrix $D(A \cup \{0\})$ from Definition 3.1 has size $(n-1) \times (n-1)$ and its last column can be chosen to include the distances from $n-1$ points of A to the origin at $0 \in \mathbb{R}^n$.

Any n vectors $v_1, \ldots, v_n \in \mathbb{R}^n$ can be written as columns in the $n \times n$ matrix whose determinant has a *sign* ± 1 or 0 if $v_1, \ldots, v_n$ are linearly dependent. Any permutation $\xi \in S_n$ on indices $1, \ldots, n$ is a composition of some t transpositions $i \leftrightarrow j$ and has $\mathrm{sign}(\xi) = (-1)^t$.

Definition 4.1 (Simplexwise Centered Distribution SCD). *Let $C \subset \mathbb{R}^n$ be any cloud of m unlabeled points. For any ordered subset A of points $p_1, \ldots, p_{n-1} \in C$, the matrix $R(C; A)$ from Definition 3.1 has a column of Euclidean distances $|q - p_1|, \ldots, |q - p_{n-1}|$. At the bottom of this column, add the distance $|q - 0|$ to the origin and the sign of the determinant of the $n \times n$ matrix consisting of the vectors $q - p_1, \ldots, q - p_{n-1}, q$. The resulting $(n+1) \times (m-n+1)$-matrix with signs in the bottom $(n+1)$-st row is the* oriented relative distance *matrix $M(C; A \cup \{0\})$.*

Any permutation $\xi \in S_{n-1}$ of $n-1$ points of A acts on $D(A)$, permutes the first $n-1$ rows of $M(C; A\cup\{0\})$ and multiplies every sign in the $(n+1)$-st row by $\mathrm{sign}(\xi)$.

The Oriented Centered Distribution $\mathrm{OCD}(C;A)$ *is the equivalence class of pairs $[D(A\cup\{0\}), M(C; A\cup\{0\})]$ considered up to permutations $\xi \in S_{n-1}$ of points of A.*

The Simplexwise Centered Distribution $\mathrm{SCD}(C)$ *is the unordered set of the distributions* $\mathrm{OCD}(C;A)$ *for all $\binom{m}{n-1}$ unordered $(n-1)$-point subsets $A \subset C$. The mirror image* $\overline{\mathrm{SCD}}(C)$ *is obtained from* $\mathrm{SCD}(C)$ *by reversing signs.*

Definition 4.1 needs no permutations for any $C \subset \mathbb{R}^2$ as $n-1 = 1$. Columns of $M(C; A\cup\{0\})$ can be lexicographically ordered without affecting the metric in Lemma 4.6. Some of the $\binom{m}{n-1}$ OCDs in $\mathrm{SCD}(C)$ can be identical as in Example 4.2(b). If we collapse any $l > 1$ identical OCDs into a single OCD with the *weight* $l/\binom{m}{h}$, SCD can be considered as a weighted probability distribution of OCDs.

Example 4.2 (SCD for clouds in Fig. 3)**.** ***(a)*** *Let $R \subset \mathbb{R}^2$ consist of the vertices $p_1 = (0,0)$, $p_2 = (4,0)$, $p_3 = (0,3)$ of the right-angled triangle in Fig. 3 (middle). Though $p_1 = (0,0)$ is included in R and is not its center of mass,* $\mathrm{SCD}(R)$ *still makes sense. In* $\mathrm{OCD}(R;p_1) = [0, \begin{pmatrix} 4 & 3 \\ 4 & 3 \\ 0 & 0 \end{pmatrix}]$*, the matrix $D(\{p_1, 0\})$ is $|p_1 - 0| = 0$, the top row has $|p_2 - p_1| = 4$, $|p_3 - p_1| = 3$. In* $\mathrm{OCD}(R;p_2) = [4, \begin{pmatrix} 4 & 5 \\ 0 & 3 \\ 0 & - \end{pmatrix}]$*, the first row has $|p_1 - p_2| = 4$, $|p_3 - p_2| = 5$, the second row has $|p_1 - 0| = 0$, $|p_3 - 0| = 3$,* $\det\begin{pmatrix} -4 & 0 \\ 3 & 3 \end{pmatrix} < 0$*. In* $\mathrm{OCD}(R;p_3) = [3, \begin{pmatrix} 3 & 5 \\ 0 & 4 \\ 0 & + \end{pmatrix}]$*, the first row has $|p_1 - p_3| = 3$, $|p_2 - p_3| = 5$, the second row has $|p_1 - 0| = 0$, $|p_2 - 0| = 4$,* $\det\begin{pmatrix} 4 & 4 \\ -3 & 0 \end{pmatrix} > 0$*. So* $\mathrm{SCD}(R)$ *consists of the three Oriented Centered Distributions* OCDs *above.*

If we reflect R with respect to the x-axis, the new cloud $\bar{R}$ of the points $p_1, p_2, \bar{p}_3 = (0,-3)$ has $\mathrm{SCD}(\bar{R}) = \overline{\mathrm{SCD}}(R)$ *with* $\mathrm{OCD}(\bar{R};p_1) = \mathrm{OCD}(R)$, $\mathrm{OCD}(\bar{R};p_2) = [4, \begin{pmatrix} 4 & 5 \\ 0 & 3 \\ 0 & + \end{pmatrix}]$, $\mathrm{OCD}(R;\bar{p}_3) = [3, \begin{pmatrix} 3 & 5 \\ 0 & 4 \\ 0 & - \end{pmatrix}]$ *whose signs changed under reflection, so* $\mathrm{SCD}(R) \neq \mathrm{SCD}(\bar{R})$.

(b) *Let $S \subset \mathbb{R}^2$ consist of $m = 4$ points $(\pm 1, 0), (0, \pm 1)$ that are vertices of the square in Fig. 3 (right). The center of mass is $0 \in \mathbb{R}^2$ and has a distance 1 to each point of S.*

For each 1-point subset $A = \{p\} \subset S$, the distance matrix $D(A\cup\{0\})$ on two points is the single number 1. The matrix $M(S; A\cup\{0\})$ has $m-n+1 = 3$ columns. For $p_1 = (1,0)$, we have $M(S; \begin{pmatrix} p_1 \\ 0 \end{pmatrix}) = \begin{pmatrix} \sqrt{2} & \sqrt{2} & 2 \\ 1 & 1 & 1 \\ - & + & 0 \end{pmatrix}$*, where the columns are ordered according to $p_2 = (0,-1)$, $p_3 = (0,1)$, $p_4 = (-1,0)$ in Fig. 3 (right). The sign in the bottom right corner is 0 because the points $p_1, 0, p_4$ are in a straight line. Due to the rotational symmetry, $M(S; \{p_i, 0\})$ is independent of $i = 1,2,3,4$. So* $\mathrm{SCD}(S)$ *can be considered as one* $\mathrm{OCD} = [1, M(S; \begin{pmatrix} p_1 \\ 0 \end{pmatrix})]$ *of weight 1.*

Example 4.2(b) illustrates the key discontinuity challenge: if $p_4 = (-1,0)$ is perturbed, the corresponding sign can discontinuously change to $+1$ or -1. To get a continuous metric on OCDs, we will multiply each sign by a continuous *strength* function that vanishes for any zero sign.

Definition 4.3 (*strength* $\sigma(A)$ of a simplex)**.** *For a set A of $n+1$ points $q = p_0, p_1, \dots, p_n$ in $\mathbb{R}^n$, let $p(A) = \frac{1}{2}\sum\limits_{i\neq j}^{n+1} |p_i - p_j|$ be half of the sum of all pairwise distances. Let $V(A)$ denote the volume the n-dimensional simplex on the set A. Define the* strength $\sigma(A) = V^2(A)/p^{2n-1}(A)$.

For $n = 2$ and a triangle A with sides a,b,c in $\mathbb{R}^2$, Heron's formula gives $\sigma(A) = \dfrac{(p-a)(p-b)(p-c)}{p^2}$, $p = \dfrac{a+b+c}{2} = p(A)$ is the half-perimeter of A.

For $n = 1$ and a set $A = p_0, p_1 \subset \mathbb{R}$, the volume is $V(A) = |p_0 - p_1| = 2p(A)$, so $\sigma(A) = 2|p_0 - p_1|$.

The strength $\sigma(A)$ depends only on the distance matrix $D(A)$ from Definition 3.1, so the notation $\sigma(A)$ is used only for brevity. In any $\mathbb{R}^n$, the squared volume $V^2(A)$ is expressed by the Cayley-Menger determinant [60] in pairwise distances between points of A. Importantly, the strength $\sigma(A)$ vanishes when the simplex on a set A degenerates.

Theorem 4.7 will need the continuity of $s\sigma(A)$, when a sign $s \in \{\pm 1\}$ from a bottom row of ORD discontinuously changes while passing through a degenerate set A. The proof of the continuity of $\sigma(A)$ in Theorem 4.4 gives an explicit upper bound for a Lipschitz constant c_n below.

Theorem 4.4 (Lipschitz continuity of the strength σ)**.** *Let a cloud A' be obtained from another $(n+1)$-point cloud $A \subset \mathbb{R}^n$ by perturbing every point within its ε-neighborhood. The strength $\sigma(A)$ from Definition 4.3 is Lipschitz continuous so that $|\sigma(A') - \sigma(A)| \leq 2\varepsilon c_n$ for a constant c_n.*

Example 4.5 (strength $\sigma(A)$ and its upper bounds)**.** *[39, Theorem 4.2] proves upper bounds for the Lipschitz constant of the strength: $c_2 = 2\sqrt{3}$, $c_3 \approx 0.43$, $c_4 \approx 0.01$, which quickly tend to 0 due to the 'curse of dimensionality'. The plots in Fig. 4 illustrate that the strength $\sigma()$ behaves smoothly in the x-coordinate of a vertex and its derivative $|\frac{\partial\sigma}{\partial x}|$ is much smaller than the proved bounds c_n above.*

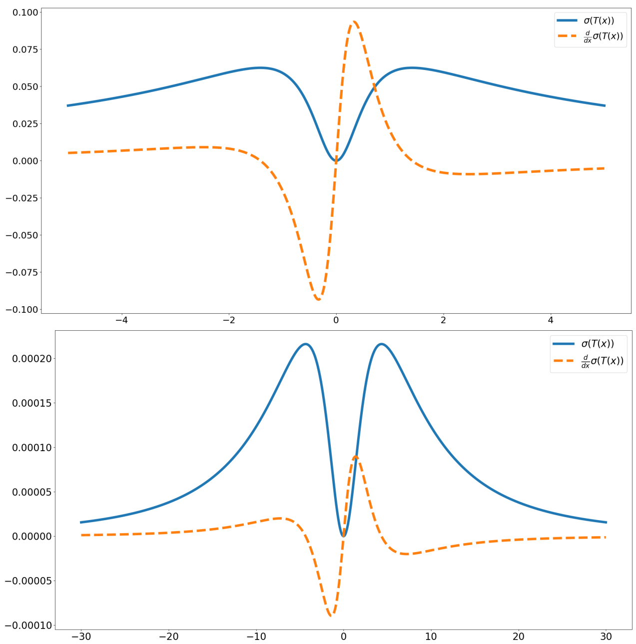

Figure 4. The strength σ (solid curve) and its derivative $\frac{\partial \sigma}{\partial x}$ (dashed curve) in the x-coordinate of a point from A were averaged over 3000 random triangles (top) and tetrahedra (bottom).

The strength $\sigma(A)$ from Definition 4.3 will take care of extra signs in ORDs and allows us to prove the analogue of Lemma 3.7 for a similar time complexity with $h = n$.

Lemma 4.6 (metric on OCDs)**.** *Using the strength σ from Definition 4.3, we consider the bottleneck distance W_∞ on the set of permutable $m-n+1$ columns of $M(C; A\cup\{0\})$ as on the set of $m-n+1$ unlabeled points $\left(v, \dfrac{s}{c_n}\sigma(A\cup\{0,q\})\right) \in \mathbb{R}^{n+1}$. For another $\mathrm{OCD}' = [D(A'\cup\{0\}); M(C'; A'\cup\{0\})]$ and any permutation $\xi \in S_{n-1}$ of indices $1,\dots,n-1$ acting on $D(A\cup\{0\})$ and the first $n-1$ rows of $M(C; A\cup\{0\})$, set $d_o(\xi) = \max\{L, W\}$,*

$$\text{where } L = L_\infty\Big(\xi(D(A\cup\{0\})), D(A'\cup\{0\})\Big),$$

$$W = W_\infty\Big(\xi(M(C; A\cup\{0\})), M(C'; A'\cup\{0\})\Big).$$

Then $M_\infty(\mathrm{OCD}, \mathrm{OCD}') = \min\limits_{\xi\in S_{n-1}} d_o(\xi)$ satisfies all metric axioms on Oriented Centered Distributions (OCDs) and is computed in time $O((n-1)!(n^2 + m^{1.5}\log^n m))$.

The coefficient $\frac{1}{c_n}$ normalizes the Lipschitz constant c_n of σ to 1 in line with changes of distances by at most 2ε when points are perturbed within their ε-neighborhoods. An equality $\mathrm{SCD}(C) = \mathrm{SCD}(C')$ is interpreted as a bijection between unordered sets $\mathrm{SCD}(C) \to \mathrm{SCD}(C')$ matching all OCDs, which is best detected by checking if metrics in Theorem 4.7 between these SCDs is 0.

Theorem 4.7 (completeness and continuity of SCD)**. *(a)*** *The Simplexwise Centered Distribution $\mathrm{SCD}(C)$ in Definition 4.1 is a complete isometry invariant of clouds $C \subset \mathbb{R}^n$ of m unlabeled points with a center of mass at the origin $0 \in \mathbb{R}^n$, and can be computed in time $O(m^n/(n-4)!)$.*

So any clouds $C, C' \subset \mathbb{R}^n$ are related by rigid motion (isometry, respectively) if and only if $\mathrm{SCD}(C) = \mathrm{SCD}(C')$ *($\mathrm{SCD}(C)$ equals $\mathrm{SCD}(C')$ or its mirror image $\overline{\mathrm{SCD}}(C')$, respectively). For any m-point clouds $C, C' \subset \mathbb{R}^n$, let $\mathrm{SCD}(C)$ and $\mathrm{SCD}(C')$ consist of $k = \binom{m}{n-1}$ OCDs.*

(b) *For the $k\times k$ matrix of costs computed by the metric M_∞ between OCDs in $\mathrm{SCD}(C)$ and $\mathrm{SCD}(C')$, LAC from Definition 3.8 satisfies all metric axioms on SCDs and needs time $O((n-1)!(n^2 + m^{1.5}\log^n m)k^2 + k^3\log k)$.*

(c) *Let SCDs have a maximum size $l \le k$ after collapsing identical OCDs. Then EMD from Definition 3.9 satisfies all metric axioms on SCDs and can be computed in time $O((n-1)!(n^2 + m^{1.5}\log^n m)l^2 + l^3\log l)$.*

(d) *Let C' be obtained from a cloud $C \subset \mathbb{R}^n$ by perturbing each point within its ε-neighborhood. Then $\mathrm{SCD}(C)$ changes by at most 2ε in the LAC and EMD metrics.*

If we estimate $l \le k = \binom{m}{n-1} = m(m-1)\dots(m-n+2)/n!$ as $O(m^{n-1}/n!)$, Theorem 4.7(b,c) gives time $O(n(m^{n-1}/n!)^3\log m)$ for metrics on SCDs, which is $O(m^3\log m)$ for $n = 2$, and $O(m^6\log m)$ for $n = 3$.

Though the above time estimates are very rough upper bounds, the time $O(m^3\log m)$ in $\mathbb{R}^2$ is faster than the only past time $O(m^5\log m)$ for comparing m-point clouds by the Hausdorff distance minimized over isometries [16].

Definition 4.8 (Centered Distance Moments CDM)**.** *For any m-point cloud $C \subset \mathbb{R}^n$, let $A \subset C$ be a subset of $n-1$ unordered points. The* Centered Interpoint Distance *list $\mathrm{CID}(A)$ is the increasing list of all $\frac{(n-1)(n-2)}{2}$ pairwise distances between points of A, followed by $n-1$ increasing distances from A to the origin 0. For each column of the $(n+1)\times(m-n+1)$ matrix $M(C; A\cup\{0\})$ in Definition 4.1, compute the average of the first $n-1$ distances. Write these averages in increasing order, append the list of increasing distances $|q-0|$ from the n-th row of $M(C; A\cup\{0\})$, and also append the vector of increasing values of $\dfrac{s}{c_n}\sigma(A\cup\{0\})$ taking signs s from the $(n+1)$-st row of $M(C; A\cup\{0\})$. Let $\vec{M}(C; A) \in \mathbb{R}^{3(m-n+1)}$ be the final vector.*

The pair $[\mathrm{CID}(A); \vec{M}(C; A)]$ is the Average Centered Vector $\mathrm{ACV}(C; A)$ *considered as a vector of length $\frac{n(n-1)}{2} + 3(m-n+1)$. The unordered set of $\mathrm{ACV}(C; A)$ for all $\binom{m}{n-1}$ unordered subsets $A \subset C$ is the Average Centered Distribution $\mathrm{ACD}(C)$. The* Centered Distance Moment *$\mathrm{CDM}(C; l)$ is the l-th (standardized for $l \ge 3$) moment of $\mathrm{ACD}(C)$ considered as a probability distribution of $\binom{m}{n-1}$ vectors, separately for each coordinate.*

Example 4.9 (CDM for clouds in Fig. 3)**.** ***(a)*** *For $n = 2$ and the cloud $R \subset \mathbb{R}^2$ of $m = 3$ vertices $p_1 = (0,0)$, $p_2 = (4,0)$, $p_3 = (0,3)$ of the right-angled triangle in Fig. 3 (middle), we continue Example 4.2(a) and flatten $\mathrm{OCD}(R;p_1) = [0, \begin{pmatrix} 4 & 3 \\ 4 & 3 \\ 0 & 0 \end{pmatrix}]$ into the vector $\mathrm{ACV}(R;p_1) = [0;3,4;3,4;0,0]$ of length $\frac{n(n-1)}{2} + 3(m-n+1) = 7$, whose four parts ($1+2+2+2=7$) are in increasing order, similarly for p_2, p_3. The Average Centered Distribution can be written as a 3×7 matrix with unordered rows:*

$$\mathrm{ACD}(R) = \left(\begin{array}{c|cc|cc|cc} 0 & 3 & 4 & 3 & 4 & 0 & 0 \\ 4 & 4 & 5 & 0 & 3 & 0 & -6/c_2 \\ 3 & 3 & 5 & 0 & 4 & 0 & 6/c_2 \end{array}\right).$$

The area of the triangle on R equals 6 and can be normalized by $c_2 = 2\sqrt{3}$ to get $6/c_2 = \sqrt{3}$, see [39, section 4]. The 1st moment is $\mathrm{CDM}(R;1) = \frac{1}{3}(7;10,14;3,11;0)$.

(b) *For $n = 2$ and the cloud $S \subset \mathbb{R}^2$ of $m = 4$ vertices of the square in Fig. 3 (right), Example 4.2(a) computed $\mathrm{SCD}(R)$ as one $\mathrm{OCD} = [1, \begin{pmatrix} \sqrt{2} & \sqrt{2} & 2 \\ 1 & 1 & 1 \\ - & + & 0 \end{pmatrix}]$, which flattens to $\mathrm{ACV} = (1;\sqrt{2},\sqrt{2},2;1,1,1;-\frac{1}{2},\frac{1}{2},0) = \mathrm{ACD}(S) = \mathrm{CDM}(S;1) \in \mathbb{R}^{10}$, where $\frac{1}{2}$ is the area of the triangle on the vertices $(0,0),(1,0),(0,1)$.*

Corollary 4.10 (time for continuous metrics on CDMs)**.** *For any cloud $C \subset \mathbb{R}^n$ of m unlabeled points, the Centered Distance Moment $\mathrm{CDM}(C;l)$ in Definition 4.8 is computed in time $O(m^n/(n-4)!)$. The metric L_∞ on CDMs needs $O(n^2+m)$ time and $\mathrm{EMD}\big(\mathrm{SCD}(C), \mathrm{SCD}(C')\big) \geq |\mathrm{CDM}(C;1) - \mathrm{CDM}(C';1)|_\infty$ holds.*

5. Experiments and discussion of future work

This paper advocates a scientific approach to any data exemplified by Problem 1.1, where rigid motion on clouds can be replaced by another equivalence of other data. The scientific principles such as axioms should be always respected. Only the first coincidence axiom in (1.1b) guarantees no duplicate data. If the triangle inequality fails with any additive error, results of clustering can be pre-determined [51].

The notorious $m!$ challenge of m *unlabeled* points in Problem 1.1 was solved in $\mathbb{R}^n$ by Theorem 4.7, also up to rigid motion by using the novel *strength* of a simplex to smooth signs of determinants due to hard Theorem 4.4.

The results above sufficiently justify re-focusing future efforts from experimental attempts at Problem 1.1 to higher level tasks such as predicting properties of rigid objects, e.g. crystalline materials, using the complete invariants with *no false negatives* and *no false positives* for all possible data since *no experiments* can beat the proved 100% guarantee.

To tackle the limitation of comparing only clouds having a fixed number m of points, the Earth Mover's Distance (EMD) continuously can compare any distributions (SDD or SCD) of different sizes. Using EMD instead of the bottleneck distance W_∞ on $m-h$ (or $m-n+1$) columns of matrices in Definitions 3.3 and 4.1 increases a time from $O(m^{1.5}\log m)$ to $O(m^3 \log m)$ but the total time remains the same due to a near cubic time in the last step.

The running time in real applications is smaller for several reasons. First, the shape (isometry class) of any rigid body in $\mathbb{R}^3$ is determined by only $m = 4$ labeled points in general position. Even when points are unlabeled, dozens of corners or feature points suffice to represent a rigid shape well enough. Second, the key size l (number of distinct Oriented Centered Distributions) in Theorem 4.7 is often smaller than m, especially for symmetric objects, see $l = 1 < m = 4$ in Example 4.2. The SCD invariants are on top of others due to their completeness and continuity.

The past work [71, 73] used the simpler Pointwise Distance Distribution (PDD) to complete 200B+ pairwise comparisons of all 660K+ periodic crystals in the world's largest database of real materials. This experiment took only a couple of days on a modest desktop and established the *Crystal Isometry Principle* saying that any real periodic crystal is uniquely determined by the geometry of its atomic centers without chemical elements. So the type of any atom is provably reconstructable from distances to atomic neighbors.

The new invariants allow us to go deeper and compare atomic clouds from higher level periodic crystals. Fig. 5 visualizes all 300K+ atomic clouds extracted from all 10K+ crystalline drugs in the Cambridge Structural Database (CSD) by using SDV invariants for $5+1$ atoms including the central one. Future maps will use stronger invariants.

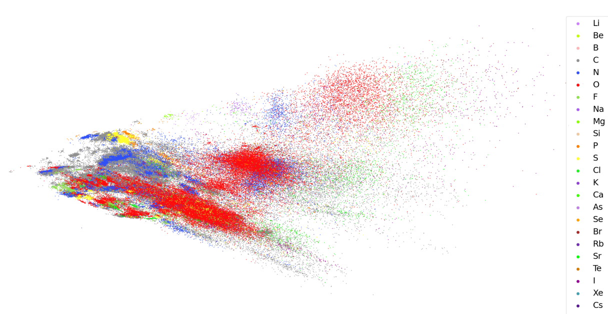

Figure 5. Two principal directions of SDVs for all 300K+ atomic clouds from all 10K+ drugs in the CSD, colored by 25 elements.

This research was supported by the Royal Academy Engineering fellowship IF2122/186, and EPSRC grants EP/R018472/1, EP/X018474/1. We thank the Data Science Theory and Applications group (Liverpool) and reviewers of this conference version [72] for 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*. 2

[3] Olga Anosova and Vitaliy Kurlin. An isometry classification of periodic point sets. In *LNCS (Proceedings of DGMM)*, volume 12708, pages 229–241, 2021. 2

[4] Olga Anosova and Vitaliy Kurlin. Algorithms for continuous metrics on periodic crystals. *arxiv:2205.15298*, 2022. 2

[5] Olga Anosova and Vitaliy Kurlin. Density functions of periodic sequences. In *LNCS (Proceedings of DGMM)*, volume 13493, pages 395–408, 2022. 2

[6] O Anosova and V Kurlin. Density functions of periodic sequences of continuous events. *arXiv:2301.05137*, 2023. 2

[7] Jonathan Balasingham, Viktor Zamaraev, and Vitaliy Kurlin. Compact graph representation of crystals using Pointwise Distance Distributions. *arXiv:2212.11246*, 2022. 2

[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. 1, 4

[10] Peter Brass and Christian Knauer. Testing congruence and symmetry for general 3-dimensional objects. *Computational Geometry*, 27(1):3–11, 2004. 2

[11] Matthew J Bright, Andrew I Cooper, and Vitaliy A Kurlin. Welcome to a continuous world of 3-dimensional lattices. *arxiv:2109.11538*, 2021. 2

[12] 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. 2

[13] Michael M Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: grids, groups, graphs, geodesics, and gauges. *arXiv:2104.13478*, 2021. 2

[14] 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

[15] 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

[16] 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, 7

[17] Paul Chew and Klara Kedem. Improvements on geometric pattern matching. In *SWAT*, pages 318–325, 1992. 2

[18] 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

[19] Yinpeng Dong, Fangzhou Liao, Tianyu Pang, Hang Su, Jun Zhu, Xiaolin Hu, and J Li. Boosting adversarial attacks with momentum. In *CVPR*, pages 9185–9193, 2018. 2

[20] 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. 2

[21] Alon Efrat, Alon Itai, and Matthew J Katz. Geometry helps in bottleneck matching and related problems. *Algorithmica*, 31(1):1–28, 2001. 5

[22] Yury Elkin. New compressed cover tree for k-nearest neighbor search (PhD thesis). *arxiv:2205.10194*, 2022. 3

[23] Y. Elkin and V. Kurlin. The mergegram of a dendrogram and its stability. In *MFCS*, pages 32:1–32:13, 2020. 2

[24] 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

[25] Michael L Fredman and Robert Endre Tarjan. Fibonacci heaps and their uses in improved network optimization algorithms. *Journal ACM*, 34:596–615, 1987. 4, 5

[26] Michael T Goodrich, Joseph SB Mitchell, and Mark W Orletsky. Approximate geometric pattern matching under rigid motions. *Transactions PAMI*, 21:371–379, 1999. 2

[27] Cosmin Grigorescu and Nicolai Petkov. Distance sets for shape filters and shape recognition. *IEEE transactions on image processing*, 12(10):1274–1286, 2003. 2

[28] Darij Grinberg and Peter J Olver. The n body matrix and its determinant. *SIAM Journal on Applied Algebra and Geometry*, 3(1):67–86, 2019. 2

[29] Chuan Guo, Jacob Gardner, Yurong You, Andrew Gordon Wilson, and Kilian Weinberger. Simple black-box adversarial attacks. In *ICML*, pages 2484–2493, 2019. 2

[30] Felix Hausdorff. Dimension und äuβeres maβ. *Mathematische Annalen*, 79(2):157–179, 1919. 2

[31] Snir Hordan, Tal Amir, Steven J Gortler, and Nadav Dym. Complete neural networks for Euclidean graphs. *arXiv:2301.13821*, 2023. 2

[32] Daniel Huttenlocher, Gregory Klanderman, and William Rucklidge. Comparing images using the Hausdorff distance. *Transactions PAMI*, 15:850–863, 1993. 2

[33] Ernest Sydney Keeping. *Introduction to statistical inference*. Courier Corporation, 1995. 3

[34] Vitaliy Kurlin. A complete isometry classification of 3-dimensional lattices. *arxiv:2201.10543*, 2022. 2

[35] Vitaliy Kurlin. Computable complete invariants for finite clouds of unlabeled points under Euclidean isometry. *arXiv:2207.08502*, 2022. 2

[36] Vitaliy Kurlin. Exactly computable and continuous metrics on isometry classes of finite and 1-periodic sequences. *arxiv:2205.04388*, 2022. 2

[37] Vitaliy Kurlin. Mathematics of 2-dimensional lattices. *Foundations of Computational Mathematics*, pages 1–59, 2022. 2

[38] Vitaliy Kurlin. Simplexwise distance distributions for finite spaces with metrics and measures. *arXiv:2303.14161*, 2023. 2, 3, 5

[39] 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. 2, 3, 6, 8

[40] Cassidy Laidlaw and Soheil Feizi. Functional adversarial attacks. *Adv. Neural Inform. Proc. Systems*, 32, 2019. 2

[41] Sushovan Majhi, Jeffrey Vitter, and Carola Wenk. Approximating Gromov-Hausdorff distance in Euclidean space. *arXiv:1912.13008*, 2019. 2

[42] Siddharth Manay, Daniel Cremers, Byung-Woo Hong, Anthony Yezzi, and Stefano Soatto. Integral invariants for shape matching. *Transactions PAMI*, 28:1602–1618, 2006. 2

[43] Facundo Mémoli. Gromov–Wasserstein distances and the metric approach to object matching. *Foundations of Computational Mathematics*, 11(4):417–487, 2011. 2

[44] Facundo Mémoli and Tom Needham. Distance distributions and inverse problems for metric measure spaces. *Studies in Applied Mathematics*, 149(4):943–1001, 2022. 2

[45] Facundo Mémoli, Zane Smith, and Zhengchao Wan. The Gromov-Hausdorff distance between ultrametric spaces: its structure and computation. *arXiv:2110.03136*, 2021. 2

[46] Marco M Mosca and Vitaliy Kurlin. Voronoi-based similarity distances between arbitrary crystal lattices. *Crystal Research and Technology*, 55(5):1900197, 2020. 2

[47] 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

[48] François Pomerleau, Francis Colas, Roland Siegwart, et al. A review of point cloud registration algorithms for mobile robotics. *Found. Trends® in Robotics*, 4:1–104, 2015. 1

[49] Helmut Pottmann, Johannes Wallner, Qi-Xing Huang, and Yong-Liang Yang. Integral invariants for robust geometry processing. *Comp. Aided Geom. Design*, 26:37–60, 2009. 2

[50] Sergey N Pozdnyakov, Michael J Willatt, Albert P Bartók, Christoph Ortner, Gábor Csányi, and Michele Ceriotti. Incompleteness of atomic structure representations. *Physical Review Letters*, 125:166001, 2020. 5

[51] Stefan Rass, Sandra König, Shahzad Ahmad, and Maksim Goman. Metricizing Euclidean space towards desired distance relations in point clouds. *arXiv:2211.03674*, 2022. 8

[52] 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

[53] 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. 2

[54] Y. Rubner, C. Tomasi, and L. Guibas. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121, 2000. 5

[55] Uwe Schmidt and Stefan Roth. Learning rotation-aware features. In *CVPR*, pages 2050–2057, 2012. 1

[56] 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

[57] Isaac Schoenberg. Remarks to Maurice Frechet's article "Sur la definition axiomatique d'une classe d'espace distances vectoriellement applicable sur l'espace de Hilbert. *Annals of Mathematics*, pages 724–732, 1935. 2

[58] Jingnan Shi, Heng Yang, and Luca Carlone. Robin: a graph-theoretic approach to reject outliers in robust estimation using invariants. In *ICRA*, pages 13820–13827, 2021. 2

[59] 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

[60] Manfred Sippl and Harold Scheraga. Cayley-Menger coordinates. *PNAS*, 83:2283–2287, 1986. 6

[61] Philip Smith and Vitaliy Kurlin. Families of point sets with identical 1D persistence,. *arxiv:2202.00577*, 2022. 2

[62] 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. 2

[63] Riccardo Spezialetti, Samuele Salti, and Luigi Di Stefano. Learning an effective equivariant 3d descriptor without supervision. In *ICCV*, pages 6401–6410, 2019. 2

[64] 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

[65] 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

[66] 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. 2

[67] Yue Wang and Justin M Solomon. Deep closest point: Learning representations for point cloud registration. In *Proceedings of CVPR*, pages 3523–3532, 2019. 1

[68] E Weisstein. Triangle. *https://mathworld. wolfram. com*. 1

[69] Carola Wenk. Shape matching in higher dimensions. *PhD thesis, FU Berlin*, 2003. 2

[70] D Widdowson and V Kurlin. Pointwise distance distributions of periodic point sets. *arxiv:2108.04798*, 2021. 2

[71] Daniel Widdowson and Vitaliy Kurlin. Resolving the data ambiguity for periodic crystals. *Advances in Neural Information Processing Systems (NeurIPS)*, 35, 2022. 2, 3, 8

[72] D Widdowson and V 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 (arxiv:2303.15385)*, 2023. 8

[73] 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. 2, 3, 8

[74] N Zava. The Gromov-Hausdorff space isn't coarsely embeddable into any Hilbert space. *arXiv:2303.04730*, 2023. 2

[75] 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. *J American Chemical Society*, 144:9893–9901, 2022. 2

[76] Wen Zhu, Lingchao Chen, Beiping Hou, Weihan Li, Tianliang Chen, and Shixiong Liang. Point cloud registration of arrester based on scale-invariant points feature histogram. *Scientific Reports*, 12(1):1–13, 2022. 2