

The Strength of a Geometric Simplex Helps Quantify Geometric Affinities of Molecules

Olga D Anosova^a, Vitaliy A Kurlin^{a,*}

^a*School of Computer Science and Informatics, University of Liverpool,
Liverpool L69 3BX, UK*

vitaliy.kurlin@gmail.com

(Received September 14, 2026)

Abstract

All molecules naturally live in a 3-dimensional space, though some of them are linear, like carbon dioxide, or flat, like water or benzene. Molecules were recently classified as clouds of unordered atoms by complete invariants with Lipschitz continuous metrics, all computable in polynomial time. The key ingredient of this new classification is the strength of a geometric simplex. We rigorously prove the Lipschitz continuity of this strength in any dimension and extend this concept to continuously quantify the degeneracy of atomic clouds when they can close to a flat or linear configuration.

1 The importance of Lipschitz continuity for quantifying molecular geometry

Molecules are naturally represented as clouds of unordered atoms or atomic centers, many of which can be chemically indistinguishable [9].

The first fundamental question that can be asked about any data objects is “Same or different?” A rigorous answer to this question requires

*Corresponding author.

an *equivalence* [12], which is any binary relation $\sim$ satisfying three axioms: (1) *reflexivity*: $A \sim A$, (2) *symmetry*: if $A \sim B$ then $B \sim A$, and (3) *transitivity*: if $A \sim B \sim C$ then $A \sim C$. The transitivity axiom guarantees a well-defined classification as a splitting into disjoint *equivalence classes* $[A] = \{\text{all objects } B \sim A\}$ so that if any classes $[A], [C]$ overlap, share a common object B , they should coincide: $[A] = [C]$.

Though one can consider many equivalences on molecules, one of them stands out due to preserving the intrinsic structure of rigid shapes, including crystalline materials: a *rigid motion* in $\mathbb{R}^n$ is a composition of translations and rotations. A crystal *structure* was defined as an equivalence class of periodic sets of atoms under rigid motion [3, Def 6].

Even if a molecule is flexible, such as a protein, its functional properties depend on a *rigid conformation* (an equivalence class under rigid motion) and should be reliably identified [4, 16, 30]. Another practical equivalence on molecules is an *isometry*, which can be defined as a composition of a rigid motion and mirror reflections in $\mathbb{R}^n$. Isometry is weaker than rigid motion in the sense that if any objects are related by rigid motion, then they are isometric, but not vice versa. Any mirror images that are not equivalent under rigid motion map to a single class under isometry [20].

A unique identification of any object under a given equivalence needs a *complete invariant* that is a function I on objects such that $A \sim B$ if and only if $I(A) = I(B)$. If points $p_1, \dots, p_m \in \mathbb{R}^n$ are ordered, they are uniquely determined under isometry by the matrix of distances $|p_i - p_j|$ or the Gram matrix of scalar products $p_i \cdot p_j$, known since 1935, see [25].

However, if the m given points are not ordered, which is the most practical case, extending this complete invariant to comparisons of point clouds requires exponentially many $m!$ permutations on m points.

The case of $m = 3$ points has been known for 2000+ years as the side-side theorem, which continuously parametrizes the space of triangles (equivalence classes of 3-point clouds) under *isometry* by three inter-point distances a, b, c as $\{(a, b, c) \in \mathbb{R}^3 \mid 0 < a \leq b \leq c \leq a + b\} \subset \mathbb{R}^3$.

The crucial obstacle to developing such parameterizations for all real

data under equivalences is the discontinuity of traditional representations.

Positions in a Euclidean space, such as atomic centers, are always uncertain due to measurement noise or thermal vibrations, see Feynman’s first lecture on physics *Atoms in motion* [14, chapter 1]. Data objects from molecules to galaxies are studied through dynamic simulations and their properties are expected to continuously change depending on rigid conformations. Moreover, most optimization methods implicitly assume the smoothness of cost functions, which is stronger than continuity.

Since all real objects are at least slightly different under noise, the second fundamental question is “If different, by how much?”

A rigorous answer to this question requires a distance *metric* d on equivalence classes of data objects satisfying the three metric axioms:

(1) $d(A, B) = 0$ if and only if $A \sim B$, (2) *symmetry*: $d(A, B) = d(B, A)$, and (3) *triangle inequality* $d(A, B) + d(B, C) \geq d(A, C)$. Clustering should always use a metric, because if the triangle axiom is ignored, outputs of k -means and DBSCAN on point clouds may not be trustworthy [24].

Though real objects are often symmetric, their noisy representations deviate from ideal symmetry. For example, a narrow triangle can degenerate to a straight line and evolve to a mirror image of the original triangle that has an opposite (sign of) orientation. This discontinuity challenge for finite and periodic point sets was formalized and extended to the geo-mapping problem [5, Problem 1.4.5], which aims to continuously parametrize spaces of real data objects under practical equivalences by complete invariants similar to geographic coordinates on Earth.

Many molecules, such as benzene C_6H_6 , are nearly flat but live and are simulated on a computer in $\mathbb{R}^3$, where their principal directions [1] from Principal Component Analysis are undefined or unstable [17]. Hence, it is important to reliably distinguish between nearly flat configurations of molecules by quickly computable invariants that continuously change under perturbations during computer simulations or phase transitions. The classical continuity in terms of ε, δ is too weak in the sense that all standard functions are continuous on the domains where they are defined.

For example, $f(x) = \frac{1}{x}$ is continuous for $x \neq 0$ because, for any $\varepsilon > 0$, there is $\delta = \frac{|x|}{2} \min\{1, \varepsilon|x|\}$ such that if $|x - y| < \delta$, then $|f(x) - f(y)| = \frac{|x - y|}{|xy|} \leq \frac{\delta}{|xy|} \leq \frac{2\delta}{x^2} \leq \varepsilon$. However, for small $x > 0$, the chosen delta δ is much smaller than ε , so $f(x) = \frac{1}{x}$ grows too fast close to 0. The Lipschitz continuity below is more practical by restricting the growth of a function via an explicit constant λ and a maximum amount ε of perturbation.

Definition 1.1 (Lipschitz continuity). A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is *Lipschitz continuous* if there is a *Lipschitz constant* $\lambda > 0$ such that $|f(x) - f(y)| \leq \lambda|x - y|$ for any $x, y \in \mathbb{R}^n$, where $|x - y|$ denotes the Euclidean distance.

Then $f(x) = \frac{1}{x}$ is not Lipschitz continuous because for any potential $\lambda > 0$, we can set $c = \max\{1, \lambda\}$, $x = \frac{1}{2c}$ and $y = \frac{1}{c}$ such that $|f(x) - f(y)| = c \geq 1 > \frac{\lambda}{2c} = \lambda|x - y|$.

Though the Lipschitz continuity makes sense for maps between any metric spaces, we consider only scalar functions on subsets of $\mathbb{R}^n$.

A simplex can be defined as a finite set of arbitrary elements whose every subset is also a simplex. We consider only geometric realizations of a simplex, called a geometric simplex on ordered points in $\mathbb{R}^n$.

Definition 1.2 (a geometric simplex T on any $n+1$ ordered points in $\mathbb{R}^n$).

(a) The *geometric simplex* T on any points $p_0, \dots, p_n \in \mathbb{R}^n$ is the subset $A = \left\{ \sum_{i=0}^n t_i p_i \mid t_i \in [0, 1], \sum_{i=0}^n t_i = 1 \right\} \subset \mathbb{R}^n$ with the ordered $p_0, \dots, p_n$.

(b) An *orientation* of a simplex T on ordered points $p_0, \dots, p_n \in \mathbb{R}^n$ is the sign of the determinant of the $n \times n$ matrix with the column vectors $\mathbf{p}_1 - \mathbf{p}_0, \dots, \mathbf{p}_n - \mathbf{p}_0$, and is denoted by $\text{sign}(T)$.

For $n = 1$, the simplex on any points $p_0, p_1 \in \mathbb{R}$ is the line segment connecting p_0 and p_1 . If points $p_0, \dots, p_n \in \mathbb{R}^n$ are *affinely independent*, i.e. there is no $(n - 1)$ -dimensional affine subspace of $\mathbb{R}^n$ containing all

$p_0, \dots, p_n$ and hence the simplex T , then T is n -dimensional. Definition 1.2 makes sense for any points, which may not be affinely independent.

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. When a simplex T goes through a degenerate configuration, an orientation of T can discontinuously change sign. This discontinuity is an obstacle to recognizing simplices (or more general molecules as clouds of atoms) that are nearly mirror-symmetric.

One attempt 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. Unfortunately, $\text{vol}(T)$ is not Lipschitz continuous for any $n \geq 2$.

Example 1.3 (the area of a triangle is not Lipschitz continuous). For any large $l > 0$ and real ε close to 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 points are not restricted to a fixed bounded region, a small change in their positions may lead to a large change in the area of a triangle, and also in $\mathbb{R}^n$.

Example 1.3 motivated the following problem, which will be solved in sections 2 and 3, with applications to affinities of molecules in section 4.

Problem 1.4 (Lipschitz continuous detection of degenerate simplices). Find a Lipschitz continuous real-valued function $f(T)$ for all geometric simplices $T \subset \mathbb{R}^n$ on $n + 1$ ordered points $p_0, \dots, p_n \in \mathbb{R}^n$ such that $f(T) = 0$ if and only if T is degenerate.

2 Lipschitz continuous strength of a simplex

We solve Problem 1.4 by introducing the strength of a geometric simplex in Definition 2.1 and proving its Lipschitz continuity in Theorem 2.4.

Definition 2.1 (the *strength* $\sigma(T)$ and *signed strength* $s(T)$ of a simplex $T \subset \mathbb{R}^n$). Let $T \subset \mathbb{R}^n$ be the simplex on any $n + 1$ ordered points $p_0, p_1, \dots, p_n$ in $\mathbb{R}^n$. The *half-perimeter* $p(T) = \frac{1}{2} \sum_{i \neq j} |p_i - p_j|$ is one half of the sum of all distances between the vertices of T . The *strength* of T is defined as $\sigma(T) = \frac{\text{vol}^2(T)}{p^{2n-1}(T)}$. The *signed strength* is $s(T) = \text{sign}(T)\sigma(T)$.

Example 2.2 (the strength of a line segment). For $n = 1$, the simplex T on any points $p_0, p_1 \in \mathbb{R}$ is the line segment with $\text{vol}(T) = |p_1 - p_0| = 2p(T)$, the strengths $\sigma(T) = \frac{\text{vol}^2(T)}{p(T)} = 2|p_1 - p_0|$, and $s(T) = 2(p_1 - p_0)$.

Example 2.3 (the strength of a triangle). **(a)** Let $T \subset \mathbb{R}^2$ be a triangle with sides a, b, c . Heron's formula for $\text{vol}(T) = \sqrt{p(p-a)(p-b)(p-c)}$, where $p = \frac{a+b+c}{2} = p(T)$ is the half-perimeter, gives the strength $\sigma(T) = \frac{\text{vol}^2(T)}{p^3(T)} = \frac{(p-a)(p-b)(p-c)}{p^2}$. The triangle $T(l, \varepsilon)$ in Example 1.3 has $p = \frac{l + \varepsilon + \sqrt{l^2 + \varepsilon^2}}{2}$. Using $l \leq \sqrt{l^2 + \varepsilon^2} \leq l + \varepsilon$, we can estimate the strength of $T(l, \varepsilon)$ as follows: $\sigma =$

$$\frac{(l - \varepsilon + \sqrt{l^2 + \varepsilon^2})(\varepsilon + \sqrt{l^2 + \varepsilon^2} - l)(l + \varepsilon - \sqrt{l^2 + \varepsilon^2})}{2(l + \varepsilon + \sqrt{l^2 + \varepsilon^2})^2} \leq \frac{2l \cdot 2\varepsilon \cdot \varepsilon}{2(2l)^2} \leq \frac{\varepsilon}{2}$$

for any $0 \leq \varepsilon \leq l$. Hence, the strength of $T(l, \varepsilon)$ is Lipschitz continuous.

(b) To visualize the strength $\sigma(T)$ of a triangle T , we assume that $0 < a \leq b \leq c$ and normalize T by the larger side c to get the sides $\tilde{a} = \frac{a}{c} \leq \tilde{b} = \frac{b}{c} \leq \tilde{c} = 1$. The resulting space of normalized triangles is parameterized by the coordinates $x = \frac{a}{c}$ and $y = 1 - \frac{b}{c}$ in the triangular region

$$\Delta = \{(x, y) \in \mathbb{R}^2 \mid x \in [0, 1], x \geq y, x + y \leq 1\},$$

where $x + y \leq 1$ means that $a \leq b$, while $x \geq y$ is equivalent to the triangle inequality $a + b \geq c$, see Fig. 1 (left). The half-perimeter is

$$\tilde{p} = \frac{1}{2}(\tilde{a} + \tilde{b} + \tilde{c}) = \frac{1}{2}(x + (1 - y) + 1) = \frac{1}{2}(2 + x - y).$$

Then $\tilde{p} - \tilde{a} = \frac{1}{2}(2 - x - y)$, $\tilde{p} - \tilde{b} = \frac{1}{2}(x + y)$, $\tilde{p} - \tilde{c} = \frac{1}{2}(x - y)$, and $\sigma(x, y) = \frac{(2 - x - y)(x^2 - y^2)}{2(2 + x - y)^2}$, see Fig. 1 (right).

In the triangular region Δ in Fig. 1 (left), the horizontal side $\{x \in (0, 1], y = 0\}$ represents all (normalized) isosceles triangles with $\tilde{a} \leq \tilde{b} = \tilde{c} = 1$ and strength $\sigma = \frac{(2 - x)x^2}{2(2 + x)^2}$ for $x \in [0, 1]$.

The right hand side $\{x \in (0.5, 1], x + y = 1\}$ of the region Δ represents all (normalized) isosceles triangles with $\tilde{a} = \tilde{b} \leq \tilde{c} = 1$ and strength $\sigma = \frac{2x - 1}{2(2x + 1)^2}$ for $x \in [\frac{1}{2}, 1]$. The vertex $(x, y) = (1, 0)$ represents all equilateral triangles with side 1 of strength $\sigma = \frac{1}{18}$. Any equilateral triangle T with sides $a = b = c$ has the strength $\sigma(T) = \frac{(a^2\sqrt{3}/4)^2}{(3a/2)^3} = \frac{a}{18}$.

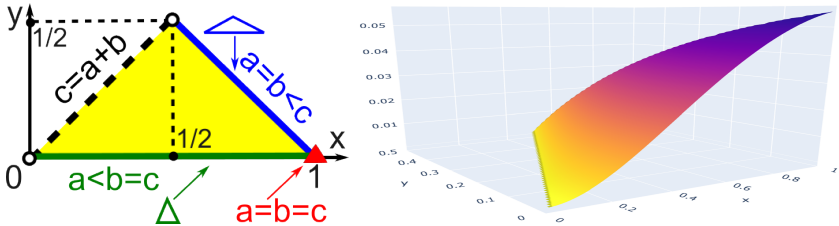


Figure 1. **Left:** Example 2.3(b) parameterizes the space of (normalized) triangles with sides $0 < a \leq b \leq c$ by $x = \frac{a}{c}$ and $y = 1 - \frac{b}{c}$ over the region $\Delta = \{(x, y) \in \mathbb{R}^2 \mid x \in [0, 1], x \geq y, x + y \leq 1\}$. **Right:** the strength $\sigma(x, y) = \frac{(2 - x - y)(x^2 - y^2)}{2(2 + x - y)^2}$ of a normalized triangle over Δ .

The strength of a simplex was used to define a Lipschitz continuous metric on invariants of n -dimensional clouds of m unordered points, which are complete under rigid motion in $\mathbb{R}^n$ and can be computed in polynomial

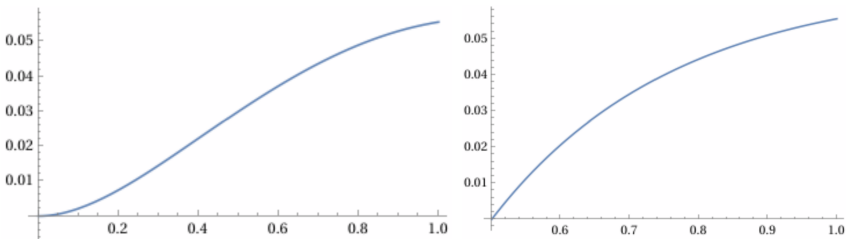


Figure 2. Vertical sections of the strength σ of normalized triangles parameterized by x, y in Fig. 1. **Left:** isosceles triangles with $\tilde{a} = \tilde{b} \leq \tilde{c} = 1$ have $y = 0$ and $\sigma = \frac{(2-x)x^2}{2(2+x)^2}$ for $x \in [0, 1]$. **Right:** isosceles triangles with sides $\tilde{a} \leq \tilde{b} = \tilde{c}$ have $y = 1 - x$ and $\sigma = \frac{2x-1}{2(2x+1)^2}$ for $x \in [\frac{1}{2}, 1]$.

time of m , for a fixed dimension n ; see [28, Theorem 4.7] and [18, Theorem 5.8]. Theorem 2.4(c) proves the Lipschitz continuity of σ for the first time, and with better constants than announced in [28, Theorem 4.4].

All time complexities are considered in the RAM model of computation so that any value in memory can be accessed in constant time.

Recall that any permutation $\xi \in S_n$ has $\text{sign}(\xi) = (-1)^t$, where t is the number of pairs of $i, j \in \{1, \dots, n\}$ such that $i < j$ but $\xi(i) > \xi(j)$.

Theorem 2.4 (strength properties). **(a)** If a permutation $\xi \in S_{n+1}$ acts on $n+1$ ordered vertices of a geometric simplex $T \subset \mathbb{R}^n$, then $\sigma(T)$ remains unchanged, but $\text{sign}(T)$ and $s(T)$ are multiplied by $\text{sign}(\xi)$.

(b) The strength $\sigma(T)$ and signed strength $s(T)$ 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 both strengths $\sigma(T)$ and $s(T)$ by c .

(c) Fix any dimension $n \geq 1$. Then there is a constant $\lambda_n > 0$ such that, for any $\varepsilon > 0$, if a simplex Q is obtained from another 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 any $n \geq 3$, e.g. $\lambda_3 < 0.37$ and $\lambda_n \rightarrow 0$ as $n \rightarrow +\infty$.

3 Proofs of strength properties for all $n \geq 1$

It was highly non-trivial to design the strength $\sigma = \frac{\text{vol}^2}{p^3}$ in Definition 2.1 after rejecting numerous conjectures. For example, the attempt to choose a Lipschitz continuous function $f(T)$ as the shortest height of the simplex $T \subset \mathbb{R}^n$ likely fails because this height is proportional to a ratio of the volume of T and the largest volume of an $(n-1)$ -dimensional subsimplex of T . Here both volumes are non-Lipschitz for $n > 2$ by Example 1.3.

Proof of Theorem 2.4(a). Any permutation ξ of vertices of a geometric simplex $T \subset \mathbb{R}^n$ preserves the volume $\text{vol}(T)$ and the half-perimeter $p(T)$, and hence the strength $\sigma(T) = \frac{\text{vol}^2(T)}{p(T)}$. Recall that $\text{sign}(T)$ is the sign of the determinant of $n \times n$ matrix $M(p_0, \dots, p_n)$ with the column vectors $\mathbf{p}_1 - \mathbf{p}_0, \dots, \mathbf{p}_n - \mathbf{p}_0$ for the ordered vertices $p_0, \dots, p_n$ of the simplex T .

If a permutation ξ keeps the vertex p_0 of T fixed, then ξ induces the permutation of the n columns of $M(p_0, \dots, p_n)$ whose sign is $\text{sign}(\xi)$, so $\text{sign}(T)$ is multiplied by $\text{sign}(\xi)$. Any permutation ξ that moves p_0 to (say) p_1 is a composition of the transposition $p_0 \leftrightarrow p_1$ and a permutation fixing p_0 . Since a composition $\xi \circ \xi'$ of permutations has $\text{sign}(\xi \circ \xi') = \text{sign}(\xi)\text{sign}(\xi')$, it remains to prove that $\text{sign}(T)$ becomes opposite under the transposition $p_0 \leftrightarrow p_1$. The matrix $M(p_1, p_0, p_2, \dots, p_n)$ is obtained from $M(p_0, p_1, p_2, \dots, p_n)$ by changing the sign of the first column, which changes the sign of the determinant, and by subtracting the first column from every other column, which preserves the determinant. ■

We will prove Theorem 2.4(b,c) using the Cayley-Menger determinant.

Definition 3.1 (Cayley-Menger determinant, [7, Chapter II, p.98]). Let the simplex T on any ordered points $p_0, \dots, p_n \in \mathbb{R}^n$ have the $(n+1) \times (n+1)$ matrix D_{ij} of squared Euclidean distances $d_{ij}^2 = |p_i - p_j|^2$ for $i, j = 0, \dots, n$. The $(n+2) \times (n+2)$ matrix $\hat{D}$ is obtained from D by adding the top row $(0, 1, \dots, 1)$ and the left column $(0, 1, \dots, 1)^T$.

The *Cayley-Menger determinant* [10, 21] expresses the squared volume of the simplex T as $\text{vol}^2(T) = \frac{(-1)^{n-1}}{2^n(n!)^2} \det \hat{D}$.

Proof of Theorem 2.4(b). Any isometry preserves all distances and hence the strength $\sigma(T)$ expressed via distances in Definition 2.1. Any special orthogonal matrix $M \in \text{SO}(\mathbb{R}^n)$ keeps the sign of a simplex T on $n + 1$ ordered vertices, i.e. multiplies $\text{sign}(T)$ by $\det M = 1$. Then any rigid motion preserves the signed strength $s(T) = \text{sign}(T)\sigma(T)$.

To compute $\sigma(T)$ and $s(T)$, we need the half-perimeter $p(T)$, which requires a quadratic time $O(n^2)$, and then $\text{sign}(T)$ and $\text{vol}^2(T)$ by using determinants of sizes up to $n + 2$, which can be calculated in time $O(n^3)$ by matrix diagonalization [23, section 11.5]. ■

We prove Theorem 2.4(c) first for $n = 1, 2$ and then for $n \geq 3$.

Proof of Theorem 2.4(c) for $n = 1$ and $\lambda_1 = 2$. For $n = 1$, a simplex $T \subset \mathbb{R}$ has two vertices $p_0, p_1 \in \mathbb{R}$ at the distance $d = |p_1 - p_0|$. Then the strength $\sigma(T) = 2d$ and $s(T) = 2(p_1 - p_0)$ have the Lipschitz constant $2\lambda_1 = 4$. Indeed, perturbing each of p_0, p_1 up to ε changes the difference $p_1 - p_0$ and the distance d up to 2ε . ■

Vertical lines denote the determinant $|M|$ of a matrix M , the Euclidean length $|v|$ of a vector $v \in \mathbb{R}^n$, and the absolute value $|r|$ of a real $r \in \mathbb{R}$.

Proof of Lipschitz continuity in Theorem 2.4(c) for $n = 2$. Let a triangle T have pairwise distances a, b, c . Using the half-perimeter $p = \frac{a+b+c}{2}$, we change the variables to $\tilde{a} = p - a$, $\tilde{b} = p - b$, $\tilde{c} = p - c$.

Then $a = \tilde{b} + \tilde{c}$, $b = \tilde{a} + \tilde{c}$, $c = \tilde{a} + \tilde{b}$, $p = \tilde{a} + \tilde{b} + \tilde{c}$, and the Jacobian of this change of variables is $\left| \frac{\partial(a, b, c)}{\partial(\tilde{a}, \tilde{b}, \tilde{c})} \right| = \begin{vmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{vmatrix} = 2$. If every point of T is perturbed up to ε , then any pairwise distance between the vertices of T changes by at most 2ε . By the mean value theorem [13], this bound 2ε gives $|\sigma(T) - \sigma(Q)| \leq 2\varepsilon \sup |\nabla \sigma|$, where $\nabla \sigma = \left(\frac{\partial \sigma}{\partial a}, \frac{\partial \sigma}{\partial b}, \frac{\partial \sigma}{\partial c} \right)$ is the gradient of the first order partial derivatives of $\sigma(T)$ with respect to the three distances between points of T . Since

$$|\nabla \sigma| = \left| \frac{\partial(a, b, c)}{\partial(\tilde{a}, \tilde{b}, \tilde{c})} \cdot \left(\frac{\partial \sigma}{\partial \tilde{a}}, \frac{\partial \sigma}{\partial \tilde{b}}, \frac{\partial \sigma}{\partial \tilde{c}} \right) \right| \leq 2 \left| \left(\frac{\partial \sigma}{\partial \tilde{a}}, \frac{\partial \sigma}{\partial \tilde{b}}, \frac{\partial \sigma}{\partial \tilde{c}} \right) \right|,$$

it remains to estimate the first order partial derivatives of the strength from Definition 2.1 $\sigma = \frac{\tilde{a}\tilde{b}\tilde{c}}{(\tilde{a} + \tilde{b} + \tilde{c})^2}$ with respect to the variables $\tilde{a}, \tilde{b}, \tilde{c}$. Since σ is symmetric in $\tilde{a}, \tilde{b}, \tilde{c}$, it suffices to consider

$$\frac{\partial \sigma}{\partial \tilde{a}} = \frac{\tilde{b}\tilde{c}}{(\tilde{a} + \tilde{b} + \tilde{c})^2} - \frac{2\tilde{a}\tilde{b}\tilde{c}}{(\tilde{a} + \tilde{b} + \tilde{c})^3} = \frac{\tilde{b}\tilde{c}(\tilde{b} + \tilde{c} - \tilde{a})}{(\tilde{a} + \tilde{b} + \tilde{c})^3} =$$

$$\frac{(p-b)(p-c)(2a-p)}{p^3} = \left(1 - \frac{b}{p}\right) \left(1 - \frac{c}{p}\right) \left(2\frac{a}{p} - 1\right).$$

By the triangle inequalities for the sides a, b, c , we have $\max\{a, b, c\} \leq p = \frac{a+b+c}{2}$. Then $\frac{a}{p}, \frac{b}{p}, \frac{c}{p} \in (0, 1]$, and $1 - \frac{b}{p}, 1 - \frac{c}{p} \in [0, 1]$, and $2\frac{a}{p} - 1 \in (-1, 1]$, so $\left|\frac{\partial \sigma}{\partial \tilde{a}}\right| \leq 1$. The similar upper bounds for the other two partial derivatives $\left|\frac{\partial \sigma}{\partial \tilde{b}}\right|, \left|\frac{\partial \sigma}{\partial \tilde{c}}\right| \leq 1$ imply that $|\nabla \sigma| \leq \sqrt{1^2 + 1^2 + 1^2} = \sqrt{3}$, so $|\sigma(T) - \sigma(Q)| \leq 2 \sup |\nabla \sigma| \varepsilon \leq 2\lambda_2 \varepsilon$ for $\lambda_2 = \sqrt{3}$. $\blacksquare$

Theorem 2.4(c) for $n \geq 3$ will be proved by Lemmas 3.2, 3.3, 3.4.

Lemma 3.2 (edge ratios). For any simplex T on points $p_0, \dots, p_n \in \mathbb{R}^n$, we have that $\frac{|p_i - p_j|}{p(T)} \leq \frac{2}{n}$ for all $i, j = 0, \dots, n$, where $p(T) = \frac{1}{2} \sum_{i,j=1}^{n+1} d_{ij}$.

Proof. Set $d_{ij} = |p_i - p_j|$ and use triangle inequalities as follows:

$$2p(T) = \sum_{k,l=0}^n d_{kl} \geq d_{ij} + \sum_{k \neq i,j} (d_{ki} + d_{kj}) \geq d_{ij} + (n-1)d_{ij} = nd_{ij}. \quad \blacksquare$$

Recall that the *rencontre* number $r_n = n! \sum_{k=0}^n \frac{(-1)^k}{k!}$ counts all permutations of $1, \dots, n$ without a fixed point [11] and equals the integer nearest to $\frac{n!}{e}$, e.g. $r_2 = 1, r_3 = 2, r_4 = 9, r_5 = 44$.

Lemma 3.3 (determinant ratio). In the notations of Definition 3.1, we have that $\frac{\det \hat{D}}{p^{2n}(T)} \leq r_{n+2} \left(\frac{2}{n}\right)^{2n}$ for any $n \geq 1$.

Proof. The determinant $\det \hat{D} = \sum_{\xi \in S_{n+2}} \text{sign}(\xi) \hat{D}_{1,\xi(1)} \dots \hat{D}_{n+2,\xi(n+2)}$ over all permutations $\xi \in S_{n+2}$ of $1, \dots, n+2$ excludes all zeros on the diagonal.

Then $k \neq \xi(k)$ for $k = 1, \dots, n+2$, i.e. we can consider only permutations ξ of $1, \dots, n+2$ that have no fixed elements. The number of such permutations is the rencontre number r_{n+2} . Then $\det \hat{D}$ is a sum of r_{n+2} non-zero terms, each being a product of n squared distances d_{ij}^2 .

The upper bound $d_{ij} \leq \frac{2}{n}p(T)$ in Lemma 3.2 implies that each term in $\det \hat{D}$ is at most $\left(\frac{2}{n}p(T)\right)^{2n}$. Then $\frac{\det \hat{D}}{p^{2n}(T)} \leq r_{n+2} \left(\frac{2}{n}\right)^{2n}$. ■

For $n = 1$ in Lemma 3.3, we have $\det \hat{D} = 2d_{01}^2$ and $p(T) = \frac{d_{01}}{2}$, so $\frac{\det \hat{D}}{p^2(T)} = 8 \leq r_3(2^2) = 8$.

Lemma 3.4. In the notations of Definition 3.1 for any distinct indices $i, j \in \{1, \dots, n\}$, we have $\left| \frac{\partial \det \hat{D}}{\partial d_{ij}} \right| \frac{1}{p^{2n-1}(T)} \leq 4(r_n + r_{n+1}) \left(\frac{2}{n}\right)^{2n-1}$.

Proof. Since $\det \hat{D}$ contains d_{ij}^2 in exactly two cells in different rows and columns, $\det \hat{D}$ is a quadratic polynomial $\alpha d_{ij}^4 + \beta d_{ij}^2 + \gamma$ for some α, β, γ that depend on other fixed distances $d_{kl} \neq d_{ij}$. Then $\frac{\partial \det \hat{D}}{\partial d_{ij}} = 4\alpha d_{ij}^3 + 2\beta d_{ij}$. The coefficient α is the determinant of the $n \times n$ submatrix (b_{ij}) obtained from $\hat{D}$ by removing two rows and columns indexed by $i+2, j+2$.

For example, fix $i = 0$ and $j = 1$. If $n = 2$, then $\alpha = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1$.

If $n = 3$, then $\alpha = \begin{vmatrix} 0 & 1 & 1 \\ 1 & 0 & d_{23}^2 \\ 1 & d_{32}^2 & 0 \end{vmatrix} = 2d_{23}^2$.

For any $n \geq 1$, since the matrix (b_{ij}) has zeros on the diagonal, its determinant $\alpha = \sum_{\xi \in S_n} \text{sign}(\xi) b_{1, \xi(1)} \dots b_{n, \xi(n)}$ is a sum over all permutations $\xi \in S_n$ with no fixed points. Then the sum α has r_n non-zero products $b_{1, \xi(1)} \dots b_{n, \xi(n)}$ and the total degree $2(n-2)$ in all distances $d_{kl} \neq d_{ij}$. After dividing the polynomial αd_{ij}^3 of the degree $2n-1$ by $p^{2n-1}(T)$, we use the upper bound $\frac{d_{kl}}{p(T)} \leq \frac{2}{n}$ in Lemma 3.2 to get $\frac{|\alpha| d_{ij}^3}{p^{2n-1}(T)} \leq r_n \left(\frac{2}{n}\right)^{2n-1}$.

The coefficient β in $\det \hat{D} = \alpha d_{ij}^4 + \beta d_{ij}^2 + \gamma$ is the sum of products from the determinants of the two $(n+1) \times (n+1)$ submatrices obtained from $\hat{D}$ by removing row $i+2$ and column $j+2$ (for one submatrix), then row $j+2$ and column $i+2$ (for another submatrix). For $n=3$, $i=0$, $j=2$, the

$$\text{determinants give the sum } \begin{vmatrix} 0 & 1 & 1 & 1 \\ 1 & d_{10}^2 & 0 & d_{13}^2 \\ 1 & d_{20}^2 & d_{21}^2 & d_{23}^2 \\ 1 & d_{30}^2 & d_{32}^2 & 0 \end{vmatrix} + \begin{vmatrix} 0 & 1 & 1 & 1 \\ 1 & d_{01}^2 & d_{02}^2 & d_{03}^2 \\ 1 & 0 & d_{12}^2 & d_{13}^2 \\ 1 & d_{31}^2 & d_{32}^2 & 0 \end{vmatrix}.$$

Since each $(n+1) \times (n+1)$ submatrix includes one entry d_{ij}^2 , we exclude all products with this entry, which are multiplied by the removed d_{ij}^2 from $\hat{D}$ and hence were counted in αd_{ij}^4 . Hence we can replace $d_{02} = d_{02}$ with

$$0 \text{ and get } \beta = \begin{vmatrix} 0 & 1 & 1 & 1 \\ 1 & d_{10}^2 & 0 & d_{13}^2 \\ 1 & 0 & d_{21}^2 & d_{23}^2 \\ 1 & d_{30}^2 & d_{32}^2 & 0 \end{vmatrix} + \begin{vmatrix} 0 & 1 & 1 & 1 \\ 1 & d_{01}^2 & 0 & d_{03}^2 \\ 1 & 0 & d_{12}^2 & d_{13}^2 \\ 1 & d_{31}^2 & d_{32}^2 & 0 \end{vmatrix}.$$

Up to a permutation of indices, each submatrix can be rewritten with the diagonal that has one original d_{ij}^2 (now replaced with 0), while all other diagonal elements are initial zeros. The example above gives $\beta =$

$$- \begin{vmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & d_{21}^2 & d_{23}^2 \\ 1 & d_{10}^2 & 0 & d_{13}^2 \\ 1 & d_{30}^2 & d_{32}^2 & 0 \end{vmatrix} - \begin{vmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & d_{12}^2 & d_{13}^2 \\ 1 & d_{01}^2 & 0 & d_{03}^2 \\ 1 & d_{31}^2 & d_{32}^2 & 0 \end{vmatrix}. \text{ Similarly to the argument}$$

for the determinant α , the sum βd_{ij} contains $2r_{n+1}$ products of the total degree $2n-1$, so $\frac{|\beta|d_{ij}}{p^{2n-1}(T)} \leq 2r_{n+1} \left(\frac{2}{n}\right)^{2n-1}$ by Lemma 3.2.

Then the required inequality follows:

$$\left| \frac{\partial \det \hat{D}}{\partial d_{ij}} \right| \frac{1}{p^{2n-1}(T)} = \frac{|4\alpha d_{ij}^3 + 2\beta d_{ij}|}{p^{2n-1}(T)} \leq (4r_n + 4r_{n+1}) \left(\frac{2}{n}\right)^{2n-1}. \quad \blacksquare$$

Proof of Theorem 2.4(c) for $n \geq 3$. For $n=3$, the squared volume is

$$\text{vol}^2(T) = \frac{1}{288} \begin{vmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & d_{01}^2 & d_{02}^2 & d_{03}^2 \\ 1 & d_{10}^2 & 0 & d_{12}^2 & d_{13}^2 \\ 1 & d_{20}^2 & d_{21}^2 & 0 & d_{23}^2 \\ 1 & d_{30}^2 & d_{31}^2 & d_{32}^2 & 0 \end{vmatrix}. \text{ Similarly to the case } n=2,$$

the mean value theorem [13] for the strength $\sigma(T) = \frac{\text{vol}^2(T)}{p^{2n-1}(T)}$, where $p(T) = \frac{1}{2} \sum_{i \neq j} d_{ij}$, implies that $|\sigma(T) - \sigma(Q)| \leq$

$$\leq 2\varepsilon \sup |\nabla \sigma| \leq 2\varepsilon \sqrt{\sum_{i \neq j} \sup \left| \frac{\partial \sigma}{\partial d_{ij}} \right|^2} \leq 2\varepsilon \sqrt{\frac{n(n+1)}{2}} \max_{i \neq j} \sup \left| \frac{\partial \sigma}{\partial d_{ij}} \right|.$$

To find an upper bound of $\left| \frac{\partial \sigma}{\partial d_{ij}} \right|$, we initially ignore the numerical factor in the squared volume $\text{vol}^2(T) = \frac{(-1)^{n-1}}{2^n(n!)^2} \det \hat{D}$ and differentiate $\det \hat{D} \cdot \frac{1}{p^{2n-1}(T)}$ by the product rule:

$$\frac{\partial}{\partial d_{ij}} \left(\frac{\det \hat{D}}{p^{2n-1}(T)} \right) = \frac{\partial \det \hat{D}}{\partial d_{ij}} \cdot \frac{1}{p^{2n-1}(T)} - \frac{\det \hat{D}}{p^{2n}(T)} \cdot \frac{2n-1}{2}.$$

Lemmas 3.3 and 3.4 imply the upper bound

$$\left| \frac{\partial}{\partial d_{ij}} \left(\frac{\det \hat{D}}{p^{2n-1}(T)} \right) \right| \leq (4r_n + 4r_{n+1}) \left(\frac{2}{n} \right)^{2n-1} + \left(n - \frac{1}{2} \right) r_{n+2} \left(\frac{2}{n} \right)^{2n} < \\ < (4r_n + 4r_{n+1} + 2r_{n+2}) \left(\frac{2}{n} \right)^{2n-1}. \text{ By taking into account the factors } \\ \sqrt{\frac{n(n+1)}{2}} \text{ and } \frac{(-1)^{n-1}}{2^n(n!)^2} \text{ in } \text{vol}^2(T) \text{ to estimate the length of the gradient } \\ \nabla \sigma \text{ of the } \frac{n(n+1)}{2} \text{ first order partial derivatives } \frac{\partial \sigma}{\partial d_{ij}}, \text{ we get} \\ |\sigma(T) - \sigma(Q)| \leq 2\varepsilon \sqrt{\frac{n(n+1)}{2}} \max_{i \neq j} \sup \left| \frac{\partial \sigma}{\partial d_{ij}} \right| \leq 2\varepsilon c_n \text{ for the upper bound}$$

$$c_n = \frac{\sqrt{n(n+1)/2}}{2^n(n!)^2} (4r_n + 4r_{n+1} + 2r_{n+2}) \left(\frac{2}{n} \right)^{2n-1} \\ = (2r_n + 2r_{n+1} + r_{n+2}) \frac{2^{n-0.5} \sqrt{n+1}}{(n!)^2 n^{2n-1.5}}.$$

The estimate $r_n \leq \frac{n!}{2}$ gives

$$c_n < \frac{n!}{2} \left(2 + 2(n+1) + (n+1)(n+2) \right) \frac{2^{n-0.5} \sqrt{n+1}}{(n!)^2 n^{2n-1.5}} = \\ (n^2 + 5n + 6) \sqrt{n+1} \frac{2^{n-1.5}}{(n!) n^{2n-1.5}} = \left(1 + \frac{5}{n} + \frac{6}{n^2} \right) \sqrt{1 + \frac{1}{n}} \frac{2^{n-1.5}}{(n!) n^{2n-4}}.$$

For $n \geq 3$, we get $1 + \frac{5}{n} + \frac{6}{n^2} \leq \frac{10}{3}$ and $\sqrt{1 + \frac{1}{n}} \leq \frac{2}{\sqrt{3}}$, so

$$c_n < \frac{20}{3\sqrt{3}} \cdot \frac{2^{n-1.5}}{(n!) n^{2n-4}} = \frac{5\sqrt{2}}{3\sqrt{3}} \frac{2^n}{(n!) n^{2n-4}} < \frac{2^{n+0.5}}{(n!) n^{2n-4}} = b_n. \quad \blacksquare$$

Proof of constants for the signed strength in Theorem 2.4(c). If T, Q have $\text{sign}(T) = \text{sign}(Q)$, then $|s(T) - s(Q)| = |\sigma(T) - \sigma(Q)| \leq 2\lambda_n \varepsilon$. If T, Q have opposite signs, we will prove that $|s(T) - s(Q)| \leq 4\lambda_n \varepsilon$. Any simplex T can be connected to its ε -perturbation Q by a straight-line deformation through simplices $Q_r \subset \mathbb{R}^n$, where $r \in [0, 1]$, $Q_0 = T$, and $Q_1 = Q$. There is an intermediate value $r \in (0, 1)$ such that $\sigma(Q_r) = 0 = \text{sign}(Q_r)$. Then both T, Q are deformations of the intermediate simplex Q_r such that all corresponding vertices are perturbed up to Euclidean distances $r\varepsilon$ and $(1-r)\varepsilon$, respectively. The inequalities $|s(T)| = |\sigma(T) - \sigma(Q_r)| \leq 2\lambda_n r\varepsilon$ and $|s(Q)| = |\sigma(Q_r) - \sigma(Q)| \leq 2\lambda_n (1-r)\varepsilon$ imply that $|s(T) - s(Q)| \leq |\sigma(T) - \sigma(Q_r)| + |\sigma(Q_r) - \sigma(Q)| \leq 2\lambda_n \varepsilon$. $\blacksquare$

Hence, the strength of a simplex provides a Lipschitz continuous analog of a volume, which is measured in the units of point coordinates.

4 Geometric affinities of molecules

This section introduces geometric affinities that continuously quantify the degeneracy of molecules, for example, how close a molecule of water H_2O is to a straight line (measured in Angstroms), and how flat (close to a 2-dimensional plane) molecules of benzene C_6H_6 and methane CH_4 are.

Recall that the *affine dimension* $\text{aff}(C)$ of a point cloud C is the minimum k such that C is contained in a k -dimensional affine subspace of $\mathbb{R}^n$.

If $\text{aff}(C) < n$, then the simplex on any $n + 1$ points of C is degenerate and has strength 0. Definition 4.1 extends the discrete invariant $\text{aff}(C)$ to continuous affinities based on the strength of simplices on points of C .

Definition 4.1 (affinities $\sigma(C; k)$ of a point cloud C). For any cloud $C \subset \mathbb{R}^n$ of m unordered points, let $O(C) = \frac{1}{m} \sum_{p \in C} p$ denote the *geometric center* of C . Fix an integer $1 \leq k \leq n$. For any points $p_1, \dots, p_k \in C$, let $T(p_1, \dots, p_k)$ denote the geometric simplex on $k+1$ points $O(C), p_1, \dots, p_k$. The k -th *Maximum Geometric Affinity* $\text{MGA}(C; k)$ and *Average Geometric Affinity* $\text{AGA}(C; k)$ are the maximum and average, respectively, of $\sigma(T(p_1, \dots, p_k))$ over all subsets of k unordered points $p_1, \dots, p_k \in C$.

Since $\sigma(T)$ is invariant under permutations of vertices of T by Theorem 2.4(a), $\sigma(T(p_1, \dots, p_k))$ is independent of an order of $p_1, \dots, p_k$.

For any point cloud $C \subset \mathbb{R}^n$, if $k = 1$, then $T(p_1)$ is the line segment $[O(C), p_1]$ whose the strength $\sigma(T(p_1)) = 2|O(C) - p_1|$ is the double radial distance from p_1 to the center $O(C)$. Then $\text{MGA}(C; 1)$ and $\text{AGA}(C; 1)$ are the maximum and average, respectively, radial distances of C .

In chemistry, the concept of strength is discussed in terms of inter-atomic bonds [22], while Definition 2.1 introduced the strength only for a geometric simplex. Geometric affinities in Definition 4.1 are motivated by the affine dimension of an atomic cloud and certainly differ from binding affinities of small-molecule ligands to macromolecular receptors [15].

Example 4.2 and Fig. 3 use inter-atomic distances from PubChem [8].

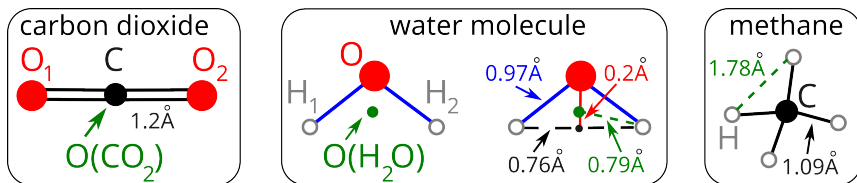


Figure 3. **Left:** a linear molecule of carbon dioxide CO_2 . **Middle:** a water molecule H_2O is an isosceles triangle. **Right:** a molecule of methane CH_4 is a regular tetrahedron with a carbon atom in the center. See details in Example 4.2.

Example 4.2 (geometric affinities of molecules CO_2 , H_2O , and CH_4). **(a)** The PubChem entry 280 of carbon dioxide CO_2 has 3 atoms in a straight line with a bond length $|CO| \approx 1.2\text{\AA}$; see Fig. 3 (left). Since the center of CO_2 coincides with atom C, three radial distances have the maximum $\text{MGA}(\text{CO}_2; 1) \approx 1.2\text{\AA}$ and the average $\text{AGA}(\text{CO}_2; 1) \approx \frac{0 + 2 \times 1.2}{3} = 0.8\text{\AA}$. All other affinities in Definition 4.1 for $k > 1$ vanish.

(b) The PubChem entry 962 of water H_2O has approximate radial distances 0.4, 0.79, 0.79 in Fig. 3 (middle). Then $\text{MGA}(\text{H}_2\text{O}; 1) \approx 0.79\text{\AA}$ and $\text{AGA}(\text{H}_2\text{O}; 1) \approx \frac{0.4 + 2 \times 0.79}{3} = 0.66\text{\AA}$. Three triangles with the vertices at the center $O(\text{H}_2\text{O})$ and two atoms of H_2O have the following normalized strengths $\bar{\sigma} = \frac{\text{vol}^2}{p^3 \lambda_2}$ with $\lambda_2 = \sqrt{3}$ in [18, Example 6.9(b)]: 0.01, 0.004, 0.004. Then we get $\text{MGA}(\text{H}_2\text{O}; 2) = 0.01\lambda_2 \approx 0.017\text{\AA}$, and $\text{AGA}(\text{H}_2\text{O}; 2) = \frac{0.01 + 2 \times 0.004}{3} \lambda_2 \approx 0.01\text{\AA}$. All other geometric affinities of H_2O for $k > 2$ are equal to zero.

(c) The PubChem entry 297 of methane CH_4 has atom C at the center $O(\text{CH}_4)$ and 4 atoms H at a radial distance of $1.09\text{\AA}$ in Fig. 3 (right). Then $\text{MGA}(\text{CH}_4; 1) \approx 1.09\text{\AA}$ and $\text{AGA}(\text{CH}_4; 1) \approx \frac{4 \times 1.09}{5} \approx 0.87\text{\AA}$.

For a fixed vertex at $C = O(\text{CH}_4)$, the 5-point cloud CH_4 has 4 degenerate triangles (of area 0) with two other vertices at C and H, and 6 non-degenerate triangles with two other vertices at 2 of 4 atoms H. Each non-degenerate triangle is isosceles with sides 1.09, 1.78, 1.78, half-perimeter $p = \frac{1.09 + 2 \times 1.78}{3} \approx 1.55$, $\text{vol} \approx 0.56$, and strength $\sigma_2 = \frac{\text{vol}^2}{p^3} \approx 0.08\text{\AA}$. Then $\text{MGA}(\text{CH}_4; 2) \approx 0.08\text{\AA}$ and $\text{AGA}(\text{CH}_4; 2) = \frac{6}{10} \sigma_2 \approx 0.05\text{\AA}$.

For a fixed vertex at C, molecule CH_4 has 6 degenerate tetrahedra (of volume 0) with three other vertices at C, H, H, and 4 non-degenerate tetrahedra with three other vertices at 3 of 4 atoms H. Each non-degenerate tetrahedron is a regular pyramid with 3 sides $b = 1.09$ and an equilateral base with 3 sides $a = 1.78$. This pyramid has the half-perimeter $p = \frac{a + b}{2} \approx 1.44$, volume $\frac{1}{3} \frac{a^2 \sqrt{3}}{4} \sqrt{b^2 - \frac{a^2}{3}} \approx 0.17$, and strength $\sigma_3 = \frac{\text{vol}^2}{p^3} \approx 0.01\text{\AA}$. Then $\text{MGA}(\text{CH}_4; 3) \approx 0.01\text{\AA}$ and $\text{AGA}(\text{CH}_4; 3) = \frac{4}{10} \sigma_3 \approx 0.004\text{\AA}$.

We could define affinities by using simplices on $k + 1$ points of C , but computations are faster if we use k points in addition to the center $O(C)$.

Theorem 4.3 (affinity properties). Let $C \subset \mathbb{R}^n$ consist of m points.

(a) The affine dimension $\text{aff}(C)$ is less than n if and only if $\text{MGA}(C; k) = 0$ for $k = \text{aff}(C) + 1$ or, equivalently, $\text{AGA}(C; k) = 0$.

(b) For any $1 \leq k \leq n$, both $\text{MGA}(C; k)$ and $\text{AGA}(C; k)$ are invariant under any isometry of $\mathbb{R}^n$ and can be computed in time $O(n^3 m^k)$.

(c) Let λ_n be a Lipschitz constant from Theorem 2.4. For any $\varepsilon > 0$, if a cloud $B \subset \mathbb{R}^n$ is obtained from C by perturbing every point of C within its ε -neighborhood, then $|\text{MGA}(B; k) - \text{MGA}(C; k)| \leq 2\lambda_n \varepsilon$ and $|\text{AGA}(B; k) - \text{AGA}(C; k)| \leq 2\lambda_n \varepsilon$.

Proof. (a) Let $k = \text{aff}(C) + 1 \leq n$. Then, for any points $p_1, \dots, p_k \in C$, the simplex $T(p_1, \dots, p_k)$ is contained in a $(k - 1)$ -dimensional affine subspace of $\mathbb{R}^n$ and hence has k -dimensional volume 0. Then $\sigma(T(p_1, \dots, p_k)) = 0$ and $\text{MGA}(C; k) = 0 = \text{AGA}(C; k)$ by Definition 4.1.

Conversely, since the strength of any geometric simplex is non-negative, each of $\text{MGA}(C; k) = 0$, $\text{AGA}(C; k) = 0$ implies that $\sigma(T(p_1, \dots, p_k)) = 0$ for any points $p_1, \dots, p_k \in C$. Then the simplex $T(p_1, \dots, p_k)$ has k -dimensional volume 0 and is contained in a $(k - 1)$ -dimensional affine subspace $H \subset \mathbb{R}^n$. This subspace H is uniquely defined by the center $O(C)$ and $k - 1$ fixed points $p_1, \dots, p_{k-1}$. By taking any other point $p_k \in C \setminus \{p_1, \dots, p_{k-1}\}$, we conclude that the whole cloud C is within H .

(b) For any points $p_1, \dots, p_k \in C$, by Theorem 2.4(b) any isometry of $\mathbb{R}^n$ preserves the strength $\sigma(T(p_1, \dots, p_k))$, which can be computed in time $O(n^3)$. Hence, any isometry preserves their maximum $\text{MGA}(C; k)$ and average $\text{AGA}(C; k)$ over all $\binom{m}{k} = O(m^k)$ subsets of k unordered points $p_1, \dots, p_k \in C$. Then the total computational time is $O(n^3 m^k)$.

(c) For any points $p_1, \dots, p_k \in C$, which are perturbed within their ε -neighborhoods, the strength $\sigma(T(p_1, \dots, p_k))$ changes by at most $2\lambda_n \varepsilon$ by Theorem 2.4(c). Hence the maximum $\text{MGA}(C; k)$ and the average $\text{AGA}(C; k)$ of all these strengths also change by at most $2\lambda_n \varepsilon$. ■

Theorem 4.3(c) allows us to physically interpret geometric affinities of a molecule with a cloud C of atoms as follows. For a flat molecule with $C \subset \mathbb{R}^2$, such as H_2O , if $\text{MGA}(C; 2) > 0$, then we need to shift at least one atom of C by Euclidean distance $\frac{\text{MGA}(C; 2)}{2\lambda_2}$ for $\lambda_3 = \sqrt{3}$ to make C linear. If $\text{MGA}(C; 3) > 0$, then we need to shift at least one atom of C by Euclidean distance $\frac{\text{MGA}(C; 3)}{2\lambda_3}$ for $\lambda_3 \approx 0.37$ to make C flat.

Here we used very rough upper bounds of Lipschitz constants λ_n from Theorem 2.4(c). Experimental bounds should be much smaller and hence improve the estimates of perturbations above. For any fixed bounds of λ_n , geometric affinities can be used for relative comparisons of molecules.

The recent classification of periodic point sets under rigid motion relied on round clusters of atoms in [2, 6] with metrics in [18] based on strengths of simplices. We plan to adapt the signed strength to quantify chirality of periodic crystals, whose deviations from higher symmetry were continuously measured in [19] only by isometry invariants from [26, 27, 29].

Acknowledgment: This research was funded by the Royal Society Short Industry Fellowship “Ultra-fast comparisons of crystals for large materials databases” (SIF/R2/252048) of the last author.

References

- [1] H. Abdi, L. J. Williams, Principal component analysis, *Wiley Interdis. Rev. Comput. Stat.* **2** (2010) 433–459.
- [2] O. Anosova, V. Kurlin, An isometry classification of periodic point sets, in: J. Lindblad, F. Malmberg, N. Sladoje (Eds.), *Discrete Geometry and Mathematical Morphology*, Springer, Cham, 2021, pp. 229–241.
- [3] O. Anosova, V. Kurlin, M. Senechal, The importance of definitions in crystallography, *Int. Union Cryst. J.* **11** (2024) 453–463.
- [4] O. Anosova, A. Gorelov, W. Jeffcott, Z. Jiang, V. Kurlin, A complete and bi-continuous invariant of protein backbones under rigid motion, *MATCH Commun. Math. Comput. Chem.* **94** (2025) 97–134.

-
- [5] O. Anosova, V. Kurlin, *Geometric Data Science*, arXiv:2512.05040, the latest version at <https://kurlin.org/Geometric-Data-Science-book.pdf>.
- [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] L. M. Blumenthal, *Theory and Applications of Distance Geometry*, Chelsea Pub. Co, 1970.
- [8] E. Bolton, J. Chen, S. Kim, L. Han, S. He, W. Shi, V. Simonyan, Y. Sun, P. A. Thiessen, J. Wang, B. Yu, J. Zhang, S. H. Bryant, Pubchem3d: a new resource for scientists, *J. Cheminf.* **3** (2011) #32.
- [9] D. Bonchev, D. H. Rouvray, *Chemical Graph Theory – Introduction and Fundamentals*, CRC Press, Boca Raton, 1991.
- [10] A. Cayley, On a theorem in the geometry of position, *Cambridge Math. J.* **2** (1841) 267–271.
- [11] C. A. Charalambides, *Enumerative Combinatorics*, CRC Press, Boca Raton, 2018.
- [12] S. El-Basil, M. Randić, Equivalence of mathematical objects of interest in chemistry and physics, *Adv. Quantum Chem.* **24** (1992) 239–290.
- [13] P. Elliott, *Probabilistic Number Theory I: Mean-Value Theorems*, Springer, New York, 1979.
- [14] R. Feynman, *The Feynman Lectures on Physics*, Basic Books, New York, 2011.
- [15] H. Gohlke, G. Klebe, Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors, *Angew. Chem. Int. Ed.* **41** (2002) 2644–2676.
- [16] V. Kurlin, Simplexwise Distance Distributions for finite spaces with metrics and measures, arXiv:2303.14161, the latest version at <http://kurlin.org/projects/cloud-isometry-spaces/SDD.pdf>.
- [17] V. Kurlin, Polynomial-time algorithms for continuous metrics on atomic clouds of unordered points, *MATCH Commun. Math. Comput. Chem* **91** (2024) 79–108.

-
- [18] V. Kurlin, Complete invariants of atomic clouds under rigid motion with Lipschitz continuous metrics in a polynomial time, *MATCH Commun. Math. Comput. Chem* **97** (2027) 379–428.
- [19] S. Majumder, Y. Elkin, D. Widdowson, O. Anosova, A. I. Cooper, G. M. Day, V. Kurlin, Continuous invariant-based asymmetries of periodic crystals quantify deviations from higher symmetry, *Int. Union Cryst. J.* **13** (2026) 456–470.
- [20] F. Mémoli, G. Sapiro, A theoretical and computational framework for isometry invariant recognition of point cloud data, *Found. Comput. Math.* **5** (2005) 313–347.
- [21] K. Menger, New foundation of Euclidean geometry, *Am. J. Math.* **53** (1931) 721–745.
- [22] P. Metola, E. V. Anslyn, What is strength?, *Supramol. Chem.* **37** (2026) 141–146.
- [23] W. Press, S. Teukolsky, W. Vetterling, B. Flannery, *Numerical Recipes: the Art of Scientific Computing*, Cambridge University Press, Cambridge, 2007.
- [24] S. Rass, S. König, S. Ahmad, M. Goman, Metricizing the euclidean space towards desired distance relations in point clouds, *IEEE Trans. Inf. Forens. Sec.* **19** (2024) 7304–7319.
- [25] I. Schoenberg, Remarks to Maurice Frechet’s article “Sur la definition axiomatique d’une classe d’espace distances vectoriellement applicable sur l’espace de Hilbert, *Ann. Math.* (1935) 724–732.
- [26] 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.
- [27] D. Widdowson, V. Kurlin, Resolving the data ambiguity for periodic crystals, *Adv. Neural Inf. Process. Sys.* **35** (2022) 24625–24638.
- [28] 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: *Computer Vision and Pattern Recognition*, 2023, pp. 1275–1284.
- [29] D. Widdowson, V. Kurlin, Pointwise distance distributions for detecting near-duplicates in large materials databases, *SIAM J. Appl. Math.* **86** (2026) 898–918.

- [30] A. Wlodawer, Z. Dauter, P. Rubach, W. Minor, M. Jaskolski, Z. Jiang, W. Jeffcott, O. Anosova, V. Kurlin, Duplicate entries in the protein data bank: how to detect and handle them, *Acta Cryst. D* **81** (2025) 170–180.