

# PERSISTENT HOMOLOGY AND THE UPPER BOX DIMENSION

BENJAMIN SCHWEINHART

ABSTRACT. We introduce a fractal dimension for a metric space defined in terms of the persistent homology of extremal subsets of that space. We exhibit hypotheses under which this dimension is comparable to the upper box dimension; in particular, the dimensions coincide for subsets of  $\mathbb{R}^2$  whose upper box dimension exceeds 1.5. These results are related to extremal questions about the number of persistent homology intervals of a set of  $n$  points in a metric space.

## 1. INTRODUCTION

Several notions of fractal dimensions based on persistent homology have been proposed in the literature, including in the PhD thesis of Vanessa Robins [25], in a paper written by Robert MacPherson and the current author [21], and by Adams et al. [1]. In that work, empirical estimates of the proposed dimensions were compared with classically defined fractal dimensions. Here, we prove the first rigorous analogue of those comparisons.

The motivation for the definition proposed here comes from the literature on minimal spanning trees. The properties of minimal spanning trees of point collections contained in a bounded subset of  $\mathbb{R}^m$  have long been of interest [28, 19]. In 2005, Kozma, Lotker, and Stupp [20] proved a relationship between the extremal behavior of these trees and the upper box dimension. To be precise, let  $T(\mathbf{y})$  denote the minimal spanning tree of a finite metric space  $\mathbf{y}$  and let

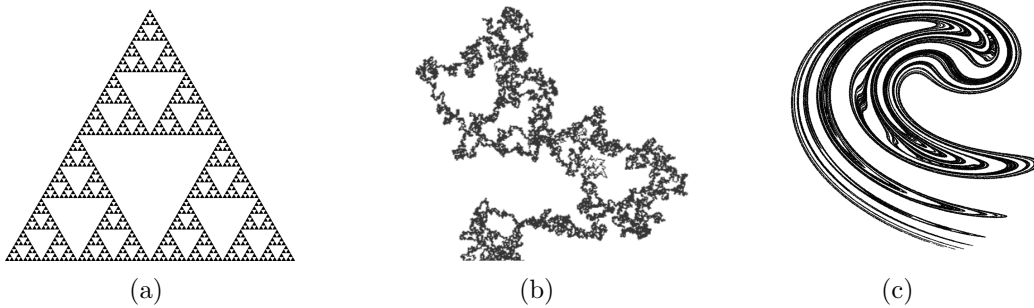
$$E_\alpha^0(\mathbf{y}) = \frac{1}{2} \sum_{e \in T(\mathbf{y})} |e|^\alpha,$$

where the sum is taken over all edges  $e$  in the tree  $T(\mathbf{y})$ , and  $|e|$  denotes the length of the edge.

**Definition 1.** *If  $X$  is a bounded metric space, let  $\dim_{MST}(X)$  be the infimal exponent  $\alpha$  so that  $E_\alpha^0(\mathbf{x})$  is uniformly bounded for all finite point sets  $\mathbf{x} \subset X$ :*

---

*Date:* July 2019.



**Figure 1.** Examples which satisfy the hypotheses of our main theorem: the Sierpinski triangle ( $\dim_{\text{box}} = \log(3) / \log(2) \approx 1.585$ ), Schramm–Loewner Evolution with  $\kappa = 6$  ( $\dim_{\text{box}} \geq \dim_{\text{Hausdorff}} = 7/4$  [2]), Figure provided by T. Kennedy [18]), and the Ikeda attractor ( $\dim_{\text{box}} \approx 1.7$  assuming that computational estimates of the box dimension of that set accurate).

$$\dim_{MST}(X) = \inf \left\{ \alpha : \exists C \text{ so that } E_{\alpha}^0(\mathbf{x}) < C \forall \text{ finite } \mathbf{x} \subset X \right\} .$$

Kozma, Lotker, and Stupp proved that  $\dim_{MST}(X)$  equals the upper box dimension [3] of  $X$ .

**Definition 2.** Let  $X$  be a bounded metric space and let  $N_{\delta}(X)$  be the maximal number of disjoint closed  $\delta$ -balls with centers in  $X$ . The **upper box dimension** of  $X$  is

$$\dim_{\text{box}}(X) = \limsup_{\delta \rightarrow 0} \frac{\log(N_{\delta}(X))}{\log(1/\delta)} .$$

**Theorem 3** (Kozma, Lotker, and Stupp [20]). For any metric space  $X$ ,

$$\dim_{MST}(X) = \dim_{\text{box}}(X) .$$

If  $\mathbf{x}$  is a finite point set contained in a bounded metric space, and  $PH_i(\mathbf{x})$  is the  $i$ -dimensional persistent homology of the Čech complex of  $\mathbf{x}$ , there is a bijection between the edges of the Euclidean minimal spanning tree of  $\mathbf{x}$  and the intervals in the canonical decomposition of  $PH_0(\mathbf{x})$ , where the length of an interval in  $PH_0(\mathbf{x})$  is half the length of the corresponding edge. (Note that if persistent homology is taken of the Rips complex of  $\mathbf{x}$ , then a similar correspondence holds without the requirement of an ambient space, where an interval corresponds to the edge of the same length.) This observation suggests a generalization of the previous result to

higher-dimensional persistent homology. Namely, let

$$E_\alpha^i(\mathbf{x}) = \sum_{(b,d) \in PH_i(\mathbf{x})} (d-b)^\alpha$$

where the sum is taken over all bounded  $PH_i$  intervals and define:

**Definition 4.** *Let  $X$  be a bounded subset of a metric space. The  $PH_i$ -dimension of  $X$  is*

$$\dim_{PH}^i(X) = \inf \left\{ \alpha : \exists C \text{ so that } E_\alpha^i(\mathbf{x}) < C \forall \text{ finite } \mathbf{x} \subset X \right\}.$$

Also, let  $\widetilde{\dim}_{PH}^i(X)$  be defined by replacing the Čech complex with the Rips complex in the previous construction.

Then

$$\dim_{PH}^0(X) = \dim_{\widetilde{PH}}^0(X) = \dim_{\text{MST}}(X).$$

Note that  $\dim_{\widetilde{PH}}^i(X)$  is defined for bounded metric spaces  $X$ , rather than bounded subsets of a metric space.

**Question 5.** *Are there hypotheses on  $X$  under which  $\dim_{PH}^i(X) = \dim_{\text{box}}(X)$  or  $\widetilde{\dim}_{PH}^i(X) = \dim_{\text{box}}(X)$  for  $i > 0$ ?*

Unlike the 0-dimensional case, equality will not always hold. For example, the  $PH_1$ -dimension of a line in  $\mathbb{R}^m$  will always be 0. Also, there are metric spaces not embeddable in any finite-dimensional Euclidean space whose  $\widetilde{PH}_1$ -dimension exceeds their upper box dimension.

**Proposition 6.** *There is a metric space  $X$  so that  $\dim_{\widetilde{PH}_1}(X) = 2$  but  $\dim_{\text{box}}(X) = 1$ .*

We prove this in Section 5.1. This behavior is related to the existence of families of point sets  $\{\mathbf{x}_n\}_{n \in \mathbb{N}}$  for which number of intervals of  $PH_i(\{\mathbf{x}_n\})$  grows slower or faster than linearly in  $|\mathbf{x}_n|$  when  $i > 0$ ; by contrast, the reduced 0-dimensional persistent homology of a finite metric space with  $k$  points always has  $k - 1$  intervals.

We restrict our attention to subsets of Euclidean space, and conjecture

**Conjecture 7.** *For any  $m \in \mathbb{N}$  and  $0 \leq i < m$ , there is a constant  $\gamma_i^m < m$  so that if  $X \subset \mathbb{R}^m$  and  $\dim_{\text{box}}(X) > \gamma_i^m$  then*

$$\dim_{PH}^i(X) = \dim_{\text{box}}(X).$$

Cohen-Steiner, Edelsbrunner, Harer, and Mileyko studied a quantity similar to  $E_\alpha^i$  in their paper [9]. Their results immediately imply that if  $X \subset \mathbb{R}^m$ , then  $\dim_{PH}^i(X) \leq m$  for all  $i \in \mathbb{N}$  (Corollary 15). If  $X$  is a subset of  $\mathbb{R}^m$  with non-empty interior the corresponding lower bound is easy to show and  $\dim_{PH}^i(X) = m$  for  $i = 1, \dots, m - 1$  ( Proposition 21). However, our focus here will be to prove results about subsets of fractional box dimension. This task is challenging, and involves difficult combinatorial problems. Our main result is:

**Theorem 8.** *Let  $X$  be a bounded subset of  $\mathbb{R}^2$ . If  $\dim_{box}(X) > 1.5$ , then*

$$\dim_{PH}^1(X) = \dim_{box}(X) .$$

See Figure 1 for images of two examples known to meet these hypotheses, and one that is believed to based on computational experiments. The upper bound  $\dim_{PH}^1(X) \leq \dim_{box}(X)$  is proven in Section 3, and the lower bound in Section 4.

We also prove partial results in more general cases:

**Theorem 9.** *Let  $X$  be a bounded subset of  $\mathbb{R}^m$ . If  $\dim_{box}(X) > m - 1/2$ , then*

$$\dim_{box}(X) \leq \dim_{PH}^1(X) \leq m .$$

Furthermore, we show that the example in Proposition 6 cannot be taken to be a subset of  $\mathbb{R}^m$ .

**Theorem 10.** *If  $X$  is a bounded subset of  $\mathbb{R}^m$  then*

$$\dim_{PH}^1(X) \leq \dim_{box}(X) .$$

In the process of proving this, we also show:

**Theorem 11.** *If  $\mathbf{x}$  is a finite subset of  $\mathbb{R}^m$  then the first-dimensional persistent homology of the Rips complex of  $\mathbf{x}$  contains  $O(|\mathbf{x}|)$  intervals.*

In previous work with MacPherson [21], we defined an alternate notion of persistent homology dimension that measures the complexity of a shape rather than a classical notion of fractal dimension. In Appendix A, we show that it is a lower bound for  $\dim_{PH}^1$ .

We prove the upper bounds for Theorems 8 and 9 in Section 3, and the lower bounds in Section 4. The results and arguments will be stated for the Čech complex, but we will indicate which ones also work for the Rips complex. We prove our results specific to the Rips complex in Section 5. First, we cover some preliminaries.

2. PRELIMINARIES

In the introduction we defined the upper box dimension in terms of  $N_\delta(X)$ , maximal number of disjoint closed balls centered at points of a bounded metric space  $X$  :

$$\dim_{\text{box}}(X) = \limsup_{\delta \rightarrow 0} \frac{\log(N_\delta(X))}{\log(1/\delta)} .$$

Note that this is equivalent to taking  $\dim_{\text{box}}(X)$  to be the unique real number  $d$  so that

- $N_\delta(X) = O(\delta^{-\alpha})$  for all  $\alpha > d$ . That is, for all  $\alpha > d$  there exists a  $C > 0$  so that  $N_\delta(X) < C\delta^{-\alpha}$  for all  $\delta > 0$ .
- $N_\delta(X) \neq O(\delta^{-\alpha})$  for all  $\alpha < d$ . That is, for all  $\alpha < d$  and all  $D > 0$  there exists a sequence  $\delta_j \rightarrow 0$  so that  $N_{\delta_j}(X) > D\delta_j^{-\alpha}$  for all  $j$ .

There are other equivalent definitions. In particular  $N_\delta(X)$  can be replaced by the minimum number of balls of radius  $\delta$  required to cover  $X$  or, if  $X \subset \mathbb{R}^m$ , the number of cubes in a regular square tessellation of  $\mathbb{R}^m$  of width  $\delta$  which intersect  $X$ . See Falconer [15] for details.

Also, if  $\Delta$  is an  $k$ -dimensional simplex in  $\mathbb{R}^m$ , the **circumsphere** of  $\Delta$  is the smallest  $m - 1$ -dimensional sphere containing the vertices of  $\Delta$ . The **circumradius** of  $\Delta$  is the radius of its circumsphere.

In the following, bold lower case letters will denote finite point sets (i.e.  $\mathbf{x}, \mathbf{y}$ ) .

**2.1. Persistent Homology.** Here, we provide a brief introduction to persistent homology. For a more thorough exposition, see see [14, 12, 4]. We assume the reader is familiar with the basics of simplicial homology, as in [17].

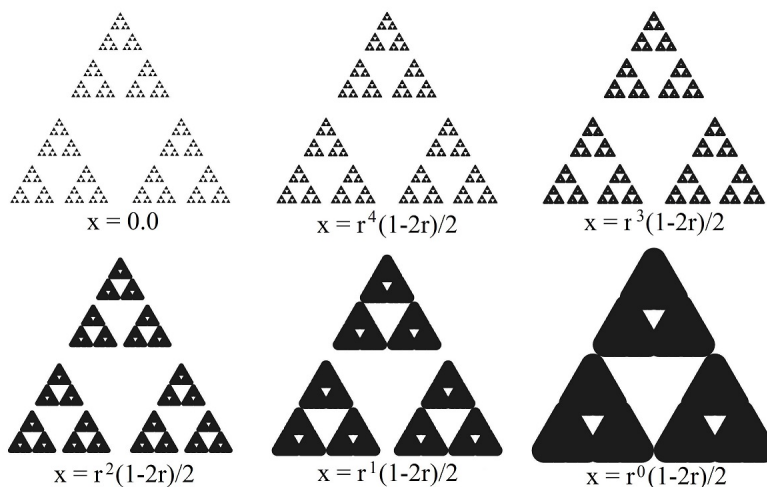
**2.1.1. Filtrations.** A **filtration** is a sequence of topological spaces  $\{S_\alpha\}_{\alpha \in I}$  together with inclusion maps  $i : S_\alpha \rightarrow S_\beta$  for  $\alpha, \beta \in I, \alpha < \beta$ , where  $I$  is an ordered index set — usually the positive real numbers, the natural numbers, or a finite set of real numbers. For example, if  $X$  is a subset of a metric space  $M$  and  $\epsilon > 0$  the  $\epsilon$ -neighborhood filtration of  $X$  is  $\{X_\epsilon\}_{\epsilon \in \mathbb{R}^+}$  where

$$X_\epsilon = \{x \in M : d(x, y) < \epsilon \text{ for some } y \in X\} .$$

See Figure 2 for an example.

We define three commonly used filtrations, each indexed by a positive real number  $\epsilon$ . If  $X$  is a subset of a metric space, the **Cech complex** on  $X$  is the filtration of simplicial complexes  $C_\epsilon(X)$ , defined by

$$(x_1, \dots, x_k) \in C_\epsilon(X) \text{ if } \bigcap_{j=1}^k B_\epsilon(x_j) \neq \emptyset ,$$



**Figure 2.** A self-similar fractal and its  $r$ -neighborhoods [21].

where  $x_1, \dots, x_k$  are points in  $X$ . For example if  $X$  is a subset of Euclidean space, the 2-simplex formed by the vertices of an acute triangle enters the Čech complex when  $\epsilon$  is the circumradius of that triangle. Also in the Euclidean case,  $C_\epsilon(X)$  is homotopy equivalent to the  $\epsilon$ -neighborhood  $X_\epsilon$ .

If  $\mathbf{x}$  is a finite subset of Euclidean space, we can define another filtration called the Alpha complex on  $\mathbf{x}$ ,  $A_\epsilon(\mathbf{x})$ . It is smaller than the Čech complex on  $\mathbf{x}$ , but contains equivalent topological information:  $A_\epsilon(\mathbf{x})$  is homotopy equivalent to  $C_\epsilon(\mathbf{x})$  for all  $\epsilon > 0$ . [13] We define the Alpha complex for point sets in general position; a more general definition is in [13]. First, we require the definition of the Delaunay triangulation [11, 7].

Let  $\mathbf{x}$  be a finite subset of  $\mathbb{R}^m$  in general position. The **Delaunay triangulation**  $DT(\mathbf{x})$  on  $\mathbf{x}$  is the unique triangulation of  $\mathbb{R}^m$  with the property that if  $\Delta$  is any  $m$ -simplex in  $DT(\mathbf{x})$  and  $S(\Delta)$  is its circumsphere, then no point of  $\mathbf{x}$  is contained in the bounded component of  $S(\Delta)$ .

Now, we can define the Alpha complex of a point set  $\mathbf{x} \subset \mathbb{R}^m$  in general position. The **Alpha complex**  $A_\epsilon(\mathbf{x})$  is a filtration of subcomplexes of the Delaunay triangulation on  $\mathbf{x}$ ,  $DT(\mathbf{x})$ . A simplex  $\sigma \in DT(\mathbf{x})$  is contained in  $A_\epsilon(\mathbf{x})$  if either of the following two conditions is met:

- The circumradius of  $\sigma$  is less than or equal to  $\epsilon$ , and no points of  $\mathbf{x}$  are contained in the bounded component of the complement of its circumsphere.
- $\sigma$  is a simplex of a higher-dimensional simplex contained in  $A_\epsilon(\mathbf{x})$ .

Finally, we define the Rips complex [30, 10] of a metric space which, unlike the two previous constructions, does not depend on an ambient space. If  $X$  is a metric space, the **Rips complex**  $R_\epsilon(X)$  is the simplicial complex with vertex set  $X$  so that

$$(x_1, \dots, x_l) \in R_\epsilon(X) \text{ if } d(x_j, x_k) \leq \epsilon \text{ for } 1 \leq j < k \leq l,$$

where  $x_1, \dots, x_l$  are points in  $X$ .

**2.1.2. Definition of Persistent Homology.** Let  $H_i(X)$  denote the reduced homology of a topological space, with coefficients in a field  $k$ .<sup>1</sup> If  $X_\alpha$  is a filtration of topological spaces, the **persistence module** of  $X_\alpha$ , is the product  $\prod_\alpha H_i(X_\alpha)$ , together with the maps  $i_{\alpha,\beta} : H_i(X_\alpha) \rightarrow H_i(X_\beta)$  for  $\alpha < \beta$ . If the rank of  $i_{\alpha,\beta}$  is finite for all  $\alpha < \beta$ , the structure of the persistent homology of  $X_\alpha$  is captured by a unique set of intervals [31, 5]  $PH_i(X)$  so that

$$\text{rank } i_{\alpha,\beta} = \left| \{I \in PH_i(X) : [\alpha, \beta] \subseteq I\} \right|.$$

If  $Y$  is a compact subset of a metric space, the Rips and Čech filtrations on  $X$  satisfy the required finiteness hypotheses [6, 4]. In the following  $PH_i(X)$  will refer to the set of intervals of the  $i$ -dimensional persistent homology of the Čech filtration of  $X$  and  $\widehat{PH}_i(X)$  will refer to the  $i$ -dimensional persistent homology of the Rips filtration of  $X$ . Note that  $PH_i(X)$  depends on the ambient space.

We will repeatedly use a stability theorem for persistent homology [8, 6]. If  $X$  and  $Y$  are filtrations, let the bottleneck distance between  $PH_i(X) = \{(b_i, d_i)\}$  and  $PH_i(Y) = \{(\hat{b}_j, \hat{d}_j)\}$  be

$$d_B(PH_i(X), PH_i(Y)) = \inf_\eta \sup_j \max \left( \left| b_j - \hat{b}_{\eta(j)} \right|, \left| d_j - \hat{d}_{\eta(j)} \right| \right)$$

where  $\eta$  ranges over all partial matchings between the intervals  $PH_i(X)$  and  $PH_i(Y)$ , allowing intervals from either set to be matched to intervals along the diagonal (that is, ones of the form  $(d, d)$ ). Also, let  $d_H$  denote the Hausdorff distance between subsets of a metric space

$$d_H(X, Y) = \max \left( \sup_{x \in X} \inf_{y \in Y} d(x, y), \sup_{y \in Y} \inf_{x \in X} d(x, y) \right).$$

**Theorem 12.** (*Stability of the Bottleneck Distance* [8]) *If  $X$  and  $Y$  are bounded subsets of a metric space and  $i \in \mathbb{N}$  then*

$$d_B(PH_i(X), PH_i(Y)) \leq d_H(X, Y).$$

<sup>1</sup>We assume the reader is familiar with the basics of simplicial homology, as in [17].

In particular, if  $I_{i,\epsilon}(X)$  is the set of  $PH_i$  intervals of  $X$  of length greater than  $\epsilon$

$$(1) \quad I_{i,\epsilon}(\mathbf{x}) = \{(b, d) : PH_i(\mathbf{x}) : d - b > \epsilon\} ,$$

then we have the following result.

**Corollary 13.** *Let  $X$  and  $Y$  be compact subsets of a metric space and  $\epsilon, \delta > 0$ . If  $d_H(X, Y) < \delta/2$  then*

$$|I_{i,\epsilon+\delta}(X)| \leq |I_{i,\epsilon}(X)| .$$

The preceding theorem and corollary are also true for  $\widetilde{PH}_i$ .

**2.2. Persistent Homology Dimension.** Recall that the  $PH_i$  dimension of a bounded subset of a metric space is

$$\dim_{PH}^i(X) = \inf \left\{ \alpha : \exists C \text{ so that } E_\alpha^i(\mathbf{x}) < C \forall \text{ finite } \mathbf{x} \subset X \right\} .$$

A straightforward argument based on stability of the bottleneck distance shows that we could modify the definition of  $\dim_{PH}^i(X)$  to consider all compact subsets of  $X$ , rather than all finite subsets:

$$\dim_{PH}^i(X) = \inf \left\{ \alpha : \exists C \text{ so that } E_\alpha^i(Y) < C \forall \text{ compact } Y \subseteq X \right\} .$$

In Proposition 17 below, we provide another equivalent definition of the  $PH_i$  dimension, in terms of the asymptotic number of “long” intervals of “well-spaced” point sets of  $X$ .

Also, for the Rips complex define

$$\widetilde{E}_\alpha^i(\mathbf{x}) = \sum_{(b,d) \in \widetilde{PH}_i(\mathbf{x})} (d - b)^\alpha ,$$

where the sum is taken over the finite intervals of  $\widetilde{PH}_i(\mathbf{x})$ . With that,

$$\dim_{\widetilde{PH}}^i(X) = \inf \left\{ \alpha : \exists C \text{ so that } \widetilde{E}_\alpha^i(\mathbf{x}) < C \forall \text{ finite } \mathbf{x} \subset X \right\} .$$

### 3. AN UPPER BOUND

Our strategy for bounding  $\dim_{PH}^i(X)$  in terms of  $\dim_{\text{box}}(X)$  is to approximate subsets of  $X$  by finite point sets whose size is controlled by the box dimension, and apply bottleneck stability. In this section, Lemmas 14 and 16 and Propositions 17 and 18 hold for the Rips complex using identical arguments, but Corollaries 15 and 20 only apply to the Čech complex.



**Lemma 14.** *Let  $X$  be a bounded metric space and suppose there are positive real numbers  $c$  and  $D_0$  so that for all  $\epsilon > 0$  and all  $\mathbf{x} \subset X$*

$$(2) \quad |I_{i,\epsilon}(\mathbf{x})| < D_0 \epsilon^{-c},$$

where  $I_{i,\epsilon}(\mathbf{x})$  is defined in Equation 1. Then  $\dim_{PH}^i(X) \leq c$ .

*Proof.* Rescale  $X$  if necessary so its diameter is less than one. Note that this implies that  $|I| \leq 1$  for all  $I \in PH_i(X)$ .

Let  $\alpha > c$  and  $\mathbf{x} \subseteq X$ . We will bound  $E_i^\alpha(\mathbf{x})$  by summing over the contributions of the intervals whose lengths are between  $2^{-k-1}$  and  $2^{-k}$ . For  $k \in \mathbb{N}$  let

$$(3) \quad J_{i,k}(\mathbf{x}) = \left\{ I \in PH_i(\mathbf{x}) : 2^{-k-1} < |I| \leq 2^{-k} \right\}.$$

Then

$$\begin{aligned} E_\alpha^i(\mathbf{x}) &= \sum_{k=0}^{\infty} \sum_{I \in J_{i,k}(\mathbf{x})} |I|^\alpha \\ &\leq \sum_{k=0}^{\infty} |J_{i,k}(\mathbf{x})| 2^{-\alpha k} && \text{by Eqn. 3} \\ &\leq \sum_{k=0}^{\infty} |I_{i,2^{-k-1}}(\mathbf{x})| 2^{-\alpha k} && \text{by Eqn. 1} \\ &\leq \sum_{k=0}^{\infty} D_0 2^{kc+c} 2^{-\alpha k} && \text{by Eqn. 2} \\ &= D_0 2^c \sum_{k=0}^{\infty} \left( 2^{c-\alpha} \right)^k \\ &= D_0 2^c \frac{1}{1 - 2^{c-\alpha}} && \text{because } \alpha > c. \end{aligned}$$

Therefore,  $E_\alpha^i(\mathbf{x})$  is uniformly bounded for all  $\mathbf{x} \subset X$ ,  $\dim_{PH}^i(X) \leq \alpha$  for all  $\alpha > c$ , and  $\dim_{PH}^i(X) \leq c$ .  $\square$

As noted by Cohen-Steiner et. al. [9], if  $\mathbf{x}$  is a subset of a triangulable metric space  $M$  then  $|I_{i,\epsilon}(\mathbf{x})|$  is less than the number of simplices in a triangulation of mesh  $\epsilon$  of  $M$ . In particular, if  $\mathbf{x}$  is a finite subset of  $\mathbb{R}^m$  (or an  $m$ -dimensional Riemannian manifold) then  $|I_{i,\epsilon}(\mathbf{x})| = O(\epsilon^{-m})$ .

**Corollary 15.** (Cohen-Steiner, Edelsbrunner, Harer, and Mileyko [9]) *Let  $X$  be a bounded subset of  $\mathbb{R}^m$ . Then*

$$\dim_{PH}^i(X) \leq m.$$

Which is the upper bound in Theorem 9.

In the following  $X$  will be a bounded metric space, and  $\mathbf{x}^\epsilon$  will be the centers of a maximal collection of disjoint balls of radius  $\epsilon/4$  centered at points of  $X$ . By the maximality of  $\mathbf{x}^\epsilon$ , the balls of radius  $\epsilon/2$  centered at the points of  $\mathbf{x}^\epsilon$  cover  $X$  and

$$(4) \quad d_H(\mathbf{x}^\epsilon, X) < \epsilon/2.$$

**Lemma 16.** *Let  $X$  be a bounded metric space and let  $c, D > 0$ . For  $\epsilon > 0$  let  $\mathbf{x}^\epsilon$  be the centers of a maximal collection of disjoint balls of radius  $\epsilon/4$  centered at points of  $X$ . If*

$$(5) \quad |I_{i,\epsilon}(\mathbf{y})| < D\epsilon^{-c} \text{ for all } \mathbf{y} \subseteq \mathbf{x}^\epsilon,$$

for all  $\epsilon > 0$  then

$$\dim_{PH}^i(X) \leq c.$$

*Proof.* Rescale the metric if necessary so that the diameter of  $X$  is less than one, and let  $\alpha > c$ . We will show that  $E_\alpha^i(\mathbf{y})$  is uniformly bounded for all  $\mathbf{y} \subset X$ .

Let  $\mathbf{y} \subset X$ ,  $\epsilon > 0$ , and

$$\mathbf{y}^\epsilon = \{x \in \mathbf{x}^\epsilon : d(x, \mathbf{y}) < \epsilon/2\}.$$

Every point of  $\mathbf{y}^\epsilon$  is within distance  $\epsilon/2$  of a point of  $\mathbf{y}$ , and every point of  $\mathbf{y}$  is within distance  $\epsilon/2$  of a point of  $\mathbf{y}^\epsilon$  because  $d_H(\mathbf{x}^\epsilon, X) < \epsilon/2$ . It follows that  $d_H(\mathbf{y}^\epsilon, \mathbf{y}) < \epsilon/2$ .

By stability (Corollary 13),

$$|I_{i,2\epsilon}(\mathbf{y})| \leq |I_{i,\epsilon}(\mathbf{y}^\epsilon)| \leq D\epsilon^{-c},$$

for all  $\epsilon > 0$ . Then

$$|I_{i,\epsilon}(\mathbf{y})| \leq 2^c D\epsilon^{-c},$$

and the desired result follows from Lemma 14.  $\square$

Next, we show that we can characterize  $\dim_{PH}^i(X)$  in terms of the number of “long” intervals of “well-spaced” point sets in  $X$ . If  $X$  is a bounded metric space and  $\mathbf{x}^\epsilon$  is a finite point set consisting of the centers of a maximal collection of balls of radius  $\epsilon/4$ , centered at points of  $X$  let

$$g(\epsilon) = \max_{\mathbf{y} \subseteq \mathbf{x}^\epsilon} |I_{i,\epsilon}(\mathbf{y})|.$$

**Proposition 17.** *Let  $X$  be a bounded subset of a metric space and let*

$$c_i(X) = \limsup_{\epsilon \rightarrow 0} \frac{\log(g(\epsilon))}{\log(1/\epsilon)}.$$

Then

$$\dim_{PH}^i(X) = c_i(X).$$

*Proof.* Let  $\alpha > c_i(X)$ , so there is a  $D_1 > 0$  so that  $g(\epsilon) < \epsilon^{-\alpha}$  for all  $\epsilon > 0$ . By the definition of  $g(\epsilon)$ , the hypotheses of Lemma 16 are satisfied with  $c = \alpha$ . Therefore,  $\dim_{PH}^i(X) \leq \alpha$  for all  $\beta > c_i(X)$  and  $\dim_{PH}^i(X) \leq c_i(X)$ .

Conversely, suppose that  $\alpha < c_i(X)$ , and let  $\alpha < \beta < c_i(X)$ . For  $D_2 > 0$  there exists a sequence  $\epsilon_j \rightarrow 0$  so that  $g(\epsilon_j) > D_2 \epsilon_j^{-\beta}$  for all  $j \in \mathbb{N}$ . For each  $j \in \mathbb{N}$  there exists a  $\mathbf{y}_j \subset \mathbf{x}^{\epsilon_j}$  so that

$$\left| I_{i, \epsilon_j}(\mathbf{y}_j) \right| > D_2 \epsilon_j^{-\beta}.$$

Then

$$E_\alpha^i(\mathbf{y}_j) \geq \epsilon_j^\alpha \left| I_{i, \epsilon_j}(\mathbf{y}_j) \right| \geq D_2 \epsilon_j^{\alpha - \beta},$$

which limits to  $\infty$  as  $\epsilon_j \rightarrow 0$  because  $\alpha < \beta$ . Therefore,  $\dim_{PH}^i(X) \geq \alpha$  for all  $\alpha > c_i(X)$  and  $\dim_{PH}^i(X) = c_i(X)$ .  $\square$

In our next proposition, we show a relationship between the  $PH_i$  dimension and the upper box dimension.

**Proposition 18.** *Let  $X$  be a bounded metric space, and let  $\phi_X^i(n)$  be the maximal number of  $PH_i$  intervals of a set of  $n$  points in  $X$ . If  $\phi_X^i(n) = O(n^\lambda)$  then*

$$\dim_{PH}^i(X) \leq \lambda \dim_{\text{box}}(X).$$

*Proof.* Let  $\alpha > \dim_{\text{box}}(X)$ , so there exists a  $C' > 0$  so that

$$(6) \quad |\mathbf{x}^\epsilon| < C' \epsilon^{-\alpha}$$

for all  $\epsilon > 0$ . Also, there is a  $D_3 > 0$  so that

$$(7) \quad \phi_X^i(n) < D_3 n^\lambda$$

for all  $n > 0$ . Then, if  $\mathbf{y} \subseteq \mathbf{x}^\epsilon$ ,

$$\begin{aligned}
 |I_{i,\epsilon}(\mathbf{y})| &\leq |PH_i(\mathbf{y})| \\
 &< D_3 |\mathbf{y}|^\lambda && \text{by Eqn. 7} \\
 &\leq D_3 |\mathbf{x}^\epsilon|^\lambda && \text{because } \mathbf{y} \subseteq \mathbf{x}^\epsilon \\
 &< D_3 (C'\epsilon^{-\alpha})^\lambda && \text{by Eqn. 6} \\
 &\leq D_3 C'\epsilon^{-\lambda\alpha},
 \end{aligned}$$

and  $g(\epsilon) < D_3 C'\epsilon^{-\lambda\alpha}$  for all  $\epsilon > 0$ . By the previous lemma,

$$\begin{aligned}
 \dim_{PH}^i(X) &= \limsup_{\epsilon \rightarrow 0} \frac{\log(g(\epsilon))}{\log(1/\epsilon)} \\
 &\leq \lim_{\epsilon \rightarrow 0} \frac{\log(D_3 C'\epsilon^{-\lambda\alpha})}{\log(1/\epsilon)} \\
 &= \lim_{\epsilon \rightarrow 0} \frac{-\lambda\alpha \log(\epsilon)}{\log(1/\epsilon)} \\
 &= \lambda\alpha
 \end{aligned}$$

for all  $\alpha > \dim_{\text{box}}(X)$ . Therefore,

$$\dim_{PH}^i(X) \leq \lambda \dim_{\text{box}}(X).$$

□

If  $X \subset \mathbb{R}^m$  we can apply the well-known Upper Bound Theorem on the maximal number of simplices of a Delaunay triangulation to bound  $\phi_X^i(n)$ .

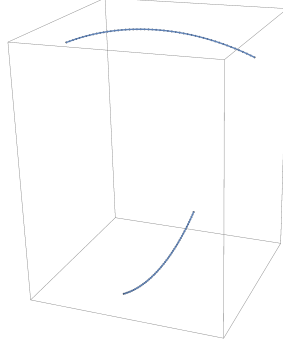
**Theorem 19** (The Upper Bound Theorem [27, 22]). *The maximum number of simplices of a Delaunay triangulation with  $n$  vertices in  $\mathbb{R}^m$  is*

$$\binom{n - \lfloor \frac{m+1}{2} \rfloor}{n-m} + \binom{n - \lfloor \frac{m+2}{2} \rfloor}{n-m} = O\left(n^{\lfloor \frac{m+1}{2} \rfloor}\right).$$

Furthermore, if  $0 < i < \lfloor \frac{m+1}{2} \rfloor$  the maximum number of  $i$ -simplices of a Delaunay triangulation with  $n$  vertices in  $\mathbb{R}^m$  is

$$f_i(m, n) = \binom{n}{i+1} = O\left(n^{i+1}\right).$$

The upper bound is sharp, and is achieved by taking the Delaunay triangulation on the vertices of a cyclic polytope [23]. We have the following corollary.



**Figure 3.** The set  $X$  : two opposing arcs from unit circles.

**Corollary 20.** *Let  $X$  be a bounded subset of  $\mathbb{R}^m$  then*

$$\dim_{PH}^i(X) \leq \min\left((i+1), \lfloor \frac{m+1}{2} \rfloor\right) \dim_{box}(X) .$$

*In particular, if  $m = 2$*

$$\dim_{PH}^1(X) \leq \dim_{box}(X) .$$

*Proof.* Let  $\mathbf{x}$  be a finite subset of  $\mathbb{R}^m$ . If necessary, we may perturb the points of  $\mathbf{x}$  by an amount less than  $\delta/2$ , where  $\delta$  is the minimum length of a  $PH_i$  interval of  $\mathbf{x}$  to place the points in general position without decreasing the number of intervals (by Corollary 13). The number of  $PH_i$  intervals of  $\mathbf{x}$  is bounded above by the number of  $i$ -simplices in the Alpha complex on  $X$ , which equals the number of  $i$ -simplices in the Delaunay triangulation. Therefore, by the Upper Bound Theorem,

$$|PH_i(\mathbf{x})| \leq \binom{n}{i+1} = O(n^\lambda) ,$$

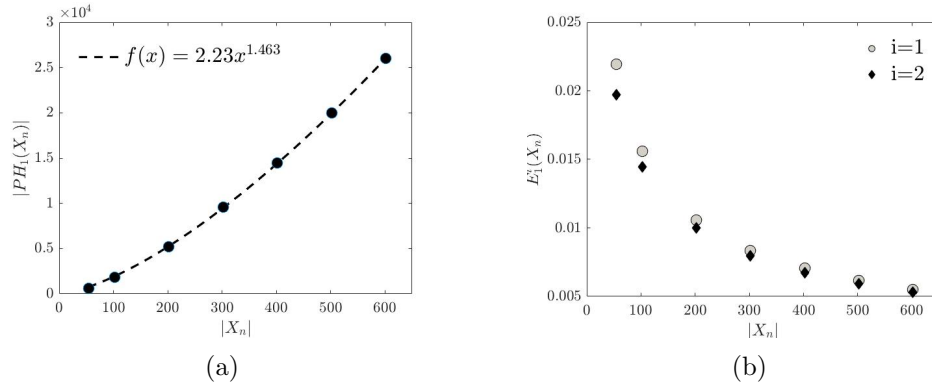
where  $\lambda = \min\left((i+1), \lfloor \frac{m+1}{2} \rfloor\right)$ .

It follows that if  $X$  is a bounded subset of  $\mathbb{R}^m$  then

$$\phi_X^i(n) = O(n^\lambda) ,$$

and the desired result follows from Proposition 18. □

This completes the proof of the upper bound in Theorem 8.



**Figure 4.** (a) The number of  $PH_1$  intervals of  $\mathbf{x}_n$ , which appears to grow approximately as  $n^{1.5}$ . The data for  $PH_2$  is similar. (b) The quantities  $E_1^1(\mathbf{x}_n)$  and  $E_1^2(X_n)$ , which appear to be bounded as a function of  $n$ . Persistent homology was computed by calculating the alpha complex with CGAL [29] and passing the resulting filtration to JPLEX [26].

**3.1. An Example with Many Intervals.** We present an example of a subset of  $\mathbb{R}^3$  with subsets of size  $n$  that appear to have  $\approx n^{1.5}$   $PH$  intervals. This example was suggested by Herbert Edelsbrunner.

Let  $C_1$  and  $C_2$  be the following arcs from unit circles in  $\mathbb{R}^3$  :

$$C_1 = \left\{ (\cos(\theta), \sin(\theta), 0) : \theta \in [-\pi/8, \pi/8] \right\}$$

$$C_2 = \left\{ (1 - \cos(\theta), 0, \sin(\theta)) : \theta \in [-\pi/8, \pi/8] \right\}$$

and let  $X = C_1 \cup C_2$ .  $X$  is shown in Figure 3. Let  $\mathbf{x}_n \subset X$  be a point set obtained by placing  $\lfloor n/2 \rfloor$  uniformly spaced points on each of the two circular arcs. Computations indicate that  $|PH_i(\mathbf{x}_n)| \approx |\mathbf{x}_n|^{1.5}$  for  $i = 1, 2$  (Figure 4(a)), but that  $E_1^1(\mathbf{x}_n)$  and  $E_1^2(\mathbf{x}_n)$  are bounded as functions of  $n$  (Figure 4(b)). As such, the question remains of whether the  $PH_i$  dimension is bounded above by the upper box dimension for any subset of Euclidean space. This example suggests that it may be necessary to bound the number of “long” intervals of a “well-spaced” point set rather than the total number of intervals to achieve a sharp result.

## 4. A LOWER BOUND

In this section, we prove the lower bounds in Theorems 8 and 9. However, we first consider the case of a bounded subset of Euclidean space with non-empty interior. Our later arguments will be more complicated, but have the same general outline.

**Proposition 21.** *If  $X$  is a bounded subset of  $\mathbb{R}^m$  with non-empty interior then*

$$\dim_{PH}^i(X) = m$$

for  $i = 0, \dots, m - 1$ .

*Proof.* The upper bound was shown in Corollary 15.

Let  $0 \leq i \leq m - 1$ . First, we compute the  $i$ -dimensional persistent homology of the vertex set of the regular  $i + 1$ -simplex  $\sigma$  with edge length  $\sqrt{2}$ . We may assume that the simplex is formed by the standard basis vectors  $e_1, \dots, e_{i+1}$  in  $\mathbb{R}^{i+1}$ . The circumcenter of  $\sigma$  is  $c := \left(\frac{1}{i+1}, \dots, \frac{1}{i+1}\right)$ , and the circumradius is

$$d(e_j, c) = \sqrt{\left(1 - 1/i\right)^2 + \frac{i}{(i+1)^2}} = \sqrt{\frac{i}{i+1}},$$

so  $\sigma$  enters the alpha complex  $A_\epsilon(\sigma)$  when  $\epsilon = \sqrt{\frac{i}{i+1}}$ . The same computation shows that the  $i$ -dimensional faces of  $\sigma$  enter  $A_\epsilon(\sigma)$  when  $\epsilon = \sqrt{\frac{i-1}{i}} = \sqrt{1 - \frac{1}{i}}$ . Therefore,  $PH_i(\sigma)$  consists of the single interval  $\left(\sqrt{1 - \frac{1}{i}}, \sqrt{\frac{i}{i+1}}\right)$ .

For convenience, let  $\tau$  be the length of the single  $PH_1$  interval of the regular  $i + 1$ -simplex with edge length 1,

$$\tau = \frac{1}{\sqrt{2}} \left( \sqrt{\frac{i}{i+1}} - \sqrt{1 - \frac{1}{i}} \right)$$

and  $\hat{d}$  be the death-time of that interval,

$$\hat{d} = \sqrt{\frac{i}{2i+2}}.$$

$X$  has non-empty interior so we can find an  $m$ -dimensional cube  $C \subset X$ . Rescale  $X$  if necessary so that  $C$  is a unit cube. Fix  $n \in \mathbb{N}$  and sub-divide  $C$  into  $n^m$  cubes of

width  $1/n$ . Let  $C_1, \dots, C_t$  be a maximal sub-collection of these cubes so that

$$(8) \quad d(C_k, C_l) > 2\frac{\hat{d}}{n}$$

for all  $k, l \in \{1, \dots, t\}$  so that  $k \neq l$  (where  $d(C_k, C_l)$  is the minimum distance between any pair of points in the two cubes). There is a constant  $F$  depending only on  $i$  and  $m$  so that  $t > Fn^m$  for all sufficiently large  $n$ .

For  $j = 1, \dots, t$  find a regular  $i + 1$ -simplex of edge length  $\frac{1}{n}$  inside each cube  $C_j$ . Let  $\mathbf{x}_j$  be the vertices of this simplex, and let  $\mathbf{z}_n = \cup_{j=1}^t \mathbf{x}_j$ . By Equation 8, the Čech complex on  $\mathbf{z}_n$  equals the disjoint union of the Čech complexes on  $\mathbf{x}_1, \dots, \mathbf{x}_t$  for any filtration value less than  $\frac{\hat{d}}{n}$  (the first time an edge between a point in  $\mathbf{x}_j$  and one in  $\mathbf{x}_k$  can enter the complex for  $j \neq k$ ).  $PH_i(\mathbf{x}_j)$  consist of a single interval of length  $\frac{1}{n}\tau$  that dies at time  $\frac{\hat{d}}{n}$ . It follows that  $PH_i(\mathbf{z}_n)$  contains at least  $Fn^m$  intervals of length  $\frac{1}{n}\tau$ .

Let  $\alpha < m$ , and compute

$$E_\alpha^i(\mathbf{z}_n) \geq Fn^m \left(\frac{\tau}{n}\right)^\alpha = F\tau^\alpha n^{m-\alpha}$$

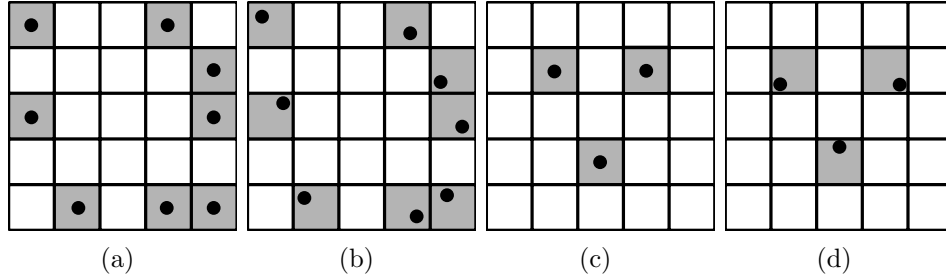
which limits to  $\infty$  as  $n \rightarrow \infty$ , because  $\alpha < m$ . Therefore,  $\dim_{PH}^i(X) \geq \alpha$  for any  $\alpha < m$  and  $\dim_{PH}^i(X) = m$ , as desired.  $\square$

In the preceding proof, we could find a cube in Euclidean space for which each sub-cube was occupied, and found point sets with non-trivial homology inside those cubes. If  $X$  is a bounded subset of  $\mathbb{R}^m$ , and  $\dim_{\text{box}} X < m$ , most cubes will not be occupied. As such, we need to proceed with care.

Let  $[N]$  denote the integers  $1, \dots, n$  and let  $[N]^m \subset \mathbb{Z}^m$  be  $[N] \times [N] \times \dots \times [N]$ . For each  $x \in \delta\mathbb{Z}^m$ , let the cube corresponding to  $x$  be the cube of width  $\delta$  centered at  $x$ . The grid of mesh  $\delta$  is the set of all cubes in this tessellation. If  $X$  is a bounded subset of  $\mathbb{R}^m$ , the upper box dimension of  $X$  controls the number of cubes in the grid of mesh  $\delta$  that intersect  $X$ . Our strategy to prove a lower bound for the  $PH$  dimension of  $X$  in terms of the upper box dimension is to show that a sufficiently large collection of points, each in a distinct cube of  $[N]^m$ , must contain a subset with non-trivial persistent homology.

**Definition 22.**  $\mathbf{x} \subset \mathbb{Z}^m$  has a **stable**  $PH_i$ -class if any point collection  $\mathbf{y}$  consisting of exactly one point in each cube corresponding to a point of  $\mathbf{x}$  has a  $PH_i$  interval  $I$  so that  $|I| \geq c$  for some constant  $c > 0$  (see Figure 5). The supremal such  $c$  is called the **size** of the stable persistence class.





**Figure 5.** The  $PH_1$  class of the lattice points corresponding to the gray cubes in (a) and (b) is stable — any choice of one point in each cube will give the vertices of an acute triangle, and therefore a set with non-trivial  $PH_1$ . The one in (c) and (d) is not, because the points in (d) form an obtuse triangle so the persistent homology of that set is trivial.

The stability of the bottleneck distance immediately implies the following.

**Proposition 23.** *If  $\mathbf{x} \subset \mathbb{Z}^m$  and there is an interval  $I \in PH_i(\mathbf{x})$  with*

$$|I| > \sqrt{m} + \epsilon,$$

*then  $\mathbf{x}$  has a stable  $PH_i$ -class of size at least  $\epsilon$ .*

*Proof.* If  $\mathbf{x}$  and  $\mathbf{y}$  are in Definition 22, then  $d_H(\mathbf{x}, \mathbf{y}) < \sqrt{m}/2$ , the maximal distance between a point in a unit cube and the center of that cube. By bottleneck stability,  $PH_i(\mathbf{y})$  has an interval of length at least  $\epsilon$ .  $\square$

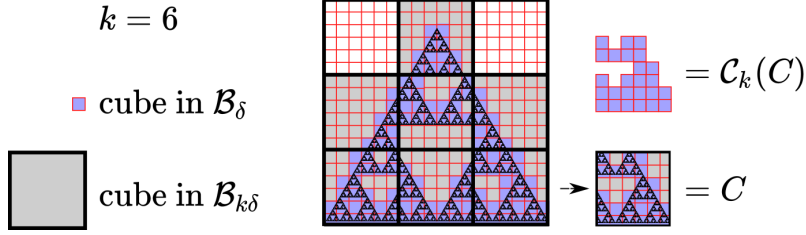
**Definition 24.** *Let  $\xi_i^m(N)$  be the size of the largest subset  $\mathbf{x}$  of  $[N]^m$  so that no subset  $\mathbf{y}$  of  $\mathbf{x}$  has a stable  $PH_i$ -class of size greater than 1. That is,  $\xi_i^m(N)$  is the smallest integer so that if  $|\mathbf{x}| > \xi_i^m(N)$ , then there exist a  $\mathbf{y} \subset \mathbf{x}$  with a stable  $PH_i$  class. Define*

$$\gamma_i^m = \liminf_{N \rightarrow \infty} \frac{\log(\xi_i^m(N))}{\log(N)}.$$

The following proposition is stated for the Čech complex and  $\gamma_i^m$  is defined for that filtration, but an analogous result holds for the Rips complex (though the corresponding constant  $\hat{\gamma}_i^m$  may be different).

**Proposition 25.** *If  $X$  is a bounded subset of  $\mathbb{R}^m$  and  $\dim_{\text{box}}(X) > \gamma_i^m$  then*

$$\dim_{PH}^1(X) \geq \dim_{\text{box}}(X).$$



**Figure 6.** The box notation used in the proofs of Proposition 25 and Lemma 26.

Before proving this proposition, we introduce notation and state a technical lemma related to box-counting. The notation is illustrated in Figure 6. For  $\delta \in \mathbb{R}$ , let  $\mathcal{B}_\delta$  be the collection of cubes in the grid of mesh  $\delta$  in  $\mathbb{R}^m$  that intersect  $X$ . Also, if  $k \in \mathbb{N}$ , consider the larger cubes  $\mathcal{B}_{k\delta}$ . Note that each cube of  $\mathcal{B}_\delta$  is contained in a unique cube of  $\mathcal{B}_{k\delta}$ , and each cube of  $\mathcal{B}_{k\delta}$  contains at least one cube of  $\mathcal{B}_\delta$ . If  $C \in \mathcal{B}_{k\delta}$ , let

$$\mathcal{C}_k(C) = \{B \in \mathcal{B}_\delta : B \subset C\} .$$

Furthermore, for  $\lambda > 0$  define

$$\mathcal{D}_{\delta,k}^\lambda = \left\{ C \in \mathcal{B}_{k\delta} : |\mathcal{C}_k(C)| > k^\lambda \right\} .$$

The idea of the proof of Proposition 25 is to choose  $k$  and  $\lambda$  so that we can find a set of points in every cube  $\mathcal{D}_{\delta,k}^\lambda$  that has a stable  $PH_i$  class, and show that  $\delta^\alpha \left| \mathcal{D}_{\delta,k}^\lambda \right| \rightarrow \infty$  for  $\alpha < \dim_{\text{box}}(X)$ . This is a little tricky as, the definition of the upper box dimension only gives us weak control over the asymptotics of  $|\mathcal{B}_\delta|$ . As such, we cannot choose a fixed value of  $k$  and must instead let  $k$  increase slowly as  $\delta \rightarrow 0$ .

**Lemma 26.** *If  $X$  is a bounded subset of  $\mathbb{R}^m$  and  $\alpha < \lambda < \dim_{\text{box}}(X)$ , there exist a sequence of real numbers  $\delta_j \rightarrow 0$  and a sequence of natural numbers  $k_j \rightarrow \infty$  so that*

$$\delta_j^\alpha \left| \mathcal{B}_{\delta_j, k_j}^\lambda \right| \rightarrow \infty .$$

We prove Proposition 25 first, then return to the proof of Lemma 26.

*Proof of Propostion 25.* Let  $d = \dim_{\text{box}}(X)$  and  $\alpha < d$ . We will find finite point collections  $\mathbf{z}_j \subset X$  so that  $E_\alpha^i(\mathbf{z}_j) \rightarrow \infty$  as  $j \rightarrow \infty$ . Choose  $\lambda$  so that

$$\max(\gamma_i^m, \alpha) < \lambda < d .$$

By the previous lemma, there are sequences of real numbers  $\delta_j \rightarrow 0$  and  $k_j \rightarrow \infty$  so that

$$\delta_j^\alpha \left| \mathcal{D}_{\delta_j, k_j}^\lambda \right| \rightarrow \infty$$

Each cube  $C \in \mathcal{D}_{\delta_j, k_j}^\lambda$  contains at least  $k_j^\lambda$  sub-cubes of width  $\delta_j$  that intersect  $X$ .  $\lambda > \gamma_i^m$ , so for sufficiently large  $k_j$  we can find a finite point set  $\mathbf{x}(C) \in X \cap C$  so that  $PH_i(\mathbf{x}(C))$  has an interval of length at least  $\delta_j$ . To combine the contributions of multiple points sets  $\mathbf{x}(C)$ , we must thin the collection  $\mathcal{D}_{\delta_j, k_j}^\lambda$ .

Let  $\{C_1, \dots, C_t\}$  be a maximal collection of cubes in  $\mathcal{D}_{\delta_j, k_j}^\lambda$  so that  $d(C_k, C_l) > 2k_j\delta_j\sqrt{m}$  for all  $k, l \in \{1, \dots, t\}$  so that  $k \neq l$  (where  $d(C_k, C_l)$  is the minimum distance between any pair of points in the two cubes). There is a constant  $F$  depending only on the ambient dimension  $m$  so that  $t \geq F \left| \mathcal{D}_{\delta_j, k_j}^\lambda \right|$  for sufficiently large  $j$ .

Let  $\mathbf{z}_j = \cup_{k=1}^t \mathbf{x}(C_k)$ . By construction, the Čech complex on  $\mathbf{z}_j$  equals the disjoint union of the Čech complexes on  $\mathbf{x}(C_1), \dots, \mathbf{x}(C_t)$  for any filtration value less than  $k_j\delta_j\sqrt{m}$ . Furthermore, the diameter of each set  $\mathbf{x}(C_k)$  is less than  $k_j\delta_j\sqrt{m}$  so any interval  $(b_0, d_0) \in PH_i(\mathbf{x}(C_k))$  satisfies  $d_0 < k_j\delta_j\sqrt{m}$ . Therefore, as sets of intervals, we have that  $\cup_k PH_i(\mathbf{x}(C_k)) \subseteq PH_i(\mathbf{z}_j)$ . It follows that  $PH_i(\mathbf{z}_j)$  contains at  $t$  intervals of length greater than  $\delta_j$ .

Therefore,

$$E_\alpha^i(\mathbf{z}_j) \geq t\delta_j^\alpha \geq F \left| \mathcal{D}_{\delta_j, k_j}^\lambda \right| \delta_j^\alpha,$$

which goes to  $\infty$  as  $j \rightarrow \infty$  by the previous lemma. Thus  $\dim_{PH}^i(X) \geq \alpha$  for all  $\alpha < d$ , and  $\dim_{PH}^i(X) \geq d$  as desired.  $\square$

*Proof of Lemma 26.* Let  $d = \dim_{\text{box}}(X)$ , and choose  $0 < p < 1$  so that  $d - pm > \alpha$ . Also, choose

$$\epsilon < \min(2(d - pm - \alpha), pd - p\lambda).$$

Let  $\beta = d + \epsilon/2$ , and note for future reference that

$$(9) \quad \alpha + pm - \beta + \epsilon = \alpha + pm - d + \epsilon/2 < 0$$

and

$$(10) \quad p\beta - p\lambda - \epsilon = pd - p\lambda - (1 - p/2)\epsilon > pd - p\lambda - \epsilon > 0.$$

$\beta > d$  so there exists a  $\delta_0 > 0$  so that  $|B_\delta| < \delta^{-\beta}$  for all  $\delta < \delta_0$ . Also,  $\beta - \epsilon < d$  so there exists a sequence  $\delta_j \rightarrow 0$  so that  $|B_{\delta_j}| > \delta_j^{-\beta+\epsilon}$  for all  $j \in \mathbb{N}$ . We may assume that  $\delta_j < \delta_0$  for all  $j$ . Let

$$(11) \quad k_j = \lfloor \delta_j^{-p} \rfloor.$$

A cube in  $\mathcal{B}_{k_j\delta_j}$  can contain at most  $k_j^m$  cubes in  $\mathcal{B}_{\delta_j}$ , and every cube in  $\mathcal{B}_{\delta_j}$  is contained in exactly one cube  $\mathcal{B}_{k_j\delta_j}$ . It follows that  $\mathcal{D}_{\delta_j, k_j}^\lambda$  is greater than or equal to the smallest integer  $a$  so that

$$ak_j^m + \left( \left| \mathcal{B}_{k_j\delta_j} \right| - a \right) k_j^\lambda \geq \left| \mathcal{B}_{\delta_j} \right|.$$

Rearranging terms, we have that

$$\begin{aligned} a &\geq \frac{\left| \mathcal{B}_{\delta_j} \right| - \left| \mathcal{B}_{k_j\delta_j} \right| k_j^\lambda}{k_j^m - k_j^\lambda} \\ &= \frac{\left| \mathcal{B}_{\delta_j} \right| k_j^{-\lambda} - \left| \mathcal{B}_{k_j\delta_j} \right|}{k_j^{m-\lambda} - 1} \\ &\geq \frac{\delta_j^{-\beta+\epsilon} k_j^{-\lambda} - k_j^{-\beta} \delta_j^{-\beta}}{k_j^{m-\lambda} - 1} \\ &= \frac{\delta_j^{-\beta+\epsilon} \lfloor \delta_j^{-p} \rfloor^{-\lambda} - \lfloor \delta_j^{-p} \rfloor^{-\beta} \delta_j^{-\beta}}{\lfloor \delta_j^{-p} \rfloor^{m-\lambda} - 1} && \text{by Eqn. 11} \\ &\approx \frac{\delta_j^{p\lambda-\beta+\epsilon} - \delta_j^{p\beta-\beta}}{\delta^{p\lambda-pm} - 1} && \text{as } \delta_j \rightarrow 0 \\ &\approx \delta_j^{pm-\beta+\epsilon}, \end{aligned}$$

where we used that  $p\lambda - \beta + \epsilon > p\beta - \beta$  by Equation 10.

Choose a  $J$  so that if  $j > J$ ,

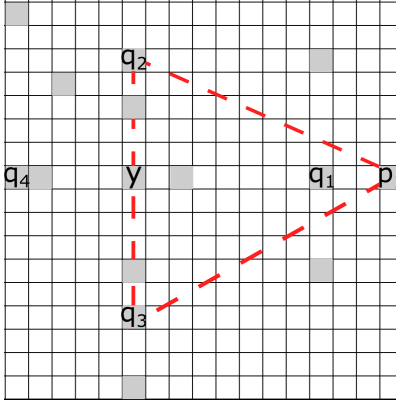
$$\left| \mathcal{D}_{\delta_j, k_j}^\lambda \right| \geq \frac{1}{2} \delta_j^{pm-\beta+\epsilon}.$$

Then, if  $j > J$ ,

$$\delta_j^\alpha \left| \mathcal{D}_{\delta_j, k_j}^\lambda \right| \geq \frac{1}{2} \delta_j^\alpha \delta_j^{pm-\beta+\epsilon} = \frac{1}{2} \delta_j^{\alpha+pm-\beta+\epsilon},$$

which by equation Equation 9, limits to  $\infty$  as  $j \rightarrow \infty$ .  $\square$

**4.1. A Bound for  $\gamma_1^m$ .** We prove an upper bound for  $\gamma_1^m$  by considering the contribution of small triangles to the persistent homology of a subset of  $\mathbb{Z}^m$ . Three points in Euclidean space give rise to a  $PH_1$ -class if and only if they are the vertices of an acute triangle, in which case the total persistence (death minus birth) equals the circumradius minus half the length of its longest edge.



**Figure 7.** The setup in Proposition 27. The points of  $X$  are shown in gray, and the triangle formed by the points  $p_1$ ,  $q_2$ , and  $q_3$  is shown by the red dashed lines. It is an acute triangle, and the triangle  $q_1, q_2$ , and  $q_3$  is either acute or right.

**Proposition 27.**

$$\gamma_1^m \geq m - 1/2.$$

*Proof.* We will prove that  $\xi_1^m(N) \in O(N^{m-1/2})$  by showing that if  $X \subset [N]^m$ ,

$$X \subset [N]^m, \quad c > \sqrt{4\sqrt{m} + 4}, \quad |X| > 8cN^{m-1/2} + 4N^{m-1},$$

and  $N$  is sufficiently large then there exist three points  $p_1, q_2$ , and  $q_3$  of  $X$  so that  $PH_1(\{p_1, q_2, q_3\})$  contains an interval of length at least  $1 + \sqrt{m}$ . The desired result will then follow from Proposition 23.

First, we reduce the problem to a two-dimensional one. If  $m > 2$ , we may find a two-dimensional slice  $S$  of  $[N]^m$  of the form  $[N]^2 \times w$ , so that the density of  $X$  in  $S$  is greater than or equal to the density of  $X$  in  $[N]^m$ . That is,

$$(12) \quad |X \cap S| \geq N^2 \frac{|X|}{N^m} > 8cN^{1.5} + 4N.$$

We will treat  $S$  as a two-dimensional grid of  $m$ -dimensional cubes, with  $N$  rows and  $N$  columns, as in Figure 7.

Let  $R_1, \dots, R_N$  be the rows of  $[N]^2$  and  $C_1 \dots C_N$  the columns. Let  $1 \leq j \leq N$ . If  $|X \cap R_j| > 4c\sqrt{N} + 2$ , let  $A_j$  be the set of  $4c\sqrt{N} + 2$  elements of  $X \cap R_j$  that contains the leftmost  $2c\sqrt{N} + 1$  elements of  $X \cap R_j$  and the rightmost  $2c\sqrt{N} + 1$  elements of  $X \cap R_j$ . Otherwise, let  $A_j = X \cap R_j$ . Similarly, if  $|X \cap C_j| > 4c\sqrt{N} + 2$ , let  $B_j$

be the set of  $4c\sqrt{N} + 2$  elements of  $X \cap C_j$  that contains the uppermost  $2c\sqrt{N} + 1$  elements of  $X \cap C_j$  and the lowermost  $2c\sqrt{N} + 1$  elements of  $X \cap C_j$ . Otherwise, let  $B_j = X \cap C_j$ . Then, if  $D = \bigcup_{j=1}^N A_j \cup B_j$ ,

$$|D| \leq 2N \left( 4c\sqrt{N} + 2 \right) = 8cN^{1.5} + 4N < |X \cap S| ,$$

by Equation 12. Therefore,  $X \cap S \setminus D \neq \emptyset$  and there is a  $y \in X \cap S \setminus D$ . The row of  $S$  containing  $y$  contains at least  $2c\sqrt{N} + 1$  additional elements of  $X \cap S$  both to the left and to the right of  $y$ , and the column of  $S$  containing  $y$  has at least  $2c\sqrt{N} + 1$  additional elements of  $X \cap S$  both above and below  $y$ . Let  $q_1, q_2, q_3$  and  $q_4$  be elements of  $X$  to the right, above, below, and to the left of  $y$  so that there  $c\sqrt{N}$  other elements of  $X \cap S$  between them and  $y$ , and at least  $c\sqrt{N}$  additional elements between them and the boundary of  $S$ . Translate  $S$  so that  $y$  is located at  $(0, 0)$  (ignoring the higher-dimensional coordinates).  $q_1, q_2, q_3$ , and  $q_4$  form a quadrilateral, so there is a non-obtuse angle at at least one of the vertices. Without loss of generality, suppose the angle at  $q_1$  is non-obtuse. Let  $q_1 = (\hat{x}, 0)$ ,  $q_2 = (0, y_1)$ , and  $q_3 = (0, y_2)$ . By construction,  $\hat{x}$  and  $y_1$  are positive, and  $y_2$  is negative. Additionally, let  $p_1 = (x, 0)$  be an element of  $X \cap S$  in the same row as  $y$  and  $q_1$  that is at least  $c\sqrt{N}$  cubes further to the right of  $q_1$ . Then  $p_1, q_2$ , and  $q_3$  form an acute triangle. See Figure 7.

$p_1, q_2$ , and  $q_3$  form an acute triangle, so  $PH_1(\{p_1, q_2, q_3\})$  consists of a single interval. We will show that the length of this interval is greater than  $1 + \sqrt{m}$  for sufficiently large  $N$ , which will imply that the  $PH_1$  class is a stable class of size greater than one. The length of this interval is equal to the circumradius of the triangle minus half the length of its maximum edge. The circumradius of a triangle with edge lengths  $a, b, c$  and area  $A$  is  $abc/4A$  so the circumradius of  $p_1, q_2, q_3$  is

$$\frac{\sqrt{(x^2 + y_1^2)(x^2 + y_2^2)}(y_1 - y_2)}{4\frac{1}{2}(y_1 - y_2)x} = \frac{\sqrt{(x^2 + y_1^2)(x^2 + y_2^2)}}{2x}.$$

We consider three cases. In the first case, the edge from  $q_2$  to  $q_3$  is longest. In this case, the total persistence equals

$$TP_1(x, y_1, y_2) = \frac{\sqrt{(x^2 + y_1^2)(x^2 + y_2^2)}}{2x} - \frac{y_1 - y_2}{2}.$$

$q_1, q_2, q_3$  is an acute triangle, so  $\hat{x} > -y_1y_2$  and  $x > \sqrt{-y_1y_2} + c\sqrt{N}$ . In Appendix B.1, we show that, subject to the constraints that  $c\sqrt{N} \leq y_1 \leq N$ ,  $-N \leq y_2 \leq -c\sqrt{N}$ , and  $x > \sqrt{-y_1y_2} + c\sqrt{N}$  (which all hold in this case),

$$TP_1(x, y_1, y_2) \geq TP_1(c\sqrt{N} + N, N, -N) = \frac{c^2 N}{2(c\sqrt{N} + N)}$$

which is an increasing function of  $N$  whose limit is  $\frac{1}{2}c^2$  as  $N$  goes to  $\infty$ . As  $c > \sqrt{4\sqrt{m} + 4} > \sqrt{2\sqrt{m} + 2}$ , the total persistence will exceed  $1 + \sqrt{m}$  for sufficiently large  $N$ , as desired.

In the second case, the edge between  $p$  and  $q_1$  is longest. In this case, the total persistence equals

$$TP_2(x, y_1, y_2) = \frac{\sqrt{(x^2 + y_1^2)(x^2 + y_2^2)}}{2x} - \frac{\sqrt{x^2 + y_1^2}}{2}.$$

As we show in Appendix B.2, for sufficiently large  $N$ , this is minimized when  $x$  is as large as possible and the magnitudes of  $y_1$  and  $y_2$  are as small as possible. Thus, for sufficiently large  $N$ ,

$$TP_2(x, y_1, y_2) \geq TP_2(N, c\sqrt{N}, -c\sqrt{N}) = \frac{1}{2}(c^2 + N - \sqrt{c^2 N + N^2})$$

which is a decreasing function of  $N$  with limiting value  $\frac{1}{4}c^2$ . Because  $c > \sqrt{4\sqrt{m} + 4}$ , the total persistence is always greater than  $1 + \sqrt{m}$ , as desired. The argument in the third case is identical to this one.  $\square$

The previous result, together with Proposition 25, completes the proofs of Theorems 8 and 9.

## 5. RESULTS FOR THE RIPS COMPLEX

In this section, we first construct an example where for which  $\dim_{\widetilde{PH}_1}(X) = 2$  but  $\dim_{\text{box}}(X) = 1$ , then we show that this cannot occur in Euclidean space by proving Theorems 10 and 11.

**5.1. An Example.** Here, we prove Proposition 6 by constructing an example of a metric space not embeddable in any Euclidean space whose  $\widetilde{PH}_1$  dimension equals 2 but whose upper box dimension equals 1. This contrasts with the  $\widetilde{PH}_0$  dimension, which was proven by Kozma, Lotker, and Stupp [20] to agree with the upper box dimension for all metric spaces. The rough idea of the construction is to build a space that includes a copies of the complete bipartite graph on  $2^n$  vertices at different scales.

*Proof of Proposition 6.* We will construct  $X$  as the union of finite point sets  $X_n$ . For each  $n \in \mathbb{N}$  let  $X_n$  be the set consisting of  $2^{n+1}$  points  $x_1^n, x_2^n, \dots, x_{2^n}^n$  and  $y_1^n, y_2^n, \dots, y_{2^n}^n$ , and let

$$d(x_i^n, y_i^n) = \frac{1}{2^{n+1}} \quad d(x_i^n, x_j^n) = (1 - \delta_{i,j}) \frac{1}{2^n} \quad d(y_i^n, y_j^n) = (1 - \delta_{i,j}) \frac{1}{2^n}.$$

Let  $X = \cup_{i=0}^{\infty} X_i$ . if  $i \neq j$ ,  $z \in X_i$  and  $w \in X_j$ , set

$$(13) \quad d(z, w) = \frac{1}{2^{\min(i,j)}}.$$

We will show the upper box dimension of  $X$  equals one by computing  $N_\delta(X)$ , the maximum number of disjoint closed balls of radius  $\delta$  centered at points of  $X$ .

Suppose  $2^{-n-1} < \delta < 2^{-n}$  and that the balls centered at  $\mathbf{x}$  are a maximal collection of disjoint balls of radius  $\delta$  centered at points of  $X$ . Let  $Y_n = \cup_{j < n} X_j$  and  $Z_n = \cup_{j > n} X_j$ . If  $y \in Y_n$ ,  $B_\delta(y)$  contains every point of  $Y_n$ , but no point of  $X \setminus Y_n$ . Therefore,  $\mathbf{x}$  contains one point of  $Y_n$ . Also, for  $1 \leq i \leq 2^n$ ,  $B_\delta(y_i^n)$  contains  $y_i^n$  and the  $2^n$  points  $x_1^n, \dots, x_{2^n}^n$ . By symmetry,  $B_\delta(x_i^n)$  contains  $x_i^n$  and the  $2^n$  points  $y_1^n, \dots, y_{2^n}^n$ . Therefore, any two  $\delta$ -balls centered at points of  $X_n$  intersect and  $\mathbf{x}$  contains either one point  $y_i^n$  or one point  $x_i^n$ . Finally, if  $z \in Z_n$ ,  $B_\delta(z) \cap X = z$ . Therefore,  $\mathbf{x}$  must contain all of the points of  $Z_n$ . It follows that

$$\begin{aligned} N_\delta(X) &= |\mathbf{x}| \\ &= 1 + 2^n + |Z_n| \\ &= 1 + \sum_{j=0}^{n-1} |X_j| \\ &= 1 + \sum_{j=0}^{n-1} 2^{j+1} \\ &= 2^{n+1} - 1 \\ &= 2^{-\lceil \frac{\log(\delta)+1}{\log(2)} \rceil} - 1. \end{aligned}$$



Therefore

$$\begin{aligned}
 \dim_{\text{box}}(X) &= \limsup_{\delta \rightarrow 0} \frac{\log(N_\delta(X))}{-\log(\delta)} \\
 &= \lim_{\delta \rightarrow 0} \frac{\log\left(2^{-\lceil \frac{\log(\delta)}{\log(2)} \rceil + 1} - 1\right)}{-\log(\delta)} \\
 &= \lim_{\delta \rightarrow 0} \frac{-\lceil \frac{\log(\delta)}{\log(2)} \rceil \log(2)}{-\log(\delta)} \\
 &= 1.
 \end{aligned}$$

On the other hand, consider the Rips complex  $\mathcal{R}_\epsilon(X_n)$  on  $X_n$ . There are three regimes. If  $\epsilon < \frac{1}{2^{n+1}}$ , then  $\mathcal{R}_\epsilon(X_n)$  consists of  $2^{n+1}$  0-cells. When  $\epsilon = \frac{1}{2^{n+1}}$ , the  $2^{2n}$  edges between the points  $\{x_i^n\}_{i=1}^{2^n}$  and  $\{y_i^n\}_{i=1}^{2^n}$  enter the complex. The complex remains unchanged until  $\epsilon = \frac{1}{2^n}$ , when all possible simplices enter the complex and it becomes contractible. As such, all intervals of  $\widetilde{PH}_1(X_n)$  are born at  $\epsilon = \frac{1}{2^{n+1}}$  and die at  $\epsilon = \frac{1}{2^n}$ . By an Euler characteristic argument, there are  $2^{2n} - 2^{n+1} + 1$  such intervals. Therefore, for  $\alpha > 0$

$$\begin{aligned}
 \widetilde{E}_\alpha^1(X_n) &= \left(2^{2n} - 2^{n+1} + 1\right) \left(\frac{1}{2^n} - \frac{1}{2^{n+1}}\right)^\alpha \\
 &= \left(2^{2n} - 2^{n+1} + 1\right) \frac{1}{2^{\alpha n + \alpha}} \\
 &= 2^{-\alpha} \left(2^{(2-\alpha)n} - 2^{(1-\alpha)n+1} + 2^{-\alpha n}\right) \\
 &\approx 2^{-\alpha} 2^{(2-\alpha)n},
 \end{aligned}$$

which is which is unbounded as  $n \rightarrow \infty$  if  $\alpha < 2$ . Thus  $\dim_{\widetilde{PH}_1}(X) \geq 2$ .

The argument that  $\dim_{\widetilde{PH}_1}(X) \leq 2$  is similar but slightly trickier. Let  $\alpha > 2$ ,  $\mathbf{x} \subseteq X$ , and  $\mathbf{x}_n = \mathbf{x} \cap X_n$ . An argument nearly identical to the previous one shows that  $\widetilde{PH}_1(\mathbf{x}_n)$  consists of at most  $2^{2n} - 2^{n+1} + 1$  intervals of the form  $\left(\frac{1}{2^{n+1}}, \frac{1}{2^n}\right)$ , and

$$(14) \quad \widetilde{E}_\alpha^1(\mathbf{x}_n) < 2^{-\alpha} \left(2^{(2-\alpha)n} + 2^{-\alpha n}\right).$$

We will show that we can write

$$\widetilde{E}_\alpha^1(\mathbf{x}) = \sum_{n=0}^{\infty} \widetilde{E}_\alpha^1(\mathbf{x}_n).$$

Let  $n \in \mathbb{N}$  and define

$$K_{i,n}(Y) = \left\{ (b, d) \in \widetilde{PH}_i(Y) : \frac{1}{2^{n+1}} \leq b < \frac{1}{2^n} \right\}.$$

Also, let

$$\frac{1}{2^{n+1}} \leq \epsilon < \frac{1}{2^n}, \quad \mathbf{y}_n = \cup_{j < n} \mathbf{x}_j, \quad \text{and} \quad \mathbf{z}_n = \cup_{j > n} \mathbf{x}_j.$$

By equation 13,  $\mathcal{R}_\epsilon(\mathbf{x})$  contains no edges between points of  $\mathbf{y}_n$  and points of  $\mathbf{x}_n$ . Furthermore, no point of  $\mathbf{z}_n$  is contained in an edge. Therefore,

$$\mathcal{R}_\epsilon(\mathbf{x}) = \mathcal{R}_\epsilon(\mathbf{x}_n) \sqcup \mathcal{R}_\epsilon(\mathbf{y}_n) \sqcup \mathcal{R}_\epsilon(\mathbf{z}_n),$$

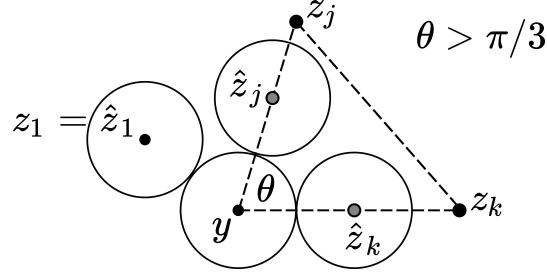
where  $\sqcup$  denotes a disjoint union.  $\mathcal{R}_\epsilon(\mathbf{y}_n)$  is contractible because it contains all possible simplices, so  $H_1(\mathcal{R}_\epsilon(\mathbf{y}_n)) = 0$ .  $H_1(\mathcal{R}_\epsilon(\mathbf{z}_n))$  is also trivial because  $\mathcal{R}_\epsilon(\mathbf{z}_n)$  contains only zero cells. Therefore,  $H_1(\mathcal{R}_\epsilon(\mathbf{x})) = H_1(\mathcal{R}_\epsilon(\mathbf{x}_n))$ . Also, if  $\delta > \frac{1}{2^n}$  then  $\mathcal{R}_\delta(\mathbf{x}_n)$  is contractible so the inclusion map  $H_1(\mathcal{R}_\epsilon(\mathbf{x})) \rightarrow H_1(\mathcal{R}_\delta(\mathbf{x}))$  is trivial. It follows that

$$(15) \quad K_{1,n}(\mathbf{x}) = K_{1,n}(\mathbf{x}_n) = \widehat{PH}_1(\mathbf{x}_n).$$

Let  $\alpha > 2$  and compute

$$\begin{aligned} \widetilde{E}_\alpha^1(\mathbf{x}) &= \sum_{n \in \mathbb{N}} \sum_{I \in K_{1,n}(\mathbf{x})} |I|^\alpha \\ &= \sum_{n \in \mathbb{N}} \sum_{I \in \widehat{PH}_1(\mathbf{x}_n)} |I|^\alpha && \text{by Eqn 15} \\ &= \sum_{n \in \mathbb{N}} \widetilde{E}_\alpha^1(\mathbf{x}_n) \\ &\leq \sum_{n \in \mathbb{N}} 2^{-\alpha} \left( 2^{(2-\alpha)n} + 2^{-\alpha n} \right) && \text{by Eqn 14} \\ &= 2^{-\alpha} \left( \frac{1}{1 - 2^{-(2-\alpha)}} + \frac{1}{1 + 2^{-\alpha}} \right) && \text{because } \alpha > 2. \end{aligned}$$

Therefore  $\widetilde{E}_\alpha^1(\mathbf{x})$  is uniformly bounded for all  $\mathbf{x} \subset X$  when  $\alpha > 2$ , and  $\dim_{\widehat{PH}_1}(X) = 2$ .  $\square$



**Figure 8.** The setup in the proof of Lemma 28, where  $z_1$  is the closest point of  $\mathbf{x} \setminus y$  to  $y$ .

5.2. **A Bound for  $\dim_{\widetilde{PH}_1}$ .** We prove an upper bound for the  $\widetilde{PH}_1$ -dimension of a subset of  $\mathbb{R}^m$  that does not depend on the ambient Euclidean dimension. Note that the proof of Proposition 18 (and the preceding lemmas) works for persistent homology defined in terms of either the Čech complex or the Rips complex.

The following argument is a modified version of the proof of Theorem 3.1 in Goff [16], which bounds the (non-persistent) first Betti number of a Rips complex of a finite point set in  $\mathbb{R}^m$  in terms of the kissing number  $K_m$  of  $\mathbb{R}^m$ . Recall that  $K_m$  is the maximum number of disjoint unit spheres that can be tangent to a shared unit sphere. For example  $K_1 = 2$ ,  $K_2 = 6$ , and  $K_3 = 12$ . There is an expansive literature on upper and lower bounds for  $K_m$  [24], but here we will just need that  $K_m$  is finite.

Let  $\mathcal{R}_\epsilon(X)$  be the Rips complex of  $X$  at parameter  $\epsilon$ . For a 0-cell  $x$  let  $S_\epsilon(x)$  denote the star of  $x$

$$S_\epsilon(x) = \{\sigma \in \mathcal{R}_\epsilon(X) : x \in \sigma\},$$

and  $L_{\epsilon(x)}$  denote the link of  $x$

$$L_\epsilon(x) = \{(z_1, \dots, z_k) \in \mathcal{R}_\epsilon(X) : (x, z_1, \dots, z_k) \in \mathcal{R}_\epsilon(X)\}.$$

$S_{\epsilon(x)}$  is contractible.

**Lemma 28.** *Let  $\mathbf{x}$  be a finite subset of  $\mathbb{R}^m$ , and  $y \in \mathbf{x}$ , and let  $\{L_\epsilon(y)\}_{\epsilon \in \mathbb{R}^+}$  be the filtration of the links of  $y$  in  $\mathcal{R}_\epsilon(\mathbf{x})$ . Then*

$$\left| PH_0 \left( \{L_\epsilon(y)\}_{\epsilon \in \mathbb{R}^+} \right) \right| \leq K_m - 1.$$

*Proof.* If necessary, perturb the points of  $\mathbf{x}$  so all pairwise distances are distinct without reducing the number of  $PH_0$  bars (by Corollary 13, this can be done by perturbing the points by a distance less than  $\delta/2$ , where  $\delta$  is the length of the shortest  $PH_0$  bar). Let the  $PH_0$  intervals of the filtration be  $\{(b_j, d_j)\}_{j=1}^k$ . For each  $j$ , the

number of components of  $L_\epsilon(y)$  increases at  $\epsilon = b_j$ . The only way a new component can be added to the link of  $y$  at  $\epsilon = b_j$  is if there is a point  $z_j$  so that  $b_j = d(z_j, y)$ . Let  $\{z_j\}_{j=1}^k$  be the collection of these points.

Let  $\epsilon = d(y, \mathbf{x} \setminus y) / 2$ . For each  $z_j$ , let  $\hat{z}_j$  be the point on the line segment from  $y$  to  $z_j$  with  $d(y, \hat{z}_j) = 2\epsilon$ . By construction, the balls of radius  $\epsilon$  centered at the points  $\hat{z}_j$  are tangent to the ball of radius  $\epsilon$  centered at  $y$ . We will show that they are disjoint by performing another computation related to the non-hatted points  $\{z_j\}$ .

Let  $z_j$  and  $z_k$  be distinct and assume without loss of generality that  $d(z_k, y) > d(z_j, y)$ . See Figure 8. Then  $z_j$  and  $z_k$  are contained in  $L_{d(z_k, y)}(y)$ , but they must be in distinct components (if not, the number of components of the link would not increase at  $\epsilon = d(z_k, y)$ ). In particular,

$$d(z_j, z_k) > d(z_k, y) > d(z_j, y) .$$

Consider the triangle  $T_0$  formed by  $z_j, z_k$ , and  $y$ . The previous equation shows the edge between  $z_j$  and  $z_k$  is the longest edge of  $T_0$ , so the angle opposite to it must be greater than  $\pi/3$ . In other words, the angle between the vectors  $z_j - y$  and  $z_k - y$  is greater than  $\pi/3$ . By construction, the angle between the vectors  $\hat{z}_j - y$  and  $\hat{z}_k - y$  equals the angle between the vectors  $z_j - y$  and  $z_k - y$ . Furthermore, these vectors have the same length,  $2\epsilon$ . Therefore, the triangle formed by  $\hat{z}_j, \hat{z}_k$  and  $y$  is an isosceles triangle so that the angle opposite the edge  $(\hat{z}_j, \hat{z}_k)$  is greater than  $\pi/3$ . It follows that this is the longest edge of the triangle and  $d(\hat{z}_j, \hat{z}_k) > 2\epsilon$ .

Therefore, the balls  $\{B_\epsilon(\hat{z}_j)\}$  are disjoint and tangent to  $B_\epsilon(y)$ , so there are at most  $K_m$  of them. Thus

$$\left| PH_0 \left( \{L_\epsilon(y)\}_{\epsilon \in \mathbb{R}^+} \right) \right| = |\{z_j\}| - 1 = \left| \{B_\epsilon(\hat{z}_j)\} \right| \leq K_m - 1 ,$$

where the “ $-1$ ” comes from the fact that are using reduced homology.  $\square$

**Lemma 29.** *Let  $\mathbf{x}$  be a finite metric space so that all pairwise distances are distinct, and let  $\epsilon_1 = 0 < \epsilon_2 < \dots < \epsilon_p$  be the values of  $\epsilon$  at which a simplex is added to the Rips complex of  $\mathbf{x}$ . Then*

$$\left| \widetilde{PH}_1(\mathbf{x}) \right| = \left| \left\{ j : \dim H_1 \left( \mathcal{R}_{\epsilon_j}(\mathbf{x}) \right) = \dim H_1 \left( \mathcal{R}_{\epsilon_{j-1}}(\mathbf{x}) \right) + 1 \right\} \right| .$$

*Proof.* We will show that there is bijection between the birth times  $PH_1$  intervals and filtration values  $\epsilon_j$  so that the rank of the first homology increases by one. If

$$\dim H_1 \left( \mathcal{R}_{\epsilon_j}(\mathbf{x}) \right) = \dim H_1 \left( \mathcal{R}_{\epsilon_{j-1}}(\mathbf{x}) \right) + 1$$

then there is persistent homology interval born at  $\epsilon_j$ , with death time at least  $\epsilon_{j+1}$ .

<sup>2</sup> Therefore

$$\left| \widetilde{PH}_1(\mathbf{x}) \right| \geq \left| \left\{ j : \dim H_1 \left( \mathcal{R}_{\epsilon_j}(\mathbf{x}) \right) = \dim H_1 \left( \mathcal{R}_{\epsilon_{j-1}}(\mathbf{x}) \right) + 1 \right\} \right|.$$

Suppose a  $PH_1$  interval is born at  $\epsilon_j$ . Note that, by the definition of the Rips complex, the  $\epsilon_k$ 's are precisely the pairwise distances between the points of  $\mathbf{x}$ . It follows that exactly one edge enters the complex at  $\epsilon_j$ : the unique edge  $(z_j, y_j)$  so that  $d(z_j, y_j) = \epsilon_j$ . The Euler characteristic of the 1-skeleton of the Rips complex increases exactly by one, so

$$\dim H_1 \left( \mathcal{R}_{\epsilon_j}(\mathbf{x}) \right) \leq \dim H_1 \left( \mathcal{R}_{\epsilon_{j-1}}(\mathbf{x}) \right) + 1.$$

A  $PH_1$  interval is born at  $\epsilon_j$  so there is a cycle  $\sigma \in H_1 \left( \mathcal{R}_{\epsilon_j}(\mathbf{x}) \right)$  that is not homologous to a cycle in  $H_1 \left( \mathcal{R}_{\epsilon_{j-1}}(\mathbf{x}) \right)$ .  $\sigma$  must contain the edge  $(y_j, z_j)$ . We will show that the contribution of  $\sigma$  to the rank of  $H_1 \left( \mathcal{R}_{\epsilon_j}(\mathbf{x}) \right)$  cannot be canceled by the simultaneous death of another  $PH_1$  interval. By way of contradiction, assume that another  $PH_1$  interval has death time equal to  $\epsilon_j$ , so a 2-simplex also enters the Rips complex at  $\epsilon_j$ . Then it must be of the form  $(y_j, z_j, w_j)$  for some  $w_j \in \mathbf{x}$ . Then  $\sigma$  is homologous to the cycle obtained by replacing the edge  $(y_j, z_j)$  with the edges  $(y_j, z_j)$  and  $(z_j, w_j)$ . This is a contradiction, so every cycle in  $H_1 \left( \mathcal{R}_{\epsilon_{j-1}}(\mathbf{x}) \right)$  is a cycle in  $H_1 \left( \mathcal{R}_{\epsilon_j}(\mathbf{x}) \right)$  and

$$\dim H_1 \left( \mathcal{R}_{\epsilon_j}(\mathbf{x}) \right) = \dim H_1 \left( \mathcal{R}_{\epsilon_{j-1}}(\mathbf{x}) \right) + 1.$$

□

*Proof of Theorem 11.* We will proceed by induction on the size of  $\mathbf{x}$ . Let  $\delta$  be the length of the shortest interval of  $\widetilde{PH}_1(\mathbf{x})$ . Perturb the points of  $\mathbf{x}$  by an amount less than  $\delta/2$  so that the lengths of all of the pairwise distances between points of  $\mathbf{x}$  are

---

<sup>2</sup>Note that a similar statement is false for general Rips complexes on infinite point sets; there can be “ephemeral” homology classes that are born and die at the same value of  $\epsilon$  and do not correspond to any persistent homology interval.

distinct. The stability of the bottleneck distance implies that no bars of  $\widetilde{PH}_1(\mathbf{x})$  are destroyed by this perturbation, so an upper bound for the perturbed set implies one for the original.

For  $\epsilon \geq 0$  and  $y \in \mathbf{x}$ , we have that

$$\mathcal{R}_\epsilon(\mathbf{x}) = \mathcal{R}_\epsilon(\mathbf{x} \setminus y) \cup S_\epsilon(y)$$

and

$$\mathcal{R}_\epsilon(\mathbf{x} \setminus y) \cap S_\epsilon(y) = L_\epsilon(y).$$

As such, there is a Mayer–Vietoris sequence.

$$\dots \rightarrow H_1(L_\epsilon(y)) \rightarrow H_1(\mathcal{R}_\epsilon(\mathbf{x} \setminus y)) \oplus H_1(S_\epsilon(y)) \rightarrow H_1(\mathcal{R}_\epsilon(\mathbf{x})) \rightarrow H_0(L_\epsilon(y)) \rightarrow \dots$$

Let  $\epsilon_1 = 0 < \epsilon_2 < \dots < \epsilon_p$  be the values of  $\epsilon$  at which a simplex is added to the Rips complex of  $\mathbf{x}$ . The Mayer–Vietoris sequences for each  $\epsilon_j$  fit into the following commutative diagram, where vertical maps are inclusions and we have suppressed  $H_1(S_\epsilon(x)) = 0$  in the left column.

$$\begin{array}{ccccc} H_1(\mathcal{R}_{\epsilon_{i-1}}(\mathbf{x} \setminus y)) & \xrightarrow{\alpha_{i-1}} & H_1(\mathcal{R}_{\epsilon_{i-1}}(\mathbf{x})) & \xrightarrow{\partial_{i-1}} & H_0(L_{\epsilon_{i-1}}(y)) \\ \downarrow & & \downarrow & & \downarrow \xi_i \\ H_1(\mathcal{R}_{\epsilon_i}(\mathbf{x} \setminus y)) & \xrightarrow{\alpha_i} & H_1(\mathcal{R}_{\epsilon_i}(\mathbf{x})) & \xrightarrow{\partial_i} & H_0(L_{\epsilon_i}(y)) \\ \downarrow & & \downarrow & & \downarrow \xi_{i+1} \\ H_1(\mathcal{R}_{\epsilon_{i+1}}(\mathbf{x} \setminus y)) & \xrightarrow{\alpha_{i+1}} & H_1(\mathcal{R}_{\epsilon_{i+1}}(\mathbf{x})) & \xrightarrow{\partial_{i+1}} & H_0(L_{\epsilon_{i+1}}(y)) \end{array}$$

Recall from the proof of the previous lemma that  $H_1(\mathcal{R}_{\epsilon_i}(\mathbf{x}))$  and  $H_1(\mathcal{R}_{\epsilon_i}(\mathbf{x} \setminus y))$  can increase at most by one from row to row, and that the number of times the rank increases is equal to the number of  $PH_1$  intervals of the column.

Homology is taken with field coefficients so

$$H_1(\mathcal{R}_{\epsilon_i}(\mathbf{x})) \cong \text{im } \alpha_i \oplus \text{im } \partial_i.$$

Therefore, if

$$\dim H_1(\mathcal{R}_{\epsilon_i}(\mathbf{x})) = \dim H_1(\mathcal{R}_{\epsilon_{i-1}}(\mathbf{x})) + 1$$

the dimension of either  $\text{im } \alpha_i$  or  $\text{im } \partial_i$  must also have increased. In the former case, commutativity of the diagram (combined with the previous lemma) implies that

$$\dim H_1 \left( \mathcal{R}_{\epsilon_i} (\mathbf{x} \setminus y) \right) = \dim H_1 \left( \mathcal{R}_{\epsilon_{i-1}} (\mathbf{x} \setminus y) \right) + 1,$$

so an interval of  $\widetilde{PH}_1 (\mathbf{x} \setminus y)$  must also be born at time  $\epsilon_i$ . In latter case, commutativity of the diagram implies that there a  $\sigma_i \in H_0 (L_{\epsilon_i} (y))$  so that

$$(16) \quad \sigma_i \in \text{im } \partial_i \text{ and } \sigma_i \notin \text{im } \xi_i \circ \xi_{i-1} \dots \circ \xi_{k+1} \circ \partial_k \quad \forall 1 \leq k < i.$$

Let

$$\hat{i} = \max \{ j : 1 \leq j \leq i \text{ and } \sigma_i \notin \text{im } \xi_i \circ \xi_{i-1} \dots \circ \xi_j \}.$$

(Note that we are using reduced homology, so the row of diagram with  $i = 1$  is all zero and  $\xi_1 = 0$ ). We may assume that  $\sigma_i$  was chosen to give the minimal possible value of  $\hat{i}$ , subject to Equation 16. By construction, there is an element of  $H_1 (L_{\epsilon_i})$  that is not in the image of  $\xi_i$  (the one that maps to  $\sigma_i$ ), so there is a  $PH_0$  interval  $I_i$  of the link filtration of  $y$  with birth time equal to  $\epsilon_i$ . This interval is unique, by the same reasoning as in Lemma 28. We will show that  $i \mapsto I_i$  gives an injection

$$\{ i : \dim \text{im } \partial_i > \dim \text{im } \partial_{i-1} \} \rightarrow \widetilde{PH}_0 \left( \{ L_\epsilon (x) \} \right).$$

By way of contradiction, suppose that that

$$\epsilon_j, \epsilon_l \in \{ i : \dim \text{im } \partial_i > \dim \text{im } \partial_{i-1} \}$$

and  $j < l$  but  $I_j = I_l$  (so  $\hat{j} = \hat{l}$ ). Choose  $\hat{\sigma}_j, \hat{\sigma}_l \in H_0 (L_{\epsilon_i} (y))$  so that

$$\xi_j \circ \dots \circ \xi_{\hat{i}} (\hat{\sigma}_j) = \sigma_j \text{ and } \xi_l \circ \dots \circ \xi_{\hat{i}} (\hat{\sigma}_l) = \sigma_l.$$

Let  $z$  be the unique point of  $\mathbf{x}$  so that  $z \in L_{\epsilon_i} (y)$  but  $z \notin L_{\epsilon_{i-1}} (y)$ . By the minimality of  $\hat{j}$  and  $\hat{l}$ , there are  $a_j, a_l$  in the coefficient field so that  $a_j, a_l \neq 0$  and

$$\hat{\sigma}_j - a_j z, \hat{\sigma}_l - a_l z \in \text{im } \xi_i.$$

$j < l$  and  $\frac{a_l}{a_j} \sigma_j \in \text{im } \partial_j$ , so it follows from commutativity of the diagram that  $\sigma_l - \frac{a_l}{a_j} \sigma_j$  must also satisfy Equation 16 for  $i = l$ , but

$$\sigma_l - \frac{a_l}{a_j} \sigma_j = \xi_l \circ \dots \circ \xi_{\hat{i}} \left( \hat{\sigma}_l - \frac{a_l}{a_j} \hat{\sigma}_j \right)$$

and  $\hat{\sigma}_l - \frac{a_l}{a_j} \hat{\sigma}_j \in \text{im } \xi_i$ , which contradicts the minimality of  $\hat{l}$ . Therefore

$$\left| \left\{ i : \dim \text{im } \partial_i > \dim (\text{im } \partial_{i-1}) \right\} \right| \leq \left| \widetilde{PH}_0 \left( \{ L_\epsilon (x) \} \right) \right|.$$

Thus, by applying the previous lemma,

$$\begin{aligned}
 \left| \widetilde{PH}_1(\mathbf{x}) \right| &= \left| \left\{ i : \dim H_1(\mathcal{R}_{\epsilon_{i+1}}(\mathbf{x})) = \dim H_1(\mathcal{R}_{\epsilon_i}(\mathbf{x})) + 1 \right\} \right| \\
 &= \left| \{i : \dim \text{im } \partial_i > \dim \text{im } \partial_{i-1}\} \right| + \left| \{i : \dim \text{im } \alpha_i > \dim \alpha_{i-1}\} \right| \\
 &\leq \left| \widetilde{PH}_1(\mathbf{x} \setminus y) \right| + \left| \widetilde{PH}_0(\{L_\epsilon(x)\}_{\epsilon \in \mathbb{R}^+}) \right| \\
 &\leq \left| \widetilde{PH}_1(\mathbf{x} \setminus y) \right| + K_m - 1 && \text{by Lemma 28} \\
 &\leq (K_m - 1) |x| && \text{by induction.}
 \end{aligned}$$

□

This, together with Proposition 18, implies Theorem 10.

## 6. CONCLUSION

We have taken the first steps toward answering Question 5, which asks for hypotheses under which the  $PH_i$ -dimension of a bounded subset of  $\mathbb{R}^m$  equals its upper box dimension. However, the question remains open for all cases with  $m > 2$  and  $i > 0$ . We suspect that even the  $n = 2, i = 1$  case can be improved, to include sets whose upper box dimension is between 1 and 1.5.

Another interesting question is whether similar results can be shown for a probabilistic version of the  $PH_i$ -dimension. That is, if  $\mathbf{x}$  is a finite point collection drawn from a probability measure on  $\mathbb{R}^m$ , can the expectation of  $E_i(\mathbf{x})$  be controlled in terms a classically defined fractal dimension for probability measures? This question would perhaps be more interesting for applications, which usually deal with random point collections rather than extremal ones. However, proving a lower bound is already difficult in the extremal case.

## 7. ACKNOWLEDGMENTS

The author would like to thank Henry Adams, Tamal Dey, Herbert Edelsbrunner, Matthew Kahle, Chris Peterson, and Shmuel Weinberger for useful discussions. In particular, he would like to thank Herbert Edelsbrunner for suggesting the example in Section 3.1 and the reference [9]. He would also like to thank the members of the Pattern Analysis Laboratory at Colorado State University, whose computational experiments involving a similar dimension inspired the ideas in this paper [1].





**Figure 9.** The Cantor set cross an interval. If  $C$  is the cantor set and  $I$  is an interval,  $\dim_{PH}^0(C \times I) = \dim_{PH}^0(C) = \log_3 2$  and  $\dim_{PH}^1(C \times I) = 0$ , but  $\dim_{PH}^0(C \times I) = \dim_{PH}^1(C \times I) = \dim_{\text{box}}(C \times I) = 1 + \log_3 2$ .

Funding for this research was provided by a NSF Mathematical Sciences Postdoctoral Research Fellowship under award number DMS-1606259.

#### APPENDIX A. PROPERTIES OF $PH$ COMPLEXITY

In previous work with MacPherson [21], we defined another notion of persistent homology dimension for a compact subset of a metric space based on the persistent homology of the subset itself rather than of the finite point sets contained within it. Here, we denote it by  $\dim_{PH}^i(X)$ .  $\dim_{PH}^i(X)$  measures the complexity of a shape rather than any classically defined fractal dimension, so we refer to it as the “ $PH_i$  complexity” of  $X$ . See Figure 9.

**Definition 30.** *The  $i$ -th dimensional persistent homology complexity of compact subset of a metric space is*

$$\dim_{PH}^i(X) = \sup \left\{ c : \lim_{x \rightarrow 0} x^c I_{i,\epsilon}(X) = \infty \right\}.$$

We show that there is an equivalent definition that looks more similar to that of  $\dim_{PH}^i$ .

**Proposition 31.** *Let  $X$  be a compact subset of a metric space. Then*

$$\dim_{PH}^i(X) = \inf \left\{ \alpha : E_\alpha^i(X) < \infty \right\}.$$

The proposition is an immediate consequence of the following lemma.

**Lemma 32.** *Suppose  $Y \subset \mathbb{R}^+$  has the property that  $\{y \in Y : y > \epsilon\}$  is finite for all  $\epsilon > 0$ . If*

$$F(\epsilon) = |\{y \in Y : y > \epsilon\}|$$

then

$$\inf \left\{ \alpha : \sum_{y \in Y} y^\alpha < \infty \right\} = \sup \left\{ c : \lim_{\epsilon \rightarrow 0} \epsilon^c F(\epsilon) = \infty \right\}.$$

*Proof.* Let  $d_1 = \inf \left\{ \alpha : \sum_{y \in Y} y^\alpha < \infty \right\}$  and  $d_2 = \sup \left\{ c : \lim_{\epsilon \rightarrow 0} \epsilon^c F_Y(\epsilon) = \infty \right\}$ .

Let  $\alpha > \beta > d_2$ ,  $I_k = \left\{ y \in Y : 2^{-(k+1)} < y \leq 2^{-k} \right\}$ , and  $C_0 = \sum_{\{y \in Y : y > 1\}} \epsilon^\alpha$ . By definition, there is a constant  $C_1 > 0$  so that  $F(\epsilon) < C_1 \epsilon^{-\beta}$  for all  $\epsilon \in \mathbb{R}$ . Then

$$\begin{aligned} \sum_{y \in Y} y^\alpha &\leq \sum_{k=0}^{\infty} |I_k| 2^{-\alpha k} + C_0 \\ &< \sum_{k=0}^{\infty} F\left(2^{-(k+1)}\right) 2^{-\alpha k} + C_0 \\ &< \sum_{k=0}^{\infty} C_1 2^{\beta(k+1)} 2^{-\alpha k} + C_0 \\ &= C_1 2^\beta \sum_{k=1}^{\infty} 2^{(\beta-\alpha)k} + C_0 \\ &< \infty, \end{aligned}$$

because  $\beta - \alpha < 0$ . Therefore,  $d_1 \leq d_2$ .

Let  $\alpha < d_2$  and  $D > 0$ .  $\lim_{\epsilon \rightarrow 0} \epsilon^\alpha F(\epsilon) = \infty$  so there is an  $\epsilon_0 > 0$  so that  $\epsilon_0^\alpha F_Y(\epsilon_0) > D$ . Then

$$\sum_{y \in Y} y^\alpha \geq \sum_{y > \epsilon_0} y^\alpha > \epsilon_0^\alpha F_Y(\epsilon_0) > D,$$

so  $\sum_{y \in Y} y^\alpha$  is greater than any positive real number and  $\sum_{y \in Y} y^\alpha = \infty$ . Therefore  $d_1 \geq d_2$ , and  $d_1 = d_2$ .  $\square$

We also show that  $\dim_{\widehat{PH}}^i$  is bounded above by  $\dim_{PH}^i$  :

**Proposition 33.** *Let  $X$  be a compact subset of a metric space. Then*

$$\dim_{\widehat{PH}}^i(X) \leq \dim_{PH}^i(X).$$

*Proof.* Let  $\alpha < \dim_{PH}^i(X)$  and  $D > 0$ . There is a finite collection of intervals  $\{b_j, d_j\}_{j=1}^n \subset PH_i(X)$  with

$$(17) \quad \sum_{j=1}^n (d_j - b_j)^\alpha > D + 1.$$

The function  $f(x) = x^\alpha$  is uniformly continuous on  $[\min_j (d_j - b_j), \max_j (d_j - b_j)]$  so there is a  $\epsilon > 0$  so that

$$(18) \quad |x^\alpha - (x + \delta)^\alpha| < \frac{1}{n}$$

for all  $x \in [\min_j (d_j - b_j), \max_j (d_j - b_j)]$  and  $|\delta| < \epsilon$ . Let  $\mathbf{x} \subset X$  be a finite point satisfying  $d_H(\mathbf{x}, X) < \epsilon$ . By bottleneck stability, the intervals  $\{(b_j, d_j)\}$  can be paired with intervals  $\{(\hat{b}_j, \hat{d}_j)\} \in PH_i(\mathbf{x})$  so that

$$(19) \quad \left| (d_j - b_j) - (\hat{d}_j - \hat{b}_j) \right| < \epsilon,$$

for  $j = 1, \dots, n$ . (Note that  $\mathbf{x}$  might have other, smaller intervals, but it doesn't matter for our computations.) Then

$$\begin{aligned} E_\alpha^i(\mathbf{x}) &= \sum_{(\hat{b}, \hat{d}) \in PH_i(\mathbf{x})} (\hat{d} - \hat{b})^\alpha \\ &\geq \sum_{j=1}^n (\hat{d}_j - \hat{b}_j)^\alpha \\ &\geq \sum_{j=1}^n \left( (d_j - b_j)^\alpha - \frac{1}{n} \right) && \text{using Eqns. 18 and 19} \\ &> D && \text{by Eqn. 17.} \end{aligned}$$

$D$  was arbitrary so  $E_\alpha^i(\mathbf{x})$  can be made arbitrarily large and  $\dim_{PH}^i(X) \geq \alpha$  for all  $\alpha < \dim_{PH}^i(X)$  and  $\dim_{PH}^i(X) \geq \dim_{PH}^i(X)$ , as desired.  $\square$

## APPENDIX B. TRIANGLE COMPUTATIONS

Here, we complete the computations in the proof of Proposition 27.

**B.1. Minimizing  $TP_1$ .** We will find the minimum of the function

$$TP_1(x, y_1, y_2) = \frac{\sqrt{(x^2 + y_1^2)(x^2 + y_2^2)}}{2x} - \frac{y_1 - y_2}{2}$$

subject to the constraints  $c\sqrt{N} \leq y_1 \leq N$ ,  $-N \leq y_2 \leq -c\sqrt{N}$ , and  $c\sqrt{N} + \sqrt{-y_1 y_2} \leq x \leq N$ . The partial derivative with respect to  $x$  is

$$\frac{\partial TP_1(x, y_1, y_2)}{\partial x} = \frac{2x^2 + y_1^2 + y_2^2}{2\sqrt{(x^2 + y_1^2)(x^2 + y_2^2)}} - \frac{\sqrt{(x^2 + y_1^2)(x^2 + y_2^2)}}{2x^2} = \frac{x^4 - y_1^2 y_2^2}{2x^2 \sqrt{(x^2 + y_1^2)(x^2 + y_2^2)}},$$

which is positive when  $x > \sqrt{-y_1 y_2}$ . Therefore, the minimum value of  $TP_1$  is achieved when  $x$  is as small as possible,  $x = c\sqrt{N} + \sqrt{-y_1 y_2}$ . Also,

$$\frac{\partial TP_1(x, y_1, y_2)}{\partial y_1} = \frac{y_1(x^2 + y_2^2)}{2x\sqrt{(x^2 + y_1^2)(x^2 + y_2^2)}} - \frac{1}{2} = \frac{-x\sqrt{(x^2 + y_1^2)(x^2 + y_2^2)} + x^2 y_1 + y_1 y_2^2}{2x\sqrt{(x^2 + y_1^2)(x^2 + y_2^2)}},$$

and (using that  $x, y_1$ , and  $-y_2$  are positive)

$$\begin{aligned} \frac{\partial TP_1(x, y_1, y_2)}{\partial y_1} = 0 &\iff -x\sqrt{(x^2 + y_1^2)(x^2 + y_2^2)} + x^2 y_1 + y_1 y_2^2 = 0 \\ &\iff (x^2 y_1 + y_1 y_2^2)^2 - x^2 (x^2 + y_1^2)(x^2 + y_2^2) = 0 \\ &\iff (x^2 + y_2^2)(x^4 - y_1^2 y_2^2) = 0 \\ &\iff x = \pm \sqrt{-y_1 y_2} \end{aligned}$$

which is also not allowed by our constraints. We can see (for example, by plugging in  $x = 2y_1, y_2 = -y_1$  above) that this partial derivative is always negative when our constraints hold, and the minimum is achieved when  $y_1$  is as large as possible,  $y_1 = N$ .

A symmetric computation shows that  $y_2 = -N$  at the minimum. Therefore,

$$\begin{aligned} TP_1(x, y_1, y_2) &\geq TP_1(c\sqrt{N} + N, N, -N) \\ &= \frac{\sqrt{\left(\left(c\sqrt{N} + N\right)^2 + N^2\right)^2}}{2\left(c\sqrt{N} + N\right)} - N \\ &= \frac{c^2N}{2\left(c\sqrt{N} + N\right)}. \end{aligned}$$

**B.2. Minimizing  $TP_2$ .** We will find the minimum of the function

$$TP_2(x, y_1, y_2) = \frac{\sqrt{(x^2 + y_1^2)(x^2 + y_2^2)}}{2x} - \frac{\sqrt{x^2 + y_1^2}}{2}$$

subject to the constraints  $c\sqrt{N} \leq y_1 \leq N$ ,  $-N \leq y_2 \leq -c\sqrt{N}$ , and  $c\sqrt{N} \leq x \leq N$ . The partial derivative of  $TP_2$  with respect to  $y_1$  is

$$\frac{\partial TP_2(x, y_1, y_2)}{\partial y_1} = \frac{y_1(x^2 + y_2^2)}{2x\sqrt{(x^2 + y_1^2)(x^2 + y_2^2)}} - \frac{y_1}{2\sqrt{x^2 + y_1^2}} = \frac{y_1(\sqrt{x^2 + y_2^2} - x)}{2x\sqrt{x^2 + y_1^2}},$$

which is always positive. Therefore the minimum will occur when  $y_1 = c\sqrt{N}$ . A symmetric computation shows that  $y_2$  will equal  $-c\sqrt{N}$  at the minimum. Plugging these values into  $TP_2$  gives

$$TP_2\left(x, c\sqrt{N}, -c\sqrt{N}\right) = \frac{c^2N + x^2}{2x} - \frac{1}{2}\sqrt{c^2N + x^2}.$$

This is a decreasing function of  $x$ . To see this, take the derivative

$$\begin{aligned} \frac{\partial TP_2(x, c\sqrt{N}, -c\sqrt{N})}{\partial x} &= \frac{1}{2} \frac{\partial \left( \sqrt{x^2 + 2c^2N + c^4N^2/x^2} - \sqrt{x^2 + c^2N} \right)}{\partial x} \\ &= \frac{1}{2} \left( \left( x - \frac{c^4N^2}{x^3} \right) \left( x^2 + 2c^2N + \frac{c^4N^2}{x^2} \right)^{-\frac{1}{2}} - x \left( x^2 + c^2N \right)^{-\frac{1}{2}} \right) \\ &< \frac{x}{2} \left( \left( x^2 + 2c^2N + \frac{c^4N^2}{x^2} \right)^{-\frac{1}{2}} - \left( x^2 + c^2N \right)^{-\frac{1}{2}} \right) \\ &< 0 \end{aligned}$$

because the function  $y \mapsto y^{-\frac{1}{2}}$  is decreasing. Therefore, the minimum value of  $TP_2$  is attained when  $x$  is as large as possible and

$$TP_2(x, y_1, y_2) \geq TP_2(N, c\sqrt{N}, -c\sqrt{N}) = \frac{1}{2} \left( c^2 + N - \sqrt{N(c^2 + N)} \right),$$

which is a decreasing function of  $N$  by an argument similar to the previous one.

## REFERENCES

- [1] Henry Adams, Manuchehr Aminian, Elin Farnell, Michael Kirby, Joshua Mirth, Rachel Neville, Chris Peterson, Patrick Shipman, and Clayton Shonkwiler. A fractal dimension for measures via persistent homology. *Abel Symposia*, 2019. arXiv:1808.01079.
- [2] Vincent Beffara. Hausdorff dimensions for sle6. *The Annals of Probability*, 2004.
- [3] M. G. Bouligand. Ensembles impropres et nombre dimensionnel. *Bulletin des Sciences Mathématiques*, 1928.
- [4] F. Chazal, V. de Silva, M. Glisse, and S. Oudot. *The Structure and Stability of Persistence Modules*. Springer, 2016.
- [5] Frederic Chazal, David Cohen-Steiner, Marc Glisse, Leonidas J. Guibas, and Steve Y. Oudot. Proximity of persistence modules and their diagrams. *Proceedings of the 25th Annual ACM Symposium on Computational Geometry (SoCG)*, 2009.
- [6] Frederic Chazal, Vin de Silva, and Steve Oudot. Persistence stability for geometric complexes. *Geometriae Dedicata*, 2014.
- [7] Siu-Wing Cheng, Tamal K. Dey, and Jonathan Shewchuk. *Delaunay Mesh Generation*. Chapman and Hall, 2012.
- [8] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete and Computational Geometry*, 2007.
- [9] David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Yuriy Mileyko. Lipschitz functions have  $l_p$ -stable persistence. *Foundations of Computational Mathematics*, 2010.
- [10] Vin De Silva and Robert Ghrist. Coverage in sensor networks via persistent homology. *Algebraic and Geometric Topology*, 2007.

- [11] Boris Delaunay. Sur la sphère vide. *Bulletin de l'Académie des Sciences de l'URSS, Classe des Sciences Mathématiques et Naturelles*, 1934.
- [12] H. Edelsbrunner and J. Harer. *Computational Topology: an Introduction*. AMS Press, 2010.
- [13] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discrete and Computational Geometry*, 2002.
- [14] Herbert Edelsbrunner and John Harer. Persistent homology — a survey. *Contemporary Mathematics*, 2008.
- [15] Kenneth Falconer. *Fractal Geometry: Mathematical Foundations and Applications*. Wiley, 2014.
- [16] Michael Goff. Extremal Betti numbers of Vietoris-Rips complexes. *Discrete and Computational Geometry*, 2011.
- [17] Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 2002.
- [18] Tom Kennedy. Numerical computations for the schramm-loewner evolution. *Journal of Statistical Physics*, 2009.
- [19] Harry Kesten and Sungchul Lee. The central limit theorem for weighted minimal spanning trees on random points. *Annals of Applied Probability*, 1996.
- [20] Gady Kozma, Zvi Lotker, and Gideon Stupp. The minimal spanning tree and the upper box dimension. *Proceedings of the American Mathematical Society*, 2006.
- [21] R. D. MacPherson and B. Schweinhart. Measuring shape with topology. *Journal of Mathematical Physics*, 53(7), 2012.
- [22] P. McMullen. The maximum number of faces of a convex polytope. *Mathematika*, 1970.
- [23] Olivier Devillers. Nina Amenta, Dominique Attali. Complexity of delaunay triangulation for points on lower-dimensional polyhedra. Technical report, INRIA, 2006.
- [24] Florian Pfender and Gunter M. Ziegler. Kissing numbers, sphere packings, and some unexpected proofs. *Notices of the AMS*, 2004.
- [25] Vanessa Robins. *Computational topology at multiple resolutions: foundations and applications to fractals and dynamics*. PhD thesis, University of Colorado at Boulder, 2000.
- [26] Harlan Sexton and Mikael Vejdemo-Johansson. jPlex, December 2008. <http://www.math.colostate.edu/~adams/jplex/index.html>.
- [27] Richard Stalney. The upper bound conjecture and Cohen-Macaulay rings. *Studies in Applied Mathematics*, 1975.
- [28] J. Michael Steele. Growth rates of Euclidean minimal spanning trees with power weighted edges. *Annals of Probability*, 1988.
- [29] The CGAL Project. *CGAL User and Reference Manual*. CGAL Editorial Board, 4.11 edition, 2017.
- [30] Leopold Vietoris. Über den Zusammenhang kompakter Räume und eine Klasse von Zusammenhangstreuen Abbildungen. *Mathematische Annalen*, 1927.
- [31] A. Zomorodian and G. Carlsson. Computing persistent homology. *Discrete and Computational Geometry*, 33:249–274, 2005.