

Complete Invariants of Atomic Clouds under Rigid Motion with Lipschitz Continuous Metrics in a Polynomial Time

Vitaliy Kurlin^{a,*}

^a*School of Computer Science and Informatics, University of Liverpool,
Liverpool L69 3BX, UK
vitaliy.kurlin@gmail.com*

(Received April 3, 2026)

Abstract

A basic representation of any real molecule is a finite cloud of unordered atoms, many of which are chemically indistinguishable. A natural equivalence on point clouds in any metric space is defined by isometries that are distance-preserving transformations.

In a Euclidean space, any isometry is a composition of translations, rotations, and reflections. If points are ordered, the isometry class of this cloud is uniquely determined by the matrix of all pairwise distances. If m points are unordered, a naive metric based on distance matrices needs exponentially many $m!$ permutations.

We define a complete invariant for n -dimensional clouds of m unordered points under rigid motion, which distinguishes all mirror images in \mathbb{R}^n . The key challenge was to design a distance on invariant values that is Lipschitz continuous under noise and computable in a polynomial time of cloud sizes, for a fixed dimension n .

*Corresponding author.

1 Introduction: clouds of unordered points

Many objects are given as collections of unordered points. For example, a molecule in \mathbb{R}^3 is often represented as a list of points at atomic centers without a canonical order. For example, the benzene molecule C_6H_6 in the first picture of Fig. 1 consists of 6 carbon atoms and 6 hydrogen atoms, which can be permuted in $(6!)^2 > 500,000$ ways. In computer vision, the ambient dimension is still small, but the size of point clouds can be much larger. Since points are often unordered in many applications, we study point *clouds* defined as finite sets of unordered (unlabelled) points in \mathbb{R}^n .

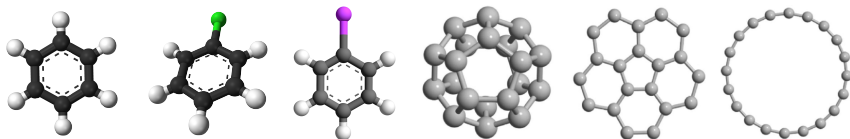


Figure 1. Benzene C_6H_6 , phenyllithium C_6H_5Li , chlorobenzene C_6H_5Cl , isomers of C_{20} have many indistinguishable atoms.

Molecules are often considered as graphs embedded in \mathbb{R}^3 with atoms connected by covalent bonds. However, these bonds are not real sticks in space, and only abstractly represent interatomic interactions that have no precise definitions but depend on many thresholds for distances and angles. Since atomic nuclei are real physical objects, a cloud of atomic centers remains the most fundamental representation of a real molecule.

In addition to the lack of point ordering, another practical challenge is the ambiguity of point representations depending on a coordinate system. Indeed, any translation or rotation changes a coordinate representation but not the rigid shape of a given cloud. Clouds $C, C' \subset \mathbb{R}^n$ are called *rigidly equivalent* (denoted by $C \cong C'$) if there is a *rigid motion* (a composition of translations and rotations) from the special Euclidean group $SE(\mathbb{R}^n)$ that maps C to C' . Rigid motion defines an equivalence relation on clouds.

An *equivalence* is any binary relation (denoted by $C \sim C'$) that satisfies the following axioms: (1) *reflexivity*: $C \sim C$, (2) *symmetry*: if $C \sim C'$, then $C' \sim C$, (3) *transitivity*: if $C \sim C' \sim C''$, then $C \sim C''$ for any given clouds. The transitivity axiom justifies any classification as a splitting into

disjoint *equivalence classes*. Any cloud C generates its class $[C] = \{C' \mid C' \sim C\}$ consisting of all C' equivalent to C . We define the *rigid shape* of any cloud $C \subset \mathbb{R}^n$ as its equivalence class under rigid motion.

Another equivalence on point clouds is defined by *isometries*, which are distance-preserving maps and form the full Euclidean group $E(\mathbb{R}^n)$. The equivalence under isometry (denoted by $C \simeq C'$) is slightly weaker than under rigid motion because any two mirror images belong to the same isometry class but may not be rigidly equivalent. Compositions of rigid motion or isometry with uniform scaling generate larger groups of transformations and hence smaller spaces of equivalence classes of m -point clouds. These spaces of classes are still high-dimensional and continuous because coordinates of given points, such as atomic centers, are real.

Rigid shapes can be distinguished by *rigid invariants* I that are descriptors preserved under any rigid motion, i.e. if $C \cong C'$, then $I(C) = I(C')$. Equivalently, if $I(C) \neq I(C')$, then $C \not\cong C'$ are not rigidly equivalent. In other words, I has *no false negatives* that are pairs of $C \cong C'$ with $I(C) \neq I(C')$. For any cloud $C \subset \mathbb{R}^n$, its number m of points is invariant, but the center of mass $O(C) = \frac{1}{m} \sum_{p \in C} \mathbf{p}$ is not invariant. Here $\mathbf{p} \in \mathbb{R}^n$ denotes the position vector from the origin $0 \in \mathbb{R}^n$ to any point $p \in \mathbb{R}^n$.

The number of points does not distinguish any non-equivalent clouds of the same size. A rigid invariant I is *complete* if the converse implication holds: any non-equivalent clouds $C \not\cong C'$ have $I(C) \neq I(C')$. In other words, I has *no false positives* that are pairs $C \not\cong C'$ with $I(C) = I(C')$.

Recall that the *affine dimension* $\text{aff}(C)$ of a cloud C of points p_1, \dots, p_m in \mathbb{R}^n is the maximum dimension of the vector space generated by all inter-point vectors $\mathbf{p}_i - \mathbf{p}_j$, $i, j \in \{1, \dots, m\}$. Then $\text{aff}(C)$ is an isometry invariant and is independent of an order of points of C . Any cloud $C \subset \mathbb{R}^n$ of 2 distinct points has $\text{aff}(C) = 1$. Any cloud $C \subset \mathbb{R}^n$ of 3 points that are not in the same straight line has $\text{aff}(C) = 2$. In the sequel, we assume that any cloud $C \subset \mathbb{R}^n$ is *n-dimensional* in the sense that $\text{aff}(C) = n$.

A complete invariant answers the first key question (*same or different?*) for objects under a given equivalence. Since atomic coordinates are

always uncertain due to thermal vibrations or experimental noise, the next practical question (*if different, by how much?*) requires the concept of a distance metric [12] and continuity, which we formalize in Problem 1.1.

Problem 1.1 (complete and continuous invariants of unordered clouds). Design an invariant I on all n -dimensional clouds of m unordered points in \mathbb{R}^n with values in a metric space to satisfy the following conditions.

(a) *Completeness*: any n -dimensional clouds $C, C' \subset \mathbb{R}^n$ are related by rigid motion (denoted by $C \cong C'$) if and only if $I(C) = I(C')$.

(b) *Reconstruction*: any n -dimensional cloud $C \subset \mathbb{R}^n$ of unordered points can be reconstructed from $I(C)$, uniquely under rigid motion.

(c) *Metric*: there is a metric d on invariant values satisfying all axioms:

(1) $d(I(C), I(C')) = 0$ if and only if $C \cong C'$ are related by rigid motion;

(2) *symmetry* : $d(I(C), I(C')) = d(I(C'), I(C))$ for any clouds C, C' ;

(3) $d(I(C), I(C')) + d(I(C'), I(C'')) \geq d(I(C), I(C''))$ for any C, C', C'' .

(d) *Lipschitz continuity*: there is a constant $\lambda > 0$ such that, for any $\varepsilon > 0$, if a cloud C' is obtained by perturbing every point of a cloud $C \subset \mathbb{R}^n$ up to Euclidean distance ε , then $d(I(C), I(C')) \leq \lambda\varepsilon$.

(e) *Computability*: for a fixed dimension n , the invariant $I(C)$, metric $d(I(C), I(C'))$, and a reconstruction of a cloud $C \subset \mathbb{R}^n$ from $I(C)$ can be found in a polynomial time of the maximum size of any given clouds C, C' .

A reconstruction in 1.1(b) is stronger than completeness in 1.1(a) because a complete invariant $I(C)$ can be too abstract and may not allow one to reconstruct $C \subset \mathbb{R}^n$. In biology, a genetic code is considered complete in practice but does not allow yet a reconstruction of a living organism.

The metric axioms in 1.1(c) are important because they are basic requirements for proofs in metric geometry [19]. If the first axiom fails, even the zero distance $d \equiv 0$ satisfies the other two axioms. If the triangle axiom fails or approximately holds with any small additive error, then the k -means and DBSCAN clustering are open to adversarial attacks [43].

The first metric axiom in 1.1(c) implies the completeness of I in 1.1(a), so the equality $I(C) = I(C')$ is can be checked by verifying if $d(I(C), I(C'))$

$= 0$. For any complete invariant I , one can define the discrete metric that takes the value of (say) 1 for any clouds $C \not\cong C'$, which unhelpfully treats all different clouds (even near-duplicates) as equally distant.

The Lipschitz continuity in 1.1(d) is much stronger than the classical $\varepsilon - \delta$ continuity because the Lipschitz constant λ should be independent of a cloud C and a threshold ε . Indeed, $\frac{1}{x}$ is theoretically continuous for all $x > 0$, but its behaviour for $x \rightarrow 0$ is an example of explosive growth.

The computability in 1.1(e) prevents brute-force attempts to define a complete but rather impractical invariant $I(C)$ by taking the infinite set of images of C under all possible rigid motions or by minimizing a distance d between infinitely many alignments of given clouds $C, C' \subset \mathbb{R}^n$. All computational complexities are considered in the Random Access Memory (RAM) model, where any value can be accessed in a constant time.

The key contribution is the new invariant that completely solves exponentially hard Problem 1.1 for all n -dimensional cloud of unordered points.

2 Past work on invariants of point clouds

This section reviews related approaches with theoretical guarantees.

The case of ordered points was resolved more than 90 years ago by reconstructing any sequence $p_1, \dots, p_m \in \mathbb{R}^n$ (uniquely under Euclidean isometry) from the matrix of Euclidean distances $|p_i - p_j|$ or, equivalently, the Gram matrix of scalar products $\mathbf{p}_i \cdot \mathbf{p}_j$. Then the difference between $m \times m$ matrices of distances (or) can be converted into a continuous metric by taking a matrix norm. In this ordered case, Kendall's shape theory [29] described shape spaces Σ_m^n of m -point sequences under isometry in \mathbb{R}^n .

In the unordered case, comparing $m \times m$ matrices needs an exponential number $m!$ of permutations acting on m given points. This action by $m!$ permutations makes invariants and metrics $m!$ times more computationally challenging in Problem 1.1 than for sequences of ordered points. Invariants based on principal directions are continuous in general position [33] but degenerate cases, such as 3 points in a line, can lead to discontinuities.

Computational geometry studied many metrics between fixed point clouds, which can be theoretically minimized under isometry [26].

Definition 2.1 (Minkowski, Hausdorff and bottleneck distances). **(a)** For any real parameter $q \in [1, +\infty)$ and vectors $\mathbf{u} = (u_1, \dots, u_n)$ and $\mathbf{v} = (v_1, \dots, v_n)$ in \mathbb{R}^n , the *Minkowski* distance is $L_q(\mathbf{u}, \mathbf{v}) = \left(\sum_{i=1}^n |u_i - v_i|^q \right)^{1/q}$. In the limit case $q = +\infty$, set $L_\infty(\mathbf{u}, \mathbf{v}) = \max_{i=1, \dots, n} |u_i - v_i|$.

(b) Let A, B be sets in a space with a metric d . The distance from $p \in A$ to B is $d(p, B) = \inf_{q \in B} d(p, q)$. Set $d_H(A, B) = \sup_{p \in A} d(p, B)$. The *Hausdorff* distance is $\text{HD}(A, B) = \max\{d_H(A, B), d_H(B, A)\}$. The *bottleneck distance* $\text{BD}(A, B) = \inf_{g: A \rightarrow B} \sup_{p \in A} d(g(p), p)$ is minimized for all bijections $g: A \rightarrow B$. If no bijections $g: A \rightarrow B$ exist, set $\text{BD}(A, B) = +\infty$. ■

The Hausdorff distance between clouds of up to m points, minimized under rigid motion in \mathbb{R}^2 , is computable in time $O(m^5 \log m)$ [13]. Approximate methods for matching clouds, such as Scale Invariant Feature Transform, sample rotations from the infinite group $\text{SO}(\mathbb{R}^n)$ or use a basis [47], which can be unstable for equal inter-point distances. The Gromov-Hausdorff distances for metric-measure spaces [39] are defined via infinitely many correspondences between points, but cannot be approximated with a factor less than 3 in polynomial time unless $\text{P}=\text{NP}$ [45, Corollary 3.8], see polynomial-time algorithms in partial cases [38, 40].

Persistent homology [20] can be considered an isometry invariant of point clouds for filtrations of Vietoris-Rips, Cech, or Delaunay complexes [8]. The persistence is Lipschitz continuous [16] in the bottleneck distance, as in condition 1.1(c), but its fastest versions in dimensions 0 and 1 do not distinguish generic families of non-isometric clouds [46]. Comparing such infinite-size invariants requires sampling the infinite group $\text{O}(\mathbb{R}^n)$.

Problem 1.1 is fully solved for $m = 3$ points by the side-side-side theorem saying that any triangles are congruent (isometric) if and only if they have the same triple of sides. In 2004, Boutin and Kemper [9] proved in the unordered case that all sorted pairwise distances in Definition 2.2 uniquely determine almost any (generic) cloud $C \subset \mathbb{R}^n$ under isometry.

Definition 2.2 (Sorted Pairwise Distances SPD). For any cloud C of m unordered points, the vector $\text{SPD}(C)$ of *Sorted Pairwise Distances* has $\frac{m(m-1)}{2}$ distances between points of C , written in increasing order. ■

The invariant $\text{SPD}(C)$ is incomplete already for $m = 4$ points in \mathbb{R}^2 , see the non-isometric 4-point clouds $T \not\cong K$ in the two middle pictures of Fig. 2, which have $\text{SPD}(T) = [\sqrt{2}, \sqrt{2}, 2, \sqrt{10}, \sqrt{10}, 4] = \text{SPD}(K)$.

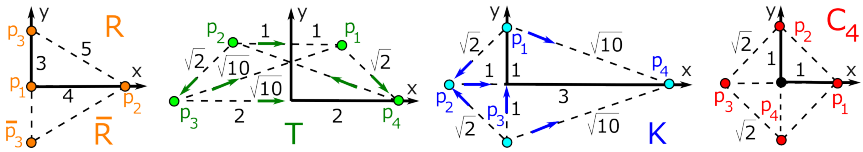


Figure 2. **1st:** the right-angled cloud $R \subset \mathbb{R}^2$ consists of points $p_1 = (0, 0)$, $p_2 = (4, 0)$, $p_3 = (0, 3)$, and its mirror image \bar{R} of p_1, p_2 , and $\bar{p}_3 = (0, -3)$ with respect to the x -axis. **2nd:** the trapezoid cloud $T \subset \mathbb{R}^2$ consists of points $p_1 = (1, 1)$, $p_2 = (-1, 1)$, $p_3 = (-2, 0)$, $p_4 = (2, 0)$. **3rd:** the kite cloud $K \subset \mathbb{R}^2$ consists of points $p_1 = (0, 1)$, $p_2 = (-1, 0)$, $p_3 = (0, -1)$, $p_4 = (3, 0)$. **4th:** the square cloud $S \subset \mathbb{R}^2$ consists of $p_1 = (1, 0)$, $p_2 = (0, 1)$, $p_3 = (-1, 0)$, $p_4 = (0, -1)$.

Problem 1.1 was solved for $m = 4$ unordered points under isometry by the stronger invariant PDD in Definition 2.3 by [52, Theorem 5.3], but the case of $m \geq 5$ unordered points remained open even in the plane \mathbb{R}^2 .

Definition 2.3 (Pointwise Distance Distribution $\text{PDD}(C; k)$). In a metric space, let C be any finite cloud of unordered points p_1, \dots, p_m , which are labeled here only for convenience. Fix an integer $1 \leq k < m$. For every point $p_i \in C$, let $d_1(p_i) \leq \dots \leq d_k(p_i)$ be ordered distances from p_i to its k nearest neighbors in C . The matrix $D(C; k)$ has m rows consisting of the distances $d_1(p_i), \dots, d_k(p_i)$ for $i = 1, \dots, m$. If any $l \geq 2$ rows coincide, we collapse them into a single row with the weight $\frac{l}{m}$. The resulting distribution of unordered rows, each having a weight and k ordered distances, is called the *Pointwise Distance Distribution* $\text{PDD}(C; k)$ [50]. ■

If we consider only point clouds in a general position away from singular configurations, one can order all points by their distinct distances to the center of mass. In this general position, Problem 1.1 reduces to the ordered case but remains exponentially hard for degenerate clouds, see Fig. 3.

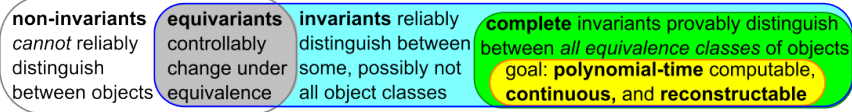


Figure 3. Non-invariant descriptors, such as point coordinates, are unreliable for comparisons under isometry. Any isometry f changes equivariant descriptors, such as the center of mass, by the same f . Only invariants preserved under all equivalences can classify objects. New invariants allow reconstructions and polynomial-time Lipschitz continuous metrics.

3 Oriented Simplexwise Distribution (OSD)

This section starts solving Problem 1.1 by introducing the Oriented Simplexwise Distribution (OSD) in Definition 3.2 after a few auxiliary concepts below. For any ordered vectors $\mathbf{v}_1, \dots, \mathbf{v}_n \in \mathbb{R}^n$, let $\text{sign}(\mathbf{v}_1, \dots, \mathbf{v}_n)$ denote the sign of the $n \times n$ determinant with the column vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$, which can be ± 1 or 0 (if the vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ are linearly dependent).

Definition 3.1 (matrix $D(A)$ and cloud $M(C; A) \subset \mathbb{R}^{n+1}$ for $A \subseteq C$). Let C be any n -dimensional cloud of m unordered points in \mathbb{R}^n . For a *base sequence* A of n ordered points $p_1, \dots, p_n \in C$, let $D(A)$ be the $n \times n$ matrix of distances $|\mathbf{p}_i - \mathbf{p}_j|$ for $i, j = 1, \dots, n$. For any point $q \in C \setminus A$, set $s(q; A) = \text{sign}(\mathbf{q} - \mathbf{p}_1, \dots, \mathbf{q} - \mathbf{p}_n)$. Let $d(q) \in \mathbb{R}^{n+1}$ be the point with the coordinates $|\mathbf{q} - \mathbf{p}_1|, \dots, |\mathbf{q} - \mathbf{p}_n|, s(q; A)$. Let $M(C; A) \subset \mathbb{R}^{n+1}$ be the cloud of $m - n$ unordered points $d(q)$ for all points $q \in C \setminus A$. ■

Since all points $d(q) \in M(C; A) \subset \mathbb{R}^{n+1}$ have fixed coordinates, it will be convenient to represent the cloud $M(C; A)$ by a matrix with $m - n$ permutable columns and $n + 1$ ordered rows. Let S_n denote the symmetric group on the indices $1, \dots, n$. Any permutation $\xi \in S_n$ is a composition of some t transpositions $i \leftrightarrow j$ and has $\text{sign}(\xi) = (-1)^t$.

Definition 3.2 (oriented distributions $\text{ORD}(C; A)$ and $\text{OSD}(C)$ for a cloud $C \subset \mathbb{R}^n$). In the notations of Definition 3.1, any permutation $\xi \in S_n$ acts on the matrix $D(A)$ by mapping $D(A)_{i,j}$ to $D(A)_{\xi(i), \xi(j)}$ and on the cloud $M(C; A)$ of points $d(q)$ for $q \in C \setminus A$ by permuting the first n coordinates of $d(q)$ and multiplying the $(n + 1)$ -st coordinate $s(q; A)$ by

sign(ξ). The *Oriented Relative Distribution* $\text{ORD}(C; A)$ is the equivalence class of the pair $[D(A); M(C; A)]$ under all permutations $\xi \in S_n$, which act on both $D(A)$ and $M(C; A)$. The *Oriented Simplexwise Distribution* $\text{OSD}(C)$ is the unordered collection of $\text{ORD}(C; A)$ for all $\binom{m}{n}$ unordered subsets $A \subseteq C$ of n points. The *mirror images* $\overline{\text{ORD}}(C; A)$ and $\overline{\text{OSD}}(C)$ are obtained from $\text{ORD}(C; A)$ and $\text{OSD}(C)$, respectively, by reversing the signs of $(n + 1)$ -st coordinates of points $d(q) \in M(C; A)$ for $q \in C \setminus A$. ■

Example 3.3 (OSD for clouds in Fig. 2). **(a)** In \mathbb{R}^2 with the counter-clockwise orientation, the right-angled cloud R consisting of the vertices $p_1 = (0, 0)$, $p_2 = (4, 0)$, $p_3 = (0, 3)$ of the triangle in Fig. 2 (1st) has

$$\text{the distribution } \text{OSD}(R) \text{ consisting of } \text{ORD}(R; (p_1, p_2)) = \left[4, \begin{pmatrix} 3 \\ 5 \\ +1 \end{pmatrix} \right],$$

$$\text{ORD}(R; (p_2, p_3)) = \left[5, \begin{pmatrix} 4 \\ 3 \\ +1 \end{pmatrix} \right], \text{ORD}(R; (p_3, p_1)) = \left[3, \begin{pmatrix} 5 \\ 4 \\ +1 \end{pmatrix} \right].$$

If we swap the points $p_1 \leftrightarrow p_3$, the last ORD above changes to the

$$\text{equivalent form } \text{ORD}(R; (p_1, p_3)) = \left[3, \begin{pmatrix} 4 \\ 5 \\ - \end{pmatrix} \right], \text{ without affecting others.}$$

If we reflect R across the x -axis, the mirror image \bar{R} of $p_1, p_2, \bar{p}_3 =$

$$(0, -3) \text{ has } \text{OSD}(\bar{R}) = \overline{\text{OSD}}(R) \text{ with } \text{ORD}(\bar{R}; (p_1, p_2)) = \left[4, \begin{pmatrix} 3 \\ 5 \\ -1 \end{pmatrix} \right],$$

$$\text{ORD}(\bar{R}; (p_2, \bar{p}_3)) = \left[5, \begin{pmatrix} 4 \\ 3 \\ - \end{pmatrix} \right], \text{ORD}(\bar{R}; (p'_3, p_1)) = \left[3, \begin{pmatrix} 5 \\ 4 \\ -1 \end{pmatrix} \right],$$

which differs from $\text{OSD}(R)$ even if we swap the points in each pair.

(b) Table 1 shows the Oriented Simplexwise Distributions for the clouds T, K in Fig. 2. Each row has similar ORDs with **highlighted** differences.

Lemma 3.4 provides a criterion for a matrix to be realizable by squared distances between any m ordered points in \mathbb{R}^n .

ORDs in OSD(T)	ORDs in OSD(K)
$\left[\sqrt{2}, \begin{pmatrix} 2 & \sqrt{10} \\ \sqrt{10} & 4 \\ -1 & -1 \end{pmatrix} \right]$	$\left[\sqrt{2}, \begin{pmatrix} 2 & \sqrt{10} \\ \sqrt{2} & 4 \\ -1 & -1 \end{pmatrix} \right]$
$\left[\sqrt{2}, \begin{pmatrix} 2 & \sqrt{10} \\ \sqrt{10} & 4 \\ +1 & +1 \end{pmatrix} \right]$	$\left[\sqrt{2}, \begin{pmatrix} 2 & \sqrt{10} \\ \sqrt{2} & 4 \\ +1 & +1 \end{pmatrix} \right]$
$\left[2, \begin{pmatrix} \sqrt{2} & \sqrt{10} \\ \sqrt{10} & \sqrt{2} \\ -1 & -1 \end{pmatrix} \right]$	$\left[2, \begin{pmatrix} \sqrt{2} & \sqrt{10} \\ \sqrt{2} & \sqrt{10} \\ -1 & +1 \end{pmatrix} \right]$
$\left[\sqrt{10}, \begin{pmatrix} \sqrt{2} & 4 \\ 2 & \sqrt{2} \\ +1 & -1 \end{pmatrix} \right]$	$\left[\sqrt{10}, \begin{pmatrix} \sqrt{2} & 2 \\ 4 & \sqrt{10} \\ -1 & -1 \end{pmatrix} \right]$
$\left[\sqrt{10}, \begin{pmatrix} \sqrt{2} & 4 \\ 2 & \sqrt{2} \\ -1 & +1 \end{pmatrix} \right]$	$\left[\sqrt{10}, \begin{pmatrix} \sqrt{2} & 2 \\ 4 & \sqrt{10} \\ +1 & +1 \end{pmatrix} \right]$
$\left[4, \begin{pmatrix} \sqrt{2} & \sqrt{10} \\ \sqrt{10} & \sqrt{2} \\ +1 & +1 \end{pmatrix} \right]$	$\left[4, \begin{pmatrix} \sqrt{2} & \sqrt{2} \\ \sqrt{10} & \sqrt{10} \\ +1 & -1 \end{pmatrix} \right]$

Table 1. The Oriented Simplexwise Distributions OSDs from Definition 3.2 for the 4-point clouds $T, K \subset \mathbb{R}^2$ in Fig. 2.

Lemma 3.4 (realization of distances). **(a)** A symmetric $m \times m$ matrix of $s_{ij} \geq 0$ with $s_{ii} = 0$ is realizable as a matrix of squared distances between $p_0 = 0, p_1, \dots, p_{m-1} \in \mathbb{R}^n$ for some n if and only if the $(m - 1) \times (m - 1)$ matrix G of $g_{ij} = \frac{s_{0i} + s_{0j} - s_{ij}}{2}$ has only non-negative eigenvalues.

(b) If G has only non-negative eigenvalues, then $\text{aff}(0, p_1, \dots, p_{m-1})$ equals the number $k \leq m - 1 \leq n$ of positive eigenvalues of G . In this case, $g_{ij} = p_i \cdot p_j$ define the *Gram matrix* of the vectors $\mathbf{p}_1, \dots, \mathbf{p}_{m-1} \in \mathbb{R}^n$, which are uniquely determined in time $O(m^3)$ under a map from $O(\mathbb{R}^n)$.

Proof. We extend [17, Theorem 1] for $m \leq n$ and find points p_1, \dots, p_{m-1} in \mathbb{R}^n in time $O(m^3)$, uniquely under an orthogonal map from $O(\mathbb{R}^n)$.

The part *only if* \Rightarrow . Let a symmetric matrix S consist of squared distances between the points $p_0 = 0, p_1, \dots, p_{m-1} \in \mathbb{R}^n$. For $i, j = 1, \dots, m-1$, $g_{ij} = \frac{s_{0i} + s_{0j} - s_{ij}}{2} = \frac{|\mathbf{p}_i|^2 + |\mathbf{p}_j|^2 - |\mathbf{p}_i - \mathbf{p}_j|^2}{2} = \mathbf{p}_i \cdot \mathbf{p}_j$ form the Gram matrix, which can be written as $G = P^T P$, where the columns of the $n \times (m-1)$ matrix P are the vectors $\mathbf{p}_1, \dots, \mathbf{p}_{m-1}$. For any vector $\mathbf{v} \in \mathbb{R}^{m-1}$, we have $0 \leq |\mathbf{P}\mathbf{v}|^2 = (\mathbf{P}\mathbf{v})^T (\mathbf{P}\mathbf{v}) = \mathbf{v}^T (\mathbf{P}^T \mathbf{P}) \mathbf{v} = \mathbf{v}^T \mathbf{G} \mathbf{v}$. Since the quadratic form $\mathbf{v}^T \mathbf{G} \mathbf{v} \geq 0$ for any $\mathbf{v} \in \mathbb{R}^{m-1}$, the matrix G is positive semi-definite, i.e. G has non-negative eigenvalues [25, Theorem 7.2.7].

The part *if* \Leftarrow . For any positive semi-definite matrix G , there is an orthogonal matrix Q such that $Q^T G Q = D$ is the diagonal matrix, whose $m-1$ diagonal elements are non-negative eigenvalues of G . The diagonal matrix \sqrt{D} consists of the square roots of the eigenvalues of G . The number of positive eigenvalues of G equals the dimension $k = \text{aff}(\{0, p_1, \dots, p_{m-1}\})$ of the subspace that is linearly spanned by $\mathbf{p}_1, \dots, \mathbf{p}_{m-1}$.

We may assume that all $k \leq n$ positive eigenvalues of G correspond to the first k coordinates of \mathbb{R}^n . Since $Q^T = Q^{-1}$, the matrix $G = Q D Q^T = (Q \sqrt{D})(Q \sqrt{D})^T$ becomes the Gram matrix of the columns of $Q \sqrt{D}$. These columns become the reconstructed points $p_1, \dots, p_{m-1} \in \mathbb{R}^n$.

If there is another diagonalization $\tilde{Q}^T G \tilde{Q} = \tilde{D}$ for $\tilde{Q} \in O(\mathbb{R}^n)$, then \tilde{D} differs from D by a permutation of eigenvalues, which is realized by an orthogonal map, so we set $\tilde{D} = D$. Then $G = \tilde{Q} D \tilde{Q}^T = (\tilde{Q} \sqrt{D})(\tilde{Q} \sqrt{D})^T$ is the Gram matrix of the columns of $\tilde{Q} \sqrt{D}$. The new columns differ from the previously reconstructed points $p_1, \dots, p_{m-1} \in \mathbb{R}^n$ by the orthogonal map $Q \tilde{Q}^T$. Hence the reconstruction is unique under orthogonal transformations from $O(\mathbb{R}^n)$. Computing $\mathbf{p}_1, \dots, \mathbf{p}_{m-1}$ as eigenvectors needs a diagonalization of G in time $O(m^3)$, see [42, section 11.5]. ■

Lemma 3.5(a) holds for all clouds, including degenerate ones, e.g., for 3 points in a straight line. Any $n-1$ points have $\text{aff}(\{p_1, \dots, p_{n-1}\}) \leq n-2$. For example, any two distinct points in $A \subset \mathbb{R}^3$ generate a straight line.

Lemma 3.5 (sequence reconstruction). **(a)** Any sequence of ordered points p_1, \dots, p_m in \mathbb{R}^n can be reconstructed (uniquely under isometry) from the matrix of the Euclidean distances $|\mathbf{p}_i - \mathbf{p}_j|$ in time $O(m^3)$. If all

distances are divided by $R = \max_{i=1, \dots, m} |\mathbf{p}_i|$, the reconstruction of points p_1, \dots, p_m is unique under isometry and uniform scaling in \mathbb{R}^n .

(b) If $m \leq n$, the uniqueness of reconstructions in part (a) remains true if we replace isometry with rigid motion in \mathbb{R}^n .

Proof. (a) By translation, we can fix the point p_1 at the origin $0 \in \mathbb{R}^n$. Let G be the $(m-1) \times (m-1)$ matrix $g_{ij} = \frac{|\mathbf{p}_i|^2 + |\mathbf{p}_j|^2 - |\mathbf{p}_i - \mathbf{p}_j|^2}{2} = \mathbf{p}_i \cdot \mathbf{p}_j$, where $i, j = 2, \dots, m$, which is obtained from the squared distances between the points $p_1 = 0, p_2, \dots, p_m$.

By Lemma 3.4 if G has $k \leq n$ positive eigenvalues, then $p_1 = 0, \dots, p_m$ can be uniquely determined under isometry in $\mathbb{R}^k \subset \mathbb{R}^n$ in time $O(m^3)$. If all distances are divided by the same radius $R = \max_{i=1, \dots, m} |\mathbf{p}_i|$, the construction guarantees uniqueness under isometry and uniform scaling.

(b) If $m \leq n$, any mirror images of $p_1, \dots, p_m \in \mathbb{R}^n$, after a suitable rigid motion, can be assumed to belong to an $(n-1)$ -dimensional hyperspace $H \subset \mathbb{R}^n$, where they are matched by a mirror reflection $H \rightarrow H$ with respect to an $(n-2)$ -dimensional subspace $S \subset H$, which is realized by the 180° orientation-preserving rotation around S . ■

Lemma 3.5(b) for $m = n = 3$ implies that any triangle is determined by its sides, uniquely under rigid motion in \mathbb{R}^3 . For example, sides 3, 4, 5 define a right-angled triangle whose mirror images are not related by rigid motion within a plane $H \subset \mathbb{R}^3$, but are matched by a rigid motion in H and a 180° rotation of \mathbb{R}^3 around a line in the plane H .

Lemma 3.6 below will help guarantee the reconstruction in Problem 1.1.

Lemma 3.6 (reconstruction from ORD). Any n -dimensional cloud $C \subset \mathbb{R}^n$ of m unordered points can be reconstructed, uniquely under rigid motion, from $\text{ORD}(C; A)$ for any base sequence $A \subseteq C$ with $\text{aff}(A) = n-1$.

Proof. By Lemma 3.5(b), any base sequence $A \subseteq C$ of n points can be reconstructed, uniquely under rigid motion in \mathbb{R}^n , from the distance matrix $D(A)$ in Definition 3.2. We may assume that the n points p_1, \dots, p_n of $A \subseteq C$ span the subspace of the first $n-1$ coordinate axes of \mathbb{R}^n .

We will prove that any point $q \in C \setminus A$ has a unique location in \mathbb{R}^n , determined by the n distances $|\mathbf{q} - \mathbf{p}_1|, \dots, |\mathbf{q} - \mathbf{p}_n|$ written in a column of the matrix $\text{ORD}(C; A)$. Since the points of A do not belong to any $(n - 1)$ -dimensional affine subspace of \mathbb{R}^n , the n spheres $S(p_i; |\mathbf{q} - \mathbf{p}_i|)$ with radii $|\mathbf{q}_i - \mathbf{p}_i|$ and centers p_i , $i = 1, \dots, n$, contain q and their full intersections consists of one or two points. We can uniquely choose q among these two options due to the sign of the determinant on the column vectors $\mathbf{q} - \mathbf{p}_1, \dots, \mathbf{q} - \mathbf{p}_n$ in the bottom row of $\text{ORD}(C; A)$. ■

Lemma 3.6 implies that $\text{ORD}(C; A)$ can have identical columns only for degenerate subsets $A \subseteq C$ with $\text{aff}(A) < n - 1$. For example, let $n = 3$ and A consist of three points p_1, p_2, p_3 in the same straight line $L \subset \mathbb{R}^3$. The three distances $|\mathbf{q} - \mathbf{p}_i|$, $i = 1, 2, 3$, to any other point $q \in C$ outside the line L define three spheres $S(p_i; |\mathbf{q} - \mathbf{p}_i|)$ that share a common circle in \mathbb{R}^3 , so the position of q is not uniquely determined in this case.

Though one $\text{ORD}(C; A)$ with $\text{aff}(A) = n - 1$ suffices to reconstruct $C \subset \mathbb{R}^n$ uniquely under rigid motion, the dependence on A requires us to consider the Oriented Simplexwise Distribution $\text{OSD}(C)$ of $\text{ORD}(C; A)$ for all n -point subsets $A \subseteq C$ to get a complete invariant in Theorem 3.7. Equality $\text{OSD}(C) = \text{OSD}(C')$ between unordered distributions is interpreted as a bijection $\text{OSD}(C) \rightarrow \text{OSD}(C')$ matching all ORDs.

Theorem 3.7 (time and completeness of OSD under rigid motion). **(a)** For any n -dimensional cloud $C \subset \mathbb{R}^n$ of m unordered points, the Oriented Simplexwise Distribution $\text{OSD}(C)$ can be computed in time $O(m^{n+1})$.

(b) Any n -dimensional clouds $C, C' \subset \mathbb{R}^n$ of m unordered points can be exactly matched by rigid motion *if and only if* $\text{OSD}(C) = \text{OSD}(C')$.

(c) Any n -dimensional clouds $C, C' \subset \mathbb{R}^n$ of m unordered points can be exactly matched by isometry *if and only if* $\text{OSD}(C)$ equals $\text{OSD}(C')$ or its mirror image $\overline{\text{OSD}(C')}$.

Proof. **(a)** To compute $\text{OSD}(C)$, we consider $\binom{m}{n} = \frac{m!}{n!(m-n)!}$ subsets $A \subseteq C$ of n points. For each fixed A , the matrix $D(A)$ of $\frac{n(n-1)}{2}$ pairwise distances needs $O(n^2)$ time. The Oriented Relative Distribution

$\text{ORD}(C; A) = [D(A); M(C; A)]$ includes $n(m - n)$ distances and $m - n$ signs, where each sign requires determinant computations in time $O(n^3)$ by Gaussian elimination. So $\text{ORD}(C; A)$ is computed in time $O(n^3 m)$. Multiplying the last time by the number $\binom{m}{n} = \frac{m!}{n!(m-n)!}$ of n -point subsets $A \subseteq C$, we estimate the final time for $\text{OSD}(C)$ as

$$O\left(\frac{m!n^3m}{n!(m-n)!}\right) = O\left(\frac{m^2(m-1)\dots(m-n+1)n^2}{(n-1)!}\right) = O(m^{n+1}).$$

(b) Part *if* \Leftarrow . Any bijection $\text{OSD}(C) \rightarrow \text{OSD}(C')$ matches $\text{ORD}(C; A)$ with $\text{ORD}(C'; A')$ for some base sequences $A \subseteq C$ and $A' \subseteq C'$. Any n -dimensional cloud $C \subset \mathbb{R}^n$ has a base sequence A with $\text{aff}(A) = n - 1$. By Lemma 3.6 an equality $\text{ORD}(C; A) = \text{ORD}(C'; A')$ for base sequences A, A' with $\text{aff} = n - 1$ guarantees that C, C' are related by rigid motion.

Part *only if* \Rightarrow . Any rigid motion f of \mathbb{R}^n bijectively maps C to $f(C) = C'$ and hence induces a bijection $\text{OSD}(C) \rightarrow \text{OSD}(C')$.

(c) follows from part (b). Indeed, any orientation-reversing isometry $C \rightarrow C'$ induces a bijection $\text{OSD}(C) \rightarrow \overline{\text{OSD}}(C') = \text{OSD}(\bar{C}')$. Conversely, $\text{OSD}(C) = \overline{\text{OSD}}(C')$ implies that $\text{ORD}(C; A) = \text{ORD}(\bar{C}'; A')$ for some base sequences $A \subseteq C$ and $A' \subset \bar{C}'$, where \bar{C}' is a mirror image of C' . Then C, \bar{C}' are related by rigid motion due to Lemma 3.6, so $C \simeq C'$. ■

For a fixed dimension n , the time $O(m^{n+1})$ is polynomial in the number m of points and will be improved to $O(m^n)$ for a smaller invariant below.

4 Simplexwise Centered Distribution (SCD)

This section simplifies the OSD invariant to the Simplexwise Centered Distribution (SCD) in Definition 4.1. The Euclidean structure of \mathbb{R}^n allows us to translate the *center of mass* $O(C) = \frac{1}{m} \sum_{p \in C} \mathbf{p}$ 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 the invariance under rotations from the special orthogonal group $\text{SO}(\mathbb{R}^n)$.

Definition 3.2 introduced the Oriented Simplexwise Distribution $\text{OSD}(C)$ as an ordered collection of ORDs depending on n -point subsets

of C . Including the center of mass $O(C)$ allows us to consider the smaller number of $\binom{m}{n-1}$ subsets $A \subseteq C$ consisting of $n-1$ points instead of n points.

Though the center of mass is uniquely determined for any cloud $C \subset \mathbb{R}^n$ of unordered points, real applications may offer one or several labelled 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 can be an extra base point in all constructions below.

BD(A, B)	Bottleneck Distance between sets A, B	Definition 2.1
HD(A, B)	Hausdorff Distance between sets A, B	Definition 2.1
OCD($C; A$)	Oriented Centered Distribution for $A \subset C$	Definition 4.1
ORD($C; A$)	Oriented Relative Distribution for $A \subset C$	Definition 3.2
OSD(C)	Oriented Simplexwise Distribution of C	Definition 3.2
PDD($C; k$)	Pointwise Distance Distribution of C	Definition 2.3
SCD(C)	Simplexwise Centered Distribution of C	Definition 4.1
SPD(C)	Sorted Pairwise Distances of a cloud C	Definition 2.2

Table 2. Acronyms and references for concepts in sections 2, 3, and 4.

Definition 4.1 (Simplexwise Centered Distribution SCD). Let $C \subset \mathbb{R}^n$ be any cloud of m unordered points. For any base sequence A of $n-1$ ordered points $p_1, \dots, p_{n-1} \in C$, set $\tilde{A} = A \cup \{O(C)\}$ by adding the center of mass $O(C)$ as the last n -th point. Following Definition 3.1, take the $n \times n$ distance matrix $D(\tilde{A})$ and the cloud $M(C; \tilde{A}) \subset \mathbb{R}^{n+1}$ of points $d(q)$ for $q \in C \setminus A$. Any permutation $\xi \in S_{n-1}$ acts on $D(\tilde{A})$ by mapping $D(\tilde{A})_{i,j}$ to $D(\tilde{A})_{\xi(i),\xi(j)}$ for $i, j = 1, \dots, n-1$, and on $M(C; \tilde{A}) \subset \mathbb{R}^{n+1}$ by permuting the first $n-1$ coordinates and by multiplying the $(n+1)$ -st coordinate by $\text{sign}(\xi)$. The *Oriented Centered Distribution* $\text{OCD}(C; A)$ is the equivalence class of pairs $[D(\tilde{A}), M(C; \tilde{A})]$ under all permutations $\xi \in S_{n-1}$ on points of A . The *Simplexwise Centered Distribution* $\text{SCD}(C)$ is the unordered set of $\text{OCD}(C; A)$ for all $\binom{m}{n-1}$ unordered $(n-1)$ -point subsets $A \subseteq C$. The *mirror image* $\overline{\text{SCD}}(C)$ is obtained from $\text{SCD}(C)$ by reversing the signs of $(n+1)$ coordinates $s(q; A)$ of $d(q) \in M(C; A)$. ■

Even if a point of A accidentally coincides with $O(C)$, Definition 4.1 still makes sense. In dimension $n = 2$, Definition 4.1 needs no permutations because $n - 1 = 1$. The points of $M(C; \tilde{A})$ can be lexicographically ordered without affecting future metrics. Some of the $\binom{m}{n-1}$ OCDs in $\text{SCD}(C)$ can be identical as in Example 4.2(b). If we collapse any $l > 1$ identical OCDs into a single OCD with the *weight* $l/\binom{m}{n-1}$, then $\text{SCD}(C)$ can be considered a weighted probability distribution of OCDs.

Example 4.2 (Simplexwise Centered Distribution SCDs for regular clouds).

(a) Let $C_3 \subset \mathbb{R}^2$ be the triangular cloud of the points $p_1 = (1, 0)$, $p_2 = (\cos \frac{2\pi}{3}, \sin \frac{2\pi}{3})$, $p_3 = (\cos \frac{4\pi}{3}, \sin \frac{4\pi}{3})$, which form an equilateral triangle with the center of mass $O(C_3) = (0, 0) = \mathbf{0}$ and side lengths $\sqrt{3}$. For the base sequence $A_1 = \{p_1\}$ and extended $\tilde{A}_1 = \{p_1, O(C)\}$, we get

$$\begin{aligned} \text{OCD}(C_3; A_1) &= \left[|\mathbf{p}_1 - \mathbf{0}|, \begin{pmatrix} |\mathbf{p}_2 - \mathbf{p}_1| & |\mathbf{p}_3 - \mathbf{p}_1| \\ |\mathbf{p}_2 - \mathbf{0}| & |\mathbf{p}_3 - \mathbf{0}| \\ \text{sign}(\mathbf{p}_2 - \mathbf{p}_1, \mathbf{p}_2) & \text{sign}(\mathbf{p}_3 - \mathbf{p}_1, \mathbf{p}_3) \end{pmatrix} \right] = \\ &= \left[1, \begin{pmatrix} \sqrt{3} & \sqrt{3} \\ 1 & 1 \\ -1 & +1 \end{pmatrix} \right], \text{ because } \det(\mathbf{p}_2 - \mathbf{p}_1, \mathbf{p}_2) = \begin{vmatrix} -\frac{3}{2} & -\frac{1}{2} \\ \frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{2} \end{vmatrix} < 0 \end{aligned}$$

and $\det(\mathbf{p}_3 - \mathbf{p}_1, \mathbf{p}_2) = \begin{vmatrix} -\frac{3}{2} & -\frac{1}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \end{vmatrix} > 0$. Due to the 3-fold rotational symmetry, $\text{OCD}(C_3; A_i)$ is independent of $i = 1, 2, 3$. Hence, $\text{SCD}(C_3)$ can be considered a distribution of a single $\text{OCD}(C_3; A_1)$ of weight 1.

(b) Let $C_4 \subset \mathbb{R}^2$ be the square cloud of the points $p_1 = (1, 0)$, $p_2 = (0, 1)$, $p_3 = (-1, 0)$, $p_4 = (0, -1)$, see the last image in Fig. 2. The center of mass is $\mathbf{0} \in \mathbb{R}^2$ and has a distance 1 to each p_i . For $A_1 = \{p_1\}$, we get

$$\begin{aligned} M(C_4; \tilde{A}_1) &= \begin{pmatrix} |\mathbf{p}_2 - \mathbf{p}_1| & |\mathbf{p}_3 - \mathbf{p}_1| & |\mathbf{p}_4 - \mathbf{p}_1| \\ |\mathbf{p}_2 - \mathbf{0}| & |\mathbf{p}_3 - \mathbf{0}| & |\mathbf{p}_4 - \mathbf{0}| \\ \text{sign}(\mathbf{p}_2 - \mathbf{p}_1, \mathbf{p}_2) & \text{sign}(\mathbf{p}_3 - \mathbf{p}_1, \mathbf{p}_3) & \text{sign}(\mathbf{p}_4 - \mathbf{p}_1, \mathbf{p}_4) \end{pmatrix} \\ &= \begin{pmatrix} \sqrt{2} & 2 & \sqrt{2} \\ 1 & 1 & 1 \\ -1 & 0 & +1 \end{pmatrix}, \text{ because } \det(\mathbf{p}_2 - \mathbf{p}_1, \mathbf{p}_2) = \begin{vmatrix} -1 & 0 \\ -1 & 1 \end{vmatrix} < 0, \text{ then} \end{aligned}$$

$$\det(\mathbf{p}_3 - \mathbf{p}_1, \mathbf{p}_3) = \begin{vmatrix} -2 & -1 \\ 0 & 0 \end{vmatrix} = 0, \text{ and } \det(\mathbf{p}_4 - \mathbf{p}_1, \mathbf{p}_4) = \begin{vmatrix} -1 & 0 \\ 1 & -1 \end{vmatrix} > 0.$$

Due to the 4-fold rotational symmetry, $\text{OCD}(C_4; A_i) = [1, M(C_4; \tilde{A}_i)]$ is independent of $A_i = \{p_i\}$ for $i = 1, 2, 3, 4$. Hence, $\text{SCD}(C_4)$ can be considered a distribution of a single $\text{OCD}(C_4; A_1)$ of weight 1. ■

Lemma 4.3 and Theorem 4.4 follow from Lemma 3.6 and Theorem 3.7, respectively, by adding the center of mass $O(C)$ as the n -th point to any base sequence $A \subset C$ of $n - 1$ points.

Lemma 4.3 (reconstruction from OCD). Any n -dimensional cloud $C \subset \mathbb{R}^n$ of m unordered points with the center of mass $O(C)$ can be reconstructed, uniquely under rigid motion, from $\text{OCD}(C; A)$ in Definition 3.2 for any base sequence $A \subseteq C$ of $n - 1$ points with $\text{aff}(\tilde{A}) = n - 1$. ■

Theorem 4.4 (complexity and completeness of SCD under rigid motion).

(a) For any n -dimensional cloud C of m unordered points, the Simplexwise Centered Distribution $\text{SCD}(C)$ is computable in time $O(m^n)$.

(b) Any n -dimensional clouds $C, C' \subset \mathbb{R}^n$ of m unordered points can be exactly matched by rigid motion *if and only if* $\text{SCD}(C) = \text{SCD}(C')$.

(c) Any n -dimensional clouds $C, C' \subset \mathbb{R}^n$ of m unordered points can be exactly matched by isometry *if and only if* $\text{SCD}(C)$ equals $\text{SCD}(C')$ or its mirror image $\overline{\text{SCD}}(C')$. ■

Lemma 4.3 and Theorem 4.4 fulfill conditions (a,b) and partially (e), which is the invariant part of Problem 1.1. Equality $\text{SCD}(C) = \text{SCD}(C')$ is interpreted as a bijection between the unordered distributions $\text{SCD}(C) \rightarrow \text{SCD}(C')$ matching all OCDs, which can be verified by checking if a distance metric between these SCDs equals 0. The metric part of Problem 1.1 requires a Lipschitz continuous metric on SCDs.

Example 4.2(b) illustrates the key discontinuity challenge: if $p_3 = (-1, 0)$ is perturbed, the corresponding sign can discontinuously change to +1 or -1. To get a continuous metric on SCDs, we will multiply each sign by a continuous *strength* of a simplex, as formalized below.

Definition 4.5 (a *simplex* T on $n + 1$ points in \mathbb{R}^n). The geometric *simplex* T on any ordered points $p_0, \dots, p_n \in \mathbb{R}^n$ is the subset $T = \left\{ \sum_{i=0}^n t_i \mathbf{p}_i \mid t_i \in [0, 1], \sum_{i=0}^n t_i = 1 \right\} \subset \mathbb{R}^n$ with ordered *vertices* p_0, \dots, p_n . Define $\text{sign}(T) = \text{sign}(\mathbf{p}_1 - \mathbf{p}_0, \dots, \mathbf{p}_n - \mathbf{p}_0)$, see Definition 3.1. ■

The *volume* $\text{vol}(T)$ of a simplex T or an arbitrary polyhedron is often used as a shape descriptor and detects affine independence in the sense that T is *degenerate* if and only if $\text{vol}(T) = 0$. However, the volume and all other distance-based descriptors do not distinguish mirror images, which have different signs of orientation. If a simplex T goes through a degenerate configuration C of vertices with $\text{aff}(C) < n$, then $\text{sign}(T)$ can discontinuously change. This discontinuity is an obstacle to recognizing nearly mirror-symmetric simplices or more general clouds.

One way to resolve this discontinuity is to consider the signed volume $\text{sign}(T)\text{vol}(T)$, because $\text{vol}(T)$ vanishes only on degenerate simplices. However, the volume $\text{vol}(T)$ is not Lipschitz continuous, as illustrated below.

Example 4.6 (the area of a triangle is not Lipschitz continuous). For any large $l > 0$ and small $\varepsilon > 0$, let $T(l, \varepsilon) \subset \mathbb{R}^2$ be the 2D simplex (triangle) on the vertices $(0, \varepsilon)$ and $(\pm l, 0)$. The signed area $l\varepsilon$ of $T(l, \varepsilon)$ distinguishes mirror images $T(l, \pm\varepsilon)$ but is not Lipschitz continuous under perturbations. Indeed, as $\varepsilon \rightarrow 0$, the triangle $T(l, \varepsilon)$ degenerates to a straight line, while the area drops to 0 too quickly so that $\frac{\text{vol}(T(l, \varepsilon)) - \text{vol}(T(l, 0))}{\varepsilon - 0} = \frac{l\varepsilon}{\varepsilon} = l$ is not bounded. Hence, if given points are not restricted to a fixed region, a small change in their positions may lead to a large change for the area of a triangle, and similarly in \mathbb{R}^n . ■

Definition 4.7 (*strength* $\sigma(T)$ of a simplex). Let T be the simplex on any $p_0, \dots, p_n \in \mathbb{R}^n$. The *half-perimeter* is $p(T) = \frac{1}{2} \sum_{i \neq j} |p_i - p_j|$. The *strength* is $\sigma(T) = \frac{\text{vol}^2(T)}{p^{2n-1}(T)}$. The *signed strength* is $s(T) = \text{sign}(T)\sigma(T)$. ■

Example 4.8 (the strength for $n = 1, 2$). (a) For $n = 1$, the simplex T on ordered points $p_0, p_1 \in \mathbb{R}$ is the line segment with the half-perimeter $p(T) = \frac{1}{2}|p_1 - p_0|$, length $\text{vol}(T) = |p_1 - p_0| = 2p(T)$, strength $\sigma(T) =$

$\frac{\text{vol}^2(T)}{p(T)} = 4p(T) = 2|p_1 - p_0|$, and signed strength $s(T) = 2(p_1 - p_0)$.

(b) For $n = 2$, any triangle $T \subset \mathbb{R}^2$ with inter-point distances a, b, c has the half-perimeter $p = \frac{a + b + c}{2}$, area $\text{vol}(T) = \sqrt{p(p-a)(p-b)(p-c)}$ by Heron's formula, strength $\sigma(T) = \frac{\text{vol}^2(T)}{p^3} = \frac{(p-a)(p-b)(p-c)}{p^2}$. ■

Theorem 4.9 (strength properties, [7, Theorem 2.4]). (a) The strengths $\sigma(T)$ and $s(T)$ of any simplex $T \subset \mathbb{R}^n$ are invariant under isometry and rigid motion in \mathbb{R}^n , respectively, and can be computed in time $O(n^3)$. The uniform scaling of \mathbb{R}^n by any factor $c > 0$ multiplies $\sigma(T)$ and $s(T)$ by c .

(b) There is a constant $\lambda_n > 0$ such that, for any $\varepsilon > 0$, if a simplex Q is obtained from a simplex $T \subset \mathbb{R}^n$ by perturbing every vertex of T within its ε -neighborhood, then $|\sigma(T) - \sigma(Q)| \leq 2\lambda_n\varepsilon$ and $|s(T) - s(Q)| \leq 2\lambda_n\varepsilon$, where $\lambda_1 = 2$, $\lambda_2 = \sqrt{3}$, and $\lambda_n \leq \frac{2^{n+0.5}}{(n!)n^{2n-4}}$ for $n \geq 3$, e.g., $\lambda_3 < 0.37$. ■

Lemma 4.10 (signs of sequences and simplices). For any cloud $C \subset \mathbb{R}^n$, a base sequence $A \subset C$ of n ordered points, and any $q \in C \setminus A$, consider the simplex T on the sequence $A \cup \{q\}$, including q as the $(n+1)$ -st point. Then $\text{sign}(T)$ from Definition 4.5 equals the sign $s(q; A)$ from Definition 3.1.

Proof. Definition 3.1 introduced $s(q; A)$ as the sign of the determinant with the columns $\mathbf{q} - \mathbf{p}_1, \dots, \mathbf{q} - \mathbf{p}_n$. Then $s(q; A) = \text{sign}(\mathbf{q} - \mathbf{p}_1, \dots, \mathbf{q} - \mathbf{p}_n) = (-1)^n \text{sign}(T')$ for the simplex T' on the ordered q, p_1, \dots, p_n , because $\text{sign}(T') = \text{sign}(\mathbf{p}_1 - \mathbf{q}, \dots, \mathbf{p}_n - \mathbf{q})$ by Definition 4.5. By moving the vertex q of T' to the last $(n+1)$ -st position, we get the simplex T on $A \cup \{q\}$ with $\text{sign}(T) = (-1)^n \text{sign}(T')$, so $s(q; A) = \text{sign}(T)$. ■

Theorem 4.9(b) and Lemma 4.10 justify replacing the discontinuous sign $s(q; A)$ with the signed strength $s(A \cup \{q\}) = s(q; A)\sigma(A \cup \{q\})$ to define Lipschitz continuous metrics in the next section.

5 Continuous metrics on complete invariants

This section introduces Lipschitz continuous metrics on OSDs and SCDs by using the signed strength of an ordered sequence from Lemma 4.10.

For simplicity and similar to the case of a metric space, we assume that a cloud $C \subset \mathbb{R}^n$ is given by a matrix of pairwise Euclidean distances. If C is given by Euclidean coordinates of points, then any distance requires $O(n)$ computations, and we should add the factor n in all complexities below, which keeps all times polynomial in the number m of points.

Recall that Definition 2.1 introduced the bottleneck distance BD between point clouds of the same size in a metric space. Definition 5.1 interprets BD as the bottleneck matching distance for a weighted complete bipartite graph, which is needed for more general distributions OSD, SCD.

Definition 5.1 (Bottleneck Matching Distance BMD). Let Γ be a complete bipartite graph with m white vertices and m black vertices so that every white vertex is connected to every black vertex by a single edge e of a weight $w(e) \geq 0$. A *vertex matching* of the graph Γ is a collection E of m disjoint edges (with distinct vertices). The *weight* $W(E) = \max_{e \in E} w(e)$ is the largest weight of an edge in E . The *Bottleneck Matching Distance* $\text{BMD}(\Gamma) = \min_E W(E)$ is the minimum weight of a vertex matching. ■

The bottleneck matching distance $\text{BMD}(\Gamma)$ for a bipartite graph Γ with V vertices and E edges is computed in time $O(E\sqrt{V})$, see [24].

Lemma 5.2 (metric axioms for the bottleneck matching distance BMD). Let $B = \{B_1, \dots, B_m\}$ and $C = \{C_1, \dots, C_m\}$ be any unordered distributions of m objects with a base metric d . Define the complete bipartite graph $\Gamma(B, C)$ with m black vertices (associated with) B_1, \dots, B_m and m white vertices (associated with) C_1, \dots, C_m . Set the weight of the edge e joining B_i, C_j as $w(e) = d(B_i, C_j)$. Then $\text{BMD}(\Gamma(B, C))$ in Definition 5.1 satisfies all metric axioms on unordered distributions of m objects.

Proof. The first axiom says that $\text{BMD}(B, C) = 0$ if and only if the weighted distributions B, C are equal in the sense that there is a bijection $g : B \rightarrow C$ so that $d(g(B_i), C_i) = 0$, i.e. $g(B_i) = C_i$ due to the first metric axiom for d , $i = 1, \dots, m$. Indeed, if the distributions B, C can be matched by a bijection, we get a vertex matching E of $\Gamma(B, C)$ whose all edges have weight 0, then $\text{BMD}(\Gamma(B, C)) = 0$ by Definition 5.1. Conversely, if $\text{BMD}(\Gamma(B, C)) = 0$, there is a vertex matching E of $\Gamma(B, C)$ with all edge weights 0, which induces a bijection $B \rightarrow C$.

The symmetry $\text{BMD}(\Gamma(B, C)) = \text{BMD}(\Gamma(C, B))$ follows from Definition 5.1 and the symmetry of d . To prove the triangle inequality

$$\text{BMD}(\Gamma(B, C)) + \text{BMD}(\Gamma(C, D)) \geq \text{BMD}(\Gamma(B, D)),$$

let E_{BC}, E_{CD} be optimal vertex matchings in the graphs $\Gamma(B, C), \Gamma(C, D)$, respectively, such that $\text{BMD}(\Gamma(B, C)) = W(E_{BC})$ and $\text{BMD}(\Gamma(C, D)) = W(E_{CD})$, see Definition 5.1. The composition $E_{BD} = E_{BC} \circ E_{CD}$ is a vertex matching in $\Gamma(B, D)$, so $W(E_{BD}) \geq \text{BMD}(\Gamma(B, D))$. It suffices to prove that $W(E_{BC}) + W(E_{CD}) \geq W(E_{BD})$. Let e_{BD} be an edge with a largest weight from E_{BD} , so $W(E_{BD}) = w(e_{BD})$. The edge e_{BD} can be considered a union of edges $e_{BC} \in E_{BC}$ and $e_{CD} \in E_{CD}$. The triangle inequality $w(e_{BC}) + w(e_{CD}) \geq w(e_{BD}) = W(E_{BD})$ for d implies that $W(E_{BC}) + W(E_{CD}) \geq W(E_{BD})$ because both terms in the left-hand side are maximized for all edges (not only e_{BC}, e_{CD}) from E_{BC}, E_{CD} . ■

Definition 5.5 will use BD for clouds of the same size in \mathbb{R}^{n+1} with the base metric L_∞ . The Earth Mover's Distance EMD in Definition 5.3 will be applied to clouds of any number of equally weighted points in \mathbb{R}^{n+1} with the base metric L_∞ . Though the EMD was introduced in [44] for all distributions, which can have different total weights, [44, appendix] proved that the EMD is a metric for distributions of the same total weight. Hence Definition 5.3 considers only normalized distributions of total weight 1.

Definition 5.3 (Earth Mover's Distance on weighted distributions). Let $B = \{B_1, \dots, B_k\}$ and $C = \{C_1, \dots, C_l\}$ be unordered sets of objects with weights $w(B_i)$, $i = 1, \dots, k$, and $w(C_j)$, $j = 1, \dots, l$, respectively such that $\sum_{i=1}^k w(B_i) = 1 = \sum_{j=1}^l w(C_j)$. Let d be a *base metric* between any objects B_i and C_j . A *flow* from B to C is a $k \times l$ matrix whose entry f_{ij} represents a *flow* from B_i to C_j . The *Earth Mover's Distance* $\text{EMD}(B, C) = \sum_{i=1}^k \sum_{j=1}^l f_{ij} d(B_i, C_j)$ is minimized for $f_{ij} \in [0, 1]$ subject to (1) $\sum_{j=1}^l f_{ij} \leq w(B_i)$ for any fixed index $i = 1, \dots, k$, (2) $\sum_{i=1}^k f_{ij} \leq w(C_j)$ for any fixed index $j = 1, \dots, l$, and (3) $\sum_{i=1}^k \sum_{j=1}^l f_{ij} = 1$. ■

The first condition $\sum_{j=1}^l f_{ij} \leq w(B_i)$ means that, for each $i = 1, \dots, k$, at most the weight $w(B_i)$ ‘flows’ into all objects C_j via f_{ij} for $j = 1, \dots, l$. The second condition $\sum_{i=1}^k f_{ij} \leq w(C_j)$ means that, for each $j = 1, \dots, l$, all f_{ij} ‘flow’ from B_i for $i = 1, \dots, k$ into C_j up to its weight $w(C_j)$. The last condition $\sum_{i=1}^k \sum_{j=1}^l f_{ij} = 1$ forces to ‘flow’ all B_i to all C_j . The EMD is a partial case of Wasserstein metrics [48] in transportation theory [27].

Lemma 5.4 (BMD \geq EMD). Let $B = \{B_1, \dots, B_m\}$, $C = \{C_1, \dots, C_m\}$ be unordered sets of equally weighted objects with a base metric d and the complete bipartite graph $\Gamma(B, C)$ in Lemma 5.2. Fix the weight of $w(e)$ of the edge between any vertices B_i and C_j of $\Gamma(B, C)$ as $d(B_i, C_j)$. Then $\text{BMD}(\Gamma(B, C)) \geq \text{EMD}(B, C)$, where $\text{EMD}(B, C)$ uses any base metric d' such that $d'(B_i, C_j) \leq d(B_i, C_j)$ for all $i, j = 1, \dots, m$.

Proof. In Definition 5.1, let E be a vertex matching of $\Gamma(B, C)$ that minimizes the Bottleneck Matching Distance $\text{BMD}(\Gamma(B, C)) = \max_{e \in E} w(e)$. This vertex matching E defines a bijection $g : B \rightarrow C$, which is a permutation of m objects. In the notations of Definition 5.3, g induces the ‘flow’ from B to C with $f_{i,g(i)} = \frac{1}{m}$ and $f_{ij} = 0$ for all $j \neq g(i)$, $i = 1, \dots, m$, with the cost $\sum_{i=1}^m \sum_{j=1}^m f_{ij} d'(B_i, C_j) = \sum_{i=1}^m \frac{1}{m} d'(B_i, C_{g(i)}) \leq \sum_{i=1}^m \frac{1}{m} d(B_i, C_{g(i)}) \leq \frac{1}{m} \sum_{i=1}^m \text{BMD}(\Gamma(B, C)) = \text{BMD}(\Gamma(B, C))$. Since $\text{EMD}(B, C)$ is the cost minimized over all $f_{ij} \in [0, 1]$ subject to the weight-respecting conditions in Definition 5.3, $\text{BMD}(\Gamma(B, C))$ is the upper bound of $\text{EMD}(B, C)$. ■

In Definition 5.5, we use the notations BD_∞ and EMD_∞ for metrics on clouds of unordered and equally weighted points with base metric L_∞ .

Definition 5.5 (metrics $\text{BD}_\infty, \text{EMD}_\infty$ on ORD, OCD). **(a)** Let $C \subset \mathbb{R}^n$ be any cloud of m unordered points with a base sequence A of n ordered points. In the notations of Definition 3.2, for each point $q \in C \setminus A$, replace the sign $s(q; A)$ in the last coordinate of $d(q) \in \mathbb{R}^{n+1}$ with $\frac{s(A \cup \{q\})}{\lambda_n}$, where $s(A \cup \{q\}) = s(q; A)\sigma(A \cup \{q\})$ is called the *signed strength*.

Denote the resulting cloud of $m - n$ unordered points $\hat{d}(q) \in \mathbb{R}^{n+1}$ by $\hat{M}(C; A)$. Any permutation $\xi \in S_n$ of $1, \dots, n$ acts on $D(A)$ and $\hat{M}(C; A)$ as in Definition 3.2. For another cloud $C' \subset \mathbb{R}^n$ of unordered points with a base sequence A' of n ordered points, similarly define the cloud $\hat{M}(C'; A')$. Let the cloud C' have the same number of points as C . Then set

$$\text{BD}_\infty[\xi] = \max \left\{ L_\infty(\xi(D(A)), D(A')), \text{BD}_\infty(\xi(\hat{M}(C; A)), \hat{M}(C'; A')) \right\}.$$

In general, even if the sizes of the clouds $C, C' \subset \mathbb{R}^n$ differ, set

$$\text{EMD}_\infty[\xi] = \max \{ L_\infty(\xi(D(A)), D(A')), \text{EMD}_\infty(\xi(\hat{M}(C; A)), \hat{M}(C'; A')) \}.$$

Define the following metrics between Oriented Relative Distributions:

$$\begin{aligned} \text{BD}_\infty(\text{ORD}(C; A), \text{ORD}(C'; A')) &= \min_{\xi \in S_n} \text{BD}_\infty[\xi], \text{ and} \\ \text{EMD}_\infty(\text{ORD}(C; A), \text{ORD}(C'; A')) &= \min_{\xi \in S_n} \text{EMD}_\infty[\xi]. \end{aligned}$$

(b) The metrics $\text{BD}_\infty, \text{EMD}_\infty$ are defined on OCDs similarly to part (a) after replacing $A \subset C$ with the extended base sequence $\tilde{A} = A \cup \{O(C)\}$, which includes the center of mass $O(C)$ of a given cloud $C \subset \mathbb{R}^n$. ■

In Definition 5.5, the subscript ∞ refers to the base Chebyshev metric L_∞ , which can be replaced with many other base metrics on distance matrices and clouds in \mathbb{R}^{n+1} . For the $(n + 1)$ -st coordinate of points in clouds $M(C; A) \subset \mathbb{R}^{n+1}$, the coefficient $\frac{1}{\lambda_n}$ in front of the signed strength $s(A \cup \{q\})$ normalises the Lipschitz constant $2\lambda_n$ from Theorem 4.9(b) to 2 in line with changes of inter-point distances by at most 2ε when given points are perturbed within their ε -neighbourhoods.

Lemma 5.6 (properties of metrics BD_∞ and EMD_∞ on ORDs). (a) For any n -dimensional clouds $C, C' \subset \mathbb{R}^n$ of m unordered points with base sequences $A \subset C$ and $A' \subset C'$ of n ordered points, the bottleneck distance $\text{BD}_\infty(\text{ORD}(C; A), \text{ORD}(C'; A'))$ in Definition 5.5(a) satisfies all metric axioms and can be computed in time $O(n!(n^2 + m^{1.5} \log^{n+1} m))$. Also, $\text{BD}_\infty(\text{ORD}(C; A), \text{ORD}(C'; A')) \geq \text{EMD}_\infty(\text{ORD}(C; A), \text{ORD}(C'; A'))$.

(b) For any n -dimensional clouds $C, C' \subset \mathbb{R}^n$ of up to m unordered

points with base sequences $A \subset C$ and $A' \subset C'$ of n ordered points, $\text{EMD}_\infty(\text{ORD}(C; A), \text{ORD}(C'; A'))$ in Definition 5.5(a) satisfies all metric axioms and can be computed in time $O(n!(n^2 + m^3 \log m))$.

Proof. The first metric axiom says that $\text{ORD}(C; A), \text{ORD}(C'; A')$ are equivalent by Definition 3.2 if and only if $\text{BD}_\infty(\text{ORD}(C; A), \text{ORD}(C'; A')) = 0$, i.e. $\text{BD}_\infty[\xi] = 0$ for some permutation $\xi \in S_n$. Due to the first metric axiom for L_∞ and BD , the equality $\text{BD}_\infty[\xi] = 0$ is equivalent to $\xi(D(A)) = D(A')$ and $\xi(M(C; A)) = M(C'; A')$ as clouds of unordered points in \mathbb{R}^{n+1} . The last two conclusions mean that the Oriented Relative Distributions $\text{ORD}(C; A), \text{ORD}(C'; A')$ are equivalent by Definition 3.2.

The symmetry axiom for BD_∞ follows because any permutation $\xi \in S_n$ is invertible. To prove the triangle inequality

$$\text{BD}_\infty(\text{ORD}(C; A), \text{ORD}(C'; A')) + \text{BD}_\infty(\text{ORD}(C'; A'), \text{ORD}(C''; A'')) \geq \text{BD}_\infty(\text{ORD}(C; A), \text{ORD}(C''; A'')),$$

let $\xi, \xi' \in S_n$ be optimal permutations for the BD_∞ 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 BD . Taking the maximum of $L_\infty, \text{BD}_\infty$ preserves the triangle inequality, see other metric transforms in [18, section 4.1]. Then $\text{BD}_\infty(\text{ORD}(C; A), \text{ORD}(C''; A'')) = \min_{\xi \in S_n} d(\xi)$ cannot be larger than $\text{BD}_\infty[(\xi')^{-1}\xi]$ for the composition of the permutations above, so the triangle inequality holds for BD_∞ . For a fixed $\xi \in S_n$, the distance $L_\infty(\xi(D(A)), D(A'))$ requires time $O(n^2)$. By Theorem 4.9(a) the signed strengths in the $(n+1)$ -st coordinates of all points of $\hat{M}(C; A), \hat{M}(C'; A')$ need time $O(mn^3)$, which is dominated by the final time. The bottleneck distance BD_∞ between the clouds $\xi(\hat{M}(C; A))$ and $\hat{M}(C'; A')$ of $m-n$ unordered points in \mathbb{R}^{n+1} needs time $O(m^{1.5} \log^{n+1} m)$ by [21, Theorem 6.5]. The minimization over all $\xi \in S_n$ gives the factor $n!$ in the final time.

To prove that $\text{BD}_\infty \geq \text{EMD}_\infty$, take a permutation $\xi \in S_n$ that minimizes $\text{BD}_\infty[\xi]$, i.e. $\text{BD}_\infty(\text{ORD}(C; A), \text{ORD}(C'; A')) = \text{BD}_\infty[\xi]$. Since

both metrics BD_∞ and EMD_∞ between clouds use the base distance L_∞ between points in \mathbb{R}^{n+1} , Lemma 5.4 implies that

$$\text{BD}_\infty(\xi(\hat{M}(C; A)), \hat{M}(C'; A')) \geq \text{EMD}_\infty(\xi(\hat{M}(C; A)), \hat{M}(C'; A')).$$

Taking the maximum of both sides with $L_\infty(\xi(D(A)), D(A'))$ gives

$$\text{BD}_\infty(\text{ORD}(C; A), \text{ORD}(C'; A')) = \text{BD}_\infty[\xi] \geq \text{EMD}_\infty[\xi].$$

Taking the minimum of the right hand side over $\xi \in S_n$ gives

$$\text{BD}_\infty(\text{ORD}(C; A), \text{ORD}(C'; A')) \geq \text{EMD}_\infty(\text{ORD}(C; A), \text{ORD}(C'; A')).$$

(b) The metric axioms for EMD_∞ are proved as in part (a) by replacing BD with EMD whose metric axioms were checked in [44, appendix]. Consider the clouds $\hat{M}(C; A)$ and $\hat{M}(C'; A')$ as unordered distributions of equally weighted $m - n$ points in \mathbb{R}^{n+1} . The EMD between weighted distributions of a maximum size m can be computed in time $O(m^3 \log m)$, see [22, 23]. Taking into account time $O(n^2)$ for $L_\infty(\xi(D(A)), D(A'))$ and the minimization over $\xi \in S_n$, we get time $O(n!(n^2 + m^3 \log m))$. ■

Definition 5.7 adapts $\text{BD}_\infty, \text{EMD}_\infty$ in Definition 5.5 to invariants.

Definition 5.7 (metrics BD_∞ and EMD_∞ on invariants OSD and SCD). (a) Let $C, C' \subset \mathbb{R}^n$ be any n -dimensional clouds of m unordered points. Let $\Gamma(C, C')$ denote the complete bipartite graph with $\binom{m}{n}$ black vertices and $\binom{m}{n}$ white vertices that are associated with $\text{ORD}(C; A)$ and $\text{ORD}(C'; A')$ for all base sequences $A \subset C$ and $A' \subset C'$ of n ordered points. For each edge of $\Gamma(C, C')$ between vertices associated $\text{ORD}(C; A)$ and $\text{ORD}(C'; A')$, define its weight as $\text{BD}_\infty(\text{ORD}(C; A), \text{ORD}(C'; A'))$ in Definition 5.5 (a). The *bottleneck distance* $\text{BD}_\infty(\text{OSD}(C), \text{OSD}(C'))$ is the bottleneck matching distance $\text{BMD}(\Gamma(C, C'))$. In general, even if the sizes of the clouds C, C' differ, we can consider $\text{OSD}(C)$ and $\text{OSD}(C')$ as weighted distributions whose objects $\text{ORD}(C; A)$ and $\text{ORD}(C'; A')$ have the base metric $\text{EMD}_\infty(\text{ORD}(C; A), \text{ORD}(C'; A'))$. Define the *Earth Mover's Distance*

$\text{EMD}_\infty(\text{OSD}(C), \text{OSD}(C'))$ as in Definition 5.3.

(b) The metrics $\text{BD}_\infty, \text{EMD}_\infty$ are defined on SCDs similarly to part (a) after replacing $A \subset C$ with the extended base sequence $\tilde{A} = A \cup \{O(C)\}$, which includes the center of mass $O(C)$ of a given cloud $C \subset \mathbb{R}^n$. ■

Theorem 5.8 (properties of BD_∞ and EMD_∞ on OSDs). (a) For any n -dimensional clouds C, C' of m unordered points, $\text{BD}_\infty(\text{OSD}(C), \text{OSD}(C'))$ in Definition 5.7(a) satisfies all metric axioms and can be computed in time $O(n!(n^2 + m^{1.5} \log^{n+1} m)k^2 + k^{2.5})$ for $k = \binom{m}{n} = O(m^n)$. Also, $\text{BD}_\infty(\text{OSD}(C), \text{OSD}(C')) \geq \text{EMD}_\infty(\text{OSD}(C), \text{OSD}(C'))$.

(b) Let $C, C' \subset \mathbb{R}^n$ be any n -dimensional clouds, which may consist of up to m unordered points. Let $\text{OSD}(C), \text{OSD}(C')$ have a maximum size $l \leq \binom{m}{n}$ after collapsing identical ORDs. Then $\text{EMD}_\infty(\text{OSD}(C), \text{OSD}(C'))$ in Definition 5.7(a) satisfies all metric axioms and can be computed in time $O(n!(n^2 + m^{1.5} \log^{n+1} m)l^2 + l^3 \log l)$.

Proof. (a) The base metrics $\text{BD}_\infty, \text{EMD}_\infty$ on ORDs in Definition 5.5(a) satisfy the axioms by Lemma 5.6(a). Then $\text{BD}_\infty, \text{EMD}_\infty$ on OSDs in Definition 5.7(a) satisfy the axioms by Lemma 5.2, [44], respectively.

In part (a), both clouds $C, C' \subset \mathbb{R}^n$ consist of m unordered points. Then both $\text{OSD}(C), \text{OSD}(C')$ consist of $k = \binom{m}{n}$ distributions ORD. The graph $\Gamma(C, C')$ in Definition 5.7 has $V = 2k$ vertices and $E = k^2$ edges whose all weights are computed in time $O(n!(n^2 + m^{1.5} \log^{n+1} m)k^2)$. Then $\text{BD}_\infty(\text{OSD}(C), \text{OSD}(C')) = \text{BMD}(\Gamma(C, C'))$ is computed by [24] in extra time $O(E\sqrt{V}) = O(k^{2.5})$. Lemma 5.4 and the inequality $\text{BD}_\infty \geq \text{EMD}_\infty$ for ORDs in Lemma 5.6(a) imply that $\text{BD}_\infty \geq \text{EMD}_\infty$ holds for OSDs.

(b) $\text{OSD}(C), \text{OSD}(C')$ consist of $l \leq \binom{m}{n}$ distributions ORD. Each EMD_∞ between any ORDs in the invariants $\text{OSD}(C)$ and $\text{OSD}(C')$ is computable in time $O(n!(n^2 + m^{1.5} \log^{n+1} m)l)$. The total time of all $E \leq l^2$ distances is $O(n!(n^2 + m^{1.5} \log^{n+1} m)l^2)$. The final extra time $O(l^3 \log l)$ is for computing EMD between distributions whose size is at most l . ■

Theorem 5.9 is proved as Theorem 5.8 by replacing n with $n-1$, because any base sequence $A \subset C$ of $n-1$ ordered points is extended by the center mass $O(C)$ to the sequence $\tilde{A} = A \cup \{O(C)\}$ of n points.

Theorem 5.9 (properties of BD_∞ and EMD_∞ on invariants SCDs). **(a)** For any n -dimensional clouds $C, C' \subset \mathbb{R}^n$ of m unordered points, the bottleneck distance $\text{BD}_\infty(\text{SCD}(C), \text{SCD}(C'))$ in Definition 5.7(b) satisfies all metric axioms and, for $k = \binom{m}{n-1} = O(m^{n-1})$, can be computed in time $O((n-1)!(n^2 + m^{1.5} \log^n m)k^2 + k^{2.5})$. Also, the inequality $\text{BD}_\infty(\text{SCD}(C), \text{SCD}(C')) \geq \text{EMD}_\infty(\text{SCD}(C), \text{SCD}(C'))$ holds.

(b) Let $C, C' \subset \mathbb{R}^n$ be any n -dimensional clouds, which can have up to m unordered points. Let $\text{SCD}(C), \text{SCD}(C')$ have a maximum size $l \leq \binom{m}{n-1}$ after collapsing identical ORDs. Then $\text{EMD}_\infty(\text{SCD}(C), \text{SCD}(C'))$ in Definition 5.7(b) satisfies all metric axioms and can be computed in time $O((n-1)!(n^2 + m^{1.5} \log^n m)l^2 + l^3 \log l)$. ■

A rough upper bound is $l \leq k = \binom{m}{n-1} = \frac{m(m-1)\dots(m-n+2)}{n!} = O(m^{n-1})$. In dimension $n = 2$, Theorem 5.9 gives time $O(m^4 \log^2 m)$ for BD_∞ and EMD_∞ . This upper bound is faster than the past time $O(m^5 \log m)$ for comparing m -point clouds by the Hausdorff distance minimized over all isometries in \mathbb{R}^2 [13]. For $n \geq 3$, BD_∞ and EMD_∞ are computable in times $O(m^{2.5(n-1)})$ and $O(nm^{3(n-1)} \log m)$, respectively.

Theorem 5.10 (Lipschitz continuity of OSD, SCD). For any cloud $C \subset \mathbb{R}^n$ of m unordered points, perturbing any point within its ε -neighborhood changes OSD(C) and SCD(C) by at most 2ε in BD_∞ and EMD_∞ .

Proof. Arbitrarily order points of C as p_1, \dots, p_m . Then C' consist of their ε -perturbations p'_1, \dots, p'_m such that $|p_i - p'_i| \leq \varepsilon$ for $i = 1, \dots, m$.

The triangle inequality implies that any inter-point distance changes by at most 2ε , i.e. $||p_i - p_j| - |p'_i - p'_j|| \leq |p_i - p'_i| + |p_j - p'_j| \leq 2\varepsilon$ for any $i, j = 1, \dots, m$. Without loss of generality, one can assume that a base sequence $A \subset C$ consists of the first n points p_1, \dots, p_n .

Let $\xi \in S_n$ be the trivial permutation keeping all $1, \dots, n$ fixed. Let $A' \subset C'$ denote the base sequence of the perturbed points p'_1, \dots, p'_n . In the

notations of Definition 5.5, any element (coordinate) of $D(A) = \xi(D(A))$ and $\hat{M}(C; A) = \xi(\hat{M}(C; A))$ differs from the corresponding element of $D(A')$ and $\hat{M}(C'; A')$ by at most 2ε . This upper bound of 2ε for normalized strengths, which are the last $(n + 1)$ -st coordinates of points in $\hat{M}(C; A)$ and $\hat{M}(C'; A')$, follows from Theorem 4.9(b) and Lemma 4.10. Then all three distances $L_\infty(\xi(D(A)), D(A'))$, $\text{BD}_\infty(\xi(\hat{M}(C; A)), \hat{M}(C'; A'))$, and $\text{EMD}_\infty(\xi(\hat{M}(C; A)), \hat{M}(C'; A'))$ have the upper bound 2ε . Minimizing over all permutations $\xi \in S_n$ preserves the upper bound 2ε , i.e.

$$\begin{aligned} \text{BD}_\infty(\text{ORD}(C; A), \text{ORD}(C'; A')) &= \min_{\xi \in S_n} \text{BD}_\infty[\xi] \leq 2\varepsilon, \\ \text{EMD}_\infty(\text{ORD}(C; A), \text{ORD}(C'; A')) &= \min_{\xi \in S_n} \text{EMD}_\infty[\xi] \leq 2\varepsilon. \end{aligned}$$

In the notations of Definition 5.7, the trivial permutation $\xi \in S_n$ induces the matching E for $k = \binom{m}{n}$ pairs of vertices of the bipartite graph $\Gamma(C, C')$, which are associated with all base sequences $A \subset C$ and their perturbations $A' \subset C'$. Since all weights of edges $e \in E$ have weights at most 2ε , we get $\text{BD}_\infty(\text{OSD}(C), \text{OSD}(C')) = \text{BMD}(\Gamma(C, C')) \leq 2\varepsilon$. Similarly, the trivial flows $\text{OSD}(C) \rightarrow \text{OSD}(C')$ with base distances at most 2ε imply that normalized distributions have $\text{EMD}_\infty(\text{OSD}(C), \text{OSD}(C')) \leq 2\varepsilon$.

The arguments above work similarly work for the SCD, because the center of mass changes under ε -perturbations by at most ε as follows: $|O(C) - O(C')| = \left| \frac{1}{m} \sum_{i=1}^m p_i - \frac{1}{m} \sum_{i=1}^m p'_i \right| \leq \frac{1}{m} \sum_{i=1}^m |p_i - p'_i| \leq \frac{1}{m} \sum_{i=1}^m \varepsilon = \varepsilon$. ■

Lemma 4.3 and Theorems 4.4, 5.9, and 5.10 imply that the Simplexwise Centered Distribution fully solves Problem 1.1.

6 Faster moment invariants and distances

This section simplifies the complete SCD to faster invariants based on averages and other statistical moments of unordered distributions.

Let R be any unordered set of $r_1, \dots, r_m \in \mathbb{R}$ with weights $w_1, \dots, w_m \geq 0$, respectively, such that $\sum_{i=1}^m w_i = 1$. The weighted average $\mu_1(R) =$

$\sum_{i=1}^m w_i r_i$ is the *1st moment*. For any $t \geq 2$, the normalized t -moment is

$$\mu_t(R) = \sqrt[t]{\sum_{i=1}^m w_i (r_i - \mu_1(R))^t}, \text{ e.g., } \mu_2(R) \text{ is the standard deviation [28,}$$

section 2.8]. If D is a distribution of unordered vectors in \mathbb{R}^k , for any $t \geq 1$, its t -th moment $\mu_t\{D\}$ is defined by computing μ_t coordinate-wise.

Definition 6.1 (Average Simplexwise Moments $\text{ASM}(C; t)$). For any cloud $C \subset \mathbb{R}^n$ of m unordered points, let $A \subset C$ be a base sequence of n points p_1, \dots, p_n . Recall that $\text{SPD}(A)$ from Definition 2.2 consists of $\frac{n(n-1)}{2}$ inter-point distances of A , written in increasing order and hence independent of an order of points in A . The vector $\vec{\mathbf{M}}(C; A)$ is formed by $m-n$ averages $\frac{1}{n} \sum_{i=1}^n |\mathbf{q} - \mathbf{p}_i|$ for all points $q \in C \setminus A$, written in increasing order of values. The vector $\vec{\mathbf{S}}(C; A)$ is formed by $m-n$ normalized strengths $\frac{\sigma(A \cup \{q\})}{\lambda_n}$, written in increasing order, for all points $q \in C \setminus A$.

(a) In the notations above, the *Average Simplexwise Vector* is the tuple

$$\text{ASV}(C; A) = [\text{SPD}(A); \vec{\mathbf{M}}(C; A); \vec{\mathbf{S}}(C; A)] \text{ of } \frac{n(n-1)}{2} + 2(m-n) \text{ values.}$$

For any other cloud $C' \subset \mathbb{R}^n$ with a base sequence A' , define the metric $\text{EMD}_\infty(\text{ASV}(C; A), \text{ASV}(C'; A'))$ as maximum of $L_\infty(\text{SPD}(A), \text{SPD}(A'))$, $\text{EMD}_\infty(\vec{\mathbf{M}}(C; A), \vec{\mathbf{M}}(C'; A'))$, and $\text{EMD}_\infty(\vec{\mathbf{S}}(C; A), \vec{\mathbf{S}}(C'; A'))$.

(b) The *Average Simplexwise Distribution* $\text{ASD}(C)$ is the unordered set of equally weighted vectors $\text{ASV}(C; A)$ for all $\binom{m}{n}$ unordered subsets $A \subseteq C$ of n points. Use the base metric EMD_∞ on ASVs in part (a) to define the Earth Mover's Distance $\text{EMD}_\infty(\text{ASD}(C), \text{ASD}(C'))$ as in Definition 5.3.

(c) For $t \geq 1$, the *Average Simplexwise Moment* $\text{ASM}(C; t) = \mu_t\{\text{ASD}(C)\}$ is the t -th moment of the distribution $\text{ASD}(C)$ of equally weighted vectors $\text{ASV}(C; A)$ over all $\binom{m}{n}$ unordered subsets $A \subseteq C$ of n points, i.e.

$$\text{ASM}(C; t) = [\mu_t\{\text{SPD}(A)\}; \mu_t\{\vec{\mathbf{M}}(C; A)\}; \mu_t\{\vec{\mathbf{S}}(C; A)\}].$$

For any clouds $C, C' \subset \mathbb{R}^n$, define $\text{EMD}_\infty(\text{ASM}(C; t), \text{ASM}(C'; t))$ as the maximum of the three metrics EMD_∞ between constituent subvectors. ■

Since $\sigma(A \cup \{q\})$ is independent of an order of points in A , the vector $\text{ASV}(C; A)$ is well-defined for any subset $A \subset C$ of n unordered points.

The metrics EMD_∞ in Definition 6.1 make sense for any clouds $C, C' \subset \mathbb{R}^n$, which might have different sizes. If underlying distributions are 1-dimensional as for Average Simplexwise Moments $\text{ASM}(C; t)$, then EMD_∞ between them can be analytically computed by Proposition 6.2 below.

Proposition 6.2 (EMD on weighted distributions in \mathbb{R}). Let $S \subset \mathbb{R}$ be a weighted distribution of m unordered numbers with a total weight 1. Let us write all values of S as a sequence \vec{S} of numbers $s_1 \leq \dots \leq s_m$ with weights w_1, \dots, w_m , respectively. Define the *Cumulative Distribution*

$$\text{Function CDF}[S](x) = \begin{cases} 0 & \text{for } x < s_1; \\ \sum_{i=1}^k w_i & \text{for } s_k \leq x < s_{k+1}, k = 1, \dots, m - 1; \\ 1 & \text{for } x \geq s_m. \end{cases}$$

Any other distribution S' of a total weight 1 satisfies $\text{EMD}_\infty(S, S') = \int_{\mathbb{R}} |\text{CDF}[S](x) - \text{CDF}[S'](x)| dx$. If S, S' consist of the same number of equally weighted values, then $\text{EMD}_\infty(S, S') = \frac{1}{m} L_1(\vec{S}, \vec{S}')$.

Proof. The equality $\text{EMD}_\infty(S, S') = \int_{\mathbb{R}} |\text{CDF}[S](x) - \text{CDF}[S'](x)| dx$ is [15, Theorem 3]. Let the distributions S, S' consist of the same number m of equally weighted (not necessarily distinct) values $s_1 \leq \dots \leq s_m$ and $s'_1 \leq \dots \leq s'_m$. Then the area $\int_{\mathbb{R}} |\text{CDF}[S](x) - \text{CDF}[S'](x)| dx$ between the graphs $y = \text{CDF}[S](x)$ and $y = \text{CDF}[S'](x)$ can be split by the lines $y = \frac{k}{m}$ for $k = 1, \dots, m - 1$ into m horizontal slices. For each $k = 1, \dots, m$, in the slice right below the line $y = \frac{k}{m}$, the difference of areas is $\frac{|s_k - s'_k|}{m}$. Hence the total difference of areas is $\sum_{k=1}^m \frac{|s_k - s'_k|}{m} = \frac{1}{m} L_1(\vec{S}, \vec{S}')$. ■

Example 6.3 (ASD and ASM for the clouds $T, K \subset \mathbb{R}^2$ in Fig. 2). For the 4-point clouds T, K in Fig. 2, a base sequence A consists of two (unordered)

points. To compute $\frac{\sigma(A \cup \{q\})}{\lambda_2}$, we use the Lipschitz constant $\lambda_2 = \sqrt{3}$ and the strength of a triangle on $A \cup \{q\}$ in Example 4.8(b).

The trapezoid T in Fig. 4 has triangles $\triangle p_1 p_2 p_3$ and $\triangle p_1 p_2 p_4$ with sides $\sqrt{2}, 2, \sqrt{10}$, $\text{vol} = 1$, half-perimeter $p = \frac{\sqrt{2} + 2 + \sqrt{10}}{2}$, and normalized strength $\sigma_1 = \frac{\text{vol}^2}{\lambda_2 p^3} = \frac{1}{\sqrt{3}} \left(\frac{2}{\sqrt{2} + 2 + \sqrt{10}} \right)^3 \approx 0.016$.

The kite K in Fig. 5 has $\triangle p_1 p_2 p_3$ with sides $\sqrt{2}, \sqrt{2}, 2$, $\text{vol} = 1$, half-perimeter $p = 1 + \sqrt{2}$, normalized strength $\sigma_3 = \frac{1}{\sqrt{3}(1 + \sqrt{2})^3} \approx 0.041$.

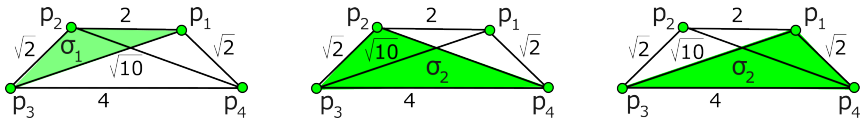


Figure 4. The trapezoid $T \subset \mathbb{R}^2$ on points $p_1 = (1, 1)$, $p_2 = (-1, 1)$, $p_3 = (-2, 0)$, $p_4 = (2, 0)$ has triangles with normalized strengths $\sigma_1 = \frac{1}{\sqrt{3}} \left(\frac{2}{\sqrt{2} + 2 + \sqrt{10}} \right)^3 \approx 0.016$ and $\sigma_2 = \frac{4}{\sqrt{3}} \left(\frac{2}{\sqrt{2} + \sqrt{10} + 4} \right)^3 \approx 0.029$, see Example 6.3.

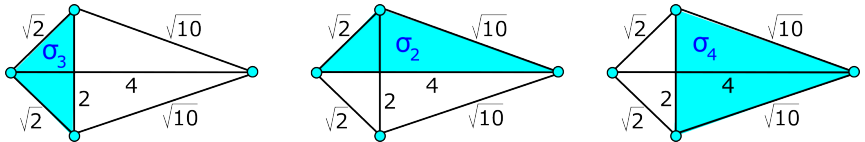


Figure 5. The kite cloud $K \subset \mathbb{R}^2$ on points $p_1 = (0, 1)$, $p_2 = (-1, 0)$, $p_3 = (0, -1)$, $p_4 = (3, 0)$ has triangles with normalized strengths $\sigma_3 = \frac{1}{\sqrt{3}(1 + \sqrt{2})^3} \approx 0.041$, $\sigma_2 = \frac{2^2}{\sqrt{3}} \left(\frac{2}{\sqrt{2} + \sqrt{10} + 4} \right)^3 \approx 0.029$, $\sigma_4 = \frac{3\sqrt{3}}{(1 + \sqrt{10})^3} \approx 0.072$.

Both T, K have triangles with sides $\sqrt{2}, \sqrt{10}, 4$, $\text{vol} = 2$, half-perimeter $p = \frac{\sqrt{2} + \sqrt{10} + 4}{2}$, and normalized strength $\sigma_2 = \frac{2^2}{\sqrt{3}} \left(\frac{2}{\sqrt{2} + \sqrt{10} + 4} \right)^3 \approx 0.029$. Finally, K has $\triangle p_1 p_3 p_4$ with sides $2, \sqrt{10}, \sqrt{10}$, $\text{vol} = 3$, $p = 1 + \sqrt{10}$ and normalized strength $\sigma_4 = \frac{3^2}{\sqrt{3}} \left(\frac{1}{1 + \sqrt{10}} \right)^3 = \frac{3\sqrt{3}}{(1 + \sqrt{10})^3} \approx 0.072$.

Table 3 summarizes Average Simplexwise Distributions ASD(T),

ASD(K), each consisting of 6 Average Simplexwise Vectors $ASV(T; A)$ and $ASV(K; A)$, respectively, of $1 + 2 + 2$ coordinates. The first coordinate is the distance between points of an unordered subset A . The two further coordinates are the average distances $\bar{d}(q)$ from Definition 6.1 in increasing order. The two last coordinates are normalized strengths, also in increasing order.

In the trapezoid T , the subset $A = \{p_1, p_4\}$ has $SPD(A)$ consisting of one distance $|p_1 - p_4| = \sqrt{2}$. The average distance to p_1 from p_2, p_3 is $\frac{2 + \sqrt{10}}{2}$. The average distance to p_4 from p_2, p_3 is $\frac{\sqrt{10} + 4}{2}$. The strengths of $\triangle p_1 p_2 p_4$ and $\triangle p_1 p_3 p_4$ are $\sigma_0 < \sigma_2$, respectively. The resulting Average Simplexwise Vector $ASV(T; \{p_1, p_4\}) \in \mathbb{R}^5$ is in the first cell of Table 3. Another subset $A = \{p_2, p_3\}$ gives the same vector $ASV(T; \{p_1, p_4\})$. This fact is marked by the ‘double’ factor $\times 2$ in front of vectors in Table 3.

T : Average Simplexwise Vectors	K : Average Simplexwise Vectors
$2 \times [\sqrt{2}; \frac{2+\sqrt{10}}{2}, \frac{\sqrt{10}+4}{2}; \sigma_1, \sigma_2]$ $\approx [1.41; 2.58, 3.58; 0.02, 0.03]$	$2 \times [\sqrt{2}; \frac{2+\sqrt{2}}{2}, \frac{\sqrt{10}+4}{2}; \sigma_2, \sigma_3]$ $\approx [1.41; 1.71, 3.58; 0.03, 0.04]$
$[2; \frac{\sqrt{2}+\sqrt{10}}{2}, \frac{\sqrt{2}+\sqrt{10}}{2}; \sigma_1, \sigma_1]$ $\approx [2; 2.58, 2.58; 0.02, 0.02]$	$[2; \sqrt{2}, \sqrt{10}; \sigma_3, \sigma_4]$ $\approx [2; 1.41, 3.16; 0.04, 0.07]$
$2 \times [\sqrt{10}; \frac{2+\sqrt{2}}{2}, \frac{\sqrt{2}+4}{2}; \sigma_1, \sigma_2]$ $\approx [3.16; 1.71, 2.71; 0.02, 0.03]$	$2 \times [\sqrt{10}; \frac{2+\sqrt{10}}{2}, \frac{\sqrt{2}+4}{2}; \sigma_2, \sigma_3]$ $\approx [3.16; 2.58, 2.71; 0.03, 0.04]$
$[4; \frac{\sqrt{2}+\sqrt{10}}{2}, \frac{\sqrt{2}+\sqrt{10}}{2}; \sigma_2, \sigma_2]$ $\approx [4; 2.29, 2.29; 0.03, 0.03]$	$[4; \frac{\sqrt{2}+\sqrt{10}}{2}, \frac{\sqrt{2}+\sqrt{10}}{2}; \sigma_2, \sigma_2]$ $\approx [4; 2.29, 2.29; 0.03, 0.03]$

Table 3. The Average Simplexwise Distributions from Definition 6.1 for the clouds $T, K \subset \mathbb{R}^2$ in Fig. 4, 5 consist of six vectors in \mathbb{R}^5 . The coordinate-wise averages of six vectors in ASD give the Average Simplexwise Moments $ASM \in \mathbb{R}^5$ in Table 4.

The kite K has two subsets subset $A = \{p_1, p_2\}$ and $A = \{p_2, p_3\}$ with $SPD(A)$ consisting of one distance $\sqrt{2}$. The average distance from p_4 and p_2 (or p_3) to p_1 is $\frac{4 + \sqrt{10}}{2}$, which appeared for T above. However, the

average distance from p_4 and p_3 to p_2 is $\frac{2 + \sqrt{2}}{2}$, so its difference with $\frac{2 + \sqrt{10}}{2}$ is **highlighted** in the first row of Table 3. Also, the normalized strengths of triangles $\triangle p_2 p_3 p_4$ and $\triangle p_1 p_2 p_3$ in K are $\sigma_2 < \sigma_1$, which differ from the strengths $\sigma_0 < \sigma_2$ for the similar subset A in the trapezoid T .

Rows 2 and 3 in Table 3 compare $ASV(T; A)$ and $ASV(K; A)$ for 2-point subsets A with inter-point distances 2 (one subset in both T, K) and $\sqrt{10}$ (two subsets in each of T, K). Row 4 in Table 3 contains the same vectors $ASV(T; A) = ASV(K; A)$ for A with inter-point distance 4, because each of T, K has two equal triangles with the largest side 4.

Table 4 shows the five coordinates of the Oriented Distance Moments $ASM(T; 1)$ for $t = 1$ in Definition 6.1 obtained by coordinate-wise averaging six vectors in Table 3. In the first row both averages are equal to $ASM_1 = \frac{1}{6}(2\sqrt{2} + 2 + 2\sqrt{10} + 4) = \frac{3 + \sqrt{2} + \sqrt{10}}{3} \approx 2.53$.

T : Average Simplexwise Moment	K : Average Simplexwise Moment
$ASM_1 = \frac{3 + \sqrt{2} + \sqrt{10}}{3} \approx 2.53$	$ASM_1 = \frac{3 + \sqrt{2} + \sqrt{10}}{3} \approx 2.53$
$ASM_2 = \frac{2 + \sqrt{2} + \sqrt{10}}{3} \approx 2.19$	$ASM_2 = \frac{8 + 5\sqrt{2} + 3\sqrt{10}}{12} \approx 2.05$
$ASM_3 = \frac{4 + \sqrt{2} + \sqrt{10}}{3} \approx 2.86$	$ASM_3 = \frac{16 + 3\sqrt{2} + 5\sqrt{10}}{12} \approx 3.01$
$ASM_4 = \frac{5\sigma_1 + \sigma_2}{6} \approx 0.018$	$ASM_4 = \frac{5\sigma_2 + \sigma_3}{6} \approx 0.031$
$ASM_5 = \frac{\sigma_1 + 5\sigma_2}{6} \approx 0.027$	$ASM_5 = \frac{\sigma_2 + 4\sigma_3 + \sigma_4}{6} \approx 0.045$

Table 4. Average Simplexwise Moments from Definition 6.1 for $T, K \subset \mathbb{R}^2$ in Fig. 4, 5 are obtained by averaging the six vectors in the Average Simplexwise Distributions ASD from Table 3.

The distance L_∞ between the vectors $ASM(T; 1)$ and $ASM(K; 1)$ in \mathbb{R}^5 is $\frac{\sigma_2 + 4\sigma_3 + \sigma_4}{6} - \frac{\sigma_1 + 5\sigma_2}{6} = \frac{\sigma_4 + 4\sigma_3 - 4\sigma_2 - \sigma_1}{6} \approx 0.018$. ■

The vector $ASD(C; A)$ in Definition 6.1 can be strengthened by including normalized strengths with signs, which distinguish mirror images. In this case, T and K would be distinguished also in last row 4 of Table 3 by the pairs of triangles with the common edge of length 4, because T has these triangles on one side of the edge, but K has them on different sides

of the edge. However, including signs creates a dependence on ordering points in a subset A of a cloud C , which makes the invariant larger.

The rest of this section will prove lower bounds for EMD_∞ on OSDs and SCDs in terms of simpler invariants from Definition 6.1 and its centered version in Definition 6.8 later. Recall that the matrix $D(A)$ of pairwise distances in Definition 3.1 and vector $\text{SPD}(A)$ of sorted pairwise distances in Definition 6.1 make sense for a sequence A of points in any metric space.

Proposition 6.4 (L_∞ bounds for sorted vectors). **(a)** For any vector $\mathbf{v} = (v_1, \dots, v_m)$, write its coordinates in increasing order $v_{\beta(1)} \leq \dots \leq v_{\beta(m)}$ for a permutation β of $1, \dots, m$. Set $\text{sort}(\mathbf{v}) = (v_{\beta(1)}, \dots, v_{\beta(m)})$. Then $L_\infty(\mathbf{u}, \mathbf{v}) \geq L_\infty(\text{sort}(\mathbf{u}), \text{sort}(\mathbf{v}))$ for any vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^m$.

(b) Any sequences A, A' of n ordered points in a metric space satisfy the inequality $L_\infty(D(A), D(A')) \geq L_\infty(\text{SPD}(A), \text{SPD}(A'))$.

Proof. **(a)** Set $\varepsilon = L_\infty(\mathbf{u}, \mathbf{v})$. For each $k = 1, \dots, m$, we will prove that the k -th sorted coordinates $\text{sort}_k(\mathbf{u})$ and $\text{sort}_k(\mathbf{v})$ are ε -close. Assume by contradiction that $\text{sort}_k(\mathbf{u}) < \text{sort}_k(\mathbf{v}) - \varepsilon$. Then $\text{sort}(\mathbf{u})$ has at least k coordinates less than $\text{sort}_k(\mathbf{v}) - \varepsilon$. Hence all their ε -close coordinates in \mathbf{v} are less than $\text{sort}_k(\mathbf{v})$, so the k -th ordered coordinate in $\text{sort}(\mathbf{u})$ is less than $\text{sort}_k(\mathbf{v})$, which contradicts the choice of $\text{sort}_k(\mathbf{v})$. The opposite assumption $\text{sort}_k(\mathbf{u}) > \text{sort}_k(\mathbf{v}) + \varepsilon$ similarly leads to a contradiction. Then $|\text{sort}_k(\mathbf{u}) - \text{sort}_k(\mathbf{v})| \leq \varepsilon$ for all $k = 1, \dots, m$ implies that $L_\infty(\mathbf{u}, \mathbf{v}) \leq \varepsilon$.

(b) Flattening the upper triangle of $D(A)$ to a vector \mathbf{v} of $m = \frac{n(n-1)}{2}$ indexed distances gives $\text{sort}(\mathbf{v}) = \text{SPD}(A)$. It remains to use part (a). ■

Proposition 6.5 (EMD respects bounds of base metrics). In the notations of Definition 5.3, let distributions $B = \{B_1, \dots, B_k\}$ and $C = \{C_1, \dots, C_l\}$ have base metrics d, d' such that $d(B_i, C_j) \geq d'(B_i, C_j)$ for all i, j . Then the Earth Mover's Distances EMD, EMD' based on the base metrics d, d' , respectively, satisfy $\text{EMD}(B, C) \geq \text{EMD}'(B, C)$.

Proof. Let $f_{ij} \in [0, 1]$ be optimal flows that minimize the Earth Mover's Distance $\text{EMD}(B, C) = \sum_{i,j} f_{ij} d(B_i, C_j)$. The given inequality $d \geq d'$ implies that $\text{EMD}(B, C) \geq \sum_{i,j} f_{ij} d'(B_i, C_j)$. Taking the minimum of the

right hand side over all $f_{ij} \in [0, 1]$ subject to the weight conditions in Definition 5.3 gives $\text{EMD}(B, C) \geq \text{EMD}'(B, C)$. \blacksquare

Recall that Definition 6.1 introduced the vectors $\vec{\mathbf{M}}(C; A)$ and $\vec{\mathbf{S}}(C; A)$ of averages $\bar{d}(q) = \frac{1}{n} \sum_{k=1}^n |\mathbf{q} - \mathbf{p}_k|$ and normalized strengths $\frac{\sigma(A \cup \{q\})}{\lambda_n}$, respectively, which are written in increasing order for all points $q \in C \setminus A$.

Lemma 6.6 (lower bounds for EMD_∞ on clouds). **(a)** Let $C \subset \mathbb{R}^n$ be any cloud of m unordered points and a base sequence $A \subset C$ of n ordered points p_1, \dots, p_n . For any point $q \in C \setminus A$, let $\check{d}(q) \in \mathbb{R}^n$ be the point with coordinates $|\mathbf{q} - \mathbf{p}_i|$ for $i = 1, \dots, n$. Denote by $\check{M}(C; A) \subset \mathbb{R}^n$ the cloud of $m - n$ unordered points $\check{d}(q)$ for all $q \in C \setminus A$. In the notations of Definition 5.5, any other cloud $C' \subset \mathbb{R}^n$ with a base sequence $A' \subset C'$ of n ordered points satisfies the inequalities

$$\text{EMD}_\infty(\hat{M}(C; A), \hat{M}(C'; A')) \geq \max \left\{ \text{EMD}_\infty(\check{M}(C; A), \check{M}(C'; A')), \text{EMD}_\infty(\vec{\mathbf{S}}(C; A), \vec{\mathbf{S}}(C'; A')) \right\},$$

$$\text{EMD}_\infty(\check{M}(C; A), \check{M}(C'; A')) \geq \text{EMD}_\infty(\vec{\mathbf{M}}(C; A), \vec{\mathbf{M}}(C'; A')), \text{ and}$$

$$\text{EMD}_\infty(\text{ORD}(C; A), \text{ORD}(C'; A')) \geq \text{EMD}_\infty(\text{ASV}(C; A), \text{ASV}(C'; A')).$$

(b) For any Minkowski metric L_q on \mathbb{R}^n for $q \in [1, +\infty]$ and clouds $C, C' \subset \mathbb{R}^n$, the Earth Mover's Distance EMD_q based on L_q has the lower bound in terms of weighted centers of mass: $\text{EMD}_q(C, C') \geq L_q(O(C), O(C'))$.

Proof. **(a)** Arbitrarily order the points of $C \setminus A$ and $C' \setminus A'$ as q_1, \dots, q_{m-n} and $q'_1, \dots, q'_{m'-n}$, respectively. Then all points of the clouds $\hat{M}(C; A)$ and $\hat{M}(C'; A')$ are ordered as $\hat{d}(q_1), \dots, q_{m-n}$ and $q'_1, \dots, q'_{m'-n}$, respectively. All values of the sets $S(C; A)$ and $S(C'; A')$ become ordered as $s_1(C; A), \dots, s_{m-n}(C; A)$ and $s_1(C'; A'), \dots, s_{m'-n}(C'; A')$, respectively.

Due to Proposition 6.5(a), to prove the first lower bound, it suffices to check two similar inequalities for the base metric in \mathbb{R}^{n+1} , i.e.

$$L_\infty(\check{d}(q_i), \check{d}(q'_j)) \leq L_\infty(\hat{d}(q_i), \hat{d}(q'_j)) \geq |s_i(C; A) - s_j(C'; A')|.$$

The former inequality holds, because each $L_\infty(\hat{d}(q_i), \hat{d}(q'_j))$ is the maximum absolute difference between corresponding coordinates of the points

$\hat{d}(q_i), \hat{d}(q'_j) \in \mathbb{R}^{n+1}$ whose first n coordinates form $\check{d}(q_i), \check{d}(q'_j) \in \mathbb{R}^n$. The latter inequality will follow below from the fact that the last $(n+1)$ -st coordinates are normalized signed strengths $\frac{s(A \cup \{q_i\})}{\lambda_n}$ and $\frac{s(A' \cup \{q'_j\})}{\lambda_n}$, respectively. By Definition 5.5, each $s(A \cup \{q_i\}) = s(q_i; A)\sigma(A \cup \{q_i\})$ equals the strength up to the sign factor $s(q_i; A)$ in Definition 3.1. Then

$$|s(A \cup \{q_i\}) - s(A' \cup \{q'_j\})| \geq |\sigma(A \cup \{q_i\}) - \sigma(A' \cup \{q'_j\})|, \text{ and hence}$$

$$L_\infty(\hat{d}(q_i), \hat{d}(q'_j)) \geq \frac{|s(A \cup \{q_i\}) - s(A' \cup \{q'_j\})|}{\lambda_n} \geq |s_i(C; A) - s_j(C'; A')|.$$

The last expression is the distance L_∞ between indexed elements of the unordered distributions $S(C; A)$ and $S(C'; A')$ of normalized strengths. Since EMD_∞ is independent of an order of objects, these distributions can be replaced with the vectors $\vec{S}(C; A)$ and $\vec{S}(C'; A')$ from Definition 6.1.

Now we will prove the lower bound involving the vectors $\vec{M}(C; A)$ and $\vec{M}(C'; A')$ of average distances. Each point $\check{d}(q_i) \in \mathbb{R}^n$ consists of n coordinates $\check{d}_k(q_i) = |\mathbf{q}_i - \mathbf{p}_k|$ for $k = 1, \dots, n$. The base metric equals Chebyshev distance $L_\infty(\check{d}(q_i), \check{d}(q_j))$ has the lower bound equal to the average of absolute differences between corresponding coordinates, i.e.

$$L_\infty(\check{d}(q_i), \check{d}(q_j)) \geq \frac{1}{n} \sum_{k=1}^n |d_k(q_i) - d_k(q'_j)| \geq \frac{1}{n} \left| \sum_{k=1}^n d_k(q_i) - \sum_{k=1}^n d_k(q'_j) \right| =$$

$|\bar{d}(q_i) - \bar{d}(q'_j)|$, which is the distance L_∞ between corresponding coordinates of $\vec{M}(C; A)$ and $\vec{M}(C'; A')$, whose re-ordering does not affect the Earth Mover's Distance. The proved inequality between the base metrics gives rise to the same inequality between EMDs by Proposition 6.5(a).

To prove the final inequality, let $\xi \in S_n$ be the permutation that minimizes $\text{EMD}_\infty[\xi]$. By re-ordering the base sequence $A \subseteq C$, we can assume that ξ is trivial. Then $\text{EMD}_\infty(\text{ORD}(C; A), \text{ORD}(C'; A'))$ equals

$$e = \max \{L_\infty(D(A), D(A')), \text{EMD}_\infty(\hat{M}(C; A), \hat{M}(C'; A'))\}.$$

The first lower bound $e \geq L_\infty(D(A), D(A')) \geq L_\infty(\text{SPD}(A), \text{SPD}(A'))$

holds by Proposition 6.4(b). The other lower bounds were proved above:

$$e \geq \text{EMD}_\infty(\hat{M}(C; A), \hat{M}(C'; A')) \geq \max \left\{ \text{EMD}_\infty(\vec{M}(C; A), \vec{M}(C'; A')), \text{EMD}_\infty(\vec{S}(C; A), \vec{S}(C'; A')) \right\}.$$

(b) It was proved in [15, Theorem 3]. ■

In Lemma 6.6(a), if given clouds $C, C' \subset \mathbb{R}^n$ consists of the same number m of points, then Proposition 6.2 provides the explicit formulas:

$$\text{EMD}_\infty(\vec{M}(C; A), \vec{M}(C'; A')) = \frac{1}{m} L_1(\vec{M}(C; A), \vec{M}(C'; A')),$$

$$\text{EMD}_\infty(\vec{S}(C; A), \vec{S}(C'; A')) = \frac{1}{m} L_1(\vec{S}(C; A), \vec{S}(C'; A')).$$

Theorem 6.7 (properties of metrics on ASDs and ASMs). (a) For any $t \geq 1$ and a cloud $C \subset \mathbb{R}^n$ of m unordered points, the Average Simplexwise Distribution $\text{ASD}(C)$ and the Average Simplexwise Moment $\text{ASM}(C; t)$ in Definition 6.1 can be computed in time $O(n^3 m^{n+1} \log m)$.

(b) For any clouds $C, C' \subset \mathbb{R}^n$ of up to m points, the Earth Mover's Distance between $\text{ASD}(C), \text{ASD}(C')$ s of a maximum size $l \leq \binom{m}{n} = O(m^n)$ in Definition 6.1(b) can be computed in time $O(l^3 \log l)$ and provides the lower bound $\text{EMD}_\infty(\text{OSD}(C), \text{OSD}(C')) \geq \text{EMD}_\infty(\text{ASD}(C), \text{ASD}(C'))$. Hence, $\text{ASD}(C)$ has Lipschitz constant 2 under any perturbations of C .

(c) For any $t \geq 1$ and clouds $C, C' \subset \mathbb{R}^n$ with up to m points and pre-computed ASMs, $\text{EMD}_\infty(\text{ASM}(C; t), \text{ASM}(C'; t))$ in Definition 6.1(c) is computed in extra time $O((n^6 + m^3) \log m)$ and provides the lower bound

$$\text{EMD}_\infty(\text{ASD}(C), \text{ASD}(C')) \geq \text{EMD}_\infty(\text{ASM}(C; 1), \text{ASM}(C'; 1)),$$

hence $\text{ASM}(C)$ has Lipschitz constant 2 under any perturbations of C .

(d) Any clouds $C, C' \subset \mathbb{R}^n$ of m unordered points satisfy the inequalities

$$\begin{aligned} \text{BD}_\infty(\text{OSD}(C), \text{OSD}(C')) &\geq \text{EMD}_\infty(\text{OSD}(C), \text{OSD}(C')) \geq \\ \text{EMD}_\infty(\text{ASD}(C), \text{ASD}(C')) &\geq \text{EMD}_\infty(\text{ASM}(C; 1), \text{ASM}(C'; 1)). \end{aligned}$$

Proof. **(a)** To compute $\text{OSD}(C)$, we consider $k = \binom{m}{n} = \frac{m!}{n!(m-n)!} = O(m^n)$ unordered subsets $A \subseteq C$ of n points. The vector $\text{SPD}(A)$ of $\frac{n(n-1)}{2}$ sorted pairwise distances can be computed in time $O(n^2 \log n)$. The vector $\vec{\mathbf{M}}(C; A)$ consists of $m-n$ sorted averages based on $n(m-n)$ distances, all computable in time $O(mn + m \log m)$. The vector $\vec{\mathbf{S}}(C; A)$ consists of $m-n$ average strengths, each computable in time $O(n^3)$ by Theorem 4.9. Then $\text{ASV}(C; A) = [\text{SPD}(A); \vec{\mathbf{M}}(C; A); \vec{\mathbf{S}}(C; A)]$ is computed in time $O(n^3 m \log m)$. Multiplying the last time by the number $k = O(m^n)$ of n -point subsets $A \subseteq C$, the time of $\text{ASD}(C)$ becomes $O(n^3 m^{n+1} \log m)$. After that, the t -th moment $\text{ASM}(C; t)$ needs extra time $O(m^n)$ for each of $\frac{n(n-1)}{2} + 2(m-n)$ coordinates, which keeps the total time the same.

(b) In the notations of Definition 6.1, the distance $L_\infty(\text{SPD}(A), \text{SPD}(A'))$ on $\frac{n(n-1)}{2}$ ordered distances needs time $O(n^2 \log n)$. The vectors $\vec{\mathbf{M}}(C; A)$, $\vec{\mathbf{M}}(C'; A')$, $\vec{\mathbf{S}}(C; A)$, $\vec{\mathbf{S}}(C'; A')$ have $O(m)$ coordinates. Hence any distance EMD_∞ between them needs time $O(m^3 \log m)$, see [22, 23]. Since $n \leq m$, the time for $\text{EMD}_\infty(\text{ASV}(C; A), \text{ASV}(C'; A'))$ is $O(m^3 \log m)$.

If the distributions $\text{ASD}(C), \text{ASD}(C')$ consist of $l \leq \binom{m}{n} = O(m^n)$ unordered ASVs based on subsets of n points, $\text{EMD}_\infty(\text{ASD}(C), \text{ASD}(C'))$ can be computed in time $O(l^3 \log l) = O(nm^{3n} \log m)$. Both $\text{OSD}(C)$ and $\text{ASD}(C)$ can be considered unordered distributions of equally weighted pairs $(C; A)$ with the base metrics $d = \text{EMD}_\infty(\text{ORD}(C; A), \text{ORD}(C'; A'))$, $d' = \text{EMD}_\infty(\text{ASV}(C; A), \text{ASV}(C'; A'))$, respectively. Then $d \geq d'$ from Lemma 6.6(a) implies the same inequality for EMD_∞ by Proposition 6.5. This inequality and continuity of EMD_∞ on OSDs with Lipschitz constant 2 in Theorem 5.10 imply the same continuity of EMD_∞ on ASDs.

(c) For a fixed $t \geq 1$, the Average Simplexwise Moment $\text{ASM}(C; t)$ consist of the vector $\mu_t\{\text{SPD}(A)\}$ of $\frac{n(n-1)}{2} = O(n^2)$ coordinates and the vectors $\mu_t\{\vec{\mathbf{M}}(C; A)\}$ and $\mu_t\{\vec{\mathbf{S}}(C; A)\}$ of $m-n = O(m)$ coordinates each. Computing EMD_∞ between vectors of k coordinates needs time $O(k^3 \log k)$, which gives the overall time $O((n^6 + m^3) \log m)$.

To prove the lower bound for EMD_∞ on $\text{ASM}(C; 1)$, separately consider the distributions $\text{SPD}(A)$, $\vec{\mathbf{M}}(C; A)$, and $\vec{\mathbf{S}}(C; A)$ of subvectors. Then apply Lemma 6.6(b) to the centers of each distribution as follows:

$$\begin{aligned} \text{EMD}_\infty(\text{SPD}(A), \text{SPD}(A')) &\geq L_\infty(\mu_1\{\text{SPD}(A)\}, \mu_1\{\text{SPD}(A')\}), \\ \text{EMD}_\infty(\vec{\mathbf{M}}(C; A), \vec{\mathbf{M}}(C'; A')) &\geq L_\infty(\mu_1\{\vec{\mathbf{M}}(C; A)\}, \mu_1\{\vec{\mathbf{M}}(C'; A')\}), \\ \text{EMD}_\infty(\vec{\mathbf{S}}(C; A), \vec{\mathbf{S}}(C'; A')) &\geq L_\infty(\mu_1\{\vec{\mathbf{S}}(C; A)\}, \mu_1\{\vec{\mathbf{S}}(C'; A')\}). \end{aligned}$$

The inequalities imply the similar required inequality for the overall distances EMD_∞ , which were introduced in Definition 6.1(b,c) as the maxima of three EMD_∞ distances for both ASD and ASM. The proved lower bounds and the continuity of EMD_∞ on OSDs with Lipschitz constant 2 in Theorem 5.10 imply the same continuity of EMD_∞ on ASMs.

(d) The first inequality appeared in Theorem 5.8(a). The other two inequalities were proved in parts (b,c) for any clouds $C, C' \subset \mathbb{R}^n$. ■

Though the rough time estimate $O(nm^{3n} \log m)$ for EMD_∞ on ASDs in Theorem 6.7(b) is asymptotically the same as for EMD_∞ on OSDs in Theorem 5.8(b), averaging over points of $C \setminus A$ likely makes $\text{ASD}(C)$ smaller than $\text{OSD}(C)$ for symmetric clouds, as shown in Table 3.

The time $O((n^6 + m^3) \log m)$ for EMD_∞ on vectors ASMs is certainly smaller than $O(nm^{3n} \log m)$ for practical dimensions $n = 2, 3$. We will reduce these times by using the center of mass as an extra base point.

Definition 6.8 (Average Centered Moments $\text{ACM}(C; t)$). Let $C \subset \mathbb{R}^n$ be a cloud of m unordered points with a center of mass $O(C)$. For a base sequence $A \subset C$ of $n - 1$ ordered points p_1, \dots, p_{n-1} , let the extended sequence $\tilde{A} = A \cup O(C)$ include $O(C)$ as the last n -th point. The vector $\vec{\mathbf{D}}(C; A)$ is formed by $n - 1$ distances $|\mathbf{p}_i - \mathbf{O}(C)|$, $i = 1, \dots, n - 1$, written in increasing order. The vector $\vec{\mathbf{O}}(C; A)$ is formed by $m - n + 1$ distances $|\mathbf{q} - \mathbf{O}(C)|$, written in increasing order, for all points $q \in C \setminus A$.

(a) In the notations of Definition 6.1, the *Average Centered Vector*

$$\text{ACV}(C; A) = [\text{SPD}(A); \vec{\mathbf{D}}(C; A); \vec{\mathbf{M}}(C; A); \vec{\mathbf{O}}(C; A); \vec{\mathbf{S}}(C; \tilde{A})]$$

has $\frac{n(n-1)}{2} + 3(m-n+1)$ coordinates. For any cloud $C' \subset \mathbb{R}^n$ of m' unordered points with a base sequence A' of $n-1$ points, define $\text{EMD}_\infty(\text{ACV}(C; A), \text{ACV}(C'; A'))$ as the maximum of the five distances

$$\begin{aligned} &L_\infty(\text{SPD}(A), \text{SPD}(A')), \\ &L_\infty(\vec{\mathbf{D}}(C; A), \vec{\mathbf{D}}(C'; A')), \\ &\text{EMD}_\infty(\vec{\mathbf{M}}(C; A), \vec{\mathbf{M}}(C'; A')), \\ &\text{EMD}_\infty(\vec{\mathbf{O}}(C; A), \vec{\mathbf{O}}(C'; A')), \\ &\text{EMD}_\infty(\vec{\mathbf{S}}(C; \tilde{A}), \vec{\mathbf{S}}(C'; \tilde{A}')). \end{aligned}$$

(b) The *Average Centered Distribution* $\text{ACD}(C)$ is the unordered set of equally weighted vectors $\text{ACV}(C; A)$ for all $\binom{m}{n-1}$ unordered subsets $A \subseteq C$ of $n-1$ points. Use the base metric EMD_∞ on ACVs in part (a) to define the Earth Mover's Distance $\text{EMD}_\infty(\text{ACD}(C), \text{ACD}(C'))$.

(c) For $t \geq 1$, the *Average Centered Moment* $\text{ACM}(C; t) = \mu_t\{\text{ACD}(C)\}$ is the t -th moment of $\text{ACD}(C)$ of equally weighted vectors $\text{ACV}(C; A)$ over all $\binom{m}{n-1}$ unordered subsets $A \subseteq C$ of $n-1$ points, i.e. $\text{ACM}(C; t) =$

$$[\mu_t\{\text{SPD}(A)\}; \mu_t\{\vec{\mathbf{D}}(C; A)\}; \mu_t\{\vec{\mathbf{M}}(C; A)\}; \mu_t\{\vec{\mathbf{O}}(C; A)\}; \mu_t\{\vec{\mathbf{S}}(C; \tilde{A})\}].$$

For any clouds $C, C' \subset \mathbb{R}^n$, define $\text{EMD}_\infty(\text{ACM}(C; t), \text{ACM}(C'; t))$ as the maximum of the five metrics EMD_∞ between constituent subvectors. ■

$\text{ACD}(C)$	Average Centered Distribution	Definition 6.8
$\text{ACM}(C; t)$	Average Centered t -th Moment	Definition 6.8
$\text{ACV}(C; A)$	Average Centered Vector for $A \subset C$	Definition 6.8
$\text{ASD}(C)$	Average Simplexwise Distribution	Definition 6.1
$\text{ASM}(C; t)$	Average Simplexwise t -th Moment	Definition 6.1
$\text{ASV}(C; A)$	Average Simplexwise Vector for $A \subset C$	Definition 6.1
$\text{BMD}(\Gamma)$	Bottleneck Matching Distance of a graph	Definition 5.1
$\text{EMD}(B, C)$	Earth Mover's Distance on distributions	Definition 5.3

Table 5. Acronyms and references for all concepts in sections 5 and 6.

$\text{ACV}(C; A)$ in Definition 6.8(a) uses the extended base sequence \tilde{A} only in the last vector $\vec{\mathbf{S}}(C; A)$ because the strengths need $n+1$ points from $\tilde{A} \cup \{q\}$. Example 6.9 simplifies the complete invariant SCD of regular clouds from Example 4.2 to distributions and moments from Definition 6.8.

Example 6.9 (distances between molecules and clouds in \mathbb{R}^2). **(a)** A linear molecule CO_2 of carbon dioxide can be represented in the horizontal x -axis of \mathbb{R}^2 by two oxygen atoms O_1, O_2 at $x \approx \pm 1.2\text{\AA}$ and one carbon atom C at the origin $x = 0$, which coincides with the center of mass. Here $1\text{\AA} = 10^{-10}$ meter, see Fig. 6 (left). We compute the Simplexwise Centered Distribution $\text{SCD}(\text{CO}_2)$ by using three 1-point base sequences A_i for $i = 0, 1, 2$ so that all triangles on atoms of CO_2 are degenerate and have zero strengths. For the base sequence $A_0 = \{C\}$, the Oriented

Centered Distribution is $\text{OCD}(\text{CO}_2; A_0) = \left[0, \begin{pmatrix} 1.2 & 1.2 \\ 1.2 & 1.2 \\ 0 & 0 \end{pmatrix} \right]$.

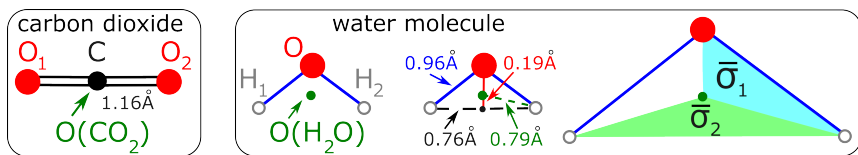


Figure 6. **Left:** a molecule of carbon dioxide CO_2 with a bond $\text{C}=\text{O}$ of length 1.16\AA , see Example 6.9 (a). **Right:** a water molecule H_2O forms an isosceles triangle with equal sides of length 0.96\AA . Triangles around the center of mass $O(\text{H}_2\text{O})$ have normalized strengths $\bar{\sigma}_1, \bar{\sigma}_2$ computed in Example 6.9 (b).

By Definition 6.8, $\text{SPD}(A_0)$ is empty, $\vec{\mathbf{D}}(\text{CO}_2; A_0)$ consists of one distance $|C| = 0$, $\vec{\mathbf{M}}(\text{CO}_2; A_0)$ consists of two distances $|O_i - C| = 1.2$, $\vec{\mathbf{O}}(\text{CO}_2; A_0)$ consists of two distances $|O_i| = 1.2$, and $\vec{\mathbf{S}}(\text{CO}_2; A_0)$ consists of two strengths 0, so $\text{ACV}(\text{CO}_2; A_0) = [0; 1.2, 1.2; 1.2, 1.2; 0, 0] \in \mathbb{R}^7$.

For $A_i = \{O_i\}$, $i = 1, 2$, we get $\text{OCD}(\text{CO}_2; A_i) = \left[1.2, \begin{pmatrix} 1.2 & 2.4 \\ 0 & 1.2 \\ 0 & 0 \end{pmatrix} \right]$ and $\text{ACV}(\text{CO}_2; A_i) = [1.2; 1.2, 2.4; 0, 1.2; 0, 0]$. For simplicity, let all atoms be equally weighted. Then $\text{SCD}(\text{CO}_2)$ consists of $\text{OCD}(\text{CO}_2; A_0)$ of weight $\frac{1}{3}$ and $\text{OCD}(\text{CO}_2; A_1)$ of weight $\frac{2}{3}$. The Average Centered Distribution $\text{ACD}(\text{CO}_2)$ consists of $\text{ACV}(\text{CO}_2; A_0)$ of weight $\frac{1}{3}$ and $\text{ACV}(\text{CO}_2; A_1)$

of weight $\frac{2}{3}$. The 1st Average Centered Moment is $\text{ACM}(\text{CO}_2; 1) =$

$$= \frac{1}{3}\text{ACV}(\text{CO}_2; A_0) + \frac{2}{3}\text{ACV}(\text{CO}_2; A_1) = [0.8; 1.2, 2; 0.4, 1.2; 0, 0].$$

To compute the 2nd moment, the distribution $S = \{(0, \frac{1}{3}), (1.2, \frac{2}{3})\}$ of distances 0, 1.2 with weights $\frac{1}{3}, \frac{2}{3}$, respectively, has the average $\mu_1 = 0.8$

and the standard deviation $\mu_2\{S\} = \sqrt{\frac{1}{3}(0 - 0.8)^2 + \frac{2}{3}(1.2 - 0.8)^2} \approx 0.57$.

The distributions $\{(1.2, \frac{1}{3}), (0, \frac{2}{3})\}$ and $\{(1.2, \frac{1}{3}), (2.4, \frac{2}{3})\}$ have the same standard deviation μ_2 , so $\text{ACM}(\text{CO}_2; 2) = [0.57; 0, 0.57; 0.57, 0; 0, 0]$.

(b) A molecule H_2O of water can be represented in \mathbb{R}^2 by two hydrogen atoms H_1, H_2 at $(x, y) \approx (\pm 0.76, 0)$ and one oxygen atom O at $(x, y) = (0, 0.59)$, so the center of mass is $O(\text{H}_2\text{O}) \approx (0, 0.2)$ and the H-O bond length is 0.96\AA , see Fig. 6. For the base sequence $A_0 = \{O\}$, the triangle on $O, H_i, O(\text{H}_2\text{O})$ has half-perimeter $p \approx \frac{0.4 + 0.96 + 0.79}{2} \approx 1.07$, vol \approx

0.15 , and normalized strength $\bar{\sigma}_1 = \frac{\text{vol}^2}{p^3 \lambda_2} = \frac{0.15^2}{1.07^3 \sqrt{3}} \approx 0.01$. The Ori-

ented Centered Distribution is $\text{OCD}(\text{H}_2\text{O}; A_0) = \left[0.4, \begin{pmatrix} 0.96 & 0.96 \\ 0.79 & 0.79 \\ +0.01 & -0.01 \end{pmatrix} \right]$.

By Definition 6.8, the vector $\text{SPD}(A_0)$ is empty, the vector $\vec{\mathbf{D}}(\text{H}_2\text{O}; A_0)$ consists of one distance 0.4, the vector $\vec{\mathbf{M}}(\text{H}_2\text{O}; A_0)$ consists of two distances $|\mathbf{H}_i - \mathbf{O}| = 0.96$, the vector $\vec{\mathbf{O}}(\text{H}_2\text{O}; A_0)$ consists of two distances 0.79, and $\vec{\mathbf{S}}(\text{H}_2\text{O}; \tilde{A}_0)$ consists of two normalized strengths 0.01, so

$$\text{ACV}(\text{H}_2\text{O}; A_0) = [0.4; 0.96, 0.96; 0.79, 0.79; 0.01, 0.01] \in \mathbb{R}^7.$$

For the base sequence $A_i = \{H_i\}$, $i = 1, 2$, the triangle on $H_1, H_2, O(\text{H}_2\text{O})$ has half-perimeter $p \approx 0.76 + 0.96 \approx 1.72$, vol ≈ 0.19 , and the normal-

ized strength $\bar{\sigma}_2 = \frac{\text{vol}^2}{p^3 \lambda_2} = \frac{0.19^2}{1.72^3 \sqrt{3}} \approx 0.004$. The Oriented Centered

Distribution is $\text{OCD}(\text{H}_2\text{O}; A_i) = \left[0.79, \begin{pmatrix} 0.96 & 1.92 \\ 0.4 & 0.79 \\ 0.01(-1)^i & 0.004(-1)^{i-1} \end{pmatrix} \right]$.

By Definition 6.8, $\text{SPD}(A_0)$ is empty, the vector $\vec{\mathbf{D}}(\text{H}_2\text{O}; A_i)$ consists

of one distance 0.79, the vector $\vec{\mathbf{M}}(\text{H}_2\text{O}; A_i)$ consists of two distances 0.96, 1.92, the vector $\vec{\mathbf{O}}(\text{H}_2\text{O}; A_0)$ consists of two distances 0.4, 0.79, and $\vec{\mathbf{S}}(\text{H}_2\text{O}; \tilde{A}_0)$ consists of the normalized strengths 0.004, 0.01, so

$$\text{ACV}(\text{H}_2\text{O}; A_i) = [0.79; 0.96, 1.92; 0.4, 0.79; 0.004, 0.01] \in \mathbb{R}^7, \quad i = 1, 2.$$

Let all atoms be equally weighted. Then $\text{SCD}(\text{H}_2\text{O})$ consists $\text{OCD}(\text{H}_2\text{O}; A_0)$ of weight $\frac{1}{3}$ and $\text{OCD}(\text{H}_2\text{O}; A_1)$ of weight $\frac{2}{3}$, similarly for the Average Centered Distribution $\text{ACD}(\text{H}_2\text{O})$. The 1st Average Centered Moment is

$$\begin{aligned} \text{ACM}(\text{H}_2\text{O}; 1) &= \frac{1}{3}\text{ACV}(\text{H}_2\text{O}; A_0) + \frac{2}{3}\text{ACV}(\text{H}_2\text{O}; A_1) = \\ &= [0.66; 0.96, 1.6; 0.53, 0.79; 0.006, 0.01]. \end{aligned}$$

The 2nd Average Centered Moment of standard deviations is

$$\text{ACM}(\text{H}_2\text{O}; 2) = [0.18; 0, 0.45; 0.18, 0; 0.003, 0].$$

(c) To compare the molecules CO_2 and H_2O in parts (a,b), we can compute EMD_∞ from Definition 6.8(c) between corresponding parts of moment vectors by Proposition 6.2 for EMD_∞ between equal-size distributions:

$$\begin{aligned} \text{EMD}_\infty(\text{ACM}(\text{CO}_2; 1), \text{ACM}(\text{H}_2\text{O}; 1)) &= \max\left\{ |0.8 - 0.66|, \right. \\ &\quad \left. \frac{|1.2 - 0.96| + |2 - 1.6|}{2}, \frac{|0.4 - 0.53| + |1.2 - 0.79|}{2}, \frac{0.006 + 0.01}{2} \right\} = \\ &= \max\{0.14, 0.32, 0.27, 0.008\} = 0.32, \end{aligned}$$

$$\begin{aligned} \text{EMD}_\infty(\text{ACM}(\text{CO}_2; 2), \text{ACM}(\text{H}_2\text{O}; 2)) &= \\ \max\left\{ |0.57 - 0.18|, \frac{|0.45 - 0.57|}{2}, \frac{|0.57 - 0.18|}{2}, \frac{0.003}{2} \right\} &= \\ = \max\{0.39, 0.006, 0.02, 0.0015\} &= 0.39. \end{aligned}$$

(d) The cloud $C_3 \subset \mathbb{R}^2$ at the vertices p_1, p_2, p_3 of an equilateral triangle has the center of mass $O(C_3) = \mathbf{0}$ at the origin. For any base sequence $A_i = \{p_i\}$ of one point, and another point $p_j \neq p_i$, the triangle $\Delta p_i p_j \mathbf{0}$ has $\text{vol} = \frac{\sqrt{3}}{4}$, half-perimeter $p = \frac{2 + \sqrt{3}}{2}$, and the normalized strength

$\bar{\sigma}_3 = \frac{\text{vol}^2}{p^3 \lambda_2} = \frac{(\sqrt{3})^2 2^3}{4^2 (2 + \sqrt{3})^3 \sqrt{3}} = \frac{\sqrt{3}}{2(2 + \sqrt{3})^3} \approx 0.017$. Example 4.2(a) has

$$\text{the Oriented Centered Distribution } \text{OCD}(C_3; A_i) = \left[1, \begin{pmatrix} \sqrt{3} & \sqrt{3} \\ 1 & 1 \\ -1 & +1 \end{pmatrix} \right].$$

By Definition 6.8, $\text{SPD}(A_i)$ is empty, $\vec{\mathbf{D}}(C_3; A_i)$ consists of one distance $|p_i| = 1$, the vector $\vec{\mathbf{M}}(C_3; A_i)$ consists of two distances $|\mathbf{q} - \mathbf{p}_i| = \sqrt{3}$, the vector $\vec{\mathbf{O}}(C_3; A_i)$ consists of two distances $|\mathbf{q}| = 1$, and $\vec{\mathbf{S}}(C_3; \tilde{A}_i)$ consists of two normalized strengths $\frac{\sigma(\tilde{A}_i \cup \{q\})}{\lambda_2} = \bar{\sigma}_3$ for both points $q \in C_3 \setminus \{p_i\}$.

The Average Centered Vector is $\text{ACV}(C_3; A_i) = [1; \sqrt{3}, \sqrt{3}; 1, 1; \bar{\sigma}_3, \bar{\sigma}_3] \in \mathbb{R}^7$. Since $\text{ACV}(C_3; A_i)$ is independent of $i = 1, 2, 3$, the Average Centered Distribution $\text{ACD}(C_3)$ and the Average Centered Moment $\text{ACM}(C_3; 1)$ coincide with $[1; \sqrt{3}, \sqrt{3}; 1, 1; \bar{\sigma}_3, \bar{\sigma}_3]$, and $\text{ACM}(C_3; t) = \mathbf{0}$ for $t > 1$.

(e) The cloud $C_4 \subset \mathbb{R}^2$ at the vertices p_1, p_2, p_3, p_4 of a square has the center of mass $O(C_4) = \mathbf{0}$ at the origin. For any base sequence $A_i = \{p_i\}$ of one point, and another point $p_{i \pm 1}$, where $i \pm 1$ is taken modulo 4 in $\{1, 2, 3, 4\}$, the triangle $\Delta p_i p_{i \pm 1} \mathbf{0}$ has $\text{vol} = \frac{1}{2}$, half-perimeter $p = \frac{2 + \sqrt{2}}{2}$, and the normalized strength $\bar{\sigma}_4 = \frac{\text{vol}^2}{p^3 \lambda_2} = \frac{2^3}{2^2 (2 + \sqrt{2})^3 \sqrt{3}} = \frac{2}{\sqrt{3}(2 + \sqrt{3})^3} \approx 0.022$. Any other triangle $\Delta p_i p_{i+2} \mathbf{0}$ is degenerate, so its strength is 0. Example 4.2(b) has the Oriented Centered Distribu-

$$\text{tion } \text{OCD}(C_4; \tilde{A}_i) = \left[1, \begin{pmatrix} \sqrt{2} & 2 & \sqrt{2} \\ 1 & 1 & 1 \\ -1 & 0 & +1 \end{pmatrix} \right]. \text{ Then } \text{SPD}(A_i) \text{ is empty,}$$

$\vec{\mathbf{D}}(C_4; A_i)$ consists of one distance $|p_i| = 1$, the vector $\vec{\mathbf{M}}(C_4; A_i)$ consists of three distances $\sqrt{2}, \sqrt{2}, 2$, the vector $\vec{\mathbf{O}}(C_3; A_i)$ consists of three distances $|\mathbf{q}| = 1$, and $\vec{\mathbf{S}}(C_3; \tilde{A}_i)$ consists of 0 and two normalized strengths $\frac{\sigma(\tilde{A}_i \cup \{q\})}{\lambda_2} = \bar{\sigma}_4$. The Average Centered Vector is $\text{ACV}(C_4; A_i) = [1; \sqrt{2}, \sqrt{2}, 2; 1, 1, 1; 0, \bar{\sigma}_4, \bar{\sigma}_4] \in \mathbb{R}^{10}$. The vector $\text{ACV}(C_4; A_i)$ is independent of i , then $\text{ACD}(C_4)$ and $\text{ACM}(C_4; 1)$ coincide with this vector.

(f) To compare the triangular and square clouds C_3, C_4 , we compute the metric $\text{EMD}_\infty(\text{ACD}(C_3), \text{ACD}(C_4))$ as the maximum of three distances

EMD_∞ , which can be found by Proposition 6.2: $\text{EMD}_\infty([1, 1], [1, 1, 1]) = 0$,

$$\begin{aligned} \text{EMD}_\infty([\sqrt{3}, \sqrt{3}], [\sqrt{2}, \sqrt{2}, 2]) &= \frac{1}{3}(2 - \sqrt{3}) + \frac{2}{3}(\sqrt{3} - \sqrt{2}) \approx 0.3, \\ \text{EMD}_\infty([\bar{\sigma}_3, \bar{\sigma}_3], [0, \bar{\sigma}_4, \bar{\sigma}_4]) &= \frac{1}{3}(\bar{\sigma}_3 - 0) + \frac{2}{3}(\bar{\sigma}_4 - \bar{\sigma}_3) \approx 0.003, \text{ so} \end{aligned}$$

$\text{EMD}_\infty(\text{ACD}(C_3), \text{ACD}(C_4)) = \text{EMD}_\infty(\text{ACM}(C_3; 1), \text{ACM}(C_4; 1)) \approx 0.3$, which slightly differs from another $\text{EMD} = \frac{5}{16}$ in [33, Example 6.7]. ■

Theorem 6.10 (properties of metrics on ACDs and ACMs). **(a)** For any $t \geq 1$ and a cloud $C \subset \mathbb{R}^n$ of m unordered points, the Average Centered Distribution $\text{ACD}(C)$ and the Average Centered Moment $\text{ACM}(C; t)$ in Definition 6.8 can be computed in time $O(n^3 m^n \log m)$.

(b) For any clouds $C, C' \subset \mathbb{R}^n$ of up to m points, the Earth Mover's Distance between $\text{ACD}(C), \text{ACD}(C')$ of a maximum size $l \leq \binom{m}{n-1} = O(m^{n-1})$ in Definition 6.8(b) can be computed in time $O(l^3 \log l)$ and satisfies $\text{EMD}_\infty(\text{OCD}(C), \text{OCD}(C')) \geq \text{EMD}_\infty(\text{ACD}(C), \text{ACD}(C'))$, hence $\text{ACD}(C)$ has Lipschitz constant 2 under any perturbations of C .

(c) For any $t \geq 1$ and clouds $C, C' \subset \mathbb{R}^n$ with up to m points and pre-computed ACMs, the distance $\text{EMD}_\infty(\text{ACM}(C; t), \text{ACM}(C'; t))$ in Definition 6.8(c) can be computed in extra time $O((n^6 + m^3) \log m)$ and satisfies

$$\text{EMD}_\infty(\text{ACD}(C), \text{ACD}(C')) \geq \text{EMD}_\infty(\text{ACM}(C; 1), \text{ACM}(C'; 1)),$$

so $\text{ACM}(C; 1)$ has Lipschitz constant 2 under any perturbations of C .

(d) Any clouds $C, C' \subset \mathbb{R}^n$ of m unordered points satisfy the inequalities

$$\begin{aligned} \text{BD}_\infty(\text{SCD}(C), \text{SCD}(C')) &\geq \text{EMD}_\infty(\text{SCD}(C), \text{SCD}(C')) \geq \\ \text{EMD}_\infty(\text{ACD}(C), \text{ACD}(C')) &\geq \text{EMD}_\infty(\text{ACM}(C; 1), \text{ACM}(C'; 1)). \end{aligned}$$

Proof. All arguments are similar to the proof of Theorem 6.7, where n points in a base sequence A are replaced with $n-1$ points, because the extra n -th point is the center of mass $O(C)$. Lemma 6.6(a) is similarly updated with five constituent vectors in Definition 6.8(a), instead of three. ■

In conclusion, this paper solved Problem 1.1 for all n -dimensional clouds $C \subset \mathbb{R}^n$ of m unordered points by the complete invariant SCD in Theorem 4.4 with Lipschitz continuous metrics due to Theorems 5.9 and 5.10, all computable in polynomial times of m , for a fixed dimension n , with faster invariants and lower bounds for distances in Theorem 6.10.

Problem 1.1 was extended with stronger conditions in [5, Problem 1.4.5] within the emerging area of Geometric Data Science, which develops complete and continuous invariants of real data beyond clouds [32], such as lattices [10, 11, 30, 34], 1-periodic sequences [2, 3, 37], periodic crystals in \mathbb{R}^3 [4, 35, 36] and general periodic sets of unordered points in \mathbb{R}^n [1, 6, 49].

Acknowledgment: The author thanks Daniel Widdowson for implementing the original invariants, see the short conference version [51] and the extended version [31] with early proofs of some results in this paper.

References

- [1] O. Anosova, V. Kurlin, An isometry classification of periodic point sets, in: J. Lindblad, F. Malmberg, N. Sladoje (Eds.), *Discrete Geometry and Mathematical Morphology – First International Joint Conference, DGMM 2021, Uppsala, Sweden, May 24–27, 2021, Proceedings*, Springer, Cham, 2021, pp. 229–241.
- [2] O. Anosova, V. Kurlin, Density functions of periodic sequences, in: É. Baudrier, B. Naegel, A. Krähenbühl, M. Tajine (Eds.), *Discrete Geometry and Mathematical Morphology Second International Joint Conference, DGMM 2022, Strasbourg, France, October 24–27, 2022, Proceedings*, Springer, Cham, 2022, pp. 395–408.
- [3] O. Anosova, V. Kurlin, Density functions of periodic sequences of continuous events, *J. Math. Imag. Vis.* **65** (2023) 689–701.
- [4] O. Anosova, V. Kurlin, M. Senechal, The importance of definitions in crystallography, *IUCrJ* **11** (2024) 453–463.
- [5] O. Anosova, V. Kurlin, *Geometric Data Science Book* (latest version at <https://kurlin.org/geometric-data-science-book.pdf>), *arXiv* (2025) doi: <https://doi.org/10.48550/arXiv.2512.05040>.

-
- [6] O. Anosova, D. Widdowson, V. Kurlin, Recognition of near-duplicate periodic patterns by continuous metrics with approximation guarantees, *Pattern Recogn.* **171** (2026) #112108.
- [7] O. Anosova, V. Kurlin, The strength of a geometric simplex is a key ingredient in a polynomial-time classification of unordered point clouds by Lipschitz continuous invariants, *arXiv* (2026) doi: <https://doi.org/10.48550/arXiv.2602.17630>.
- [8] U. Bauer, H. Edelsbrunner, The Morse theory of Cech and Delaunay filtrations, in: S. W. Cheng, O. Devillers (Eds.), *SOCG'14: Annual Symposium on Computational Geometry*, ACM, New York, 2014, pp. 484–490.
- [9] M. Boutin, G. Kemper, On reconstructing n-point configurations from the distribution of distances or areas, *Adv. Appl. Math.* **32** (2004) 709–735.
- [10] M. Bright, A. Cooper, V. Kurlin, Geographic-style maps for 2-dimensional lattices, *Acta Cryst. A* **79** (2023) 1–13.
- [11] M. Bright, A. Cooper, V. Kurlin, Continuous chiral distances for 2-dimensional lattices, *Chirality* **35** (2023) 920–936.
- [12] D. Burago, Y. Burago, S. Ivanov, *A Course in Metric Geometry*, AMS, 2001.
- [13] P. Chew, M. Goodrich, D. Huttenlocher, K. Kedem, J. Kleinberg, D. Kravets, Geometric pattern matching under Euclidean motion, *Comput. Geom.* **7** (1997) 113–124.
- [14] P. Chew, D. Dor, A. Efrat, K. Kedem, Geometric pattern matching in d-dimensional space, *Discr. Comput. Geom.* **21** (1999) 257–274.
- [15] S. Cohen, L. Guibas, *The Earth Mover's Distance: lower bounds and invariance under translation*, Tech. rep., Stanford University, 1997.
- [16] D. Cohen-Steiner, H. Edelsbrunner, J. Harer, Stability of persistence diagrams, *Discr. Comput. Geom.* **37** (2007) 103–120.
- [17] B. V. Dekster, J. B. Wilker, Edge lengths guaranteed to form a simplex, *Archiv der Mathematik* **49** (1987) 351–366.
- [18] E. Deza, M. M. Deza, *Encyclopedia of Distances*, Springer, 2009.
- [19] L. Dorst, D. Fontijne, S. Mann, *Geometric Algebra for Computer Science: An Object-Oriented Approach to Geometry*, Elsevier, 2010.

-
- [20] H. Edelsbrunner, J. L. Harer, *Computational Topology: An Introduction*, AMS, 2022.
- [21] A. Efrat, A. Itai, M. J. Katz, Geometry helps in bottleneck matching and related problems, *Algorithmica* **31** (2001) 1–28.
- [22] M. L. Fredman, R. E. Tarjan, Fibonacci heaps and their uses in improved network optimization algorithms, *J. ACM* **34** (1987) 596–615.
- [23] A. Goldberg, R. Tarjan, Solving minimum-cost flow problems by successive approximation, in: *Proc. STOC*, 1987, pp. 7–18.
- [24] J. E. Hopcroft, R. M. Karp, An $n^5/2$ algorithm for maximum matchings in bipartite graphs, *SIAM J. Comput.* **2** (1973) 225–231.
- [25] R. Horn, C. Johnson, *Matrix Analysis*, Cambridge Univ. Press, Cambridge, 2012.
- [26] D. P. Huttenlocher, G. A. Klanderman, W. J. Rucklidge, Comparing images using the Hausdorff distance, *Trans. Patt. Anal. Mach. Intell.* **15** (1993) 850–863.
- [27] L. Kantorovich, Mathematical methods of organizing and planning production, *Manag. Sci.* **6** (1960) 366–422.
- [28] E. S. Keeping, *Introduction to Statistical Inference*, Courier Corporation, 1995.
- [29] D. G. Kendall, D. Barden, T. K. Carne, H. Le, *Shape and Shape Theory*, Wiley, New York, 2009.
- [30] V. Kurlin, A complete isometry classification of 3-dimensional lattices, *arXiv* (2022) **doi: <https://doi.org/10.48550/arXiv.2201.10543>**.
- [31] V. Kurlin, Complete invariants of atomic clouds under rigid motion with Lipschitz continuous metrics in a polynomial time, *arXiv:2303.13486* (2023).
- [32] V. Kurlin, Simplexwise distance distributions for finite spaces with metrics and measures, *arXiv* (2023) **doi: <https://doi.org/10.48550/arXiv.2303.14161>**.
- [33] V. Kurlin, Polynomial-time algorithms for continuous metrics on atomic clouds of unordered points, *MATCH Commun. Math. Comput. Chem.* **91** (2024) 79–108.
- [34] V. Kurlin, Mathematics of 2-dimensional lattices, *Found. Comput. Math.* **24** (2024) 805–863.

-
- [35] D. Widdowson, V. Kurlin, Continuous invariant-based maps of the Cambridge structural database, *Crystal Growth Design* **24** (2024) 5627–5636.
- [36] D. Widdowson, V. Kurlin, Geographic-style maps with a local novelty distance help navigate in the materials space, *Sci. Rep.* **15** (2025) 27588.
- [37] V. Kurlin, Complete and continuous invariants of 1-periodic sequences in polynomial time, *SIAM J. Math. Data Sci.* **7** (2025) 1643–1663.
- [38] S. Majhi, J. Vitter, C. Wenk, Approximating Gromov-Hausdorff distance in Euclidean space, *Comput. Geom.* **116** (2024) #102034.
- [39] F. Mémoli, Gromov–Wasserstein distances and the metric approach to object matching, *Found. Comput. Math.* **11** (2011) 417–487.
- [40] F. Mémoli, Z. Smith, Z. Wan, The Gromov-Hausdorff distance between ultrametric spaces: its structure and computation, *arXiv* (2021) doi: <https://doi.org/10.48550/arXiv.2603.00796>.
- [41] F. Mémoli, T. Needham, Distance distributions and inverse problems for metric measure spaces, *Stud. Appl. Math.* **149** (2022) 943–1001.
- [42] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, Cambridge Univ. Press, Cambridge, 2007.
- [43] S. Rass, S. König, S. Ahmad, M. Goman, Metricizing the euclidean space towards desired distance relations in point clouds, *IEEE Trans. Inf. Forensics Secur.* **19** (2024) 7304–7319.
- [44] Y. Rubner, C. Tomasi, L. Guibas, The Earth Mover’s Distance as a metric for image retrieval, *Int. J. Comput. Vis.* **40** (2000) 99–121.
- [45] F. Schmedl, Computational aspects of the Gromov–Hausdorff distance and its application in non-rigid shape matching, *Discr. Comp. Geom.* **57** (2017) 854–880.
- [46] P. Smith, V. Kurlin, Generic families of finite metric spaces with identical or trivial 1-dimensional persistence, *J. Appl. Comput. Topology* **8** (2024) 839–855.
- [47] M. Toews, W. Wells, Efficient and robust model-to-image alignment using 3d scale-invariant features, *Med. Image Anal.* **17** (2013) 271–282.

-
- [48] L. N. Vaserstein, Markov processes over denumerable products of spaces, describing large systems of automata, *Problemy Peredachi Informatsii* **5** (1969) 64–72.
- [49] D. Widdowson, M. M. Mosca, A. Pulido, A. I. Cooper, V. Kurlin, Average minimum distances of periodic point sets - foundational invariants for mapping all periodic crystals, *MATCH Commun. Math. Comput. Chem.* **87** (2022) 529–559.
- [50] D. Widdowson, V. Kurlin, Resolving the data ambiguity for periodic crystals, in: S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, A. Oh (Eds.), *Advances in Neural Information Processing Systems 35*, NeurIPS, 2022, pp. 1–14.
- [51] D. Widdowson, 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, 2023*, pp. 1275–1284.
- [52] D. Widdowson, V. Kurlin, Pointwise distance distributions for detecting near-duplicates in large materials databases, *SIAM J. Appl. Math.* (2026) accepted, **doi:** <https://doi.org/10.1137/25M1736657>.