arXiv:2303.06118v1 [math.AT] 10 Mar 2023

通过有根子集分解零维持久模

´Ángel Javier Alonso 和 Michael Kerber

Abstract. 我们研究零维持久模的分解问题。零维持久模可视为取值于向量空间范畴、且经由集合分解的函子。我们不直接在向量空间层面开展研究，而是退一步，先考察集合层面的分解问题。

这种方法使我们能够定义 *rooted subsets*这一组合概念。对于滤过度量空间 M，有根子集将 M中各点的聚类行为与相应持久模的分解联系起来。特别是，我们可以快速识别这类分解中的区间。此外，有根子集可以理解为长者规则的一种推广，并且也与 Cai、Kim、Mémoli 和 Wang 提出的常值征服者概念有关。作为应用，我们给出欧几里得空间中密度-Rips 过滤的零维持久模分解中区间数量的一个下界：在极限情况下，并且在非常一般的条件下，我们可以预期至少 25% 的不可分解直和项是区间模。

1. Introduction

多参数持久同调是拓扑数据分析中的一个活跃研究领域。其动机在于，在许多数据集中，多尺度分析中有多个值得关注的参数 [11, 18, 32]。具体来说，在分析点云时，我们希望考察点与点之间的距离，也可能需要移除密度较低的点。

持久同调的核心对象是持久模，它以代数方式追踪数据的拓扑特征如何随着我们在参数空间中移动而变化。在单参数情形下，每个持久模都可分解为一组区间，称为持久条形码 [20]，其中每个区间表示数据中某个拓扑特征的持续时间。在多参数情形中，区间有一种广义概念，同样表示某个拓扑特征的持续时间；然而，多参数持久模并非总能分解为区间，因此可能会出现非区间型不可分解持久模，从而在理论上 [12, 13, 18, 33]和计算上 [1, 4, 23]带来复杂问题。

事实上，这类不可分解持久模的分类被认为难以企及：即使考虑某些简化情形，某些相关偏序集仍属于野表示型 [2]。此外，简单的几何构造 [13], 就能实现无穷多个复杂的不可分解持久模族；最近， [3] 还证明了，在交错度量下，多参数持久模通常接近不可分解（精确表述请参见 [3] ）。

尽管如此，前述复杂性并不意味着实际中出现的持久模接近不可分解，也不意味着它们无法分解为区间。事实上，多参数持久模的分解在实践中的表现，是否真的如理论所预期的那样糟糕？ [2]的作者和 [3] 的作者提出了类似的问题。

Institute of Geometry, Graz University of Technology, Austria.
E-mail addresses: alonsohernandez@tugraz.at, kerber@tugraz.at。

TABLE 1. 密度-Rips过滤中零维持久模分解的区间数量。我们尝试了两种样本：点根据围绕 5 个峰的多元高斯分布进行采样的簇集样本，以及单位正方形中的均匀样本。密度参数通过高斯核密度估计（kde）计算，或指定一个随机密度。该表显示了 5 次独立测试运行的区间数量；对于 n 点，如果区间数量为n，则该模是区间可分解的。这只发生在一次运行中。

		100 个点					500 个点				
样本	密度	运行 1	2	3	4	5	运行 1	2	3	4	5
簇状	kde	100	98	95	98	98	474	487	478	479	479
一致的	kde	**88**	**88**	**86**	**88**	**86**	**444**	**447**	**433**	453	457
成簇的	随机	77	86	87	88	76	397	381	390	380	386
一致的	随机	76	79	75	75	70	**376**	**361**	**366**	**355**	**377**

作为初步测试，我们使用 Dey 和 Xin 算法的一个原型实现，计算了一个标准零维构造中的持久模分解 [23]（该实现将在另一篇论文中讨论）。如表1所示，至少在这种情况下，持久模可以完全分解为区间这一假设似乎大多数时候都是错误的。然而，表1 还显示，在所有测试实例中，大多数不可分解直和项确实都是区间。

这不禁让人思考：一般情况下，我们能否有充分理由预期会出现许多区间？此外，了解这些区间可以极大地简化并加快持久模的计算任务：例如，分析二参数持久模的一种常见方法是考察一维限制，即所谓的切片，从而得到持久条形码的参数化族 [29,32,34,35]。二维持久模中的每个区间与切片相交后，都会在该切片的条形码中产生一条。因此，了解这些区间后，现有算法便可以专注于问题中非区间的“核心”部分，而这部分通常规模小得多。

所述方法在实践中的问题在于，尽管人们一直在努力 [23]，分解多参数持久模仍然代价高昂。然而，要利用区间方面的知识，无须计算完全分解，甚至无须识别所有区间。只要有一种能够快速从持久模中“剥离”区间的方法就足够了。因此，我们提出这样一个问题：是否存在一些方法，既能在实践中运行得非常快，又能检测出许多区间？

贡献。我们聚焦于零维持久模这一情形。仅这一情形就具有实际意义，因为它与层次聚类方法相关（见下文“相关工作”一节），并且近期已受到关注 [2,10,14,35]。在此背景下，我们对上述问题作出了一些回答：

对于点云 M，最近邻对 是指一对 $(x,\ y) \in M \times M$，满足 y是 x 的最近邻，且 x 是 y 的最近邻（使用固定全序打破平局）。我们发展的理论表明，对于密度-Rips 双滤过的零维持久模（无论采用何种密度估计函数），其区间数至少与 M中的最近邻对数相同。这些区间可根据最近邻对轻松确定，我们称之为*NN* 区间。由于所有最近邻对都能在 $O(n \log\ n)$ 时间内计算出来 [19,38],，因此这提供了一种快速计算所有 NN 区间的方法

分解中的区间。此外，我们可以预期会有许多 NN 区间：利用关于最近邻图的已有结果，我们证明，如果 M 是从任意一个几乎连续密度函数中独立采样得到的，那么分解中的直和项至少有四分之一是形如 $n \to \infty$的区间。据我们所知，这是首个证明分解中区间数具有非恒定下界的结果。

为得出这一结果，我们采用以下核心思路：我们不直接在（分级）向量空间范畴中研究持久模的分解，而是在持久集范畴中进行研究，其对象可以解释为双参数层次聚类。持久模的分解由其幂等自同态决定，因此我们寻找的不是持久模的幂等自同态，而是更简单的持久集的幂等自同态。我们证明，这类幂等自同态可以转化为有根子集，即在层次聚类中始终与某个固定点合并的点集。此外，只有一个元素的有根子集对应于相关持久模中的区间。

我们不从持久模中逐个剥离区间，而是从持久集中逐个剥离有根子集。这样做的好处是，剩余结构仍然是层次聚类，因此这一过程可以迭代进行。

相关工作。本文研究的多参数持久集合和零维持久模，与 Carlsson 和 M´emoli [16]最初探讨的聚类问题的多参数方法有关。通过公理化方法论证了采用多个参数——密度和尺度——的必要性 [15, 17, 28]。近年来，将多参数持久同调中的技术（如持久模和交错）应用于这一问题，已引起人们的关注 [2, 14, 32, 34, 35]。

Cai、Kim、M´emoli 和 Wang [14]为由 density-Rips 得到的零维持久模定义了一种有用的概括，称为 长者规则阶梯码，其灵感来自长者规则 [24]。他们还引入了相关概念“常值征服者”，并提出一个问题：常值征服者是否会在相关持久模的分解中诱导出一个区间。我们通过示例 4.3给出了否定答案；相反，我们证明了本文引入的 有根生成元确实会在分解中诱导出一个区间（推论 3.7）。

Brodzki、Burfitt 和 Pirashvili [10] 也研究了零维持久模的分解。他们确定了一类称为半分量模的持久模，这类模可能作为直和项出现在零维模的分解中，但仍然难以分类。他们的方法给了我们很大的启发，在定理 5.1 中，我们在所发展的理论框架内给出了他们一个定理的另一种证明。

致谢。作者感谢 Jan Jendrysiak 提供有益的讨论。我们也感谢匿名审稿人认真审阅稿件，并提出详细的意见和建议。本研究得到奥地利科学基金（FWF）资助，项目编号为 P 33765-N。

2. 预备知识

持久集与持久模。下文中，令 P 为一个有限偏序集，并将其视为一个范畴。一个**持久模**（在 P 上）是一个从 P 到有限维向量空间范畴 Vec 的函子，其中向量空间定义在固定域 K 上。这样的函子 $F\colon P \to \mathrm{Vec}$ 为每个**等级** $p \in P$ 关联一个有限维向量空间 F_p，并为 P 中的每个态射$p \leq q$ 关联一个线性映射 $F_{p\to q}\colon F_p \to F_q$，且满足 $F_{p\to p} = \mathrm{id}$ 和复合运算保持不变。我们将持久模视为函子范畴 Vec^P的对象，其中自然变换是态射。从这个意义上说，持久模之间的态射 $f\colon F \to G$

模是由映射 $\{f_p\colon F_p \to G_p\}_{p\in P}$ 组成的族，使得对于任意两个 $p \le q$，下图都交换

$$\begin{array}{ccc} F_p & \xrightarrow{F_{p\to q}} & F_q \\ \downarrow{\scriptstyle f_p} & & \downarrow{\scriptstyle f_q} \\ G_p & \xrightarrow{G_{p\to q}} & G_q. \end{array}$$

类似地，**持久集**（定义在 P 上）是一个从 P 到 Set 的函子；Set 是有限集范畴，而持久集之间的态射则是如上所述的自然变换。

通过应用 **线性化函子** $\mathrm{Set} \to \mathrm{Vec}$，可以从持久集得到持久模。该函子将每个集合映射到由其生成的自由向量空间。这个线性化函子通过后复合诱导出函子 $\mathcal{L}\colon \mathrm{Set}^P \to \mathrm{Vec}^P$。

从几何到持久集。设 (M,d) 是一个有限度量空间，并考虑函数 $f\colon M \to \mathbb{R}$。我们可以将 f 理解为给 M 中的每个点赋予一个密度；也就是说，它是一个密度估计函数 [37]。我们假设 f 对密度较高的点赋予较低的值。遵循 [14],，我们将三元组 (M,d,f) 称为增广度量空间。我们构造一个持久集，即 (M,d,f) 的密度-Rips 持久集，用以追踪当密度和尺度参数发生变化时，M 中各点的聚类如何变化；这一点我们很快会给出精确定义。

首先，对于固定的尺度参数 $\varepsilon \ge 0$，我们定义 M ε 处的**几何图**，记作 $\mathcal{G}_\varepsilon(M)$。它是顶点集为 M、边为 $(x,\ y)$ 的无向图，其中 $d(x,\ y) \le \varepsilon$。当 ε 从 0 变化到 ∞ 时，$\mathcal{G}_\varepsilon(M)$ 的连通分支构成了通过单链接聚类方法得到的树状图中的簇。

为引入密度，对于每个 $\sigma \in \mathbb{R}$，我们令 $M_\sigma := \{x \in M \mid f(x) \le \sigma\} \subseteq M$ 表示由（余）密度低于 σ 的点组成的度量子空间。对于任意两个 $\sigma \le \sigma'$、$M_\sigma \subseteq M_{\sigma'}$，将每个 $(\varepsilon,\ \sigma)$ 映射到图 $\mathcal{G}_\varepsilon(M_\sigma)$，即可得到函子 $\mathcal{G}(M,\ f)\colon \mathbb{R}_{\ge 0}\times\mathbb{R} \to \mathrm{Graph}$，其中 $\mathbb{R}_{\ge 0}\times\mathbb{R}$ 中的序关系定义为：当且仅当 $\varepsilon \le \varepsilon'$ 和 $\sigma \le \sigma'$ 时，$(\varepsilon,\ \sigma) \le (\varepsilon',\ \sigma')$。接着，我们考虑**连通分支函子** $\pi_0\colon \mathrm{Graph} \to \mathrm{Set}$，它将每个图映射到其连通分支集合。由此，我们得到函子 $\pi_0 \circ \mathcal{G}(M,\ f)\colon \mathbb{R}_{\ge 0}\times\mathbb{R} \to \mathrm{Set}$。

注记 2.1. 线性化持久模 $\mathcal{L}(\pi_0 \circ \mathcal{G}(M,f))\colon \mathbb{R}_{\ge 0}\times\mathbb{R} \to$ 向量空间范畴同构于在图层级应用零维同调所得到的持久模 $H_0 \circ \mathcal{G}(M,f)\colon \mathbb{R}_{\ge 0}\times\mathbb{R} \to$ 向量空间范畴。从这个意义上说，我们所描述的构造是密度-Rips 过滤的零维层级，这是多参数持久同调中的标准做法（参见 [5,18] 以及 [14]）。

我们可以如下理解函子 $\pi_0 \circ \mathcal{G}(M,f)\colon \mathbb{R}_{\ge 0}\times\mathbb{R} \to \mathrm{Set}$：它是一个由有限网格 $P \subseteq \mathbb{R}_{\ge 0}\times\mathbb{R}$ 索引的持久集 $S\colon P \to \mathrm{Set}$。考虑距离集 $D := \{d(x,y) \mid x,y \in M\}$ 以及密度集 $T := \{f(x) \mid x \in M\}$，并定义有限网格 $P := D \times T \subset \mathbb{R}_{\ge 0}\times\mathbb{R}$。最后，将每个 $(\varepsilon,\sigma) \in P$ 映射为 $(\pi_0 \circ \mathcal{G}(M,f))(\varepsilon,\sigma)$，从而定义持久集 $S\colon P \to \mathrm{Set}$；态射也以类似方式定义。

定义 2.2. 设 (M,d,f) 为一个增广度量空间。我们将其**密度-Rips 持久集** 定义为函子 $S\colon P \to \mathrm{Set}$，构造方式如上。

持久模的分解。我们可以通过分解来研究持久模。对于两个持久模 F 和 G，它们的直和 $F \oplus G$ 是逐点取直和得到的持久模，$(F \oplus G)_p = F_p \oplus G_p$。如果 $F \cong F_1 \oplus F_2$ 蕴含 $F_1 = 0$ 或 $F_2 = 0$，则称持久模为**不可分解**。由于持久模是实际的模（例如参见 [8, 引理 2.1]），根据 Krull-Schmidt 定理，持久模 $F = F_1 \oplus F_2 \oplus \ldots F_n$ 分解为不可分解直和项的方式在直和项的排列和同构意义下是唯一的。

令 I 是偏序集 P 的一个非空连通子偏序集，满足：对于任意两个 i、$j \in I$ 以及任意$l \in P$，若 $i \le l \le j$，则 $l \in I$。以**为支撑的区间模** I，$\mathcal{I}(I)$：$P \to$向量空间范畴是由下式给出的不可分解（例如，参见 [7, 命题 2.2]）持久模

$$\mathcal{I}(I)_p = \begin{cases} K, & \text{if } p \in I, \\ 0, & \text{otherwise,} \end{cases} \quad \text{with internal maps} \quad \mathcal{I}(I)_{p\to q} = \begin{cases} \mathrm{id}, & \text{if } p, q \in I, \\ 0, & \text{otherwise.} \end{cases}$$

如果 P 是全序集，那么每个定义在 P 上的持久模都可分解为区间模的直和 [6], ，但对于其他偏序集，一般不存在如此优美的分解。

分解与自同态。持久模的直和 $X = X_1 \oplus X_2$ 在同构意义下由态射 $\iota_i \colon X_i \to X$ 和 $\pi_i \colon X \to X_i$（其中 $i = 1, 2$）刻画，且满足 $\pi_i \circ \iota_i = \mathrm{id}_{X_i}$ 和 $\iota_1 \circ \pi_1 + \iota_2 \circ \pi_2 = \mathrm{id}_X$（例如参见 [31]）。在这种情况下，对于每个 $i = 1, 2$，映射 ι_i 和 π_i 诱导出 X 的一个自同态 $\varphi_i := \iota_i \circ \pi_i$。这样的自同态 $\varphi_i \colon X \xrightarrow{\pi_i} X_i \xrightarrow{\iota_i} X$ 也可分裂：

定义 2.3. 在任意范畴中，若自同态 $\varphi \colon X \to X$ 存在一个对象 Y和一个分解 $\varphi \colon X \xrightarrow{\pi} Y \xrightarrow{\iota} X$，使得 $\pi \circ \iota = \mathrm{id}_Y$，则称其为**可分裂的**。

我们将使用以下关于分裂自同态的标准事实：

引理 2.4. 设 $\varphi \colon X \to X$ 是一个可分裂自同态，有两个分解 $X \xrightarrow{\pi} Y \xrightarrow{\iota} X$和 $X \xrightarrow{\pi'} Y' \xrightarrow{\iota'} X$，且满足 $\pi \circ \iota = \mathrm{id}_Y$和 $\pi' \circ \iota' = \mathrm{id}_{Y'}$。则 Y和 Y' 同构。

证明。考虑复合 $\pi' \circ \iota \colon Y \xrightarrow{\iota} X \xrightarrow{\pi'} Y'$和 $\pi \circ \iota' \colon Y' \xrightarrow{\iota'} X \xrightarrow{\pi} Y$。根据可分裂自同态的定义，我们计算得

$$(\pi \circ \iota') \circ (\pi' \circ \iota) = \pi \circ (\iota' \circ \pi') \circ \iota = \pi \circ \varphi \circ \iota = \pi \circ \iota \circ \pi \circ \iota = \mathrm{id}_Y,$$

并且，同样地，$(\pi' \circ \iota) \circ (\pi \circ \iota') = \mathrm{id}_{Y'}$。我们得出结论：$Y$和 Y' 同构。 □

每个可分裂自同态 $\varphi \colon X \xrightarrow{\pi} Y \xrightarrow{\iota} X$ 也都是 **幂等元**，也就是说，$\varphi \circ \varphi = \varphi$。此外，在我们关注的范畴中，即持久集 Set^P 和持久模 向量空间范畴^P，每个幂等自同态都通过其 **像** 分裂，见下文。在这两个范畴中，我们定义态射 f 的像 $\mathrm{img}\ f$，方法是逐点取像，即 $(\mathrm{img}\ f)_p = f_p(S_p)$。以下两个引理是标准结果。

引理 2.5。设 $\varphi \colon X \to X$ 是 向量空间范畴^{P}或 Set^P中的一个幂等自同态。那么 f 可通过其像分裂：存在一个分解 $f \colon X \xrightarrow{\pi} \mathrm{img}\ \varphi \xrightarrow{\iota} X$，其中 $\pi \circ \iota = \mathrm{id}_{\mathrm{img}\,\varphi}$。

证明。在任意阿贝尔范畴中，例如 Vec^P，每个态射 $f \colon X \to Y$都有一个分解 $X \xrightarrow{\pi} \mathrm{img}\, f \xrightarrow{\iota} Y$，其中 π 是一个满态射，而 ι 是一个单态射（见 [31, 命题 VIII.3.1]）。很容易验证，在 Set 和 Set^P 中也有同样的结论。

现在，考虑一个幂等自同态 $\varphi \colon X \to X$ 及其如上所述的满单分解 $\varphi = \iota \circ \pi$。由于它是幂等的，我们有 $\varphi \circ \varphi = (\iota \circ \pi) \circ (\iota \circ \pi) = \iota \circ \pi$。由于 π 是一个满态射，可以右消去；又由于 ι 是一个单态射，可以左消去，因此从 $\iota \circ (\pi \circ \iota) \circ \pi = \iota \circ \pi$ 可以得到 $\pi \circ \iota = \mathrm{id}_{\mathrm{img}\,\varphi}$，如所需。□

引理 2.6。设 $F \colon P \to$向量空间范畴 是一个持久模，且 $\varphi \colon F \to F$是一个幂等自同态。那么 F 分解为 $\mathrm{img}(\mathrm{id}_F - \varphi) \oplus \mathrm{img}\,\varphi$。

证明。应用引理 2.5 于 φ，我们得到一个分解 $\varphi \colon F \xrightarrow{\pi} \text{像}\ \varphi \xrightarrow{\iota} F$，其中 $\iota \circ \pi = \varphi$且 $\pi \circ \iota = \mathrm{id}$。反过来，$\mathrm{id} - \varphi$ 也是一个幂等元，它可分裂且满足 $\mathrm{id} - \varphi + \varphi = \mathrm{id}$。根据直和的刻画（定义 2.3上方的段落），可得如下形式的分解：像$(\mathrm{id} - \varphi) \oplus$ 像 φ。 □

3. Endomorphisms of persistent sets and rooted subsets

如上所见，持久模的分解与其幂等自同态密切相关。我们的主要思路是，在研究形式为 $\mathcal{L}S$ 的持久模时，对于持久集 $S\colon P \to \mathsf{Set}$，我们寻找 S 的幂等自同态，并通过线性化函子 $\mathcal{L}$ 研究它们的像。

定义3.1. 给定一个持久集 S，**生成元** 是一个对 (p_x, x)，其中 $x \in S_{p_x}$ 满足：x 不属于任何态射 $S_{q\to p_x}$ 的像，对任意 $q < p_x$均如此。显然时，我们通常会在记法中省略等级 p_x，直接称 $x \in S_{p_x}$ 为生成元。

在 S 的生成元上可诱导出一个预序：对于两个生成元 $x \in S_{p_x}$ 和 $y \in S_{p_y}$，我们称 $(p_x, x) \leq (p_y, y)$ 当且仅当 $p_x \leq p_y$。这个关系未必具有反对称性，因此一般而言，生成元的预序集并非偏序集。

生成元很有用，因为持久集 $S\colon P \to \mathsf{Set}$ 的自同态 φ 由其生成元的像唯一确定：根据交换性，对于每个 $z \in S_q$，都有 $\varphi_q(z) = S_{p_x\to q} \circ \varphi_{p_x}(x)$，其中 $x \in S_{p_x}$ 是某个生成元。

在线性代数中，幂等自同态可以看作投影到其像上，也就是投影到其不动点上。这一观点以及上文的生成元概念，启发了下面的定义；该定义在我们的工作中起着基础性作用。

定义 3.2. 根子集 A 是 S 的生成元的一个非空子集，并且存在一个 S 的幂等自同态 φ，使得其中未被固定的生成元集合恰好为 A。若根子集是单元素集 $A = \{x\}$，则称 x 为**有根生成元**。

注 3.3。 对于增广度量空间（M， d， f）及其定义 2.2 中的密度-Rips 持久集 S， M 的点与S的生成元之间存在双射。点 $x \in M$ 首次出现在图 $\mathcal{G}_0(M_{f(x)})$ 中，其中 x 始终是其自身的一个连通分支。下文中，我们常将点 $x \in M$ 与其生成元 $x \in S_{p_x}$视为同一对象。从这个意义上说，我们可以将 S 的自同态理解为点集上的一个自同态，并且该自同态与所有图 $\mathcal{G}_\varepsilon(M_\sigma)$ 的连通分支相容。

我们尤其关注由（增广）度量空间得到的持久集，目标是将有根生成元与这些对象的几何联系起来。考虑一个增广度量空间（M， d， f）及其密度-Rips 持久集，下面的命题 3.4 通过 M 中各点的聚类行为来刻画有根生成元。

命题 3.4. 设（M， d， f）是一个增广度量空间，并考虑一个点 $x \in M$。如果存在另一个点 $y \in M$，使得

(i) $f(y) \leq f(x)$（即 y 比 x“更稠密”），并且 *(ii)* 每当 x 属于一个包含多个点的簇时， $y \in M$ 也属于同一个簇：对于每个 $\mathcal{G}_\varepsilon(M_\sigma)$，如果 x 与另一个点路径连通，那么 x 与 y路径连通，

th那么密度-*Rips* 持久集 S *of* M 的生成元 (p_x, x) 是一个有根生成元。

$Conv$反之，若 x 是 S 的有根生成元，则存在一个点 $y \in M$ 满足条件 ns *(i)*和 *(ii)* 如上。

证明。 在进入证明之前，回顾一下，根据我们构造 S 和包含关系 $P \hookrightarrow \mathbb{R}_{\geq 0} \times \mathbb{R}$ 的方式，对于每个 $q \in P$，都存在一个相应的图 $\mathcal{G}_\varepsilon(M_\sigma)$，其中某个 $(\varepsilon, \sigma) \in \mathbb{R}_{\geq 0} \times \mathbb{R}$。每个元素 $z \in S_q$ 都是该图 $\mathcal{G}_\varepsilon(M_\sigma)$ 的一个连通分支，而满足 $S_{p_x\to q}(x) = z$ 的生成元 $x \in S_{p_x}$ 恰好是该连通分支中的点。

第一部分由下文的命题 3.5 得出，该命题在更一般的情形下证明了这一点 y.

反过来，设 φ 是 S 的一个幂等元，其唯一一个未被固定的生成元是 $x \in S_{p_x}$。这意味着存在一个生成元 $y \in S_{p_y}$，它不同于 x，且 $\varphi_{p_x}(x) =$

$S_{p_y \to p_x}(y)$。显然，对于任何其他生成元 $z \in S_{p_z}$，都有 $\varphi_{p_z}(z) = z$。由 $\varphi_{p_x}(x) = S_{p_y \to p_x}(y)$ 这一事实可推出 $f(y) \le f(x)$，因为 $p_y \le p_x$ 属于 P。为验证第二个条件，取一个 $q \ge p_x$，并假设存在一个生成元 $w \in S_{p_w}$，使得 $S_{p_x \to q}(x) = S_{p_w \to q}(w)$。这意味着，在与 q 相关联的图 $\mathcal{G}_\varepsilon(M_\sigma)$ 中，x 和 w 位于同一个连通分支中；我们断言 y 也在这个分支中。事实上，根据 φ 的定义，有 $S_{p_x \to q} \circ \varphi_{p_x}(x) = S_{p_y \to q}(y) = S_{p_w \to q}(w)$。□

命题 3.5. 设(M, d, f)为一个增广度量空间。若对于一组点 $A \subset M$，存在一点 $y \notin A$，使得对于每个 $x \in A$, $f(y) \le f(x)$，以及每个 $\mathcal{G}_\varepsilon(M_\sigma)$，以下条件之一成立：

- x与 y路径连通，或
- 与 x 路径连通的点集包含于 A,

那么生成元 $\{(p_x, x) \mid x \in A\}$ 是 M的密度-*Rips* 持久集S 中的一个根子集。

证明. 为证明 A 是一个根子集，我们需要定义一个适当的幂等元 φ，它属于S。回想一下，自同态由其生成元的像唯一确定，我们通过如下设定来定义 φ

$$\varphi_{p_x}(x) = \begin{cases} S_{p_y \to p_x}(y), & \text{if } x \in A, \\ x, & \text{otherwise}, \end{cases} \tag{1}$$

对于每个生成元 $x \in S_{p_x}$。我们需要证明 φ 确实定义良好，这意味着无论选择哪个生成元 $x \in S_{p_x}$，$z \in S_q$, $\varphi_q(z) = S_{p_x \to q} \circ \varphi_{p_x}(x)$ 的像都相同。固定一个 $q \in P$ 和一个 $z \in S_q$，并考虑生成元集合 G，其中的生成元在 S_q 中的像为 z,$G := \{(p_x, x) \mid p_x \le q, \quad S_{p_x \to q}(x) = z\}$。

接下来，为了检验 φ是否定义良好，对于任意两个 (p_x, x)，$(p_w, w) \in G$，必须满足

$$S_{p_x \to q} \circ \varphi_{p_x}(x) = S_{p_w \to q} \circ \varphi_{p_w}(w). \tag{2}$$

如果 (p_x, x) 和 (p_w, w) 均不属于 A，或者 (p_x, x) 和 (p_w, w) 均属于 A，那么根据我们在公式 (2) 中对 φ 的定义，以上公式显然成立（参见公式 (1)）。

因此，唯一有意义的情况是 (p_x, x) 或 (p_w, w) 中只有一个属于 A。设 $(p_x, x) \in A$ 且 $(p_w, w) \notin A$。那么，根据假设，x 和 w 在与 q 相关的图 $\mathcal{G}_\varepsilon(M_\sigma)$ 上都需要路径连通到 y，这意味着，如所需，

$$S_{p_x \to q} \circ \varphi_{p_x}(x) = S_{p_y \to q}(y) = S_{p_w \to q}(w) = S_{p_w \to q} \circ \varphi_{p_w}(w).$$

现在，φ 是幂等元，因为对于每个 $x \in A$，我们都有 $\varphi^2_{p_x}(x) = \varphi_{p_x}(S_{p_y \to p_x}(y)) = S_{p_y \to p_x}(\varphi_{p_y}(y)) = S_{p_y \to p_x} y$。显然，唯一不被 φ 固定的生成元是 A 中的那些。由此可知，实际上，A 是一个根子集。□

由根子集诱导的分解。正如我们所见，根子集与点的聚类行为有关。它们也与线性化持久模的分解有关：它们会诱导出直和项。

定理 3.6。设 φ 是持久集 S的一个幂等自同态。那么，持久模 $\mathcal{L}S$ 分解为

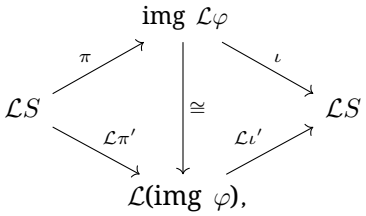
$$\operatorname{img}(\mathrm{id}_{\mathcal{L}S} - \mathcal{L}\varphi) \oplus \mathcal{L}(\operatorname{img}\varphi).$$

证明。由于 φ 是幂等元，$\mathcal{L}\varphi$ 是幂等元。由引理2.6，这诱导出一个分解

$$\mathcal{L}S \cong \operatorname{img}(\mathrm{id} - \mathcal{L}\varphi) \oplus \operatorname{img}\mathcal{L}\varphi.$$

尚需证明 $\operatorname{img}\mathcal{L}\varphi \cong \mathcal{L}(\operatorname{img}\varphi)$。应用引理 2.5 于 $\mathcal{L}\varphi$，得到一个分解$\mathcal{L}\varphi$: $\mathcal{L}S \xrightarrow{\pi} \operatorname{img}\mathcal{L}\varphi \xrightarrow{\iota} \mathcal{L}S$，其中 $\pi \circ \iota = \mathrm{id}$。再次应用引理 2.5，这次应用于 φ，得到一个分解 φ: $S \xrightarrow{\pi'} \operatorname{img}\varphi \xrightarrow{\iota'} S$，其中 $\pi' \circ \iota' = \mathrm{id}$。现在，可分裂自同态被

每个函子保持：在图式 $\mathcal{L}S \xrightarrow{\mathcal{L}\pi'} \mathcal{L}(\mathrm{img}\,\varphi) \xrightarrow{\mathcal{L}\iota'} \mathcal{L}S$ 中，有 $\mathcal{L}\iota' \circ \mathcal{L}\pi' = \mathcal{L}\varphi$ 且 $\mathcal{L}\pi' \circ \mathcal{L}\iota' = \mathrm{id}$。因此，自同态 $\mathcal{L}\varphi$ 以两种方式分裂：



wh中间的箭头存在，并且根据引理 2.4它是一个同构，证明完毕。 □

Co结合上述定理和 Krull-Schmidt 定理，我们得到以下结论：

推论 3.7。 持久集 S 的一个根子集在 $\mathcal{L}S$的分解中诱导出一个直和项。一个有根生成元 $x \in S_{p_x}$ 诱导出一个区间直和项，而所有其他直和项都可以通过分解 $\mathcal{L}(\mathrm{img}\,\varphi)$得到，其中 φ 是与 x相关联的自同态。

这使我们能够逐步剥离形如 $\mathcal{L}S$的持久模区间：找到 S的一个有根生成元，其关联的幂等元为 φ，然后继续考虑 φ 的像，而非 S。在增广度量空间（M， d， f）及其密度-Rips 持久集的设定下，根据命题 M3.4，被剥离的区间可以通过点 M的聚类行为轻松解读。此外，我们描述的条件在实践中确实会出现，正如第5节所示。

邻近根点。事实上，通过考察最近邻，可以看出增广度量空间（M， d， f）中的某些点是有根的，这将在第5节中派上用场。下文中，我们在 M 上固定一个与 f所诱导的序相容的全序。回顾一下，**点**的最近邻 x 是 $x' \neq x$ 中距离 x最小的元素；若出现并列，则按 M上的固定全序打破平局。

定义 3.8。 设 (M, d, f) 是一个增广度量空间。若元素 x 的最近邻 $y \in M$ 满足 $f(y) \leq f(x)$，则称其为**邻域有根的** 。

引理 3.9。 沿用上述记号，若点 $x \in M$ 是邻域有根的，则 x 是(M, d, f)的密度-*Rips* 持久集中的有根生成元。

*Pr*明。 显然， x 的最近邻满足命题 3.4 的条件。 . □

注 3.10。 识别所有邻域有根的点所需的时间与解决全最近邻问题所需的时间相同。显然，通过检查所有可能的点对，可以在 $O(n^2)$ 时间内解决全最近邻问题，其中 n 是点的数量。当点位于欧几里得空间中时，运行时间可改进为 $O(n \log n)$ [19, 38]时间。

分解中的两个显著区间。 根生成元这一概念使我们能够证明，在某些情况下，可以像下面的定理 $\mathcal{L}S$ 所示，找到其分解中至少两个区间。我们先证明定理 3.12 ，再证明定理 3.11；后者已见于 [10, 定理 5。3], 其证明方法是直接构造持久模的一个自同态；在与线性化函子复合后，我们也采用了这一方法。

定理 3.11. 设 S 是一个持久集。假设 S的生成元预序集有一个底元 $\perp$ （也就是说，对于任意其他生成元 x，都有 $\perp \leq x$ ）。那么 $\mathcal{L}S$ 的分解至少包含一个由 $\perp$诱导的区间。

证明. 设 $\perp \in S_{p_\perp}$ 是一个底元，且 $x \in S_{p_x}$ 是 S的一个生成元。由于 $\perp \in S_{p_\perp}$ 是底元，因此有 $p_\perp \leq p_x$。我们可以定义一个幂等元 $\varphi\colon S \to S$，对 S的每个生成元 $x \in S_{p_x}$，令 $\varphi_{p_x}(x) = S_{p_\perp \to p_x}(\perp)$。这个自同态定义良好，且其像只有一个生成元，即 $\perp$，因此 $\mathcal{L}(\operatorname{img}\varphi)$ 同构于一个区间模。□

定理 3.12. 设 $M, d, f)$ 是一个增广度量空间，且令 $S\colon P \to \mathsf{Set}$ 为其密度 *Rips* 持久集，如定义 2.2所述。若 $|M| \geq 2$ 则 $\mathcal{L}S$ 分解为不可分解直和项后，至少包含两个区间。

证明。 考虑函数值最大的点 $\top \in M$，即 $f(\top) \geq f(x)$，对任意其他$x \in M$均成立。令 y为 $\top$ 的最近邻。由于 $M_{f(\top)} = M_\sigma$对任意 $\sigma \geq f(\top)$ 均成立，因此显然$\top$ 及其最近邻 y 满足命题 3.4的条件，因此 $\top$ 是一个有根生成元，给出第一个区间。对于第二个区间，我们注意到至少存在一个密度值最小的点 $\perp \in M$，并应用定理 3.11。 □

例 3.13. 不可分解分解中的直和项并非都能通过取有根子集并应用推论 3.7得到。举例来说，考虑由平面中的六个点 $\{x_0, \ldots, x_5\}$ 构成的增广度量空间，如图 1所示。注意， x_4和 x_5在相应的密度-Rips 持久集中是有根的，并且可以将它们剥离。剥离后，我们得到持久集 $S\colon \ P \to \mathsf{Set}$，其中 $P\colon = \{0,2,3,4\} \times \{0,1,2,3,4,5\} \subset \mathbb{R}^2$，如图 2所示。此例改编自 [14, 例 4。12], 该例是在讨论 征服者的背景下引入的，我们将在第 4 节讨论征服者。

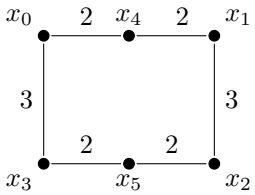


FIGURE 1. 例 M， d， f）构成的增广度量空间（3.13，其中f（x_i） $= i$。这是平面中的六个点 $\{x_0, \ldots, x_5\}$，每条线旁的数字表示相应的距离。

我们声称，持久模 $\mathcal{L}S\colon P \to \mathsf{Vec}$ 可分解为四个直和项，且它们都是区间模。我们用 I_0, I_1, I_2 和 I_3 表示这些直和项，其中每个 I_i 都与 S 的生成元 (p_i, x_i) 相关联，其中 $p_i = (0, i) \in P$。对于每个 $i = 0, \ldots, 3$，我们令 $(I_i)_p = 0$，其中任意 $p < p_i$ 均成立，并令 $(I_i)_{p_i} = K$，再定义 $\iota_i\colon I_i \to \mathcal{L}S$ 如下：

$$(\iota_0)_{p_0}(1) = [x_0], \qquad (\iota_1)_{p_1}(1) = [x_1] - [x_0],$$
$$(\iota_2)_{p_2}(1) = [x_2] - [x_1], \qquad (\iota_3)_{p_3}(1) = [x_3] - [x_0] + [x_1] - [x_2].$$

上确界 是满足 $((\mathcal{L}S)_{p_i \to p} \circ (\iota_i)_{p_i})(1)$ 非零的等级 $p \geq p_i$。可以看出 这些映射诱导出一个分解 $\mathcal{L}S \cong I_0 \oplus I_1 \oplus I_2 \oplus I_3$。

然而，除 $\{x_1, x_2, x_3\}$ 以外，生成元的任何子集都不是有根的，因为由 $\{x_1$给出的每个连通分支 $, x_0\}, \{x_1, x_2\}, \{x_2, x_3\}$, 以及 $\{x_0\, x_3\}$ 出现在 S .

4. ROOTED GENERATORS AS A GENERALIZATION OF THE ELDER RULE

单参数情形。 现在假设偏序集 P 是一个有限全序偏序集。在此情形下，有根生成元理论使我们能够恢复长者规则 [24] （另见 [22]和 [14]）。

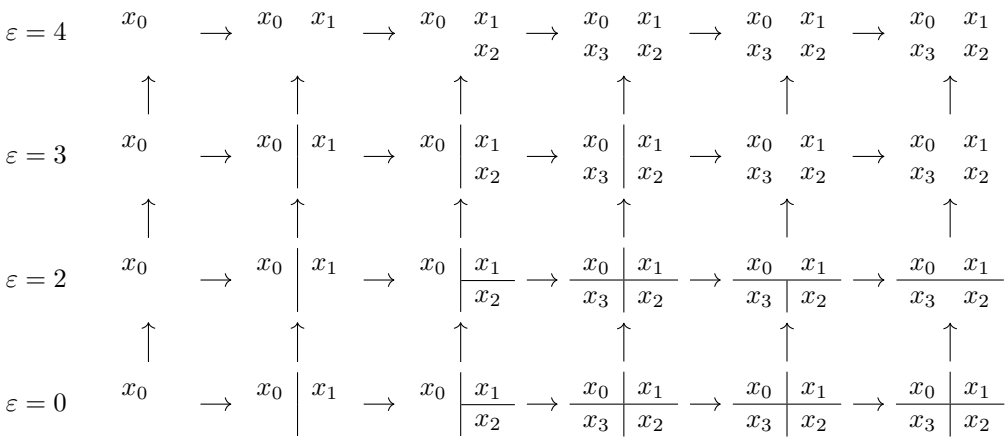


FIGURE 2. 持久集 $S\colon \quad P \to \mathsf{Set}$ 示例3.13 通过取图1 中的密度-Rips持久集并移除 x_4 和 x_5而得到。网格中的每个节点都表示 x_i的一个划分，其中 x_i 和 x_j 若未被一条线分隔，则属于同一划分。箭头表示这样的函数：将一个节点中的 x_i 的划分映射到另一个节点中的 x_i 的划分。

命题4.1。设 P 为有限全序偏序集，且 $S\colon P \to \mathsf{Set}$ 为持久集。假设 S 至少有两个生成元，且 $S_\top$ 是单元素集，其中 $\top$ 是 P的最大元。那么每个极大生成元（按照定义 *3.1*中的预序）都是有根的。

证明。令 $x \in S_{p_x}$ 为一个极大生成元，并定义

$$I_x \coloneqq \{q \in P \mid q \geq p_x \text{ and, for any other generator } w \in S_{p_w},\ S_{p_w \to q}(w) \neq S_{p_x \to q}(x)\}.$$

由于 $p_x \in I_x$, I_x 非空，因此我们可以考虑 I_x的上界集合 $U \subset P$。此外，由于 $S_\top = \{*\}$ 且根据假设至少存在两个生成元，集合 $U \setminus I_x$非空。令 $U \setminus I_x$中的最小元素为 α。根据 I_x 和 $U \setminus I_x$ 的构造，存在一个生成元 $y \in S_{p_y}$，使得 $S_{p_y \to \alpha}(y) = S_{p_x \to \alpha}(x)$。现在，由于 x 是极大的，因此有$p_y \leq p_x$，我们可以定义一个幂等元 $\varphi\colon S \to S$，使得 $\varphi_{p_x}(x) = S_{p_y \to p_x}(y)$，且对于任意其他生成元 $z \in S_{p_z}$，$\varphi_{p_z}(z) = z$。根据我们对α的定义方式，这样的幂等元是良定义的：如果存在另一个生成元 $w \in S_{p_w}$，使得 $S_{p_w \to q}(w) = S_{p_x \to q}(x)$，那么 $\alpha \leq q$，并且$S_{p_y \to q}(y) = S_{p_x \to q}(x)$。因此，$x$ 是有根的，证毕。□

因此，当 P 是全序时，我们可以通过剥离有根的生成元，将任意持久模 $\mathcal{L}S$ 分解；具体做法是遵循定理3.6 并迭代地考察极大生成元。

与常值征服者的关系。设 $(M,\ d,\ f)$ 是一个增广度量空间。Cai、Kim、M´emoli 和 Wang [14]将常值征服者定义如下。首先，在 M 上定义一个超度量：$u(x,\ x') \coloneqq \min\{\varepsilon \in [0, \infty)$，并且 $\mid x$ 和 x' 在 $\mathcal{G}_\varepsilon(M)\}$ 中路径连通。

现在固定 $\prec$ M 上的一个全序，并设 $x \in M$ 是相对于此序的非极小元。x 在 M 中的一个**征服者** 是另一个点 $x' \in M$，满足：**(1)** $x' \prec x$；**(2)** 对任意满足 $x'' \prec x$ 的 x''，都有 $u(x,\ x') \leq u(x,\ x'')$。给定函数 $f\colon M \to \mathbb{R}$，相对于 $\prec$的非极小 $x \in M$的**征服函数**是一个函数 $c_x\colon [f(x)\ \infty) \to M$，它将每个 σ

映射到 M_σ中 x的一个征服者。对于 M的极小元 $\perp$，我们定义 $c_\perp\colon [f(\perp)\ \infty) \to M$为取值恒为 $\perp$的常值函数。

此外，在同一篇论文中 [14], 给定一点 $x \in M$，并假设 $f\colon\ M \to \mathbb{R}$ 是单射，作者将**阶梯码** x 定义为由下式给出的集合

$$I_x := \{(\varepsilon, \sigma) \in \mathbb{R}_{\geq 0} \times \mathbb{R} \mid x \in M_\sigma \text{ and } x \text{ is the oldest in } [x]_{(\varepsilon,\sigma)}\},$$

其中，$[x]_{(\varepsilon,\sigma)}$ 是与 x 在 $\mathcal{G}_\varepsilon(M_\sigma)$中路径连通的点的集合，而“最老”意味着对于任意其他 $x' \in [x]_{(\varepsilon,\sigma)}$，$f(x) < f(x')$。作者还定义了 f 不是单射时的类似概念，此处不再赘述。

最后，作者提出了以下问题：

问题4.2。设 (M, d, f) 是一个增广度量空间。若 $x \in M$ 具有常值征服者函数，那么由 I_x 支撑的区间模是否是其密度-Rips持久模的一个直和项?

如果我们用有根的生成元替换常值征服者，那么根据推论 3.7，答案是肯定的。下一个例子表明，不能像上文问题中原先的表述那样得到同样的结论。

例 4.3。考虑 M 中由以下点组成的子集：$\mathbb{R}$ $x_0 = 0$，$x_1 = 7.5$，$x_2 = 3$ 和$x_3 = 5$。在 $\mathbb{R}$ 上由欧几里得距离诱导的度量下，M 是一个度量空间，并且通过定义 $f(x_i) = i$，可将其构造成一个增广度量空间，见图 3。

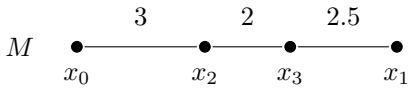


FIGURE 3. 例 M、d、f) 构成的增广度量空间，见例 4.3，其中 $M \subset \mathbb{R}$和 $f(x_i) = i$。

考虑与 f, $x_0 \prec x_1 \prec x_2 \prec x_3$ 兼容的 M 上唯一的全序 $\prec$。点 x_1 有一个常值征服者：x_0 是唯一的候选者；显然，对于每个 $i = 1, \ldots, 3$

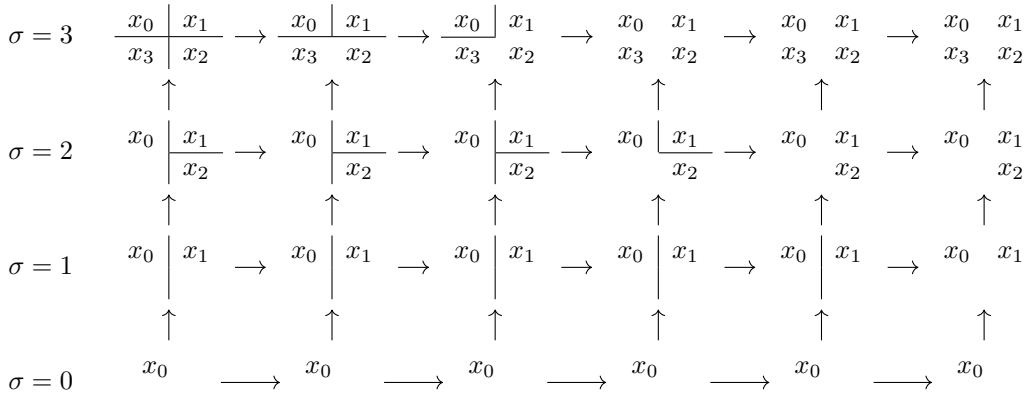


FIGURE 4. 我们将图 3 中的密度-Rips 持久集描绘如下。

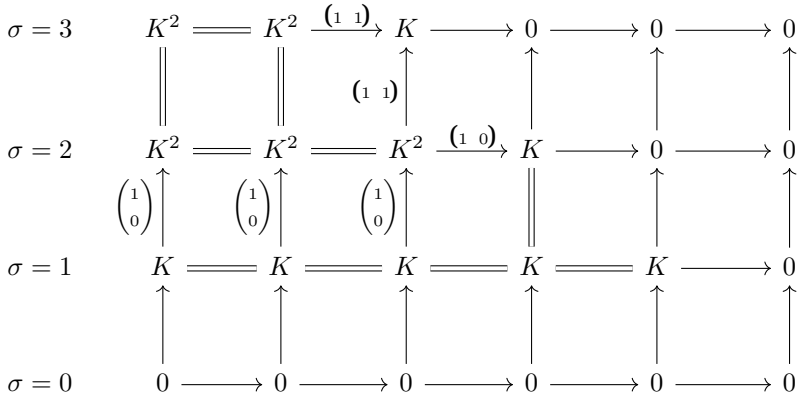


FIGURE 5. 一个不可分解持久模 $F\colon\ P \to$ 向量空间范畴，如例 4.3所示。

以及 $x' \prec x_1$, $u_i(x_1, x_0) \leq u_i(x_1, x')$，其中 u_i 是 M_i 的超度量，恰恰是因为 x_0 是唯一满足 $x' \prec x_1$ 的点。

令 $S\colon P \to \mathsf{Set}$ 为由增广度量空间 (M, d, f) 构造的密度-Rips 持久集。这里，偏序集 P 是由网格 $\{0, 2, 2.5, 3, 4.5, 7.5\} \times \{0, 1, 2, 3\}$给出的 $\mathbb{R}^2$ 的子偏序集，其中第一个坐标表示距离，第二个坐标表示密度。我们在图 4 中描绘了 S 。现在我们继续分解 $\mathcal{L}S$。首先，注意到 x_3 是一个有根生成元，并考虑与之关联的幂等元 $\varphi\colon\ S \to S$。根据定理 3.6，分解中存在一个区间 $I \coloneqq \mathsf{img}(\mathrm{id}_{\mathcal{L}S} - \mathcal{L}\varphi)$，我们可以继续考察持久集 $\mathsf{img}\,\varphi$。在 $\mathsf{img}\ \varphi$中， x_0 是一个极小生成元。根据定理 3.11（及其证明），存在一个幂等元 $\psi\colon \mathsf{img}\ \varphi \to \mathsf{img}\ \varphi$，使得 $I' \coloneqq \mathsf{img}\ \mathcal{L}\psi$ 是一个区间。再次应用定理 3.6，我们得到 $\mathcal{L}S$ 的如下形式的分解

$$I \oplus I' \oplus \mathrm{img}(\mathrm{id}_{\mathcal{L}\,\mathrm{img}\,\varphi} - \mathcal{L}\psi).$$

通过直接计算可以看出，$\mathsf{img}(\mathrm{id}_{\mathcal{L}\,\mathrm{img}\,\varphi} - \mathcal{L}\psi)$ 与图 5所描述的持久模同构。此外，该持久模是不可分解的，这可以通过考察其自同态来验证：当且仅当持久模 F 的每个自同态 F 要么是幂零的，要么是一个同构时， [9], 就 [10]是不可分解的（参见 和）。

注意， x_1 不是 $\mathcal{L}S$中的有根生成元。在 M_1中，在 $\varepsilon \in [0, 7.5)$期间， x_1 是一个独立的连通分支，直到 x_0 加入该连通分支。在 M_3 中，它在$\varepsilon \in [0, 2.5)$期间也是独立的，随后加入 x_3 所在的连通分支，而该分支在那个时刻与 x_0 不连通。同样， x_2 也不是有根的。

注记 4.4。请注意，在征服者定义中的条件 (2) 中，我们要求 $x'' \prec x$。对于增广度量空间而言，这项要求体现了常值征服者函数与有根生成元之间部分差异。若去掉这项要求，并将由此得到的概念称为 **征服者***，假设 f 是单射，且 $\prec$ 与 f所诱导的序相容，那么相对于 $\prec$而言非极小的点 $x \in M$ 有一个常值**征服者*** 函数，当且仅当 x 是一个有根生成元，如命题 3.4所述。

5. A lower bound on the number of expected intervals

我们将所建立的理论应用于研究由密度-Rips 得到的持久模，其典型分解可能呈现怎样的形式。具体而言，假设我们从 $\mathbb{R}^d$ 中的同一个密度函数 $f(x)$ 独立抽取 n 个点，得到一个有限度量空间 $M \subset \mathbb{R}^d$。接着可以考虑增广度量空间 $(M,\ d,\ f)$，其中 f 并非估计得到的密度，而是真实的底层密度函数。这种设定与实际操作相似，但更适合理论研究。令 S 为 M 的密度-Rips 持久集。那么，我们可以预期 $\mathcal{L}S$ 的分解中有多少个区间？下述定理表明，在对 f 的条件非常一般的情况下，无论 d 如何，随着 n 趋于无穷，至少可以预期有 25% 的直和项是区间。

定理 5.1. 设 $X_1, \ldots, X_n$ 是取值于 $\mathbb{R}^d$ 中、从共同密度函数 $f(x)$ 独立同分布抽取的点；该密度函数相对于勒贝格测度几乎处处连续。

考虑有限增广度量空间 $(M = \{X_1, \ldots, X_n\}, d_M, f)$，其中 d_M 由 $\mathbb{R}^d$ 中的欧几里得度量诱导，并令 S 为其密度-*Rips* 持久集。

令 $\mathfrak{I}_n$ 为随机变量，表示 $\mathcal{L}S$ 的不可分解分解中区间的数量；令 $\mathfrak{S}_n$ 为随机变量，表示同一分解中直和项的总数。我们有

$$\liminf_{n\to\infty} \mathrm{E}\left[\frac{\mathfrak{I}_n}{\mathfrak{S}_n}\right] \geq c(d), \tag{3}$$

其中 $c(d)$ 是依赖于 d 的常数，而 $c(1) = \frac{1}{3}$，$c(2) \approx 0.31$ 和 $c(d) \downarrow \frac{1}{4}$，当 $d \to \infty$ 时。

本节余下部分将用于证明该定理。度量空间的最近邻图起着根本性的作用。

定义 5.2。 M 的**最近邻图** 是定义在 M 上的有向图，其有向边形如 $(x,\ x')$，其中 x' 是 x的最近邻。

现在，我们感兴趣的是估计邻近有根元素的数量，如定义 3.8所述，因为它们在 $\mathcal{L}S$的分解中诱导出一个区间。不过，一般而言，是否邻近有根取决于 f。为去掉对 f 的条件，我们有：

引理 5.3。 设 (M, d_M, f) 为一个增广度量空间，且 S为其密度-*Rips* 持续集。 $\mathcal{L}S$ 的不可分解分解中的区间数至少等于 M的最近邻图中的2-环数。

证明。 不失一般性，设 $|M| \geq 2$。令 G 为 M的最近邻图。该图中的环恰好是 2-循环，并且 G 的每个弱连通分量都恰好包含一个 2-环（见 [25]）。

设 $C_1, \ldots, C_k$ 是 G 的弱连通分量。固定 $i \in \{1, \ldots, k\}$，并令 $x, y \in M$ 满足 (x, y) 和 (y, x) 是 C_i 中的 2- 循环。要么 $f(y) \leq f(x)$，要么 $f(x) \leq f(y)$；并且 x 是邻域有根的、 y 是邻域有根的，或两者都是邻域有根的。设 x 是邻域有根的，并定义自同态 $\varphi_i\colon S \to S$，令

$$(\varphi_i)_{p_z}(z) = \begin{cases} S_{p_y \to p_x}(y), & \text{if } x = z, \\ z, & \text{otherwise,} \end{cases}$$

对每个生成元 $z \in S_{p_z}$。由命题 3.4 可知，这样的自同态是良定义的。对每个 i，按上述方式构造一个幂等元 φ_i，显然我们可以迭代地剥离相应的区间，从而得到所需结论。□

显然，2-环的数量是其最近邻的最近邻也以该点为最近邻的点数的一半。估计在随机点过程下，一个点成为最近

邻的最近邻的概率这一问题，已有多位作者研究过（参见 [21, 25, 26, 27, 36]）。

在我们的情形中，当我们有 $X_1, \dots, X_n$ 个来自同一密度函数 f 的独立同分布点，且满足定理 5.1 的条件时，根据 [27, 定理 1.1], ，并令 $N_{i,n}$ 表示 X_i 是其最近邻的最近邻这一概率事件，我们有

$$\lim_{n\to\infty} \mathrm{P}(N_{i,n}) = b(d), \tag{4}$$

其中，$b(d)$ 是单位 d-球的体积除以两个中心距离为 1 的单位球面的并集体积。事实上，$b(1) = \frac{2}{3}$，$b(2) \approx 0.621$，并且当 $d \to \infty$ 时，$b(d) \downarrow \frac{1}{2}$（见 [36, 表 2]）；我们定义 $c(d) := \frac{b(d)}{2}$。

现在我们可以完成本节开头定理 5.1 的证明了。应用引理 5.3以及期望的线性性，可得

$$\mathrm{E}[\mathfrak{I}_n] \geq \mathrm{E}\left[\sum_{i=1}^{n} \frac{I(N_{i,n})}{2}\right] = \sum_{i=1}^{n} \frac{\mathrm{E}[I(N_{i,n})]}{2} = \sum_{i=1}^{n} \frac{\mathrm{P}(N_{i,n})}{2},$$

其中 $I(N_i)$ 是 $N_{i,n}$ 的示性随机变量。根据式 (4)，我们有

$$\liminf_{n\to\infty} \mathrm{E}\left[\frac{\mathfrak{I}_n}{n}\right] \geq \frac{b(d)}{2} = c(d).$$

最后，注意到分解中的直和项数不超过点的数量 $\mathfrak{S}_n \leq n$，由下面的引理可知，定理 5.1 的式 (3) 得证，从而完成证明。

引理 5.4。设 (M, d_M, f) 是一个增广度量空间，并设 $S\colon P \to \mathsf{Set}$ 是其密度 *Rips* 持久集。那么，$\mathcal{L}S$ 的任意分解至多包含 $n = |M|$ 个直和项。

证明。给定一个持久模 $F\colon P \to$向量空间范畴，记 $\beta_0(F)\colon P \to \mathbb{N}$为将每个 $p \in P$映射到 F在 $p \in P$处的多重分次零阶贝蒂数的函数（其定义参见 [30] ）。正是因为 S 有 n 个生成元，不难看出 $\sum_{p\in P} \beta_0(\mathcal{L}S)(p) = n$。

考虑一个分解 $\mathcal{L}S \cong X_1 \oplus \cdots \oplus X_k$。由于贝蒂数具有可加性，我们有$\beta_0(\mathcal{L}S) = \sum_{i=1}^{k} \beta_0(X_i)$。因此，我们可以写成

$$\sum_{p\in P} \beta_0(\mathcal{L}S)(p) = \sum_{i=1}^{k} \sum_{p\in P} \beta_0(X_i)(p) = n. \tag{5}$$

现在，由于 $\sum_{p\in P} \beta(X_i)(p) \geq 1$ 当且仅当 $X_i \neq 0$，因此可得 $k \leq n$。 □

6. 讨论

尽管我们关注的是增广度量空间和 density-Rips，但有根子集也可应用于其他持久集。我们特别感兴趣的是度量空间的度数-Rips滤过 [5]，在该滤过中，我们按照底层几何图中顶点的度数进行过滤。为适应这种情况，可以修改命题 3.4 的条件 1，以考虑度数的演变，而非密度。我们将对这一情形进行深入探讨，留待未来工作。

我们在第 5 节的下界推导以及初步实验评估中都已看到，至少在本文考察的情形下，我们可以预期，在源自几何的那些持久模的分解中会发现许多区间。这与纯代数情形形成对比；鉴于近期的发展， [2,3], 寻求分解或许难以奏效。

References

[1] Hideto Asashiba、Micka

¨

el Buchet、Emerson G. Escolar、Ken Nakashima 和 Michio Yoshiwaki，，《计算几何》（），第 33 号论文，doi:10.1016/j.comgeo.2022.101879。[2] Ulrich Bauer、Magnus B. Botnan、Steffen Oppermann 和 Johan Steen，*Cotorsion torsion triples and the representation theory of filtered hierarchical clustering*，《数学进展》369（2020），107171，51 页，doi:10.1016/j.aim.2020.107171。[3] Ulrich Bauer 和 Luis Scoccola，*Generic two-parameter persistence modules are nearly indecomposable*，2022 年 11 月，arXiv:2211.15306。[4] H

¨

avard Bakke Bjerkevik、Magnus Bakke Botnan 和 Michael Kerber，，《计算数学基础》（），第 – 期，doi:10.1007/s10208-019-09442-y。[5] Andrew J. Blumberg 和 Michael Lesnick，*Stability of 2-parameter persistent homology*，《计算数学基础》（2022），doi:10.1007/s10208-022-09576-6。[6] Magnus Bakke Botnan 和 William Crawley-Boevey，*Decomposition of persistence modules*，《美国数学会会刊》148（2020），第 11 期，4581–4596，doi:10.1090/proc/14790。[7] Magnus Bakke Botnan 和 Michael Lesnick，*Algebraic stability of zigzag persistence modules*，《代数与几何拓扑》18（2018），第 6 期，3133–3204，doi:10.2140/agt.2018.18.3133。[8] Magnus Bakke Botnan、Steffen Oppermann、Steve Oudot 和 Luis Scoccola，*On the bottleneck stability of rank decompositions of multi-parameter persistence modules*，2022，arXiv:2208.00300，7 月，。[9] Michel Brion，*Representations of quivers*，《表示论中的几何方法 I》，S´emin. Congr.，第 24 卷，法国数学会，巴黎，2012 年，第 103–144 页。[10] Jacek Brodzki、Matthew Burfitt 和 Mariam Pirashvili，*On the complexity of zero-dimensional multiparameter persistence*，2020，arXiv:2008.11532，8 月，。[11] Micka

¨

el Buchet、Frederic Chazal、Tamal K. Dey、Fengtao Fan、Steve Y. Oudot 和 Yusu Wang，，《第 1 届国际计算几何研讨会》，LIPIcs. Leibniz Int. Proc. Inform.，第 34 卷，2015 年，第 827–841 页，doi:10.4230/LIPIcs.SOCG.2015.827。[12] Micka

¨

el Buchet 和 Emerson G. Escolar，，《应用与计算拓扑学杂志》（），第 – 期，doi:10.1007/s41468-020-00053-z。[13] Micka

¨

el Buchet 和 Emerson G. Escolar，，《计算几何杂志》（），第 – 期，doi:10.20382/jocg.v13i1a12。[14] Chen Cai、Woojin Kim、Facundo M

´

emoli 和 Yusu Wang，，《SIAM应用代数与几何杂志》（），第 – 期，。[15] Gunnar Carlsson 和 Facundo M

´

emoli，，《机器学习研究杂志》（），–。[16]，*Multiparameter hierarchical clustering methods*，作为研究工具的分类，《Stud. Classifica- tion Data Anal. Knowledge Organ.》，Springer，柏林，2010 年，第 63–70 页，doi:10.1007/978-3-642-10745-0_6。[17] Gunnar Carlsson 和 Facundo M´emoli，*Classifying Clustering Schemes*，《计算数学基础》13（2013），第 2 期，221–252，doi:10.1007/s10208-012-9141-9。[18] Gunnar Carlsson 和 Afra Zomorodian，*The Theory of Multidimensional Persistence*，《离散与计算几何》42（2009），第 1 期，71–93，doi:10.1007/s00454-009-9176-0。[19] Kenneth L. Clarkson，*Fast algorithms for the all nearest neighbors problem*，第 24 届计算机科学基础年度研讨会，IEEE，1983 年，第 226–232 页，doi:10.1109/SFCS.1983.16。[20] Anne Collins、Afra Zomorodian、Gunnar Carlsson 和 Leonidas J. Guibas，*A barcode shape descriptor for curve point cloud data*，**28** 2004 6, 881 894 doi:10.1016/j.cag.2004.08.015，《计算机与图形学》（），第 – 期，。[21] Trevor F. Cox，*Reflexive nearest neighbours*，《生物统计学》37（1981），第 2 期，367，doi:10.2307/2530424。[22] Justin Curry，*The fiber of the persistence map for functions on the interval*，《应用与计算拓扑学杂志》2（2018），第 3-4 期，301–321，doi:10.1007/s41468-019-00024-z。[23] Tamal K. Dey 和 Cheng Xin，*Generalized persistence algorithm for decomposing multiparameter persistence modules*，**6** 2022 3, 271 322，《应用与计算拓扑学杂志》（），第 – 期，doi:10.1007/s41468-022-00087-5。

[24] Herbert Edelsbrunner 和 John L. Harer，*Computational topology: an introduction*，美国数学学会，普罗维登斯，罗得岛州，2010 年，doi:10.1090/mbk/069。[25] David Eppstein、Michael S. Paterson 和 Frances F. Yao，*On nearest-neighbor graphs*，《离散与计算几何》17（1997），第 3 期，263–282，doi:10.1007/PL00009293。

[26] Norbert Henze, *On the probability that a random point is the th nearest neighbour to its own th nearest* j k，《应用概率学杂志》（），第 – 期，[27] *On the fraction of random points by specified nearest-neighbour interrelations and degree of attraction* **23** 1986 1, 221 226 doi:10.2307/3214132 **19** 1987 4, 873 895 doi:10.2307/1427106 《应用概率进展》（），第 – 期，.[28] Jon Kleinberg, *An impossibility theorem for clustering*，《神经信息处理系统进展》，第15卷，麻省理工学院出版社，2002年。[29] Michael Lesnick 和 Matthew Wright, *Interactive visualization of 2-d persistence modules*，2015年12月，arXiv:1512.00180。[30]，*Computing minimal presentations and bigraded Betti numbers of 2-parameter persistent homology*《SIAM应用代数与几何杂志》6（2022），第2期，267–298页，doi:10.1137/20M1388425。[31] Saunders Mac Lane, *Categories for the working mathematician*，《数学研究生教材》，第5卷，施普林格纽约，1978年，doi:10.1007/978-1-4757-4721-8。[32] Leland McInnes 和 John Healy, *Accelerated hierarchical density clustering*，2017年 IEEE 国际数据挖掘会议研讨会，2017年，第33–42页，doi:10.1109/ICDMW.2017.12。[33] Samantha Moore, *Hyperplane restrictions of indecomposable n-dimensional persistence modules*，《同调、同伦与应用》24（2022），第2期，291–305页，doi:10.4310/HHA.2022.v24.n2.a14。[34] Alexander Rolle 和 Luis Scoccola, *Persistable: Persistent and stable clustering* https://github.，网址：。[35]，com/LuisScoccola/persistable *Stable and consistent density-based clustering*，2021年7月，arXiv:2005.09048。[36] Mark F. Schilling, *Mutual and shared neighbor probabilities: finite- and infinite-dimensional results*，《应用概率进展》18（1986），第2期，388–405页，doi:10.2307/1427305。[37] Bernard W. Silverman, *Density estimation for statistics and data analysis*，《统计学与应用概率专著》，Chapman & Hall，伦敦，1986年。[38] Pravin M. Vaidya, *An O*（*n*log *n*）*algorithm for the all-nearest-neighbors problem*，《离散与计算几何》4（1989），第2期，101–115页，doi:10.1007/BF02187718。