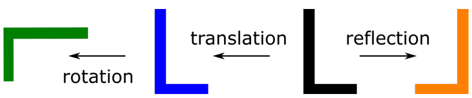
Recognizing rigid patterns of Euclidean clouds of unordered points by complete and continuous isometry invariants with **no false negatives** and **no false positives** for all possible data

Daniel Widdowson, *Vitaliy Kurlin*'s Data Science group
Materials Innovation Factory, University of Liverpool, UK

A **point cloud** consists of m *unordered* points. An **isometry** is any map preserving inter-point distances. In Euclidean $\mathbb{R}^n$, all isometries are compositions of translations, rotations, reflections, and form the group $\mathrm{E}(n)$.



If reflections are excluded, the resulting *orientation-preserving isometries* are rigid motions that form the special Euclidean group $\mathrm{SE}(n)$.

If m points are *unordered*, such clouds can be represented by $m!$ distance matrices obtained by $m!$ permutations of given points, better than any infinite-size representation but impractical.

Geometric Deep Learning experimentally outputs *invariants* preserved by the actions of $\mathrm{E}(3)$ or $\mathrm{SE}(3)$, optimized for specific data without using *stronger* invariants [1,2].

$\frac{1}{2}m(m-1)$ sorted pairwise distances between all points are *generically complete*: distinguish all m-point clouds in *general* position in $\mathbb{R}^n$ [1].

Problem. Design a practical invariant I : {all unordered point clouds in $\mathbb{R}^n$} $\to$ {a simpler space} satisfying

(a) **completeness**: any A, B are related by rigid motion in $\mathbb{R}^n$ *if and only if* $I(A) = I(B)$;

(b) **Lipschitz continuity**: there is a constant λ such that if any point of A is perturbed within its ε-neighborhood, then $I(A)$ changes by at most $\lambda\varepsilon$ in a metric d satisfying all the metric axioms below:

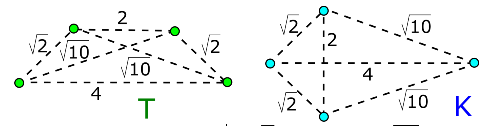
1) $d(I(A), I(B)) = 0$ if and only if clouds A, B are related by rigid motion in $\mathbb{R}^n$,

2) $d(I_1, I_2) = d(I_2, I_1)$,

3) $\triangle$ *triangle inequality*: $d(I_1, I_2) + d(I_2, I_3) \geq d(I_1, I_3)$ for any invariant values;

(c) **computability** : I, d and a reconstruction of a cloud $A \subset \mathbb{R}^n$ from $I(A)$ are obtained in polynomial time in the size $|A|$ for a fixed dimension n.

For any point $p \in C$, write distances $d_1 \leq \cdots \leq d_{m-1}$ to all points in $C - \{p\}$. The *Pointwise Distance Distribution* [2] is the unordered set of all such distance rows in the $m \times (m-1)$-matrix $\mathrm{PDD}(C)$.



$$\mathrm{PDD}(T) = \left(\begin{array}{c|ccc} 1/2 & \sqrt{2} & 2 & \sqrt{10} \\ 1/2 & \sqrt{2} & \sqrt{10} & 4 \end{array}\right)$$

$$\mathrm{PDD}(K) = \left(\begin{array}{c|ccc} 1/4 & \sqrt{2} & \sqrt{2} & 4 \\ 1/2 & \sqrt{2} & 2 & \sqrt{10} \\ 1/4 & \sqrt{10} & \sqrt{10} & 4 \end{array}\right)$$

New invariants [3]: the *Simplexwise Centered Distribution* (SCD) solves the problem for all n-dimensional clouds $C \subset \mathbb{R}^n$. Firstly, shift the center of C to the origin $p_0 = 0$ in $\mathbb{R}^n$. $\mathrm{SCD}(C)$ is the unordered set of pairs $[D(A'), M(C; A')]$ for all subsets $A \subset C$ of permutable points $p_1, \dots, p_{n-1}$, $D(A')$ is the distance matrix of $A' = A \cup \{0\}$, where $M(C; A')$ is the $(n+1) \times (m-n+1)$-matrix with permutable columns for points $q \in C - A$, each consisting of n distances $|q - p_i|$, sign of determinant on the vectors $q - p_i$ for $i = 0, \dots, n-1$.

[1] M.Boutin, G.Kemper. Advances in Applied Math. 32 (2004), 709-735.

[2] D.Widdowson, V.Kurlin. Resolving the data ambiguity for periodic crystals. Proceedings of NeurIPS 2022.

[3] D.Widdowson, V.Kurlin. Proceedings of CVPR 2023.

Details of continuous *complete isometry invariants* in the CVPR 2023 paper

A new area of **Geometric Data Science** develops continuous parametrizations and computable metrics on geographic-style maps of data objects modulo practical equivalences. This work studies finite clouds of unordered points under Euclidean isometry. The past work in NeurIPS 2022 established the *Crystal Isometry Principle* (CRISP): all real periodic crystals live in one *Crystal Isometry Space* continuously extending Mendeleev's table of elements.

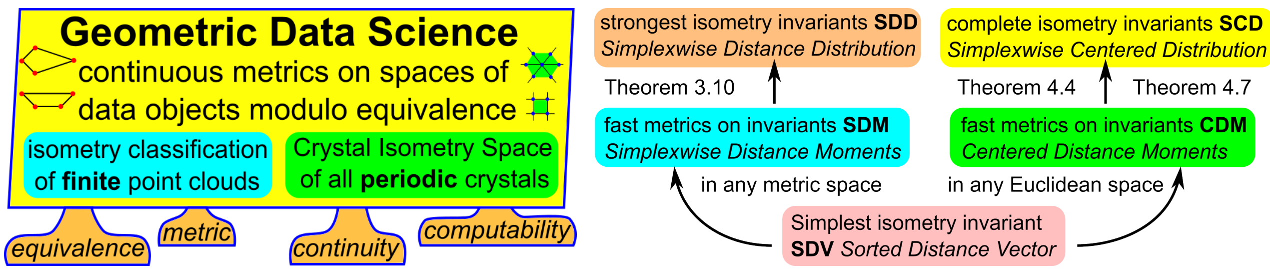
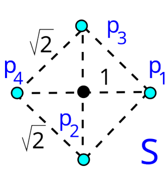


Figure 1: **Left**: the key concepts of Geometric Data Science (GDS) are equivalence, metric, continuity, and computability. **Right**: a hierarchy of isometry invariants from ordered pairwise distances to the SDD (strongest known in metric spaces) and SCD (proved complete in $\mathbb{R}^n$).

Metric space with a distance d. Let C be a cloud of m unordered points. Let $A = (p_1, \dots, p_h) \subset C$ be an ordered subset of $1 \leq h < m$ points. Let $D(A)$ be the *triangular distance* matrix whose entry $D(A)_{i,j-1}$ is $d(p_i, p_j)$ for $1 \leq i < j \leq h$, all other entries are filled by zeros. Any permutation $\xi \in S_h$ acts on $D(A)$ by mapping $D(A)_{ij}$ to $D(A)_{kl}$, where $k \leq l$ is the pair of indices $\xi(i), \xi(j) - 1$ written in increasing order. The $h \times (m-h)$-matrix $R(C; A)$ is formed by $m-h$ permutable columns of distances from $q \in C - A$ to $p_1, \dots, p_h$. Any $\xi \in S_h$ acts on rows of $R(C; A)$. The *Relative Distance Distribution* $\text{RDD}(C; A)$ is the equivalence class of $[D(A), R(C; A)]$ up to permutations $\xi \in S_h$. The *Simplexwise Distance Distribution* $\text{SDD}(C; h)$ is the unordered set of $\text{RDD}(C; A)$ for all unordered h-point subsets $A \subset C$.

Euclidean cloud $C \subset \mathbb{R}^n$. Fix the center of mass $p_0 = \frac{1}{m}\sum_{p \in A}$ at the origin $0 \in \mathbb{R}^n$. In $R(C; \{0\} \cup A)$ for $q \in C - A$, to each column of n Euclidean distances $|q - p_0|, \dots, |q - p_{n-1}|$, add the sign of the determinant of the $n \times n$ matrix consisting of the vectors $q - p_0, \dots, q - p_{n-1}$. Any $\xi \in S_{n-1}$ permutes the first $n-1$ rows of the resulting $(n+1) \times (m-n+1)$-matrix $M(C; \{0\} \cup A)$ and multiplies every sign in the $(n+1)$-st row by $\text{sign}(\xi)$. The *Oriented Centered Distribution* $\text{OCD}(C; A)$ is the equivalence class of $[D(A \cup \{0\}), M(C; A \cup \{0\})]$ up to permutations $\xi \in S_{n-1}$ of 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 $A \subset C$.



$S \subset \mathbb{R}^2$ consists of 4 vertices $(\pm 1, 0)$, $(0, \pm 1)$ of a square. For each 1-point subset $A = \{p\} \subset S$, the distance matrix $D(A \cup \{0\})$ on two points is one number 1. The matrix $M(S; A \cup \{0\})$ has $m - n + 1 = 3$ columns and $n + 1 = 3$ rows.

$$M(S; \begin{pmatrix} p_1 \\ 0 \end{pmatrix}) = \begin{pmatrix} \sqrt{2} & \sqrt{2} & 2 \\ 1 & 1 & 1 \\ - & + & 0 \end{pmatrix}. \text{ Then SCD}(S) \text{ is one OCD} = [1, \begin{pmatrix} \sqrt{2} & \sqrt{2} & 2 \\ 1 & 1 & 1 \\ - & + & 0 \end{pmatrix}].$$

Theorem 4.7: $\text{SCD}(C)$ is a complete invariant for all n-dimensional clouds $C \subset \mathbb{R}^n$ of m unordered points computable in time $O(m^n/(n-4)!)$, so any clouds C, C' are related by $\text{SO}(n)$ rotation around their common center of mass if and only if the Earth Mover's Distance $\text{EMD}(\text{SCD}(C), \text{SCD}(C')) = 0$. Any mirror reflection changes only the signs of $\text{SCD}(C)$. This EMD is Lipschitz continuous, needs time $O((n-1)!(n^2 + m^{1.5} \log^n m) l^2 + l^3 \log l)$, $l = $size(SCDs).