arXiv:2303.14161v1 [math.MG] 24 Mar 2023

带度量和测度的有限空间的单形距离分布

Vitaliy Kurlin 利物浦大
学计算机科学系，英国
vitaliy.kurlin@gmail.com

摘要

欧几里得空间中的有限无标记点集，是许多真实对象最简单的表示形式，从矿物岩石到雕塑皆是如此。由于大多数实体对象都是刚性的，它们的自然等价关系是保持所有点间距离的刚性运动或等距变换。更一般地，任何有限度量空间都是度量测度空间的一个例子，它具有一个概率测度和一个满足所有公理的度量。

本文为任意有限度量空间和度量测度空间构建了单形距离分布（*SDD*）。这些 *SDD* 能够区分所有已知的非等价空间，而这些空间此前无法通过更简单的不变量加以区分。我们定义了 *SDD* 上的度量，这些度量具有 *Lipschitz* 连续性，并支持精确计算；其参数化复杂度关于给定点的数量呈多项式增长。

1. 对度量空间进行分类的动机

任何刚性物体（例如汽车或雕塑）最简单的表示方式是一个由 $S \subset \mathbb{R}^n$ 个 m 无标记点构成的有限集（点云），其中 $n = 2$，3 是最实用的维数。

许多固体对象具有刚性，这促使我们研究它们在刚性运动下的性质。刚性运动是欧几里得空间中的平移和旋转的复合 $\mathbb{R}^n$。我们可以考虑任意带有度量的有限集 X，该度量是一个距离函数d_X： $X \times X \to [0, +\infty)$，满足所有度量公理。度量空间上的自然等价关系是等距变换，即保持所有点间距离的任意映射 f： $X \to Y$，使得 $d_X(p,\ q) = d_Y(f(p)\ f(q))$，其中 $p,\ q \in X$。

在 $\mathbb{R}^n$ 中，任意等距变换都是某个镜像反射与刚性运动的复合。任何保向等距变换都可以实现为连续的刚性运动。

刚性物体的形状 在数学上定义为其等距类。任何非刚性变形都会定义一种比等距关系更弱的等价关系，从而得到一个包含更少可变形状的空间。要在等距意义下比较形状，需要更精细的不变量来区分更多等距类。

在数学上，要区分等距空间，需要使用不变量，即在任何等距变换下都保持不变的性质。任何不变量 I 都会将所有等距空间映射为相同的值，因此不会出现这样的假阴性： 两个等距空间 $S \cong Q$满足 $I(S) \neq I(Q)$。

一个完备的不变量 I 应能区分所有非等距点云，因此，若 $S \ncong Q$，则 $I(S) \neq I(Q)$。换言之，I不会出现假阳性，即不会有非等距空间点对 $S \ncong Q$ 满足 $I(S) = I(Q)$。那么，I就是一种 DNA式编码，可以在等距意义下唯一标识任意空间 S。

由于真实数据总是带有噪声，有用的完备不变量还必须在点的移动下保持连续。对于由 m 无标记点构成的集合，要同时满足完备性和连续性极具挑战性，因为存在 $m!$ 可能的置换，能够匹配所有 m 点。

一个适用于 $m = 3$ 个点的完备且连续的不变量由三个两两距离构成，分别是 a、b、c（三角形的边），在学校里称为边边边定理 [68]，用于判断三角形的全等（等距）。因此，三点集的等距空间可连续映射为一个四边锥 $\{0 < a \leq b \leq c \leq a + b\}$，其参数为 $a \leq b \leq c$，并满足一个三角不等式 $a + b \geq c$。

上述完整描述在 $m \geq 4$ 个点的 $\mathbb{R}^n$中没有简单的对应物。一个障碍是$\mathbb{R}^2$ 中的一族四点集，它们具有相同的六个两两距离，见图1。

问题 1.1（具有可计算连续度量的完备等距不变量）。设计一个针对有限度量空间的不变量 I，满足以下性质：

(a) completeness *: S, Q are isometric $\Leftrightarrow I(S) = I(Q)$;*

（***b***）利普希茨连续性：若对 C 中任意一点进行扰动，且扰动幅度不超过其 ε 邻域范围，则 I（S）的变化至多为 $\lambda\varepsilon$，其中常数为 λ，度量 d 满足所有公理：

(1) d（I（S） I（Q））$= 0$ 当且仅当 $S \cong Q$ 等距，

(2) 对称性：d（I（S） I（Q））$= d$（I（Q） I（S））,

(3) $d(I(S), I(Q)) + d(I(Q), I(T)) \geq d(I(S), I(T))$*;*

(c) 可计算性*：* $I(S)$ 和 d 可在关于给定空间中点数 m 的多项式时间内计算。

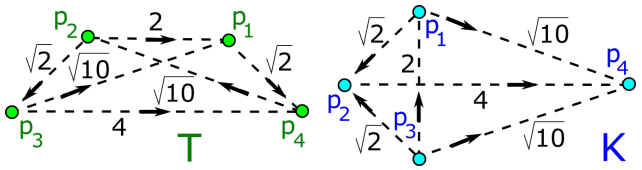

图 1。**左**：梯形 $T = \{(1, 1), (-1, 1), (-2, 0), (2, 0)\}$。**右**：风筝形 $K = \{(0, 1), (-1, 0), (0, -1), (3, 0)\}$。两者 T和 K 具有相同的六个两两距离 $\sqrt{2}$，$\sqrt{2}$，2，$\sqrt{10}$，$\sqrt{10}$，4。

条件（1.1b）要求连续度量，这比（1.1a）中的完备性要求更强。检测等距变换 $S \cong Q$ 会得到一个不连续度量，例如 $d = 1$对于所有非等距的 $S \ncong Q$，即使 S，Q 也几乎完全相同。任何 d 满足（1.1b）中第一公理的度量，都可以通过检查 $S \cong Q$ 来检测等距变换 $d(I(S)\ I(Q)) = 0$。

问题 1.1至少自 1974 年起就是一个未解决的问题，当时 Gilbert 和 Shepp [31] 描述了实数集 $\mathbb{R}^2$ 中的一个四参数族四点集，这些点集具有相同的六个两两距离。问题 1.1 也受到 [26, 27, 62] 拓扑数据分析中持久同调的局限性的启发。

第 2 节回顾了与点集不变量及更一般度量空间密切相关的研究。第 3 节介绍单形距离分布（SDD），它对以往所有基于距离的有限度量空间不变量作了实质性推广。第 4 节表明，SDD足够简单，便于手工计算，也能对无法通过较简单的距离分布加以区分的点云无穷族进行分类。第 5 节发展了定义在 SDD上的 Lipschitz 连续度量，其计算时间关于点数为参数化多项式时间 m 。

我们考虑问题 1.1 ，这是理解任意数据对象的模空间的重要第一步。度量空间和等距变换分别可以由其他数据和等价关系取代，从而得到问题 1.1的类似版本。本文扩展了 [72], 的第 3 节；该文不含证明和大型示例的 8 页版本将很快发表。在论文 [70–72] 中，第一作者实现了所有算法，第二作者设计了全部理论、证明和示例。

2. 关于等距变换和度量空间的既有工作

本节回顾相关工作，从问题 1.1 的一个简化版本说起，该版本只要求检测点云之间是否可能存在等距变换，其中点为 m 无标记点

等距检测指的是问题1.1 的一种更简单的形式，旨在通过算法检测给定点云之间是否存在潜在的等距变换 m 中的点 $\mathbb{R}^n$。Brass 和 Knauer 提出的最佳算法 [10] 需要 $O(m^{\lceil n/3 \rceil} \log m)$时间，因此在 $O(m\log\ m)$ 中 $\mathbb{R}^3$ [11]。最新进展是 $O(m\log\ m)$ 算法，见 $\mathbb{R}^4$ [41]。这些算法只输出二元答案（是/否），无法通过连续度量量化非等距点云之间的相似性。

多维尺度分析 (MDS)。对于任意 $m \times m$ 点云 A的给定 m 距离矩阵，多维尺度分析 [60] 会寻找一个嵌入 $A \subset \mathbb{R}^k$（如果存在），使其在维数 $k \leq m$下保留M 的所有距离。最终的嵌入 $A \subset \mathbb{R}^k$使用特征向量，而其符号上的不确定性会导致指数级的比较时间，可能接近 $O(2^m)$。

热核特征 (HKS) 是流形的一种完备等距变换不变量 M，其拉普拉斯-贝尔特拉米算子具有互异特征值；这一结论见 [65, 定理 1]。如果 M由点采样得到，热核特征可以离散化并保持连续性 [65,节 4]，但其完备性尚不明确。

豪斯多夫距离 [37]可以为环境度量空间中的任意子集 A, B 定义为 $d_H(A, B)$ $=\max\{\vec{d}_H(A, B)\ \vec{d}_H(B, A)\}$，其中有向豪斯多夫距离为 $\vec{d}_H(A, B)$ $= \sup_{p\in A} \inf_{q\in B} |p - q|$。为了将.....纳入考
虑，等距变换可以通过对所有等距变换 [17, 19, 39]取豪斯多夫距离的最小值来实现。对于 $n = 2$，对于至多含有m 个点的集合，在 $\mathbb{R}^2$ 中对等距变换取最小值的豪斯多夫距离问题需要 $O(m^5 \log m)$时间 [18]。给定 $\varepsilon > 0$和 $n > 2$，判定是否 $d_H \leq \varepsilon$ 在平移意义下成立的相关问题，其时间复杂度为 $O(m^{\lceil (n+1)/2 \rceil})$ [69,第4章，推论 6]。对于一般的等距变换，最初只有近似算法处理了包含无限多个旋转的最小化问题，见 $\mathbb{R}^3$ [33] 和 $\mathbb{R}^n$ [4, 引理5.5]。

Gromov-瓦瑟斯坦距离可定义于度量测度空间，而这些空间不必嵌入同一个环境空间。最简单的格罗莫夫-豪斯多夫距离（GH）在多项式时间内无法以小于 3 的任何因子进行近似，除非 P =NP [59, 推论 3.8]。针对超度量空间的 GH 多项式时间算法已经被设计出来 [50]。然而，即便对于直线上的有限点集，GH 空间也颇具挑战性 $\mathbb{R}$，见 [48] 和 [75]。

实验方法涵盖了种类繁多的描述符，这些描述符由人工设计或通过机器学习优化，例如尺度不变特征变换 [56, 64, 66, 77]。其中一些描述符旨在对点的置换保持不变 [55, 74], ，也旨在对等距变换保持不变 [16, 52, 61], ，例如几何深度学习中 [14, 15]。在诸多障碍中 [1, 21, 22, 36, 47], 棘手之处在于，如何从理论上保证这类描述符在扰动下的完备性和利普希茨连续性，正如问题 1.1所述。

距离的局部分布在 M 中´emoli 的开创性研究[? ， 49] 对于度量测度空间，或形状分布[8, 28, 29, 34, 53], 是新型 SDD 的一阶版本。问题 1.1 的另一种解决方法采用基于方向的不变量 [43], ，这启发了完全神经网络[38]。利普希茨连续性已在 [43, 定理 49] 中对一般位置得到证明，但对于近奇异构形则尚未证明，例如三角形退化为直线时。这些退化情形将在即将发表的研究中讨论

续篇中 [46]，将 SDD扩展为点云的完备不变量 $\mathbb{R}^n$。

3. 单纯形距离分布（SDD）

本节介绍任意度量空间 X 中由无标记点构成的有限点云 C 的等距不变量 SDD。向量 $u=(u_1,\ldots,u_h)$ 和 $v=(v_1,\ldots,v_h)$ 上的字典序 顺序 $u<v$ 表示：如果 u,v 的前 i 个（可能为 $i=0$）坐标相同，那么 $u_{i+1}<v_{i+1}$。令 S_h 表示作用于索引 $1,\ldots,h$ 的置换群。

定义 *3.1*（*RDD(C; A)*）。设 C 是一个由 m 个无标记点组成的点云，位于带有度量 d 的空间中。基序列$A=(p_1,\ldots,p_h)\in C^h$ 由 $1\le h<m$ 个互不相同的点组成。令 *D(A)* 表示三角距离矩阵，其元素 $D(A)_{i,j-1}$ 为 $d(p_i,p_j)$（$1\le i<j\le h$），其余元素均填为零。任意置换 $\xi\in S_h$ 通过将 $D(A)_{ij}$ 映射为 $D(A)_{kl}$，作用于*D(A)*；其中 $k\le l$ 是索引对 $\xi(i)$ $\xi(j)$ 按递增顺序写成的形式。

对于任意其他点 $q\in C-A$，将从q到 p_1、……、p_h 的距离写成一列。 $h\times(m-h)$-矩阵*R(C; A)*由这些按字典序排列的 $m-h$ 列构成。 ξ 对 *R(C; A)*的作用会将任意第 i行映射到第 $\xi(i)$行，之后可以将所有列重新按字典序排列。相对距离分布 *RDD(C; A)*是矩阵[*D(A) R(C; A)*] 在置换 $\xi\in S_h$下的等价类。

对于 $h=1$ 和基序列 $A=(p_1)$，矩阵*D(A)*为空，而 *R*(*C*; *A*)是一行距离，这些距离按递增顺序排列，表示从 p_1 到所有其他点 $q\in C$的距离。对于$h=2$ 和基序列 $A=(p_1,p_2)$，矩阵 *D*(*A*)是单个数值 $d(p_1,p_2)$，而 *R*(*C*; *A*)由两行距离组成，表示从 p_1,p_2 到所有其他点 $q\in C$的距离。

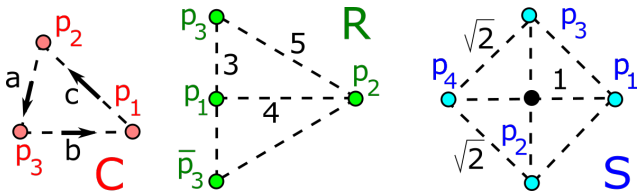

图 2 $C=\{p_1,p_2,p_3\}$ $a\le$。**左侧**：一个包 $b\le c$ $R=\{0,0,\ 4,0,\ 0,3\}$ 含距离的点云。**中间**：三角形点云 () () ()。**右侧**：正方形点云 $S=\{(1,0),(-1,0),(0,1),(-1,0)\}$。

例 3.2 (RDD 针对一个 3点云 *C*)。设 $C\subset\mathbb{R}^2$由 p_1、p_2、p_3 构成，其点间距离 $a\le b\le c$按逆时针顺序排列，如图 *2*(左)所示。则

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

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

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

对于由列表示的有序点基序列 $A=(p_i,p_j)$，我们将其写作 RDD(*C*; *A*)。交换点 $p_1\leftrightarrow p_2$ 后，上述最后一个 RDD 等价于 $\mathrm{RDD}(C;\begin{pmatrix}p_2\\p_1\end{pmatrix})=[c;\begin{pmatrix}a\\b\end{pmatrix}]$。

尽管 RDD(*C*; *A*) 在对 $A\in C^h$ 中的点进行置换 $\xi\in S_h$ 后才有定义，但比较 RDD 在度量与 ξ 无关时，对于 $h=2,3$ 是可行的。

定义 3.3 (单纯形距离分布SDD$(C;h)$)。设 C 是度量空间中由 m 个无标记点组成的点云。对于整数 $1\le h<m$，单纯形距离分布 $\mathrm{SDD}(C;h)$ 是所有无序的 h 点子集 $A\subset C$ 所对应的RDD$(C;A)$ 构成的无序集合。

对于 $h=1$任意 m点点云 C，分布SDD$(C;1)$ 可以视为一个矩阵，其中包含 m 行，每行按序排列了每个点到 $p\in C$ 所有其他 $m-1$点的距离。若按字典序排列这些 m 行，并将所有 $l>1$ 相同行合并为一行，权重为 l/m，便得到PDD$(C;m-1)$，即 [71, 定义 31]中引入的点态距离分布。

PDD 被简化为更易于比较的平均最小距离向量 [73]: $\mathrm{AMD}_k(C)=\frac{1}{m}\sum_{i=1}^{m}d_{ik}$，其中表示点 $p_i\in C$到其 k第 C 近邻的距离。这些基于邻居的不变量可在 m [25] 近线性时间内计算，并对全球最大真实材料数据库中的全部 66 万种+周期晶体进行了两两比较 [71]。定义3.4 类似地将 SDD映射为一个更小的不变量。

Recall that the 1st moment of a set of numbers $a_1,\ldots,a_k$ is the *average* $\mu=\frac{1}{k}\sum_{i=1}^{k}a_i$. The 2nd moment is the *standard deviation* $\sigma=\sqrt{\frac{1}{k}\sum_{i=1}^{k}(a_i-\mu)^2}$. For $l\ge3$, the l-th *standardized moment* [40, section 2.7] is $\frac{1}{k}\sum_{i=1}^{k}\left(\frac{a_i-\mu}{\sigma}\right)^l$.

定义 *3.4*（单形距离矩 *SDM*）。对于度量空间中任意由 m 个点组成的点云 C，令 $A\subset C$ 为 h 个无序点的子集。排序距离向量 *SDV(A)* 是由 A 中各点之间所有 $^h(_{h-1})$ *2* 个两两距离按递增顺序排列而成的列表。向量 $\vec{R}$*(C; A)* $\in\mathbb{R}^{m-h}$ 是根据定义 *3.1* 中的 $h\times(m-h)$ 矩阵 *R(C; A)*，将列平均值向量按递增顺序排列后得到的。

这一对 $[\mathrm{SDV}(A);\ \vec{R}(C;A)]$ 是平均距离分布ADD$(C;A)$，视为长度为$h(h-3)_2$ $+m$的向量。对于以下对象，~~ADD~~$(C;A)$ 的无序集合为

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

表1. 定义3.3中的单纯形距离分布3.3 适用于图1中的4点云 T, $K \subset \mathbb{R}^2$, 见图1。符号×2 表示重复的RDD。底部三行显示定义3.4中SDM(C;2, 1) $\in \mathbb{R}^3$的坐标，适用于3.4 中的 $h = 2$, $l = 1$ $C = T$ K 以及两者，。不同元素已高亮显示。。

所有 $\binom{m}{h}$ 无序子集 $A \subset C$ 均为平均单纯形分布 ASD(C; h)。单纯形距离矩 SDM(C; h, l) 是 ASD(C; h) 的 l 阶（针对 $l \geq 3$ 标准化）矩，将其视为 $\binom{m}{h}$ 向量的概率分布，并对每个坐标分别考察。

例*3.5*（T、K的*SDD*和*SDM*）。图*1*展示了具有相同有序成对距离的非等距*4*点云 T、K：*SDV* $= \{\sqrt{2}, \sqrt{2}, 2, \sqrt{10}, \sqrt{10}, 4\}$，见 *[9]*中的无穷多个示例。$T$、$K$ 边上的箭头表示*RDD*中每对顶点的点的顺序。于是，表*1* 中的*SDD(T; 2)* $\neq$ *SDD(K; 2)*可将 T、K 区分开来，且区分结果在等距变换意义下成立。*SDM(C; 2, 1)* $\in \mathbb{R}^3$ 的第*1*个坐标是*SDV*中六个距离的平均值（T、K也相同），但另外两个坐标（*R(C; A)*矩阵的列平均值）不同。

有些 $\binom{m}{h}$ RDD在SDD(C; h) 中可能会重合，如例3.5所示。如果我们将任何 $l > 1$ 相同的RDD合并为一个 RDD，权重为 $l/\binom{m}{h}$，则可以将 SDD 视为 RDD 的加权概率分布。

所有时间复杂度均在随机存取机（RAM）模型下证明。在一般度量空间中，点云 C 通常由 C 中（顺序任意的）点的距离矩阵给出。因此，我们假设 C 中任意两点之间的距离都可以在常数时间内获取。

定理3.6（SDD的不变性与时间复杂度）。对于 $h \geq 1$ 以及度量空间中由 m 个无标记点构成的任意点云 C，SDD(C; h) 是一个等距不变量，其计算时间为 $O(m^{h+1}/(h-1)!)$。对于任意 $l \geq 1$，不变量 SDM(C; h, l) $\in \mathbb{R}^{m+\binom{h(h-3)}{2}}$ 所需的时间相同。

证明。任意等距变换 $S \to Q$ 都保持距离不变，因此会为 $h \geq 1$ 诱导出一个双射 SDD(S; h) $\to$ SDD(Q; h)。

根据定义 3.3，对于度量空间中的任意 $h \geq 1$ 个未标记点组成的点云 C m，其单纯形距离分布 SDD(C; h) 由 $\binom{m}{h} = {}_h\ {}^{m}!/(m-h)!$ $= O(m^h/h!)$ 个相对距离分布 RDD(C; A) 构成，每个对应于 h个点的任意无序子集 $A \subset C$。

对于 A中点的任意排列，每个 RDD(C; A) 由距离矩阵 $D(A)$构成，这需要 $O(h^2)$ 时间；还包括矩阵 $R(C; A)$，这需要 $h(m-h)$ 时间。由于 $h \leq m$，额外的因子 $O(hm)$ 使得 SDD(C; h) 的最终时间复杂度为$O(m^{h+1}/(h-1)!)$。

对于一个固定的 h 点子集 $A \subset C$，定义 3.4 中的向量 $\vec{R}(C; A)$ 需要 $O(hm)$ 时间来计算 $m - h$ 列中 h 个距离的平均值，并需要 $O(m \log m)$ 时间对这些平均值排序。通过在 $O(h^2 \log h)$ 时间内对来自$D(A)$ 的所有两两距离排序，即可得到有序两两距离列表 SDV(A)。因此，将有序向量 SDV(A) $\in \mathbb{R}^{\binom{h(h-1)}{2}}$ 与 $\vec{R}(C; A) \in \mathbb{R}^{m-h}$ 拼接得到的平均距离分布 ADD(C; A)，仅需额外 $O((h^2+m) \log m)$ 时间。因此，对所有 h 点子集 $A \subset C$，平均单纯形分布 ASD(C; h) 所需时间为 $O(m^{h+1}/(h-1)!)$，包括$O((h^2+m) \log m)$，与初始 SDD(C; h) 的时间复杂度相同。

对于 $l = 1$，一阶原始矩 SDM(C; h, 1) 是所有长度为 $O(hm)$ 的 $k = \binom{m}{h}$ 向量 ADD(C; A) 的简单平均值，因此需要 $O(m^{h+1}/(h-1)!)$ 时间。对于$l = 2$，所有向量 ADD(C; A) 中每个坐标的标准差 σ 需要相同的时间。然后，对于任意固定的 $l \geq 3$，l 阶标准化矩 $\frac{1}{k}\sum_{i=1}^{k}\left(\frac{a_i - \mu}{\sigma}\right)^l$ 同样需要时间 $O(m^{h+1}/(h-1)!)$。□

我们猜想，SDD(C; h) 对于某个 $C \subset \mathbb{R}^n$ 点云 $h \leq n$ 而言，是一个完备等距不变量。第4 节表明，SDD(C;2) 能够区分所有已知的无穷多个点对 [54,图 S4] 中的非等距 m点集 S，$Q \subset \mathbb{R}^3$ 这些点集具有相同的PDD(S) PDD(Q)

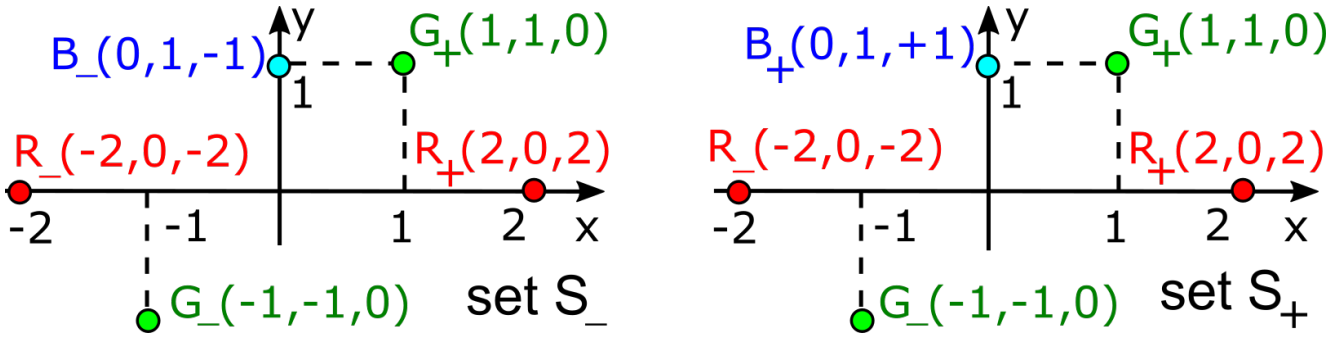

图3. **左图**：5点集 $S_- \subset \mathbb{R}^3$ 的 (x,y) 投影，其中包含绿色点 $G_- = (-1,-1,0)$ 和 $G_+ = (1,1,0)$、红色点 $R_- = (-2,0,-2)$ 和 $R_+ = (2,0,2)$，以及蓝色点 $B_- = (0,1,-1)$。**右图**：要从 S_- 得到集合 $S_+ \subset \mathbb{R}^3$，将点 B_- 替换为新点 $B_+ = (0,1,1)$。

4. 等距不变量 SDD 的强度

例4.1和4.2 分别通过比较实数集 $\mathbb{R}^3$ 中SDD的2阶形式，区分了由5个点和7个点组成的点云。例4.3中，SDD$(C;2)$ 能够区分一个由三个参数决定的点对族中的6点云。

例4.1 (5点云)。 图 3展示了取自 $S_\pm \subset \mathbb{R}^3$ 的 *[54, 图 S4(A)]*。

S_-的距离	R_-	R_+	G_-	G_+	B_-
$R_-(-2,\ 0,\ -2)$	0	$\sqrt{32}$	$\sqrt{6}$	$\sqrt{14}$	$\sqrt{6}$
$R_+(+2,0,+2)$	$\sqrt{32}$	0	$\sqrt{14}$	$\sqrt{6}$	$\sqrt{14}$
$G_-(-1,\ -1,\ 0)$	$\sqrt{6}$	$\sqrt{14}$	0	$\sqrt{8}$	$\sqrt{6}$
$G_+(+1,+1,0)$	$\sqrt{14}$	$\sqrt{6}$	$\sqrt{8}$	0	$\sqrt{2}$
$B_-(0,\ +1,\ -1)$	$\sqrt{6}$	$\sqrt{14}$	$\sqrt{6}$	$\sqrt{2}$	0

到 S_+的距离	R_-	R_+	G_-	G_+	B_+
$R_-(-2,\ 0,\ -2)$	0	$\sqrt{32}$	$\sqrt{6}$	$\sqrt{14}$	$\sqrt{14}$
$R_+(+2,\ 0,\ +2)$	$\sqrt{32}$	0	$\sqrt{14}$	$\sqrt{6}$	$\sqrt{6}$
$G_-(-1,-1,0)$	$\sqrt{6}$	$\sqrt{14}$	0	$\sqrt{8}$	$\sqrt{6}$
$G_+(+1,\ +1,\ 0)$	$\sqrt{14}$	$\sqrt{6}$	$\sqrt{8}$	0	$\sqrt{2}$
$B_+(0,\ +1,\ +1)$	$\sqrt{14}$	$\sqrt{6}$	$\sqrt{6}$	$\sqrt{2}$	0

表2. 集合中所有点之间的距离 $S_\mp$ 见图3。

这些集合 $S_\pm$ 并非等距，因为 S_+ 有三个点 B_+、G_+、R_+，它们的两两距离分别为 $\sqrt{2}$、$\sqrt{6}$、$\sqrt{6}$，但 S_- 不存在这样的一组三点。表2 展示了距[illegible]间的差异。若按与邻点的距离排序，表3 中的矩阵仅有一对不同。

如果忽略点的标签， $S_\pm$ 具有相同的点逐一距离分布（PDD），即定义 SDD中的单纯形逐一距离分布 （3.3）。

为便于直观展示，下面的矩阵通过以下方式得到：

S_- 到	第 1 个邻居	2nd	3rd	4th
$R_- = (-2,\ 0,\ -2)$	$\sqrt{6}$	$\sqrt{6}$	$\sqrt{14}$	$\sqrt{32}$
$R_+ = (+2,\ 0,\ +2)$	$\sqrt{6}$	$\sqrt{14}$	$\sqrt{14}$	$\sqrt{32}$
$G_- = (-1,-1,0)$	$\sqrt{6}$	$\sqrt{6}$	$\sqrt{8}$	$\sqrt{14}$
$G_+ = (+1,+1,0)$	$\sqrt{2}$	$\sqrt{6}$	$\sqrt{8}$	$\sqrt{14}$
$B_- = (0,+1,-1)$	$\sqrt{2}$	$\sqrt{6}$	$\sqrt{6}$	$\sqrt{14}$

S_+ 到.....的距离	第 1 个邻居	2nd	3rd	4th
$R_- = (-2,\ 0,\ -2)$	$\sqrt{6}$	$\sqrt{14}$	$\sqrt{14}$	$\sqrt{32}$
$R_+ = (+2,\ 0,\ +2)$	$\sqrt{6}$	$\sqrt{6}$	$\sqrt{14}$	$\sqrt{32}$
$G_- = (-1,-1,0)$	$\sqrt{6}$	$\sqrt{6}$	$\sqrt{8}$	$\sqrt{14}$
$G_+ = (+1,+1,0)$	$\sqrt{2}$	$\sqrt{6}$	$\sqrt{8}$	$\sqrt{14}$
$B_+ = (0,\ +1,\ -1)$	$\sqrt{2}$	$\sqrt{6}$	$\sqrt{6}$	$\sqrt{14}$

表3。对于图 S_+ 中5点集的每个点 3，表 2中的邻点距离按行排序。

按字典序对表格的行进行排序3：

$$\mathrm{PDD}(S_\pm) = \mathrm{SDD}(S_\pm;1) = \begin{pmatrix} \sqrt{2} & \sqrt{6} & \sqrt{6} & \sqrt{14} \\ \sqrt{2} & \sqrt{6} & \sqrt{8} & \sqrt{14} \\ \sqrt{6} & \sqrt{6} & \sqrt{8} & \sqrt{14} \\ \sqrt{6} & \sqrt{6} & \sqrt{14} & \sqrt{32} \\ \sqrt{6} & \sqrt{14} & \sqrt{14} & \sqrt{32} \end{pmatrix}$$

现在我们证明 $\mathrm{SDD}(S_-;2) \neq \mathrm{SDD}(S_+;2)$。对于 $h=2$，单纯形距离分布 $\mathrm{SDD}(C;h)$ 由 2 点子集 $A \subset C$ 对应的 $\mathrm{RDD}(C;A)$ 组成。两个集合 $S_\pm$中都只有一对点 $(G_+, B-)$ 和 $(G+, B_+)$，其距离为 $\sqrt{2}$。因此，只需证明这对点对应的相对距离分布不同：

$$\mathrm{RDD}(S_-, \begin{pmatrix} G_+ \\ B_- \end{pmatrix}) = \left[\sqrt{2}, \begin{pmatrix} \sqrt{8} & \sqrt{14} & \sqrt{6} \\ \sqrt{6} & \sqrt{6} & \sqrt{14} \\ G_- & R_- & R_+ \end{pmatrix}\right]$$

$$\mathrm{RDD}(S_+, \begin{pmatrix} G_+ \\ B_+ \end{pmatrix}) = \left[\sqrt{2}, \begin{pmatrix} \sqrt{8} & \sqrt{14} & \sqrt{6} \\ \sqrt{6} & \sqrt{14} & \sqrt{6} \\ G_- & R_- & R_+ \end{pmatrix}\right]$$

上面 3×3 矩阵的最后一行表示一个用于对2×3 矩阵 $R(C; A)$的列进行索引的补充点 $q \in C - A$，如定义 *3.1* 所示。所得的 *RDD* 不同，因为对 $R(S_+; \{G_+, B_+\})$ 的行或列进行任何置换，点对 $\sqrt{6}, \sqrt{6}$ 仍位于同一列；而 $R(S_-; \{G_+, B_+\})$中没有点对 $\sqrt{6}, \sqrt{6}$ 位于同一列。因此，$SDD(S_-; 2) \neq SDD(S_+; 2)$。

例 4.2 (7点集)。图 $Q_\pm$ 中的集合 *4* 取自 *[54, 图 S4(B)]*，其距离见表 *4*。两个集合中，距离为 $\sqrt{6}$ 的点对都只有两对。因此，只需比较RDD，如下所示。

$$R(Q_-; \begin{pmatrix} G \\ B_{+1} \end{pmatrix}) = \begin{pmatrix} \sqrt{32} & \sqrt{14} & \sqrt{17} & 3 & \sqrt{13} \\ \sqrt{6} & \sqrt{8} & \sqrt{5} & 1 & \sqrt{3} \\ R & B_{-1} & B_{-2} & B_{+2} & O_- \end{pmatrix},$$

$$R(Q_-; \begin{pmatrix} R \\ B_{-1} \end{pmatrix}) = \begin{pmatrix} \sqrt{32} & \sqrt{14} & 3 & \sqrt{17} & \sqrt{5} \\ \sqrt{14} & \sqrt{8} & 3 & \sqrt{13} & \sqrt{3} \\ G & B_{+1} & B_{-2} & B_{+2} & O_- \end{pmatrix}$$

The pair above has submatrices $\begin{pmatrix} 3 & \sqrt{13} \\ 1 & \sqrt{3} \end{pmatrix}$ *and* $\begin{pmatrix} 3 & \sqrt{5} \\ 3 & \sqrt{3} \end{pmatrix}$ *but the pair below has no such submatrices.*

$$R(Q_+; \begin{pmatrix} G \\ B_{+1} \end{pmatrix}) = \begin{pmatrix} \sqrt{32} & \sqrt{14} & \sqrt{17} & 3 & \sqrt{5} \\ \sqrt{6} & \sqrt{8} & \sqrt{5} & 1 & \sqrt{3} \\ R & B_{-1} & B_{-2} & B_{+2} & O_+ \end{pmatrix},$$

$$R(Q_+; \begin{pmatrix} R \\ B_{-1} \end{pmatrix}) = \begin{pmatrix} \sqrt{32} & \sqrt{14} & 3 & \sqrt{17} & \sqrt{13} \\ \sqrt{14} & \sqrt{8} & 3 & \sqrt{13} & \sqrt{3} \\ G & B_{+1} & B_{-2} & B_{+2} & O_+ \end{pmatrix}$$

The pair of $\mathrm{RDD}(Q_-; \begin{pmatrix} G \\ B_{+1} \end{pmatrix})$ *and* $\mathrm{RDD}(Q_-; \begin{pmatrix} R \\ B_{-1} \end{pmatrix})$ *differs from the pair* $\mathrm{RDD}(Q_+; \begin{pmatrix} G \\ B_{+1} \end{pmatrix})$ *and* $\mathrm{RDD}(Q_+; \begin{pmatrix} R \\ B_{-1} \end{pmatrix})$. *Hence* $\mathrm{SDD}(Q_-; 2) \neq \mathrm{SDD}(Q_+; 2)$.

例*4.3*（*6* 点集）。图 *5* 中的集合 $T_\pm$ 受到 *[54, 图 S4(C)]*，的启发，其中包含来自例 *4.2* 中集合 $Q_\pm$ 的点 R、G、$O_\pm$，以及三个新点 C_1（x_1，y_1，*0*）、C_2（x_2，y_2，*0*）、C_3（x_3，y_3，*0*），使得$|RC_1| = |GC_2|$、$|RC_2| = |GC_3|$、$|RC_3| = |GC_1|$。

将这三对线段投影到 xy 平面后的长度分别记为 $2l_1$、$2l_2$、$2l_3$，则

$$(4.3.1)\begin{cases} (x_2+2)^2 + y_2^2 = |RC_2|^2 - 4 = (2l_1)^2, \\ (x_3-2)^2 + y_3^2 = |GC_3|^2 - 4 = (2l_1)^2; \end{cases}$$

$$(4.3.2)\begin{cases} (x_3+2)^2 + y_3^2 = |RC_3|^2 - 4 = (2l_2)^2, \\ (x_1-2)^2 + y_1^2 = |GC_1|^2 - 4 = (2l_2)^2; \end{cases}$$

$$(4.3.3)\begin{cases} (x_1+2)^2 + y_1^2 = |RC_1|^2 - 4 = (2l_3)^2, \\ (x_2-2)^2 + y_2^2 = |GC_2|^2 - 4 = (2l_3)^2. \end{cases}$$

比较（*4.3.1*）的第一部分和（*4.3.3*）的第二部分可得 $(2l_1)^2 - 4x_2 = (2l_3)^2 + 4x_2$，因此$x_2 = \frac{l_1^2 - l_3^2}{2}$。类似地，$x_3 = \frac{l_2^2 - l_1^2}{2}$，$x_1 = \frac{l_3^2 - l_2^2}{2}$，因此$x_1 + x_2 + x_3 = 0$。由（*4.3.2*）的第二部分可得 $x_1^2 - 4x_1 + 4 + y_1^2 = 4l_2^2$，因此

$$|O_\pm C_1|^2 = x_1^2 + y_1^2 + 1 = 4l_2^2 + 4x_1 - 3 = 2l_2^2 + 2l_3^2 - 3,$$

类似地 $|O_\pm C_2|^2 = 2l_3^2 + 2l_1^2 - 3$，$|O_\pm C_3|^2 = 2l_1^2 + 2l_2^2 - 3$。那么 $|C_1C_2|^2 = (x_1 - x_2)^2 + (y_1 - y_2)^2 = x_1^2 + y_1^2$。

表 *5* 列出了 $T_\mp$ 中各点之间的所有两两距离。我们将在下文证明，$T_\pm$ 可由简化后的不变量 $SDD(T_\pm; 2)$ 区分。对于 $R(C; A)$ 的每一列，我们还允许元素独立于其他列进行任意置换，因此可以将每列（一个距离对）按字典序排列。将 *RDD* 简化后的结果记为 *RDD*。那么，对于列表 $\{R, G\}$、$\{O_\pm, C_i\}$、$\{C_i, C_j\}$ 中互不相同的 i、$j = 1$、*2*、*3*，$SDD(T_\pm; 2)$ 在 *2* 点子集 A 上具有相同的 *RDD*。

例如，两个 $RDD(T_\pm; \{R, G\})$ 都以距离 $|R - G| = \sqrt{32}$ 开头，随后包含相同的四个点对 $(\sqrt{5}, \sqrt{13})$、$(2\sqrt{l_i^2} + 1, 2\sqrt{l_{i-1}^2 + 1})$（模*3*），这些点对应按字典序排列并书写。因此，对于表 *6* 中列表 $\{R, O_\pm\}$、$\{G, O_\pm\}$、$\{R, C_i\}$、$\{G, C_j\}$ 中的 A，比较 $SDD(T_\pm; 2)$ 时，只需比较剩余的 $RDD(T_\pm; A)$。

不失一般性，假设 $l_1 \geq l_2 \geq l_3$。如果所有长度都互不相同，那么 $l_1 > l_2 > l_3$。在表 *6* 中，即使将每个点对都按距离从小到大的顺序排列，并将所有点对按字典序排列，$\{R, O_-\}$ 和 $\{G, O_+\}$ 对应的行仍然不同。因此 $\mathrm{SDD}(T_-; 2) \neq \mathrm{SDD}(T_+; 2)$，除非 l_i 中有两个相等。

若（例如）$l_1 = l_2$，则在 $\mathrm{SDD}(T_\pm; 2)$ 中，$\{R, O_-\}$ 和 $\{G, O_+\}$ 的按字典序排列的行重合；$\{G, O_-\}$ 和 $\{R, O_+\}$ 的行也同样如此。因此，只需在 $\mathrm{SDD}(T_\pm; 2)$ 中比较剩余点对 $\{R, C_i\}, \{G, C_j\}$ 所对应的六行。

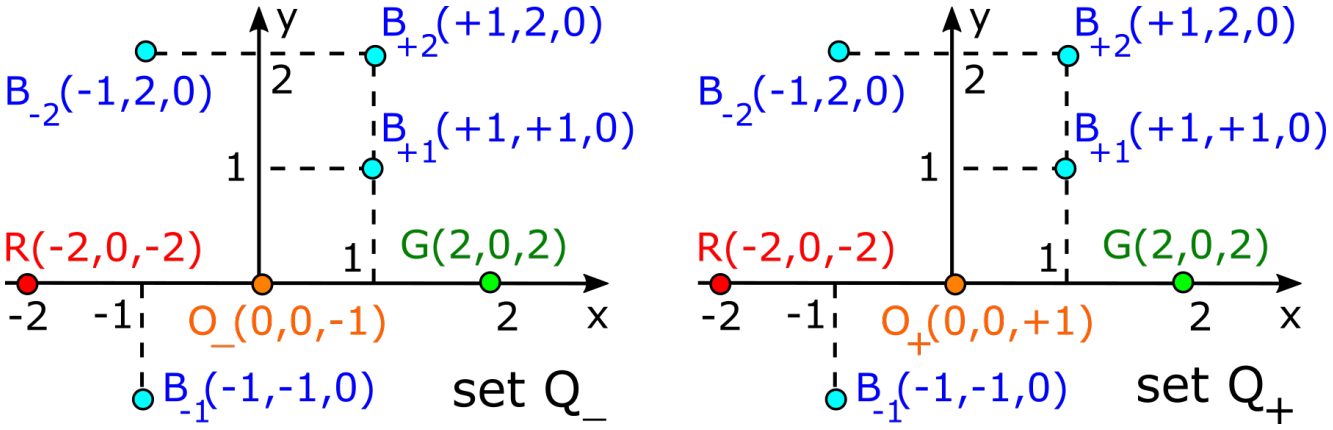

图4. **左图**：7点集 $Q_- \subset \mathbb{R}^3$ 关于红色点 $R=(-2,\ 0,\ -2)$、绿点 $G=(2,\ 0,\ 2)$和四个蓝点 $B_{\pm 1}=(\pm 1,\ \pm 1,\ 0)$、 $B_{\pm 2}=(\pm 1,\ 2,\ 0)$的 $(x,\ y)$-投影，以及橙色点 $O_-=(0,\ 0,\ -1)$。**右图**：要从 Q_- 得到 Q_+，将 O_- 替换为 $O_+=(0,\ 0,\ +1)$。

的距离	R	G	B_{-1}	B_{+1}	B_{-2}	B_{+2}	O_-
$R=(-2,\ 0,\ -2)$	0	$\sqrt{32}$	$\sqrt{6}$	$\sqrt{14}$	3	$\sqrt{17}$	$\sqrt{5}$
$G=(+2,\ 0,\ +2)$	$\sqrt{32}$	0	$\sqrt{14}$	$\sqrt{6}$	$\sqrt{17}$	3	$\sqrt{13}$
$B_{-1}=(-1,\ -1,\ 0)$	$\sqrt{6}$	$\sqrt{14}$	0	$\sqrt{8}$	3	$\sqrt{13}$	$\sqrt{3}$
$B_{+1}=(+1,+1,0)$	$\sqrt{14}$	$\sqrt{6}$	$\sqrt{8}$	0	$\sqrt{5}$	1	$\sqrt{3}$
$B_{-2}=(-1,\ 2,\ 0)$	3	$\sqrt{17}$	3	$\sqrt{5}$	0	2	$\sqrt{6}$
$B_{+2}=(+1,\ 2,\ 0)$	$\sqrt{17}$	3	$\sqrt{13}$	1	2	0	$\sqrt{6}$
$O_-=(0,\ 0,\ -1)$	$\sqrt{5}$	$\sqrt{13}$	$\sqrt{3}$	$\sqrt{3}$	$\sqrt{6}$	$\sqrt{6}$	0

的距离	R	G	B_{-1}	B_{+1}	B_{-2}	B_{+2}	O_+
$R=(-2,\ 0,\ -2)$	0	$\sqrt{32}$	$\sqrt{6}$	$\sqrt{14}$	3	$\sqrt{17}$	$\sqrt{13}$
$G=(+2,\ 0,\ +2)$	$\sqrt{32}$	0	$\sqrt{14}$	$\sqrt{6}$	$\sqrt{17}$	3	$\sqrt{5}$
$B_{-1}=(-1,-1,0)$	$\sqrt{6}$	$\sqrt{14}$	0	$\sqrt{8}$	3	$\sqrt{13}$	$\sqrt{3}$
$B_{+1}=(+1,+1,0)$	$\sqrt{14}$	$\sqrt{6}$	$\sqrt{8}$	0	$\sqrt{5}$	1	$\sqrt{3}$
$B_{-2}=(-1,\ 2,\ 0)$	3	$\sqrt{17}$	3	$\sqrt{5}$	0	2	$\sqrt{6}$
$B_{+2}=(+1,\ 2,\ 0)$	$\sqrt{17}$	3	$\sqrt{13}$	1	2	0	$\sqrt{6}$
$O_+=(0,\ 0,\ +1)$	$\sqrt{13}$	$\sqrt{5}$	$\sqrt{3}$	$\sqrt{3}$	$\sqrt{6}$	$\sqrt{6}$	0

表 4. 取自 $Q_\mp$ 图 4 的 7点集所有点之间的距离矩阵 [54, 图 S4(B)]。

对于 $l_1=l_2$，我们得到 $x_3=\dfrac{l_2^2-l_1^2}{2}=0$ 和 $x_1=-x_2=\dfrac{l_3^2-l_2^2}{2}$。在等式 *(4.3.3)* 中，等式 $(x_1+2)^2+y_1^2=(x_2-2)^2+y_2^2$ 与 $x_1=-x_2$ 意味着 $y_1^2=y_2^2$。更为退化的情形 $l_1=l_2=l_3$ 意味着$x_1=x_2=x_3=0$ 和 $y_1^2=y_2^2=y_3^2$，因此 C_1、C_2、C_3 中至少有两个应当重合。上述矛盾表明，只需考虑以下情形：$l_1=l_2>l_3$，其中 $x_1=-x_2\neq 0=x_3$ 且 $y_1=\pm y_2$，见图 *5*。

若 $y_1=y_2$，集合 $T_\pm$ 可通过 $(x,y,z)\ \mapsto$ 等距 $(-x,y,-z)$。若 $y_1=-y_2$ 且 $y_3=0$，集合 $T_\pm$ 可通过 $(x,y,z)\mapsto(-x,-y,-z)$ 等距。若 $y_1=-y_2$ 且$y_3\neq 0$，则 $C_1=(x_1,y_1,0)$，$C_2=(-x_1,-y_1,0)$，$C_3\neq(0,0,0)$。那么在剩余六行中，只有 $\{R,C_1\},\{G,C_2\}$ 所在的行含有距离为 $2\sqrt{l_3^2+1}$ 的点；关于模 *3* 考虑的 $i=3$，请参见表 *6*。于是 $i+1\equiv 1$(模 *3*)，$i-1\equiv 2$ (模 *3*)，因此 $l_{i+1}=l_1=l_2=l_{i-1}$。

观察 $\{R,C_1\},\{G,C_2\}$ 的各行可见，$\mathrm{SDD}(T_\pm;2)$ 中每行的三个公共点对包含相同的距离$2\sqrt{l_1^2+1}=2\sqrt{l_2^2+1}$，但在 $|C_{i-1}C_i|=|C_2C_3|\neq$ 上有所不同

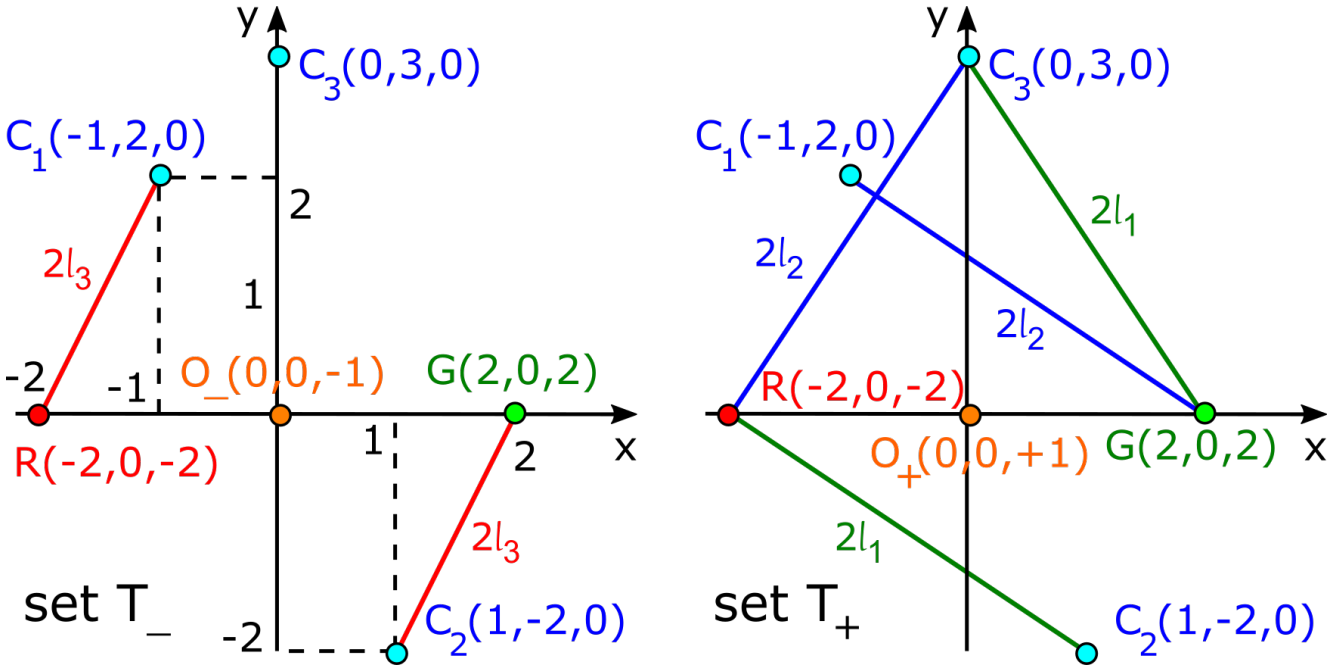

图 5. **左侧**：6点集 $T_- \subset \mathbb{R}^3$ 的 (x,y) 投影，其中包括红色点 $R=(-2,\ 0,\ -2)$、绿色点 $G=(2,\ 0,\ 2)$、三个蓝点 $C_1=(x_1,\ y_1,\ 0)$、$C_2=(x_2,\ y_2,\ 0)$、$C_3=(x_3,\ y_3,\ 0)$ 以及橙色点 $O_-=(0,\ 0,\ -1)$，使得 $|RC_1|=2l_3=|GC_2|$，$|RC_2|=2l_1=|GC_3|$ $|RC_3|=2l_2=|GC_1|$ $T_+ \subset \mathbb{R}^3$ T_- O_- $O_+=\ 0\ 0\ +1$ ，。**右侧**：要从 中得到该集合，将 替换为 (，，)。

T_- 的距离	R	G	C_1	C_2	C_3	O_-
$R=(-2,\ 0,\ -2)$	0	$\sqrt{32}$	$2\sqrt{l_3^2+1}$	$2\sqrt{l_1^2+1}$	$2\sqrt{l_2^2+1}$	$\sqrt{5}$
$G=(+2,\ 0,\ +2)$	$\sqrt{32}$	0	$2\sqrt{l_2^2+1}$	$2\sqrt{l_3^2+1}$	$2\sqrt{l_1^2+1}$	$\sqrt{13}$
$C_1=(x_1,y_1,0)$	$2\sqrt{l_3^2+1}$	$2\sqrt{l_2^2+1}$	0	$\lvert C_1C_2\rvert$	$\lvert C_3C_1\rvert$	$\sqrt{2l_2^2+2l_3^2-3}$
$C_2=(x_2,y_2,0)$	$2\sqrt{l_1^2+1}$	$2\sqrt{l_3^2+1}$	$\lvert C_1C_2\rvert$	0	$\lvert C_2C_3\rvert$	$\sqrt{2l_3^2+2l_1^2-3}$
$C_3=(x_3,y_3,0)$	$2\sqrt{l_2^2+1}$	$2\sqrt{l_1^2+1}$	$\lvert C_3C_1\rvert$	$\lvert C_2C_3\rvert$	0	$\sqrt{2l_1^2+2l_2^2-3}$
$O_-=(0,\ 0,\ -1)$	$\sqrt{5}$	$\sqrt{13}$	$\sqrt{2l_2^2+2l_3^2-3}$	$\sqrt{2l_3^2+2l_1^2-3}$	$\sqrt{2l_1^2+2l_2^2-3}$	0

T的距离 $_+$	R	G	C_1	C_2	C_3	O_+
$R=(-2,\ 0,\ -2)$	0	$\sqrt{32}$	$2\sqrt{l_3^2+1}$	$2\sqrt{l_1^2+1}$	$2\sqrt{l_2^2+1}$	$\sqrt{13}$
$G=(+2,\ 0,\ +2)$	$\sqrt{32}$	0	$2\sqrt{l_2^2+1}$	$2\sqrt{l_3^2+1}$	$2\sqrt{l_1^2+1}$	$\sqrt{5}$
$C_1=(x_1,y_1,0)$	$2\sqrt{l_3^2+1}$	$2\sqrt{l_2^2+1}$	0	$\lvert C_1C_2\rvert$	$\lvert C_3C_1\rvert$	$\sqrt{2l_2^2+2l_3^2-3}$
$C_2=(x_2,y_2,0)$	$2\sqrt{l_1^2+1}$	$2\sqrt{l_3^2+1}$	$\lvert C_1C_2\rvert$	0	$\lvert C_2C_3\rvert$	$\sqrt{2l_3^2+2l_1^2-3}$
$C_3=(x_3,y_3,0)$	$2\sqrt{l_2^2+1}$	$2\sqrt{l_1^2+1}$	$\lvert C_3C_1\rvert$	$\lvert C_2C_3\rvert$	0	$\sqrt{2l_1^2+2l_2^2-3}$
$O_+=(0,\ 0,\ +1)$	$\sqrt{13}$	$\sqrt{5}$	$\sqrt{2l_2^2+2l_3^2-3}$	$\sqrt{2l_3^2+2l_1^2-3}$	$\sqrt{2l_1^2+2l_2^2-3}$	0

表 5. 6点集中所有点之间的距离矩阵 $T_\mp$ 见图 5 受 [54, 图 S4(C)]。

$|C_3C_1| = |C_iC_{i+1}|$ *as* $C_1 = \pm C_2$, $C_3 \neq (0,0,0)$.　　$l_1 = l_2 = \frac{\sqrt{13}}{2}$, $l_3 = \frac{\sqrt{5}}{2}$.

这两行的差异意味着 SDD(T_-;2) ≠SDD(T_+; 2)，这是由于其余点对中的距离互换所致 $\sqrt{5}$, $\sqrt{13}$，参见表 7 中的集合 $T_\pm$，见图5，其中

T_- 点对	距离	其中的公共对 $\widetilde{\mathrm{SDD}}(T_\pm;2)$	其中不同的对 $\widetilde{\mathrm{SDD}}(T_+;2)$
$\{R,O_-\}$	$\sqrt{5}$	$(\sqrt{13},\ \sqrt{32})$ 至 G	$(2\sqrt{l_3^2+1},\sqrt{2l_2^2+2l_3^2-3})$ 到 C_1, $(2\sqrt{l_1^2+1},\sqrt{2l_3^2+2l_1^2-3})$ 到 C_2, $(2\sqrt{l_2^2+1},\ \sqrt{2l_1^2+2l_2^2-3})$ 至 C_3
$\{G,O_-\}$	$\sqrt{13}$	$(\sqrt{5},\ \sqrt{32})$ 至 R	$(2\sqrt{l_2^2+1},\sqrt{2l_2^2+2l_3^2-3})$ 到 C_1, $(2\sqrt{l_3^2+1},\sqrt{2l_3^2+2l_1^2-3})$ 到 C_2, $(2\sqrt{l_1^2+1},\ \sqrt{2l_1^2+2l_2^2-3})$ 至 C_3
$\{R,C_{i+1}\}$	$2\sqrt{l_i^2+1}$	$(2\sqrt{l_{i-1}^2+1},\sqrt{32})$ 到 G, $(2\sqrt{l_{i+1}^2+1},\ \lvert C_{i+1}C_{i-1}\rvert)$ 到 C_{i-1}, $(2\sqrt{l_{i-1}^2+1},\ \lvert C_iC_{i+1}\rvert)$ 至 C_i	$(\sqrt{5},\ \sqrt{2l_{i-1}^2+2l_i^2-3})$ 至 O_-
$\{G,C_{i-1}\}$	$2\sqrt{l_i^2+1}$	$(2\sqrt{l_{i+1}^2+1},\ \sqrt{32})$ 到 R, $(2\sqrt{l_{i-1}^2+1},\ \lvert C_{i+1}C_{i-1}\rvert)$ 到 C_{i+1}, $(2\sqrt{l_{i+1}^2+1},\ \lvert C_{i-1}C_i\rvert)$ 到 C_i	$(\sqrt{13},\ \sqrt{2l_i^2+2l_{i+1}^2-3})$ 到 O_-

T_+ 对	距离	SDD 中的公共对 $\widetilde{\mathrm{SDD}}(T_\pm;2)$	SDD 中不同的对 $\widetilde{\mathrm{SDD}}(T_-;2)$
$\{G,O_+\}$	$\sqrt{5}$	$(\sqrt{13},\ \sqrt{32})$ 至 R	$(2\sqrt{l_2^2+1},\sqrt{2l_2^2+2l_3^2-3})$ 到 C_1, $(2\sqrt{l_3^2+1},\sqrt{2l_3^2+2l_1^2-3})$ 到 C_2, $(2\sqrt{l_1^2+1},\ \sqrt{2l_1^2+2l_2^2-3})$ 至 C_3
$\{R,O_+\}$	$\sqrt{13}$	$(\sqrt{5},\ \sqrt{32})$ 至 G	$(2\sqrt{l_3^2+1},\sqrt{2l_2^2+2l_3^2-3})$ 到 C_1, $(2\sqrt{l_1^2+1},\ \sqrt{2l_3^2+2l_1^2-3})$ 到 C_2, $(2\sqrt{l_2^2+1},\ \sqrt{2l_1^2+2l_2^2-3})$ 至 C_3
$\{R,C_{i+1}\}$	$2\sqrt{l_i^2+1}$	从 $(2\sqrt{l_{i-1}^2+1},\ \sqrt{32})$ 到 G, 从 $(2\sqrt{l_{i+1}^2+1},\ \lvert C_{i+1}C_{i-1}\rvert)$ 到 C_{i-1}, 从 $(2\sqrt{l_{i-1}^2+1},\ \lvert C_iC_{i+1}\rvert)$ 到 C_i	从 $(\sqrt{13},\ \sqrt{2l_{i-1}^2+2l_i^2-3})$ 到 O_+
$\{G,C_{i-1}\}$	$2\sqrt{l_i^2+1}$	从 $(2\sqrt{l_{i+1}^2+1},\ \sqrt{32})$ 到 R, 从 $(2\sqrt{l_{i-1}^2+1},\ \lvert C_{i+1}C_{i-1}\rvert)$ 到 C_{i+1}, $(2\sqrt{l_{i+1}^2+1},\ \lvert C_{i-1}C_i\rvert)$到 C_i	$(\sqrt{5},\ \sqrt{2l_i^2+2l_{i+1}^2-3})$ 至 O_+

表 6. 对于集合 $T_\pm$，分布 $\widetilde{\mathrm{SDD}}(T_\pm;2)$ 只能在上方所示点对 $\{R,O_\pm\},\{G,O_\pm\},\{R,C_i\},\{G,C_i\}$ 的 $\widetilde{\mathrm{RDD}}$ 上有所不同，其中 $i\in\{1,2,3\}$ 按模 3 考虑，因此 $1-1\equiv 3$(模 3)。在对应点对的各行中，两个 $\widetilde{\mathrm{SDD}}(T_\pm;2)$ 中有些距离点对相同，但其他点对可能不同。如果参数 l_1,l_2,l_3 两两不同，那么行$\{R,O_-\},\{G,O_+\}$ 包含三个不同的距离点对，因此 $\widetilde{\mathrm{SDD}}(T_-;2)\neq\widetilde{\mathrm{SDD}}(T_+;2)$，参见例4.3。

T_- 对	距离	到邻居1的距离	到邻居2的距离	到邻居3的距离	到邻居4的距离
$\{R, C_1\}$	3	$(\sqrt{2},\ \sqrt{17})$ 至 C_3	$(\sqrt{5},\ \sqrt{6})$ 至 O_-	$(\sqrt{17},\ \sqrt{20})$ 至 C_2	$(\sqrt{17},\ \sqrt{32})$ 至 G
$\{G, C_2\}$	3	$(\sqrt{6},\ \sqrt{13})$ 至 O_-	$(\sqrt{17},\ \sqrt{20})$ 至 C_1	$(\sqrt{17},\ \sqrt{26})$ 至 C_3	$(\sqrt{17},\ \sqrt{32})$ 至 R

T_+ 对	距离	到邻居1的距离	到邻居2的距离	到邻居3的距离	到邻居4的距离
$\{R, C_1\}$	3	$(\sqrt{2},\ \sqrt{17})$ 至 C_3	$(\sqrt{6},\ \sqrt{13})$ 至 O_+	$(\sqrt{17},\ \sqrt{20})$ 至 C_2	$(\sqrt{17},\ \sqrt{32})$ 至 G
$\{G, C_2\}$	3	$(\sqrt{5},\ \sqrt{6})$ 到 O_+	$(\sqrt{17},\ \sqrt{20})$ 到 C_1	$(\sqrt{17}, \sqrt{26})$ 到 C_3	$(\sqrt{17},\ \sqrt{32})$ 至 R

表 7。上述各行表明 $\mathrm{SDD}(T_-;2) \neq \mathrm{SDD}(T_+;2)$ 对于集合 $T_\pm$ 满足 $C_1 = (-1, 2, 0)$, $C_2 = (1, -2, 0)$, $C_3 = (0, 3, 0)$因此 $l_1 = l_2 = \frac{\sqrt{5}}{2}$, $l_3 = \frac{\sqrt{13}}{2}$ 见表 6。

5. SDD 上的连续可计算度量

这些 $m - h$ 矩阵的可置换列 $R(C; A)$在RDD 中，根据定义3.1，可以解释为 $m - h$中的无标记点 $\mathbb{R}^h$。由于任何等距变换都是双射，尊重双射的最简单度量是瓶颈距离（也称为 Wasserstein 距离 W_∞）。

定义 *5.1*（瓶颈距离 W_∞）。对于任意向量$v = (v_1, \ldots, v_n) \in \mathbb{R}^n$，闵可夫斯基范数为 $||v||_\infty = \max_{i=1,\ldots,n} |v_i|$。对于任意同尺寸的向量或矩阵 N, N'，闵可夫斯基距离为 $L_\infty(N, N') = \max_{i,j} |N_{ij} - N'_{ij}|$。对于由 m 个无标记点组成的点云 $C, C' \subset \mathbb{R}^n$，瓶颈距离 $W_\infty(C, C') = \inf_{g:C\to C'} \sup_{p\in C} ||p - g(p)||_\infty$，取遍所有双射 g 时的最小值：$C \to C'$。

引理 *5.2*（*RDD* 上的最大度量 M_∞）。对于任意 m 个点云和有序的 h 点序列 $A \subset C$ 与 $A' \subset C'$，令 $d(\xi) = \max\{L_\infty(\xi(D(A)), D(A')), W_\infty(\xi(R(C; A)), R(C'; A'))\}$，其中置换 $\xi \in S_h$ 作用于 h 个点。则最大度量 $M_\infty(\mathrm{RDD}(C; A), \mathrm{RDD}(C'; A')) = \min_{\xi\in S_h} d(\xi)$ 满足定义 *3.1* 中 *RDD* 的所有度量公理，且可在 $O(h!(h^2 + m^{1.5}\log^h m))$ 时间内计算。

引理 5.2 的证明。第一条度量公理指出，根据定义 3.1，$\mathrm{RDD}(C; A)$ 与 $\mathrm{RDD}(C'; A')$ 等价，当且仅当 $M_\infty(\mathrm{RDD}(C; A), \mathrm{RDD}(C'; A')) = 0$ or$d(\xi) = 0$ 对某个置换 $\xi \in S_h$ 成立。于是，由于 W_∞ 的第一条公理，$d(\xi) = 0$ 与 $\xi(D(A)) = D(A')$ 等价，并且在列置换意义下，$\xi(R(C; A)) = R(C'; A')$ 也等价。这两项结论意味着，根据定义 3.1，相对距离分布 $\mathrm{RDD}(C; A)$ 与 $\mathrm{RDD}(C'; A')$ 等价。对称性公理成立，因为任意置换 ξ 都可逆。为证明三角不等式 $M_\infty(\mathrm{RDD}(C; A), \mathrm{RDD}(C'; A')) +$ $M_\infty(\mathrm{RDD}(C''; A''), \mathrm{RDD}(C'; A')) \geq$ $M_\infty(\mathrm{RDD}(C; A), \mathrm{RDD}(C''; A''))$，令 $\xi, \xi' \in S_h$ 为左侧 M_∞ 个值的最优排列

如上所述。L_∞ 的三角不等式表明$L_\infty(\xi(D(A)), \xi(D(A'')))$ $+L_\infty(\xi(D(A)), \xi(D(A')))$ $\geq L_\infty(\xi(D(A)), \xi'(D(A'')))$ $= L_\infty(\xi'^{-1}\xi(D(A)), D(A''))$，定义 5.1 中的瓶颈距离 W_∞ 也同理。取 L_∞、W_∞ 的最大值仍满足三角不等式。于是，对于上述置换的复合，$M_\infty(\mathrm{RDD}(C; A), \mathrm{RDD}(C''; A''))$ $= \min_{\xi\in S_h} d(\xi)$ 不可能大于 $d(\xi'^{-1}\xi)$，因此 M_∞ 满足三角不等式。

对于固定置换 $\xi \in S_h$，距离$L_\infty(\xi(D(A)), D(A'))$需要 $O(h^2)$ 时间。瓶颈距离 $W_\infty(\xi(R(C; A)), R(C'; A'))$可视为具有可置换列的 $h \times (m - h)$ 矩阵 $\xi(R(C; A))$ 和 $R(C'; A')$ 上的距离，也就是 $\mathbb{R}^h$ 中 $(m - h)$ 个无标记点的点云之间的瓶颈距离。因此，根据 [24, 定理 6.5]，$W_\infty(\xi(R(C; A)), R(C'; A'))$ 只需 $O(m^{1.5}\log^h m)$ 时间。对所有置换 $\xi \in S_h$ 取最小值，会使最终运行时间增加因子 $h!$ □

对于 $h = 1$ 和一个含 1 个点的子集 $A \subset C$，矩阵 $D(A)$ 为空，因此 $d(\xi) = W_\infty(\xi(R(C; A)), R(C'; A'))$。RDD 上的度量 M_∞ 将用作中间代价，再运用下文定义 5.3 和 5.4 中的标准工具，为 RDD 的无序集合（SDD）构造度量。

定义 *5.3*（线性指派代价 *LAC* [30]）。对于任意 $k \times k$ 代价矩阵 $c(i, j) \geq 0$, $i, j \in \{1, \ldots, k\}$, 线性指派代价 $LAC = \frac{1}{k}\min_g \sum_{i=1} c(i, g(i))$在所有指标 $1, \ldots, k$上的双射 g 中取最小值。

LAC 中的归一化因子 $\frac{1}{k}$使该度量更便于与权重总和为 1 的 EMD 进行比较。

定义 5.4（分布上的推土机距离）。设 $B = \{B_1, \ldots, B_k\}$ 是一个带有权重 $w(B_i)$、$i = 1$、……、k的有限无序对象集。考虑另一个集合

$D=\{D_1,\ldots,D_l\}$，其权重为 w（D_j）、 $j=1$、……、 l。假设 B_i与 D_j 之间的距离由度量 d（B_i， D_j）测量。从 B 到 D 的流是一个 $k\times l$ 矩阵，其中元素 $f_{ij}\in[0,1]$ 表示从对象 B_i到 D_j的部分流。推土机距离 [58]是满足k lEMD(B, D) $=\sum\sum f_{ij}d(B_i, D_j)$ 的 $f_{ij}\in[0,1]$ 子$i=1$ $j=1$ l k的最小值，并满足$\sum f_{ij}\le w$（B_i），其中 $i=1$、……、 k；$\sum f_{ij}\le w$（D_j） $j=1$ $i=1$ k l，其中 $j=1$、……、 l；以及$\sum\sum f_{ij}=1$。$i=1$ $j=1$

l第一个条件 $\sum f_{ij}\le w(B_i)$ 表示，通过流 f_{ij}，$j=1,\ldots,l$，流入所有D_j的对象 B_i 的权重 $w(B_i)$不会超过$j=1$。第二个条件 $\sum f_{ij}\le w(D_j)$ 表示，所有来自 B_i 的流 f_{ij} 都会流向$i=1,\ldots,k$，且总量不超过其权重 $w(D_j)$。最后一个条件 $\sum\sum f_{ij}=1$ 要求所有 B_i 共同“流入”所有 D_j。对于给定对象集合，LAC [30] 和 EMD [58] 可在近似立方时间内计算完成。

定理 5.5(b) 将 $O(m^{1.5}\log^n m)$ 算法从固定的 m 个无标记点的点云推广到 [24, 定理 6。5]中的等距类这一更难的情形，同时仍保持关于 m 的多项式时间复杂度，其中维数 n固定。

定理 5.5 (SDD上的度量的时间复杂度)。 对于各自在自身度量空间中的任意 m 个点云 C, C' 以及 $h\ge 1$，令单形距离分布 SDD$(C;h)$ 和SDD$(C';h)$ 由 $k=\binom{m}{h}$ 个权重相等的RDD构成，且不合并相同的RDD。

(a) 使用 $k\times k$ 由度量计算得到的代价矩阵M_∞，计算 RDD在 SDD$(C;h)$ 与 SDD$(C';h)$之间的线性指派代价 LAC 满足定义 *5.3* 中的所有度量公理，适用于 SDD，且可在 $O(h!(h^2+m^{1.5}\log^h m)k^2+k^3\log k)$ 时间内计算。

(b) 设 SDD$(C;h)$ 和 SDD$(C';h)$ 在合并相同的 RDD 后的最大规模为$l\le k$ EMD 满足定义 *5.4* 中的所有度量公理，适用于 SDD，且计算时间为 $O(h!(h^2+m^{1.5}\log^h m)l^2+l^3\log l)$。

证明。定义 5.3 中的线性指派代价（LAC）是对称的，因为任何双射匹配都可以反向进行。LAC 的三角不等式由引理 5.2 中度量 M_∞ 的三角不等式推出：通过复合双射 SDD$(C;h)\to$SDD$(C';h)\to$SDD$(C'';h)$，按照引理 5.2 证明中的类似方式匹配所有 RDD。LAC 的第一条度量公理意味着，LAC $=0$ 当且仅当存在一个双射g：SDD$(C;h)\to$SDD$(C';h)$，使得所有匹配的 RDD 之间的距离为 $M_\infty=0$，因此这些 RDD 等价（从而 SDD 相等），这是由 $M_\infty=0$ 的第一条公理得出的，该公理已在引理 5.2 中得到证明。

推土机距离（EMD）的度量公理由 [58] 的附录中证明，假设底层距离满足度量公理 d，它是引理 M_∞ 中的度量5.2 （在我们的情形下）。

LAC和EMD的时间复杂度可由引理 $O((h^2+m^{1.5}\log^h m)h!)$ 推出，其中 M_∞ 见引理5.2；再乘以成本矩阵规模的二次因子，并加上近三次时间 [30, 32]。 □

定理5.8 中SDD的利普希茨连续性需要引理5.7 由引理5.6中的部分情形推出。

引理 5.6。对于任意 a、 b、 c、 $d\in\mathbb{R}$，若 $a\le b$ 且 $c\le d$，则最大值$\{|a-c|,|b-d|\}\le$ 最大值$\{|a-d|,|b-c|\}$。

证明。 我们考虑直线 $\mathbb{R}$ 上点对 $a\le b$ 与 $c\le d$ 的相对位置的几种情况。
情形 $a\le b\le c\le d$。所需的不等式由 max $\{c-a,d-b\}\le d-a=\max\{d-a,c-b\}$ 得出。
情形 $a\le c\le b\le d$。所需的不等式由 max $\{c-a,d-b\}\le d-a=\max\{d-a,b-c\}$ 得出。
情形 $a\le c\le d\le b$。由于 $c-a\le d-a, b-d\le b-c$，不等式 $\max\{c-a,b-d\}\le\max\{d-a,b-c\}$ 成立。所有其他情形均可通过变换 $(a,b)\mapsto(-b,-a)$, $(c,d)\mapsto(-d,-c)$ 化归为上述情形，该变换保持给定条件和所需结论不变。 □

引理 5.7 (度量 L_∞ 保持排序不变)。 对于任意向量 $v=(v_1,\ldots,v_k)\in\mathbb{R}^k$，向量 $\vec{v}\in\mathbb{R}^k$ 由 v 将所有坐标按递增顺序排列得到。于是，对于任意向量 u、 $v\in\mathbb{R}^k$，都有 $|\vec{u}-\vec{v}|_\infty\le|u-v|_\infty$。

证明。 由于 $|u-v|_\infty=\max_{i=1,\ldots,k}|u_i-v_i|$，度量 L_∞在对 u,v 的坐标同时应用任意置换 $\xi\in S_k$ 后保持不变。因此，不失一般性，我们可以假设其中一个向量 u 的所有坐标已经按递增顺序排列，所以 $u=\vec{u}$。对于任意一对相邻坐标 $u_i\le u_{i+1}$，设其在 v 中对应的一对坐标顺序相反 $v_i>v_{i+1}$。

根据引理 5.6 交换 $v_i\leftrightarrow v_{i+1}$ 不会增大 L_∞ 之间的距离(u_i, u_{i+1}) 与 (v_i,v_{i+1})，因此也不会增大 $u,v\in\mathbb{R}^k$之间的距离。应用这类交换，可将 v 的所有坐标按递增顺序排列，且不会增大 L_∞。 □

定理5.8 大幅推广了这样一个事实：在两个点各自的 ε-邻域内扰动这两个点，其间的欧几里得距离至多改变 2ε。

定理 5.8 （SDD的利普希茨连续性）。在任意度量空间中，设 C' 是由点云 C 得到的，其中每个点都在其 ε-邻域内受到扰动。对于

任意 $h \geq 1$，$\mathrm{SDD}(C;\ h)$ 在LAC 和EMD 度量下的变化至多为 2ε 。下界成立：$\mathrm{EMD}\big(\mathrm{SDD}(C;\ h),\ \mathrm{SDD}(C';\ h)\big) \geq |\mathrm{SDM}(C;\ h,1) - \mathrm{SDM}(C';\ h,1)|_\infty$。

证明。将给定点云 C, C' 中的所有点排序，使每个点 $p_i \in C$ 与其扰动点 $p'_i \in C'$ 的索引相同。在给定度量空间中，C 中任意两点之间的距离 $d(p_i, p_j)$ 在扰动后至多变化 2ε，因此 $|d(p_i, p_j) - d(p'_i, p'_j)| \leq 2\varepsilon$。根据引理 5.2，这个上界 2ε 对最大度量 M_∞ 仍然成立；对于 LAC 和 EMD 度量，这也成立，因为定义 5.3 和 5.4 中所有代价的总权重均为 1。

引理5.7 表明，将向量的坐标按递增顺序重新排列不会增大度量 L_∞，因此$L_\infty(\xi(D(A))\ D(A')) \geq |\mathrm{SDV}(A) - \mathrm{SDV}(A')|_\infty$ 对任意指标置换 $\xi \in S_h$ 均成立，其中指标为 $1, \ldots, h$。

瓶颈距离 $W_\infty(\xi(R(C;A))\ R(C';\ A'))$ 是在矩阵 $h \times (m-h)$ $\xi(R(C;\ A))$ 和 $R(C';\ A')$ 中对应列向量之间计算出的度量 L_∞ 的最大值。设 $d = (d_1, \ldots, d_h)$ 和 $d' = (d'_1, \ldots, d'_h)$ 是这样两列。三角不等式 h 表明 $|d - d'|_\infty = \max_{i=1,\ldots,h} |d_i - d'_i| \geq \frac{1}{h} \sum_{i=1}^{h} |d_i - d'_i| \geq |\frac{1}{h} \sum d_i - \frac{1}{h} \sum d'_i|$。因此，对所有$i=1$向量坐标取平均值不会增大度量 L_∞。于是$W_\infty(\xi(R(C;\ A))\ R(C';A'))$ 的下界等于矩阵 $\xi(R(C;A))$ 和 $R(C';A')$ 中列平均值向量之间的度量 L_∞。将引理 5.7 应用于 $\mathbb{R}^{m-h}$ 中的这些向量可得 $W_\infty(\xi(R(C;\ A))\ R(C';\ A')) \geq |\vec{R}(C;\ A) - \vec{R}(C';\ A')|_\infty$

取上述度量 L_∞ 和 W_∞的最大值，我们得到以定义 3.4 中的平均距离分布表示的下界：$d(\xi) = \max\{L_\infty(\xi(D(A))\ D(A'))\ W_\infty(\xi(R(C;A))\ R(C';\ A'))\} \geq |\mathrm{ADD}(C;\ A) - \mathrm{ADD}(C';A')|_\infty$

由于上述论证适用于任意置换$\xi \in S_h$，我们得到 $M_\infty(\mathrm{RDD}(C;A),\ \mathrm{RDD}(C';A')) = \min d(\xi) \geq |\mathrm{ADD}(C;\ A) - \mathrm{ADD}(C';\ A')|_\infty$。

$\mathrm{SDD}(C;\ h)$ 和 $\mathrm{ASD}(C;h)$ 分别是 $\binom{m}{h}$ RDD 和向量的无序集合。若使用最优流矩阵 f_{ij} 来计算 $\mathrm{EMD}\big(\mathrm{SDD}(C;h),$ 并将定义 5.4 中的 $\mathrm{SDD}(C';\ )$ 用于计算 ASD 向量上的 EMD，我们便得到 $\mathrm{EMD}\big(\mathrm{ASD}(C;\ h), \mathrm{ASD}(C';\ h)\big)$ 的一个上界。该上界可能更小（对于另一个流矩阵而言），但不会更大，因此 $\mathrm{EMD}\big(\mathrm{SDD}(C;\ h), \mathrm{SDD}(C';\ h)\big) \geq \mathrm{EMD}\big(\mathrm{ASD}(C;\ h), \mathrm{ASD}(C';h)\big)$。将 $\mathrm{ASD}(S;\ h)$ 视为向量的加权分布，则 $\mathrm{SDM}(C;h;1)$ 是其在 [20, 第 3] 节中的质心。下界 $\mathrm{EMD}\big(\mathrm{ASD}(C;\ h), \mathrm{ASD}(C';h)\big) \geq |\mathrm{SDM}(C;\ h,\ 1) - \mathrm{SDM}(C';\ h,1)|_\infty$ 可由 [20, 定理 1] 推得。 □

6. 测度单纯形分布（MSD）

本节将 SDD 推广到度量测度空间。

定义 6.1（度量测度空间）。根据 *Gromov* [35, 节 $3\frac{1}{2}.1$]，度量测度空间（X，d_X，μ_X）是一个紧致空间 X，配有度量 d_X和博雷尔测度 μ_X，满足 $\mu_X(X) < +\infty$。度量测度空间之间的同构是保测度的等距变换 $f: X \to Y$，其意义是，对于任意子集 $U \subset Y$，都有 $\mu_Y(U) = \mu_X(f^{-1}(X))$。

将 $\mu_X(U)$ 除以总测度 $\mu_X(X) < +\infty$（对任意 $U \subset X$），我们可以假设 $\mu_X(X) = 1$，因此 μ_X 是一个概率测度。任何含有 m 个点的度量空间 X 都可以视为配备均匀测度 $\mu_X(p)$ 的度量测度空间，其中对所有点 $p \in X$，$= \frac{1}{m}$。

在欧几里得直线 $\mathbb{R}$ 上的两个点 0，1 上，度量测度空间 $X = (\{0,1\},\ 1,\ \{\frac{1}{2}, \frac{1}{2}\})$ 和 $Y = (\{0,1\},\ 1,\ \{\frac{1}{3}, \frac{2}{3}\})$是等距的，但不同构，因为权重不同。

定义6.2 将距离的局部分布从 [49, 定义 5.5] 推广到更高阶 $h > 1$。

定义 *6.2*（测度单纯形分布 *MSD*）。设 (X, d_X, μ_X) 为任意度量测度空间。对于任意基序列 $A = (p_1, \ldots, p_h) \in X^h$（由 $h \geq 1$ 个有序点组成），将定义 *3.1* 中的三角距离矩阵 $D(A)$ 按行写出，记为点间距离向量 $VID(A) \in \mathbb{R}_+^{h(h-1)/2}$，使得对于$k = h(i-1) + j - 1$，$1 \leq i < j \leq h$，$VID_k = d_X(p_i, p_j)$。对于距离阈值向量$\vec{d} = (d_1, \ldots, d_h) \in \mathbb{R}_+^h$，基于序列的测度向量 $VSM(A;\ \vec{d}) \in \mathbb{R}_+^h$ 由 h 个值 $\mu_X(\{q \in X:\ d_X(q, p_i) \leq d_i\})$ 组成，其中 $i = 1, \ldots, h$。阶数为 $h \geq 1$ 的测度单纯形分布是函数 $MSD[X;\ h]: X^h \times \mathbb{R}_+^h \to \mathbb{R}_+^{h(h+1)/2}$，它将任意 $A \in X^h$ 和 $\vec{d} \in \mathbb{R}_+^h$ 映射到有序对 $[VID(A), VSM(A;\ \vec{d})]$，并将其视为 $\mathbb{R}_+^{h(h+1)/2}$ 中的拼接向量。

对于 $h = 1$，向量$\mathrm{VID}(A)$ 为空，而阶数为 $h = 1$ 的测量单纯形分布与距离局部分布 [49,定义 5.5]$\mathrm{MSD}[X;\ 1]: X \times \mathbb{R}_+ \to \mathbb{R}_+$ 将任意点 $p \in X$和阈值 $d \in \mathbb{R}_+$映射到 $\mu_X(\{q \in X:\ d_X(q,\ p) \leq d\})$。

指标 $1, \ldots, h$ 上的任意置换 ξ都会自然地置换 $\mathrm{MSD}[X;h]$的各个分量。若 X 由 m 个点组成，则 $\mathrm{MSD}[X;\ h]$ 可化为由向量 $\mathrm{VID}(A)$ 与场 $\mathrm{VSM}(A;\ \vec{d})$ 配对构成的有限集合 $\binom{m}{h}: \mathbb{R}_+^h \to \mathbb{R}_+^h$仅针对无序的 h 点子集 $A \subset X$，之后可将其细化为下文所述、类似于 SDD 的更强不变量。

定义 6.3 (加权单纯形分布 WSD)。设 X 是一个有限度量测度空间，其中任意点 p 都具有权重$w(p)$。对于 $h \geq 1$和一个序列 $A = (p_1, \ldots, p_h) \in X^h$

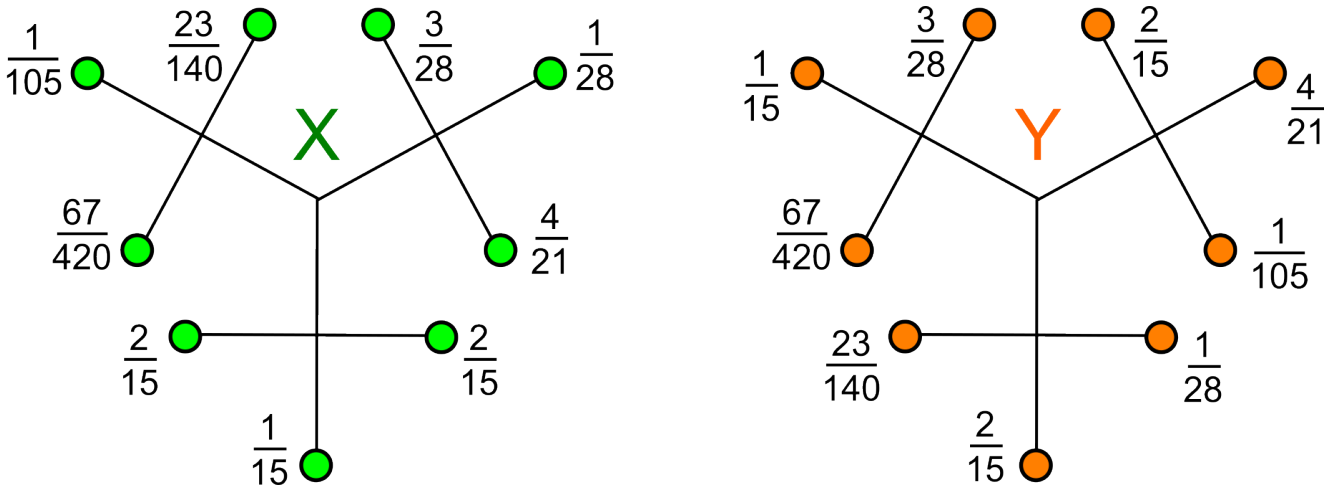

图 6. 非同构的度量测度空间 X、Y 来自 [49, 图 8] 的点间距离局部分布相同，但可由 n 区分加权单纯形分布 间 1阶分布与 2阶测度单纯形分布，详情见例 6.4。 ew

在定义 *3.1* 中，为 $D(A)$ 中的任意距离 $d(p,q)$ 配上权重的无序对 $w(p)$ $w(q)$。对于每个点$q \in X - A$，将权重 $w(q)$ 放入矩阵 $M(X;A)$ 的额外第 $(h+1)$ 行，该矩阵的列以无序 $q \in X - A$为索引。若 $h = 1$ 且 $A = p_1$，则令 $D(A) = w(p_1)$。加权距离分布 $WDD(X;A)$ 是二元组 $[D(A); M(X;A)]$ 在作用于 A 的排列 $\xi \in S_h$ 下的等价类。加权单纯形分布 $WSD(X)$ 是所有由无序 h 个点构成的子集 $A \subset X$ 对应的 $WDD(X;A)$ 的无序集合。

对于有限度量测度空间，可以通过结合权重和距离，在 WDD上定义一种度量，方式类似于引理 M_∞ 中的方法5.2 。然后可以按照定理 LAC 和 EMD 中的定义 5.3和 5.4 计算，见定理 5.5。

例 6.4 (WSD 的强度WSD)。 图6展示了由 *9* 个点组成的度量测度空间 X、Y，它们以树的形式呈现 *[49, 图 8]*。所有边的长度均为 $1\,2$，并导出最短路径度量 d_X, d_Y。每个由 *3* 个节点构成的小分支中的权重之和均为 $1\,3$。这些度量测度空间 X, Y的所有点间距离都只有 *1* 和 *2*，且距离的局部分布相同$\mathrm{MSD}[X;\ 1] = \mathrm{MSD}[Y;\ 1]$ 依据 *[49, 例 5.6]*。

事实上，这两个MSD都可以看作由相同的 *9* 个分段常数函数组成的集合 $\mu(p)$，它们分别在区间 $[0,1)$、$[1,2)$、$[2,+\infty)$上取值 $w(p)$、$1\,3$ 和 1。

然而，WSD包含更多逐点数据：$\mathrm{WSD}[X;1]$具有 A $(D) = w(p) = \frac{23}{140}$ 以及下列矩阵 2×8

$$M(X;p) = \begin{pmatrix} 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 \\ \frac{1}{105} & \frac{67}{240} & \frac{2}{15} & \frac{1}{15} & \frac{2}{15} & \frac{4}{21} & \frac{1}{28} & \frac{3}{28} \end{pmatrix},$$

但 $\mathrm{WSD}[Y;\ 1]$ 对于 $w(p) = \frac{23}{140}$有另一个矩阵。

$$M(Y;p) = \begin{pmatrix} 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 \\ \frac{2}{15} & \frac{1}{28} & \frac{1}{105} & \frac{4}{21} & \frac{2}{15} & \frac{3}{28} & \frac{1}{15} & \frac{67}{420} \end{pmatrix}.$$

列可任意置换的矩阵并不相同，因此 X、Y在 $h = 1$ 时可由WSD 区分

此外，$MSD[X;\ 2] \neq MSD[Y;\ 2]$ 因为对于任意基序列 $A = (p,q) \in X^2$，我们有 $VSM[X;\ 2](A;\ d_1, d_2) = (w(p)\ w(q))$ 对于 $d_1, d_2 < 1$，因为所有其他点到 p, q的最小距离均为 *1*，Y也同理。唯一的点 p, q 的权重分别为 $w(p) = \frac{23}{140}$和 $w(q) = \frac{67}{420}$，它们的距离 $d_X(p,q) = 1$和$d_Y(p,q) = 2$不同。于是，$MSD[X;\ 2] \neq MSD[Y;\ 2]$ 在将可唯一识别的场映射 $[0,1)^2$ 到常向量 $(w(p)\ w(q))$ 时有所不同，其中 $VID_X(A) = 1 \neq 2 = VID_Y(A)$。

我们猜想，对于足够大的 h，任意度量测度空间 X, Y 都能通过测度单纯形分布在同构意义下区分开来，其中 h 取决于 X, Y。

本文的后续更新将介绍度量测度空间上的测度单纯形分布之间的连续度量。我们欢迎新的想法与合作。

定义 3.3、6.2 及主要定理 3.6、5.5、5.8 中的新不变量，实质上推动了几何数据科学 这一新领域的发展，旨在解决所有瓶颈类似于问题 1.1的数据挑战。

此前的研究探讨了问题1.1中的以下重要情形：1-周期离散序列 [5,6,43], 2D 格点 [13, 45], 3D 格点 [12, 42, 44, 51], 中的周期点集 $\mathbb{R}^3$ [23, 63] 以及高维空间中的周期点集 [2–4]。

几何数据科学在晶态材料中的应用 [7, 57, 67, 76] 促成了晶体等距原理 [70, 71, 73]，将门捷列夫化学元素周期表扩展到由完全不变量连续参数化的所有周期晶体的晶体等距空间 。

本研究获得英国皇家工程院奖学金“面向下一代固态晶态材料工程的数据科学”（2021年-2023年，IF2122/186），以及EPSRC资助项目“应用驱动的拓扑数据分析”（2018年-2023年，EP/R018472/1）和“周期晶体的逆向设计”（2022年-2024年，EP/X018474/1）的支持。

作者感谢英国利物浦材料创新工厂数据科学理论与应用小组的所有成员，尤其是 Daniel Widdowson、MatthewBright、YuryElkin 和 Olga Anosova；也感谢 Justin Solomon（MIT）、Steven Gortler（哈佛大学）和 Nadav Dym（Technion）进行富有成效的讨论，并感谢所有审稿人付出的宝贵时间和提出的有益建议。

参考文献

[1] Naveed Akhtar 和 Ajmal Mian。《计算机视觉中深度学习面临的对抗攻击威胁：综述》。IEEE Access，6：14410–14430，2018年。2[2] Olga Anosova 和 Vitaliy Kurlin。《周期几何与拓扑导论》。arXiv:2103.02749。13[3] Olga Anosova 和 Vitaliy Kurlin。《周期点集的等距分类》。载于《离散几何与数学形态学会议论文集》，2021年。13[4] Olga Anosova 和 Vitaliy Kurlin。《周期晶体上的连续度量算法》。arxiv:2205.15298，2022年。2，13[5] Olga Anosova 和 Vitaliy Kurlin。《周期序列的密度函数》。《计算机科学讲义》（DGMM 会议论文集），2022年。13[6] O Anosova 和 V Kurlin。《连续事件周期序列的密度函数》。arXiv:2301.05137，2023年。13[7] Jonathan Balasingham、Viktor Zamaraev 和 Vitaliy Kurlin。《利用逐点距离分布表示晶体的紧凑图表示》。arXiv:2212.11246，2022年。13[8] Serge Belongie、Jitendra Malik 和 Jan Puzicha。《利用形状上下文进行形状匹配与对象识别》。模式分析与机器智能汇刊，24(4)：509–522，2002年。2[9] Mireille Boutin 和 Gregor Kemper。《从距离或面积分布重建 n 点构形》。Adv. Appl. Math.，32(4)：709–735，2004年。4[10] Peter Brass 和 Christian Knauer。《检验 d 维点集的全等性》。载于计算几何研讨会，页 310–314，2000年。2[11] Peter Brass 和 Christian Knauer。《检验一般三维对象的全等性与对称性》。计算几何，27(1)：3–11，2004年。2[12] Matthew J Bright、Andrew I Cooper 和 Vitaliy A Kurlin。《欢迎来到三维格的连续世界》。arxiv:2109.11538，2021年。13[13] Matthew J Bright、Andrew I Cooper 和 Vitaliy A Kurlin。《二维格的地理风格地图》。Acta Crystallographica Section A，79(1)：1–13，2023年。13[14] Michael M Bronstein、Joan Bruna、Taco Cohen 和 Petar Veli ˇckovi´c。《几何深度学习：网格、群、图、测地线与规范》。arXiv:2104.13478，2021年。2[15] Michael M Bronstein、Joan Bruna、Yann LeCun、Arthur Szlam 和 Pierre Vandergheynst。几何深度学习：超越欧几里得数据。IEEE Signal Processing Mag-azine，34(4):18–42，2017年。2[16] Haiwei Chen、Shichen Liu、Weikai Chen、Hao Li 和 Randall Hill。用于三维点云分析的等变点网络。载于 CVPR，第 14514–14523 页，2021年。2[17] Paul Chew、Dorit Dor、Alon Efrat 和 Klara Kedem。d维空间中的几何模式匹配。《离散与计算几何》，21(2):257–274，1999。2[18] Paul Chew、Michael Goodrich、Daniel Huttenlocher、Klara Kedem、Jon Kleinberg 和 Dina Kravets。欧几里得运动下的几何模式匹配。《计算几何》，7(1-2):113–124，1997。2[19] Paul Chew 和 Klara Kedem。几何模式匹配问题的改进。载于《斯堪的纳维亚算法理论研讨会》，第 318–325 页，1992。2[20] Scott Cohen 和 Leonidas Guibas。地球移动距离：下界及其在平移下的不变性。技术报告，斯坦福大学，1997。12[21] Matthew J Colbrook、Vegard Antun 和 Anders C Hansen。计算稳定且准确的神经网络之难：深度学习的障碍与 Smale第18问题。《美国国家科学院院刊》，119(12):e2107151119，2022年。2[22] Yinpeng Dong、Fangzhou Liao、Tianyu Pang、Hang Su、Jun Zhu、Xiaolin Hu 和 Jianguo Li。利用动量增强对抗攻击。载于《计算机视觉与模式识别》，第 9185–9193 页，2018年。2[23] H Edelsbrunner、T Heiss、V Kurlin、P Smith 和 M Wintraecken。周期点集的密度指纹。载于计算几何研讨会论文集，第 32:1–32:16 页，2021年。13[24] Alon Efrat、Alon Itai 和 Matthew J Katz。几何学有助于瓶颈匹配及相关问题。Algorithmica，31(1):1–28，2001。10, 11[25] Yury Elkin。用于 k 近邻搜索的新型压缩覆盖树（博士论文）。arxiv:2205.10194，2022年。3[26] Y. Elkin 和 V. Kurlin。树状图的合并图及其稳定性。载于数学基础、计算理论与软件研讨会论文集，2020年。2[27] Y. Elkin 和 V. Kurlin。通过扩展零维持久性的合并图，对射影扰动点云进行等距不变形状识别。《数学》，9(17)，2021年。2[28] H Pottmann 等。用于鲁棒几何处理的积分不变量。《计算机辅助几何设计》，26(1):37–60，2009年。2[29] S Manay 等。用于形状匹配的积分不变量。《模式分析与机器智能汇刊》，28:1602–1618，2006年。2[30] Michael L Fredman 和 Robert Endre Tarjan。斐波那契堆及其在改进网络优化算法中的应用。《ACM学报》，34:596–615，1987年。10, 11[31] EN Gilbert 和 LA Shepp。用于判别实验的纹理，1974年。2[32] Andrew Goldberg 和 Robert Tarjan。通过逐次逼近求解最小费用流问题。载于计算理论研讨会论文集，第 7–18 页，1987年。11[33] Michael T Goodrich、Joseph SB Mitchell 和 Mark W Orletsky。刚体运动下的近似几何模式匹配。《模式分析与机器智能汇刊》，21:371–379，1999年。2

[34] Cosmin Grigorescu 和 Nicolai Petkov。用于形状滤波器和形状识别的距离集。《IEEE图像处理汇刊》，12(10):1274–1286，2003年。2[35] Mikhael Gromov、Misha Katz、Pierre Pansu 和 Stephen Semmes。黎曼与非黎曼空间的度量结构，第152卷。施普林格，1999年。12[36] Chuan Guo、Jacob Gardner、Yurong You、Andrew Gordon Wilson 和 Kilian Weinberger。简单的黑盒对抗攻击。载于国际机器学习大会，第 2484 2493, 2019 2 – 页。Felix [37] Hausdorff。维数与外测度。《数学年刊》，79(2):157–179，1919年。2[38]Snir Hordan、Tal Amir、Steven J Gortler 和 Nadav Dym。用于欧几里得图的完全神经网络。arXiv:2301.13821，2023年。2[39] Daniel Huttenlocher、Gregory Klanderman 和 William Rucklidge。使用豪斯多夫距离比较图像。《模式分析与机器智能汇刊》，15:850–863，1993年。2[40] Ernest Sydney Keeping。《统计推断导论》。Courier Corporation，1995年。3[41] Heuna Kim 和 G¨unter Rote。四维点集的全等性检验。arXiv:1603.07269，2016年。2[42] Vitaliy Kurlin。三维格的完备等距分类。arxiv:2201.10543，2022年。13[43] Vitaliy Kurlin。欧几里得等距变换下有限无标记点云的可计算完全不变量。arXiv:2207.08502，2022年。2, 13[44] Vitaliy Kurlin。有限序列和1周期序列等距类上可精确计算且连续的度量。arxiv:2205.04388，2022年。13[45] Vitaliy Kurlin。二维格的数学。《计算数学基础》，第 Dec 7 1 59, 2022 13 页：–。[46] Vitaliy Kurlin。单纯形的强度是对无标记欧几里得点云进行连续等距分类的关键。arXiv: *2303 13486*, 2023 3 。Cassidy Laidlaw [47] 和 Soheil Feizi。函数型对抗攻击。《神经信息处理系统进展》，32，2019年。2[48] Sushovan Majhi、Jeffrey Vitter 和 Carola Wenk。欧几里得空间中格罗莫夫-豪斯多夫距离的近似计算。arXiv:1912.13008，2019年。2[49] Facundo M´emoli。格罗莫夫-瓦瑟斯坦距离与对象匹配的度量方法。《计算数学基础》，11(4):417–487，2011年。2, 12, 13[50] Facundo M´emoli、Zane Smith 和 Zhengchao Wan。超度量空间之间的格罗莫夫-豪斯多夫距离：其结构与计算。arXiv: *2110 03136*, 2021 2 。Marco [51] M Mosca 和 Vitaliy Kurlin。基于 Voronoi 图的任意晶体格之间的相似性距离。《晶体研究与技术》，55 5 1900197, 2020 13 ( ):。Jigyasa [52] Nigam、Michael J Willatt 和 Michele Ceriotti。分子哈密顿量及n中心原子尺度性质的等变表示。《化学物理学杂志》，156(1):014115，2022年。2[53] Robert 等人 Osada。形状分布。《图形学汇刊》，21:807–832，2002年。2[54] Sergey N Pozdnyakov、Michael J Willatt、Albert P Bart´ok、Christoph Ortner、G´abor Cs´anyi 和 Michele Ceriotti。原子结构表示的不完备性。《物理评论快报》，125:166001，2020年。4, 5, 6, 7, 8

[55] Charles Ruizhongtai Qi、Li Yi、Hao Su 和 Leonidas J Guibas。Pointnet++：在度量空间中的点集上进行深度分层特征学习。《神经信息处理系统进展》，30，2017。2[56] Blaine Rister、Mark A Horowitz 和 Daniel L Rubin。基于不变关键点的体积图像配准。《图像处理汇刊》，26(10):4900–4910，2017。2[57] Jakob Ropers、Marco M Mosca、Olga D Anosova、Vitaliy A Kurlin 和 Andrew I Cooper。通过晶体结构的连续等距不变量快速预测晶格能量。载于《数据密集型领域数据分析与管理国际会议》，页码 178 192, 2022 13 – 。[58] Y. Rubner、C. Tomasi 和 L. Guibas。推土机距离作为图像检索的度量。《国际计算机视觉期刊》，40(2):99–121，2000。11[59] Felix Schmiedl。格罗莫夫-豪斯多夫距离的计算方面及其在非刚性形状匹配中的应用。《离散与计算几何》，57:854–880，2017。2[60] Isaac Schoenberg。对 Maurice Frechet 的文章《关于一类可向量映射到希尔伯特空间的距离空间的公理化定义》的评论。《数学年刊》，页码 724–732，1935年。2[61]Anthony Simeonov、Yilun Du、Andrea Tagliasacchi、Joshua B Tenenbaum、Alberto Rodriguez、Pulkit Agrawal 和 Vincent Sitzmann。神经描述符场：用于操作的 SE(3) 等变物体表示。载于 ICRA，页码 6394 6400, 2022 2 – 。Philip Smith 和 Vitaliy Kurlin。[62] 具有相同 1 *2202 00577*, 2022 2 d 持续性的点集族，arxiv：。。Phil Smith 和 Vitaliy Kurlin。三维周期点集 k 阶沃罗诺伊域的实用算法。载于《计算机科学讲义》（ISVC 会议论文集），第 13599, 377 391, 2022 13 卷，页码 –[63] 。Riccardo Spezialetti 、 Samuele Salti 和 Luigi Di Stefano。无需监督地学习有效的等变三维描述符。载于 ICCV，页码 6401 6410, 2019 2 – 。Jian Sun、[64] Maks Ovsjanikov 和 Leonidas Guibas。一种基于热扩散的简洁且可证明具有信息性的多尺度签名。《计算机图形论坛》，28 1383 1392, 2009 2 : – 。[65] Matthew Toews 和 William M Wells III。使用三维尺度不变特征进行高效、稳健的模型到图像对齐。《医学图像分析》，17(3):271–282，2013年。2[66] Aikaterini Vriza、Ioana Sovago、Daniel Widdowson、Peter Wood、Vitaliy Kurlin 和 Matthew Dyer。分子集合变换器：关注剑桥结构数据库中的共晶。《数字发现》，1 834 850, 2022 13 : – 。[67] E Weisstein。三角形。https://mathworld. wolfram. com。[68]

[69]

Carola Wenk。高维形状匹配。博士论文，柏林自由大学，2003年。2[70] D Widdowson 和 V Kurlin。周期点集的逐点距离分布。arxiv: *2108 04798*, 2021 2, 13 。。[71] Daniel Widdowson 和 Vitaliy Kurlin。消解周期晶体的数据歧义。《神经信息处理系统进展》（神经信息处理系统大会），35，2022年。2、3、13[72] Daniel Widdowson 和 Vitaliy Kurlin。通过完备且连续的等距不变量识别无标签点云中的刚性模式，且无假阴性和假阳性。载于 CVPR 论文集，2023年。2

[73] Daniel Widdowson、Marco M Mosca、Angeles Pulido、Andrew I Cooper 和 Vitaliy Kurlin。周期点集的平均最小距离——映射所有周期晶体的基础不变量。《*MATCH* 数学与计算化学通讯》，87:529–559，2022年。3，13[74] Manzil Zaheer、Satwik Kottur、Siamak Ravanbakhsh、Barnabas Poczos、Russ R Salakhutdinov 和 Alexander J Smola。深度集合。《神经信息处理系统进展》，30，2017年。2[75] N Zava。格罗莫夫-豪斯多夫空间无法粗嵌入任何希尔伯特空间。*arXiv:2303.04730*，2023年。2[76] Q Zhu、J Johal、D Widdowson、Z Pang、B Li、C Kane、V Kurlin、G Day、M Little 和 A Cooper。由预测与结构不变量驱动的类比：计算引导发现介孔氢键有机笼晶体。《美国化学会志》，144:9893–9901，2022年。13[77] Wen Zhu、Lingchao Chen、Beiping Hou、Weihan Li、Tianliang Chen 和 Shixiong Liang。基于尺度不变点特征直方图的避雷器点云配准。《科学报告》，12(1):1–13，2022年。2