Complete and Lipschitz continuous invariants of finite and 1-periodic sequences in polynomial time

Vitaliy Kurlin
Department of Computer Science, University of Liverpool, Liverpool, United Kingdom

Abstract

The inevitable noise in real measurements motivates the challenging problem to continuously quantify the similarity between rigid objects such as periodic time series and 1-dimensional materials considered up to isometry maintaining inter-point distances. The past work developed many Hausdorff-like distances, which have slow or approximate algorithms due to minimizations over infinitely many isometries. For all finite and 1-periodic sequences under isometry and rigid motion in any high-dimensional Euclidean space, we introduce complete invariants and Lipschitz continuous metrics whose time complexities are polynomial in both input size and ambient dimension.

The key novelty in the periodic case is the Lipschitz continuity under perturbations that change a minimum period. This continuity is practically important for maintaining scientific integrity by real-time detection of near-duplicate structures in experimental and simulated materials datasets.

2012 ACM Subject Classification Theory of computation → Computational geometry

Keywords and phrases 1-periodic sequence, isometry, invariant, classification, metric, continuity

1 Motivations, problem statement, and overview of new results

This paper studies high-dimensional data that is periodic in one direction, motivated by applications to periodic time series [19] and 1-dimensional materials [34], e.g. nanotubes [23]. Though these data sequences are periodic in one direction, say along the first coordinate axis, the underlying points can live in a high-dimensional space $\mathbb{R}\times\mathbb{R}^{n-1}$ for any $n\geq 1$.

▶ **Definition 1.1** (*1-periodic sequences* in $\mathbb{R}\times\mathbb{R}^{n-1}$)**.** Let $\vec e_1$ be the unit vector along the first axis in $\mathbb{R}\times\mathbb{R}^{n-1}$ for $n\geq 1$. For a *period* $l>0$, a *motif* M is a set of points $p_1,\dots,p_m$ in the *slice* $[0,l)\times\mathbb{R}^{n-1}$ of the width $l>0$. We assume that the *time projections* $t(p_1),\dots,t(p_m)$ under $t:[0,l)\times\mathbb{R}^{n-1}\to[0,l)$ are distinct, while $v(p_1),\dots,v(p_m)$ under the *value projection* $v:[0,l)\times\mathbb{R}^{n-1}\to\mathbb{R}^{n-1}$ are arbitrary. A *1-periodic sequence* $S=M+l\vec e_1\mathbb{Z}$ is the infinite sequence of points $p(i+mj)=p_i+jl\in\mathbb{R}^n$ indexed by $i+mj$, where $j\in\mathbb{Z}$, $i=1,\dots,m$. ■

The slice $[0,l)\times\mathbb{R}^{n-1}$ excludes all points with $t=l$, which are equivalent to points with $t=0$ by translation in the time factor $\mathbb{R}$. So all motif points $p_1,\dots,p_m\in[0,l)\times\mathbb{R}^{n-1}$ are counted once and naturally ordered under the time projection $t:[0,l)\times\mathbb{R}^{n-1}\to[0,l)$.

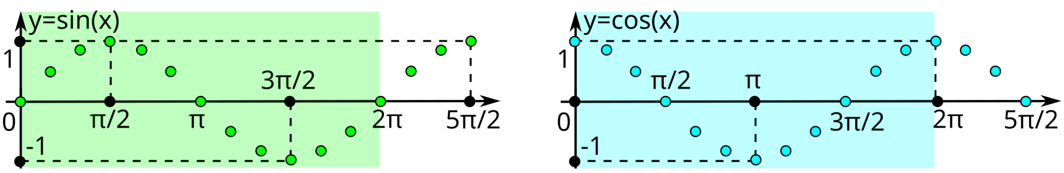

Figure 1 The periodic sequences $C,S\subset\mathbb{R}\times\mathbb{R}$ are sampled from the sine and cosine graphs. The motifs in the shaded slice $[0,2\pi)\times\mathbb{R}$ are non-isometric, but S and C are related by translation.

▶ **Example 1.2** (1-periodic sequences in $\mathbb{R}\times\mathbb{R}$)**.** Fig. 1 (left) shows the 1-periodic sequence S in $\mathbb{R}\times\mathbb{R}$ (sampled from the sine graph) with the period $l=2\pi$ and motif M_S of the points $(0,0)$, $(\frac{\pi}{6},\frac12)$, $(\frac{\pi}{3},\frac{\sqrt3}{2})$, $(\frac{\pi}{2},1)$, $(\frac{2\pi}{3},\frac{\sqrt3}{2})$,$(\frac{5\pi}{6},\frac12)$, $(\pi,0)$,$(\frac{7\pi}{6},-\frac12)$, $(\frac{4\pi}{3},-\frac{\sqrt3}{2})$, $(\frac{3\pi}{2},-1)$,$(\frac{5\pi}{3},-\frac{\sqrt3}{2})$, $(\frac{11\pi}{6},-\frac12)$. Similarly, measurements of many oscillating systems [28] generate sequences that are periodic in a single time direction and non-periodic in many other directions. Fig. 1 (right) shows another sequence C with the same period $l = 2\pi$ and a different motif $M_C \neq M_S$. However, S and C become identical under translation in the x-axis: $\sin(x - \frac{\pi}{2}) = \cos(x)$. ∎

This basic example illustrates a widespread ambiguity of digital representations when many real objects look different in various coordinate systems despite being equivalent, for example, as rigid objects. Recall that a *rigid motion* in $\mathbb{R}^n$ is any composition of translations and rotations. If we also allow compositions with mirror reflections, we get any distance-preserving transformation in $\mathbb{R}^n$, which is called an *isometry.* A linear map $f : \mathbb{R}^n \to \mathbb{R}^n$ preserves orientation if, for any linear basis $v_1, \dots, v_n$ of $\mathbb{R}^n$, the two $n \times n$ determinants with the columns $v_1, \dots, v_n$ and $f(v_1), \dots, f(v_n)$ have the same sign. Any rigid motion is an orientation-preserving isometry. We adapt these equivalences to the product $\mathbb{R} \times \mathbb{R}^{n-1}$.

▶ **Definition 1.3** (*cyclic* vs *dihedral* isometries and rigid motions in $\mathbb{R} \times \mathbb{R}^{n-1}$)**.** A *cyclic* isometry of $\mathbb{R} \times \mathbb{R}^{n-1}$ is a composition of a translation in the time factor $\mathbb{R}$ and an isometry in the value factor $\mathbb{R}^{n-1}$. If we allow compositions of a translation and mirror symmetry $x \mapsto -x$ in the time factor $\mathbb{R}$, the resulting isometry of $\mathbb{R} \times \mathbb{R}^{n-1}$ is called *dihedral.* If we allow only isometries that preserve orientation in the value factor $\mathbb{R}^{n-1}$, the resulting equivalences are called cyclic and dihedral *rigid motions* in the former and latter cases, respectively. ∎

The adjectives *cyclic* and *dihedral* are motivated by the traditional names of the cyclic group C_m and the dihedral group D_m consisting of orientation-preserving isometries and all isometries in $\mathbb{R}^2$, respectively, that map the regular polygon on m vertices to itself.

The equivalences in Definition 1.3 make sense for any finite sequence $T \subset \mathbb{R} \times \mathbb{R}^{n-1}$ but the periodicity worsens the ambiguity of representations via a period l and a motif M as follows. A translation in the time factor $\mathbb{R}$ allows us to fix any point p of a motif M at $t = 0$, but this choice of p is arbitrary, so a motif M is defined only up to cyclic permutations.

The set of integers can be defined as $\mathbb{Z}$ with period 1 or as $\{0, 1\} + 2\mathbb{Z}$ with period 2, and also with any integer period $l > 0$. For any given sequence $S = \{p_1, \dots, p_m\} + l\vec{e}_1\mathbb{Z}$, we can choose a *minimum* period l such that S can not be represented with a smaller period.

This classical approach in crystallography leads to an *invariant* I based on a minimum period (primitive cell) and defined as a set of numerical properties preserved under any rigid motion. Choosing standard settings [37] for a reduced cell [35] of 3-periodic crystals theoretically defines a *complete* invariant that unambiguously identifies any rigid crystal.

However, fixing a minimum period creates the following discontinuity. For any small $\varepsilon > 0$ and integer m, any point of $\mathbb{Z}$ is ε-close to a unique point of the sequence $\{0, 1 + \varepsilon, \dots, m + \varepsilon\} + (m + 1)\mathbb{Z}$, though their minimum periods 1 and $m + 1$ are arbitrarily different. Hence comparing periodic sequences by their given (minimum) motifs can miss near-duplicates.

Perturbations of points up to ε in the Euclidean distance are motivated by noise in real measurements. Though many materials look rigid, atoms always vibrate above the absolute zero temperature [17, chapter 1]. When the same material is characterized at different temperatures, its structure can have arbitrarily different periods (primitive cells) [42].

As a result, many experimental databases do not recognize such near-duplicates [50, 47]. More importantly, any known material can be disguised as 'new' [10] by a slight perturbation that substantially changes a primitive cell with many more options for periodicity in 3 directions. Simulated materials are even more vulnerable under perturbations because any iterative optimization always stops at some approximation to a local optimum. These slightly different approximations can accumulate around the same optimum as in Google's GNoME database [33] whose thousands of unexpected duplicates were recently exposed [2, 11].

The discontinuity of material representations threatens the public trust in science and motivates the following problem, which is stated for cyclic isometries below for simplicity but will be solved for 1-periodic sequences under all equivalences in Definition 1.3.

We assume that the input for a 1-periodic sequence S consists of a period l and a motif of $m = |S|$ points in the slice $[0, l) \times \mathbb{R}^{n-1}$. All complexities are for the real RAM model.

▶ **Problem 1.4** (complete and continuous invariants of 1-periodic sequences in $\mathbb{R} \times \mathbb{R}^{n-1}$). Find an invariant I of all 1-periodic sequences in $\mathbb{R} \times \mathbb{R}^{n-1}$ satisfying the following conditions.

(a) *Completeness* : any 1-periodic sequences $S, Q \subset \mathbb{R} \times \mathbb{R}^{n-1}$ are related by cyclic isometry (denoted as $S \cong Q$) in Definition 1.3 if and only if they have equal invariants $I(S) = I(Q)$.

(b) *Reconstruction* : any $S \subset \mathbb{R} \times \mathbb{R}^{n-1}$ is reconstructable from $I(S)$ up to cyclic isometry.

(c) *Lipschitz continuity* : there is a constant $\lambda > 0$ and a metric d on invariant values such that the metric axioms hold: (1) $d(I(S), I(Q)) = 0$ if and only if $I(S) = I(Q)$, (2) $d(I(S), I(Q)) = d(I(Q), I(S))$, (3) $d(I(S), I(Q)) + d(I(Q), I(T)) \geq d(I(S), I(T))$; and if every point of Q is obtained by perturbing a point of S up to ε, then $d(I(S), I(Q)) \leq \lambda\varepsilon$.

(d) *Computability* : the invariant I, metric d, and a reconstruction of S from $I(S)$ can be computed in a time that depends polynomially on the motif size m and dimension n. ■

Due to the first metric axiom, the equality $I(S) = I(Q)$ between complete invariants can be checked by comparing $d(I(S), I(Q))$ with 0. Hence condition 1.4(d) for a metric guarantees a polynomial-time algorithm for detecting a cyclic isometry $S \cong Q$. All axioms in 1.4(c) imply the positivity of a metric d because $2d(a, b) = d(a, b) + d(b, a) \geq d(a, a) = 0$.

The Lipschitz continuity in 1.4(c) is stronger than the classical $\varepsilon - \delta$ continuity because a constant λ should be independent of S, ε. Conditions 1.4(b,d) require a polynomial-time inverse function I^{-1}, which is stronger than the completeness (bijectivity) of an invariant I.

The main contribution is the full solution of Problem 1.4 in Theorem 4.7 by the new complete invariants and Lipschitz continuous metrics in Definitions 4.2 and 4.5 for all 1-periodic sequences under cyclic and dihedral isometries and rigid motions in $\mathbb{R} \times \mathbb{R}^{n-1}$.

2 Related work on isometry invariants and metrics on point sets

For a finite sequence of ordered points, the complete invariant under isometry is the classical distance matrix [45], see relevant Lemma A.3 based on more recent [15, Theorem 1] in appendix A, which proves all results. To distinguish mirror images, a sign of orientation can be enough, but this sign vanishes for all degenerate sets of $n+1$ points living in a hyperspace of dimension $n-1$ in $\mathbb{R}^n$. The even harder obstacle is the discontinuity of signs when a sequence of points passes through a degenerate configuration and changes its orientation. Though the volume of a simplex changes continuously there, this continuity is not Lipschitz. In $\mathbb{R}^2$, the signed area of a triangle with the base $[-x, x] \times \{0\}$ and top vertex at $(0, \varepsilon)$ is εx and hence changes by $2\varepsilon x$ when the vertex degenerates to $(0, 0)$ and then to the symmetric position $(0, -\varepsilon)$. For any fixed $\varepsilon > 0$, the change $2\varepsilon x$ can be arbitrarily large without restrictions on x and hence not Lipschitz continuous as in the sense of condition 1.4(c).

The case of m *unordered* points $T \subset \mathbb{R}^n$ is much harder because considering $m!$ distance matrices is impractical already for $m = 4$. The case of $m = 3$ is the SSS theorem saying that the triangles are isometric if and only if they have the same triple of side lengths considered up to $3! = 6$ permutations. Though all pairwise distances uniquely determine any *generic*

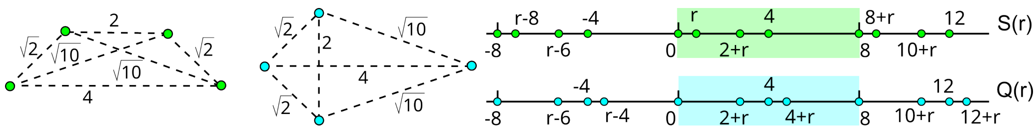

Figure 2 Left: sets $K = \{(\pm 2, 0),\ (\pm 1, 1)\}$ and $T = \{(\pm 2, 0), (-1, \pm 1)\}$ can not be distinguished by pairwise distances $\sqrt{2}, \sqrt{2}, 2, \sqrt{10}, \sqrt{10}, 4$. **Right**: sequences $S(r) = \{0, r, 2 + r, 4\} + 8\mathbb{Z}$ and $Q(r) = \{0, 2 + r, 4, 4 + r\} + 8\mathbb{Z}$ for $0 < r \leq 1$ have the same Patterson function [38, p. 197, Fig. 2].

set of m points under isometry in $\mathbb{R}^n$ [4], Fig. 2 (left) shows non-isometric sets of $m = 4$ unordered points (from an infinite family) that are indistinguishable by 6 pairwise distances.

If we need only a binary answer, [1, Theorem 1] already in 1988 checked the existence of an isometry between two m-point sets in $\mathbb{R}^n$ in time $O(m^{n-2} \log m)$. The latest algorithm [5] checks this isometry in time $O(m^{\lceil n/3 \rceil} \log m)$, which becomes $O(m \log m)$ in $\mathbb{R}^3$ [6]. If we need only a metric, distances between fixed clouds extend to their rigid classes by minimization over infinitely many rigid motions [26, 14, 13]. In $\mathbb{R}^2$, the time $O(m^5 \log m)$ [12] for the Hausdorff distance [24], see approximations in [20]. The Gromov-Wasserstein metrics [31] are defined for metric-measure spaces also by minimizing over infinitely many correspondences between points, but cannot be approximated with a factor less than 3 in polynomial time unless P=NP, see [44, Corollary 3.8] and polynomial algorithms for partial cases in [32, 29].

Mémoli's work on *local distributions of distances* [31], also known as *shape distributions* [36, 3, 21, 30, 39], for metric spaces is closest to the new invariants of 1-periodic sequences. These distributions were adapted to any number of periodic directions as Pointwise Distance Distributions (PDD) and distinguished (together with underlying lattices) any periodic sets in general position [47, Theorem 4.4] but not infinitely many examples in [40, Fig. 4].

In crystallography, the simpler invariants such as diffraction patterns consisting of all inter-point distances considered with frequencies had earlier counter-examples even in dimension 1, see Fig. 2 (right). Patterson [38] visualized any periodic sequence $S = \{p_1, \dots, p_m\} + l\mathbb{Z} \subset \mathbb{R}$ in a circle of a length l but described its isometry classes by the complicated *distance array* defined as the anti-symmetric $m \times m$ matrix of differences $p_i - p_j$ for $i, j \in \{1, \dots, m\}$. Grünbaum and Moore considered rational-valued periodic sequences given by complex numbers on the unit circle and proved [22, Theorem 4] that the combinations of k-factor products of complex numbers up to $k = 6$ suffice to distinguish all such sequences up to translation. This approach fixes a period and hence leads to a discontinuous metric.

Atomic vibrations are natural to measure by the maximum deviation of atoms from their initial positions as in 1.4(c), though the Euclidean metric can be replaced with more general Minkowski metrics without affecting the Lipschitz continuity. The maximum deviation of atoms is usually small, but the full sum over infinitely many perturbed points as in the bottleneck distance $d_B(S, Q)$ is often infinite. If we consider only periodic point sets $S, Q \subset \mathbb{R}^n$ with the same density (or primitive cells of the same volume), $d_B(S, Q)$ becomes a well-defined wobbling distance [9], which is still discontinuous under perturbations by [47, Example 2.2]. The Lipschitz continuity and polynomial-time computability remained hard conditions in Problem 1.4, which were not previously proved for the past complete invariants.

3 Isometry invariants and continuous metrics for finite sequences in $\mathbb{R}^n$

This section studies complete invariants and metrics for isometry classes of finite sequences of ordered points in $\mathbb{R}^n$, which will be later extended to 1-periodic sequences in $\mathbb{R} \times \mathbb{R}^{n-1}$.

▶ **Definition 3.1** (distance matrices DM and CDM)**.** Let $T = \{p_1, \dots, p_m\}$ be an ordered sequence of m points in $\mathbb{R}^n$. In the *distance matrix* $\mathrm{DM}(T)$ of the size $m \times m$, each element $\mathrm{DM}_{ij}(T)$ is the Euclidean distance $|p_j - p_j|$ for $i, j \in \{1, \dots, m\}$, so $d_{ii} = 0$ for $i = 1, \dots, m$.

In the *cyclic distance matrix* $\mathrm{CDM}(T)$ of the size $(m-1) \times m$, each element $\mathrm{CDM}_{ij}(T)$ is the Euclidean distance $|p_j - p_{i+j}|$ for $i \in \{1, \dots, m-1\}$ and $j \in \{1, \dots, m\}$, where all indices are considered modulo m, for example, $p_{m+1} = p_1$. ■

Any $m = 3$ points in $\mathbb{R}^n$ with pairwise distances d_{ij} have the distance matrix $\mathrm{DM} = \begin{pmatrix} 0 & d_{12} & d_{13} \\ d_{12} & 0 & d_{23} \\ d_{13} & d_{23} & 0 \end{pmatrix}$ and the cyclic distance matrix $\mathrm{CDM} = \begin{pmatrix} d_{12} & d_{23} & d_{13} \\ d_{13} & d_{12} & d_{23} \end{pmatrix}$. $\mathrm{CDM}(T)$ is obtained from $\mathrm{DM}(T)$ by removing the zero diagonal and cyclically shifting each column so that the first row of $\mathrm{CDM}(T)$ has distances from p_i to the next point p_{i+1} in T.

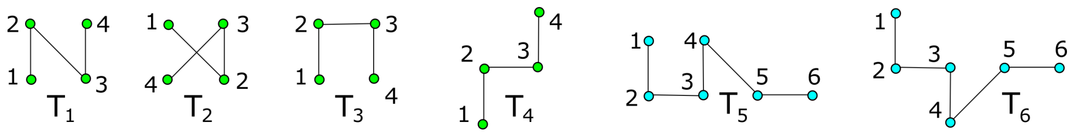

Figure 3 These sequences are distinguished by their cyclic distance matrices in Example 3.2.

▶ **Example 3.2** (cyclic distance matrices)**.** Fig. 3 shows the sequences $T_1, \dots, T_6 \subset \mathbb{R}^2$ whose points are in the integer lattice $\mathbb{Z}^2$ so that the minimum inter-point distance is 1. In each sequence, the points are connected by straight lines in the order $1 \to 2 \to \cdots \to m$. $\mathrm{CDM}(T_1) = \begin{pmatrix} 1 & \sqrt{2} & 1 & \sqrt{2} \\ 1 & 1 & 1 & 1 \\ \sqrt{2} & 1 & \sqrt{2} & 1 \end{pmatrix}$, $\mathrm{CDM}(T_2) = \begin{pmatrix} \sqrt{2} & 1 & \sqrt{2} & 1 \\ 1 & 1 & 1 & 1 \\ 1 & \sqrt{2} & 1 & \sqrt{2} \end{pmatrix}$ are different but related by a cyclic shift of columns. This shift of indices in T_1 gives a sequence isometric to T_2. Then $\mathrm{CDM}(T_3) = \begin{pmatrix} 1 & 1 & 1 & 1 \\ \sqrt{2} & \sqrt{2} & \sqrt{2} & \sqrt{2} \\ 1 & 1 & 1 & 1 \end{pmatrix}$, $\mathrm{CDM}(T_4) = \begin{pmatrix} 1 & 1 & 1 & \sqrt{5} \\ \sqrt{2} & \sqrt{2} & \sqrt{2} & \sqrt{2} \\ \sqrt{5} & 1 & 1 & 1 \end{pmatrix}$.
The CDMs of the sets T_5, T_6 differ only by distances $|p_1 - p_4| = 1$ in T_5 and $|p_1 - p_4| = \sqrt{5}$ in the highlighted cells below. If reduce the number $m-1$ of rows in CDM to the dimension $n = 2$, the smaller matrices fail to distinguish the non-isometric sequences $T_5 \not\cong T_6$.

$$T_5 : \begin{pmatrix} 1 & 1 & 1 & \sqrt{2} & 1 & \sqrt{10} \\ \sqrt{2} & \sqrt{2} & 1 & \sqrt{5} & \sqrt{5} & 3 \\ 1 & 2 & 2 & 1 & 2 & 2 \\ \sqrt{5} & 3 & \sqrt{2} & \sqrt{2} & 1 & \sqrt{5} \\ \sqrt{10} & 1 & 1 & 1 & \sqrt{2} & 1 \end{pmatrix}, \ T_6 : \begin{pmatrix} 1 & 1 & 1 & \sqrt{2} & 1 & \sqrt{10} \\ \sqrt{2} & \sqrt{2} & 1 & \sqrt{5} & \sqrt{5} & 3 \\ \sqrt{5} & 2 & 2 & \sqrt{5} & 2 & 2 \\ \sqrt{5} & 3 & \sqrt{2} & \sqrt{2} & 1 & \sqrt{5} \\ \sqrt{10} & 1 & 1 & 1 & \sqrt{2} & 1 \end{pmatrix}.$$

▶ **Definition 3.3** (*strength* of a simplex and *cyclic distances with signs* CDS)**.** For the simplex A on any set of $n+1$ points $q_0, q_1, \dots, q_n \in \mathbb{R}^n$, the *strength* is $\sigma(A) = \dfrac{V^2(A)}{p^{2n-1}(A)}$, where $V(A)$ is the volume of A, $p(A) = \dfrac{1}{2} \sum\limits_{0 \le i < j \le n} |q_i - q_j|$ is the half-perimeter.

For any sequence T of $p_1, \dots, p_m \in \mathbb{R}^n$ and $i = 1, \dots, m$, let $\sigma_i(T)$ be the strength of the simplex on the points $p_i, \dots, p_{i+n}$, where all indices are modulo m. Let $\mathrm{sign}_i(T)$ be the sign (± 1 or 0) of the $n \times n$ determinant with the columns $p_{i+1} - p_i, p_{i+2} - p_{i+1}, \dots, p_{i+n} - p_{i+n-1}$. The matrix $\mathrm{CDS}(T)$ of *cyclic distances with signs* is obtained from $\mathrm{CDM}(T)$ in Definition 3.1 by attaching the $(m+1)$-st row $\mathrm{sign}(T) = (\mathrm{sign}_1(T), \dots, \mathrm{sign}_m(T))$. ■

For a triangle A with 3 pairwise distances a, b, c in $\mathbb{R}^2$, Heron's formula gives the squared area $p(p-a)(p-b)(p-c)$, where the half-perimeter is $p = \dfrac{a+b+c}{2}$, so the strength is $\sigma(A) = \dfrac{(p-a)(p-b)(p-c)}{p^2}$. Similarly to the volume $V(A)$, the strength $\sigma(A)$ vanishes on degenerate simplices but is Lipschitz continuous [48, Theorem 4.4] with a constant λ_n, e.g. $\lambda_2 \leq 2\sqrt{3}$, while the volume of a simplex is not Lipschitz continuous over the whole $\mathbb{R}^n$.

Because the sign of a determinant discontinuously changes when a point set passes through a degenerate configuration, this sign will be multiplied by the Lipschitz continuous strength to get a metric satisfying condition 1.4(c). Section 4 will adapt the matrices from Definitions 3.1 and 3.3 to 1-periodic sequences whose motifs of points should be considered under cyclic permutations. The *cyclic* group C_m consists of m permutations on $1, \dots, m$ generated by the *shift permutation* $\gamma_m : (1, 2, \dots, m) \mapsto (2, \dots, m, 1)$. The *dihedral* group D_m consists of $2m$ permutations generated by γ_m and the *reverse permutation* $\iota_m : (1, 2, \dots, m) \mapsto (m, \dots, 2, 1)$.

▶ **Lemma 3.4** (actions on vectors and matrices)**.** The shift permutation $\gamma_m \in C_m$ acts on the cyclic distance matrix $\mathrm{CDM}(T)$ by cyclically shifting its m columns and keeping all rows. The reverse permutation $\iota_m \in D_m$ reverses the order of columns and rows in $\mathrm{CDM}(T)$. These permutations act on the row of signs in Definition 3.3 as $\gamma_m(s_1, s_2 \dots, s_m) = (s_2, \dots, s_m, s_1)$ and $\iota_m(s_1, s_2 \dots, s_m) = (-1)^{[3n/2]}(s_m, \dots, s_2, s_1)$. For any mirror image $\bar{T}$ of T, the matrix $\mathrm{CDS}(\bar{T})$ is obtained from $\mathrm{CDS}(T)$ by reversing all signs in the last row. Any element of the groups C_m, D_m acts on any sequence of m numbers as a composition of γ_m, ι_m. ■

Any matrix $k \times m$ can be rewritten row-by-row as a vector $v \in \mathbb{R}^{km}$. For any $q \in [1, +\infty]$, the Minkowski norm is $||v||_q = \left(\sum\limits_{i=1}^{km} |v_i|^q\right)^{1/q}$, where the limit case is $||v||_\infty = \max\limits_{i=1,\dots,km} |v_i|$. In the sequel, any power $a^{1/q}$ for $a > 0$ is interpreted as 1 in the limit case $q = +\infty$.

▶ **Definition 3.5** (metrics $\mathrm{MCD}_q, \mathrm{MCS}_q$ for finite sequences in $\mathbb{R}^n$)**.** For any Minkowski norm with a parameter $q \in [1, +\infty]$ and ordered sequences $T, S \subset \mathbb{R}^{n-1}$ of m points, define the metrics $\mathrm{MCD}_q(S,T) = \dfrac{||\mathrm{CDM}(S) - \mathrm{CDM}(T)||_q}{\big(m(m-1)\big)^{1/q}}$ on cyclic distance matrices from Definition 3.1 and $\mathrm{MCS}_q(S,T) = \max\left\{\mathrm{MCD}_q(S,T), \dfrac{2}{\lambda_n} \max\limits_{i=1,\dots,m} \left|\mathrm{sign}_i(S)\sigma_i(S) - \mathrm{sign}_i(T)\sigma_i(T)\right|\right\}$. ■

We use the extra factors $\big(m(m-1)\big)^{1/q}$ and $\dfrac{2}{\lambda_n}$ in the definition above, where λ_n is a Lipschitz constant of the strength σ from [48, Theorem 4.4], only to guarantee the standard Lipschitz constant 2 for the new metrics. Indeed, perturbing any points up to ε changes the distance between them up to 2ε. Instead of each maximum in the formula for $\mathrm{MCS}_q(S,T)$, one can consider other metric transforms from [16, section 4.1], for example, sums of metrics.

▶ **Theorem 3.6** (solution to the analog of Problem 1.4 for finite sequences)**.** **(a)** For any sequence $T \subset \mathbb{R}^n$ of m points, $\mathrm{CDM}(T)$ and $\mathrm{CDS}(T)$ are complete invariants of T under isometry and rigid motion in $\mathbb{R}^n$, computable in times $O(m^2 n)$ and $O(m^2 n + mn^3)$, respectively.

(b) Any sequence $T \subset \mathbb{R}^n$ of m points can be reconstructed from the complete invariant matrix $\mathrm{CDM}(T)$ and $\mathrm{CDS}(T)$ up to isometry and rigid motion, respectively, in time $O(m^3)$.

(c) For any sequences $S, T \subset \mathbb{R}^n$ of m points, the distances $\mathrm{MCD}_q(S,T), \mathrm{MCS}_q(S,T)$ satisfy all metric axioms and are computable in time $O(m^2 n)$ and $O(m^2 n + mn^3)$, respectively.

(d) If S is obtained from any finite sequence $T \subset \mathbb{R}^n$ by perturbing every point up to Euclidean distance ε, then $\mathrm{MCD}_q(S,T) \leq 2\varepsilon$ and $\mathrm{MCS}_q(S,T) \leq 2\varepsilon$ for any $q \in [1, +\infty]$. ■

4 Isometry invariants and metrics for 1-periodic sequences in $\mathbb{R} \times \mathbb{R}^{n-1}$

The invariants and metrics from section 3 will be used for a motif of a 1-periodic sequence S projected to the value factor $\mathbb{R}^{n-1}$. To solve Problem 1.4, we first resolve the discontinuity of a period under perturbations of S by considering projections to the time factor $\mathbb{R}$.

▶ **Definition 4.1** (time shift TS)**.** Let $S \subset \mathbb{R} \times \mathbb{R}^{n-1}$ be a 1-periodic sequence with a period l and a motif M of points $p_1, \dots, p_m$, which have ordered time projection $t(p_1) < \dots < t(p_m)$ in $[0, l)$ under $t : \mathbb{R} \times \mathbb{R}^{n-1} \to \mathbb{R}$, see Definition 1.1. Set $d_i = t(p_{i+1}) - t(p_i)$ for $i = 1, \dots, m$, $t(p_{m+1}) = t(p_1) + l$. The *time shift* of the pair (motif,period) is $\mathrm{TS}(M; l) = (d_1, \dots, d_m)$. ■

The sequences $S_2 = \{0, 1\} + 3\mathbb{Z}$ and $3 - S_2 = \{0, 2\} + 3\mathbb{Z}$ are related by translation but have different time shifts $\mathrm{TS}(\{0, 1\}; 3) = (1, 2)$ and $\mathrm{TS}(\{0, 2\}; 3) = (2, 1)$. To get isometry invariants, these shifts are considered modulo cyclic or dihedral permutations below.

▶ **Definition 4.2** (cyclic and dihedral invariants under isometry and rigid motion)**.** For any 1-periodic sequence $S = M + l\vec{e}_1\mathbb{Z} \subset \mathbb{R} \times \mathbb{R}^{n-1}$ with a minimum motif M of m points, let $v(M) \subset \mathbb{R}^{n-1}$ be the image of M under the value projection $v : \mathbb{R} \times \mathbb{R}^{n-1} \to \mathbb{R}^{n-1}$.

The *cyclic* and *dihedral isometry* invariants $\mathrm{CI}(S)$ and $\mathrm{DI}(S)$ are the classes of the pair $(\mathrm{TS}(M; l), \mathrm{CDM}(v(M)))$ considered up to permutations γ from the groups C_m, D_m, respectively, acting simultaneously on the time shift $\mathrm{TS}(M; l)$ and the matrix $\mathrm{CDM}(v(M))$.

The *cyclic* and *dihedral rigid* invariants $\mathrm{CR}(S)$ and $\mathrm{DR}(S)$ are the classes of the pair $(\mathrm{TS}(M; l), \mathrm{CDS}(v(M)))$ considered up to permutations γ from the groups C_m, D_m, respectively, acting simultaneously on the time shift $\mathrm{TS}(M; l)$ and the matrix $\mathrm{CDS}(v(M))$. ■

The matrices $\mathrm{CDM}, \mathrm{CDS}$ are used for the projected motif $v(M) \subset \mathbb{R}^{n-1}$ and do not depend on a period l, because a shift along the time direction $\vec{e}_1$ keeps the value projection.

In the partial case $n = 1$, when a periodic sequence $S = \{p_1, \dots, p_m\} + l\mathbb{Z}$ is in the line $\mathbb{R}$, Definition 4.2 simplifies to a single time shift obtained by lexicographic ordering.

Recall that the *lexicographic order* on vectors is defined so that $(d_1, \dots, d_m) < (d'_1, \dots, d'_m)$ if $d_1 = d'_1, \dots, d_i = d'_i$ for some $0 \le i < m$, where $i = 0$ means no identities, and $d_{i+1} < d'_{i+1}$.

▶ **Definition 4.3** (time invariants CT, DT)**.** Let $S = \{p_1, \dots, p_m\} + l\mathbb{Z} \subset \mathbb{R}$ be a periodic sequence with a minimum period $l > 0$. Set $d_i = p_{i+1} - p_i$ for $i = 1, \dots, m$, where $p_{m+1} = p_1 + l$. The *cyclic* and *dihedral time* invariants $\mathrm{CT}(S), \mathrm{DT}(S)$ are the lexicographically smallest lists obtained from $(d_1, \dots, d_m)$ by the action of C_m, D_m, respectively. ■

The periodic sequences $S = \{0, 1, 3\} + 6\mathbb{Z}$ and $Q = 6 - S = \{0, 3, 5\} + 6\mathbb{Z}$ are related by reflection $x \mapsto 6 - x$ and not by translation. Their time shifts are $\mathrm{TS}(\{0, 1, 3\}; 6) = (1, 2, 3)$ and $\mathrm{TS}(\{0, 3, 5\}; 6) = (3, 2, 1)$. So the dihedral time invariants are equal to $\mathrm{DT} = (1, 2, 3)$, but their cyclic time invariants differ: $\mathrm{CT}(S) = (1, 2, 3) \ne (1, 3, 2) = \mathrm{CT}(Q)$.

Though the time invariants from Definition 4.3 can be proved complete for periodic sequences in $\mathbb{R}$, Example 4.4 and Fig. 4 show their discontinuity under tiny perturbations.

▶ **Example 4.4.** The periodic sequence $S_0 = \{0, 1, 3, 4\} + 7\mathbb{Z}$ has two perturbations $S_{\pm\varepsilon} = \{0, 1 \pm \varepsilon, 3 \pm \varepsilon, 4\} + 7\mathbb{Z}$ for any small $\varepsilon > 0$. Rewriting the time shifts $\mathrm{TS}(\{0, 1-\varepsilon, 3-\varepsilon, 4\}; 7) = (1-\varepsilon, 2, 1+\varepsilon, 3)$ and $\mathrm{TS}(\{0, 1+\varepsilon, 3+\varepsilon, 4\}; 7) = (1+\varepsilon, 2, 1-\varepsilon, 3)$ in increasing order does not make them close, because the minimum distance $1 - \varepsilon$ is followed by the different distances $2 < 3$ in the nearly identical $S_{\pm\varepsilon}$ for any $\varepsilon > 0$, see Fig. 4 (left). This discontinuity will be resolved by minimizing over cyclic permutations but there is one more obstacle below. ■

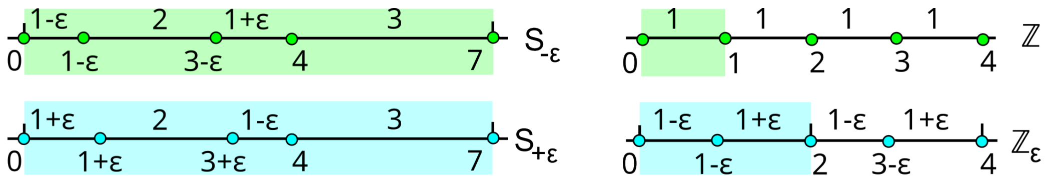

Figure 4 Left: the nearly identical sequences $S_{\pm\varepsilon} = \{0, 1\pm\varepsilon, 3\pm\varepsilon, 4\} + 7\mathbb{Z}$ have distant time invariants from Definition 4.1, see Example 4.4. **Right**: the sequence $\mathbb{Z}$ and its perturbation $\mathbb{Z}_\varepsilon$ have incomparable time shifts $\mathrm{TS}(\{0\}; 1) = (1)$, $\mathrm{TS}(\{0, 1-\varepsilon\}; 2) = (1-\varepsilon, 1+\varepsilon)$ of different lengths.

It seems natural to always use a minimum period $l > 0$ of $S = \{p_1, \dots, p_m\} + l\vec{e}_1\mathbb{Z} \subset \mathbb{R}\times\mathbb{R}^{n-1}$. However, the time shift $\mathrm{TS} = (d_1, \dots, d_m)$ of a fixed size m cannot be directly used for comparing sequences that have different sizes of motifs, see Fig. 4 (right).

Definition 4.5 defines continuous metrics after extending given motifs to a common size.

▶ **Definition 4.5** (cyclic and dihedral metrics under isometry and rigid motion)**.** For any 1-periodic sequences $S = M_S + l_S\vec{e}_1\mathbb{Z}$ and $Q = M_Q + l_Q\vec{e}_1\mathbb{Z}$ in $\mathbb{R}\times\mathbb{R}^{n-1}$, let $m = \mathrm{lcm}(|M_S|, |M_Q|)$ be the lowest common multiple of their motif sizes. For the integers $k_S = \dfrac{m}{|M_S|}$, $k_Q = \dfrac{m}{|M_Q|}$, the extended motifs defined as $k_S M_S = \bigcup\limits_{i=1,\dots,k_S} \big(M_S + i l_S\vec{e}_1\big)$, $k_Q M_Q = \bigcup\limits_{i=1,\dots,k_Q} \big(M_Q + i l_Q\vec{e}_1\big)$ have the same number $k_S|M_S| = m = k_Q|M_Q|$ of points.

Any permutation γ from C_m, D_m acts on the projected motif $v(k_Q M_Q) \subset \mathbb{R}^{n-1}$ as in Lemma 3.4. For any Minkowski norm with $q \in [1, +\infty]$, the *cyclic* and *dihedral isometry* metrics are $\mathrm{CIM}_q(S,Q) = \min\limits_{\gamma\in C_m} \max\{d_t, d_v\}$ and $\mathrm{DIM}_q(S,Q) = \min\limits_{\gamma\in D_m} \max\{d_t, d_v\}$, where

$$d_t = m^{-1/q}\big|\big|\mathrm{TS}(k_S M_S; k_S l_S) - \mathrm{TS}(\gamma(k_Q M_Q); k_Q l_Q)\big|\big|_q,\ d_v = \mathrm{MCD}_q\big(v(k_S M_S), \gamma(v(k_Q M_Q))\big).$$

The *cyclic* and *dihedral rigid* metrics $\mathrm{CRM}_q, \mathrm{DRM}_q$ are defined by the same formulae as $\mathrm{CIM}_q, \mathrm{DIM}_q$ above after replacing MCD_q with the metric MCS_q from Definition 3.5. ■

In the limit case $q = +\infty$, any factor $a^{\pm 1/q}$ for $a > 0$ is interpreted as $\lim\limits_{q\to+\infty} a^{\pm 1/q} = 1$. In Definition 4.5, the extended periods $k_S l_S$ and $k_Q l_Q$ can be different. For simplicity, the metrics $\mathrm{MCD}_q, \mathrm{MCS}_q$ were written via projected motifs as in Definition 3.5 but will be computable via the complete invariants (under relevant equivalences) from Definition 4.2.

In the partial case $n = 1$, the projected motifs are empty, so the cases of rigid motion and isometry in $\mathbb{R}^0$ trivially coincide. In both cases, the metrics are obtained by minimizing only the differences d_t between time shifts under cyclic and dihedral permutations.

▶ **Example 4.6.** The periodic sequences $S = \{0, 1\} + 3\mathbb{Z}$ and $Q = \{0, 1, 3\} + 6\mathbb{Z}$ have motifs $M_S = \{0, 1\}$ and $M_Q = \{0, 1, 3\}$ of different sizes $m_S = 2$ and $m_Q = 3$ whose lowest common multiple is $m = 6$. In the notations of Definition 4.5, we get $k_S = \dfrac{m}{|M_S|} = 3$, $k_Q = \dfrac{m}{|M_Q|} = 2$. The extended motifs and periods are $3M_S = \{0, 1, 3, 4, 6, 7\}$, $3l_S = 9$, $2M_Q = \{0, 1, 3, 6, 7, 9\}$, $2l_Q = 12$. Then $\mathrm{TS}(3M_S; 9) = (1, 2, 1, 2, 1, 2)$ and $\mathrm{TS}(2M_Q; 12) = (1, 2, 3, 1, 2, 3)$. Any cyclic or dihedral permutation of $\mathrm{TS}(3M_S; 9)$ relative to $\mathrm{TS}(2M_Q; 12)$ gives the maximum component-wise distance $|1-3| = 2$, so $\mathrm{CIM}_{+\infty}(S,Q) = 2 = \mathrm{DIM}_{+\infty}(S,Q)$. ■

▶ **Theorem 4.7** (solution to Problem 1.4 for 1-periodic sequences)**. (a)** For any 1-periodic sequence $S \subset \mathbb{R}\times\mathbb{R}^{n-1}$ with a motif of m points, $\mathrm{CI}(S), \mathrm{DI}(S)$ from Definition 4.2 are complete invariants under cyclic and dihedral isometry in $\mathbb{R}\times\mathbb{R}^{n-1}$, respectively, and computable in time $O(m^3n)$. The invariants $\mathrm{CR}(S), \mathrm{DR}(S)$ are complete under cyclic and dihedral rigid motion in $\mathbb{R}\times\mathbb{R}^{n-1}$, respectively, and computable in time $O(m^3n+m^2n^3)$.

(b) Any 1-periodic sequence $S\subset\mathbb{R}\times\mathbb{R}^{n-1}$ with a motif of m points can be reconstructed from its complete invariant up to a relevant equivalence from part (a) in time $O(m^3n)$.

(c) The metrics in Definition 4.5 remain invariant if any 1-periodic sequence $S=M+l\vec{e}_1\mathbb{Z}$ is given by its extended motif kM and period kl for any integer $k>0$. For any 1-periodic sequences $S,Q\subset\mathbb{R}\times\mathbb{R}^{n-1}$ with a lowest common multiple m of their motifs sizes, the metrics $\mathrm{CIM}_q, \mathrm{DIM}_q, \mathrm{CRM}_q, \mathrm{DRM}_q$ in Definition 4.5 satisfy all axioms and are computable in times $O(m^3n)$ and $O(m^3n+m^2n^3)$ for isometry and rigid motion, respectively.

(d) Let Q denote a 1-periodic sequence $S\subset\mathbb{R}\times\mathbb{R}^{n-1}$ after perturbing every point of S up to some Euclidean distance ε that is smaller than a half-distance between any points of $t(S)$ and of $t(Q)$. Then $\mathrm{CIM}_q(S,T), \mathrm{DIM}_q(S,Q), \mathrm{CRM}_q(S,Q), \mathrm{DRM}_q(S,Q)\leq 2\varepsilon$. ■

▶ **Example 4.8** (challenging 1-periodic sequences)**.** The infinite family of counter-examples in [40, Fig. 4] to the completeness of past distance-based invariants includes the pairs of the 1-periodic sequences $A^{\pm}\subset\mathbb{R}\times\mathbb{R}^2$ with a period $l>0$ and 6-point motifs $M^+=\{W',C_+,V,W,C'_+,V'\}$ and $M^-=\{W',C_-,V,W,C'_-,V'\}$, where $V=(v_x,v_y,0)$, $W=(\frac{l}{2},w_y,w_z)$, $C_{\pm}=(\frac{l}{4},c_y,\pm c_z)$, and $l,w_y,w_z,c_y,c_z>0$, $v_x,v_y\in[0,\frac{l}{2}]$ are free parameters.

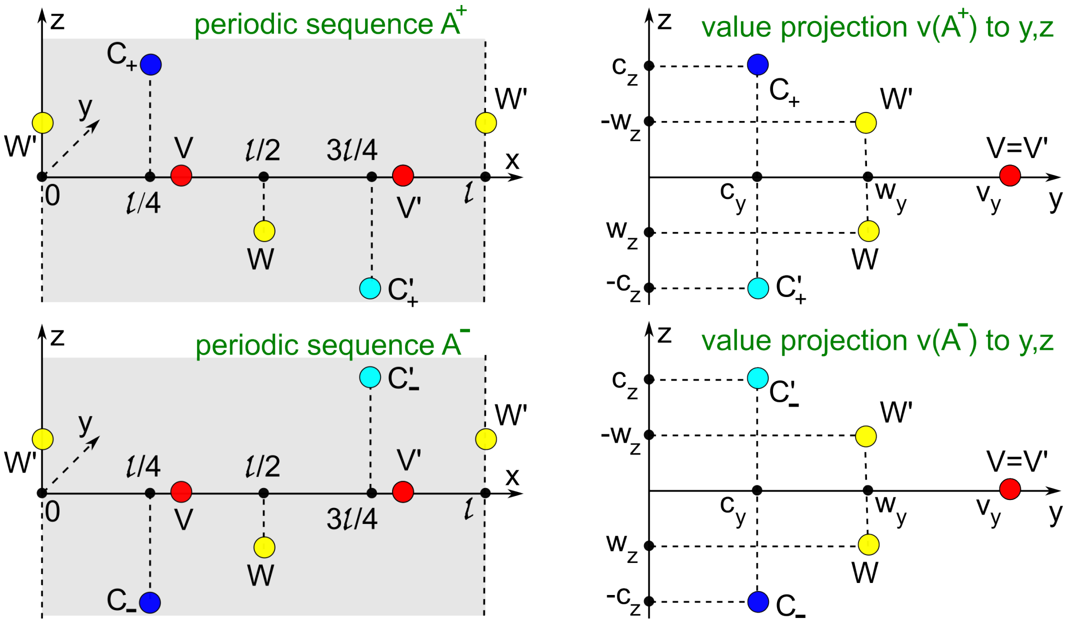

Figure 5 These periodic sequences $A^{\pm}\subset\mathbb{R}\times\mathbb{R}^2$ from [40, Fig. 2] have identical past invariants.

Any point with a dash is obtained by $g(x,y,z)=(x+\frac{l}{2},y,-z)$. The time projections are identical: $t(M^{\pm})=(0,\frac{l}{4},v_x,\frac{l}{2},\frac{3l}{4},\frac{l}{2}+v_x)$. Assuming that $v_x\in(\frac{l}{4},\frac{l}{2})$ as in Fig. 5, the time shifts are $\mathrm{TS}(M^{\pm};l)=(\frac{l}{4},v_x-\frac{l}{4},\frac{l}{2}-v_x,\frac{l}{4},v_x-\frac{l}{4},\frac{l}{2}-v_x)$. The ordered value projections are $v(M^{\pm})=\{(w_y,-w_z),(c_y,\pm c_z),(v_y,0),(w_y,w_z),(c_y,\mp c_z),(v_y,0)\}$. The cyclic distance matrices of M^+ and M^- are on the left and right hand sides below, respectively:

$$\begin{pmatrix} d_{11} & d_{12} & d_{21} & d_{11} & d_{12} & d_{21} \\ d_{21} & d_{22} & d_{12} & d_{21} & d_{22} & d_{12} \\ 2|w_z| & 2|c_z| & 0 & 2|w_z| & 2|c_z| & 0 \end{pmatrix} \neq \begin{pmatrix} d_{22} & d_{12} & d_{21} & d_{22} & d_{12} & d_{21} \\ d_{21} & d_{11} & d_{12} & d_{21} & d_{11} & d_{12} \\ 2|w_z| & 2|c_z| & 0 & 2|w_z| & 2|c_z| & 0 \end{pmatrix}.$$

The differences are highlighted: $d_{11} = \sqrt{(w_y - c_y)^2 + (w_z + c_z)^2}$, $d_{12} = \sqrt{(c_y - v_y)^2 + c_z^2}$, $d_{22} = \sqrt{(w_y - c_y)^2 + (w_z - c_z)^2}$, $d_{21} = \sqrt{(w_y - v_y)^2 + w_z^2}$. The matrix difference has the Minkowski norm $||\mathrm{CDM}(M^+) - \mathrm{CDM}(M^-)||_\infty = |d_{11} - d_{22}| > 0$ unless $c_z = 0$ or $w_z = 0$. If $c_z = 0$, $A^\pm$ are identical. If $w_z = 0$, then $A^\pm$ are isometric by $g(x, y, z) = (x + \frac{l}{2}, y, -z)$.

If both $c_z, w_z \neq 0$, then $\mathrm{CIM}_{+\infty}(A^+, A^-)$ is obtained by minimizing over 6 cyclic permutations $\gamma \in C_6$. The trivial permutation and the shift by 3 positions give $|d_{11} - d_{12}|$. Any other permutation gives $d_t = \max\{v_x - \frac{l}{4}, \frac{l}{2} - v_x\}$ from comparing $\mathrm{TS}(M^+; l)$ with $\gamma(\mathrm{TS}(M^-; l))$ and $d_v = \max\{|a - b|\}$ maximized for all pairs of $a, b \in \{d_{11}, d_{12}, d_{21}, d_{22}\}$.

In all cases, the metric is positive: $\mathrm{CIM}_{+\infty}(A^+, A^-) \geq |d_{11} - d_{22}| > 0$. Hence the invariant CI from Definition 4.2 distinguished these challenging 1-periodic sequences $A^+ \not\cong A^-$. ■

5 Discussion: the importance of Lipschitz continuity for data integrity

This paper rigorously stated Problem 1.4 to design complete invariants and Lipschitz continuous metrics for 1-periodic sequences in $\mathbb{R} \times \mathbb{R}^{n-1}$ for any high-dimension $n \geq 1$. Simpler versions of this problem were previously studied mostly for finitely many points. Even in the simpler case of ordered points, the Lipschitz continuity around degenerate configurations needs the recent strength of a simplex, so Theorem 3.6 is new to the best of our knowledge.

The 1-periodic case is much harder because a minimum period arbitrarily scales up under almost any perturbation of points within the space of 1-periodic sequences. Infinite non-periodic sequences are currently studied through finite subsets, which we considered above, but even the 1-periodic case remained open. Main Theorem 4.7 solved Problem 1.4 for the four equivalences that maintain all distances but can change orientation in a factor of the product $\mathbb{R} \times \mathbb{R}^{n-1}$. All invariants and metrics easily extend to compositions of these equivalences with uniform scaling in any of the factors. It suffices to normalize all distances and strengths by the diameter of a finite set or a minimum period l of a 1-periodic sequence.

Though the invariants in Definition 4.2 are introduced as classes under cyclic or dihedral permutations γ, any 1-periodic sequence $S \subset \mathbb{R} \times \mathbb{R}^{n-1}$ can be reconstructed from any representative time shift TS and a suitable matrix (CDM or CDS) of a finite motif, which requires less space in computer memory. Applying permutations γ is needed only for metric computations in Definition 4.5. Example 4.8 illustrates that the new invariants and metrics can be manually computed even for infinitely many periodic sequences in [40, Fig. 4] that were not distinguishable by generically complete past invariants such as PDD [47].

The Lipschitz continuity is practically important for detecting near-duplicates that have very different periods (primitive cells) because any known material can be easily perturbed with an extended motif and claimed as 'new' especially if some atoms were artificially replaced. Such duplicates were found in the well-curated and world's largest collection of real materials [46] CSD (Cambridge Structural Database) because past comparisons based on finite subsets are slow and unreliable [50]. As a result, five journals are investigating the underlying publications for data integrity [47, section 6]. The simulated data can be much worse because iterative optimizations are expected to approximate the same local optima on different runs, see [2, Tables 1-2]. Hence the Lipschitz continuity helps maintain the public trust in science.

The polynomial-time complexities in Theorems 3.6 and 4.7 suffice in practice because the new invariants form a hierarchy from the easy and fast invariants to the slower but complete. For example, we should first compare real 1-periodic sequences by their time shifts

TS in linear time $O(m)$ and continue below only for pairs with very close time shifts. After computing the matrix CDM in time $O(m^2)$ for a smaller number of potential near-duplicates, we compare simpler subinvariants such as the first rows of CDMs (distances to the next neighbor in time) or column averages still in time $O(m^2)$ by applying $O(m)$ permutations from C_m or D_m to vectors of length m. Such a hierarchical filtering was done for 200+ billion pairwise comparisons of all periodic materials in the CSD within two days on a modest desktop [47], though the underlying Earth Mover's Distance [43] has a cubic complexity.

Problem 1.4 provides a practical alternative to artificial networks, which 'horizontally explore' continuously infinite data spaces through finite datasets, which are discrete samples of measure 0. Solutions to analogs of Problem 1.4 will 'fly vertically' to map continuous data spaces from a 'satellite' point of view as in [8, 49]. Indeed, 1-periodic sequences and isometries can be replaced with any real objects and equivalences but the conditions of completeness, reconstruction, Lipschitz continuity, and polynomial-time computability remain essential.

The implementation of the new invariants and metrics can be released in October 2024. We thank all reviewers in advance for their valuable time and helpful suggestions.

A Appendix: detailed proofs of Lemma 3.4 and Theorems 3.6, 4.7

▶ **Example A.1** (strengths and signs)**.** For the sequence T_1 in Fig. 3 with the points $p_1 = (0,0)$, $p_2 = (0,1)$, $p_3 = (1,0)$, $p_4 = (1,1)$, the first 2×2 determinant with the columns $p_2 - p_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and $p_3 - p_2 = (1,-1)$ is $\det\begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix}$ has sign -1. The further determinants for $i = 2,3,4$ are $\det\begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix} = +1$, $\det\begin{pmatrix} 1 & -1 \\ 0 & -1 \end{pmatrix} = -1$, $\det\begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix} = +1$, so $\mathrm{sign}(T_1) = (-1,+1,-1,+1)$. All triangles on 4 triples p_i, p_{i+1}, p_{i+2} for $i = 1,2,3,4$ have the sides $1, 1, \sqrt{2}$, half-perimeter $p = 1 + \frac{1}{\sqrt{2}}$, area $V = \frac{1}{2}$, and strength $\sigma = \frac{1}{\sqrt{2}(1+\sqrt{2})^3}$. ■

Table 1 Acronyms and references for the new invariants and metrcis from sections 3 and 4.

$\mathrm{CDM}(T)$	Cyclic Distance Matrix of a finite sequence $T \subset \mathbb{R}^n$	Definition 3.1
$\mathrm{CDS}(T)$	matrix of Cyclic Distances and Signs of a finite sequence $T \subset \mathbb{R}^n$	Definition 3.3
MCD_q	Metric on Cyclic Distance matrices (CDM)	Definition 3.5
MCS_q	Metric on matrices of Cyclic distances and Signs (CDS)	Definition 3.5
$\mathrm{TS}(M;l)$	Time Shift for a motif M and period l of a 1-periodic sequence	Definition 4.1
$\mathrm{CI}(S)$	Cyclic Isometry invariant of a 1-periodic sequence $S \subset \mathbb{R}\times\mathbb{R}^{n-1}$	Definition 4.2
$\mathrm{DI}(S)$	Dihedral Isometry invariant of a 1-periodic sequence $S \subset \mathbb{R}\times\mathbb{R}^{n-1}$	Definition 4.2
$\mathrm{CR}(S)$	Cyclic Rigid invariant of a 1-periodic sequence $S \subset \mathbb{R}\times\mathbb{R}^{n-1}$	Definition 4.2
$\mathrm{DR}(S)$	Dihedral Rigid invariant of a 1-periodic sequence $S \subset \mathbb{R}\times\mathbb{R}^{n-1}$	Definition 4.2
$\mathrm{CI}(S)$	Cyclic Isometry invariant of a 1-periodic sequence $S \subset \mathbb{R}\times\mathbb{R}^{n-1}$	Definition 4.2
$\mathrm{DI}(S)$	Dihedral Isometry invariant of a 1-periodic sequence $S \subset \mathbb{R}\times\mathbb{R}^{n-1}$	Definition 4.2
CIM_q	Cyclic Isometry Metric between 1-periodic sequences in $\mathbb{R}\times\mathbb{R}^{n-1}$	Definition 4.5
DIM_q	Dihedral Isometry Metric between 1-periodic sequences in $\mathbb{R}\times\mathbb{R}^{n-1}$	Definition 4.5
CRM_q	Cyclic Rigid Metric between 1-periodic sequences in $\mathbb{R}\times\mathbb{R}^{n-1}$	Definition 4.5
DRM_q	Dihedral Rigid Metric between 1-periodic sequences in $\mathbb{R}\times\mathbb{R}^{n-1}$	Definition 4.5

▶ **Example A.2** (metric MCD_q)**.** For any $q \in [1,+\infty)$, we use cyclic distance matrices from Example 3.2 to compute $\mathrm{MCD}_q(T_1,T_3) = (\frac{2}{3})^{1/q}(\sqrt{2}-1)$, $\mathrm{MCD}_q(T_3,T_4) = (\frac{1}{6})^{1/q}(\sqrt{5}-1)$, and $\mathrm{MCD}_q(T_1, T_4) = (\frac{1}{2}(\sqrt{2}-1)^q + \frac{1}{6}(\sqrt{5}-\sqrt{2})^q)^{1/q}$. The triangle inequality holds for $q \geq 1$ as follows: $\big(\mathrm{MCD}_q(T_1, T_3) + \mathrm{MCD}_q(T_3, T_4)\big)^q = \big((\frac{2}{3})^{1/q}(\sqrt{2}-1) + (\frac{1}{6})^{1/q}(\sqrt{5}-1)\big)^q \geq \big((\frac{1}{2})^{1/q}(\sqrt{2}-1) + (\frac{1}{6})^{1/q}(\sqrt{5}-\sqrt{2})\big)^q \geq \frac{1}{2}(\sqrt{2}-1)^q + \frac{1}{6}(\sqrt{5}-\sqrt{2})^q = \big(\mathrm{MCD}_q(T_1, T_4)\big)^q$ due to $(a+b)^q \geq a^q + b^q$ for $a, b > 0$ and $q \geq 1$. For $q = +\infty$, the inequality becomes $(\sqrt{2}-1) + (\sqrt{5}-1) \geq \sqrt{5}-\sqrt{2}$. Finally, $T_5 \not\cong T_6$ have $\mathrm{MCD}_q(T_5, T_6) = 2^{1/q}(\sqrt{5}-1)$. ■

Proof of Lemma 3.4. The shift permutation γ_m increments each index $1, 2, \ldots, m$ (modulo m), so the columns of $\mathrm{CDM}(T)$ are shifted in the same way, also the signs $s(T)$, while row indices are differences between point indices and remain the same under γ_m.

The reverse permutation ι_m reverses the order of points and hence the columns of $\mathrm{CDM}(T)$. The rows are also reversed under ι_m because the next point for p_i in the reversed sequence $p_m, \ldots, p_1$ is the previous point of p_i in the original list. Also, under ι_m, the n difference vectors $p_{i+1} - p_i, p_{i+2} - p_{i+1}, \ldots, p_{i+n} - p_{i+n-1}$ reverse all their n signs order and also the order. The reverse permutation $(s_1, \ldots, s_n) \mapsto (s_n, \ldots, 1)$ decomposes into $[n/2]$ transpositions, where $[n/2]$ is the largest integer not greater than $n/2$. Hence the $n \times n$ determinant under ι_m changes its sign by the factor $(-1)^n(-1)^{[n/2]} = (-1)^{[3n/2]}$.

Any mirror reflection in $\mathbb{R}^n$ keeps all distances and reverses all signs in the row $\mathrm{sign}(T)$. ◀

The *affine dimension* $0 \leq \mathrm{aff}(A) \leq n$ of a point set $A = \{p_1, \ldots, p_m\} \subset \mathbb{R}^n$ is the maximum dimension of the vector space generated by all inter-point vectors $p_i - p_j$, $i, j \in \{1, \ldots, m\}$. The isometry invariant $\mathrm{aff}(A)$ is independent of an order of points. Any 2 distinct points have $\mathrm{aff} = 1$. Any 3 points that are not in the same straight line have $\mathrm{aff} = 2$.

Lemma A.3 provides a criterion for a matrix to be realizable by squared distances in $\mathbb{R}^n$.

▶ **Lemma A.3** (distance realization). **(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 points $p_0 = 0, p_1, \ldots, p_{m-1} \in \mathbb{R}^n$ *if and only if* the $(m-1) \times (m-1)$ matrix $g_{ij} = \dfrac{s_{0i} + s_{0j} - s_{ij}}{2}$ has non-negative eigenvalues.
(b) If the condition in (a) holds, $\mathrm{aff}(0, p_1, \ldots, p_{m-1})$ equals the number $k \leq m-1 \leq n$ of positive eigenvalues. Then $g_{ij} = p_i \cdot p_j$ define the *Gram matrix* G of the vectors $p_1, \ldots, p_{m-1} \in \mathbb{R}^n$, which are reconstructable in time $O(m^3)$ up to an orthogonal map in $\mathbb{R}^n$.

Proof of Lemma A.3. **(a)** We extend [15, Theorem 1] to the case $m < n+1$ and also justify the reconstruction of $p_1, \ldots, p_{m-1}$ in time $O(m^3)$ uniquely in $\mathbb{R}^n$ up to an orthogonal map from the orthogonal group $\mathrm{O}(n)$.

The part *only if* $\Rightarrow$. Let a symmetric matrix S consist of squared distances between points $p_0 = 0, p_1, \ldots, p_{m-1} \in \mathbb{R}^n$. For $i, j = 1, \ldots, m-1$, the matrix with the elements

$$g_{ij} = \frac{s_{0i} + s_{0j} - s_{ij}}{2} = \frac{p_i^2 + p_j^2 - |p_i - p_j|^2}{2} = p_i \cdot p_j$$

is 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 $p_1, \ldots, p_{m-1}$. For any vector $v \in \mathbb{R}^{m-1}$, we have

$$0 \leq |Pv|^2 = (Pv)^T(Pv) = v^T(P^T P)v = v^T G v.$$

Since the quadratic form $v^T G v \geq 0$ for any $v \in \mathbb{R}^{m-1}$, the matrix G is positive semi-definite meaning that G has only non-negative eigenvalues by [25, Theorem 7.2.7].

The part *if* $\Leftarrow$. For any positive semi-definite matrix G, there is an orthogonal matrix B such that $B^T G B = D$ 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 eigenvalues of G.

(b) The number of positive eigenvalues of G equals the dimension $k = \mathrm{aff}(\{0, p_1, \dots, p_{m-1}\})$ of the subspace in $\mathbb{R}^n$ linearly spanned by $p_1, \dots, p_{m-1}$. We may assume that all $k \leq n$ positive eigenvalues of G correspond to the first k coordinates of $\mathbb{R}^n$. Since $B^T = B^{-1}$, the matrix $G = BDB^T = (B\sqrt{D})(B\sqrt{D})^T$ becomes the Gram matrix of the columns of $B\sqrt{D}$. These columns become the reconstructed vectors $p_1, \dots, p_{m-1} \in \mathbb{R}^n$.

If there is another diagonalization $\tilde{B}^T G \tilde{B} = \tilde{D}$ for $\tilde{B} \in \mathrm{O}(n)$, 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 $G = \tilde{B} D \tilde{B}^T = (\tilde{B}\sqrt{D})(\tilde{B}\sqrt{D})^T$ is the Gram matrix of the columns of $\tilde{B}\sqrt{D}$.

The new columns are obtained from the previously reconstructed vectors $p_1, \dots, p_{m-1} \in \mathbb{R}^n$ after multiplying by the orthogonal matrix $B\tilde{B}^T$. Hence the reconstruction is unique up to an orthogonal transformation from $\mathrm{O}(n)$. Computing eigenvectors $p_1, \dots, p_{m-1}$ needs a diagonalization of G in time $O(m^3)$, see [41, section 11.5]. ◀

Proof of Theorem 3.6. **(a,b)** Any isometry in $\mathbb{R}^n$ maintains all interpoint distances and hence preserves $\mathrm{CDM}(T)$. Any rigid motion (orientation-preserving isometry) in $\mathbb{R}^n$ preserves the signs of $n \times n$ determinants, hence the row $\mathrm{sign}(T)$ and matrix $\mathrm{CDS}(T)$ from Definition 3.3. Each of $O(m^2)$ Euclidean distances in $\mathrm{CDM}(T)$ depends on n coordinates and needs $O(n)$ time. Each of m signs in the row $\mathrm{sign}(T)$ of $\mathrm{CDS}(T)$ needs $O(n^3)$ time by Gaussian elimination. So $\mathrm{CDM}(T), \mathrm{CDS}(T)$ are computable in times $O(m^2 n)$ and $O(m^2 n + mn^3)$, respectively.

For any finite sequence $T = (p_1, \dots, p_m)$, the cyclic distance matrix $\mathrm{CDM}(T)$ uniquely determines the classical distance matrix $\mathrm{DM}(T)$ and hence (after shifting p_1 to the origin) the Gram matrix of scalar products $p_i \cdot p_j$ for $1 < i, j \leq m$, which suffices to reconstruct T uniquely up to isometry by Lemma A.3(b) in time $O(m^3)$. If $\mathrm{CDS}(T)$ contains at least one non-zero sign, then $\mathrm{CDS}(\bar{T}) \neq \mathrm{CDS}(T)$, so T is distinguished from its mirror image $\bar{T}$ and hence uniquely determined from $\mathrm{CDS}(T)$ up to rigid motion in $\mathbb{R}^n$. If the row $\mathrm{sign}(T)$ consists of zeros, then T is contained within an $(n-1)$-dimensional subspace of $\mathbb{R}^n$. Indeed, $\mathrm{sign}_1(T) = 0$ means that the first $n+1$ points p_{n+1} is in the $(n-1)$-dimensional subspace S that is affinely spanned by $p_1, \dots, p_n$. Then by induction on $i = 2, \dots, m-n$, $\mathrm{sign}_i(T) = 0$ implies that p_{n+i} is in the same subspace S. Within S, the mirror images $\bar{T}$ and T with respect to any $(n-2)$-dimensional subspace $L \subset S$ are related by a high-dimensional rotation around L in $\mathbb{R}^n$, so T is uniquely determined by $\mathrm{CDS}(T)$ also in any degenerate case.

(c) The metric axioms for the distances $\mathrm{MCD}_q(S,T), \mathrm{MCS}_q(S,T)$ follow from these axioms for the Minkowski metric [18]. Taking the maximum respects the axioms as a metric transform by [16, section 4.1]. After computing the invariants $\mathrm{CDM}(S)$ and $\mathrm{CDM}(T)$ in time $O(m^2 n)$, the metric MCD_q needs only $O(n)$ extra time. Each of $2m$ strengths for the metric MCS_q needs time $O(n^3)$ for an $n \times n$ determinant, hence only $O(mn^3)$ extra time, followed by $O(m)$ time to take the maxima in the formula for MCS_q from Definition 3.5.

(d) We are given a bijection $\beta : T \to S$ that shifts every point up to ε in Euclidean distance. Then the distances between any points $p_i, p_j \in T$ and their ε-close images $\beta(p_i), \beta(p_j) \in S$ differ by at most 2ε. The matrix CDM contains $m(m-1)$ distances. By Definition 3.5, $\mathrm{MCD}_q(S,T) = \dfrac{||\mathrm{CDM}(S) - \mathrm{CDM}(T)||_q}{\big(m(m-1)\big)^{1/q}} \leq \dfrac{\big(m(m-1)(2\varepsilon)^q\big)^{1/q}}{\big(m(m-1)\big)^{1/q}} = 2\varepsilon$. The Lipschitz continuity $|\sigma_i(S) - \sigma_i(T)| \leq \lambda_n \varepsilon$ by [48, Theorem 4.4] was proved in [27]. If $\mathrm{sign}_i(S)\mathrm{sign}_i(T) \geq 0$, then $\frac{2}{\lambda_n}\big|\mathrm{sign}_i(S)\sigma_i(S) - \mathrm{sign}_i(T)\sigma_i(T)\big| = \frac{2}{\lambda_n}|\sigma_i(S) - \sigma_i(T)| \leq 2\varepsilon$. If $\mathrm{sign}_i(S) = -\mathrm{sign}_i(T)$, the straight-line deformation of the points $p_j(t) = (1-t)p_j + t\beta(p_j)$, $t \in [0,1]$, $j = i, \dots, i+n$, passes through a degenerate subsequence A with $\sigma = 0$. Each $p_j(t)$

shifts from T by at most $t\varepsilon$ to the degenerate subsequence A, then by at most $(1-t)\varepsilon$ to S. The Lipschitz continuities $|\sigma_i(S) - 0| \leq \lambda_n t\varepsilon$ and $|0 - \sigma_i(T)| \leq \lambda_n(1-t)\varepsilon$ imply that

$$\frac{2}{\lambda_n}\left|\mathrm{sign}_i(S)\sigma_i(S) - \mathrm{sign}_i(T)\sigma_i(T)\right| = \frac{2}{\lambda_n}(\sigma_i(S) + \sigma_i(T)) \leq \frac{2}{\lambda_n}(\lambda_n(1-t)\varepsilon + \lambda_n t\varepsilon) = 2\varepsilon.$$

The maxima in Definition 3.5 guarantee that $\mathrm{MCD}_q(S,T) \leq 2\varepsilon$ as required. ◀

Lemma A.4 was inspired by [7, Propositions 8.5(2) and 8.6], which were proved briefly.

▶ **Lemma A.4** (metric on a quotient space under action)**.** Let a finite group G act on a space X with a metric d_X by isometries so that $d_X(f(a), f(b)) = d_X(a,b)$ for any $a,b \in X$ and $f \in G$. Then the *quotient space* X/G consisting of equivalence classes $[a] = \{f(a) \in X \mid f \in G\}$ has the *quotient distance* $d([a],[b]) = \min\limits_{f\in G} d_X(f(a), b)$ satisfying all metric axioms. ■

Proof. All axioms for d follow from the axioms for d_X. The coincidence axiom means that $d([a],[b]) = \min\limits_{f\in G} d_X(f(a),b) = 0$ if and only if $d_X(f(a),b) = 0$ for some $f \in G$, so $f(a) = b$ and hence $[a] = [b]$. The symmetry axiom follows by using the inverse operation in G, i.e. $d([a],[b]) = \min\limits_{f\in G} d_X(f(a),b) = \min\limits_{f\in G} d_X(b, f(a)) = \min\limits_{f^{-1}\in G} d_X(f^{-1}(b), a) = d([b],[a])$.

To prove the triangle inequality $d([a],[b]) + d([b],[c]) \geq d([a],[c])$, take $f,g \in G$ such that $d([a],[b]) = d_X(f(a),b)$ and $d([b],[c]) = d_X(g(b),c)$. Then $d([a],[b]) + d([b],[c]) = d_X(g\circ f(a), g(b)) + d_X(g(b),c) \geq d_X(g\circ f(a), c) \geq \min\limits_{h\in G} d_X(h(a),c) = d([a],[c])$ as required. ◀

Proof of Theorem 4.7. (a,b) Any cyclic and dihedral isometry and rigid motion of $\mathbb{R} \times \mathbb{R}^{n-1}$ from Definition 1.3 preserve the class of the time shift TS, which is the vector of differences between successive time projections in Definition 4.1, under the actions of C_m, D_m, respectively. For a motif of m points, TS needs only $O(m)$ time. Hence the invariance of $\mathrm{CI}(S), \mathrm{DI}(S), \mathrm{CR}(S), \mathrm{DR}(S)$ and their times follow from Theorem 3.6(a) for the projected motif $v(M) \subset \mathbb{R}^{n-1}$. The completeness and reconstruction in time $O(m^3 n)$ follow from Theorem 3.6(b), which reconstructs $v(M) \subset \mathbb{R}^{n-1}$ uniquely up to a relevant equivalence after assigning the time projections $0, d_1, \dots, d_{m-1}$ to the ordered points $p_1, \dots, p_m \in v(M)$, respectively, where $\mathrm{TS} = (d_1, \dots, d_m)$ is the correspondingly ordered time shift.

(c) Let a 1-periodic sequence $S = M + l\vec{e}_1\mathbb{Z}$ be given by its extended motif kM and period kl for any integer $k > 0$. The time shift $\mathrm{TS}(kM; kl)$ is a concatenation of m identical vectors $\mathrm{TS}(M; l)$. The projected motif $v(kM) \subset \mathbb{R}^{n-1}$ is the set of k identical copies of $v(M)$.

Hence the $km \times (km-1)$ matrix $\mathrm{CDM}(v(kM))$ consists of k^2 identical $m(m-1)$ matrices $\mathrm{CDM}(v(M))$ separated by extra $k-1$ rows of zeros, which represent the zero distances from each point $p \in v(M)$ to its other $k-1$ copies in $v(kM)$ at the same location in $\mathbb{R}^{n-1}$.

Any cyclic permutation $\gamma \in C_m$ defined to the extended permutation $k\gamma \in C_{km}$ that shifts all km elements by the same number of positions as γ. Applying such an extended permutation γ to a block vector $\mathrm{TS}(kM; kl)$ or a block matrix $\mathrm{CDM}(v(kM))$ described above is equivalent to applying γ to the original vector or matrix, and then extending the output by the factor k. In other words, the minimization of differences with a block vector or a block matrix over km cyclic permutations from the larger group C_{km} is equivalent to the minimization of the differences with a smaller original vector or a matrix over k cyclic permutations from C_m. Then the metric CIM is invariant under any extension of a motif and a period. The same arguments apply to dihedral permutations and matrix CDS.

Hence, to justify the metric axioms below, we can assume that all involved 1-periodic sequences are scaled up to a common size of their extended motifs. The auxiliary distances d_t, d_v in Definition 4.5 are standard Minkowski metrics. Taking the maximum of several metrics respects all axioms as a standard metric transform [16, section 4.1]. The final operation of distance minimization over the actions of the groups C_m, D_m allows us to consider the outputs as quotient distances, which satisfy the metric axioms by Lemma A.4.

Due to the minimization by the actions of the groups C_m, D_m, each of the metrics $\mathrm{CIM}_q, \mathrm{DIM}_q, \mathrm{CRM}_q, \mathrm{DRM}_q$ requires only an extra factor $O(m)$ in comparison with the times $O(m^2 n)$ and $O(m^2 n + mn^3)$ from Theorem 3.6(c) for the relevant metrics MCD_q (under isometry) and MCS_q (under rigid motion) between the projected motifs in $\mathbb{R}^{n-1}$. Indeed, the Minkowski metric between time shifts adds only an additive time $O(m)$, which is dominated by the time $O(m^2 n)$ for the metric between cyclic distance matrices CDM.

(d) Any perturbation of points up to Euclidean distance ε in $\mathbb{R} \times \mathbb{R}^{n-1}$ changes their time projections by at most ε. Then any difference between successive time projections changes by at most 2ε, which is less than the distance between any successive points in the time projection $t(S)$ and in the time projection $t(Q)$. Hence there is a bijection $S \to Q$ respecting the time order of all points. When computing the Minkowski metric between the time shifts for the identity permutation $\gamma = \mathrm{id}$, the maximum deviation 2ε emerges m times and hence leads to the overall factor $(2\varepsilon)m^{1/q}$. The extra factor $m^{-1/q}$ in the formula for the distance b in Definition 4.5 gives the final factor 2ε. The Lipschitz constant 2 is guaranteed for the metrics $\mathrm{MCD}_q, \mathrm{MCS}_q$ by Theorem 3.6(a). The minimization over permutations γ from C_m or D_m can make the final distance only smaller. So the final Lipschitz contant is 2. ◀

References

1. Helmut Alt, Kurt Mehlhorn, Hubert Wagener, and Emo Welzl. Congruence, similarity, and symmetries of geometric objects. *Discrete and Computational Geometry*, 3:237–256, 1988.
2. Olga Anosova, Vitaliy Kurlin, and Marjorie Senechal. The importance of definitions in crystallography. *IUCrJ*, 11:453–463, 2024. `doi:10.1107/S2052252524004056`.
3. Serge Belongie, Jitendra Malik, and Jan Puzicha. Shape matching and object recognition using shape contexts. *Transactions PAMI*, 24(4):509–522, 2002.
4. Mireille Boutin and Gregor Kemper. On reconstructing n-point configurations from the distribution of distances or areas. *Advances in Applied Mathematics*, 32(4):709–735, 2004.
5. Peter Brass and Christian Knauer. Testing the congruence of d-dimensional point sets. In *Proceedings of SoCG*, pages 310–314, 2000.
6. Peter Brass and Christian Knauer. Testing congruence and symmetry for general 3-dimensional objects. *Computational Geometry*, 27(1):3–11, 2004.
7. Martin R Bridson and André Haefliger. *Metric spaces of non-positive curvature*, volume 319. Springer Science & Business Media, 2013.
8. Matthew J Bright, Andrew I Cooper, and Vitaliy A Kurlin. Geographic-style maps for 2-dimensional lattices. *Acta Crystallographica Section A*, 79(1):1–13, 2023.
9. Hans-Georg Carstens, Walter Deuber, Wolfgang Thumser, and Elke Koppenrade. Geometrical bijections in discrete lattices. *Combinatorics, Probability and Computing*, 8:109–129, 1999.
10. Dalmeet Singh Chawla. Crystallography databases hunt for fraudulent structures. `https://cen.acs.org/research-integrity/Crystallography-databases-hunt-fraudulent-structures/102/i8`, 2024.
11. Anthony K Cheetham and Ram Seshadri. Artificial intelligence driving materials discovery? perspective on the article: Scaling deep learning for materials discovery. *Chemistry of Materials*, 36(8):3490–3495, 2024.

12 L Paul Chew, Michael T Goodrich, Daniel P Huttenlocher, Klara Kedem, Jon M Kleinberg, and Dina Kravets. Geometric pattern matching under euclidean motion. *Computational Geometry*, 7(1-2):113–124, 1997.

13 Paul Chew, Dorit Dor, Alon Efrat, and Klara Kedem. Geometric pattern matching in d-dimensional space. *Discrete & Computational Geometry*, 21(2):257–274, 1999.

14 Paul Chew and Klara Kedem. Improvements on geometric pattern matching problems. In *Scandinavian Workshop on Algorithm Theory*, pages 318–325, 1992.

15 Boris V Dekster and John B Wilker. Edge lengths guaranteed to form a simplex. *Archiv der Mathematik*, 49(4):351–366, 1987.

16 Elena Deza and Michel Marie Deza. *Encyclopedia of distances.* Springer, 2009.

17 Richard Feynman. *The Feynman lectures on physics*, volume 1. 1971.

18 Thomas Foertsch and Anders Karlsson. Hilbert metrics and minkowski norms. *Journal of Geometry*, 83(1-2):22–31, 2005.

19 Philip Hans Franses and Richard Paap. *Periodic time series models.* Oxford Univ. Press, 2004.

20 Michael T Goodrich, Joseph SB Mitchell, and Mark W Orletsky. Approximate geometric pattern matching under rigid motions. *Transactions PAMI*, 21(4):371–379, 1999.

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

22 Francisco Grünbaum and Calvin Moore. The use of higher-order invariants in the determination of generalized Patterson cyclotomic sets. *Acta Cryst. A*, 51:310–323, 1995.

23 Joong Tark Han, Jeong In Jang, Joon Young Cho, Jun Yeon Hwang, Jong Seok Woo, Hee Jin Jeong, Seung Yol Jeong, Seon Hee Seo, and Geon-Woong Lee. Synthesis of nanobelt-like 1-dimensional silver/nanocarbon hybrid materials for flexible and wearable electronics. *Scientific reports*, 7(1):4931, 2017.

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

25 Roger A Horn and Charles R Johnson. *Matrix analysis.* Cambridge University Press, 2012.

26 Daniel P Huttenlocher, Gregory A. Klanderman, and William J Rucklidge. Comparing images using the Hausdorff distance. *Transactions on pattern analysis and machine intelligence*, 15(9):850–863, 1993.

27 Vitaliy Kurlin. The strength of a simplex is the key to a continuous isometry classification of euclidean clouds of unlabelled points. *arXiv:2303.13486*, 2023.

28 Ido Levin, Robert Deegan, and Eran Sharon. Self-oscillating membranes: Chemomechanical sheets show autonomous periodic shape transformation. *Physical Review Letters*, 125(17):178001, 2020.

29 Sushovan Majhi, Jeffrey Vitter, and Carola Wenk. Approximating gromov-hausdorff distance in euclidean space. *Computational Geometry*, 116:102034, 2024.

30 Siddharth Manay, Daniel Cremers, Byung-Woo Hong, Anthony J Yezzi, and Stefano Soatto. Integral invariants for shape matching. *IEEE Transactions on pattern analysis and machine intelligence*, 28(10):1602–1618, 2006.

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

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

33 Amil Merchant, Simon Batzner, Samuel S Schoenholz, Muratahan Aykol, Gowoon Cheon, and Ekin Dogus Cubuk. Scaling deep learning for materials discovery. *Nature*, 624(7990):80–85, 2023.

34 Hadeel Moustafa, Peder Meisner Lyngby, Jens Jørgen Mortensen, Kristian S Thygesen, and Karsten W Jacobsen. Hundreds of new, stable, one-dimensional materials from a generative machine learning model. *Physical Review Materials*, 7(1):014007, 2023.

35 P. Niggli. *Krystallographische und strukturtheoretische Grundbegriffe*, volume 1. Akademische verlagsgesellschaft mbh, 1928.

36 Robert Osada, Thomas Funkhouser, Bernard Chazelle, and David Dobkin. Shape distributions. *ACM Transactions on Graphics*, 21(4):807–832, 2002.

37 Erwin Parthé, Louise Gelato, Bernard Chabot, Marinella Penzo, Karin Cenzual, and Roman Gladyshevskii. *TYPIX standardized data and crystal chemical characterization of inorganic structure types*. Springer Science & Business Media, 2013.

38 A Patterson. Ambiguities in the X-ray analysis of crystal structures. *Physical Reviews*, 65:195, 1944.

39 Helmut Pottmann, Johannes Wallner, Qi-Xing Huang, and Yong-Liang Yang. Integral invariants for robust geometry processing. *Comp. Aided Geom. Design*, 26(1):37–60, 2009.

40 Sergey N Pozdnyakov and Michele Ceriotti. Incompleteness of graph neural networks for points clouds in three dimensions. *Machine Learning: Science and Technology*, 3(4):045020, 2022.

41 William H Press, Saul A Teukolsky, William T Vetterling, and Brian P Flannery. *Numerical recipes: the art of scientific computing*. Cambridge University Press, 2007.

42 A Pulido, L Chen, T Kaczorowski, D Holden, M Little, S Chong, B Slater, D McMahon, B Bonillo, C Stackhouse, A Stephenson, C Kane, R Clowes, T Hasell, A Cooper, and G Day. Functional materials discovery using energy–structure–function maps. *Nature*, 543:657–664, 2017.

43 Y. Rubner, C. Tomasi, and L. Guibas. The Earth Mover's Distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121, 2000.

44 Felix Schmiedl. Computational aspects of the Gromov–Hausdorff distance and its application in non-rigid shape matching. *Discrete Comp. Geometry*, 57:854–880, 2017.

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

46 Suzanna C Ward and Ghazala Sadiq. Introduction to the cambridge structural database–a wealth of knowledge gained from a million structures. *CrystEngComm*, 22(43):7143–7144, 2020.

47 Daniel Widdowson and Vitaliy Kurlin. Resolving the data ambiguity for periodic crystals. *Advances in Neural Information Processing Systems (NeurIPS)*, 35:24625–24638, 2022.

48 Daniel Widdowson and Vitaliy Kurlin. Recognizing rigid patterns of unlabeled point clouds by complete and continuous isometry invariants with no false negatives and no false positives. *Proceedings of Computer Vision and Pattern Recognition*, pages 1275–1284, 2023.

49 Daniel Widdowson and Vitaliy Kurlin. Continuous invariant-based maps of the cambridge structural database. *Crystal Growth and Design*, 2024. `doi:10.1021/acs.cgd.4c00410`.

50 Daniel Widdowson, Marco Mosca, Angeles Pulido, Vitaliy Kurlin, and Andrew Cooper. Average minimum distances of periodic point sets - fundamental invariants for mapping all periodic crystals. *MATCH Comm. in Mathematical and in Computer Chemistry*, 87:529–559, 2022.