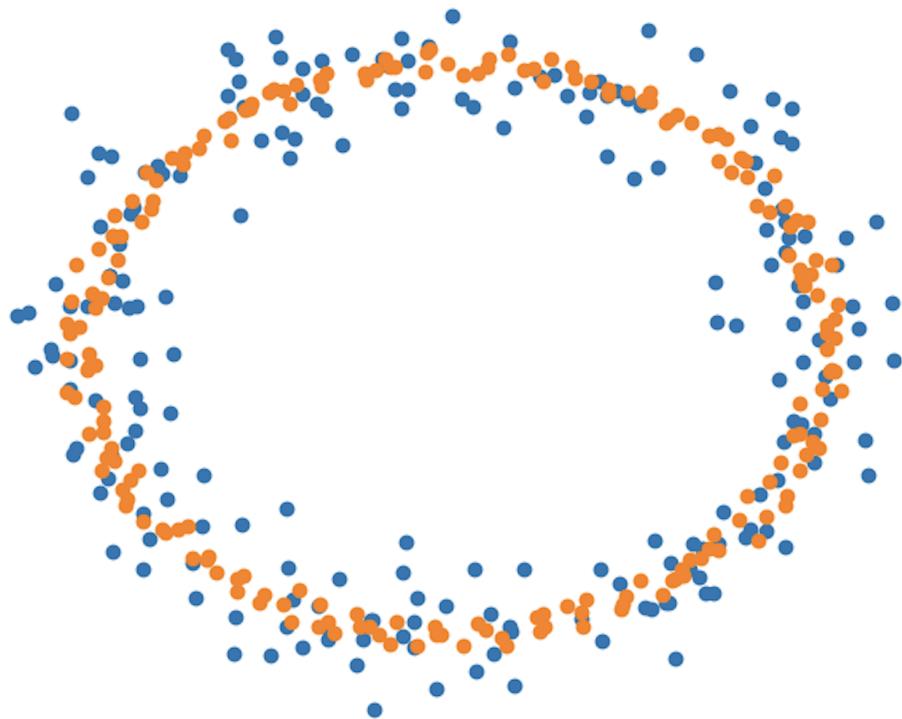


# Topological Data Analysis

## Mastermath

Magnus Bakke Botnan  
<https://www.few.vu.nl/~botnan/>

Version: June 17, 2024



# Contents

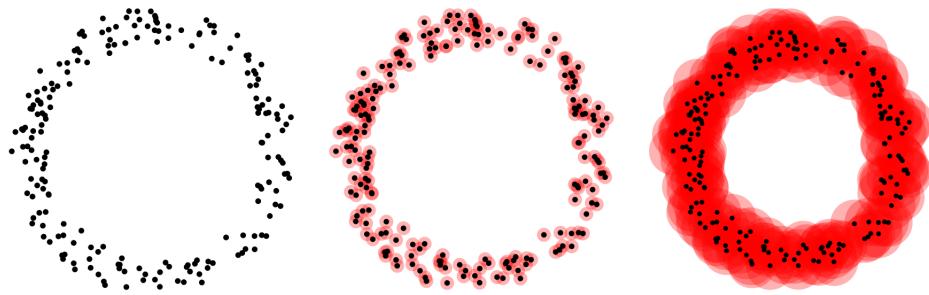
<b>1 Motivation and Simplicial Complexes</b>	<b>4</b>
1.1 Simplices . . . . .	4
1.2 Simplicial Complexes . . . . .	5
1.3 Abstract Simplicial Complexes . . . . .	6
1.4 Triangulations . . . . .	7
1.5 The Čech and Vietoris–Rips Simplicial Complexes . . . . .	7
1.6 The Alpha Complex . . . . .	10
1.7 Exercises . . . . .	12
<b>2 The Euler Characteristic and Simplicial Homology</b>	<b>14</b>
2.1 Free Vector Spaces . . . . .	16
2.2 The Betti Number and the Euler Characteristic . . . . .	16
2.3 A Remark on the Choice of Coefficients . . . . .	19
2.4 Simplicial Homology . . . . .	20
2.5 Exercises . . . . .	22
<b>3 Computation, Induced Maps and Invariance</b>	<b>23</b>
3.1 The Standard Algorithm . . . . .	23
3.1.1 The Homology of a Triangle . . . . .	23
3.1.2 The Algorithm . . . . .	24
3.2 Simplicial Maps and their Induced Maps in Homology . . . . .	25
3.3 Simplicial Approximations and Invariance of Homology . . . . .	26
3.3.1 Simplicial Maps Revisited . . . . .	26
3.3.2 Simplicial Approximations . . . . .	27
3.4 Exercises . . . . .	30
<b>4 Homotopy Equivalence, Relative Homology and Sensor Networks</b>	<b>31</b>
4.1 Homotopy Equivalence . . . . .	31
4.2 The Nerve Theorem . . . . .	32
4.2.1 The Čech Complex . . . . .	34
4.3 Relative Homology . . . . .	34
4.4 Sensor Networks . . . . .	36
4.4.1 The Covering Problem . . . . .	36
4.5 Exercises . . . . .	38
<b>5 Topology &amp; Homology Inference and Persistent Betti Numbers</b>	<b>39</b>
5.1 Topology & Homology Inference . . . . .	39
5.2 Persistent Betti Numbers . . . . .	41
5.2.1 The Čech and Vietoris–Rips Filtrations . . . . .	44
5.3 Exercises . . . . .	46
<b>6 Algebraic Foundations</b>	<b>47</b>
6.1 Motivation . . . . .	47
6.2 Persistence Modules . . . . .	48
6.3 Exercises . . . . .	52

<b>7</b>	<b>The Persistence Algorithm</b>	<b>53</b>
7.1	The standard algorithm . . . . .	53
7.2	Clearing . . . . .	56
7.3	Exercises . . . . .	57
<b>8</b>	<b>Cohomology</b>	<b>58</b>
8.1	Dual Persistence Modules . . . . .	58
8.2	Simplicial Cohomology . . . . .	58
8.3	Persistent Cohomology . . . . .	61
8.3.1	Computation . . . . .	62
8.3.2	Clearing . . . . .	63
8.4	Exercises . . . . .	64
<b>9</b>	<b>Stability of Persistent Homology</b>	<b>65</b>
9.1	The Bottleneck Distance . . . . .	65
9.1.1	Stability of Persistent Homology . . . . .	66
9.1.2	Generalizations . . . . .	68
9.2	Interleavings . . . . .	69
9.3	Exercises . . . . .	71
<b>10</b>	<b>Bjerkevik's Proof of the Isometry Theorem</b>	<b>72</b>
10.1	Part 1: $d_I \leq d_B$ . . . . .	72
10.2	Part 2: $d_B \leq d_I$ (the algebraic stability theorem) . . . . .	73
10.2.1	Hall's Theorem . . . . .	77
10.2.2	Wrapping Up . . . . .	77
10.3	Exercises . . . . .	79
<b>11</b>	<b>Clustering</b>	<b>80</b>
11.1	Examples of Clustering Methods . . . . .	80
11.2	Kleinberg's Theorem . . . . .	81
11.3	ToMATo . . . . .	82
11.3.1	The Discrete Case . . . . .	83
11.4	Exercises . . . . .	87
<b>12</b>	<b>Reeb Graphs and Mapper</b>	<b>88</b>
12.1	Reeb Graphs . . . . .	88
12.2	Morse Functions . . . . .	90
12.3	The Continuous Mapper . . . . .	91
12.4	Mapper in Practice . . . . .	92
12.5	Exercises . . . . .	95
<b>13</b>	<b>Zigzag Persistent Homology</b>	<b>96</b>
13.1	Zigzag Persistence Modules . . . . .	96
13.2	The diamond principle . . . . .	97
13.3	Levelset Zigzag Persistent Homology . . . . .	98
13.3.1	Reeb Graphs . . . . .	100
13.4	Computing the Zigzag Barcode . . . . .	102
13.5	An Algorithm for Zigzag Persistent Homology . . . . .	102
13.6	Exercises . . . . .	105

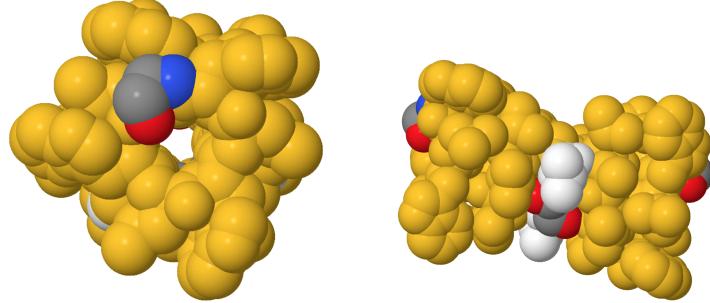
<b>14 Multiparameter Persistent Homology</b>	<b>106</b>
14.1 Multiparameter Persistence . . . . .	106
14.2 The Fibered Barcode and RIVET . . . . .	108
14.3 Distances . . . . .	109
14.4 The Rank Invariant . . . . .	111
14.5 Signed Barcodes . . . . .	111
14.6 Exercises . . . . .	114
<b>15 Non-linear dimensionality reduction</b>	<b>115</b>
15.1 ISOMAP . . . . .	115
15.1.1 Principal component analysis . . . . .	117
15.1.2 Multidimensional scaling . . . . .	117
15.1.3 The ISOMAP algorithm . . . . .	118
15.2 UMAP . . . . .	118
15.2.1 Step 1: the local fuzzy graph . . . . .	119
15.2.2 Step 2: gluing fuzzy graphs . . . . .	119
15.2.3 Step 3: finding a low-dimensional embedding . . . . .	120
15.3 Exercises . . . . .	121

# 1 Motivation and Simplicial Complexes

The idea of *Topological Data Analysis* (TDA) is to assign topological invariants to *data*. Data however is typically presented as a discrete sample whose topology is rather uninteresting. Therefore we must transform the data into a "continuous" object which (hopefully) is topologically similar to the shape of the underlying geometric object from which the data is sampled. Take for example the discrete set of data points in the figure below. Clearly, there is a "circular structure" to the data and our pattern-seeking brains have no trouble inferring such information. Topology per se does not measure this circularity but after thickening every point by a well-chosen thickness an annulus appears. This is a topological object with a well-defined "circular structure".



Such circular structures appear in one of the early motivations behind TDA. Namely, the need to robustly detect tunnels in protein molecules [22].



Ultimately, the resulting continuous spaces need to be discretized in order to process them on a computer, and this discretization process is typically done by means of *simplicial complexes*. In this section we discuss simplicial complexes in general and three constructions in particular: the *Cech*, *Vietoris–Rips*, and *alpha* complexes

## 1.1 Simplices

A set of points  $\{p_0, \dots, p_n\} \subset \mathbb{R}^d$  is said to be **geometrically independent**<sup>1</sup> if for any (real) scalars  $\{t_i\}_{i=0}^n$ , the equations

$$\sum_{i=0}^n t_i = 0 \quad \sum_{i=0}^n t_i p_i = 0 \quad (1)$$

imply that  $t_0 = t_1 = \dots = t_n = 0$ . This is clearly equivalent to the vectors  $\{p_1 - p_0, \dots, p_n - p_0\}$  being linearly independent (exercise).

---

<sup>1</sup>Or *affinely* independent.

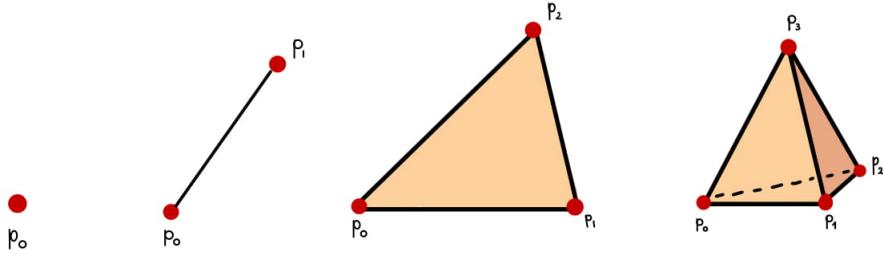


Figure 1: Examples of simplices of dimensions 0, 1, 2 and 3.

**Definition 1.1.** Let  $\{p_0, \dots, p_n\}$  be a geometrically independent set in  $\mathbb{R}^d$ . We define the  **$n$ -simplex  $\sigma$  spanned by the points  $p_i$**  to be the set of all points  $x \in \mathbb{R}^d$  of the form

$$x = \sum_{i=0}^n t_i p_i$$

where  $\sum_{i=0}^n t_i = 1$ , and  $t_i \geq 0$  for all  $i$ . The point  $x$  is a **convex combination** of the points  $p_0, \dots, p_n$ , and the uniquely determined numbers  $t_i$  are called the **barycentric coordinates**.

One immediately observes that any point is a 0-simplex, any two distinct points form a 1-simplex (a line), three non-collinear points form a 2-simplex (a triangle), four non-coplanar points form a 3-simplex (a tetrahedron) and so forth. Note that there can be no  $n$ -simplex in  $\mathbb{R}^d$  for  $n > d$ . The points  $p_0, \dots, p_n$  that span a simplex  $\sigma$  are called the **vertices** of  $\sigma$ ; the number  $n$  is the **dimension** of  $\sigma$ . A simplex is **regular** if all its edges have the same length. The **standard  $n$ -simplex  $\Delta^n$**  spanned by the endpoints of the unit vectors along the coordinate axes in  $\mathbb{R}^{n+1}$  is a regular simplex.

By definition, any subset of  $\{p_0, \dots, p_n\}$  is again geometrically independent and thus defines a simplex in its own right. We say that  $\tau$  is a **face** of  $\sigma$  if  $\tau$  is spanned by a subset of the vertices defining  $\sigma$ . We denote this by  $\tau \subseteq \sigma$  or  $\tau \subset \sigma$  if  $\tau$  is a **proper** face of  $\sigma$ , i.e.  $\tau \neq \sigma$ . The union of the proper faces of  $\sigma$  is called the **boundary** of  $\sigma$  and is denoted by  $\text{Bd}(\sigma)$ . The simplex  $\sigma$  is a (proper) **coface** of  $\tau$  if  $\tau$  is a (proper) face of  $\sigma$ .

## 1.2 Simplicial Complexes

By gluing simplices along faces we obtain topological spaces whose properties are determined by their combinatorial properties.

**Definition 1.2.** A **simplicial complex**  $K$  in  $\mathbb{R}^d$  is a finite<sup>2</sup> collection of simplices in  $\mathbb{R}^d$  such that:

1. Every face of a simplex of  $K$  is in  $K$ .
2. The non-empty intersection of any two simplices of  $K$  is a face of each of them.

The **dimension** of  $K$  is the maximum dimension of its simplices and is denoted by  $\dim K$ . If  $L$  is a subcollection of simplices that contains all faces of its elements, then  $L$  is a simplicial complex in its own right; it is called a **subcomplex** of  $K$ . The  **$p$ -skeleton**  $K^{(p)}$  of  $K$  is the collection of all simplices of  $K$  of dimension at most  $p$ . In particular,  $K^{(0)}$  is the vertex set of  $K$ .

Let  $|K|$  denote the subset of  $\mathbb{R}^d$  given by the union of all the simplices in  $K$ , equipped with the subspace topology. This space is called the **underlying space** of  $K$ . A **polyhedron** is a subset of  $\mathbb{R}^d$  that can be realized as the underlying space of a simplicial complex.

---

<sup>2</sup>We will only consider finite simplicial complexes in this course.

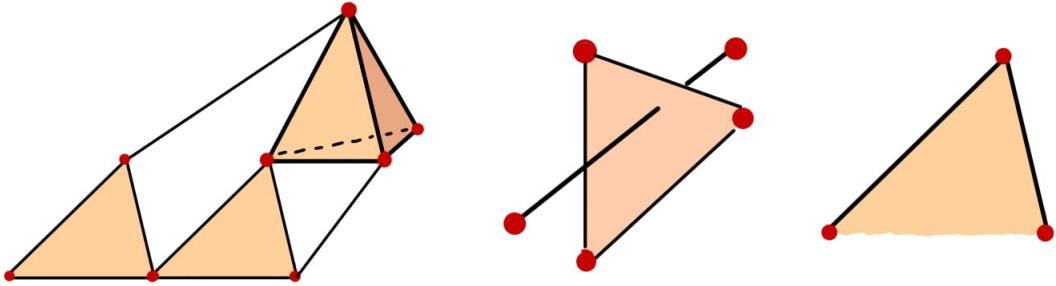


Figure 2: Left: A simplicial complex of dimension 3. Middle: This is not a simplicial complex as the intersection of the two simplices is not a face. Right: Not a simplicial complex.

### 1.3 Abstract Simplicial Complexes

Often the precise geometry of the simplicial complex is not important. We now introduce a purely combinatorial description of simplicial complexes under which two objects are isomorphic if and only if their geometric counterparts are linearly isomorphic.

**Definition 1.3.** An **abstract simplicial complex** is a finite collection  $A$  of finite non-empty sets, such that if  $\alpha$  is an element of  $A$ , then so is every nonempty subset of  $\alpha$ .

We call the elements of  $A$  **simplices** and the **dimension** of  $\alpha \in A$  is  $\dim \alpha = |\alpha| - 1$ . As above the dimension of  $A$  is given by  $\dim A = \max_{\alpha \in A} \dim \alpha$ . We say that  $B \subseteq A$  is a **subcomplex** of  $A$  if  $B$  is an abstract simplicial complex. The **vertices**  $V_A$  of  $A$  is the union of the one-point elements of  $A$ . We say that two simplicial complexes  $A$  and  $B$  are **isomorphic** if there is a bijective correspondence  $f: V_A \rightarrow V_B$  satisfying  $\{a_0, \dots, a_n\} \in A$  if and only if  $\{f(a_0), \dots, f(a_n)\} \in B$ .

A (geometric) simplicial complex  $K$  defines an abstract simplicial complex  $A$  by identifying any simplex in  $K$  with its vertices. We say that  $A$  is the **vertex scheme** of  $K$ .

**Example 1.4.** The 1-simplex connecting  $p_0$  and  $p_1$  has vertex scheme  $\{\{p_0\}, \{p_1\}, \{p_0, p_1\}\}$ .

If  $A$  is the vertex scheme of  $K$ , then we say that  $K$  is a **geometric realization** of  $A$ . It is easy to see that any abstract simplicial complex on  $d + 1$  vertices can be geometrically realized as subcomplex of any  $d$ -simplex in  $\mathbb{R}^d$ . The next lemma shows that we often can do considerably better.

A finite subset  $P \subset \mathbb{R}^d$  is in **general position** if any subset of at most  $d + 1$  points is geometrically independent.

*Remark 1.5.* Most point-sets are in general position, and those that are not can be brought into general position by an arbitrary small perturbation. One may verify that any subset of at most  $r + 1$  points in the image of the curve  $\gamma: \mathbb{R} \rightarrow \mathbb{R}^r$  defined by  $\gamma(t) = (t, t^2, \dots, t^r)$  is in general position (exercise).

**Lemma 1.6.** An abstract simplicial complex  $A$  with  $\dim A = d$  has a geometric realization in  $\mathbb{R}^{2d+1}$ .

*Proof.* Let  $h: V_A \rightarrow \mathbb{R}^{2d+1}$  be any function such that  $h(V_A)$  is in general position, e.g. let  $h$  be such that  $h(V_A) = \gamma(S)$  where  $\gamma$  is the function from Remark 1.5 and  $|S| = |V_A|$ . Any simplex  $\alpha \in A$  has at most  $d + 1$  vertices, and thus its vertices correspond to a geometrically independent set under the function  $h$ . We denote this set by  $h(\alpha)$ , and the associated geometric simplex by  $\sigma_\alpha$ . It remains to show that the intersection of any two simplices is a face of both the simplices.

Let  $\alpha$  and  $\beta$  be any two simplices and observe that  $|\alpha \cup \beta| \leq 2(d+1)$ , and therefore  $h(\alpha) \cup h(\beta)$  forms a geometrically independent set. It follows that any point  $x \in \sigma_\alpha \cap \sigma_\beta$  can be uniquely represented as a convex combination of the vertices  $h(\alpha) \cup h(\beta)$ . Hence the barycentric coordinates of  $x$  can be non-zero only for  $t_i \in h(\alpha \cap \beta)$ . We conclude that  $x \in \sigma_{\alpha \cap \beta}$ .  $\square$

## 1.4 Triangulations

**Definition 1.7.** A topological space  $X$  is **triangulable** if there exists a simplicial complex  $K$  whose underlying space is homeomorphic to  $X$ . Any homeomorphism  $h: |K| \rightarrow X$  is a **triangulation**.

**Example 1.8.** Abstract simplicial complexes can be very useful in defining triangulations of spaces; see Fig. 3 for an example of how to specify a triangulation of the cylinder (of finite height).

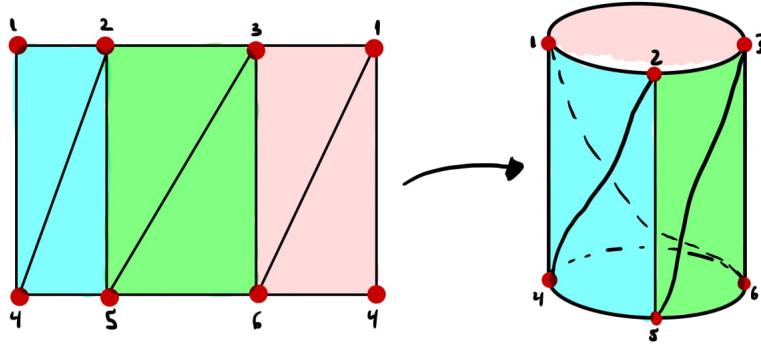


Figure 3: Using an abstract simplicial complex to generate a triangulation of the cylinder.

*Remark 1.9.* It is a well-known fact that any closed surface can be triangulated and the same holds true for 3-dimensional manifolds. The situation becomes significantly more complicated in dimension 4 where manifolds need not admit a triangulation. The *comb space* is an example of a compact, path-connected subset of the plane that cannot be triangulated. To see this, note that a simplicial complex is locally connected, whereas the comb space is not.

## 1.5 The Čech and Vietoris–Rips Simplicial Complexes

We now turn our attention to two of the most widely used simplicial complexes constructed in topological data analysis. We shall return to them for a more formal treatment later in the course. In what follows  $P = \{p_0, \dots, p_n\} \subset \mathbb{R}^d$ , and  $B_r(p) := \{x' \in \mathbb{R}^d : \|x - x'\| \leq r\}$  denotes the closed ball of radius  $r$  centered at  $p$ .

**Definition 1.10.** The **Čech complex** of  $P$  at scale  $r$  is the simplicial complex

$$\text{Čech}_r(P) = \{\sigma \subseteq P : \bigcap_{p \in \sigma} B_r(p) \neq \emptyset\}.$$

*Remark 1.11.*  $|\text{Čech}_r(P)|$  is homotopy equivalent to  $\bigcup_{p \in P} B_r(P)$ . This will be proven later in the course.

The Čech complex may contain high dimensional simplices even for planar point sets; see Fig. 4. Determining if  $\sigma \in \text{Čech}_r(P)$  may appear unwieldy for simplices of large cardinality. The following theorem shows that we need not consider all possible intersections as long as the embedding dimension is low.

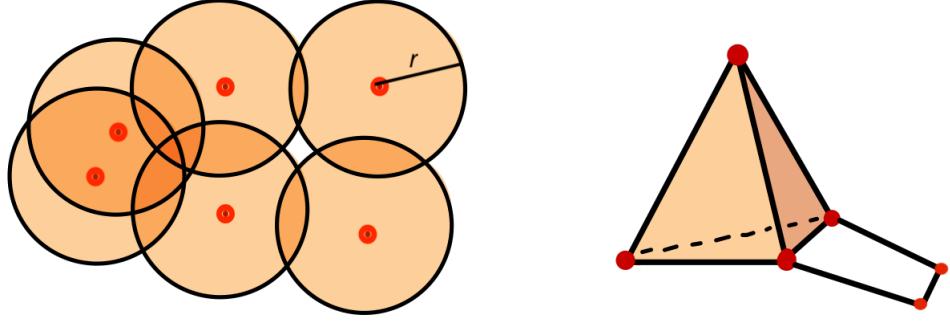


Figure 4: The Čech complex at scale  $r$  of a planar point set.

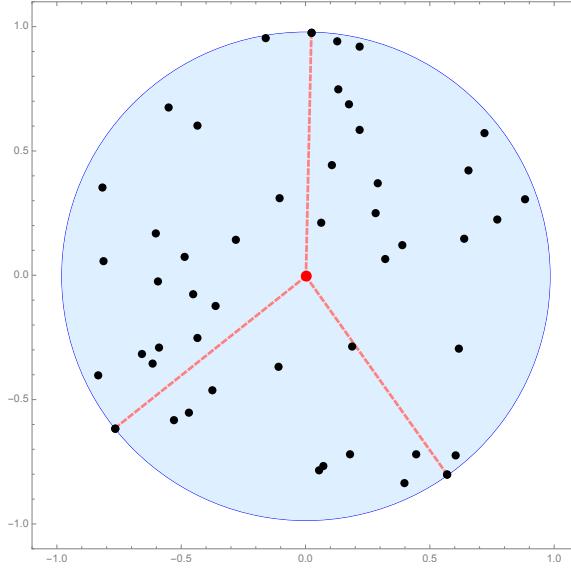


Figure 5: A set of points in the plane and the minimal enclosing ball.

**Theorem 1.12** (Helly's Theorem). *Let  $F$  denote a finite collection of closed, convex sets in  $\mathbb{R}^d$ . If the intersection of any  $d + 1$  sets is non-empty, then the whole collection has non-empty intersection.*

*Proof.* See e.g. [17, III.2]. □

Hence, if  $P \subset \mathbb{R}^3$ , then computing all quadruple intersections will be sufficient for determining  $\text{Čech}_r(P)$ . Furthermore, one observes that  $\sigma \in \text{Čech}_r(P)$  if and only if there exists a ball of radius  $r$  containing all the elements of  $\sigma$ . The unique (exercise) smallest enclosing ball containing a set of points  $\sigma \subseteq P$  is called the **minimal enclosing ball**, and it can be effectively computed with the *miniball algorithm*; see e.g. [17, III.2].

Recall that the **diameter** of a subset  $\sigma \subseteq P$  is given by  $\text{diam}(\sigma) = \max_{p_i, p_j \in \sigma} \|p_i - p_j\|$ .

**Definition 1.13.** The **Vietoris–Rips complex** of  $P$  at scale  $r$  is the simplicial complex

$$\text{VR}_r(P) = \{\sigma \subseteq P : \text{diam}(\sigma) \leq 2r\}.$$

It is straightforward to verify that  $\text{Čech}_r(P) \subseteq \text{VR}_r(P) \subseteq \text{Čech}_{2r}(P)$ . The latter bound can however be significantly sharpened.

**Theorem 1.14** (Jung's Theorem). *Let  $Q$  be a finite point set in  $\mathbb{R}^d$ . Then  $Q$  is contained in a closed ball with radius  $r \leq \delta \cdot \text{diam}(Q)$  where  $\delta = \sqrt{\frac{d}{2(d+1)}}$ .*

*Proof.* Working inductively, we first observe that the result is true for  $d = 1$ .

To prove the result in dimension  $d$ , it suffices to consider the case where  $|Q| = d + 1$ , and the points are geometrically independent. Indeed, if the points of  $Q$  are not geometrically independent, or  $|Q| < d + 1$ , then  $Q$  can be isometrically embedded in  $\mathbb{R}^{d-1}$ . The result then follows from the inductive hypothesis. For  $|Q| > d + 1$  we note that Jung's theorem is equivalent to the assertion that

$$\bigcap_{q \in Q} B_{\delta \cdot \text{diam}(Q)}(q) \neq \emptyset.$$

From Helly's theorem (Theorem 1.12) it is thus sufficient to verify that the intersection is non-trivial for subsets of  $Q$  containing  $d + 1$  points.

Assume that the minimal enclosing ball  $B$  of  $Q$  has center  $z$  and radius  $r$ . Observe that  $z$  is contained in the simplex  $\sigma$  spanned by  $Q$ . If it were not, then there would exist a point  $q \in Q$ , such that  $z$  and  $q$  are separated by the affine  $(d - 1)$ -dimensional plane spanned by the points in  $Q - \{q\}$  (see Exercise 2(c)). The orthogonal projection of  $z$  onto this plane would be strictly closer to all of the points of  $Q$ , contradicting that the minimal enclosing ball has center  $z$ . Hence,  $z$  can be written (uniquely) as a convex combination of the points  $\{q_1, \dots, q_{d+1}\} = Q$  as follows

$$z = \sum_{i=1}^{d+1} t_i q_i, \quad \sum_{i=1}^{d+1} t_i = 1, \quad t_i \geq 0.$$

Furthermore, we may assume that every point  $p_i$  is contained in the boundary of  $B$ . To see this, let  $Q' = Q \cap \partial B$  denote the set of boundary points, and let  $H$  be the hyperplane spanned by the points in  $Q'$ . Then a sufficiently small translation of  $z$  in the direction of the orthogonal projection onto  $H$  gives a point  $z'$  satisfying  $\|z' - q_i\| < r$  for all  $q_i \in Q$ , contradicting that  $B$  was minimal.

Without loss of generality we will assume that  $z = 0$ . For a fixed  $1 \leq k \leq d + 1$  we get

$$\begin{aligned} \sum_{i=1, i \neq k}^{d+1} t_i \|q_i - q_k\|^2 &= \sum_{i=1}^{d+1} t_i \|q_i - q_k\|^2 = \sum_{i=1}^{d+1} t_i \|q_i\|^2 - 2q_k \cdot \sum_{i=1}^{d+1} t_i q_i + \|q_k\|^2 \sum_{i=1}^{d+1} t_i \\ &= r^2 - 0 + r^2 = 2r^2. \end{aligned}$$

Summing over all  $1 \leq k \leq d + 1$  gives

$$\begin{aligned} (d + 1) \cdot (2r^2) &= \sum_{k=1}^{d+1} \sum_{i=1, i \neq k}^{d+1} t_i \|q_i - q_k\|^2 \leq \sum_{k=1}^{d+1} \sum_{i=1, i \neq k}^{d+1} t_i (\text{diam}(Q))^2 \\ &= \sum_{k=1}^{d+1} (1 - t_k) (\text{diam}(Q))^2 = d(\text{diam}(Q))^2. \end{aligned}$$

We conclude that

$$r \leq \sqrt{\frac{d}{2(d+1)}} \cdot \text{diam}(Q).$$

□

The bound is tight for regular  $d$ -simplices in  $\mathbb{R}^d$  (exercise).

**Corollary 1.15.**  $\text{Cech}_r(P) \subseteq \text{VR}_r(P) \subseteq \text{Cech}_{2\delta r}(P)$  where  $2\delta = 2\sqrt{\frac{d}{2(d+1)}} = \sqrt{\frac{2d}{d+1}}$ .

## 1.6 The Alpha Complex

For a set of points  $P \subset \mathbb{R}^d$ , the alpha complex at scale  $r$  is a comparatively small subcomplex of  $\text{Cech}_r(P)$  capturing the "same topology"; see Theorem 4.11. While the construction of the alpha complex suffers from "curse of dimensionality", i.e., its complexity grows exponentially in the ambient dimension, it is by far the most popular construction in TDA for  $d \in \{2, 3\}$ .

The **Voronoi cell** of  $p$  is

$$V(p) := \{x \in \mathbb{R}^d : \|x - p\| \leq \|x - q\| \text{ for all } p \neq q \in P\}.$$

Closely related to the Čech complex, we have the following definition.

**Definition 1.16.** The **alpha complex** of  $P \subset \mathbb{R}^n$  at scale  $r \geq 0$  is the simplicial complex

$$\text{Alpha}_r(P) = \{\sigma \subseteq P : \bigcap_{p \in \sigma} (V(p) \cap B_r(p)) \neq \emptyset\},$$

and

$$\text{Alpha}_\infty(P) = \{\sigma \subseteq P : \bigcap_{p \in \sigma} V(p) \neq \emptyset\}.$$

The latter case is also known as the *Delaunay triangulation*.

The efficacy of the alpha complex stems from the following theorem which we shall not prove; see e.g., [5] for a complete discussion.

**Theorem 1.17.** Let  $P \subset \mathbb{R}^d$  and assume that no  $d + 2$  points lie on the same  $(d - 1)$ -sphere<sup>3</sup>. Then, the following holds:

1.  $\text{Alpha}_\infty(P)$  has a geometric realization in  $\mathbb{R}^d$  defined by mapping an abstract  $k$ -simplex  $\sigma$  to the geometric  $k$ -simplex spanned by the points in  $\sigma$ ,
2. the number of simplices in  $\text{Alpha}_\infty(P)$  grows as  $O(|P|^{\lceil \frac{d}{2} \rceil})$ ,
3.  $\text{Alpha}_\infty(P)$  can be computed in time  $O(|P| \log |P| + |P|^{\lceil \frac{d}{2} \rceil})$ .

Note that the same theorem applies to any finite  $r$  by restriction. Furthermore, for planar points, the number of simplices in the alpha complex grows linearly in the data points. This is in contrast to the Čech and Vietoris–Rips complexes, for which there for any point set  $P$  is  $2^{|P|} - 1$  simplices in total at large scales, of which  $\binom{|P|}{3}$  are 2-simplices.

---

<sup>3</sup>This condition is known as *spherical general position*.

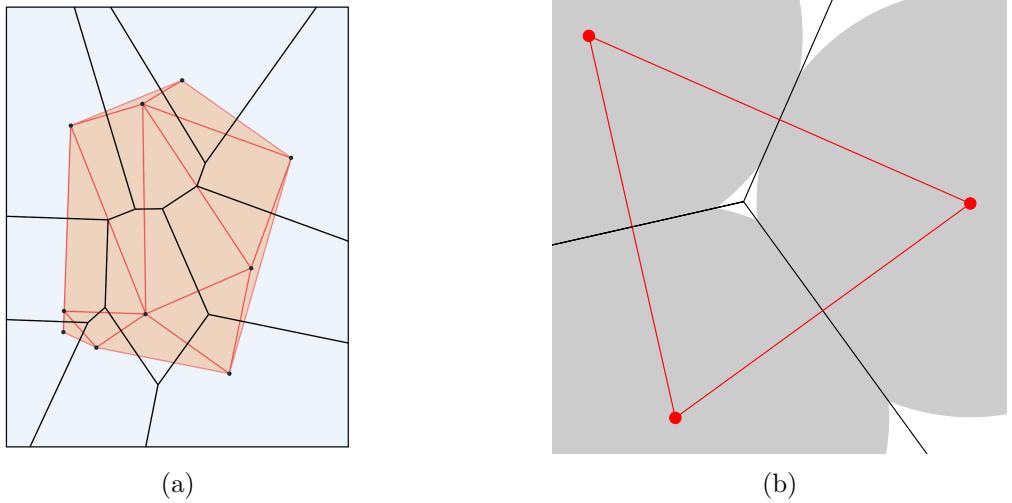


Figure 6: (a) The alpha complex of a point cloud superimposed on the Voronoi cells. (b) Intersecting the Voronoi cells with balls of radius  $r$  results in the regions colored gray. As there is no triple intersection, the triangle is not filled in.

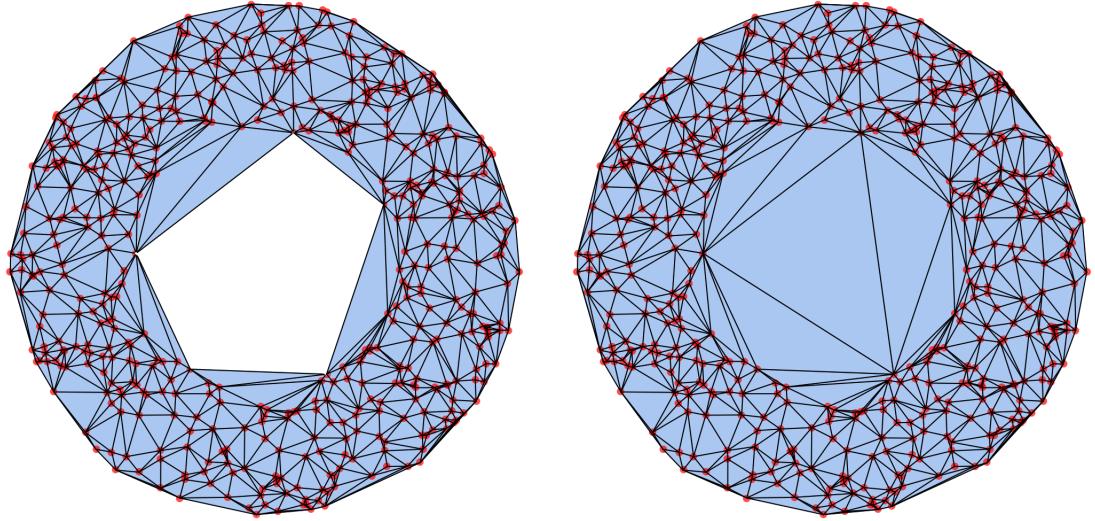


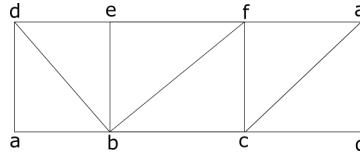
Figure 7: Alpha complexes of a point cloud  $P$  for two radii  $r_1 < r_2$ .

## 1.7 Exercises

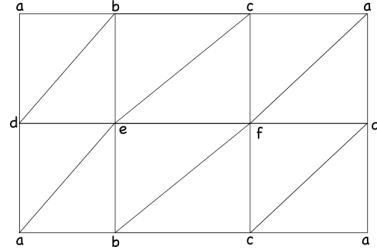
1. Verify that the points in  $\{p_0, \dots, p_n\}$  are geometrically independent if and only if  $\{p_1 - p_0, \dots, p_n - p_0\}$  is a linearly independent set.
2. Let  $P = \{p_0, \dots, p_n\} \subset \mathbb{R}^d$  be a geometrically independent set.
  - (a) Show that every point  $x$  in the affine plane spanned by the points in  $P$  can be written uniquely as  $x = \sum_{i=0}^n p_i t_i$  where  $\sum_{i=0}^n t_i = 1$ .
  - (b) Show that all  $t_i \geq 0$  if and only if  $x$  is contained in the  $n$ -simplex  $\sigma$  spanned by  $P$ .
  - (c) Assume  $n = d$ , and observe that for each  $i$ , the affine plane  $E_i$  spanned by the points  $P - \{p_i\}$  defines two halfspaces whose union is  $\mathbb{R}^d$  and whose intersection is  $E_i$ . Let  $H_i$  denote the halfspace containing the point  $p_i$ , and denote by  $\sigma$  the  $n$ -simplex spanned by the points in  $P$ . Show that

$$\sigma = \bigcap_{i=0}^n H_i.$$

3. Which topological space is this?



4. ... and what about this space? (It's not the torus!)



5. Verify that the Čech and Vietoris–Rips complexes are abstract simplicial complexes.
6. Show that any  $r + 1$  points are the image of the curve  $\gamma: \mathbb{R} \rightarrow \mathbb{R}^r$  defined by  $\gamma(t) = (t, t^2, \dots, t^r)$  are geometrically independent.
7. Let  $P$  be the vertex set of a regular  $d$ -simplex in  $\mathbb{R}^d$  with edge lengths equal to  $2r$ . Show that  $\text{VR}_r(P) \not\subseteq \text{Čech}_{r'}(P)$  for  $r' < r\sqrt{\frac{2d}{d+1}}$ . In other words, show that the bound in Jung's theorem is tight.
8. Give an example of a point cloud  $P \subset \mathbb{R}^2$  for which  $\dim \text{Alpha}_\infty(P) > 2$ .
9. (a) Define the **join** of two (disjoint) abstract simplicial complexes  $K$  and  $L$  to be the simplicial complex  $K * L$  with vertex set  $V_K \cup V_L$  and all simplices  $\{\sigma \cup \tau : \sigma \in K, \tau \in L\}$ . Verify that  $K * L$  is a simplicial complex, and show that  $\Delta^n * \Delta^m$  is isomorphic to  $\Delta^{n+m+1}$ .

- (b) Show that if  $K$  and  $L$  admit geometrical realizations  $|K| \subset \mathbb{R}^m$ ,  $|L| \subset \mathbb{R}^n$ , then  $K * L$  has a geometric realization in  $\mathbb{R}^{m+n+1}$ . (Hint: embed  $|K|$  in  $\mathbb{R}^m \times \{0, \dots, 0\}$  and  $|L|$  in  $\{0, \dots, 0\} \times \mathbb{R}^n \times \{1\}$ . ) Note: One can also check that  $|K * L| \cong |K| * |L|$  where  $*$  is the join as defined for topological spaces; this gives a combinatorial way to compute the join of triangulable spaces.
10. Show that the minimal enclosing ball of a finite set of points in  $\mathbb{R}^d$  always exists (hint: use a compactness argument). Show uniqueness for  $d = 2$ .

## 2 The Euler Characteristic and Simplicial Homology

Distinguishing topological spaces is a notoriously difficult problem: deciding if two simplicial complexes are homeomorphic (or homotopy equivalent) is an *undecidable problem*. Roughly speaking, this means that there exists no algorithm which takes as input two simplicial complexes and outputs "true" if and only if they are homeomorphic. This is even the case if the simplicial complexes are assumed to be triangulations of four-dimensional manifolds. Hence, to show that two spaces are *not* homeomorphic, it is often more fruitful to show that they look different under the lens of a tool which is insensitive to homeomorphism (or other forms of equivalence). *Algebraic topology* offers one such approach: algebraic invariants such as *homology* and *homotopy groups* are associated to topological spaces. We shall focus on the former of these two as its computation reduces to simple linear algebra whereas the latter is exceptionally hard to compute. Although the outputted homology invariants *can* be hard to interpret, we shall see that for low-dimensional simplicial complexes, the invariants have a clear interpretation in terms of connectivity, "circularity" and cavities. Our first topological invariant, the Euler characteristic, will illustrate this.

**Definition 2.1.** Let  $K$  be a simplicial complex and let  $K_i$  denote the number of  $i$ -simplices in  $K$ . The **Euler characteristic of  $K$**  is the integer

$$\chi(K) = \sum_{i \geq 0} (-1)^i K_i.$$

It is not at all clear that this number depends only on the homemorphism type of the underlying space  $|K|$ , and we will return to that fact in the next lecture. For now, we will try to see what it represents by considering some familiar topological spaces.

**Example 2.2.** Consider the planar simplicial complex of Fig. 8. Its Euler characteristic is

$$\chi = \#\text{vertices} - \#\text{edges} + \#\text{triangles} = 11 - 17 + 3 = -3.$$

We see that  $|\chi|$  is one less than the number of "holes" in the graph. The reader may verify that the same relation holds for other planar simplicial complexes. This is no coincidence: in the exercises you will prove that for a connected planar simplicial complex the following is always true

$$2 = \#\text{vertices} - \#\text{edges} + \#\text{faces}.$$

where number of faces is given by:

$$\#\text{faces} = \#\text{triangles} + \#\text{holes} + 1(\text{the unbouded component}).$$

Hence,

$$1 - \#\text{holes} = \#\text{vertices} - \#\text{edges} + \#\text{triangles} = \chi.$$

**Example 2.3.** The regular tetrahedron and regular icosahedron are two of the five platonic solids (see Fig. 9). They are both triangulations of the sphere and one can easily verify that their Euler characteristics coincide.

$$\begin{aligned}\chi(\text{tetrahedron}) &= 4 - 6 + 4 = 2 \\ \chi(\text{icosahedron}) &= 12 - 30 + 20 = 2.\end{aligned}$$

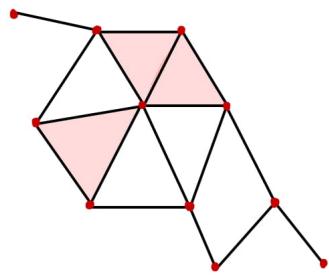


Figure 8: The Euler characteristic counts the number of circular holes.

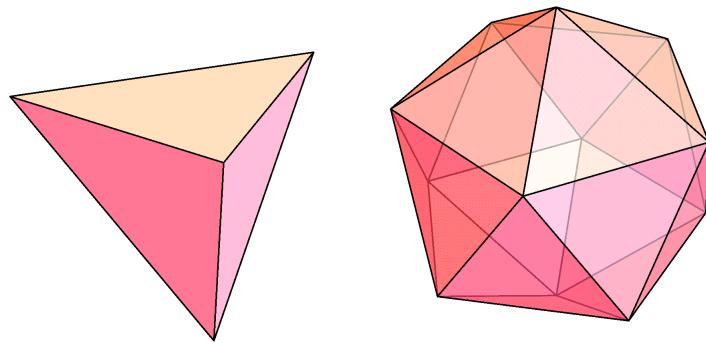


Figure 9: Left: tetrahedron. Right: icosahedron.

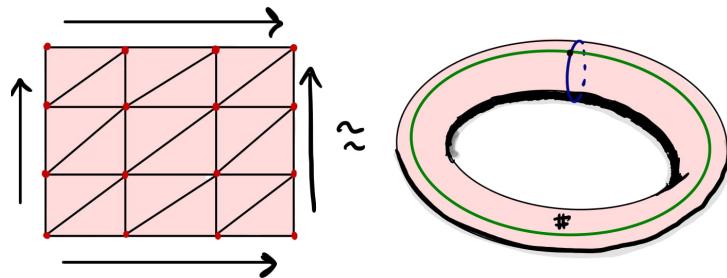


Figure 10: Left: A triangulation of the torus. Right: The torus and two independent loops

The torus on the other hand has a triangulation as shown in Fig. 10. The Euler characteristic of the triangulation is  $\chi = 9 - 27 + 18 = 0$ . Again this is no coincidence: for a simplicial complex  $K$  embedded in  $\mathbb{R}^3$  the following relation holds:

$$\chi(K) = \#\text{connected components} - \#\text{independent loops} + \#\text{voids}.$$

There are no non-trivial loops on the sphere, whereas the torus has two, as depicted in Fig. 10.

The fact that the Euler Characteristic is invariant under homeomorphism allows us one talk about *the* Euler characteristic of a triangulated topological space. Hence, the previous example shows that the torus and the sphere are not homeomorphic.

*Remark 2.4.* Remarkably, two closed, oriented surfaces are homeomorphic if and only if they have the same Euler characteristic: the Euler characteristic of a *genus*  $g$  surface is  $2 - 2g$ . In light of the introduction above, this can clearly not be the case for manifolds of arbitrary dimension (why?).

## 2.1 Free Vector Spaces

We briefly recap the construction of a free vector space on a finite set. Let  $S$  be a finite set and  $\mathbf{k}$  an arbitrary field. We define the **free  $\mathbf{k}$ -vector space generated by  $S$**  to be the vector space  $F(S)$  with elements given by formal linear combinations  $\sum_{s_i \in S} a_i s_i$  where each  $a_i \in \mathbf{k}$  and  $s_i \in S$ . Here addition is defined component-wise in the obvious way:

$$\sum_{s_i \in S} a_i s_i + \sum_{s_i \in S} b_i s_i = \sum_{s_i \in S} (a_i + b_i) s_i,$$

where  $a_i + b_i \in \mathbf{k}$ . The resulting  $\mathbf{k}$ -vector space has a canonical vector space basis given by the elements in  $S$ .

**Example 2.5.** Let  $\mathbf{k} = \mathbb{Z}_2$  be the field of two elements and  $S = \{a, b, c\}$  a set of three elements. Then  $F(S)$  is the 3-dimensional vector space with elements  $\{0, a, b, c, a + b, a + c, b + c\}$  and a basis  $\{a, b, c\}$ . Addition is straight-forward:  $(a + b) + (a + c) = (2a) + b + c = b + c$ .

**Example 2.6.** If we instead let  $\mathbf{k} = \mathbb{Z}_3$  and  $S = \{a, b\}$  then

$$F(S) = \{0, a, 2a, b, 2b, a + b, a + 2b, 2a + b, 2a + 2b\}.$$

We observe that  $F(S)$  has dimension  $|S|$  and cardinality  $|S|^{|k|}$ .

## 2.2 The Betti Number and the Euler Characteristic

We shall now construct a family of topological invariants which measure "connectedness" in a given dimension. By means of this we can make the interpretation of the Euler characteristic precise. For the time being we shall assume that  $\mathbf{k} = \mathbb{Z}_2$ .

**Definition 2.7.** Let  $n \geq 0$  be an integer and  $K$  a simplicial complex. The **vector space of  $n$ -chains in  $K$**  is the free  $\mathbb{Z}_2$ -vector space  $C_n(K)$  generated by the  $n$ -simplices of  $K$ .

By defining the boundary of an  $n$ -simplex to be the formal sum of its  $(n - 1)$ -dimensional faces, we get a linear transformation  $\partial_n: C_n(K) \rightarrow C_{n-1}(K)$ . Formally, for a simplex  $\{p_0, \dots, p_n\}$ , let  $\{p_0, \dots, \hat{p}_i, \dots, p_n\}$  denote the  $(n - 1)$ -simplex obtained by omitting the vertex  $p_i$ .

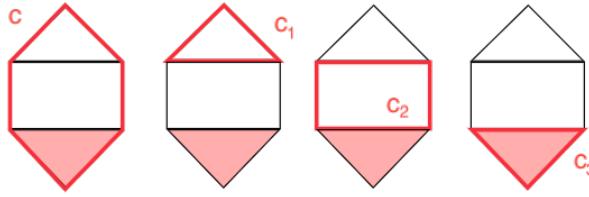


Figure 11: The cycle to the left is the sum of three other cycles.

**Definition 2.8.** The **boundary operator**  $\partial_n: C_n(K) \rightarrow C_{n-1}(K)$  is the linear transformation defined on simplices by

$$\partial_n(\{p_0, \dots, p_n\}) = \sum_{i=0}^n \{p_0, \dots, \hat{p}_i, \dots, p_n\}$$

The following lemma is what makes the theory work. Intuitively, it states that a boundary has no boundary.

**Lemma 2.9.** *The composition  $\partial_n \circ \partial_{n+1} = 0$  for all  $n \geq 0$ .*

*Proof.* By linearity it suffices to show that  $\partial_n \circ \partial_{n+1}(\sigma) = 0$  for an  $(n+1)$ -simplex  $\sigma = \{p_0, \dots, p_{n+1}\}$ .

$$\begin{aligned} \partial_n \circ \partial_{n+1}(\sigma) &= \partial_n \left( \sum_{i=0}^{n+1} \{p_0, \dots, \hat{p}_i, \dots, p_{n+1}\} \right) = \sum_{i=0}^{n+1} \partial_n(\{p_0, \dots, \hat{p}_i, \dots, p_{n+1}\}) \\ &= \sum_{i=0}^{n+1} \sum_{j=0, j \neq i}^{n+1} \{p_0, \dots, \hat{p}_j, \dots, \hat{p}_i, \dots, p_{n+1}\}. \end{aligned}$$

For  $i \neq j$  we see that the  $(n-1)$ -simplex  $\{p_0, \dots, \hat{p}_j, \dots, \hat{p}_i, \dots, p_{n+1}\}$  appears precisely two times in the sum. Since we are working over  $\mathbb{Z}_2$ , the result follows.  $\square$

We say that  $c \in C_n(K)$  is an  **$n$ -cycle** if  $\partial_n(c) = 0$  and we denote the associated vector space of  $n$ -cycles by  $Z_n(K) = \ker \partial_n = \{c \in C_n(K) : \partial_n(c) = 0\}$ . If  $c$  is in the image of  $\partial_{n+1}$ , that is,  $c = \partial_{n+1}(d)$  for some  $(n+1)$ -chain  $d$ , then  $c$  is an  **$n$ -boundary**. We denote the associated vector space of  $n$ -boundaries by  $B_n(K) = \text{Im } \partial_{n+1} = \{\partial_{n+1}(d) : d \in C_{n+1}(K)\}$ .

**Example 2.10.** Consider the simplicial complex in Fig. 11. Observe that the cycle  $c$  to the left can be written as a linear combination  $c = c_1 + c_2 + c_3$ . In fact, it is not hard to see that  $\{c_1, c_2, c_3\}$  forms a basis for the cycles in  $Z_1(K)$ ; any "hole" in the simplicial complex can be expressed as a linear combination of those cycles. We are however interested in counting the number of "unbounded holes" and therefore we ignore any contribution from the cycles of  $B_1(K) = \{c_3\}$ . The number of "circular holes" is thus  $\dim Z_1(K) - \dim B_1(K) = 3 - 1 = 2$ .

With the previous example fresh in mind we define

**Definition 2.11.** The  **$n$ -th Betti number** of  $K$  is  $\beta_n(K) = \dim Z_n(K) - \dim B_n(K)$ .

For simplicial complexes in  $\mathbb{R}^3$  the interpretation of the Betti numbers is clear:  $\beta_1$  is the number of loops, and  $\beta_2$  is the number of cavities. In general, interpreting the Betti numbers can be quite subtle. The following is however always true.

**Definition 2.12.** We say that a simplicial complex  $K$  is **path-connected** if any two vertices in  $K$  are connected by an edge path.

*Remark 2.13.* A simplicial complex  $K$  is path-connected if and only if  $|K|$  is path-connected in the topological sense. Proof: Assume that  $|K|$  is path-connected. Fix a vertex  $p$  in  $K$  and let  $K_0$  denote the simplicial complex (verify this) consisting of all simplices with the property that all of its vertices can be joined to  $p$  by an edge path. We let  $K_1$  denote the subcomplex of  $K$  consisting of all simplices that are not in  $K_0$ . Hence,  $K = K_0 \cup K_1$  and  $K_0 \cap K_1 = \emptyset$ . Since  $|K_0|$  and  $|K_1|$  are closed subsets of  $|K|$ , it follows from connectivity that  $K_1 = \emptyset$ . The converse statement is immediate.

**Lemma 2.14.**  $\beta_0(K)$  equals the number of path-connected components of  $|K|$ .

*Proof.* Assume that  $K$  is path-connected, and let  $\{p_1, \dots, p_m\}$  denote the vertex set of  $K$ . A basis for  $Z_0(K)$  is given by  $\cup_{i=1}^m \{p_1, p_2+p_1, \dots, p_m+p_1\}$ . Since  $K$  is path-connected, we can find, for every  $j \neq 1$ , a  $c \in C_1(K)$  such that  $\partial_1(c) = p_1 + p_j$ . This shows that  $\beta_0(K) \leq 1$ . To show that  $\beta_0 = 1$ , observe that a vertex cannot be in the image of  $\partial_1$ : for any  $c \in C_1(K)$ ,  $\partial_1(c)$  is either 0 or a non-trivial sum of an even number of vertices. Hence,  $\beta_0(K) = \dim Z_0(K) - \dim B_0(K) = 1$ . For the more general case, observe that if  $K = K^1 \cup K^2$  where  $K^1$  and  $K^2$  are disjoint, then the boundary operator decomposes as a direct sum

$$\partial_1: C_1(K) = C_1(K^1) \oplus C_1(K^2) \xrightarrow{\partial_1^1 \oplus \partial_1^2} C_0(K^1) \oplus C_0(K^2) = C_0(K).$$

Where  $\partial_1^i: C_1(K^i) \rightarrow C_0(K^i)$  is the first boundary operator for  $K_i$ . In particular,

$$\beta_0(K) = K_0 - \dim \text{Im } \partial_1 = K_0^1 + K_0^2 - \dim \text{Im } \partial_1^1 - \dim \text{Im } \partial_1^2 = \beta_0(K^1) + \beta_0(K^2).$$

The result follows by induction.  $\square$

**Theorem 2.15** (Euler-Poincare formula). *For a simplicial complex  $K$ ,*

$$\chi(K) = \sum_{i=0}^{\infty} (-1)^i K_i = \sum_{i=0}^{\infty} (-1)^i \beta_i(K).$$

*Proof.* From linear algebra we remember that for a linear transformation  $T: V \rightarrow W$  we have  $\dim V = \dim \ker T + \dim \text{Im } T$ . In our context this implies  $K_i = \dim C_i(K) = \dim Z_i(K) + \dim B_{i-1}(K)$ . Letting  $B_{-1}(K) = 0$  we get the following:

$$\begin{aligned} \chi(K) &= \sum_{i=0}^{\infty} (-1)^i K_i = \sum_{i=0}^{\infty} (-1)^i (\dim Z_i(K) + \dim B_{i-1}(K)) \\ &= \sum_{i=0}^{\infty} (-1)^i (\dim Z_i(K) - \dim B_i(K)) \\ &= \sum_{i=0}^{\infty} (-1)^i \beta_i(K). \end{aligned}$$

$\square$

**Computation** Computing  $\beta_n(K)$  comes down to elementary linear algebra: for  $i \in \{n, n+1\}$  represent  $\partial_i$  in the canonical bases of  $C_i(K)$  and  $C_{i-1}(K)$  and reduce the resulting matrix to diagonal form using elementary row and column operations. From the diagonal forms it is easy to read off  $\dim \text{Im } \partial_{n+1} = \dim B_n(K)$  as the number of non-zero diagonal terms, and  $\dim \ker \partial_n = \dim Z_n(K)$  as the number of zero columns. The following example will make it clear.

**Example 2.16.** Let  $K$  be the faces of a 2-simplex  $\{1, 2, 3\}$ . By identifying a simplex with its vertices and representing the matrices  $\partial_2$  and  $\partial_1$  in their standard bases we obtain

$$\partial_2 = \begin{matrix} & & 123 \\ & 12 & \left( \begin{array}{c} 1 \\ 1 \\ 1 \end{array} \right) \\ 13 & & \\ 23 & & \end{matrix}$$

$$\partial_1 = \begin{matrix} & 12 & 13 & 23 \\ 1 & \left( \begin{array}{ccc} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{array} \right) \\ 2 & & \sim \text{row/column operations} \sim & \left( \begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{array} \right) \\ 3 & & \end{matrix}$$

We conclude that  $\dim B_1(K) = 1$  and  $\dim Z_1(K) = 1$ , and therefore  $\beta_1(K) = 1 - 1 = 0$ .

### 2.3 A Remark on the Choice of Coefficients

The previous discussion has been for the field  $\mathbf{k} = \mathbb{Z}_2$ . Homology is well-defined for coefficients in any other field, or more generally, in any *abelian group*. For simplicial complexes in  $\mathbb{R}^3$  the Betti numbers are independent of the choice of coefficients. In higher dimensions however, the Betti numbers may depend on the underlying field. Instead of developing a full theory of simplicial homology in this more general setting, we shall limit ourselves to an example illustrating the subtleties arising.

**The boundary operator** Let  $\mathbf{k}$  be any field and let  $C_n(K; \mathbf{k})$  denote the free  $\mathbf{k}$ -vector space on the  $n$ -simplices of  $K$ . In order for the boundary operator to satisfy  $\partial_n \partial_{n+1} = 0$  the simplices need to be ordered. Intuitively, this means that we add a direction to every edge, and a left- or right-handed coordinate system to any 2-simplex, and so forth for higher dimensions. One way of obtaining such an orientation is to give a total order to the vertices of  $K$ :

$$p_0 < p_1 < \dots < p_m.$$

The boundary of the simplex  $\{p_0, p_1\}$  is then  $\partial_1(\{p_0, p_1\}) = p_1 - p_0$ , and more generally,  $\partial_n(\{p_0, \dots, p_n\}) = \sum_{i=0}^n (-1)^i \{p_0, \dots, \hat{p}_i, \dots, p_n\}$ . It is now a simple exercise to show that the boundary of the boundary vanishes. Hence, one gets cycle and boundary vector spaces satisfying  $B_n(K; \mathbf{k}) \subseteq Z_n(K; \mathbf{k})$ .

**Example 2.17.** Let  $K$  be the faces of a 2-simplex with vertex set  $\{1, 2, 3\}$  and  $\mathbf{k}$  any field. By identifying a simplex with its vertices and representing the matrices  $\partial_2$  and  $\partial_1$  in their standard bases we obtain

$$\partial_2 = \begin{matrix} & & 123 \\ & 12 & \left( \begin{array}{c} 1 \\ -1 \\ 1 \end{array} \right) \\ 13 & & \\ 23 & & \end{matrix}$$

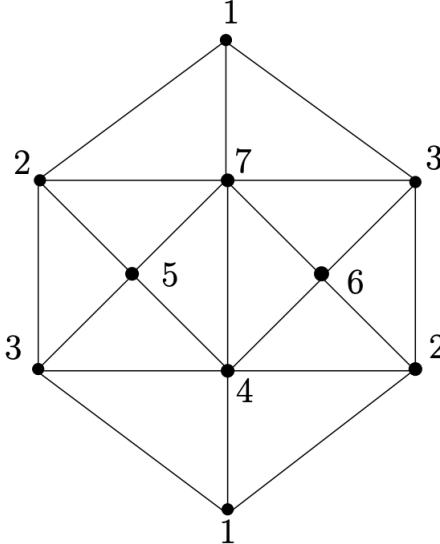


Figure 12: The projective plane,  $\mathbb{R}P^2$ .

$$\partial_1 = \begin{pmatrix} 1 & 2 & 13 & 23 \\ 2 & -1 & -1 & 0 \\ 3 & 1 & 0 & -1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \sim \text{row/column operations} \sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Hence the betti numbers are independent of the choice of coefficients.

**Example 2.18.** The *projective plane*  $\mathbb{R}P^2$  is an unoriented closed surface which does not admit an embedding in  $\mathbb{R}^3$ . A triangulation of  $\mathbb{R}P^2$  can be seen in Fig. 12. Note that for  $\mathbf{k} = \mathbb{Z}_2$ , the boundary operator of the sum of all the 2-simplices is trivial. Since there are no 3-simplices, this implies that  $\beta_2(\mathbb{R}P^2) \geq 1$  but it does not contain a "cavity" in the intuitive sense. Indeed, working with  $\mathbf{k} = \mathbb{Z}_3$  this "cavity" disappears!

From what we have learned, it is possible to write down the associated boundary matrices (preferably on a computer!) and verify that the Betti numbers differ between the choices  $\mathbf{k} = \mathbb{Z}_3$  and  $\mathbf{k} = \mathbb{Z}_2$ :

$$\begin{array}{lll} \beta_0(\mathbb{R}P^2; \mathbb{Z}_3) = 1 & \beta_1(\mathbb{R}P^2; \mathbb{Z}_3) = 0 & \beta_2(\mathbb{R}P^2; \mathbb{Z}_3) = 0 \\ \beta_0(\mathbb{R}P^2; \mathbb{Z}_2) = 1 & \beta_1(\mathbb{R}P^2; \mathbb{Z}_2) = 1 & \beta_2(\mathbb{R}P^2; \mathbb{Z}_2) = 1. \end{array}$$

This computation could be significantly simplified if we worked with *CW-complexes* rather than simplicial complexes. Note that the Euler characteristic is obviously independent of the field.

## 2.4 Simplicial Homology

The modern approach to defining the Betti numbers is by means of vector spaces. Working with vector spaces, rather than mere integers, allow us to push homological information along continuous maps. This property, called *functoriality*, is instrumental in showing invariance under topological invariance and in the development of *persistent homology*.

**Definition 2.19.** The  $n$ -th simplicial homology vector space of a simplicial complex  $K$  is the quotient space  $H_n(K) = Z_n(K)/B_n(K)$ .

It follows from the definition that  $\beta_n(K) = \dim H_n(K)$ . We say that two cycles  $z_1$  and  $z_2$  are **homologous** if  $z_1 = z_2 + b$  where  $b$  is a boundary. We write  $[z_1] = [z_2]$  for the associated equivalence class in  $H_n(K)$ . Intuitively, two homologous cycles represent the same "hole" in the sense that their sum is either trivial or cycles around boundaries.

**Lemma 2.20.** Let  $V \subseteq W$  be two finite-dimensional vector spaces (over some field). A basis  $\mathcal{B}'$  for  $V$  extends to a basis  $\mathcal{B}$  for  $W$ . Furthermore, the elements of  $\mathcal{B} \setminus \mathcal{B}'$  are representatives of a basis for  $W/V$ .

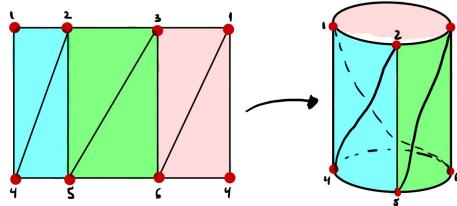
*Proof.* Left to the reader. □

The previous lemma provides us with an algorithm for computing a basis for  $H_n(K)$ : first find a basis  $\mathcal{B}'$  for  $B_n(K)$  and then extend this basis to a basis  $\mathcal{B}$  for  $Z_n(K)$ . Then the elements of  $\mathcal{B} \setminus \mathcal{B}'$  provide us with a basis for  $H_n(K)$ . Doing so by hand however can involve a significant amount of row reductions even for rather small simplicial complexes. Therefore we shall let the computer do the job!

**Example 2.21.** We return to the simplicial complex depicted in Fig. 11. The vector space  $H_0(K)$  is generated by  $[p_i]$  for any vertex  $p_i$ . Furthermore,  $H_i(K) = 0$  for any  $i > 1$ . We have already seen that  $B_1(K) = \{c_3\}$ . Hence  $\{[c_1], [c_2]\}$  is a basis for  $H_1(K) \cong \mathbb{Z}_2^2$ .

## 2.5 Exercises

1. Give an example of two non-homeomorphic triangulated spaces with Betti numbers  $\beta_0 = 1$ ,  $\beta_1 = 2$  and  $\beta_2 = 1$ .
2. Let  $G$  be a finite, connected, and piecewise-linear planar graph, i.e. a connected 1-dimensional simplicial complex geometrically realized in  $\mathbb{R}^2$ . Let  $v$  denote the number of vertices,  $e$  the number of edges and  $f$  the number of faces (regions bounded by edges, including the outer, infinitely large region).
  - (a) Show that  $v - e + f = 2$ . (Hint: start by first considering a spanning tree and then inductively add edges.) Observe from the Euler-Poincaré formula that  $\beta_1$  counts the number of unbounded faces.
  - (b) Use a) to construct a 1-dimensional simplicial complex that cannot be geometrically realized in  $\mathbb{R}^2$ .
3. Show that if  $K$  is a simplicial complex with subcomplexes  $K_1$  and  $K_2$  such that  $K_1 \cup K_2 = K$ , then  $\chi(K) = \chi(K_1) + \chi(K_2) - \chi(K_1 \cap K_2)$ . Does the same relation hold for the individual Betti numbers  $\beta_i(K)$ ?
4. Let  $K$  be the disjoint union of two simplicial complexes  $K_1$  and  $K_2$ . Show that  $H_n(K) \cong H_n(K_1) \oplus H_n(K_2)$ .
5. Construct a triangulation of the  $n$ -ball  $B^n = \{p \in \mathbb{R}^n : \|p\| \leq 1\}$  and find a basis for its homology vector spaces in all dimensions. Do the results resonate with your intuition?
6. Let  $T$  be a simplicial complex that triangulates the torus. Define  $T \# T$  to be the simplicial complex obtained by removing a 2-simplex from two copies of  $T$ , and then identifying the resulting simplicial complexes along the resulting boundary. What surface in  $\mathbb{R}^3$  does  $T \# T$  triangulate? Compute  $\chi(T \# T)$  and interpret the results geometrically.
7. Consider the following triangulation of the cylinder. Denote the associated simplicial complex by  $K$ , and show that  $H_1(K) \cong \mathbb{Z}_2$ . What are the shortest cycles (in terms of number of edges)  $z \in Z_1(K)$  for which  $0 \neq [z] \in H_1(K)$ ?



### 3 Computation, Induced Maps and Invariance

This week will discuss the "standard algorithm" for computing simplicial homology, as well as induced maps, and invariance of homology.

#### 3.1 The Standard Algorithm

Denote the  $i$ -th column of a matrix  $M$  by  $M_i$ . If  $M_i$  is non-zero, then we define  $\text{low}(M_i) = \max\{j : M_{ji} \neq 0\}$ . For a zero-column  $M_i$  we let  $\text{low}(M_i) = 0$ . We say that  $M$  is **reduced** if  $i \neq j$  implies  $\text{low}(M_i) \neq \text{low}(M_j)$  or  $0 = \text{low}(M_i) = \text{low}(M_j)$ .

The following algorithm reduces the matrix  $M$  to a reduced matrix  $R$  with  $R = MV$  where  $V$  is a upper-triangular and of full rank.

---

**Algorithm 1:** The standard algorithm.

---

**Data:**  $m_1 \times m_2$  matrix  $M$ .

**Result:** Reduced  $m_1 \times m_2$  matrix  $R$ , full-rank upper-triangular  $V$ , such that  $R = MV$ .

$R \leftarrow M$ ;

$V \leftarrow I_{m_2}$  (the  $m_2 \times m_2$  identity matrix);

**for**  $j \in \{1, 2, \dots, m_2\}$  **do**

**while** there exists  $i < j$  with  $0 \neq \text{low}(R_i) = \text{low}(R_j)$  **do**

| add column  $i$  of  $R$  to column  $j$  of  $R$ ;

| add column  $i$  of  $V$  to column  $j$  of  $V$ ;

**end**

**end**

---

Observe that the algorithm runs in  $O(m_1 m_2^2)$ .

#### 3.1.1 The Homology of a Triangle

We illustrate the algorithm on the simplicial complex given by the boundary of a triangle,  $K = \{\{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}\}$ . The first step is to stack all the boundary operators together in one big boundary matrix  $D$  and define a corresponding identity matrix  $V$ :

$$R := D = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 12 & 13 & 23 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 12 \\ 13 \\ 23 \end{matrix} & \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{matrix}, \quad V = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Observe that  $D$  is a linear transformation  $D: C_0(K) \oplus C_1(K) \rightarrow C_0(K) \oplus C_1(K)$  and that we may identify every simplex with its corresponding basis vector in  $C_0(K) \oplus C_1(K)$  in a natural way, i.e. the column vector  $(100000)^T$  corresponds to the simplex  $\{1\}$ ,  $(000100)^T$  corresponds to  $\{1, 2\}$  and  $(000111)^T$  corresponds to the linear combination  $\{1, 2\} + \{1, 3\} + \{2, 3\}$ . Likewise, we see that  $\ker D = Z_0(K) \oplus Z_1(K)$  and  $\text{Im } D = B_0(K) \oplus B_1(K)$  as subspaces of  $C_0(K) \oplus C_1(K)$ .

We will now bring  $D$  into reduced form by adding columns from *left to right*, and at the same time adding the corresponding columns in  $V$ . By iterating from the left to right, we check for every column  $j$ , if there exists a non-zero column  $i < j$  with  $\text{low}(R_i) = \text{low}(R_j)$ . By adding

column 5 to column 6, and then adding column 4 to column 6, we obtain the following two matrices

$$R := D = \begin{pmatrix} 1 & 2 & 3 & 12 & 13 & 13 + 12 + 23 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 2 & 0 & 0 & 0 & 1 & 0 \\ 3 & 0 & 0 & 0 & 0 & 1 \\ 12 & 0 & 0 & 0 & 0 & 0 \\ 13 & 0 & 0 & 0 & 0 & 0 \\ 23 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad V = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

where  $R = DV$ . The non-zero columns of a reduced matrix must necessarily be linearly independent, and therefore the non-zero columns corresponding to elements of  $C_1(K)$  will define a basis for  $B_0(K)$ . Indeed, following the notation introduced above we see that  $\{(110000)^T, (101000)^T\}$  forms a basis for  $B_0(K)$ .

By construction, the  $i$ -th column of  $V$  corresponds precisely to the linear combination of simplices whose images under  $D$  adds up to the  $i$ -th column of  $R$ . In other words, a basis for  $Z_0(K) \oplus Z_1(K)$  is given by the columns of  $V$  corresponding to 0-vectors in  $R$ . Hence,  $\{(100000)^T, (010000)^T, (001000)^T\} \cup \{(000111)^T\}$  constitutes a basis for  $Z_0(K) \oplus Z_1(K)$ . Since  $B_1(K) = 0$  it follows that  $H_1(K) \cong \mathbb{Z}_2$  with a basis  $\{[(000111)^T]\}$ . Next step is to extend the basis for  $B_0(K)$  into a basis for  $Z_0(K)$ , and then use Lemma 2.20 to obtain a basis for  $H_0(K)$ . Consider the two bases:

$$\begin{aligned} \Sigma_{B_0} &= \{(110000)^T, (101000)^T\} \\ \widetilde{\Sigma}_{Z_0} &= \{(100000)^T, (010000)^T, (001000)^T\} \end{aligned}$$

Since  $\text{low}((100000)^T) \notin \{\text{low}((11000)^T), \text{low}((101000)^T)\}$ , the set  $\Sigma_{B_0} \cup \{(100000)^T\}$  must necessarily be linearly independent. (This can be seen by constructing a matrix with the three vectors as column vectors and observing that the resulting matrix is reduced.) Furthermore, it contains the same number of basis elements as  $\widetilde{\Sigma}_{Z_0}$  which spans  $Z_0(K)$ . Hence,  $\Sigma_{Z_0} := \Sigma_{B_0} \cup \{(100000)^T\}$  forms a basis for  $Z_0(K)$  and  $H_0 \cong \mathbb{Z}_2$  has a basis  $\{[(100000)^T]\}$ .

### 3.1.2 The Algorithm

Let  $K$  be a simplicial complex of dimension  $d$ , and let

$$D: C_0(K) \oplus \dots \oplus C_d(K) \rightarrow C_0(K) \oplus \dots \oplus C_d(K)$$

denote the boundary operator represented in the standard basis given by the simplices of all dimensions.

1. Use Algorithm 1 to obtain a decomposition  $R = DV$ .
2. The columns  $R_i$  with  $\text{low}(i) \neq 0$  form a basis  $\Sigma_B = \Sigma_{B_0} \cup \dots \cup \Sigma_{B_d}$  for  $B_0(K) \oplus \dots \oplus B_d(K)$ ,
3. The columns  $V_i$  such that  $R_i = 0$  form a basis  $\widetilde{\Sigma}_Z = \widetilde{\Sigma}_{Z_0} \cup \dots \cup \widetilde{\Sigma}_{Z_d}$  for  $Z_0(K) \oplus \dots \oplus Z_d(K)$ ,
4. Let  $\Sigma_E = \widetilde{\Sigma}_Z \setminus \{V_i \in \widetilde{\Sigma}_Z : \exists R_j \text{ with } \text{low}(R_j) = \text{low}(V_i) = i\}$ .
  - (a)  $\Sigma_B \cup \Sigma_E$  is a basis for  $Z_0(K) \oplus \dots \oplus Z_d(K)$ ,
  - (b)  $\{[V_i] : V_i \in \Sigma_E\}$  is a basis for  $H_0(K) \oplus \dots \oplus H_d(K)$ .

Although the steps are justified in the example above, we briefly re-iterate the explanations.

1. -

2. The columns are linearly independent and all other columns are trivial. Hence they must span the image of the linear transformation.
3. The zero-columns of  $R$  correspond to the kernel of  $D$ . By construction, the basis elements for the kernel can be found in the corresponding columns of  $V$ .
4. (a) From the correctness of Algorithm 1,  $\text{low}(R_i) \neq \text{low}(R_j)$  for any  $R_i \neq R_j \in \Sigma_B$ , and  $\text{low}(V_i) \neq \text{low}(V_j)$  for any  $V_i \neq V_j \in \widetilde{\Sigma}_Z$ . Since  $\Sigma_B$  spans a subset of  $\widetilde{\Sigma}_Z$  in  $C_0(K) \oplus \dots \oplus C_d(K)$ , there must exist *exactly*  $|\widetilde{\Sigma}_Z| - |\Sigma_B|$  vectors  $V_i$  in  $\widetilde{\Sigma}_Z$  such that  $\text{low}(V_i) \neq \text{low}(R_j)$  for all  $R_j \in \Sigma_B$ .
- (b) By Lemma 2.20.

The total running time is  $O(N^3)$  where  $N$  is the number of simplices in  $K$ .

### 3.2 Simplicial Maps and their Induced Maps in Homology

Let  $K$  and  $L$  be abstract simplicial complexes with vertex sets  $V_K$  and  $V_L$ , respectively. A **simplicial map**  $f: K \rightarrow L$  is a function such that for every simplex  $\{p_0, \dots, p_n\}$  in  $K$  its image  $\{f(p_0), \dots, f(p_n)\}$  is a simplex in  $L$ . Associated to a simplicial map we get an induced linear map in all dimensions:

$$f_{\#}: C_n(K) \rightarrow C_n(L), \quad f_{\#}(\sigma) = \begin{cases} f(\sigma) & \text{if } \dim f(\sigma) = \dim \sigma, \\ 0 & \text{otherwise.} \end{cases}.$$

**Example 3.1.** Let  $K$  denote the 2-simplex defined on the vertices  $\{a, b, c\}$  and let  $L = K - \{a, b, c\}$ . Then  $f: K \rightarrow L$  defines a simplicial map by  $f(a) = a$ ,  $f(b) = f(c) = b$ . It follows that  $f_{\#}: C_2(K) \rightarrow C_2(L)$  is the zero-map, and that  $f_{\#}: C_1(K) \rightarrow C_1(L)$  has matrix representation

$$\begin{matrix} ab & ac & bc \\ ab & \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \end{matrix}$$

**Lemma 3.2.** We have the following commutative diagram

$$\begin{array}{ccc} C_n(K) & \xrightarrow{\partial_n} & C_{n-1}(K) \\ \downarrow f_{\#} & & \downarrow f_{\#} \\ C_n(L) & \xrightarrow{\partial_n} & C_{n-1}(L) \end{array}$$

I.e.,  $f_{\#} \circ \partial_n = \partial_n \circ f_{\#}$ .

*Proof.* It suffices to prove the result for a single simplex  $\sigma = \{p_0, \dots, p_n\}$ .

**Case 1:**  $\dim f(\sigma) = \dim \sigma$ . In this case we have that

$$f_{\#}(\partial_n(\sigma)) = \sum_{i=0}^n f_{\#}(\{p_0, \dots, \hat{p}_i, \dots, p_n\}) = \sum_{i=0}^n \{f(p_0), \dots, \widehat{f(p_i)}, \dots, f(p_n)\} = \partial_n \circ f_{\#}(\sigma).$$

**Case 2:**  $\dim f(\sigma) \leq \dim \sigma - 2$ . It is immediate that  $\partial_n \circ f_{\#}(\sigma) = 0$ . Furthermore,  $\dim\{f(p_0), \dots, \widehat{f(p_i)}, \dots, f(p_n)\} \leq \dim \sigma - 2$ , and therefore

$$\sum_{i=0}^n f_{\#}(\{p_0, \dots, \widehat{p}_i, \dots, p_n\}) = \sum_{i=0}^n 0 = 0.$$

**Case 3:**  $\dim f(\sigma) = \dim \sigma - 1$ . It is immediate that  $\partial_n \circ f_{\#}(\sigma) = 0$ . Assume without loss of generality that  $f(p_0) = f(p_1)$ . Then  $\dim\{f(p_0), \dots, \widehat{f(p_i)}, \dots, f(p_n)\} \leq \dim \sigma - 2$  for all  $2 \leq i \leq n$ . Thus,

$$\sum_{i=0}^n f_{\#}(\{p_0, \dots, \widehat{p}_i, \dots, p_n\}) = f_{\#}(\{p_1, p_2, \dots, p_n\}) + f_{\#}(\{p_0, p_2, \dots, p_n\}) \stackrel{f(p_0)=f(p_1)}{=} 0.$$

□

**Corollary 3.3.** A simplicial map  $f: K \rightarrow L$  induces a well-defined map  $f_*: H_n(K) \rightarrow H_n(L)$  defined by  $f_*([c]) = [f_{\#}(c)]$ .

*Proof.* Since  $H_n(K) = Z_n(K)/B_n(K)$  and  $H_n(L) = Z_n(L)/B_n(L)$ , it suffices to show that  $f_{\#}(Z_n(K)) \subseteq Z_n(L)$  and  $f_{\#}(B_n(K)) \subseteq B_n(L)$ .

Assume that  $\partial_n(c) = 0$  for  $c \in C_n(K)$ . Then by Lemma 3.2 we have  $\partial_n(f_{\#}(c)) = f_{\#}(\partial_n(c)) = 0$ . We conclude that  $f_{\#}(c) \in Z_n(L)$ . Now assume that  $c = \partial_{n+1}(d)$  for some  $d \in C_{n+1}(K)$ . Then  $f_{\#}(c) = f_{\#}(\partial_{n+1}(d)) = \partial_{n+1}(f_{\#}(d))$ . Hence  $f_{\#}(c) \in B_n(L)$ . □

### 3.3 Simplicial Approximations and Invariance of Homology

In order to put the homology vector spaces on a firm footing we need to prove that if  $|K|$  is homeomorphic to  $|L|$ , then  $H_n(K) \cong H_n(L)$ . A complete proof of this fact is no easy task and we will not discuss the proof in detail. The goal of this part is merely to outline how a continuous map  $h: |K| \rightarrow |L|$  induces a map  $h_*: H_n(K) \rightarrow H_n(L)$  in homology, and that this construction is *functorial*. An easy consequence will be that  $h_*$  is an isomorphism whenever  $h$  is a homeomorphism. The full story can be found in [25].

#### 3.3.1 Simplicial Maps Revisited

For geometric simplicial complexes  $K$  and  $L$  we define a **simplicial map**  $f: K \rightarrow L$  to be a function  $f: K^{(0)} \rightarrow L^{(0)}$  such that if  $\sigma$  is spanned by  $p_0, \dots, p_n$ , then the simplex spanned by  $f(p_0), \dots, f(p_n)$  is contained in  $L$ .

*Remark 3.4.* Recall that any geometric simplicial complex  $K$  can be identified with an abstract simplicial complex through its vertex scheme, i.e. by identifying every geometric simplex with the vertices that span it. A simplicial map of geometric simplicial complexes induces a simplicial map between the associated vertex schemes. Hence, a simplicial map  $f: K \rightarrow L$  between geometric simplicial complexes induces a map  $f_*: H_n(K) \rightarrow H_n(L)$  in homology.

The simplicial map  $f: K \rightarrow L$  induces a continuous (affine) map  $|f|: |K| \rightarrow |L|$  given by

$$f \left( \sum_{i=0}^n t_i p_i \right) = \sum_{i=0}^n t_i f(p_i), \quad \text{where} \quad \sum_{i=0}^n t_i = 1, \quad \text{and} \quad 0 \leq t_i \leq 1.$$

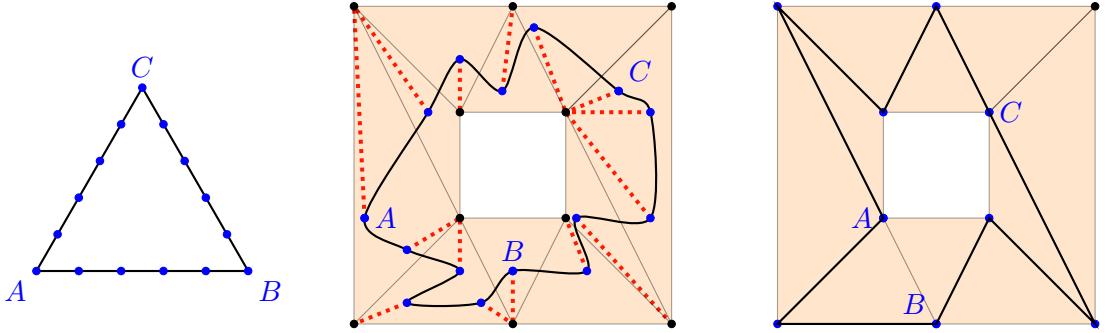


Figure 13: A continuous map and a simplicial approximation; see Fig. 13

### 3.3.2 Simplicial Approximations

We shall now see how a well-defined morphism  $h_*: H_n(K) \rightarrow H_n(L)$  can be associated to any continuous map  $h: |K| \rightarrow |L|$ .

**Definition 3.5.** Let  $K$  and  $L$  be geometric simplicial complexes and  $h: |K| \rightarrow |L|$ . We say that  $f: K \rightarrow L$  is a **simplicial approximation to  $h$** , if for each  $x \in |K|$ ,  $|f|(x)$  is contained in the minimal simplex which contains  $h(x)$ .

The following theorem makes it clear why simplicial approximations are relevant.

**Theorem 3.6.** If  $f_1, f_2: K \rightarrow L$  are simplicial approximations to  $h: |K| \rightarrow |L|$ , then  $f_{1,*} = f_{2,*}: H_n(K) \rightarrow H_n(L)$ .

In particular, for any continuous map  $h$  that admits a simplicial approximation  $f$ , we have a well-defined map  $f_*: H_n(K) \rightarrow H_n(L)$  in homology. However, not all maps admit such approximations.

**Example 3.7.** Let  $K$  be the triangulation of the triangle shown in Fig. 13, and let  $L$  be the triangulation of a filled rectangle with a hole in the same figure. The curve in  $|L|$  shows the image of  $|K|$  under a continuous mapping  $h$ . We see that  $h$  admits a simplicial approximation  $f: K \rightarrow L$  which is obtained by pushing the images of the vertices of  $K$  onto the corresponding vertices of  $L$  as indicated by arrows. Note that if  $|K|$  had a less refined triangulation given by  $\{\{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}\}$  then  $h$  would not admit a simplicial approximation.

The previous example shows that for a given  $K$  and  $L$ , and a function  $h: |K| \rightarrow |L|$ , there need not exist a simplicial approximation  $f: K \rightarrow L$  to  $h$ . However, it seems plausible that one can always "subdivide" the simplices of  $K$  to be so "small" that we can push the image of its vertices onto the vertices of  $L$  in a way that gives a well-defined simplicial map.

**Definition 3.8.** A simplicial complex  $K'$  is a **subdivision** of a simplicial complex  $K$  if  $|K'| = |K|$  and each simplex of  $K'$  is contained in a simplex of  $K$ .

Subdivisions are particularly useful because of the following non-trivial result.

**Theorem 3.9.** Let  $K'$  be a subdivision of  $K$ , and let  $g: K' \rightarrow K$  be a simplicial approximation to the identity on  $|K|$ . Then  $g_*: H_n(K') \rightarrow H_n(K)$  is an isomorphism.

One way to subdivide a simplicial complex  $K$  is through **barycentric subdivision**. Let  $\sigma \in K$  be a simplex in  $\mathbb{R}^k$  spanned by vertices  $\{p_0, \dots, p_n\}$ . The **barycenter** of  $\sigma$  is

$$\hat{\sigma} = \frac{1}{n+1}(p_0 + \dots + p_n).$$

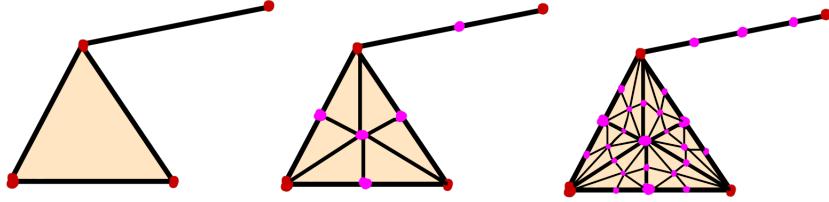


Figure 14: A simplicial complex together with its first two barycentric subdivisions.

The **first barycentric subdivision of  $K$**  is the simplicial complex  $\text{Sd}(K)$  consisting of all simplices spanned by vertices  $\hat{\sigma}_0, \dots, \hat{\sigma}_r$  where  $\sigma_0 \subset \sigma_1 \dots \subset \sigma_r$  is a sequence of different simplices in  $K$ . In particular, the vertex set of  $\text{Sd}(K)$  is  $\{\hat{\sigma} : \sigma \in K\}$ .

*Remark 3.10.*  $\text{Sd}(K)$  is a simplicial complex and  $|\text{Sd}(K)| = |K|$ .

For  $j > 1$  we define the  $j$ -th barycentric subdivision  $\text{Sd}^j(K)$  to be the barycentric subdivision of  $\text{Sd}^{j-1}(K)$ .

**Example 3.11.** Fig. 14 shows a simplicial complex  $K$  together with its first two barycentric subdivisions  $\text{Sd}(K)$  and  $\text{Sd}^2(K)$ . Observe that any function  $f: \text{Sd}(K)^{(0)} \rightarrow K^{(0)}$  which maps a vertex  $\hat{\sigma}$  to any vertex of  $\sigma$  defines a simplicial map.

We state the following important theorem due to Alexander without a proof.

**Theorem 3.12** (Simplicial Approximation Theorem). *Let  $K$  and  $L$  be simplicial complexes and  $h: |K| \rightarrow |L|$  a continuous map. Then  $h$  admits a simplicial approximation  $f: \text{Sd}^j(K) \rightarrow L$  for some sufficiently large  $j$ .*

In conjunction with Theorem 3.6 this shows that  $h$  induces a map  $f_*: H_n(\text{Sd}^j(K)) \rightarrow H_n(L)$  for some sufficiently large  $j$ . To complete our construction of an induced map in homology, we need the following lemma whose proof is left as an exercise.

**Lemma 3.13.** *Let  $f: \text{Sd}(K)^{(0)} \rightarrow K^{(0)}$  denote any map which sends a vertex  $\hat{\sigma}$  in  $\text{Sd}(K)$  to any vertex of  $\sigma$ . Then  $f$  is a simplicial map and an approximation to the identity  $\text{id}_{|K|}: |K| \rightarrow |K|$ .*

We are now ready to associate a map in homology to a continuous map  $h: |K| \rightarrow |L|$ . Choose a subdivision  $K'$  of  $K$  and simplicial maps  $g: K' \rightarrow K$  and  $f: K' \rightarrow L$  that approximate  $\text{id}_{|K|}$  and  $h$ , respectively. Such  $f, g$  and  $K'$  exist by Theorem 3.12 and repeated application of Lemma 3.13. This yields the following zigzag of vector spaces and linear maps:

$$H_n(K) \xleftarrow{g_*} H_n(K') \xrightarrow{f_*} H_n(L)$$

From Theorem 3.9,  $g_*: H_n(K') \rightarrow H_n(K)$  is an isomorphism, and we thus obtain a map in homology:

$$h_* := f_* \circ (g_*)^{-1}: H_n(K) \rightarrow H_n(L). \quad (2)$$

**Theorem 3.14.** *Let  $h_*$  be as in Eq. (2).*

- *The map  $h_*$  is well-defined, i.e. it does not depend of the choice of  $K'$ ,  $f$  and  $g$ .*
- *The construction is functorial, i.e.  $(\text{id}_{|K|})_* = \text{id}_{H_n(K)}$  and  $(h' \circ h)_* = h'_* \circ h_*$  for any pair of continuous maps  $h: |K| \rightarrow |L|$  and  $h': |L| \rightarrow |M|$ .*

We end this section by observing that homology is invariant under homeomorphism. In particular, the homology of a topological space is independent of the triangulation.

**Corollary 3.15.** *If  $h: |K| \rightarrow |L|$  is a homeomorphism, then  $h_*$  is an isomorphism.*

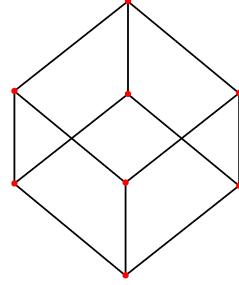
*Proof.* If  $h: |K| \rightarrow |L|$  is a homeomorphism, then

$$\begin{aligned} \text{id}_{H_n(K)} &= (\text{id}_{|K|})_* = (h^{-1} \circ h)_* = (h^{-1})_* \circ h_*: H_n(K) \rightarrow H_n(K), \\ \text{id}_{H_n(L)} &= (\text{id}_{|L|})_* = (h \circ h^{-1})_* = h_* \circ (h^{-1})_*: H_n(L) \rightarrow H_n(L). \end{aligned}$$

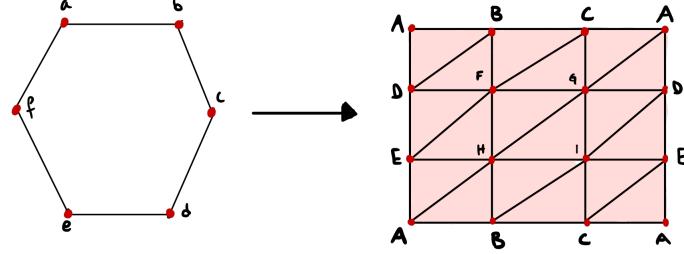
Hence,  $f_*$  is an isomorphism. □

### 3.4 Exercises

1. Compute  $\beta_1$  for the cube below. How does  $\beta_1$  correspond to the number of "holes"?



2. Show that  $Sd(K)$  is a simplicial complex.
3. Prove Lemma 3.13.
4. Consider the triangulation  $K$  of the circle and the triangulation  $L$  of the torus shown below.



A basis for  $H_1(L)$  is given by  $\{[w], [z]\}$  where  $w = \{A, D\} + \{D, E\} + \{A, E\}$  and  $z = \{A, B\} + \{B, C\} + \{C, A\}$ . Let  $[\gamma]$  be the generator of  $H_1(K)$ . Represent  $g_*([\gamma])$ ,  $h_*([\gamma])$  and  $s_*([\gamma])$  in the basis  $\{[w], [z]\}$ , where  $g, h, s: K \rightarrow L$  are simplicial given by

$$\begin{aligned} g(a) &= A, g(b) = B, g(c) = F, g(d) = E, g(e) = H, g(f) = B \\ h(a) &= A, h(b) = B, h(c) = C, h(d) = A, h(e) = E, h(f) = D \\ s(a) &= A, s(b) = G, s(c) = F, s(d) = H, s(e) = H, s(f) = A \end{aligned}$$

5. Given a finite set of points  $P$  in  $\mathbb{R}^n$ , one common approach to analyzing its "topological structure" is by means of the associated *Betti curve in dimension i*. The Betti curve in dimension 1 is the function  $b_1: \mathbb{R}_{\geq 0} \rightarrow \{0, 1, 2, \dots\}$  given by  $b_1(r) = \beta_1(\text{Cech}_r(P))$ <sup>thm. not yet proved</sup>  $= \beta_1(\cup_{p \in P} B_r(p))$ . This construction does not capture the "size" of the "holes" in the data and can be very sensitive to perturbations. Construct examples illustrating this and try to think of ways this can be rectified.

## 4 Homotopy Equivalence, Relative Homology and Sensor Networks

### 4.1 Homotopy Equivalence

In the last lecture we outlined how a continuous map  $h: |K| \rightarrow |L|$  between polyhedra induces a linear map in homology, and that this map is an isomorphism if  $h$  is a homeomorphism. We shall now discuss a weaker condition on  $h$  for which the induced map is an isomorphism. All functions are assumed to be continuous.

**Definition 4.1.** A **homotopy** between two maps  $f, g: X \rightarrow Y$  is a map  $F: X \times [0, 1] \rightarrow Y$ , such that  $F(x, 0) = f(x)$  and  $F(x, 1) = g(x)$  for all  $x \in X$ . ( $X \times [0, 1]$  is equipped with the product topology.)

If there is a homotopy between  $f$  and  $g$  then we say that  $f$  and  $g$  are **homotopic** and denote this by  $f \simeq g$ .

*Remark 4.2.* Homotopy defines an equivalence relation.

**Example 4.3.** Fig. 15 shows an example of a homotopy between two functions  $f, g: [0, 1] \rightarrow \mathbb{R}^2$ .

**Example 4.4.** The two functions  $f, g: [0, 1] \rightarrow \{0, 1\}$  defined by  $f(x) = 0$  and  $g(x) = 1$  for all  $x$  are not homotopic.

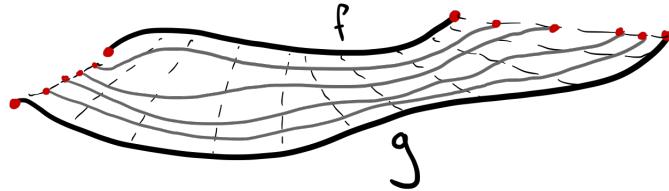


Figure 15: A homotopy between two functions  $f$  and  $g$ .

We will however need a stricter version of homotopy. Let  $A \subseteq X$  and  $B \subseteq Y$ . A **map of pairs**  $f: (X, A) \rightarrow (Y, B)$  is a map  $f: X \rightarrow Y$  such that  $f(A) \subseteq B$ . A **homotopy of pairs** between  $f, g: (X, A) \rightarrow (Y, B)$  is a homotopy  $F: X \times [0, 1] \rightarrow Y$  between  $f$  and  $g$  such that  $F_t(A) \subseteq B$  for all  $t \in [0, 1]$ .

**Example 4.5.** Let  $X = [0, 1]$ ,  $A = \{0, 1\}$ ,  $Y = \mathbb{R}^2 - \{0\}$  and  $B = \{p\}$  where  $0 \neq p \in \mathbb{R}^2$ . The equivalence classes of homotopic maps constitute the fundamental group of  $\mathbb{R}^2 - \{0\}$  (w.r.t point  $p$ ); see Fig. 16a.

**Definition 4.6.** The spaces  $(X, A)$  and  $(Y, B)$  are **homotopy equivalent** if there exists maps  $f: (X, A) \rightarrow (Y, B)$  and  $g: (Y, B) \rightarrow (X, A)$  such that  $g \circ f \simeq \text{id}_{(X, A)}$  and  $f \circ g \simeq \text{id}_{(Y, B)}$ .

We write  $(X, A) \simeq (Y, B)$  if  $(X, A)$  and  $(Y, B)$  are homotopy equivalent. If  $i: (Y, B) \hookrightarrow (X, A)$  denotes an inclusion of a subspace, and  $r: (X, A) \rightarrow (Y, B)$  is such that  $r \circ i = \text{id}_{(Y, B)}$  and  $i \circ r \simeq \text{id}_{(X, A)}$ , then  $(Y, B)$  is a **deformation retract** of  $(X, A)$ . If  $(X, A) \simeq \{pt\}$  then  $(X, A)$  is **contractible**; see Fig. 16b.

Our interest in homotopic maps stems from the following theorem.

**Theorem 4.7.** If  $f \simeq g: |K| \rightarrow |L|$  then  $f_* = g_*: H_n(K) \rightarrow H_n(L)$ .

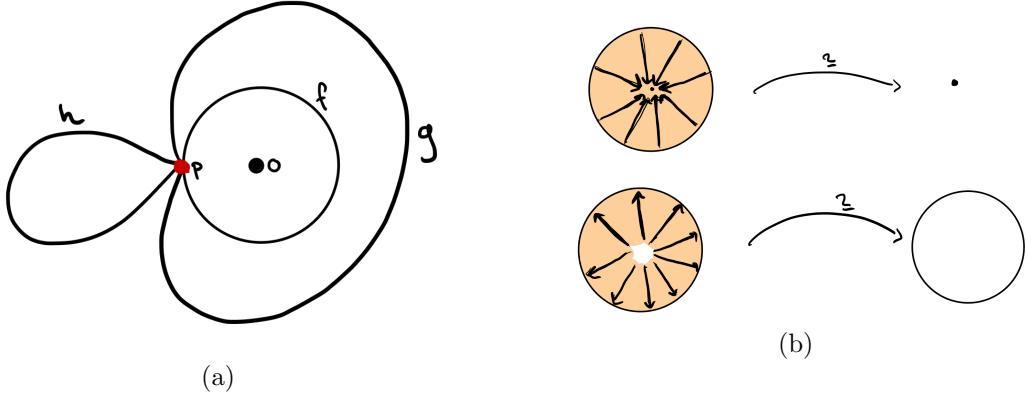


Figure 16: a) The functions  $f$  and  $g$  are homotopic, but neither of them is homotopic to  $h$ . (b) The disk deformation retracts down onto its origin and is thus contractible. The disk with a hole in the middle deformation retracts onto the circle (and is therefore not contractible).

**Corollary 4.8.** *If  $f: |K| \rightarrow |L|$  is a homotopy equivalence of polyhedra, then is  $f_*: H_n(K) \rightarrow H_n(L)$  an isomorphism.*

*Proof.* Let  $g$  be such that  $g \circ f \simeq \text{id}_{|K|}$  and  $f \circ g \simeq \text{id}_{|L|}$ . Since homology is functorial:

$$\begin{aligned} \text{id}_{H_n(K)} &= (\text{id}_{|K|})_* = (g \circ f)_* = g_* \circ f_*: H_n(K) \rightarrow H_n(K), \\ \text{id}_{H_n(L)} &= (\text{id}_{|L|})_* = (f \circ g)_* = f_* \circ g_*: H_n(L) \rightarrow H_n(L). \end{aligned}$$

Hence,  $h_*$  is an isomorphism. □

*Remark 4.9.* While it is possible to show that simplicial homology defines a *homology theory* for triangulated spaces (see e.g. [25, Section 27] for a detailed discussion), there is a more natural construction that applies to any topological space. This homology theory is called *singular homology*. With a bit of work, one can show that singular and simplicial homology coincide when restricted to polyhedra. That is, whenever  $X \simeq |K|$ , we can compute its singular homology  $H_n^S(X)$  by means of the simplicial homology  $H_n(K)$ . Furthermore, if  $f: X \rightarrow Y$  with  $|K| \simeq X$  and  $|L| \simeq Y$  then there is the following commutative diagram where  $\hat{f}$  is the map given by the composition  $|K| \rightarrow X \xrightarrow{f} Y \rightarrow |L|$ ,

$$\begin{array}{ccc} H_n^S(X) & \xrightarrow{f_*} & H_n^S(Y) \\ \downarrow \cong & & \downarrow \cong \\ H_n(K) & \xrightarrow{\hat{f}_*} & H_n(L). \end{array}$$

In conclusion: from time to time we need to consider the homology of a topological space  $X$  which is not a polyhedron and we will simply denote its homology by  $H_n(X)$  where this is to be understood as the simplicial homology of any polyhedron homotopy equivalent to  $X$ . We will *not* consider spaces that are not homotopy equivalent to a simplicial complex.

## 4.2 The Nerve Theorem

A particular instance of a homotopy equivalence is the *nerve theorem*.

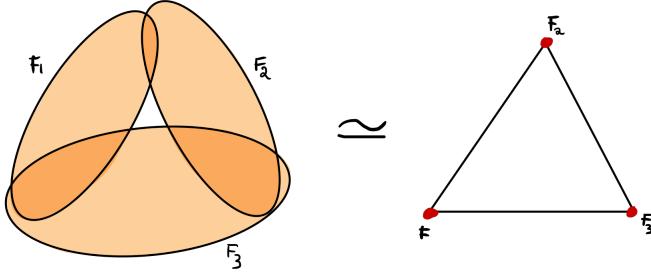


Figure 17: The Nerve (right) associated to the cover to the left.

Let  $F$  be a finite collection of closed, convex sets in Euclidean space. Define the **nerve of  $F$**  to be the abstract simplicial complex

$$\mathcal{N}(F) := \{G \subseteq F : \bigcap_{F_i \in G} F_i \neq \emptyset\}.$$

**Example 4.10.** Consider the set  $F = \{F_1, F_2, F_3\}$  in Fig. 17. Then

$$\mathcal{N}(F) = \{\{F_1\}, \{F_2\}, \{F_3\}, \{F_1, F_2\}, \{F_2, F_3\}, \{F_1, F_3\}\}$$

which is (isomorphic to) the abstract simplicial complex associated to the boundary of a 2-simplex. Clearly,  $F_1 \cup F_2 \cup F_3 \simeq |\mathcal{N}(F)|$ .

**Theorem 4.11** (The (Persistent) Nerve Theorem).

1. Let  $F = \{F_1, \dots, F_m\}$  denote a collection of closed and convex sets in Euclidean space. Then,

$$\bigcup_{F_i \in F} F_i \simeq |\mathcal{N}(F)|.$$

2. Let  $F$  be as above and let  $F' = \{F'_1, \dots, F'_m\}$  be a collection of closed and convex sets such that  $F_i \subseteq F'_i$  for all  $i$ . Denote by  $j$  the inclusion  $\bigcup_{F_i \in F} F_i \hookrightarrow \bigcup_{F'_i \in F'} F'_i$ , and let  $\sigma: \mathcal{N}(F) \rightarrow \mathcal{N}(F')$  be the simplicial map that sends  $F_i$  to  $F'_i$ . The following diagram commutes:

$$\begin{array}{ccc} H_n(\bigcup_{F_i \in F} F_i) & \xrightarrow{j_*} & H_n(\bigcup_{F'_i \in F'} F'_i) \\ \downarrow \cong & & \downarrow \cong \\ H_n(\mathcal{N}(F)) & \xrightarrow{\sigma_*} & H_n(\mathcal{N}(F')) \end{array}$$

*Proof.* The idea of the proof is the following. Consider the subspace  $\Delta F$  of  $|\mathcal{N}(F)| \times \bigcup_{F_i \in F} F_i$  defined by the disjoint union

$$\Delta F = \bigsqcup_{S \in \mathcal{N}(F)} |S| \times \bigcap_{F_i \in S} F_i,$$

under the equivalence relation  $(s, x) \sim (t, x)$  if  $s \in |S|$ ,  $t \in |T|$ ,  $S \subseteq T$ , and  $s = t$ . We have continuous projection maps  $\pi_1: \Delta F \rightarrow |\mathcal{N}(F)|$  and  $\pi_2: \Delta F \rightarrow \bigcup_i F_i$ , given by projecting onto the first and second component, respectively. This is illustrated in Fig. 18. With some work,

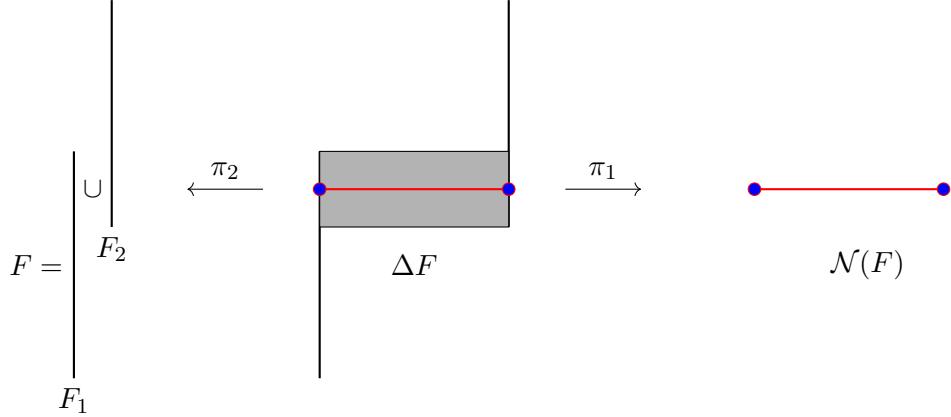


Figure 18: The space  $\Delta F$  in the proof of Theorem 4.11 for two overlapping segments in  $\mathbb{R}$ .

one can show that both of these projections are homotopy equivalences; see for instance [20] for a proof<sup>4</sup>.

This gives the following commutative diagram

$$\begin{array}{ccc}
 \bigcup_i F_i & \xhookrightarrow{j} & \bigcup_i F'_i \\
 \pi_2 \uparrow \simeq & & \pi_2 \uparrow \simeq \\
 \Delta F & \xhookrightarrow{\quad} & \Delta F' \\
 \simeq \downarrow \pi_1 & & \simeq \downarrow \pi_1 \\
 \mathcal{N}(F) & \xrightarrow{\sigma} & \mathcal{N}(F')
 \end{array}$$

Passing to homology, the projection maps become isomorphisms and we obtain the commutative diagram in the statement.  $\square$

#### 4.2.1 The Čech Complex

Let  $P \subset \mathbb{R}^n$  be a finite set of points. By definition of the Čech complex we have that  $\text{Čech}_r(P) = \mathcal{N}(F)$  where  $F = \{B_r(p) : p \in P\}$ , and it follows from Theorem 4.11 that the geometric realization of the Čech complex is homotopy equivalent to the union of closed  $r$ -balls. This shows that the Betti numbers associated to the Čech complex can be interpreted as "holes" of various dimension formed by the associated union of balls in Euclidean space; see Fig. 19. We leave it as an exercise to prove that the Alpha complex is homotopy equivalent to the Čech complex.

### 4.3 Relative Homology

Let  $K_0 \subseteq K$  be a subcomplex. The vector space of **relative  $n$ -chains** is the quotient vector space  $C_n(K, K_0) = C_n(K)/C_n(K_0)$ . Since  $\partial_n(C_n(K_0)) \subseteq C_{n-1}(K_0)$  by construction, it follows that  $\partial$  descends to a boundary map

$$\partial_n: C_n(K, K_0) \rightarrow C_{n-1}(K, K_0)$$

---

<sup>4</sup>This proof is given for an open cover, but we can reduce our problem to this by replacing the closed convex sets with slightly larger open sets such that the nerve remains the same.

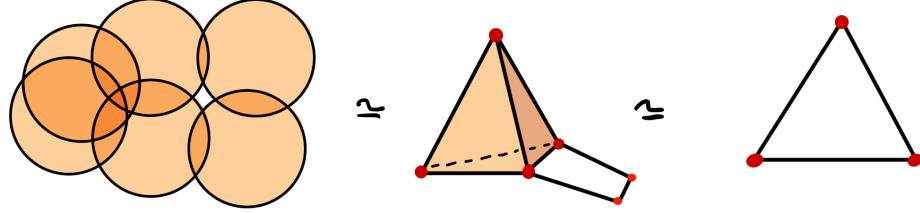


Figure 19: The union of balls is homotopy equivalent to the associated nerve (middle).

such that  $\partial_n \circ \partial_{n+1} = 0$ . Hence, we get associated **relative homology groups** given by

$$H_n(K, K_0) = \frac{Z_n(K, K_0)}{B_n(K, K_0)} := \frac{\ker \partial_n: C_n(K, K_0) \rightarrow C_{n-1}(K, K_0)}{\text{Im } \partial_{n+1}: C_{n+1}(K, K_0) \rightarrow C_n(K, K_0)}.$$

**Example 4.12.** Let  $K$  denote the simplicial complex given by a 2-simplex and all its faces, and let  $K_0$  denote its 1-skeleton. It follows that:

$$\begin{array}{ll} \dim C_2(K) = 1 & \dim C_2(K_0) = 0 \\ \dim C_1(K) = 3 & \dim C_1(K_0) = 3 \\ \dim C_0(K) = 3 & \dim C_0(K_0) = 3 \end{array}$$

We conclude that  $H_2(K, K_0) \cong \mathbb{Z}_2$ ,  $H_1(K, K_0) = H_0(K, K_0) = 0$ .

The previous example generalizes to any  $n$ -simplex: let  $K$  be denote a  $n$ -simplex with its faces, and  $K_0$  its  $(n-1)$ -skeleton. Then  $H_n(K, K_0) \cong \mathbb{Z}_2$  and all other homology vector spaces are trivial. The following theorem explains why.

**Theorem 4.13.** *Let  $K_0 \subseteq K$ . Then,  $H_n(K, K_0) \cong H_n(|K|/|K_0|)$  for all  $n \geq 1$ .*

Here  $|K|/|K_0|$  need not be a simplicial complex; see Remark 4.9.

**Example 4.14.** Returning to the previous example we see that  $|K|/|K_0| \simeq S^2 \simeq \text{Bd}(\Delta^3)$  (the boundary of the standard 3-simplex). Thus,  $H_2(K, K_0) \cong H_2(|K|/|K_0|) \cong H_2(\text{Bd}(\Delta^3)) \cong \mathbb{Z}_2$ .

Next we observe that if  $f: (K, K_0) \rightarrow (L, L_0)$  is a (geometric) simplicial map of pairs, then  $f_\#(C_n(K_0)) \subseteq C_n(L_0)$ . Therefore the induced map  $f_\#: C_n(K, K_0) \rightarrow C_n(L, L_0)$  defined by  $f_\#([c]) = [f(c)]$  is well-defined. Furthermore, it follows from Lemma 3.2 that

$$\partial_n \circ f_\#([c]) = [\partial_n \circ f_\#(c)] = [f_\# \circ \partial_n(c)] = f_\# \circ \partial_n([c]).$$

Just as in the proof of Corollary 3.3 this gives an induced map in relative homology

$$f_*: H_n(K, K_0) \rightarrow H_p(L, L_0).$$

**Theorem 4.15.** *If  $f, g: (|K|, |K_0|) \rightarrow (|L|, |L_0|)$  are homotopic, then  $f_* = g_*: H_n(K, K_0) \rightarrow H_n(L, L_0)$ .*

**Corollary 4.16.** *If  $f: (|K|, |K_0|) \rightarrow (|L|, |L_0|)$  is a homotopy equivalence, then  $f_*$  an isomorphism.*

*Proof.* See the proof of Corollary 4.8. □

**Example 4.17.** Let  $S^1 \subseteq D^2$  denote the unit circle and disk, respectively. Then  $(D^2 - \{0\}, S^1) \simeq (S^1, S^1)$  and therefore  $H_2(D^2 - \{0\}, S^1) \cong 0$ . On the other hand,  $(D^2, S^1) \simeq (\Delta^2, \text{Bd}(\Delta^2))$  and therefore

$$H_2(D, S^1) \cong H_2(\Delta^2 / \text{Bd}(\Delta^2)) \cong H_2(S^2) \cong H_2(\text{Bd}(\Delta^3)) \cong \mathbb{Z}_2.$$

The following linear transformation will be important in the next section.

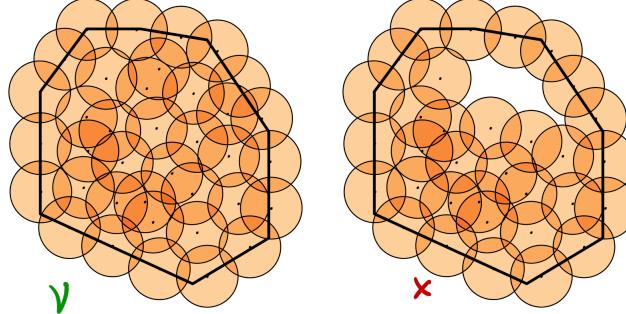
**Theorem 4.18.** Let  $K_0 \subseteq K$ . The linear map  $\delta_*: H_n(K, K_0) \rightarrow H_{n-1}(K_0)$  given by  $\delta_*([c]) = [\partial_n(c)]$  is well-defined, and if  $f: (K, K_0) \rightarrow (L, L_0)$  is a simplicial map of pairs, then the following diagram commutes:

$$\begin{array}{ccc} H_n(K, K_0) & \xrightarrow{\delta_*} & H_{n-1}(K_0) \\ \downarrow f_* & & \downarrow f_* \\ H_n(L, L_0) & \xrightarrow{\delta_*} & H_{n-1}(L_0). \end{array}$$

*Proof.* Left as an exercise. □

## 4.4 Sensor Networks

We now show how relative homology can be used to determine if a region is covered by a finite set of simple sensors. We will prove [13, Theorem 1]. For a beautiful overview of homological sensor networks see [14].



### 4.4.1 The Covering Problem

The problem is as follows: assume that you are given a collection of nodes  $X$  (i.e. a finite point set) contained in a domain  $D \subset \mathbb{R}^2$ , where each node can sense, broadcast, and cover a region of a given radius. Does the collection of coverage disks of  $X$  cover the domain? In the illustration above we see that the answer is "yes" in the left case and "no" in the right case. Assuming the location of every sensor is known, then this is a trivial task. But we will work under the assumption that we do not know the specific location of the sensors - merely which sensors are close to each other.

We will work under four assumptions:

1. The nodes broadcast their unique ID numbers. Each node can detect the identity of any node within broadcast radius  $r_b$ .
2. The nodes have radially symmetric covering domains of cover radius  $r_c \geq r_b/\sqrt{3}$ .

3. The nodes  $X$  lie in a compact connected domain  $D \subset \mathbb{R}^2$  whose boundary  $\partial D$  is connected and piecewise-linear with vertices marked fence nodes  $X_f$ .
4. Each fence node  $x \in X_f$  knows the identities of its neighbors on  $\partial D$  and these neighbors both lie within distance  $r_b$  of  $x$ .

For simplicity we use the notation  $R := \text{VR}_{r_b/2}(X)$  and we let  $F = \text{VR}_{r_b/2}(X_f)$  denote the 1-dimensional cycle associated to the boundary nodes, i.e.  $\partial D$  is a geometric realization of  $F$ . Lastly, recall from Theorem 4.18 that there is a well-defined morphism  $\delta_*: H_2(R, F) \rightarrow H_1(F)$ .

**Theorem 4.19.** *If there exists  $[\alpha] \in H_2(R, F)$  such that  $0 \neq \delta_*([\alpha]) = [\delta_2(\alpha)] \in H_1(F)$ , then the union  $U = \bigcup_{x \in X} B_{r_c}(x)$  covers  $D$ .*

*Proof.* Let  $z: |R| \rightarrow \mathbb{R}^2$  be the affine map which sends a vertex  $v_i$  of  $|R|$  to its corresponding node in the plane, and which is extended linearly to all higher-dimensional simplices:  $z(\sum_{i=0}^n t_i v_i) = \sum_{i=0}^n t_i z(v_i)$ . First we observe that  $z(|R|) \subseteq U$ . Indeed, let  $Q = \{q_1, \dots, q_n\}$  be a set of nodes which form a simplex  $\tau$  in  $R$ , i.e.,  $\text{diam}(Q) \leq r_b$ . Then, by Theorem 1.14, the intersection  $\bigcap_{x \in Q} B_{r_b/\sqrt{3}}(x)$  is non-empty. This implies that any point in  $z(\tau)$  is at distance at most  $r_b/\sqrt{3}$  from a point in  $Q$ . To see this, let  $p = t_0 q_0 + \dots + t_n q_n$  for  $\sum_i t_i = 1$  and  $t_i \geq 0$ , and let  $\tilde{c} = c - p$  where  $c$  is the center of the minimal enclosing ball of  $Q$ . For  $\tilde{q}_i = q_i - p$ , we have  $0 = t_0 \tilde{q}_0 + \dots + t_n \tilde{q}_n$ , and

$$0 = t_0 \tilde{q}_0 \cdot \tilde{c} + \dots + t_n \tilde{q}_n \cdot \tilde{c}.$$

In particular, for some  $i$ ,  $\tilde{q}_i \cdot \tilde{c} < 0$ . In that case,

$$(r_b/\sqrt{3})^2 = \|q_i - c\|^2 = \|\tilde{q}_i - \tilde{c}\|^2 = \|\tilde{q}_i\|^2 - 2\tilde{q}_i \cdot \tilde{c} + \|\tilde{c}\|^2 \geq \|\tilde{q}_i\|^2 = \|q_i - p\|^2.$$

From here we get

$$z(\tau) \subseteq \bigcup_{x \in Q} B_{r_b/\sqrt{3}}(x) \subseteq \bigcup_{x \in Q} B_{r_c}(x) \subseteq U.$$

Since this holds for every simplex  $\tau$ , it follows that  $z(|R|) \subseteq U$ .

From Theorem 4.18 we have the following commutative diagram:

$$\begin{array}{ccc} H_2(R, F) & \xrightarrow{\delta_*} & H_1(F) \\ \downarrow z_* & & \downarrow z_* \cong \\ H_2(\mathbb{R}^2, \partial D) & \xrightarrow{\delta_*} & H_1(\partial D). \end{array}$$

In particular,  $z_*([\alpha]) = z_* \circ \delta_*([\alpha]) \neq 0$ . Assume for the sake of contradiction that there is a  $p \in D$  which is not contained in  $U$ . Then, the map  $z$  factors as

$$\begin{array}{ccc} (|R|, |F|) & \xrightarrow{z} & (\mathbb{R}^2, \partial D) \\ \downarrow z & & \uparrow \\ (U, \partial D) & \longrightarrow & (\mathbb{R}^2 \setminus \{p\}, \partial D) \end{array}$$

and applying homology yields the following commutative diagram of vector spaces,

$$\begin{array}{ccc} H_2(R, F) & \xrightarrow{z_*} & H_2(\mathbb{R}^2, \partial D) \\ \downarrow z_* & & \uparrow \\ H_2(U, \partial D) & \longrightarrow & H_2(\mathbb{R}^2 \setminus \{p\}, \partial D) \end{array} .$$

Observe that  $(\mathbb{R}^2 \setminus \{p\}, \partial D) \simeq (D \setminus \{p\}, \partial D) \simeq (\partial D, \partial D)$  and thus

$$H_2(D \setminus \{p\}, \partial D) \cong H_2(\partial D, \partial D) = 0.$$

This contradicts that  $z_*([\alpha]) \neq 0$ .  $\square$

**Corollary 4.20.** *Let  $\alpha$  be as in Theorem 4.19. The restriction of  $U$  to those nodes which make up  $\alpha$  covers  $D$ .*

## 4.5 Exercises

1. Construct two topological spaces which have the same Betti numbers but which are not homotopy equivalent.
2. Let  $K$  consist of a  $p$ -simplex and its faces. How many  $p$ -simplices belong to the barycentric subdivision of  $K$ ?
3. There is a stronger version of the Nerve theorem that allows one to replace the condition that the sets are convex with the following: all possible intersections of the sets  $F_i$ 's are contractible. This is clearly satisfied if the sets  $F_i$  are convex.

Let  $P$  be a finite collection of paths that cover a tree. That is, each node and each edge of the tree belongs to at least one path. Prove that the nerve of  $P$  is contractible. Is the result still true if we allow subtrees in the collection? What about subforests?

4. Let  $v$  be a vertex of  $K$ . Show that  $H_n(K) \cong H_n(K, v)$  for all  $n \geq 1$  without using the theorem that says  $H_n(K, v) \cong H_n(|K|/|v|)$ .
5. Verify that the linear map of Theorem 4.18 is well-defined.
6. The condition for sensor coverage is an  $[\alpha] \in H_2(R, F)$  such that  $0 \neq [\partial\alpha] \in H_1(F)$ . What does this condition mean intuitively? Furthermore, give an example showing that the condition  $[\partial\alpha] \neq 0$  is necessary.
7. Let  $r \geq 0$  and  $P \subset \mathbb{R}^n$  a finite set. Show that  $|\text{Alpha}_r(P)| \simeq |\text{Cech}_r(P)|$ .

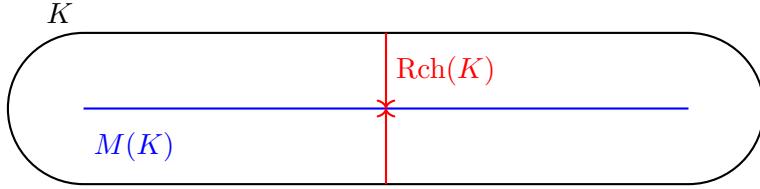


Figure 20: The medial axis and reach.

## 5 Topology & Homology Inference and Persistent Betti Numbers

### 5.1 Topology & Homology Inference

Given a compact subset  $K \subseteq \mathbb{R}^d$  and a set of points  $P$  which lie "close" to  $K$ , is it possible to infer the topology of  $K$  from  $P$  assuming the sample is dense enough and with little noise? We shall state one theorem which answers that question to the affirmative under quite strong assumptions on  $K$ .

For  $K$  and  $P$  as above and  $x \in \mathbb{R}^d$ , define

$$d_K(x) = \min_{y \in K} \|x - y\|, \quad d_P(x) = \min_{y \in P} \|x - y\|$$

where the norm is the standard 2-norm. For a given  $x \in \mathbb{R}^d$  we define

$$\Pi_K(x) = \{y \in K : d_K(x) = \|x - y\|\}.$$

Note that  $|\Pi_K(x)| \geq 1$  for all  $x$ .

**Definition 5.1.**

- The **medial axis of  $K$**  is the set  $M(K) := \{x : |\Pi_K(x)| \geq 2\}$ .
- The **reach** of  $K$  is given by

$$\text{Rch}(K) := \min_{y \in K} d(y, \overline{M(K)})$$

where  $\overline{M(K)}$  denotes the topological closure of  $M(K)$ .

See Fig. 20 for an illustration.

**Definition 5.2.** The **Hausdorff distance** between two non-empty subsets  $X, Y \subseteq \mathbb{R}^d$  is defined by

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

Associated to the functions  $d_K, d_P : \mathbb{R}^d \rightarrow \mathbb{R}$  we have the following filtrations of  $\mathbb{R}^d$ :

$$K_r := d_K^{-1}(-\infty, r], \quad P_r := d_P^{-1}(-\infty, r] = \bigcup_{p \in P} B_r(p).$$

*Remark 5.3.* If  $K$  has positive reach, then one can show (not surprisingly) that the map  $K_r \rightarrow K$  defined by mapping  $x$  to  $\Pi_K(x)$  is a homotopy equivalence for all  $r < \text{Rch}(K)$ .

**Theorem 5.4** (Niogi, Smale, Weinberger (2006)). *Let  $K$  be a compact submanifold of  $\mathbb{R}^d$  with positive reach. Let  $P$  be a point set such that  $d_H(K, P) = \epsilon < \sqrt{\frac{3}{20}}\text{Rch}(K)$ . Then for any  $r \in (2\epsilon, \sqrt{\frac{3}{5}}\text{Rch}(K))$ , the offset  $P_r$  deformation retracts onto  $K$ .*

This shows that the topology of  $K$  can be recovered under fairly strong assumptions. By replacing the reach with the *weak feature size* these results can be generalized substantially, but that is beyond the scope of this course; see Chapter 4 and 5 of [26].

We now proceed to give an elementary proof of the fact that the homology of a (not necessarily compact) space can be recovered from the *persistent homology* of a sufficiently good sampling.

**Theorem 5.5.** *Let  $K \subseteq \mathbb{R}^d$ ,  $P \subseteq \mathbb{R}^d$  a finite point set and  $\epsilon, \delta > 0$  such that*

1.  $K \subseteq P_\delta$  (sampling density),
2.  $P \subseteq K_\epsilon$  (sampling error: points need not lie on  $K$ ),
3. The inclusion  $K \hookrightarrow K_{\delta+\epsilon}$  induces an isomorphism  $H_n(K) \xrightarrow{\cong} H_n(K_{\delta+\epsilon})$ ,
4. The inclusion  $K_{\delta+\epsilon} \hookrightarrow K_{2(\epsilon+\delta)}$  induces a monomorphism  $H_n(K_{\delta+\epsilon}) \hookrightarrow H_n(K_{2(\epsilon+\delta)})$ .

Then  $H_n(K) \cong \text{Im}(H_n(P_\delta) \rightarrow H_n(P_{2\delta+\epsilon}))$ .

*Proof.* Assumptions 1. and 2. yield the following commutative diagram of inclusions

$$\begin{array}{ccccc} K & \hookrightarrow & K_{\delta+\epsilon} & \hookrightarrow & K_{2(\delta+\epsilon)} \\ \downarrow & \nearrow & \downarrow & & \nearrow \\ P_\delta & \hookrightarrow & P_{2\delta+\epsilon} & & \end{array}$$

which induces the following commutative diagram in homology (see Remark 4.9 and Theorem 3.6)

$$\begin{array}{ccccc} H_n(K) & \xrightarrow{\cong} & H_n(K_{\delta+\epsilon}) & \xrightarrow{\text{mono}} & H_n(K_{2(\delta+\epsilon)}) \\ \downarrow & \nearrow a & \downarrow b & & \nearrow \\ H_n(P_\delta) & \longrightarrow & H_n(P_{2\delta+\epsilon}) & & \end{array}$$

Since the top-left morphism is an isomorphism, we must have that  $a$  is surjective. Similarly,  $b$  must be a monomorphism. We conclude that

$$H_n(K) \cong H_n(K_{\delta+\epsilon}) \cong \text{Im } b = \text{Im } b \circ a = \text{Im}(H_n(P_\delta) \rightarrow H_n(P_{2\delta+\epsilon})).$$

□

Fig. 21 shows a point sample from an embedding of the circle that fails to recover the right homology dimensions for any fixed scale. However, we see that  $1 = \dim H_1(\mathbb{S}^1) = \text{Im}(H_n(P_{r_1}) \rightarrow H_n(P_{r_2}))$  for reasonable choices of  $r_1$  and  $r_2$ ; see Theorem 5.5.

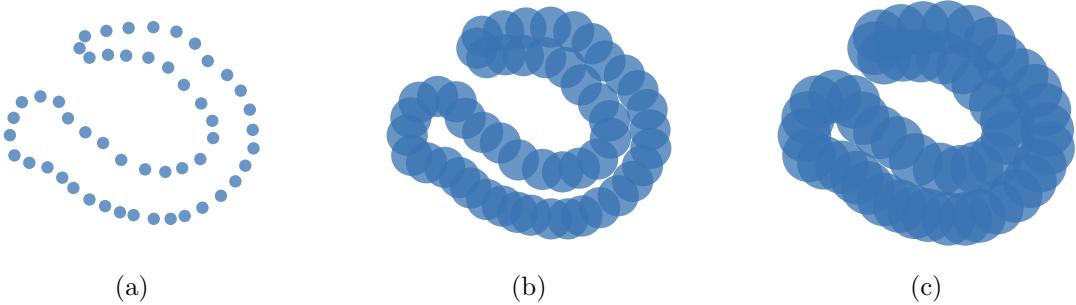


Figure 21: There exists no "right" scale that captures the hole encircled by the data points. Persistent homology on the other hand detects this feature. The figure is due to Ulrich Bauer.

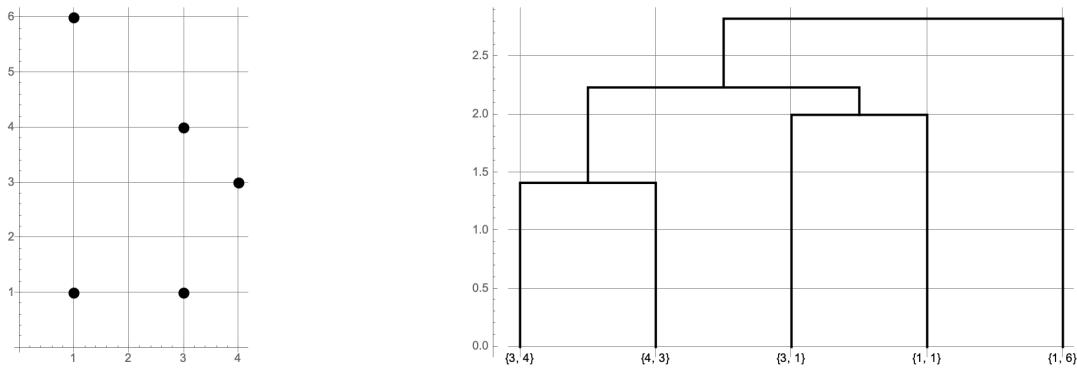


Figure 22: A point set together with its (single-linkage) dendrogram.

## 5.2 Persistent Betti Numbers

For a point set  $P \subseteq \mathbb{R}^d$ , the Betti numbers  $\beta_i(P_r)$  are very unstable in the scale parameter  $r$ . That is, a small perturbation of  $P$  may result in completely different Betti numbers. Furthermore, while a "right" scale exists for a good enough sampling of a sufficiently nice space (Theorem 5.4), fixing a scale may not be reasonable in practice (Fig. 21).

As suggested by Theorem 5.5 and Fig. 21, these issues can be rectified by considering a multi-scale view of the data. We shall see later in the course that the resulting invariants are stable with respect to perturbation. Let us first consider  $\beta_0$  - the connected components. In this setting things are particularly simple and we can use a tool from statistics called a *dendrogram*. The dendrogram is a planar tree that tracks the evolution of the connected components as the scale parameter increases.

**Example 5.6.** Consider the points  $P = \{(1, 1), (1, 6), (3, 1), (3, 4), (4, 3)\}$  of Fig. 22. When considered at a sufficiently small scale,  $P_r$  consists of 5 connected components (clusters). For  $r = \sqrt{2}$ , there are 4 connected components, and in the end there is one giant component. The corresponding dendrogram is shown in the same figure.

**Example 5.7.** Fig. 23 shows a set of points in the plane together with its associated single-linkage dendrogram. The dendrogram suggests the existence of three clusters.

Persistent homology in dimension 0 will turn out to be closely related to that of a dendrogram provided that you forget precisely which points belong to each cluster.

**Definition 5.8.** A **filtration** of a simplicial complex  $K$  is a collection of subcomplexes  $\{K_i \subseteq$

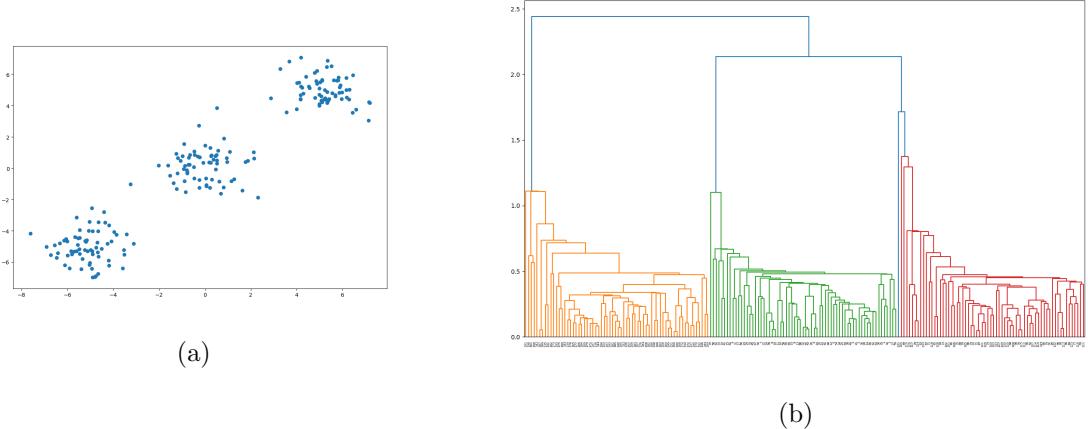


Figure 23: (a) Data. (b) Single-linkage dendrogram.

$K\}_{i=0}^m$  such that

$$K_0 \subseteq K_1 \subseteq \dots \subseteq K_m = K.$$

Applying simplicial homology to a filtration of  $K$  yields the following induced maps in homology

$$f_*^{i,j}: H_n(K_i) \rightarrow H_n(K_j)$$

for all  $i \leq j$ . The associated  **$n$ -th persistent Betti numbers** are given by  $\beta_n^{i,j} = \dim \text{Im } f_*^{i,j}$ .

**Example 5.9.** Consider the filtration depicted in Fig. 24. Some persistent Betti numbers are:  $\beta_0^{1,2} = 1$ ,  $\beta_0^{2,4} = 2$ ,  $\beta_0^{2,5} = 2$ ,  $\beta_0^{2,6} = 2$ ,  $\beta_0^{2,7} = 1$ ,  $\beta_1^{7,8} = \beta_1^{7,9} = 1$ ,  $\beta_1^{8,9} = 1$  and  $\beta_1^{7,10} = 0$ .

Let  $0 \neq [c] \in H_n(K_i)$ . Then we say that  $c$  is **born** at  $K_i$  if  $[c] \notin \text{Im } f_*^{i-1,i}$ . If  $[c]$  is born at  $K_i$ , then it **dies** at  $K_j$  if it merges with an older class as we go from  $K_{j-1}$  to  $K_j$ . That is,  $f_*^{i,j-1}([c]) \notin \text{Im } f_*^{i-1,j-1}$  but  $f_*^{i,j}([c]) \in \text{Im } f_*^{i-1,j}$ .

**Example 5.10** (Example 5.9 continued). We see that there is a class  $[c] \in H_1(K_7)$  born at  $K_7$  that dies at  $K_{10}$ . Likewise, there are two classes born at  $H_1(K_8)$  that both die in  $K_9$ . One of them gets filled in, and the other one coincides with an older class. This is the *elder rule*: precedence is given to the class with the earliest birth time.

To capture the evolution of the homological features along the filtration of a simplicial complex we introduce the following two families of integers:

$$\mu_n^{i,j} = (\beta_n^{i,j-1} - \beta_n^{i,j}) - (\beta_n^{i-1,j-1} - \beta_n^{i-1,j}), \quad \mu_n^{i,\infty} = \beta_n^{i,m} - \beta_n^{i-1,m}. \quad (3)$$

The integer  $\mu_n^{i,j}$  counts the number of *linearly independent* homology classes in dimension  $n$  born at index  $i$  that die at index  $j$ . Equivalently,

$$\mu_n^{i,j} = \dim \frac{\text{Im } f_*^{i,j-1} \cap \ker f_*^{j-1,j}}{\text{Im } f_*^{i-1,j-1} \cap \ker f_*^{j-1,j}}.$$

Likewise,  $\mu_n^{i,\infty}$  counts the number of linearly independent homology classes born at index  $i$  that never vanish. Observe that,

$$\dim H_n(K) = \sum_{i=0}^m \mu_n^{i,m}.$$

The integers  $\mu_n^{i,j}$  are usually visualized the following two ways:

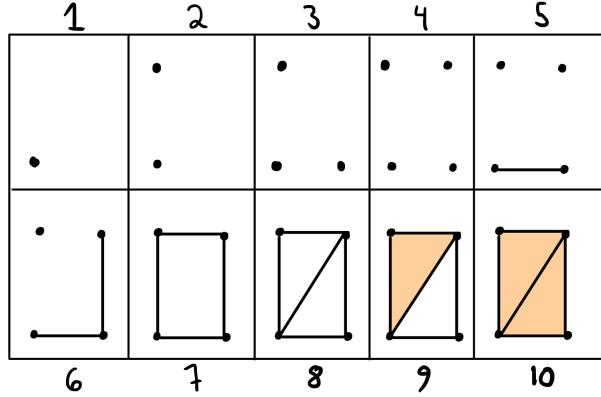


Figure 24: A filtered simplicial complex.

## Persistence Diagrams

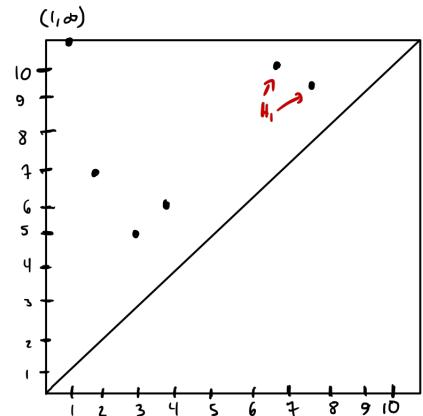
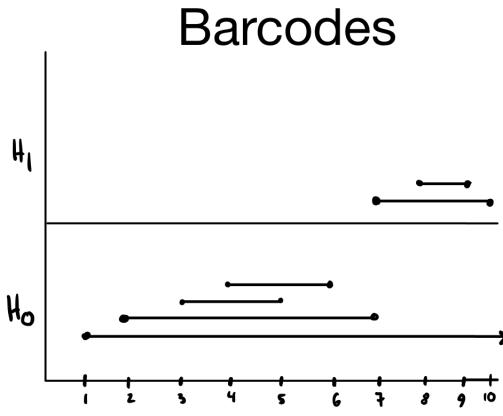


Figure 25: The barcodes and persistence diagrams of Example 5.9.

- The **barcode in dimension  $n$**  is obtained by interpreting every non-zero  $\mu_n^{i,j}$  as the interval  $[i, j)$  (representing the time when the "feature" was alive), and stacking all the intervals on top of each other in the plane; see Fig. 25.
- The **persistence diagram in dimension  $n$**  is obtained by plotting  $(i, j)$  (with multiplicity) in the plane for every non-zero  $\mu_n^{i,j}$ ; see Fig. 25.

**Example 5.11** (Example 5.10 continued). We see that the only non-zero  $\mu_n^{i,j}$ 's are the following:

$$\begin{array}{llll} \mu_0^{1,\infty} = 1 & \mu_0^{2,7} = 1 & \mu_0^{3,5} = 1 & \mu_0^{4,6} = 1 \\ \mu_1^{7,10} = 1 & \mu_1^{8,9} = 1 & & \end{array}$$

The associated barcodes and persistence diagrams are shown in Fig. 25.

### 5.2.1 The Čech and Vietoris–Rips Filtrations

**The Čech Filtration** Let  $P \subset \mathbb{R}^d$ . Then, by the Nerve lemma (Theorem 4.11), we have the following commutative diagram of vector spaces and linear morphisms

$$\begin{array}{ccccccc} H_n(P_{r_0}) & \longrightarrow & H_n(P_{r_1}) & \longrightarrow & H_n(P_{r_2}) & \longrightarrow & \cdots \longrightarrow H_n(P_{r_m}) \\ \downarrow \cong & & \downarrow \cong & & \downarrow \cong & & \downarrow \cong \\ H_n(\text{Cech}_{r_0}(P)) & \longrightarrow & H_n(\text{Cech}_{r_1}(P)) & \longrightarrow & H_n(\text{Cech}_{r_2}(P)) & \longrightarrow & \cdots \longrightarrow H_n(\text{Cech}_{r_m}(P)) \end{array}$$

It is not difficult to prove that the persistent Betti numbers for the two rows coincide, and therefore the barcode/persistence diagram of the Čech filtration has a clear geometric interpretation: it captures the evolution of "holes" of a given dimension formed by balls of increasing radius around the points in  $P$ . The same holds true for the alpha complex.

**The Vietoris–Rips Filtration** For a given scale  $r$ , it need not be the case that  $\text{VR}_r(P) \simeq P_r$ , and therefore it may appear unclear how to interpret the persistent Betti numbers associated to a filtration of the form

$$\text{VR}_{r_0}(P) \hookrightarrow \text{VR}_{r_1}(P) \hookrightarrow \text{VR}_{r_2}(P) \hookrightarrow \cdots \hookrightarrow \text{VR}_{r_m}(P).$$

However, from Corollary 1.15 we have the following chain of inclusions

$$\text{Cech}_r(P) \subseteq \text{VR}_r(P) \subseteq \text{Cech}_{2\delta r}(P) \subseteq \text{VR}_{2\delta r}(P), \quad \text{where} \quad \delta = \sqrt{\frac{d}{2(d+1)}}.$$

Applying homology yields the following commutative diagram of vector spaces:

$$\begin{array}{ccccc} H_n(\text{Cech}_r(P)) & \xrightarrow{\hspace{3cm}} & H_n(\text{Cech}_{2\delta r}(P)) & & \\ \searrow & & \nearrow & & \searrow \\ H_n(\text{VR}_r(P)) & \xrightarrow{\hspace{3cm}} & & & H_n(\text{VR}_{2\delta r}(P)) \end{array}$$

This diagram shows that any class in  $H_n(\text{VR}_r(P))$  which has non-zero image in  $H_n(\text{VR}_{2\delta r}(P))$  must be non-zero in  $H_n(\text{Cech}_{2\delta r}(P))$ . Importantly, this means that it represents a *true* topological feature of  $P_{2\delta r}(P)$ . Conversely, any class of  $H_n(\text{Cech}_r(P))$  which has non-zero image in  $H_n(\text{Cech}_{2\delta r}(P))$  must factor through  $H_n(\text{VR}_r(P))$ . Summarized: sufficiently long intervals in the barcode of the Čech filtration give rise to intervals in the barcode of the Vietoris–Rips filtration, and conversely, any sufficiently long interval in the barcode of the Vietoris–Rips filtration corresponds to a true geometrical feature. This relation will be made precise later in the course when we discuss the theory of *interleavings*.

With this obvious drawback of the Vietoris–Rips complex, one may wonder why consider it at all. There are primarily two reasons for doing so:

- Computation from a distance matrix. To construct the Čech complex one needs the coordinates of the points in  $\mathbb{R}^d$ , whereas the Vietoris–Rips is built exclusively from pairwise distances. Furthermore, this allows us to construct filtered simplicial complexes for points that do not embed in  $\mathbb{R}^d$  for any  $d$ . An example could be a distance matrix obtained through correlations of genes expression data, neural cells, etc.
- Computations. Working with the Rips complex allows for certain heuristics to be implemented which in practice yield significant speed-ups.

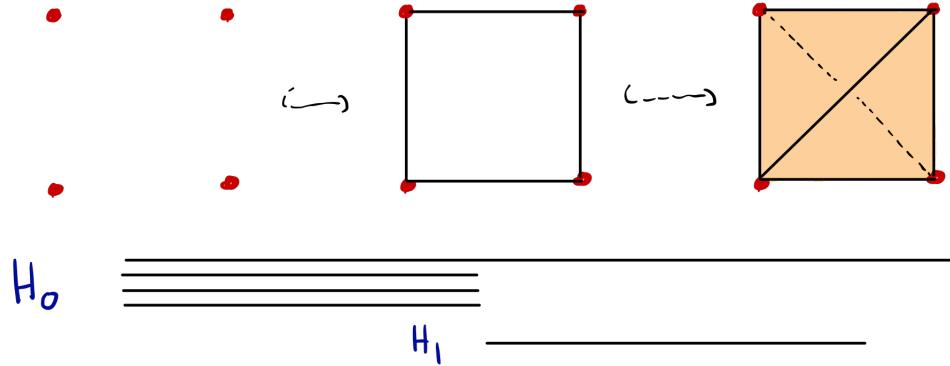


Figure 26: Vietoris–Rips filtration w/ associated barcode.

**Example 5.12.** Any finite metric space has an associated Vietoris–Rips filtration. As an example, consider the metric space consisting of three points  $\{p_1, p_2, p_3, p_4\}$  and distance matrix

$$\begin{bmatrix} 0 & 1 & 1 & 2 \\ 1 & 0 & 2 & 1 \\ 1 & 2 & 0 & 1 \\ 2 & 1 & 1 & 0 \end{bmatrix}$$

By defining  $\text{VR}_r(P) = \{\sigma \subseteq P : \text{diam}(\sigma) \leq 2r\}$  as before, we obtain a filtration:

$$\text{VR}_0(P) \subseteq \text{VR}_{1/2}(P) \subseteq \text{VR}_1(P).$$

See Fig. 26 for a (non-isometric) visualization of this filtration.

**Example 5.13.** Fig. 27 shows the persistence diagram of a the Vietoris–Rips filtration associated to a planar point set.

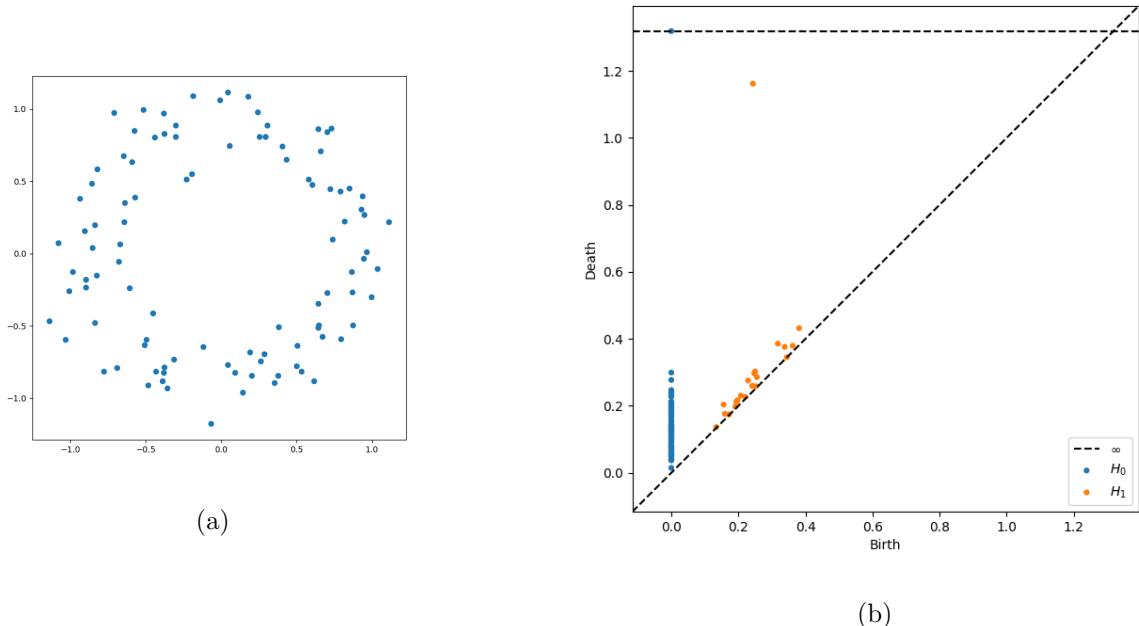


Figure 27: (a) A planar point set. (b) The persistence diagrams in dimension 0 and 1 of the associated Vietoris–Rips filtration. Computed using Python Ripser [30, 1].

### 5.3 Exercises

1. Let  $P \subset \mathbb{R}^d$ . Show that the barcodes of the Čech and alpha filtrations of  $P$  coincide (for every homology degree  $p$ ).
2. Construct a finite metric space whose associated Vietoris–Rips filtration has the following barcodes in dimensions 0,1 and 2 respectively:  $\{[0, \infty), [0, 2)_5\}$ ,  $\{[2, 3)\}$ ,  $\{[3, 4)\}$ .
3. Let  $K$  be a triangulation of the Möbius strip, and let  $K'$  denote the boundary of  $K$  (which is a triangulation of the circle). What is the rank of the map  $f_*: H_1(K') \rightarrow H_1(K)$  induced by the inclusion  $K' \subseteq K$ ? What is the rank if you work with coefficients in  $\mathbb{Z}_3$ ?
4. Let  $P$  be six equidistant points on the unit circle in  $\mathbb{R}^2$ .
  - (a) Describe  $|\text{VR}_r(P)|$  up to homeomorphism for every radius  $r \geq 0$ .
  - (b) Make barcode plots in all dimensions for the associated Vietoris–Rips and Čech filtrations.
5. Let  $P$  be such that  $H_2(\text{VR}_r(P)) \neq 0$ . Prove that  $|P| \geq 6$ . Is it possible that  $H_2(\text{Čech}_r(Q)) \neq 0$  for  $|Q| < 6$ ?

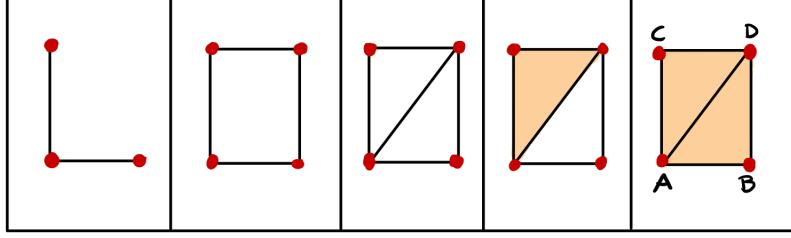


Figure 28: A filtered simplicial complex.

## 6 Algebraic Foundations

Last week we introduced the notion of persistent Betti numbers, and used them to compute the barcode. In this lecture we will give an algebraic characterization of the barcode. A more thorough discussion of the topics in this lecture can be found in any book on Quiver Representations.

### 6.1 Motivation

Consider the filtered simplicial complex  $K$  shown in Fig. 28, together with the associated collection of vector spaces and linear maps

$$H_1(K_1) \rightarrow H_1(K_2) \rightarrow H_1(K_3) \rightarrow H_1(K_4) \rightarrow H_1(K_5).$$

From the inclusion-exclusion formula in Eq. (3) it is an easy task to compute the barcode  $B_1(K)$  of  $K$  in dimension 1:

$$B_1(K) = \{[2, 5), [3, 4)\}.$$

This barcode clearly does not depend on the particular basis we choose at each index as the computation only amounts to rank considerations. From the algebraic point of view, however, there are some bases which are to be preferred. We will now illustrate this with two examples. In the first case, let  $\{AB + BD + CD + AC\}$ ,  $\{AB + BD + CD + AC, AC + CD + AD\}$  and  $\{AB + BD + CD + AC\}$  be the bases for  $K_2, K_3$ , and  $K_4$ . In the second case we replace  $\{AB + BD + CD + AC, AC + CD + AD\}$  with  $\{AB + BD + CD + AC, AB + BD + AD\}$ . Representing the linear maps with respect to these bases yields:

$$\begin{aligned} 0 \rightarrow \mathbb{Z}_2 &\xrightarrow{\begin{bmatrix} 1 \\ 0 \end{bmatrix}} \mathbb{Z}_2 \oplus \mathbb{Z}_2 \xrightarrow{\begin{bmatrix} 1 & 0 \end{bmatrix}} \mathbb{Z}_2 \rightarrow 0. \\ 0 \rightarrow \mathbb{Z}_2 &\xrightarrow{\begin{bmatrix} 1 \\ 0 \end{bmatrix}} \mathbb{Z}_2 \oplus \mathbb{Z}_2 \xrightarrow{\begin{bmatrix} 1 & 1 \end{bmatrix}} \mathbb{Z}_2 \rightarrow 0. \end{aligned}$$

The former of these two splits into a direct sum:

$$(0 \rightarrow \mathbb{Z}_2 \xrightarrow{1} \mathbb{Z}_2 \xrightarrow{1} \mathbb{Z}_2 \rightarrow 0) \bigoplus (0 \rightarrow 0 \rightarrow \mathbb{Z}_2 \rightarrow 0 \rightarrow 0),$$

whereas the latter does not "split up". An advantage of such a decomposition is that the barcode can be easily read off from the support of its components (summands):

$$\begin{aligned} \mu_1^{i,j} \left( (0 \rightarrow \mathbb{Z}_2 \xrightarrow{1} \mathbb{Z}_2 \xrightarrow{1} \mathbb{Z}_2 \rightarrow 0) \bigoplus (0 \rightarrow 0 \rightarrow \mathbb{Z}_2 \rightarrow 0 \rightarrow 0) \right) &= \\ \mu_1^{i,j} \left( 0 \rightarrow \mathbb{Z}_2 \xrightarrow{1} \mathbb{Z}_2 \xrightarrow{1} \mathbb{Z}_2 \rightarrow 0 \right) + \mu_1^{i,j} \left( 0 \rightarrow 0 \rightarrow \mathbb{Z}_2 \rightarrow 0 \rightarrow 0 \right) &= \begin{cases} 1 & \text{if } (i, j) \in \{(2, 5), (3, 4)\} \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

The goal of this section will be to prove that such a choice of basis *always* exists. This algebraic characterization of the barcode will be important when proving stability later in the course.

## 6.2 Persistence Modules

**Definition 6.1.** A **partially ordered set** (poset)  $P$  is a set  $P$  together with a binary relation  $\leq$  such that

1.  $p \leq p$ , for all  $p \in P$ ,
2.  $p \leq q$  and  $q \leq p$  implies  $p = q$ ,
3.  $p \leq q$  and  $q \leq r$  implies  $p \leq r$ .

If  $p \leq q$  and  $p \neq q$ , then we write  $p < q$ .

Here are some posets that appear in topological data analysis.

**Example 6.2.** The set  $[n] = \{1, 2, \dots, n\}$  is a partially ordered set under the standard ordering of the natural numbers. It will be convenient to represent this poset as a graph in the following way:

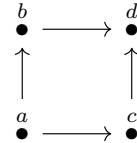
$$\bullet^1 \longrightarrow \bullet^2 \longrightarrow \cdots \longrightarrow \bullet^{n-1} \longrightarrow \bullet^n.$$

More generally,  $\mathbb{N}, \mathbb{Z}$  and  $\mathbb{R}$  are posets in the obvious way.

**Example 6.3.** One can also assign a zigzag structure to  $\{1, 2, \dots, n\}$ :

$$\bullet^1 \longleftarrow \bullet^2 \longrightarrow \cdots \longleftarrow \bullet^{n-1} \longrightarrow \bullet^n$$

**Example 6.4.** The poset  $P = \{a, b, c, d\}$  with the binary relation  $a \leq b, c \leq d$ :



**Definition 6.5.** A  **$P$ -indexed persistence module** ( $P$ -module)  $V$  is a collection of vector spaces  $\{V_p\}_{p \in P}$  and linear maps  $\{V(p \leq q): V_p \rightarrow V_q\}_{p \leq q}$  such that

1.  $V(p \leq p) = \text{id}: V_p \rightarrow V_p$ ,
2.  $V(q \leq r) \circ V(p \leq q) = V(p \leq r)$ .

*Remark 6.6.* A  $P$ -module is a functor from the partially ordered set  $P$  seen as a category in the obvious way, to the category of vector spaces.

**Example 6.7.** A filtered simplicial complex  $K_1 \subseteq K_2 \subseteq \cdots \subseteq K_n$  yields an  $[n]$ -module for every integer  $i$ ,

$$H_i(K_1) \rightarrow H_i(K_2) \rightarrow \cdots \rightarrow H_i(K_n).$$

**Example 6.8.** We can also consider a zigzag of inclusions:  $K_1 \subseteq K_2 \supseteq K_3 \subseteq \cdots \supseteq K_n$ ,

$$H_i(K_1) \rightarrow H_i(K_2) \leftarrow H_i(K_3) \rightarrow \cdots \leftarrow H_i(K_n).$$

**Example 6.9.** Given a topological space  $X$  and a real-valued function  $f: X \rightarrow \mathbb{R}$ , we get an  $\mathbb{R}$ -module  $M$  by

$$M_t = H_i(f^{-1}(-\infty, t])$$

together with maps  $M_t \rightarrow M_{t'}$  in homology induced by the inclusion  $f^{-1}(-\infty, t] \subseteq f^{-1}(-\infty, t']$ .

A **morphism**  $f: V \rightarrow W$  between  $P$ -modules is a collection of linear maps  $\{f_p: V_p \rightarrow W_p\}_{p \in P}$  such that the following diagram commutes for all  $p \leq q$ :

$$\begin{array}{ccc} V_p & \xrightarrow{V(p \leq q)} & V_q \\ \downarrow f_p & & \downarrow f_q \\ W_p & \xrightarrow{W(p \leq q)} & W_q. \end{array}$$

We say that  $f$  is an **epi-/mono-/isomorphism** if  $f_p$  is an epi-/mono-/isomorphism for all  $p$ .

**Example 6.10.** The following is an example of an epimorphism:

$$\begin{array}{ccc} V: & \mathbb{Z}_2 & \xrightarrow{1} \mathbb{Z}_2 \\ \downarrow f & \downarrow 1 & \downarrow 0 \\ W: & \mathbb{Z}_2 & \xrightarrow{0} 0. \end{array}$$

Note that there are no non-zero morphisms from  $W$  to  $V$ .

The **direct sum** of two  $P$ -modules  $V$  and  $W$  is the  $P$ -module  $V \oplus W$  defined pointwise by  $(V \oplus W)_p = V_p \oplus W_p$  and on linear maps by:

$$V_p \oplus W_p \xrightarrow{\begin{bmatrix} V(p \leq q) & 0 \\ 0 & W(p \leq q) \end{bmatrix}} V_q \oplus W_q.$$

**Definition 6.11.** We say that  $V$  is **decomposable** if there exist non-zero  $W, W'$  such that  $V \cong W \oplus W'$ . If no such  $W, W'$  exist, then we say that  $V$  is **indecomposable**.

For a  $P$ -module  $V$ , we define the **total dimension** of  $V$  by  $\dim V = \sum_{p \in P} \dim V_p$  whenever the sum exists.

**Lemma 6.12.** Let  $P$  be finite and assume that  $V_p$  is finite-dimensional for all  $p$ . Then

$$V \cong W^1 \oplus \cdots \oplus W^k$$

where each  $W^i$  is indecomposable.

*Proof.* We will work inductively on the total dimension of  $V$ . If  $\dim V = 1$ , then  $V$  is necessarily indecomposable. So assume that the statement holds for  $\dim V \leq n$  and consider  $V$  with  $\dim V = n + 1$ . If  $V$  is indecomposable, then we are done. If not, we can write  $V = W \oplus W'$  where

$$\dim V = \dim W + \dim W'$$

and both of the terms to the right are non-zero. Hence, by the inductive property,  $W$  and  $W'$  decompose into a direct sums of indecomposables.  $\square$

A **submodule**  $V'$  of  $V$  is a collection of subspaces  $V'_p \subseteq V_p$  such that the following diagram commutes for all  $p \leq q$

$$\begin{array}{ccc} V'_p & \longrightarrow & V'_q \\ \downarrow \subseteq & & \downarrow \subseteq \\ V_p & \longrightarrow & V_q \end{array}$$

If there exists a non-trivial submodule  $V'' \subseteq V$  such that  $V = V' \oplus V''$ , then we say that  $V'$  and  $V''$  are **summands** of  $V$ . For a morphism  $f: V \rightarrow W$  we get submodules  $\ker f \subseteq V$  and  $\text{Im } f \subseteq W$  by taking pointwise kernels and images. These submodules are in general not summands.

**Example 6.13.**

$$\begin{array}{ccccccc} \ker(f) & & 0 & \longrightarrow & 0 & \longrightarrow & \mathbb{Z}_2 \\ \downarrow \subseteq & & \downarrow & & \downarrow & & \downarrow 1 \\ V & & 0 & \longrightarrow & \mathbb{Z}_2 & \xrightarrow{1} & \mathbb{Z}_2 \\ \downarrow f & & \downarrow & & \downarrow 1 & & \downarrow \\ W & & \mathbb{Z}_2 & \xrightarrow{1} & \mathbb{Z}_2 & \longrightarrow & 0 \\ \uparrow \subseteq & & \uparrow & & \uparrow 1 & & \uparrow \\ \text{Im}(f) & & 0 & \longrightarrow & \mathbb{Z}_2 & \longrightarrow & 0 \end{array}$$

**Lemma 6.14** (Fitting Lemma). *Let  $P$  be finite and  $\dim V_p < \infty$  for all  $p$ . For every morphism  $f: V \rightarrow V$  there exists an  $n \geq 1$  such that*

$$V = \text{Im } f^n \oplus \ker f^n.$$

*Proof.* Since  $f(V) \supseteq f^2(V) \supseteq f^3(V) \supseteq \dots$  and  $\dim V$  is finite, there must exist an  $l$  such that  $f^l(V) = f^{l+1}(V) = f^{l+2}(V) = \dots$ . Define  $\phi := f^l: V \rightarrow V$ . As we saw above,  $\ker \phi$  and  $\text{Im } \phi$  are both submodules of  $V$ . It remains to show that they are complementary summands.

For every  $v \in V_p$ , there exists a  $v' \in V_p$  such that  $\phi_p(v) = \phi_p^2(v')$ , and thus  $v - \phi_p(v') \in \ker \phi_p$ . This shows that  $v = (v - \phi_p(v')) + \phi_p(v')$  where  $v - \phi_p(v') \in \ker \phi_p$  and  $\phi_p(v') \in \text{Im } \phi_p$ .

To see that  $\ker \phi \cap \text{Im } \phi = 0$ , let  $v \in \ker \phi \cap \text{Im } \phi$ , and observe that  $v = \phi_p(v')$  and  $0 = \phi_p(v) = \phi_p^2(v')$ . By our choice of  $\phi$ , we know that  $\phi_p: \phi_p(V_p) \rightarrow \phi_p^2(V_p)$  is an isomorphism, and therefore  $\phi_p^2(v') = 0$  implies that  $v = \phi_p(v') = 0$ .  $\square$

Now we restrict our attention to the poset  $[n]$ . Let  $\mathbf{k}$  be any field, and for  $a < b \in [n] \cup \{\infty\}$ , define  $[a, b)$  to be the subset of  $[n]$  given by the elements  $\{i: a \leq i < b\}$ . We define the **interval module** associated to  $[a, b)$  to be the  $[n]$ -module  $I^{[a, b)}$  defined by

$$I_i^{[a, b)} = \begin{cases} \mathbf{k} & \text{if } i \in [a, b) \\ 0 & \text{otherwise} \end{cases},$$

together with the identity morphism  $\text{id}_{\mathbf{k}}: I_i^{[a, b)} \rightarrow I_j^{[a, b)}$  whenever  $i, j \in [a, b)$ .

**Example 6.15.** For  $n = 2$  we have the following interval modules

$$\mathbf{k} \rightarrow 0 \quad 0 \rightarrow \mathbf{k} \quad \mathbf{k} \xrightarrow{\text{id}_{\mathbf{k}}} \mathbf{k}.$$

**Theorem 6.16.** Let  $V$  be an  $[n]$ -module such that  $\dim V_p < \infty$  for all  $p \in [n]$ . Then

$$V \cong \bigoplus_{[a,b] \in B(V)} I^{[a,b)}$$

where  $B(V)$  is a multiset of intervals in  $[n]$  called the **barcode** of  $V$ .

*Proof.* We know from Lemma 6.12 that  $V$  decomposes into a direct sum of indecomposables. Hence it suffices to show that if  $V$  is indecomposable, then  $V \cong I^{[a,b)}$  for some interval  $[a, b)$ . Assume without loss of generality that  $V_1 \neq 0$  and choose  $0 \neq v \in V_1$ . Let  $b$  denote the smallest index for which  $V(1 \leq b)(v) = 0$  and let  $j = b - 1$ . If no such  $b$  exists, then let  $b = \infty$  and  $j = n$ .

First we define a (well-defined) monomorphism  $f: I^{[1,b)} \rightarrow V$  by  $f_i(1_k) = V(1 \leq i)(v)$ . Since  $V_j$  is a vector space, we can extend  $\{f(1_k)\}$  to a basis for  $V_j$  and thus define an epimorphism  $g_j: V_j \rightarrow k$  which satisfies  $g_j(f_j(1_k)) = \text{id}_k$ . This extends to an (well-defined) epimorphism  $g: V \rightarrow I^{[1,b)}$  by defining  $g_i(w) = g_j(V(i \leq j)(w))$  for  $i \leq j$ , and 0 otherwise. Importantly,

$$g \circ f = \text{id}_{I^{[1,b)}}.$$

From Lemma 6.14 there exists an  $m \geq 1$  such that

$$V = \text{Im}(f \circ g)^m \oplus \ker(f \circ g)^m.$$

Furthermore,

$$(f \circ g)^m(V) = f \circ (g \circ f) \circ (g \circ f) \circ \cdots \circ (g \circ f) \circ g(V) = f \circ g(V) = f(I^{[1,b)}),$$

where the last equality follows from  $g$  being an epimorphism. Since  $V$  was assumed to be indecomposable, and since  $f$  is a monomorphism, it follows that

$$V = \text{Im}(f \circ g)^m = f(I^{[1,b)}) \cong I^{[1,b)}.$$

□

**Proposition 6.17.** Let  $V$  be as in Theorem 6.16. Then  $\mu_n^{i,j}$  counts the number of occurrences of  $[i, j)$  in  $B(V)$ . In particular.

*Proof.* This follows from the following two observations

$$\mu_n^{i,j}(V) = \mu^{i,j} \left( \bigoplus_{[a,b] \in B(V)} I^{[a,b)} \right) = \sum_{[a,b] \in B(V)} \mu^{i,j}(I^{[a,b)}),$$

and

$$\mu^{i,j}(I^{[a,b)}) = \begin{cases} 1 & \text{if } (i, j) = (a, b) \\ 0 & \text{otherwise.} \end{cases}$$

□

Just as the dimension of a vector space is independent of the choice of basis, the previous result shows that although the particular decomposition in Theorem 6.16 is not unique, the associated multiset  $B(V)$  is in fact unique. We state this as a corollary.

**Corollary 6.18.** Let  $V$  be as in Theorem 6.16, and assume that

$$V \cong \bigoplus_{[a,b) \in \mathcal{I}} I^{[a,b)} \cong \bigoplus_{[a,b) \in \mathcal{J}} I^{[a,b)},$$

Then  $\mathcal{I} = \mathcal{J}$ .

We state the following generalization of Theorem 6.16 without proof.

**Theorem 6.19** ([6]). Let  $T$  be a totally ordered set and  $V$  a  $T$ -module such that  $\dim V_t < \infty$  for all  $t \in T$ . Then

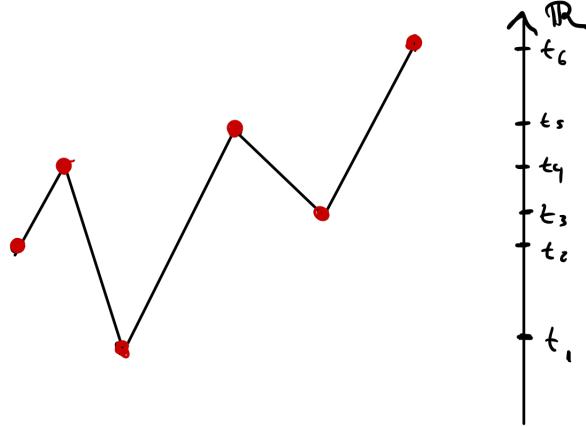
$$V \cong \bigoplus_{J \in B(V)} I^J$$

where  $B(V)$  is a unique multiset of intervals in  $T$ .

Here an *interval* in  $T$  is a subset  $J$  with the property that if  $x < y < z$  and  $x, z \in J$  then  $y \in J$ . Interval modules are defined accordingly.

### 6.3 Exercises

1. Let  $f : X \rightarrow \mathbb{R}$  be the height-function shown in the figure below.
  - (a) Define bases for  $H_0(f^{-1}(-\infty, t_i])$  such that the persistence module  $H_0(f^{-1}(-\infty, t_1]) \rightarrow \dots \rightarrow H_0(f^{-1}(-\infty, t_5]) \rightarrow H_0(f^{-1}(-\infty, t_6])$  decomposes as a direct sum of submodules isomorphic to interval modules.
  - (b) Use (a) to decompose the  $\mathbb{R}$ -module given by  $M_t = H_0(f^{-1}(-\infty, t])$  into interval modules (up to isomorphism).



2. Describe a zigzag of simplicial complexes  $K_1 \hookleftarrow K_2 \hookrightarrow K_3$  such that  $H_1(K_1) \leftarrow H_1(K_2) \rightarrow H_1(K_3)$  is isomorphic to

$$\mathbb{Z}_2 \xleftarrow{[11]} \mathbb{Z}_2 \oplus \mathbb{Z}_2 \xrightarrow{[01]} \mathbb{Z}_2,$$

and choose a basis for the middle vector space such that the persistence modules decompose non-trivially. Conclude that a submodule need not have support contained in any one summand. In particular, the *barcode* of the zigzag module does not contain an interval supported on all of the vertices, although there exists a feature in the middle whose image is non-trivial in both vector spaces.

3. Can you think of a scenario where zigzag persistent homology could be useful?

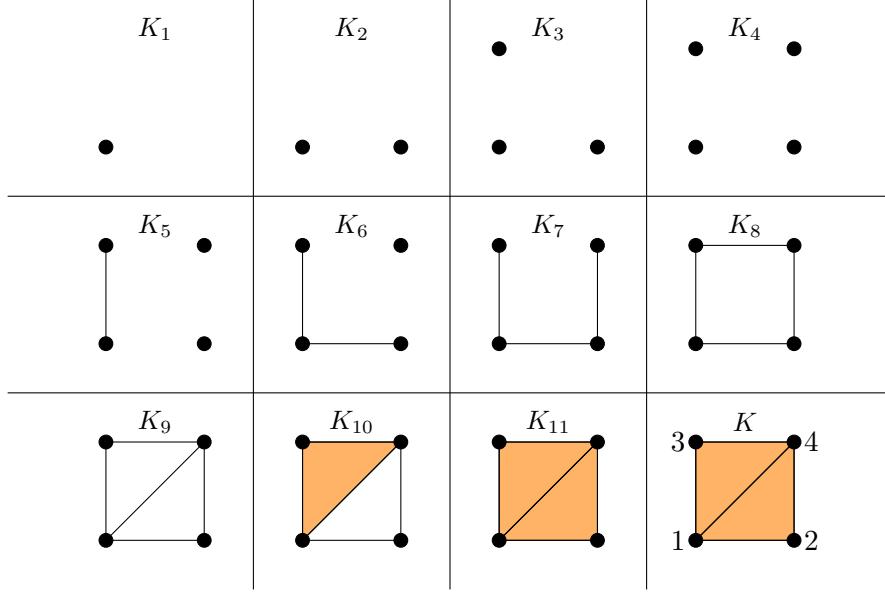


Figure 29: Yet another filtered simplicial complex.

## 7 The Persistence Algorithm

The algorithm for computing simplicial homology can be straightforwardly applied to compute persistent homology. In this lecture we will have a look at the algorithm and an improvement which makes computations faster in practice.

### 7.1 The standard algorithm

The *standard algorithm* for computing persistent homology is nothing more than the standard algorithm for simplicial homology with the ordering of the simplices given by the time they appear in the filtration. We illustrate this by computing the barcode of the filtration in Fig. 29. Expressing the boundary matrix in the order given by the simplices yields:

$$D = \begin{pmatrix} 1 & 2 & 3 & 4 & 13 & 12 & 24 & 34 & 14 & 134 & 124 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 13 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 12 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 24 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 34 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 14 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 134 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 124 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and the corresponding identity matrix  $V = I_{12 \times 12}$ . The resulting matrices  $R$  and  $V$  after

applying Algorithm 1 to  $D$  are:

$$R = \begin{pmatrix} 1 & 2 & 3 & 4 & 13 & 12 & 24 & 34 & 14 & 134 & 124 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 13 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 12 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 24 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 34 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 14 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 134 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 124 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and

$$V = \begin{pmatrix} 1 & 2 & 3 & 4 & 13 & 12 & 24 & 34 & 14 & 134 & 124 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 13 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 12 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 24 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 34 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 14 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 134 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 124 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Observe that the simplicial homology of  $K_i$  - the filtered simplicial complex at time  $i$  - can be obtained by restricting  $R$  and  $V$  to the first  $i$  columns.

Following the steps in Section 3.1.2:

$$\begin{aligned} \Sigma_{B_0} &= \{(10100000000)^T, (11000000000)^T, (01010000000)^T\} \\ \widetilde{\Sigma_{Z_0}} &= \{(10000000000)^T, (01000000000)^T, (00100000000)^T, (00010000000)^T\} \\ \Sigma_{B_1} &= \{(00001001100)^T, (00001111000)^T\} \\ \widetilde{\Sigma_{Z_1}} &= \{(00001111000)^T, (00000110100)^T\}. \end{aligned}$$

The next step is to pair the vectors of  $\widetilde{\Sigma_{Z_0}}$  to those of  $\Sigma_{B_0}$  based on their lowest entries:

$$\begin{array}{ll} (10000000000)^T \leftrightarrow \emptyset & (01000000000)^T \leftrightarrow (11000000000)^T \\ (00100000000)^T \leftrightarrow (10100000000)^T & (00010000000)^T \leftrightarrow (01010000000)^T \end{array}$$

and in dimension 1:

$$(00001111000)^T \leftrightarrow (00001111000)^T \quad (00000110100)^T \leftrightarrow (00001001100)^T$$

From this we conclude that  $H_0(K_{11}) \cong \mathbb{Z}_2$  is generated by the vertex  $v_1$ , and that  $H_i(K_{11}) = 0$  for  $i \geq 1$ . However, from the point of view of persistent homology, we are interested in the

particular pairings: a pairing  $v \leftrightarrow w$  yields a bar  $[t_v, t_w)$  in the barcode, where  $t_v$  is the time at which the cycle  $v$  appears in the filtration, and  $t_w$  is the time where  $w$  becomes a boundary. We see that  $t_v = \text{low}(v)$  and  $t_w = j$  where  $\text{low}(R_j) = t_v$ . We summarize our findings in the following table:

Dimension	Birth	Death	Rep. cycle	Vertex notation
0	1	$\infty$	$(10000000000)^T$	1
0	2	6	$(01000000000)^T$	2
0	3	5	$(00100000000)^T$	3
0	4	7	$(00010000000)^T$	4
1	8	11	$(00001111000)^T$	$13 + 12 + 24 + 34$
1	9	10	$(00000110100)^T$	$12 + 24 + 14$

This completes the computation of persistent homology. We summarize our findings:

1. Write down the boundary matrix  $D$  with the simplices ordered by their appearance in the filtration of a simplicial complex  $K$ .
2. Use Algorithm 1 to obtain a decomposition  $R = DV$ .
3. The columns  $R_i$  with  $\text{low}(i) \neq 0$  form a basis  $\Sigma_B = \Sigma_{B_0} \cup \dots \cup \Sigma_{B_d}$  for  $B_0(K) \oplus \dots \oplus B_d(K)$ ,
4. The columns  $V_i$  such that  $R_i = 0$  form a basis  $\widetilde{\Sigma}_Z = \widetilde{\Sigma}_{Z_0} \cup \dots \cup \widetilde{\Sigma}_{Z_d}$  for  $Z_0(K) \oplus \dots \oplus Z_d(K)$ ,
5. For every  $n$ -cycle  $V_i \in \widetilde{\Sigma}_Z$  do the following:
  - (a) If there exists  $R_j \in \Sigma_B$  with  $\text{low}(R_j) = \text{low}(V_i) = i$ , then output an interval  $[i, j)$  in dimension  $n$ .
  - (b) If there exists no such  $R_j$ , then output an interval  $[i, \infty)$  in dimension  $n$ .

**Bases and correctness** At the algebraic level we also get bases at every filtration step:

1. Let  $\Sigma_{H_j} = \{V_i \in \widetilde{\Sigma}_Z : i \leq j \text{ and } \exists R_l \text{ where } j < l \text{ and } \text{low}(R_l) = i\}$ ,
2. Let  $\Sigma_{H_{j_1, j_2}} = \Sigma_{H_{j_1}} \cap \Sigma_{H_{j_2}}$ ,
3. Let  $\Sigma_{E_j} = \{V_i \in \widetilde{\Sigma}_Z : i \leq j \text{ and there exists no } R_l \text{ such that } \text{low}(R_l) = i\}$ .

For each of the above sets, let  $\Sigma^n$  denote the subset of  $n$ -cycles.

It is a good exercise to verify the following steps used to prove that the algorithm outputs the correct barcode.

1. The set  $\Sigma_{H_j}^n \cup \Sigma_{E_j}^n$  forms a basis for  $H_n(K_j)$ .
2. A basis for  $\text{Im}(H_n(K_{j_1}) \rightarrow H_n(K_{j_2}))$  is given by  $\Sigma_{H_{j_1, j_2}}^n \cup \Sigma_{E_{j_1}}^n$ , and thus

$$\beta_n^{j_1, j_2} = |\Sigma_{H_{j_1, j_2}}^n \cup \Sigma_{E_{j_1}}^n|,$$

and if  $m$  denotes the maximal filtration value,

$$\beta_n^{j_1, m} = |\Sigma_{E_{j_1}}^m|.$$

3. Since  $\Sigma_{H_{j_1,j_2}}^n \subseteq \Sigma_{H_{j_1,j_2-1}}^n$  for  $j_1 < j_2$ ,

$$\beta_n^{j_1,j_2-1} - \beta_n^{j_1,j_2} = |\Sigma_{H_{j_1,j_2-1}}^n \setminus \Sigma_{H_{j_1,j_2}}^n| \in \{0, 1\}$$

which is non-zero precisely if  $R_{j_2} \neq 0$  corresponds to an  $(n+1)$ -simplex and  $\text{low}(R_{j_2}) \leq j_1$ .

4. It follows that

$$\mu_n^{j_1,j_2} = (\beta_n^{j_1,j_2-1} - \beta_n^{j_1,j_2}) - (\beta_n^{j_1-1,j_2-1} - \beta_n^{j_1-1,j_2})$$

is non-zero and equal to 1 if and only  $R_{j_2} = 0$  corresponds to an  $(n+1)$ -simplex, and  $\text{low}(R_{j_2}) = j_1$ . Hence, the algorithm computes all finite bars correctly.

5. Similarly, for  $j_1 < j_2$ ,  $\Sigma_{E_{j_1}}^n \subseteq \Sigma_{E_{j_2}}^n$ , and for  $m$  the maximal filtration value,

$$\mu_n^{j_1,\infty} = \beta_n^{j_1,m} - \beta_n^{j_1-1,m} = |\Sigma_{E_{j_1}}^n \setminus \Sigma_{E_{j_1-1}}^n| \in \{0, 1\}$$

which is precisely the number of infinite bars in homology degree  $n$  born at  $j_1$  computed by the algorithm.

**Algebraic Justification** The standard algorithm actually provides us with enough data to decompose the homology persistence modules into summands isomorphic to interval modules. We saw above that the  $n$ -cycles of  $\Sigma_{H_j} \cup \Sigma_{E_j}$  form a basis for  $H_n(K_j)$ , and that every  $V_i \in \Sigma_{H_j}$  is paired to an  $R_l$ , where  $R_l = V_i + \sum_{k=0}^{i-1} c_k V_k$  for constants  $c_k \in \{0, 1\}$ . We can thus replace the basis element  $V_i$  with  $R_l$ , and then the morphism  $H_n(K_{l-1}) \rightarrow H_n(K_l)$  maps the basis element  $R_l$  of  $H_n(K_{l-1})$  to 0. Hence, all the morphisms diagonalize and the persistence module decomposes into indecomposable summands. Let us illustrate this for the case  $n = 1$  above. Using the basis as given by the persistence algorithm we get the bases (using vertex notation)  $\{12 + 13 + 24 + 34, 12 + 24 + 14\}$  and  $\{12 + 13 + 24 + 34\}$ , for  $H_1(K_9)$  and  $H_1(K_{10})$ , respectively. However, the cycle  $12 + 24 + 14$  becomes equivalent to  $12 + 13 + 24 + 14$  in homology as we enter  $K_{10}$ , and therefore the linear maps take the form:

$$\text{span}\{12 + 13 + 24 + 34, 12 + 24 + 14\} \xrightarrow{[1 \ 1]} \text{span}\{12 + 13 + 24 + 34\}.$$

If we, however, replace  $12 + 24 + 14$  with the boundary to which it gets paired, then we obtain a new basis  $\{12 + 13 + 24 + 34, 13 + 34 + 14\}$  and the matrix diagonalizes to:

$$\text{span}\{12 + 13 + 24 + 34, 13 + 34 + 14\} \xrightarrow{[1 \ 0]} \text{span}\{12 + 13 + 24 + 34\}.$$

And therefore  $0 \rightarrow H_1(K_8) \rightarrow H_1(K_9) \rightarrow H_1(K_{10}) \rightarrow 0$  decomposes as

$$(0 \rightarrow \mathbb{Z}_2 \xrightarrow{1} \mathbb{Z}_2 \xrightarrow{1} \mathbb{Z}_2 \rightarrow 0) \bigoplus (0 \rightarrow 0 \rightarrow \mathbb{Z}_2 \rightarrow 0 \rightarrow 0),$$

See Section 6.1 for how this decomposition relates to the barcode.

We conclude that a pairing  $V_i \leftrightarrow R_j$  corresponds to an interval summand with support  $[i, j]$ , and that an unpaired  $V_i$  corresponds to an interval summand with support  $[0, \infty)$ .

## 7.2 Clearing

Consider the reduced matrix  $R$  in the previous section. Since the column corresponding to the simplex 134 is non-zero, we know from the previous discussion that column number  $i = \text{low}(134)$ , which in this case is corresponds to the simplex 14, must be zeroed out. Likewise, we know that the column number  $\text{low}(124)$ , which corresponds to simplex 34, must be zeroed out. Hence, if we first reduced the 2-simplices, then we could skip the columns 14 and 34 when reducing the 1-simplices, as we know that they will be zeroed out. This is the whole idea behind **clearing** and can be summarized as follows:

- Let  $K$  be a filtered simplicial complex with  $\dim K = d$ .
- Use the standard algorithm to reduce the columns of  $D$  corresponding to  $d$ -simplices in order to obtain  $\Sigma_{B_{d-1}}$  and  $\widetilde{\Sigma_{Z_d}}$ .
- For every  $R_j \in \Sigma_{B_{d-1}}$ , zero out column number  $\text{low}(R_j)$  in  $D$ . Reduce the  $(d-1)$ -simplices of the resulting boundary matrix to obtain the set  $\Sigma_{B_{d-2}}$ . A basis  $\widetilde{\Sigma_{Z_{d-1}}}$  for  $Z_{d-1}(K)$  is given by the set  $\Sigma_{B_{d-1}}$  together with all columns  $V_i$  corresponding to zero columns  $R_i$  that were not zeroed out prior to the reduction.
- Continue the same way until a bases for  $B_0(K)$  and  $Z_1(K)$  are computed.

The total number of rows that need not be reduced are

$$\dim B_{d-1}(K) + \dim B_{d-2}(K) + \dots + \dim B_1(K). \quad (4)$$

We do however need to reduce all columns corresponding to  $d$ -simplices, and in applications this number typically is much larger than the sum in Eq. (4). But the crux of the matter lies with the fact that reducing a column to zero is typically more expensive than reducing a column to become a boundary. And in the above case we still need to zero out a large number of columns corresponding to  $d$ -simplices. This turns out to be a major bottleneck and working with *cohomology* avoids this issue in practice.

**Example 7.1.** Let  $P$  be a set of 100 points in Euclidean space and say we want to compute the persistent homology of the Vietoris–Rips filtration in dimensions 0, 1 and 2. Counting the total number of 1, 2 and 3-simplices gives the total number of columns to be reduced in the naïve approach (0-simplices are already reduced):

$$\binom{100}{2} + \binom{100}{3} + \binom{100}{4} = 4087875.$$

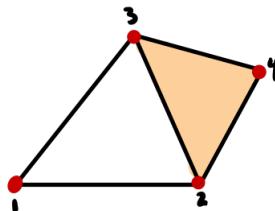
Now (exercise)

$$\dim B_2(\text{VR}_\infty(P)) + \dim B_1(\text{VR}_\infty(P)) = \binom{99}{3} + \binom{99}{2} = 161700.$$

Clearing only marginally improves the computation time and the reason for this is the substantial number of 3-simplices that still need to be reduced in the initial step. Indeed, the number of columns which need to be zeroed out equals  $\dim Z_3(\text{VR}_\infty(P)) = \binom{99}{4} = 3764376$ .

### 7.3 Exercises

1. Use the standard algorithm for persistent homology to compute the barcode of following simplicial complex with simplices ordered by 1, 2, 3, 12, 13, 24, 23, 34, 234. Find representative cycles for each interval. Use the output of the algorithm to choose bases  $H_1(K_r)$  such that every linear map is represented by a diagonal matrix.



2. Verify the claims in Example 7.1.

## 8 Cohomology

*Cohomology* is a prominent tool in algebraic topology that carries more structure than homology. Importantly, cohomology comes equipped with a *cup product* that endows cohomology with a graded ring structure. Furthermore, cohomology is a *representable functor*, meaning that cohomology groups can be identified with homotopy classes of maps in a natural way. The latter forms the foundation of *circular coordinates*, a way of parametrizing the data with respect to a circular feature. In this lecture we shall see how cohomology often leads to faster computations.

### 8.1 Dual Persistence Modules

Recall that the **dual** of a vector space  $V$  over a field  $\mathbf{k}$  is the vector space of linear maps  $V^* := \text{Hom}(V, \mathbf{k})$ . It is well-known (and easy to prove) that if  $\{v_1, \dots, v_d\}$  is a basis for  $V$ , then  $\{\epsilon_1, \dots, \epsilon_d\}$  is a basis for  $V^*$ , where  $\epsilon_i(v_i) = 1$  and  $\epsilon_i(v_j) = 0$  for  $i \neq j$ . Therefore  $\dim V = \dim V^*$  if  $V$  is finite-dimensional.<sup>5</sup> Furthermore, if  $f: V \rightarrow W$  is a linear map, then we get an induced linear map  $f^*: W^* \rightarrow V^*$  given by  $f^*(\alpha)(x) = \alpha(f(x))$ . If the matrix representation of  $f$  with respect to a choice of bases is the matrix  $A$ , then  $f^*$  is represented in the dual bases by the matrix  $A^T$ . It is clear that  $(g \circ f)^* = f^* \circ g^*$ . Hence, for an  $[m]$ -module  $V$

$$V_1 \xrightarrow{V(1 \leq 2)} V_2 \xrightarrow{V(2 \leq 3)} \cdots \xrightarrow{V(m-1 \leq m)} V_m$$

we get an  $[m]$ -module  $V^*$

$$V_m^* \xrightarrow{V(m-1 \leq m)^*} V_{m-1}^* \xrightarrow{V(m-2 \leq m-1)^*} \cdots \xrightarrow{V(1 \leq 2)^*} V_1^*.$$

Observe that dualization respects direct sums of persistence modules: for two  $[m]$ -modules  $V$  and  $W$ , there is an isomorphism  $V^* \oplus W^* \cong (V \oplus W)^*$ .

**Example 8.1.** The dual of  $0 \rightarrow \mathbf{k} \xrightarrow{1_{\mathbf{k}}} \mathbf{k} \xrightarrow{1_{\mathbf{k}}} \mathbf{k}$  is isomorphic to  $\mathbf{k} \xrightarrow{1_{\mathbf{k}}} \mathbf{k} \xrightarrow{1_{\mathbf{k}}} \mathbf{k} \rightarrow 0$ .

More generally, consider the interval module  $I^{[a,b)}$  for an interval  $[a, b)$  in  $[m]$ ,

$$0 \rightarrow \cdots \rightarrow 0 \rightarrow \mathbf{k} \xrightarrow{1_{\mathbf{k}}} \mathbf{k} \xrightarrow{1_{\mathbf{k}}} \cdots \xrightarrow{1_{\mathbf{k}}} \mathbf{k} \rightarrow 0 \rightarrow \cdots \rightarrow 0.$$

Its dual  $(I^{[a,b)})^*$  is isomorphic to the interval module  $I^{(m+1-b, m+1-a]}$  where  $(m+1-b, m+1-a]$  is the interval  $\{i \in [m] : m+1-b < i \leq m+1-a\}$ .

The following proposition follows from the previous two observations and Theorem 6.16.

**Corollary 8.2.** Let  $V$  be an  $[m]$ -module such that  $\dim V_i < \infty$  for all  $i \in [m]$ . Then

$$V^* \cong \bigoplus_{[a,b) \in B(V)} I^{(m+1-b, m+1-a]}.$$

In particular,  $B(V^*) = \{(m+1-b, m+1-a) : [a, b) \in B(V)\}$ .

### 8.2 Simplicial Cohomology

In the following we are working over the field  $\mathbf{k} = \mathbb{Z}_2$ .

**Definition 8.3.** Let  $n \geq 0$  be an integer and  $K$  a simplicial complex. The **vector space of  $n$ -cochains in  $K$**  is the dual vector space  $C^n(K) = C_n(K)^* = \text{Hom}(C_n(K), \mathbb{Z}_2)$ .

---

<sup>5</sup>This is in fact an if and only if statement but the converse is non-trivial to prove.

There is a **coboundary operator**  $\partial^n: C^n(K) \rightarrow C^{n+1}(K)$  defined on  $\alpha: C_n(K) \rightarrow \mathbb{Z}_2$  by

$$\partial^n(\alpha)(c) = \alpha(\partial_{n+1}(c)),$$

where  $\partial_n$  is the boundary operator in homology. If  $\partial^n(\alpha) = 0$ , then we say that  $\alpha$  is an  **$n$ -cocycle**, and if  $\alpha = \partial^{n-1}(\beta)$ , then we say that  $\alpha$  is an  **$n$ -coboundary**. The  $n$ -cocycles and  $n$ -coboundaries define vector spaces  $Z^n(K)$  and  $B^n(K)$ , respectively.

We immediately observe that

$$\partial^{n+1} \circ \partial^n(\alpha)(c) = \alpha(\partial_{n+1} \circ \partial_{n+2}(c)) = \alpha(0) = 0. \quad (5)$$

It follows that  $B^n(K) \subseteq Z^n(K)$ .

**Definition 8.4.** The  **$n$ -th simplicial cohomology vector space** of a simplicial complex  $K$  is the quotient vector space  $H^n(K) = Z^n(K)/B^n(K)$ .

For every  $n$ -simplex  $\sigma$  in  $K$  we have a dual  $\sigma^*: C_n(K) \rightarrow \mathbb{Z}_2$  given by

$$\sigma^*(\tau) = \begin{cases} 1 & \text{if } \sigma = \tau, \\ 0 & \text{otherwise.} \end{cases}$$

By representing the coboundary operator in all dimensions in the bases given by the duals of simplices, we can apply Algorithm 1 to compute  $H^n(K)$  in all dimensions. For the moment we shall order the columns and rows in the opposite order of what we would do in homology. The reason for doing this will become clear when we turn to *persistent* cohomology.

**Example 8.5.** Consider the simplicial complex in Fig. 30. For  $\sigma = \{3, 4\} \in K$  we get  $\partial^1(\sigma^*) = \tau = \{2, 3, 4\}^*$ , and for  $v = \{2\}$  we get  $\partial^0(v^*) = \{2, 3\}^* + \{2, 4\}^* + \{1, 2\}^*$ . Representing the coboundary operator as mentioned above yields,

$$M = \begin{matrix} & \begin{matrix} 234 & 34 & 23 & 24 & 13 & 12 & 4 & 3 & 2 & 1 \end{matrix} \\ \begin{matrix} 234 \\ 34 \\ 23 \\ 24 \\ 13 \\ 12 \\ 4 \\ 3 \\ 2 \\ 1 \end{matrix} & \left( \begin{matrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{matrix} \right) \end{matrix}$$

Applying Algorithm 1 to the above matrix yields the following  $R$  and  $V$ , respectively:

$$M = \begin{matrix} & \begin{matrix} 234 & 34 & 23 & 24 & 13 & 12 & 4 & 3 & 2 & 1 \end{matrix} \\ \begin{matrix} 234 \\ 34 \\ 23 \\ 24 \\ 13 \\ 12 \\ 4 \\ 3 \\ 2 \\ 1 \end{matrix} & \left( \begin{matrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{matrix} \right), \quad \begin{matrix} & \begin{matrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{matrix} \end{matrix}$$

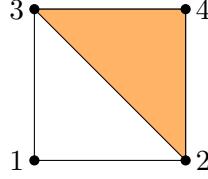


Figure 30: The simplices appear in the following order:  $\{1\}, \{2\}, \{3\}, \{4\}, \{1, 2\}, \{1, 3\}, \{2, 4\}, \{2, 3\}, \{3, 4\}, \{1, 2, 3, 4\}$ . See Example 8.5.

Using vector notation as in Section 3.1.2, we find bases for  $Z^2(K), B^2(K), Z^1(K), B^1(K), Z^0(K)$ :

$$\begin{aligned}\widetilde{\Sigma_{Z^2}} &= \{(1000000000)^T\} \\ \Sigma_{B^2} &= \{(1000000000)^T\} \\ \widetilde{\Sigma_{Z^1}} &= \{(0110000000)^T, (0101000000)^T, (0000100000)^T, (0000010000)^T\} \\ \Sigma_{B^1} &= \{(0101000000)^T, (0110100000)^T, (0011010000)^T\} \\ \widetilde{\Sigma_{Z^0}} &= \{(0000001111)^T\}.\end{aligned}$$

By means of Lemma 2.20 we extend  $\Sigma_{B^i}$  to a basis  $\Sigma_{Z^i}$  for  $Z^i(K)$ :

$$\Sigma_{Z^2} = \Sigma_{B^2}, \quad \Sigma_{Z^1} = \{(0110000000)^T\} \cup \Sigma_{B^1}, \quad \Sigma_{Z^0} = \{(0000001111)^T\}.$$

We conclude that  $H^2(K) = 0$ ,  $H^1(K) \cong \mathbb{Z}_2$ , and  $H^0(K) \cong \mathbb{Z}_2$ .

The fact that  $H^n(K) \cong H_n(K)$  can easily be proven using linear algebra:

$$\dim H^n(K) = \dim Z^n(K) - \dim B^n(K) = \dim Z_n(K) - \dim B_n(K) = \dim H_n(K). \quad (6)$$

We shall give an alternative proof. First we need the following characterization of cycles and cocycles.

### Lemma 8.6.

1.  $\gamma \in Z^n(K)$  if and only if  $\gamma(c) = 0$  for all  $c \in B_n(K)$ ,
2.  $\gamma \in B^n(K)$  if and only if  $\gamma(c) = 0$  for all  $c \in Z_n(K)$ .

*Proof.*

1. Assume  $\gamma \in Z^n(K)$  and  $c = \partial_{n+1}(d)$ . Then  $\gamma(c) = \gamma(\partial_{n+1}(d)) = \partial^n(\gamma)(d) = 0$ . Conversely, if  $\gamma$  is not in  $Z^n(K)$ , then there must exist some  $d$  such that  $0 \neq \partial^n(\gamma)(d) = \gamma(\partial_n(d))$ .
2. Assume  $\gamma(c) = 0$  for all  $c \in Z_n(K)$ . Define  $\beta: B_{n-1}(K) \rightarrow \mathbb{Z}_2$  by  $\beta(d) = \gamma(c)$  where  $d = \partial_n(c)$ . This is well-defined, because if  $d = \partial_n(c) = \partial_n(c')$ , then  $\partial_n(c - c') = 0$  and thus  $c - c' \in Z_n(K)$ . It follows that  $\gamma(c) = \gamma(c')$ . By extending a basis for  $B_{n-1}(K)$  to a basis for  $C_{n-1}(K)$ , we can extend  $\beta$  to a linear map  $\beta: C_{n-1}(K) \rightarrow \mathbb{Z}_2$  satisfying

$$\partial^{n-1}(\beta)(c) = \beta(\partial_n(c)) = \gamma(c).$$

We conclude that  $\gamma = \partial^{n-1}(\beta)$ . The converse statement is straightforward. □

**Lemma 8.7.** *The morphism  $h: H^n(K) \rightarrow H_n(K)^*$  given by  $h([\alpha])[c] = \alpha(c)$  is well-defined and an isomorphism.*

*Proof.* Let  $\alpha = \alpha' + \partial^{n-1}(\beta)$ . Then,

$$h([\alpha])[c] = \alpha(c) = \alpha'(c) + \partial^{n-1}(\beta(c)) = \alpha'(c) + \beta(\partial_n(c)) = \alpha'(c),$$

where the last equality follows from  $c$  being an  $n$ -cycle. This shows that  $h$  is independent of the representative cocycle. Next, let  $c = c' + \partial_{n+1}(d)$ ,

$$\alpha(c) = \alpha(c') + \alpha(\partial_{n+1}(d)) = \alpha(c') + \partial^n(\alpha)(d) = \alpha(c'),$$

since  $\alpha$  is an  $n$ -cocycle. This shows that  $h$  is well-defined.

- $h$  is injective: If  $h([\alpha]) = 0$ , then  $\alpha(c) = 0$  for all  $c \in Z_n(K)$ . It follows from Lemma 8.6 that  $\alpha \in B^n(K)$ , hence  $[\alpha] = 0$ .
- $h$  is surjective: Let  $f: H_n(K) \rightarrow \mathbb{Z}_2$ . Pre-composing  $f$  with the linear map  $Z_n(K) \rightarrow H_n(K)$ , defines a linear map  $f': Z_n(K) \rightarrow \mathbb{Z}_2$ . By extending a basis for  $Z_n(K)$  to a basis for  $C_n(K)$ , we can lift  $f'$  to a linear map  $f': C_n(K) \rightarrow \mathbb{Z}_2$ . Since  $f'(b) = f([b]) = f(0) = 0$  for  $b \in B_n(K)$ , it follows that  $f' \in Z^n(K)$  by Lemma 8.6. Hence  $h([f']) = f$ .

□

*Remark 8.8.* The above morphism fails to be an isomorphism when working with homology and cohomology over the integers. Can you see what part of the argument fails?

Since  $\dim V = \dim V^*$  for  $V$  finite-dimensional, it follows that

$$H_n(K) \cong H_n(K)^* \cong H^n(K).$$

### 8.3 Persistent Cohomology

For an inclusion  $f: K_i \subseteq K_j$  of simplicial complexes, there is an induced map  $f^*: H^n(K_j) \rightarrow H^n(K_i)$  given by

$$f^*([\alpha])(c) = \alpha(f(c)).$$

We leave it as an exercise to verify that this map is well-defined.

To a filtration  $K_1 \subseteq K_2 \subseteq \dots \subseteq K_m$  we thus get the following **cohomology persistence module**:

$$H^n(K) : \quad H^n(K_m) \rightarrow H^n(K_{m-1}) \rightarrow \dots \rightarrow H^n(K_2) \rightarrow H^n(K_1).$$

Let  $H_n(K)$  denote the associated homology persistence module,

$$H_n(K) : \quad H_n(K_1) \rightarrow H_2(K) \rightarrow \dots \rightarrow H_n(K_{m-1}) \rightarrow H_n(K_m).$$

**Proposition 8.9.** *The persistence modules  $H^n(K)$  and  $H_n(K)^*$  are isomorphic.*

*Proof.* For  $i \in [m]$ , define  $h_i: H^n(K_i) \rightarrow H_n(K_i)^*$  by  $h_i([\alpha])(c) = \alpha(c)$ . That yields the following commutative diagram for all  $i \leq j$

$$\begin{array}{ccc} H^n(K_j) & \longrightarrow & H^n(K_i) \\ \cong \downarrow h_j & & \cong \downarrow h_i \\ H_n(K_j)^* & \longrightarrow & H_n(K_i)^* \end{array}$$

where the vertical morphisms are isomorphisms by Lemma 8.7. □

The following theorem is an immediate consequence of the previous lemma and Corollary 8.2.

**Theorem 8.10.** *There is a bijection between the barcodes  $B(H^n(K))$  and  $B(H_n(K))$  given by*

$$B(H_n(K)) \ni [a, b] \leftrightarrow (m + 1 - b, m + 1 - a) \in B(H^n(K)).$$

*Remark 8.11.* In fact, if we view  $H^n(K)$  as an persistence module indexed over  $[m]$  with the opposite ordering,

$$H^n(K_1) \leftarrow H^n(K_2) \leftarrow \cdots H^n(K_{m-1}) \leftarrow H^n(K_m),$$

then the barcode of  $H^n(K)$  coincides with the barcode of  $H_n(K)$ .

### 8.3.1 Computation

It is tempting to think that persistent cohomology can be computed by applying the standard algorithm to the coboundary matrix as in Example 8.5. This turns out to almost be the case.

**Example 8.12.** Returning to Example 8.5 and pairing vectors based on their lowest non-zero entry as in the standard algorithm for persistent homology we get the pairs

$$\begin{array}{ll} (1000000000)^T \leftrightarrow (1000000000)^T \\ (0101000000)^T \leftrightarrow (0101000000)^T & (0000100000)^T \leftrightarrow (0110100000)^T \\ (0000010000)^T \leftrightarrow (0011010000)^T & (0110000000)^T \leftrightarrow \emptyset \\ (0000001111)^T \leftrightarrow \emptyset \end{array}$$

Let us start with the first pairing, which is between  $\{2, 3, 4\}^*$  and the column corresponding to  $\{3, 4\}^*$ . Since the duals of simplices appear in the opposite order of the simplices, this results in an interval  $[1, 2)$  in dimension 2 (no, this is not a typo). Similarly we get intervals  $[4, 7), [5, 8), [6, 9)$  and  $[3, \infty)$  in dimension 1. Lastly we get an interval  $[10, \infty)$  in dimension 0.

The barcode of the previous example can easily be transformed into the barcode of persistent cohomology, and thus persistent homology. One can prove that applying the standard algorithm to the boundary matrix as above computes the barcode of the following persistence module in *relative cohomology*:

$$H^n(K, K_{m-1}) \rightarrow H^n(K, K_{m-2}) \rightarrow \cdots \rightarrow H^n(K, K_1) \rightarrow H^n(K).$$

The barcode of relative cohomology corresponds to the barcode of  $H^n(K)$  according to the following table:

Relative Cohomology	Cohomology	Homology
$[a, b)$ in dim. $i$	$(a, b]$ in dim. $i - 1$	$[m + 1 - b, m + 1 - a)$ in dim. $i - 1$
$[a, \infty)$ in dim. $i$	$(-\infty, a]$ in dim. $i$	$[m + 1 - a, \infty)$ in dim. $i$

The reader may easily verify that transforming the barcode of Example 8.12 into the homology barcode yields the expected result.

### 8.3.2 Clearing

The salient point of the last subsection is that the persistent homology barcode can be computed by means of cohomology. A naïve implementation will not improve computational speed much compared to the standard algorithm. However, the speed improves dramatically if we employ the clearing technique discussed in the previous lecture.

Just as in the homology case, we know that if the column corresponding to the  $n$ -cochain  $\sigma^*$  forms an  $n$ -coboundary with lowest non-zero entry at row  $i$ , then the column corresponding to the  $(n+1)$ -cochain added at time  $i$  will be reduced to 0 under Algorithm 1. This means that if we have reduced all the  $(n-1)$ -cochains to form a basis for  $B^n(K)$ , then we avoid reducing  $|B^n(K)|$  columns when computing bases for  $Z^n(K)$  and  $B^{n+1}(K)$ . This is the whole idea behind **clearing** and can be summarized as follows:

- Let  $K$  be a filtered simplicial complex with  $\dim K = d$ .
- Let  $D$  be the coboundary operator in matrix form, with the columns ordered in the opposite order of the filtration (as in Example 8.5).
- Use the standard algorithm to reduce the columns of  $D$  corresponding to 0-simplices in order to obtain  $\widehat{\Sigma_{B^1}}$  and  $\widehat{\Sigma_{Z^0}}$ .
- For every  $R_j \in \widehat{\Sigma_{B^1}}$ , zero out column number  $\text{low}(R_j)$  in  $D$ . Reduce the 1-cochains of the resulting boundary matrix to obtain the set  $\widehat{\Sigma_{B^2}}$ . A basis  $\widehat{\Sigma_{Z^1}}$  for  $Z^1(K)$  is given by the set  $\widehat{\Sigma_{B^1}}$  together with all columns  $V_i$  corresponding to zero columns  $R_i$  that were not zeroed out prior to the reduction.
- Continue the same way until a bases for  $Z^{d-1}(K)$  and  $B^d(K)$  are computed.

The total number of columns that need not be reduced are

$$\dim B^1(K) + \dim B^2(K) + \dots + \dim B^{d-1}(K). \quad (7)$$

The difference with homology is that we avoid zeroing out a large number of columns corresponding to  $d$ -simplices. Let us return to Example 7.1 in the case of cohomology.

**Example 8.13.** When working with cohomology we know that  $Z^3(K) = C^3(K)$  if restricted to a simplicial complex without 4-simplices. Hence, if we were to naïvely reduce the coboundary matrix without using clearing we would have to do a total number of

$$\dim C^0(\text{VR}_\infty(P)) + \dim C^1(\text{VR}_\infty(P)) + \dim C^2(\text{VR}_\infty(P)) = \binom{100}{1} + \binom{100}{2} + \binom{100}{3} = 166750$$

column reductions. This number is much lower than for the naïve homology algorithm, but the computations are not faster in practice because the columns to be reduced are much larger. Furthermore, the total number of columns to be zeroed out equals

$$\dim Z^2(\text{VR}_\infty(P)) + \dim Z^1(\text{VR}_\infty(P)) + \dim Z^0(\text{VR}_\infty(P)) = \binom{99}{2} + \binom{99}{1} + 1 = 4951.$$

With clearing on the other hand, we need not zero out the following number of columns:

$$\dim B^1(\text{VR}_\infty(P)) + \dim B^2(\text{VR}_\infty(P)) = \binom{99}{1} + \binom{99}{2} = 4950.$$

In conclusion: 161800 column reductions but only a single column will be zeroed out.

The previous example illustrates one of the reasons why cohomology *with clearing* typically performs much faster than homology with clearing: the number of columns to zero out becomes small, and it is comparatively much cheaper to reduce a column to a boundary. Furthermore, it turns out that in practice a large number of the boundary columns need no reduction. More on this can be found in [1].

#### 8.4 Exercises

1. Verify that a simplicial map  $f: K \rightarrow L$  induces a well-defined map  $f^*: H^n(L) \rightarrow H^n(K)$  given by  $f^*([\alpha])(c) = \alpha(f(c))$ . Show that the diagram in Proposition 8.9 commutes.
2. Let  $K$  be the simplicial complex defined by an  $n$ -simplex and all its faces. Determine  $\dim Z_i(K)$ ,  $\dim B_i(K)$ ,  $\dim Z^i(K)$  and  $\dim B^i(K)$  for all  $i$ .
3. Show that  $H^n(K) \cong H_n(K)$  by proving Eq. (6).
4. Consider the filtration  $K_1 \subseteq K_2 \subseteq \cdots \subseteq K_m = K$ , where  $K_{i+1} = K_i \cup \{\sigma_{i+1}\}$ , together with the associated relative cohomology persistence module

$$H^n(K, K_{m-1}) \rightarrow H^n(K, K_{m-2}) \rightarrow \cdots \rightarrow H^n(K, K_2) \rightarrow H^n(K, K_1).$$

By definition,  $C^n(K, K_i) := C_n(K, K_i)^*$ .

- (a) Show that  $C^n(K, K_l) = \text{Span}\{\sigma_t^* \mid \dim \sigma_t = n \text{ and } t > l\}$
- (b) Order the basis elements as  $\sigma_m^* < \sigma_{m-1}^* \cdots < \sigma_1^*$  and consider the matrix representation  $D$  of the coboundary operator

$$\partial^*: C^0(K, K_l) \oplus \cdots \oplus C^{\dim K}(K, K_l) \rightarrow C^0(K, K_l) \oplus \cdots \oplus C^{\dim K}(K, K_l)$$

with respect to this ordering (as in Example 8.5). Convince yourself that running the standard algorithm on  $D$  computes a basis for  $H^0(K, K_l) \oplus \cdots \oplus H^{\dim K}(K, K_l)$ .

- (c) Let  $l = 1$  in (b), and follow the line of reasoning given in Section 7.1 to conclude that the pairings obtained from the applying the standard algorithm to  $D$  produces the barcode for the relative cohomology persistence module.

## 9 Stability of Persistent Homology

In this lecture we shall discuss metrics on barcodes/persistence diagrams and discuss to what extent persistent homology is stable under perturbation of the input data.

### 9.1 The Bottleneck Distance

Let  $\mathcal{C}$  and  $\mathcal{D}$  be multisets of intervals  $\langle a, b \rangle$  in  $\mathbb{R}$ . Here the notation  $\langle a, b \rangle$  denotes that the interval can be any well-defined member of  $\{[a, b], [a, b), (a, b], (a, b)\}$ . We define a matching between  $\mathcal{C}$  and  $\mathcal{D}$  to be a collection of pairs  $\chi = \{(I, J) \in \mathcal{C} \times \mathcal{D}\}$  where each  $I$  and each  $J$  can occur in at most one pair. Equivalently, a matching is a bijection between a subset of  $\mathcal{C}$  and a subset of  $\mathcal{D}$ . If  $(I, J) \in \chi$ , then we say that  $I$  is **matched** to  $J$ . If  $I$  does not appear in a pair, then we say that  $I$  is **unmatched**.

**Example 9.1.** Let  $\mathcal{C} = \{I_1, I_2\}$  and  $\mathcal{D} = \{J\}$ . Then the pair  $(I_1, J)$  defines a matching where  $I_2$  is unmatched. Note that  $\{(I_1, J), (I_2, J)\}$  does not define a matching.

Define the **cost**  $c(I, J)$  of matching  $I = \langle a, b \rangle$  to  $J = \langle c, d \rangle$  to be

$$c(I, J) = \max(|c - a|, |d - b|),$$

and define the cost  $c(I)$  of not matching  $I$  to be

$$c(I) = (b - a)/2.$$

Given this we define the **cost of a matching**  $\chi$  to be

$$c(\chi) := \max \left( \sup_{(I, J) \in \chi} c(I, J), \sup_{\text{unmatched } I \in \mathcal{C} \cup \mathcal{D}} c(I) \right). \quad (8)$$

We say that  $\chi$  is an  $\epsilon$ -**matching** if  $c(\chi) \leq \epsilon$ .

*Remark 9.2.* A geometric interpretation of this is as follows: consider the intervals of  $\mathcal{C}$  and  $\mathcal{D}$  as points in  $\mathbb{R}^2$ . An  $\epsilon$ -matching then pairs up points  $p$  and  $q$ , corresponding to intervals of  $\mathcal{C}$  and  $\mathcal{D}$  respectively, such that  $\|p - q\|_\infty \leq \epsilon$ , and such that any unmatched point is at most  $\epsilon$  away from the diagonal in the  $l_\infty$ -norm.

**Lemma 9.3.** *If there exists an  $\epsilon$ -matching between  $\mathcal{C}$  and  $\mathcal{D}$ , and an  $\epsilon'$ -matching between  $\mathcal{D}$  and  $\mathcal{E}$ , then the compositions of matching yields an  $(\epsilon + \epsilon')$ -matching between  $\mathcal{C}$  and  $\mathcal{E}$ .*

*Proof.* Let  $I = \langle a_1, b_1 \rangle \in \mathcal{C}$ ,  $J = \langle a_2, b_2 \rangle \in \mathcal{D}$  and  $K = \langle a_3, b_3 \rangle \in \mathcal{E}$  such that  $I$  is matched to  $J$ , and  $J$  is matched to  $K$ . Then  $|a_3 - a_1| \leq |a_3 - a_2| + |a_2 - a_1| \leq \epsilon + \epsilon'$ , and similarly for  $|b_3 - b_1|$ .

If  $I$  and  $J$  are as above, but  $J$  is unmatched in the second matching, then by assumption we have  $b_2 \leq a_2 + 2\epsilon'$ . It follows that

$$a_1 + 2\epsilon + 2\epsilon' \geq a_2 + \epsilon + 2\epsilon' \geq b_2 + \epsilon \geq b_1,$$

and thus  $c(I) \leq \epsilon + \epsilon'$ .

A similar argument applies if  $J$  is matched to  $K$  under the second matching, but unmatched under the first matching.  $\square$

It follows from the definition of a matching and the previous lemma that the following defines a distance.

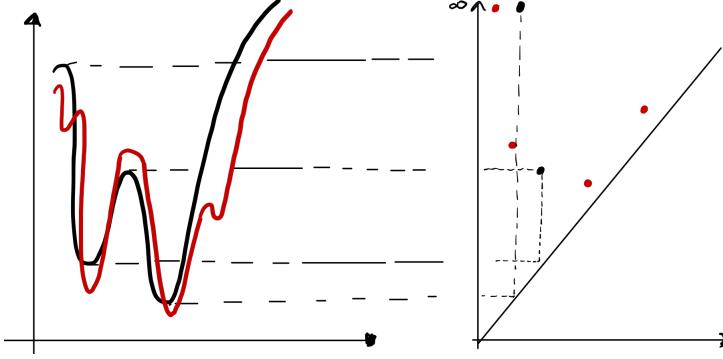


Figure 31: Left: A graph (black) and its perturbation (red). Right: The persistence diagrams of the 0-th sublevel set persistent homology of the two graphs.

**Definition 9.4.** The **bottleneck distance** between  $\mathcal{C}$  and  $\mathcal{D}$  is

$$d_B(\mathcal{C}, \mathcal{D}) = \inf\{c(\chi) : \chi \text{ is a matching between } \mathcal{C} \text{ and } \mathcal{D}\}.$$

**Example 9.5.** Let  $\mathcal{C} = \{[0, 6), [4, 6)\}$  and  $\mathcal{D} = \{[2, 8), [10, 12)\}$ . The trivial matching which leaves everything unmatched has a cost of the half the length of the longest interval:  $(8 - 2)/2 = 3$ . Hence,  $d_B(\mathcal{C}, \mathcal{D}) \leq 3$ . In order to improve this bound we need to match the interval  $[2, 8)$  to either  $[0, 6)$  or  $[4, 6)$ . These two pairings come at the same expense

$$c([0, 6), [2, 8)) = c([4, 6), [2, 8)) = 2.$$

If we match  $[0, 6)$  with  $[2, 8)$ , then we can either match  $[4, 6)$  to  $[10, 12)$  or leave both of the unmatched. We see that leaving them both unmatched comes at the lowest cost of

$$\max\{c([0, 6), [2, 8)), c([4, 6), [2, 8)), c([10, 12))\} = 2.$$

If we however match  $[4, 6)$  to  $[2, 8)$ , then the best we can do is

$$\max\{c([4, 6), [2, 8)), c([0, 6)), c([10, 12))\} = 3.$$

We conclude that  $d_B(\mathcal{C}, \mathcal{D}) = 2$ .

The previous example was somewhat tedious but luckily there are fast algorithms for computing the bottleneck distance in practice. By recasting the problem as an instance of a *bi-partite matching problem*, state-of-the-art algorithms compute  $d_B(\mathcal{C}, \mathcal{D})$  in  $O(n^{1.5} \log n)$  where  $n = |\mathcal{C}| + |\mathcal{D}|$  [18].

### 9.1.1 Stability of Persistent Homology

One of the key properties of persistent homology is the fact that it is *stable* in the bottleneck distance. As an example, consider the two graphs shown to the left in Fig. 31. Associated to these graphs we get persistence diagrams in dimension 0 by considering sublevel sets and applying homology in dimension 0. The persistence diagram of the perturbed graph has two points close to the points of the original persistence diagrams, together with two points close to the diagonal (noise). This suggests a matching where the latter two points (intervals) are left unmatched, and the two other points (intervals) will be matched with a cost bounded by the perturbation in the  $l_\infty$ -norm.

More generally, let  $X$  be a topological space equipped with two functions  $f, g: X \rightarrow \mathbb{R}$ . Associated to these functions we have  $\mathbb{R}$ -modules  $M^f$  and  $M^g$  given by

$$M_t^f = H_i(f^{-1}(-\infty, t]) \quad M_t^g = H_i(g^{-1}(-\infty, t]). \quad (9)$$

Assuming that  $\dim M_t^f < \infty$  and  $\dim M_t^g < \infty$  for all  $t$ , which ensures that we have well-defined barcodes by Theorem 6.19, the following is true

**Theorem 9.6.**

$$d_B(B(M^f), B(M^g)) \leq \|f - g\|_\infty.$$

We shall prove this theorem for sufficiently *tame* functions.

Similar stability results apply to point sets: Fig. 32 shows a noisy sample from the circle and its perturbation, together with the persistence diagrams in dimension 1 of the associated to the Vietoris–Rips filtrations. For a finite point set  $P$ , let  $H_i(\text{VR}(P))$  denote the  $\mathbb{R}$ -indexed persistence module defined at  $t$  by  $H_i(\text{VR}_t(P))$  and for which the morphism  $H_i(\text{VR}_s(P)) \rightarrow H_i(\text{VR}_t(P))$  is induced by the inclusion  $\text{VR}_s(P) \subseteq \text{VR}_t(P)$ .

**Theorem 9.7.** *Let  $P$  and  $Q$  be finite sets of points in  $\mathbb{R}^d$ , and  $\sigma: P \rightarrow Q$  a bijection such that  $\|p - \sigma(p)\| \leq \epsilon$  for all  $p \in P$ . Then,*

$$d_B(B(H_i(\text{VR}(P))), B(H_i(\text{VR}(Q)))) \leq \epsilon.$$

By means of the *Gromov–Hausdorff distance* this theorem can be strengthened to include point sets that need not be embedded in a common space nor have the same cardinality. We will not concern ourselves with the more general result.

A stability theorem for the Čech filtration in the case that  $P$  and  $Q$  need not have the same cardinality follows from Theorem 9.6. Let  $H_i(\text{Cech}(P))$  denote the  $\mathbb{R}$ -module associated to the Čech filtration. Recall the notation  $P_r := d_P^{-1}(-\infty, r]$  introduced in Section 5.1.

**Theorem 9.8.** *Let  $P$  and  $Q$  be finite sets of points in  $\mathbb{R}^d$ . If there exists an  $\epsilon \geq 0$  such that  $P \subseteq Q_\epsilon$  and  $Q \subseteq P_\epsilon$ , then*

$$d_B(B(H_i(\text{Cech}(P))), B(H_i(\text{Cech}(Q)))) \leq \epsilon.$$

*Proof.* Define functions  $d_P, d_Q: \mathbb{R}^d \rightarrow \mathbb{R}$  by

$$d_P(x) = \min_{p \in P} \|x - p\| \quad d_Q(x) = \min_{q \in Q} \|x - q\|.$$

Fix an  $x \in \mathbb{R}^2$  and let  $p \in P$  be a point closest to  $x$ . Then, from the assumption that  $P \subseteq Q_\epsilon$ , there exists some  $q \in Q$  with  $\|p - q\| \leq \epsilon$ . This shows that  $d_P(x) \leq d_Q(x) + \epsilon$ , and symmetrically,  $d_Q(x) \leq d_P(x) + \epsilon$ . We conclude that  $\|d_P - d_Q\|_\infty \leq \epsilon$ .

Now define  $\mathbb{R}$ -modules  $M^{d_P}$  and  $M^{d_Q}$  as above. From Theorem 4.11 we know that  $M^{d_P} \cong H_i(\text{Cech}(P))$  and  $M^{d_Q} \cong H_i(\text{Cech}(Q))$ . Since isomorphic  $\mathbb{R}$ -modules necessarily have the same barcode, we conclude that

$$d_B(B(H_i(\text{Cech}(P))), B(H_i(\text{Cech}(Q)))) = d_B(B(M^{d_P}), B(M^{d_Q})) \leq \|d_P - d_Q\|_\infty \leq \epsilon.$$

□

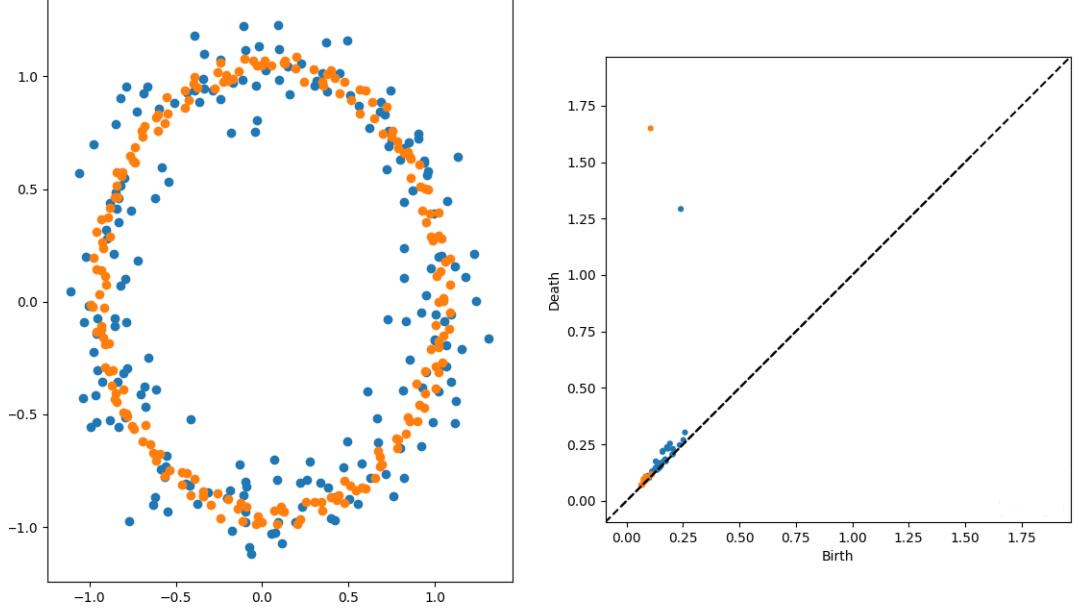


Figure 32: Left: A sample from a circle and its perturbation. Right: The associated persistence diagrams in dimension 1 computed using Ripser.

### 9.1.2 Generalizations

The bottleneck distance is only sensitive to the maximum over a matching. There are other "more discriminative" distances which are more frequently used in data analysis, although persistence diagrams do not exhibit the same type of general stability in those cases as with the bottleneck distance. One such distance is the **degree  $q$  Wasserstein distance**,

$$W_q(\mathcal{C}, \mathcal{D}) = \inf_{\text{matchings } \chi \text{ between } \mathcal{C} \text{ and } \mathcal{D}} \left( \sum_{(I,J) \in \chi} c(I, J)^q + \sum_{\text{unmatched } I \in \mathcal{C} \cup \mathcal{D}} c(I)^q \right)^{1/q},$$

where  $q$  is some positive real number. The bottleneck distance can be obtained as a limit in the following way:

$$\lim_{q \rightarrow \infty} W_q(\mathcal{C}, \mathcal{D}) = d_B(\mathcal{C}, \mathcal{D}).$$

**Stability.** We conclude by stating a recent result on stability. Let  $K$  be a simplicial complex, and let  $f: K \rightarrow \mathbb{R}$  be a **monotone** function defined on the simplices of  $K$ , i.e.,  $f(\sigma) \leq f(\tau)$  if  $\sigma \subseteq \tau$ . This defines a filtration of  $K$  by letting  $K_t = \{\sigma \in K : f(\sigma) \leq t\}$ , and a persistence module  $M_t^f := H_i(K_t)$  as above. Examples of such filtrations are the Čech, Vietoris–Rips, and alpha complexes.

**Theorem 9.9** ([28, 4]). *For monotone functions  $f, g: K \rightarrow \mathbb{R}$ ,*

$$W_q(B(M^f), B(M^g)) \leq \left( \sum_{\sigma \in K, \dim \sigma \in \{i, i+1\}} |f(\sigma) - g(\sigma)|^q \right)^{1/q}.$$

## 9.2 Interleavings

We will now introduce the theory of interleavings and an associated interleaving distance between persistence modules. Importantly, and as we will see, the interleaving distance coincides with the bottleneck distance, and it is easy to show stability in the interleaving distance. Hence, our stability results will follow from this equivalence.

**Discrete setting** For the moment, let us assume that we are working with persistence module indexed by the integers. A **0-interleaving** is nothing more than an isomorphism, i.e. a collection of morphisms  $\{M_i \rightarrow N_i\}$  and a collection of morphisms  $\{N_i \rightarrow M_i\}$  such that the following diagrams commute:

$$\begin{array}{ccccccc} \cdots & \longrightarrow & M_i & \longrightarrow & M_{i+1} & \longrightarrow & M_{i+2} \longrightarrow \cdots \\ & & \textcolor{red}{\downarrow} & & \textcolor{blue}{\downarrow} & & \textcolor{blue}{\downarrow} \\ \cdots & \longrightarrow & N_i & \longrightarrow & N_{i+1} & \longrightarrow & N_{i+2} \longrightarrow \cdots \end{array}$$

A 1-interleaving is a collection of slanted morphisms such that the following diagram commutes:

$$\begin{array}{ccccccc} \cdots & \longrightarrow & M_i & \longrightarrow & M_{i+1} & \longrightarrow & M_{i+2} \longrightarrow \cdots \\ & \textcolor{red}{\searrow} & \textcolor{blue}{\swarrow} & \textcolor{red}{\searrow} & \textcolor{blue}{\swarrow} & \textcolor{red}{\searrow} & \textcolor{blue}{\swarrow} \\ \cdots & \longrightarrow & N_i & \longrightarrow & N_{i+1} & \longrightarrow & N_{i+2} \longrightarrow \cdots \end{array}$$

Note that a 1-interleaving requires that the morphism  $M_i \rightarrow M_{i+2}$  factors through  $N_{i+1}$  in a natural way. Intuitively, this means that every “2-persistent” feature of  $M_i$ —a bar of length at least two that overlaps the index  $i$ —is also present in  $N_i$ , and vice versa. A 2-interleaving is a collection of morphisms  $\{M_i \rightarrow N_{i+2}\}$  and  $\{N_i \rightarrow M_{i+2}\}$  making a similar diagram commute. It should now be possible to deduce what an  $n$ -interleaving is.

**The formal definition** Now we shall assume that the modules  $M$  and  $N$  are indexed over the real numbers. Define the  $\epsilon$ -shift of  $M$  to be the persistence module  $M^\epsilon$  defined by  $M_t^\epsilon = M_{t+\epsilon}$  and  $M^\epsilon(s \leq t) = M(s + \epsilon \leq t + \epsilon)$  for all  $s \leq t \in \mathbb{R}$ . If  $f: M \rightarrow N$  is a morphism, then we get an  $\epsilon$ -shifted morphism  $f^\epsilon: M^\epsilon \rightarrow N^\epsilon$  defined by  $f_t^\epsilon = f_{t+\epsilon}$ .

**Example 9.10.** The  $\epsilon$  shift of the interval module  $I^{[a,b)}$  is  $I^{[a-\epsilon,b-\epsilon)}$ .

Let  $\eta_M^\epsilon: M \rightarrow M^\epsilon$  be the morphism whose restriction to each  $M_t$  is the internal morphism  $M(t \leq t + \epsilon)$ .

**Definition 9.11.** Given  $\epsilon \in [0, \infty)$ , an  $\epsilon$ -interleaving between  $M$  and  $N$  is a pair of morphisms  $\psi: M \rightarrow N^\epsilon$  and  $\varphi: N \rightarrow M^\epsilon$  such that  $\varphi^\epsilon \circ \psi = \eta_M^{2\epsilon}$  and  $\psi^\epsilon \circ \varphi = \eta_N^{2\epsilon}$ . We say that  $M$  and  $N$  are  $\epsilon$ -interleaved.

Explicitly, the last condition states the following two diagrams commute for all  $t \in \mathbb{R}$ :

$$\begin{array}{ccc} M_t & \xrightarrow{M(t \leq t+2\epsilon)} & M_{t+2\epsilon} \\ \psi_t \searrow & & \nearrow \varphi_{t+\epsilon} \\ & N_{t+\epsilon} & \end{array} \quad \begin{array}{ccc} & M_{t+\epsilon} & \\ \phi_t \nearrow & & \searrow \psi_{t+\epsilon} \\ N_t & \xrightarrow{N(t \leq t+2\epsilon)} & N_{t+2\epsilon} \end{array}$$

**Lemma 9.12.** *If  $M$  and  $N$  are  $\epsilon$ -interleaved, and  $N$  and  $L$  are  $\epsilon'$ -interleaved, then  $M$  and  $L$  are  $(\epsilon + \epsilon')$ -interleaved.*

*Proof.* Let  $\psi: M \rightarrow N^\epsilon$ ,  $\varphi: N \rightarrow M^\epsilon$ ,  $\psi': N \rightarrow L^{\epsilon'}$  and  $\varphi': L \rightarrow N^{\epsilon'}$  be the respective morphisms satisfying the interleaving condition. They yield a morphism  $\psi'': M \rightarrow L^{\epsilon+\epsilon'}$  through the composition

$$M_t \xrightarrow{\psi_t} N_{t+\epsilon} \xrightarrow{\psi'_{t+\epsilon}} L_{t+\epsilon+\epsilon'},$$

and a morphism  $\varphi'' \circ M \rightarrow L^{\epsilon+\epsilon'}$  given by

$$L_t \xrightarrow{\varphi'_t} N_{t+\epsilon'} \xrightarrow{\varphi_{t+\epsilon'}} M_{t+\epsilon+\epsilon'}.$$

It remains to check that  $(\varphi'')^{\epsilon+\epsilon'} \circ \psi'' = \eta_M^{2(\epsilon+\epsilon')}$  and  $(\psi'')^{\epsilon+\epsilon'} \circ \varphi'' = \eta_L^{2(\epsilon+\epsilon')}$ . The first equality follows from the following commutative diagram:

$$\begin{array}{ccccccc} M_t & \xrightarrow{M(t \leq t+2\epsilon)} & M_{t+2\epsilon} & \xrightarrow{M(t+2\epsilon \leq t+2\epsilon+2\epsilon')} & M_{t+2\epsilon+2\epsilon'} \\ & \searrow \psi_t & \nearrow \varphi_{t+\epsilon} & & \nearrow \varphi_{t+\epsilon+2\epsilon'} \\ & N_{t+\epsilon} & \xrightarrow{N(t+\epsilon \leq t+\epsilon+2\epsilon')} & N_{t+\epsilon+2\epsilon'} & & \\ & \searrow \psi'_{t+\epsilon} & \nearrow \varphi_{t+\epsilon+\epsilon'} & & & \end{array}$$

The second equality follows from a similar commutative diagram.  $\square$

It follows from the definition of an interleaving and the previous lemma that the following indeed defines a metric.

**Definition 9.13.** The **interleaving distance** between two  $\mathbb{R}$ -modules  $M$  and  $N$  is

$$d_I(M, N) = \inf\{\epsilon : \text{there exists an } \epsilon\text{-interleaving between } M \text{ and } N\}.$$

The following result is fundamental to topological data analysis and non-trivial. We defer its proof to a later lecture.

**Theorem 9.14.** *Let  $M$  and  $N$  be  $\mathbb{R}$ -modules such that  $M_t$  and  $N_t$  are finite-dimensional for all  $t \in \mathbb{R}$ . Then,*

$$d_B(B(M), B(N)) = d_I(M, N).$$

An immediate corollary is a proof of Theorem 9.6.

*Proof of Theorem 9.6.* Since  $\|f - g\|_\infty \leq \epsilon$ , we have the following commutative diagram of inclusions for every  $t \in \mathbb{R}$ :

$$\begin{array}{ccccc} f^{-1}(-\infty, t] & \longrightarrow & f^{-1}(-\infty, t+\epsilon] & \longrightarrow & f^{-1}(-\infty, t+2\epsilon] \\ \searrow & & \nearrow & & \searrow \\ g^{-1}(-\infty, t] & \longrightarrow & g^{-1}(-\infty, t+\epsilon] & \longrightarrow & g^{-1}(-\infty, t+2\epsilon]. \end{array}$$

Applying  $H_i$  yields the following commutative diagram

$$\begin{array}{ccccc}
 M_t^f & \longrightarrow & M_{t+\epsilon}^f & \longrightarrow & M_{t+2\epsilon}^f \\
 \searrow & \nearrow & \searrow & \nearrow & \searrow \\
 M_t^g & \longrightarrow & M_{t+\epsilon}^g & \longrightarrow & M_{t+2\epsilon}^g.
 \end{array}$$

It is easy to see that this defines an  $\epsilon$ -interleaving pair.  $\square$

*Proof of Theorem 9.7.* This proof is analogous to the previous proof. Label the points of  $P = \{p_1, \dots, p_m\}$  and  $Q = \{q_1, \dots, q_m\}$  such that  $\|p_i - q_i\| \leq \epsilon$ . If  $\sigma = \{p_{i_1}, \dots, p_{i_m}\}$  is a simplex of  $\text{VR}_r(P)$ , then we have by definition of the Vietoris–Rips complex that  $\text{diam}(\sigma) \leq 2r$ . From the bijection we get that the corresponding set  $\tau = \{q_{i_1}, \dots, q_{i_m}\}$  satisfies  $\text{diam}(\tau) \leq 2r + 2\epsilon$ . Hence, we have an inclusion  $\text{VR}_r(P) \subseteq \text{VR}_{r+\epsilon}(Q)$  defined by the simplicial mapping which sends the vertex  $p_i$  to  $q_i$ . Symmetrically we have an inclusion  $\text{VR}_r(Q) \subseteq \text{VR}_{r+\epsilon}(P)$ . Putting these two relations together we find

$$\begin{aligned}
 \text{VR}_r(P) &\subseteq \text{VR}_{r+\epsilon}(Q) \subseteq \text{VR}_{r+2\epsilon}(P) \\
 \text{VR}_r(Q) &\subseteq \text{VR}_{r+\epsilon}(P) \subseteq \text{VR}_{r+2\epsilon}(Q)
 \end{aligned}$$

Applying  $H_i$  to the above relations defines an  $\epsilon$ -interleaving between  $H_i(\text{VR}(P))$  and  $H_i(\text{VR}(Q))$ . It follows from Theorem 9.14 that

$$d_B(B(H_i(\text{VR}(P))), B(H_i(\text{VR}(Q)))) = d_I(H_i(\text{VR}(P)), H_i(\text{VR}(Q))) \leq \epsilon.$$

$\square$

### 9.3 Exercises

1. Let  $\mathcal{C} = \{[0, 10), [2, 8), [1, 2), [13, 14)\}$  and  $\mathcal{D} = \{[1, 8), [1, 9), [17, 18)\}$ . Compute  $d_B(\mathcal{C}, \mathcal{D})$ .
2. Verify that  $d_B(B(M), B(N)) = d_I(M, N)$  when  $M$  and  $N$  are interval modules over  $\mathbb{R}$ .
3. Show that if  $M$  and  $N$  are  $\epsilon$ -interleaved, then they are  $\epsilon'$ -interleaved for all  $\epsilon' \geq \epsilon$ .
4. Let  $P$  be a finite set of points in  $\mathbb{R}^n$  and let  $Q$  be an  $\epsilon$ -perturbation of  $P$ . That is, there exists a bijection  $\sigma: P \rightarrow Q$  such that  $\|p - \sigma(p)\| \leq \epsilon$  for all  $p$ . Show that  $d_B(H_0(\text{VR}(P)), H_0(\text{VR}(Q))) \leq \epsilon$  without using Theorem 9.7.

## 10 Bjerkevik's Proof of the Isometry Theorem

In this section we will prove Theorem 9.14 by showing that  $d_I \leq d_B$  and  $d_B \leq d_I$ . The former inequality first appeared in [21] and is rather straightforward. For the latter we will be following the work of Bjerkevik [2]. Other proofs do exist in the literature but we will focus on this particular approach because 1) it is combinatorial in nature, 2) it generalizes to multiparameter modules in a way that other approaches do not. For the **remainder of this section** we shall work under the additional assumption of a finite number of intervals. This is just for notational simplicity as essentially the same proof works in the more general setting of Theorem 9.14. Furthermore, given that assumption, we may further assume that:

1. The intervals have finite support.
2. Every interval is of the form  $[a, b)$ .

The first of these can be made without loss of generality as one can replace infinite intervals with finite intervals:  $(-\infty, b) \leftrightarrow [-u, b)$ ,  $(-\infty, \infty) \leftrightarrow [-u, u)$  and  $[a, \infty) \leftrightarrow [a, u)$  for  $u$  sufficiently large. The latter assumption can also be made without loss of generality as the following lemma shows.

**Lemma 10.1.** *Let  $M \cong \oplus_{\langle a, b \rangle \in B(M)} I^{\langle a, b \rangle}$  and let  $\widehat{M} = \oplus_{\langle a, b \rangle \in B(M)} I^{[a, b)}$ . Then,*

$$d_B(B(M), B(\widehat{M})) = 0, \quad d_I(M, \widehat{M}) = 0.$$

*Proof.* Exercise. □

Hence, from the triangle inequality it follows that  $d_B(B(M), B(N)) = d_B(B(\widehat{M}), B(\widehat{N}))$  and  $d_I(M, N) = d_I(\widehat{M}, \widehat{N})$ .

### 10.1 Part 1: $d_I \leq d_B$

Recall that for an interval  $J \subseteq \mathbb{R}$ ,  $I^J$  is defined by

$$I_t^J = \begin{cases} \mathbf{k} & \text{if } t \in J \\ 0 & \text{otherwise.} \end{cases}$$

and with  $I^J(t \leq t')$  the identity morphism whenever  $t, t' \in J$ .

**Lemma 10.2.** *Let  $J = [a, b)$  and  $K = [c, d)$  be (possibly empty) intervals in  $\mathbb{R}$ . Then, the following two statements are equivalent:*

1. *There exists a matching  $\chi$  between  $\{J\}$  and  $\{K\}$  with a cost  $c(\chi) \leq \epsilon$ ,*
2. *There exists an  $\epsilon$ -interleaving between  $I^J$  and  $I^K$ .*

*Proof.* Exercise. □

**Lemma 10.3.** *If  $M$  and  $N$  are  $\epsilon$ -interleaved, and  $M'$  and  $N'$  are  $\epsilon$ -interleaved, then  $M \oplus M'$  and  $N \oplus N'$  are  $\epsilon$ -interleaved.*

*Proof.* Let  $f: M \rightarrow N^\epsilon$  and  $g: N \rightarrow M^\epsilon$ , and  $f': M' \rightarrow (N')^\epsilon$  and  $g': N' \rightarrow (M')^\epsilon$  constitute  $\epsilon$ -interleavings. Taking the direct sums yields:

$$A := \begin{bmatrix} f & 0 \\ 0 & f' \end{bmatrix} : M \oplus M' \rightarrow N^\epsilon \oplus (N')^\epsilon$$

and

$$B := \begin{bmatrix} g & 0 \\ 0 & g' \end{bmatrix} : N \oplus N' \rightarrow M^\epsilon \oplus (M')^\epsilon.$$

It remains to check that  $A$  and  $B$  define an  $\epsilon$ -interleaving. That is straightforward:

$$B^\epsilon \circ A = \begin{bmatrix} g^\epsilon & 0 \\ 0 & (g')^\epsilon \end{bmatrix} \cdot \begin{bmatrix} f & 0 \\ 0 & f' \end{bmatrix} = \begin{bmatrix} g^\epsilon \circ f & 0 \\ 0 & (g')^\epsilon \circ f' \end{bmatrix} = \begin{bmatrix} \eta_M^{2\epsilon} & 0 \\ 0 & \eta_{M'}^{2\epsilon} \end{bmatrix} = \eta_{M \oplus M'}^{2\epsilon}.$$

Symmetrically,

$$A^\epsilon \circ B = \begin{bmatrix} f^\epsilon & 0 \\ 0 & (f')^\epsilon \end{bmatrix} \cdot \begin{bmatrix} g & 0 \\ 0 & g' \end{bmatrix} = \eta_{N \oplus N'}^{2\epsilon}.$$

□

Combining the two previous lemmas we arrive at our result:

**Theorem 10.4.** *Let  $B(M)$  and  $B(N)$  be finite multisets of finite intervals of the form  $[a, b)$ . Then  $d_I(M, N) \leq d_B(B(M), B(N))$ .*

*Proof.* Let  $\chi$  be a matching between  $d_B(B(M), B(N))$  with cost  $c(\chi) = \epsilon$ . If  $J \in B(M)$  is matched to  $K \in B(N)$ , then  $d_B(\{J\}, \{K\}) \leq c(J, K) = \epsilon$ , and by Lemma 10.2 there is an  $(\epsilon)$ -interleaving between  $I^J$  and  $I^K$ . Likewise, if  $J \in B(M) \cup B(N)$  is unmatched, then  $I^J$  is  $\epsilon$ -interleaved with the 0-module. Hence, by invoking Lemma 10.3 we conclude that

$$M \cong \left( \bigoplus_{J \in B(M) \text{ matched}} I^J \right) \bigoplus \left( \bigoplus_{J' \in B(M) \text{ unmatched}} I^{J'} \right) \bigoplus 0$$

is  $\epsilon$ -interleaved with

$$N \cong \left( \bigoplus_{K \in B(N) \text{ matched}} I^K \right) \bigoplus 0 \bigoplus \left( \bigoplus_{K' \in B(N) \text{ unmatched}} I^{K'} \right)$$

□

## 10.2 Part 2: $d_B \leq d_I$ (the algebraic stability theorem)

Proving this part is significantly more involved. We start with the following simple observation.

**Lemma 10.5.** *If  $I_1 = [a_1, b_1)$  and  $I_2 = [a_2, b_2)$  are intervals and  $f: I^{I_1} \rightarrow I^{I_2}$  is non-zero, then  $a_2 \leq a_1$  and  $b_2 \leq b_1$ .*

*Proof.* Assume  $a_2 > a_1$ . The following diagram must necessarily commute for any  $t \in [a_1, b_1]$ :

$$\begin{array}{ccc} I_{a_1}^{I_1} = \mathbf{k} & \xrightarrow{1} & \mathbf{k} = I_t^{I_1} \\ \downarrow f_{a_1} & & \downarrow f_t \\ I_{a_1}^{I_2} = 0 & \xrightarrow{0} & I_t^{I_2}. \end{array}$$

Since the morphism  $f_{a_1} = 0$ , it follows that  $f_t = 0$ . Hence  $f$  is the zero-morphism. A similar argument shows that  $b_2 \leq b_1$ . □

**Lemma 10.6.** *The morphism  $f: I^{[a_1, b_1)} \rightarrow I^{[a_2, b_2)}$  is completely determined by  $f_{a_1} \in \mathbf{k}$ .*

*Proof.* The morphism is trivial outside of  $[a_1, b_2)$  by the previous lemma. For any  $t \in [a_1, b_2)$  we have the following commutative diagram

$$\begin{array}{ccc} I_{a_1}^{[a_1, b_1)} = \mathbf{k} & \xrightarrow{1} & \mathbf{k} = I_t^{[a_1, b_1)} \\ \downarrow f_{a_1} & & \downarrow f_t \\ I_{a_1}^{[a_1, b_2)} = \mathbf{k} & \xrightarrow{1} & \mathbf{k} = I_t^{[a_1, b_2)}. \end{array}$$

Hence  $f_t = f_{a_1}$ . □

For an interval  $I = [a, b)$ , define  $\alpha(I) = a + b$ , and observe that this defines a *preorder*<sup>6</sup>  $\leq_\alpha$  on intervals:  $I \leq_\alpha J$  if  $\alpha(I) \leq \alpha(J)$ .

We shall make repeated use of the following observation:

$$(I^{[a, b)})^\epsilon = I^{[a-\epsilon, b-\epsilon]}.$$

**Lemma 10.7.** *Let  $I_1, I_2$  and  $I_3$  be intervals satisfying the following:  $\alpha(I_1) \leq \alpha(I_3)$ ,  $0 \neq f: I^{I_1} \rightarrow (I^{I_2})^\epsilon$ , and  $0 \neq g: I^{I_2} \rightarrow (I^{I_3})^\epsilon$ . Then,  $I^{I_2}$  is  $\epsilon$ -interleaved with  $I^{I_1}$  or  $I^{I_3}$ .*

*Proof.* Write  $I_1 = [a_1, b_1)$ ,  $I_2 = [a_2, b_2)$  and  $I_3 = [a_3, b_3)$ . From the above observation and Lemma 10.5 we find that,

$$\begin{aligned} a_2 &\leq a_1 + \epsilon \\ b_2 &\leq b_1 + \epsilon. \end{aligned}$$

Assume that  $I^{I_1}$  and  $I^{I_2}$  are not  $\epsilon$ -interleaved. The modules  $I^{[a_1, b_1)}$  and  $I^{[a_2, b_2)}$  are  $\epsilon$ -interleaved if  $|a_1 - a_2| \leq \epsilon$  and  $|b_1 - b_2| \leq \epsilon$ , and thus we may assume that  $a_2 < a_1 - \epsilon$  or  $b_2 < b_1 - \epsilon$ . Assuming the former:

$$\alpha(I_1) = a_1 + b_1 > a_2 + \epsilon + b_1 \geq a_2 + b_2 = \alpha(I_2),$$

and assuming the latter:

$$\alpha(I_1) = a_1 + b_1 > a_1 + b_2 + \epsilon \geq a_2 + b_2 = \alpha(I_2).$$

In either case,  $\alpha(I_1) > \alpha(I_2)$ . Assuming that  $I^{I_2}$  and  $I^{I_3}$  are not  $\epsilon$ -interleaved and by substituting  $I_2$  for  $I_1$ , and  $I_3$  for  $I_2$  in the above argument, we get  $\alpha(I_2) > \alpha(I_3)$ . In conclusion:

$$\alpha(I_1) > \alpha(I_2) > \alpha(I_3),$$

a contradiction. □

**Lemma 10.8.** *Let  $I_1 = [a_1, b_1)$ ,  $I_2 = [a_2, b_2)$ , and  $I_3 = [a_3, b_3)$  be such that  $b_1 - a_1 > 2\epsilon$ ,  $b_3 - a_3 > 2\epsilon$ , and  $\alpha(I_1) \leq \alpha(I_3)$ . If  $0 \neq f: I^{I_1} \rightarrow (I^{I_2})^\epsilon$  and  $0 \neq g: I^{I_2} \rightarrow (I^{I_3})^\epsilon$ . Then  $g^\epsilon \circ f \neq 0$ .*

---

<sup>6</sup>A partial order without the antisymmetry condition.

*Proof.* Applying Lemma 10.5 twice,

$$a_3 \leq a_1 + 2\epsilon \quad b_3 \leq b_1 + 2\epsilon.$$

Assuming for the sake of contradiction that  $g^\epsilon \circ f = 0$ , we must have that

$$[a_1, b_1) \cap [a_3 - 2\epsilon, b_3 - 2\epsilon) = \emptyset,$$

and thus  $b_3 \leq a_1 + 2\epsilon$ . Hence,  $b_1 > a_1 + 2\epsilon \geq b_3 > a_3 + 2\epsilon$ , where the first and last inequalities follow from the assumptions of the lemma. In particular,  $a_1 > a_3$  and  $b_1 > b_3$ . We conclude that

$$\alpha(I_1) = a_1 + b_1 > a_3 + b_3 = \alpha(I_3).$$

A contradiction.  $\square$

For an interval  $J \in B(M)$ , let  $\mu(J) := \{K \in B(N) : I^J \text{ and } I^K \text{ are } \epsilon\text{-interleaved}\}$ , and for a collection of intervals  $A \subseteq B(M)$ , let  $\mu(A) = \cup_{J \in A} \mu(J)$ .

**Lemma 10.9.** *Assume that  $M$  and  $N$  are  $\epsilon$ -interleaved, and let  $A \subseteq B(M)$  be a collection of intervals  $\{[a_j, b_j]\}$  satisfying  $b_j > a_j + 2\epsilon$ . Then*

$$|A| \leq |\mu(A)|.$$

Before proving this lemma we need to introduce some notation. Write,  $M = \oplus_{I \in B(M)} I^I$  and  $N = \oplus_{J \in B(N)} I^J$ , and let  $\iota_I^M : I^I \hookrightarrow M$  and  $\iota_J^N : I^J \hookrightarrow N$  denote the obvious inclusions into the respective summands. Dually, we also have projections  $\pi_I^M : M \twoheadrightarrow I^I$  and  $\pi_J^N : N \twoheadrightarrow I^J$ . For a morphism  $f : M \rightarrow N^\epsilon$ , we let  $f|_I = f \circ \iota_I^M$  and  $f_{I,J} = (\pi_J^N)^\epsilon \circ f \circ \iota_I^M$ , and similarly we define  $g|_J$  and  $g_{J,I}$  for  $g : N \rightarrow M^\epsilon$ . Schematically we have the following commutative diagram,

$$\begin{array}{ccc} M & \xrightarrow{f} & N^\epsilon \\ \iota_I^M \uparrow & f|_I \nearrow & \downarrow (\pi_J^N)^\epsilon \\ I^I & \xrightarrow{f_{I,J}} & (I^J)^\epsilon \end{array}$$

Recall that  $\eta_M^\epsilon : M \rightarrow M^\epsilon$  is the morphism defined by  $(\eta^\epsilon)_t = M(t \leq t + \epsilon)$ . Observing that  $\eta_{I^I}^{2\epsilon}$  is equal to the composition  $(\pi_I^M)^{2\epsilon} \circ \eta_M^{2\epsilon} \circ \iota_I^M$ , we arrive at the following commutative diagram whenever  $f$  and  $g$  constitute an  $\epsilon$ -interleaving between  $M$  and  $N$

$$\begin{array}{ccccc} & & \eta_M^{2\epsilon} & & \\ & M & \xrightarrow{f} & N^\epsilon & \xrightarrow{g^\epsilon} M^{2\epsilon} \\ \iota_I^M \uparrow & & & & \downarrow (\pi_I^M)^{2\epsilon} \\ I^I & \xrightarrow{\eta_{I^I}^{2\epsilon}} & & & \end{array}$$

Ultimately we get the following expression

$$\eta_{II}^{2\epsilon} = (\pi_I^M)^{2\epsilon} \circ g^\epsilon \circ f|_I \tag{10}$$

$$= (\pi_I^M)^{2\epsilon} \circ \left( \sum_{J \in B(N)} g|_J \circ \pi_J^N \right)^\epsilon \circ f|_I \tag{11}$$

$$= \sum_{J \in B(N)} (\pi_I^M)^{2\epsilon} \circ (g|_J)^\epsilon \circ (\pi_J^N)^\epsilon \circ f|_I \tag{12}$$

$$= \sum_{J \in B(N)} g_{J,I}^\epsilon \circ f_{I,J}. \tag{13}$$

Furthermore, by construction,  $0 = (\pi_{I'}^M)^{2\epsilon} \circ \eta_M^{2\epsilon} \circ \iota_I^M$  whenever  $I \neq I'$ . Hence,

$$0 = (\pi_{I'}^M)^{2\epsilon} \circ g^\epsilon \circ f|_I \quad (14)$$

$$= (\pi_{I'}^M)^{2\epsilon} \circ \left( \sum_{J \in B(N)} g|_J \circ \pi_J^N \right)^\epsilon \circ f|_I \quad (15)$$

$$= \sum_{J \in B(N)} (\pi_{I'}^M)^{2\epsilon} \circ (g|_J)^\epsilon \circ (\pi_J^N)^\epsilon \circ f|_I \quad (16)$$

$$= \sum_{J \in B(N)} g_{J,I'}^\epsilon \circ f_{I,J}. \quad (17)$$

*Proof of Lemma 10.9.* Order the elements of  $A = \{I_1, \dots, I_r\}$  such that  $\alpha(I_i) \leq \alpha(I_{i'})$  for  $i \leq i'$ , and let  $\mu(A) = \{J_1, \dots, J_s\}$ . By combining Lemma 10.7 with Eq. (13),

$$\eta_{II}^{2\epsilon} = \sum_{J \in B(N)} g_{J,I}^\epsilon \circ f_{I,J} = \sum_{J \in \mu(A)} g_{J,I}^\epsilon \circ f_{I,J} = \sum_{J \in \mu(A)} g_{J,I}^\epsilon \circ f_{I,J}$$

The second and third equality hold as  $f_{I,J} = 0$  or  $g_{J,I} = 0$  for every  $J \notin \mu(I)$ . Likewise, by assuming  $i < i'$ , Lemma 10.7 and invoking Eq. (17),

$$0 = \sum_{J \in B(N)} g_{J,I_{i'}}^\epsilon \circ f_{I_{i'},J} = \sum_{J \in \mu(A)} g_{J,I_{i'}}^\epsilon \circ f_{I_{i'},J}.$$

For  $I = [a, b] \in A$ , Lemma 10.6 states that every morphism  $h: I^I \rightarrow I^K$  is determined by the constant  $h_a \in \mathbf{k}$ . For notational convenience, let  $w(h)$  denote this constant. Clearly,  $w(f_{I,J}^\epsilon) = w(f_{I,J})$ . If  $i \leq i'$ , then Lemma 10.8 gives the first of the following two equalities

$$w(g_{J,I_{i'}}^\epsilon \circ f_{I_{i'},J}) = w(g_{J,I_{i'}}^\epsilon) \cdot w(f_{I_{i'},J}) = w(g_{J,I_{i'}}^\epsilon) \cdot w(f_{I_{i'},J}).$$

Hence,

$$1 = w(\eta_{II}^{2\epsilon}) = \sum_{J \in \mu(A)} w(g_{J,I}^\epsilon) \cdot w(f_{I,J})$$

and

$$0 = \sum_{J \in \mu(A)} w(g_{J,I_{i'}}^\epsilon) \cdot w(f_{I_{i'},J})$$

for  $i < i'$ . Putting this in matrix form:

$$\begin{bmatrix} w(g_{J_1,I_1}) & \cdots & w(g_{J_s,I_1}) \\ \vdots & \ddots & \vdots \\ w(g_{J_1,I_r}) & \cdots & w(g_{J_s,I_r}) \end{bmatrix} \cdot \begin{bmatrix} w(f_{I_1,J_1}) & \cdots & w(f_{I_r,J_1}) \\ \vdots & \ddots & \vdots \\ w(f_{I_1,J_s}) & \cdots & w(f_{I_r,J_s}) \end{bmatrix} = \begin{bmatrix} 1 & ? & ? & \cdots & ? \\ 0 & 1 & ? & \cdots & ? \\ 0 & 0 & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}.$$

The right hand side has rank  $r = |A|$ , whereas the the rank of the left hand side is bounded by  $s = |\mu(A)|$ . We conclude that  $|A| \leq |\mu(A)|$ .  $\square$

The previous lemma together with *Hall's theorem* proves that  $d_I \leq d_B$ .

### 10.2.1 Hall's Theorem

Consider a hypothetical course where students have to present a particular paper in pairs, but no two pairs can work on the same paper. Each pair naturally has a preference, and the problem is whether or not one can assign papers such that all pairs are satisfied. Let  $P$  denote the set of papers and let  $A_j \subseteq P$  denote the set of papers Pair  $j$  would like to study.

**Example 10.10.** Let  $P = \{1, 2, 3, 4, 5\}$  and  $A_1 = \{1, 2, 3\}$ ,  $A_2 = \{1, 4, 5\}$  and  $A_3 = \{3, 5\}$ . In this case Pair 1 can work on Paper 1, Pair 2 can work on Paper 4, and Pair 3 can work on Paper 5.

**Example 10.11.** Let  $P = \{1, 2, 3, 4, 5\}$  and  $A_1 = \{1, 2, 3\}$ ,  $A_2 = \{4, 5\}$ ,  $A_3 = \{4\}$  and  $A_4 = \{5\}$ . In this case no assignment is possible.

Let  $G$  be a bipartite graph with vertex set  $X \sqcup Y$ , and for a subset of vertices  $X' \subseteq X$ ,

$$N_G(X') = \{y \in Y : \text{there is an edge in } G \text{ between } y \text{ and } x \in X'\}.$$

**Theorem 10.12** (Hall's Theorem). *The following two statements are equivalent:*

1. *There exists an injection  $\chi: X \hookrightarrow Y$  such that  $\chi(x) = y$  only if there is an edge between  $x$  and  $y$ .*
2.  $|X'| \leq |N_G(X')|$  for all  $X' \subseteq X$ .

**Example 10.13.** Returning to Example 10.11, we let  $X = \{S_1, S_2, S_3, S_4\}$ ,  $Y = \{1, 2, 3, 4, 5\}$ , and we connect  $S_i \in X$  with  $j \in Y$  if  $j \in A_i$ . There is no way of assigning papers as  $3 = |\{S_2, S_3, S_4\}| > |N_G(\{S_2, S_3, S_4\})| = |\{4, 5\}| = 2$ .

### 10.2.2 Wrapping Up

Let  $B(M)_{2\epsilon} = \{[a, b] \in B(M) : b - a > 2\epsilon\}$  and  $B(N)_{2\epsilon} = \{[a, b] \in B(N) : b - a > 2\epsilon\}$ . Recalling the definition of the cost of a matching from Eq. (8) in Section 9.1, we see that for intervals  $I$  and  $J$ ,

$$d_B(\{I\}, \{J\}) = \min\{\max\{c(I), c(J)\}, c(I, J)\}.$$

If  $I \in B(M)_{2\epsilon}$  or  $J \in B(N)_{2\epsilon}$ , and  $d_B(\{I\}, \{J\}) \leq \epsilon$ , it must thus be true that

$$c(I, J) \leq \epsilon. \tag{18}$$

**Theorem 10.14.** *Let  $B(M)$  and  $B(N)$  be finite multisets of finite intervals  $[a, b]$ . Then,*

$$d_B(B(M), B(N)) \leq d_I(M, N).$$

*Proof.* Assume that  $M$  and  $N$  are  $\epsilon$ -interleaved. Combing Lemma 10.9 with Theorem 10.12 we get an injection  $\chi_1: B(M)_{2\epsilon} \hookrightarrow B(N)$

$$\chi_1(I) = J \Rightarrow I^I \text{ and } I^J \text{ are } \epsilon\text{-interleaved.}$$

Symmetrically, we get an injection  $\chi_2: B(N)_{2\epsilon} \hookrightarrow B(M)$  such that

$$\chi_2(J) = I \Rightarrow I^I \text{ and } I^J \text{ are } \epsilon\text{-interleaved.}$$

In either case,  $c(I, J) \leq \epsilon$  by 10.2 and Eq. (18).

We now explain how these two injections combine to yield a matching  $\chi$  between  $B(M)$  and  $B(N)$  with cost at most  $\epsilon$ . We start by forming a *directed* bipartite graph  $G$  with vertex set  $B(M) \sqcup B(N)$  and with a *directed* edge from  $I \rightarrow J$  if  $\chi_1(I) = J$  or  $\chi_2(I) = J$ . Note that there is exactly one edge starting at every vertex in  $B(M)_{2\epsilon} \cup B(N)_{2\epsilon}$ .

We define a matching for each connected component  $C_i$  of  $G$  as follows:

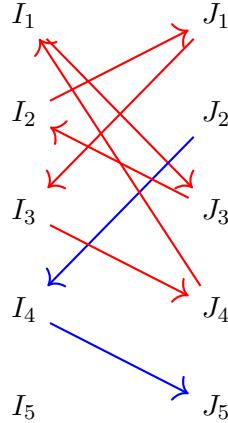
1. If  $C_i$  forms a directed cycle, add  $(I, \chi_1(I))$  to  $\chi$  for all  $I \in C_i \cap B(M)$ . This matches every element of  $C_i$ , and  $c(I, \chi_1(I)) \leq \epsilon$ .
2. If  $C_i$  does not form a directed cycle, then  $C_i$  must contain an initial interval  $J$  and a terminal interval  $K$ . If  $J \in B(M)$ , then add  $(I, \chi_1(I))$  to  $\chi$  for all  $I \in C_i \cap B(M)$ . Otherwise, add  $(\chi_2(I), I)$  to  $\chi$  for all  $I \in C_i \cap B(N)$ . This matches all intervals in  $C_i$  with the possible exception of  $K$ . As  $K$  is not the source of a directed edge,  $K \in B(M) \setminus B(M)_{2\epsilon} \cup B(N) \setminus B(N)_{2\epsilon}$ . Therefore  $K$  can be left unmatched at a cost of  $c(K) \leq \epsilon$ .

The resulting matching  $\chi$  between  $B(M)$  and  $B(N)$  is an  $\epsilon$ -matching. We conclude that

$$d_B(B(M), B(N)) \leq d_I(M, N).$$

□

**Example 10.15.** Let  $B(M)_{2\epsilon} = \{I_1, \dots, I_4\}$ ,  $B(M) \setminus B(M)_{2\epsilon} = \{I_5\}$ ,  $B(N)_{2\epsilon} = \{J_1, \dots, J_4\}$  and  $B(N) \setminus B(N)_{2\epsilon} = \{J_5\}$ , together with the injections  $B(M)_{2\epsilon} \rightarrow B(N)$  and  $B(N)_{2\epsilon} \rightarrow B(M)$  shown in the following diagram.



The above algorithm for producing a matching gives

$$\chi = \{(I_1, J_3), (I_2, J_1), (I_3, J_4), (I_4, J_2)\}.$$

### 10.3 Exercises

1. Prove Lemma 14.1.
2. Prove Lemma 14.2.
3. Give examples showing that the assumption  $\alpha(I_1) \leq \alpha(I_3)$  is necessary in Lemma 10.7 and Lemma 10.8.

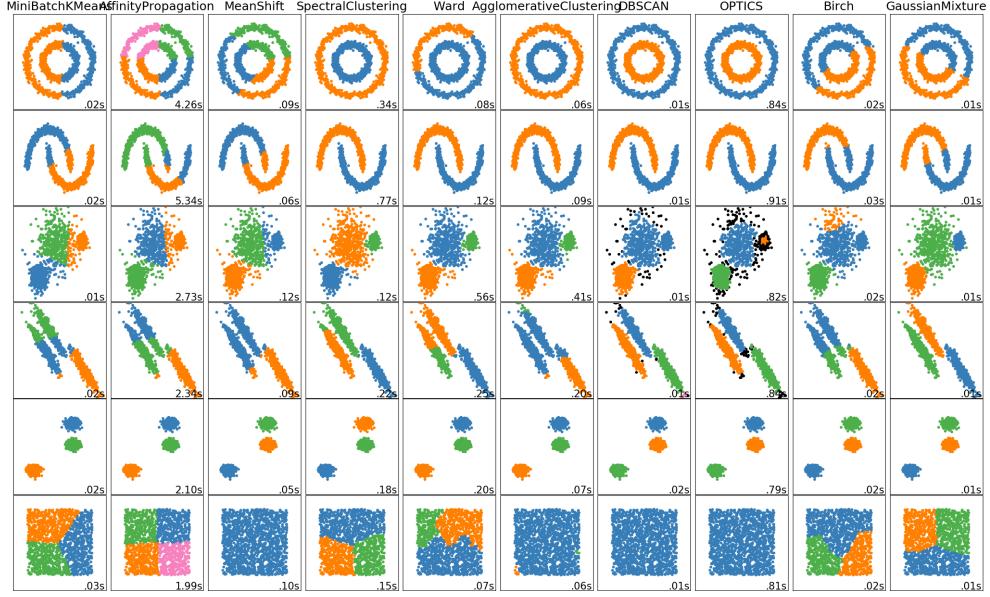


Figure 33: Points belonging to the same cluster drawn are in the same color. The figure can be generated with the code found here: [https://scikit-learn.org/stable/auto\\_examples/cluster/plot\\_cluster\\_comparison.html](https://scikit-learn.org/stable/auto_examples/cluster/plot_cluster_comparison.html)

## 11 Clustering

Clustering is the task of partitioning data such that data belonging to different partitions differ among themselves more than data which belong to the same partition. In the language of topology, clustering is the task of inferring the connected components of data. There are many different clustering methods; see Fig. 33 for a few examples. In this section we will focus on a clustering method based on persistent homology and a foundation result due to Kleinberg.

### 11.1 Examples of Clustering Methods

We briefly list some classical clustering methods.

**The  $\epsilon$ -neighborhood graph** Given a finite metric space  $(P, d)$ , we define the  **$\epsilon$ -neighborhood graph** to be the graph on the vertex set  $P$ , with an edge connecting  $p_i$  and  $p_j$  if  $d(p_i, p_j) \leq \epsilon$ . In other words, the  $\epsilon$ -neighborhood graph coincides with the 1-skeleton of  $\text{VR}_{\epsilon/2}(P)$ . The set of points  $P$  is then partitioned according to the following rule:  $p_i$  and  $p_j$  belong to the same partition if and only if they belong to the same connected component of the neighborhood graph. This method suffers from the *chaining effect* as the following example illustrates.

**Example 11.1.** Consider two dense point sets  $P_1$  and  $P_2$  (corresponding to the two clusters) that are connected by a path of low density. The  $\epsilon$ -neighborhood graph will fail to separate the two clusters. (Figure coming)

**$k$ -th nearest neighbor graph ( $k$ NN)** For a finite metric space  $(P, d)$  the  $k$ -th nearest neighbor graph has vertex set  $P$ , and an edge connecting  $p_i$  and  $p_j$  if  $p_j$  is amongst the  $k$  closest points to  $p_i$ . The clusters correspond to the connected components of the resulting graph.

**$k$ -means clustering** Assuming that  $P$  is a finite set of points in  $\mathbb{R}^d$  and  $k > 0$  is an integer,  **$k$ -means clustering** seeks to partition the points of  $P$  into  $k$  clusters such that intra-cluster variance is minimal, i.e. a partitioning of  $P$  into  $P_1, \dots, P_k$  such that

$$\sum_{i=1}^k \sum_{p \in P_i} \|p - \mu_i\|^2$$

is minimal. Here  $\mu_i$  denotes the mean of the points in  $P_i$ . Computing such an optimal partition is an NP-hard problem, and iterative techniques are employed to find local minima. This technique is ill-suited if your clusters are non-convex; see the first column of Fig. 33.

**DBSCAN** A density-based refinement of the  $\epsilon$ -neighborhood graph which is less sensitive to the chaining effect is defined as follows: let  $m$  be a positive integer and let  $C$  denote the set of points in  $P$  which have at least  $m$  points within distance  $\epsilon$ . The elements of  $C$  are called the **core points**. Now, construct the  $\epsilon$ -neighborhood graph on the core points, and use the  $\epsilon$ -neighborhood graph approach to cluster the points of  $C$ . Any other point  $p \in P$  which is not a core point, is then added to the partition containing the core point  $p_i$  if  $d(p_i, p) \leq \epsilon$ . If  $p$  is not within  $\epsilon$  distance to any core point, then  $p$  is labelled as noise. The 7th column of Fig. 33 shows that DBSCAN does quite well at separating the data when the data comes in (possibly non-convex) geometrically disjoint clusters.

## 11.2 Kleinberg's Theorem

Let  $P$  be a finite set with  $|P| \geq 2$ . A **clustering function** is a function  $f$  that associates to any metric  $d: P \times P \rightarrow \mathbb{R}_{\geq 0}$  a partition  $f(d)$  of  $P$ , i.e.

$$f: \{\text{metrics on } P\} \rightarrow \{\text{partitions of } P\}.$$

We say that a clustering function  $f$  is **isometry-invariant** if for any isometry  $\sigma: (P, d) \rightarrow (P, d')$  the points  $p_i$  and  $p_j$  belong to the same subset of  $f(d)$  if and only if  $\sigma(p_i)$  and  $\sigma(p_j)$  belong to the same subset of  $f(d')$ . If  $f(\alpha \cdot d) = f(d)$  for all  $\alpha > 0$ , then  $f$  is **scale-invariant**. Lastly, let  $d'$  be such that  $d'(p_i, p_j) \leq d(p_i, p_j)$  whenever  $p_i, p_j$  are members of the same subset of  $f(d)$ , and  $d'(p_i, p_j) \geq d(p_i, p_j)$  whenever  $p_i$  and  $p_j$  do not belong to the same subset. If  $f(d) = f(d')$  for all such  $d$  and  $d'$ , then we say that  $f$  is **consistent**.

**Theorem 11.2** (Kleinberg [19]). *If  $f$  is isometry-invariant, scale-invariant and consistent, then*

1. (**lump**)  $f(d) = \{P\}$  for all metrics  $d$ , or
2. (**discrete**)  $f(d) = \{\{p\} : p \in P\}$  for all metrics  $d$ .

*Proof.* Let  $|P| = n$  and let  $d$  be such that  $d(p_i, p_j) = 1$  for all  $p_i \neq p_j$ . By isometry invariance we must have that  $f(d) = \{P\}$  or  $f(d) = \{\{p\} : p \in P\}$ . Assume that the former is the case, let  $d'$  be any other distance on  $P$ , and rescale such that  $\alpha d' < d$ . Then

$$f(d') \stackrel{\text{scale inv}}{=} f(\alpha d') \stackrel{\text{consistency}}{=} f(d) = \{P\}.$$

If  $f(d) = \{\{p\} : p \in P\}$ , then we may rescale such that  $\alpha d' > d$ . A similar argument then applies.  $\square$



Figure 34: Example illustrating the multiscale nature of the clustering problem. At a coarse scale, two clusters (dark shading) are apparent. At a finer scale, each of these two clusters appear to decompose further into two subclusters (light shading).

*Remark 11.3.* This formulation of the theorem diverges slightly from the statement in Kleinberg's paper where isometry invariance is replaced with a **richness** assumption. The assumption is that there for every partition of  $P$  exists a metric  $d$  such that  $f(d)$  equals that partition. In other words,  $f$  is a surjective map. Kleinberg then goes on to show that there is no clustering function simultaneously satisfying scale-invariance, consistency, and richness.

**Example 11.4.** The  $k$ -th nearest neighbor graph is an example of a clustering method which is both isometry- and scale-invariant.

**Example 11.5.** The  $\epsilon$ -neighborhood graph is an example of a clustering method which is both isometry-invariant and consistent.

**Example 11.6.** Assume that  $(P, <)$  is an ordered set, and let  $d$  be any metric on  $P$ . Consider the total order on pairs of points  $\{(i, j) : i < j\}$  defined by  $(i, j) < (i', j')$  if  $d(i, j) < d(i', j')$  or  $d(i, j) = d(i', j')$  and  $(i, j) < (i', j')$  in the lexicographical ordering induced by the order on  $P$ . This gives,

$$(i_1, j_1) < (i_2, j_2) < \cdots < (i_{n(n-1)/2}, j_{n(n-1)/2}).$$

Fix an integer  $k > 0$  and let  $G_d$  be the graph on  $P$  with edges  $\{(i_1, j_1), \dots, (i_m, j_m)\}$  where  $m$  is the smallest integer such that  $G_d$  has precisely  $m$  connected components. The clustering function  $f$  that maps  $d$  to the partition defined by the connected components of  $G_d$  is both scale-invariant and consistent, but it depends on the ordering of the points and is therefore not isometry-invariant.

An issue with clustering is that there might be no unique correct scale at which the data should be considered. Indeed, Fig. 34 illustrates that what appears to be well-defined clusters at one scale, may reveal a finer structure upon inspection at a smaller scale. One may attempt to rectify these issues by considering **hierarchical clustering methods**. Such methods do not assign a single partition to the input metric space, but rather a one-parameter family of clusters capturing the evolution of the clusters as the connectivity parameter increases - much like persistent homology. The output of such a method is typically a graph like the one shown in Fig. 35. These methods do however also have their deficiencies but we will not discuss this here.

### 11.3 ToMATo

ToMATo - Topological Mode Analysis Tool - is the result of combining the classical *hill climbing algorithm* with persistent homology. The algorithm first appeared in [12] and in this section we follow their exposition.

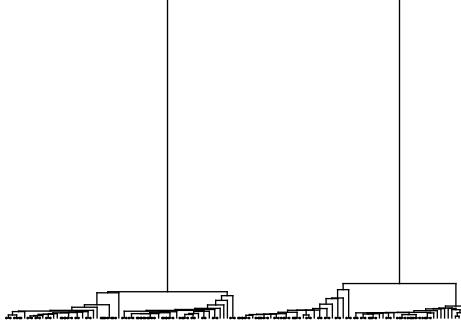


Figure 35: Dendrogram for the point set shown in Fig. 34.

To illustrate the idea of ToMAtO in the continuous setting, consider the graph shown to the left in Fig. 36 which depicts an example density-function of an assumed underlying distribution from which we sample. As there are two clear maxima, we say that each of them represent a cluster, and the points belonging to the respective clusters are those points belonging to their respective **ascending regions**. That is, the ascending region of a critical point  $m$  is the set of all points  $x$  in the domain which end up at  $m$  by following the flow induced by the gradient vector field of the density function. In the example in the question, this means that every point to the left of the valley where the two peaks meet are clustered to one cluster, and all the points to the right of the valley are clustered to the other cluster. But what if we consider the density function to the right in Fig. 36 - defining every local peak to be a cluster seems counter-productive. To proceed one chooses a threshold  $\tau > 0$  and defines a local maxima to define a cluster if the associated peak has a **prominence** of at least  $\tau$ . That is, one needs to descend a height of at least  $\tau$  in order to reach a higher peak. Persistent homology is in turn used to select the appropriate  $\tau$ : let  $f: X \rightarrow \mathbb{R}$  denote the density function and consider the persistent homology of the **superlevel filtration** given by

$$H_0(f^{-1}[s, \infty)) \rightarrow H_0(f^{-1}[t, \infty))$$

for  $s \geq t$ . In Fig. 37 we show the right-most function of Fig. 36 together with its persistence diagram. Note that the diagram is drawn upside down as we are considering the superlevel filtration and not the sublevel filtration we have frequently encountered. The blue line in the figure is the separating line which is used to distinguish the noisy peaks from the peaks that are to define clusters. Note that the line defines the parameter  $\tau$  by means of the equation  $y = x - \tau$ . Once we have settled on such a  $\tau$  we select the prominent peaks  $m_p$  and for a peak  $m$  of prominence less than  $\tau$ , we consider the minimum  $\epsilon$  such that  $m$  belongs to the same connected component as a  $\tau$ -prominent peak  $m_p$  in  $f^{-1}[m - \epsilon, \infty)$ . Any point which belongs to the ascending region of  $m$  will now be associated to the point  $m_p$ , and the clusters are defined accordingly. E.g, in Fig. 37 the points to left of the deepest valley define a cluster, and the points to the right define another cluster.

### 11.3.1 The Discrete Case

For the purpose of this lecture we shall assume that we are given a point sample  $P$  and an estimate  $\hat{f}(p)$  of the density at every point  $p \in P$ . Such an estimate can be obtained in numerous ways, the most naïve being a simple normalized count of the number of points of  $P$  within a  $\delta$ -ball around  $p$  for some  $\delta > 0$ .

Fix an  $\epsilon > 0$  and consider the  $\epsilon$ -neighborhood graph  $G$  constructed from the points in  $P$ . With this information we can form our initial clusters: every local maximum, i.e., all points

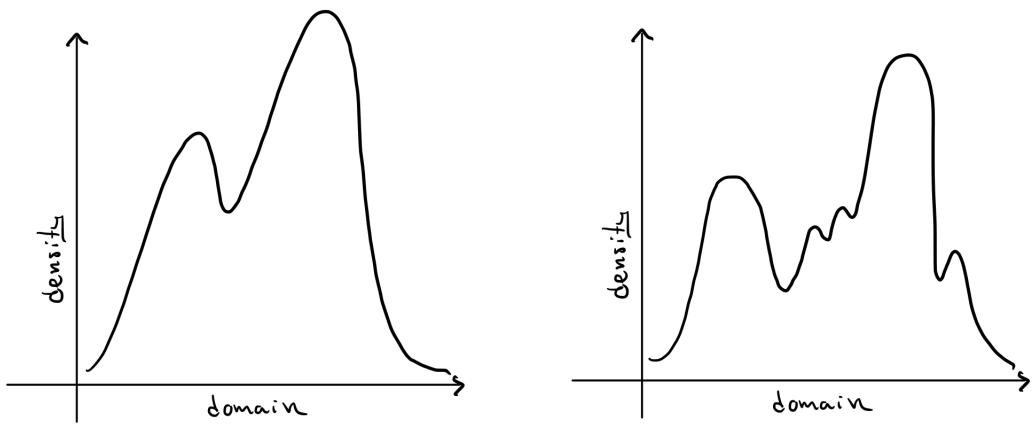


Figure 36: Two density functions on the same domain.

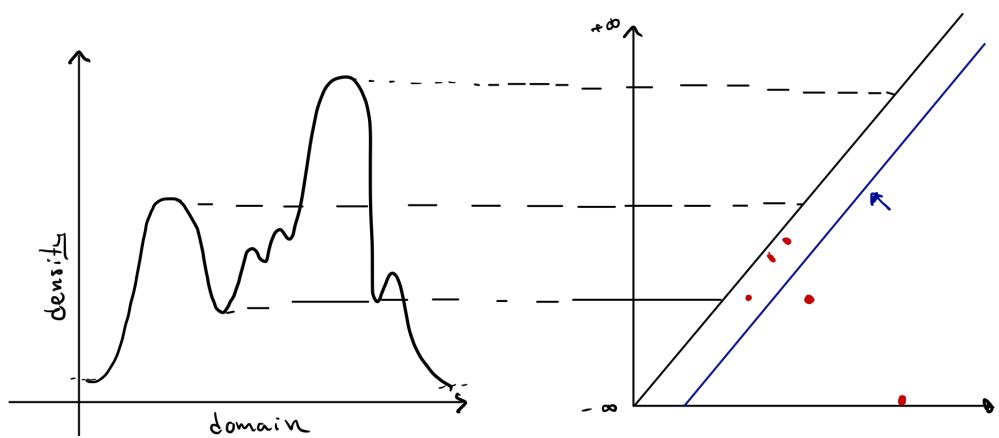
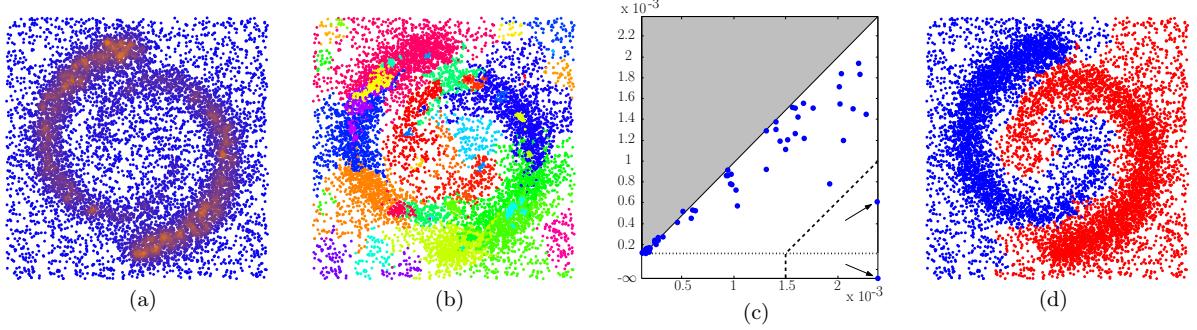


Figure 37: A density function and the persistence diagram of its superlevel filtration in dimension 0.



**Figure 2:** Our approach in a nutshell: (a) estimation of the underlying density function  $f$  at the data points; (b) result of the basic graph-based hill-climbing step; (c) approximate PD showing 2 points far off the diagonal corresponding to the 2 prominent peaks of  $f$ ; (d) final result obtained after merging the clusters of non-prominent peaks.

Figure 38: The pipeline of the discrete algorithm (copied from [12]).

$p_i$  satisfying  $\hat{f}(p_i) > \hat{f}(p_j)$  for all neighbors of  $p_i$  in  $G$ , represent a cluster. If  $p_j$  is not a local maximum then we associate  $p_j$  to the same cluster as its neighbor with the greatest filtration value. If there are  $n$  local maxima, then we get a collection of clusters  $\{C_1, \dots, C_n\}$ .

By defining the filtration value  $\hat{f}(e)$  of an edge  $e$  connecting  $p_i$  and  $p_j$  to be  $\max(\hat{f}(p_i), \hat{f}(p_j))$  we obtain a superlevel filtration of the graph  $G$ . The algorithm proceeds precisely as in the continuous case: compute the persistence diagram in degree 0 of the superlevel filtration of  $\hat{f}: G \rightarrow \mathbb{R}$  and use the persistence diagram to determine a threshold parameter  $\tau$ . Upon choosing a  $\tau$ , iterate over the points in the order given by  $\hat{f}(p_1) > \hat{f}(p_2) > \dots > \hat{f}(p_n)$  and merge clusters as follows:

- If  $\hat{f}(p_i)$  is a local maximum in  $G$ , then do nothing.
- If  $\hat{f}(p_i)$  is not a local maximum, then  $p_i$  belongs to a cluster  $C_k$ . Iterating (in an arbitrary order) over the neighbors  $p_j$  of  $p_i$  in  $G$  with  $\hat{f}(p_j) \geq \hat{f}(p_i)$ , we merge the cluster  $C_l$  containing  $p_j$  with  $C_k$  if

$$\min(\max\{\hat{f}(p) : p \in C_l\}, \max\{\hat{f}(p) : p \in C_k\}) < \hat{f}(p_i) + \tau$$

That is, we replace  $C_k$  and  $C_l$  in  $C$  by  $C_l \cup C_k$ .

The resulting  $C$  contains the clusters. The four steps of the discrete algorithm and the end result is shown in Fig. 38. This illustration is copied from [12] and the reader is encouraged to have a look at that paper for other nice examples, and theoretical guarantees for point samples on Riemannian manifolds. Note that the above explanation was written in a conceptual way and that an implementation would iterate over the points using a union-find structure. We include the algorithm from [12] in Fig. 39.

---

**ALGORITHM 1:** Clustering

---

**Input:** simple graph  $G$  with  $n$  vertices,  $n$ -dimensional vector  $\tilde{f}$ , real parameter  $\tau \geq 0$ .

Sort the vertex indices  $\{1, 2, \dots, n\}$  so that  $\tilde{f}(1) \geq \tilde{f}(2) \geq \dots \geq \tilde{f}(n)$ ;  
Initialize a union-find data structure  $\mathcal{U}$  and two vectors  $g, r$  of size  $n$ ;  
**for**  $i = 1$  to  $n$  **do**  
    Let  $\mathcal{N}$  be the set of neighbors of  $i$  in  $G$  that have indices lower than  $i$ ;  
    **if**  $\mathcal{N} = \emptyset$  **then**  
        // vertex  $i$  is a peak of  $\tilde{f}$  within  $G$   
        Create a new entry  $e$  in  $\mathcal{U}$  and attach vertex  $i$  to it;  
         $r(e) \leftarrow i$ ; //  $r(e)$  stores the root vertex associated with the entry  $e$   
    **else**  
        // vertex  $i$  is not a peak of  $\tilde{f}$  within  $G$   
         $g(i) \leftarrow \text{argmax}_{j \in \mathcal{N}} \tilde{f}(j)$ ; //  $g(i)$  stores the approximate gradient at vertex  $i$   
         $e_i \leftarrow \mathcal{U}.\text{find}(g(i))$ ;  
        Attach vertex  $i$  to the entry  $e_i$ ;  
        **for**  $j \in \mathcal{N}$  **do**  
             $e \leftarrow \mathcal{U}.\text{find}(j)$ ;  
            **if**  $e \neq e_i$   $\min\{\tilde{f}(r(e)), \tilde{f}(r(e_i))\} < \tilde{f}(i) + \tau$  **then**  
                 $\mathcal{U}.\text{union}(e, e_i)$ ;  
                 $r(e \cup e_i) \leftarrow \text{argmax}_{\{r(e), r(e_i)\}} \tilde{f}$ ;  
                 $e_i \leftarrow e \cup e_i$ ;  
            **end**  
        **end**  
    **end**  
**end**

**Output:** the collection of entries  $e$  of  $\mathcal{U}$  such that  $\tilde{f}(r(e)) \geq \tau$ .

---

Figure 39: The algorithm copied from the journal version of [12].

## 11.4 Exercises

1. Complete the proof of Kleinberg's theorem.
2. Show that the  $k$ -th nearest neighbor graph clustering method is not consistent.
3. Show that the clustering method in Example 11.6 is consistent and scale-invariant.
4. Let  $P = \{-5, -4, \dots, 4, 5\}$  and let  $f(-5) = 0, f(-4) = 1, f(-3) = 2, f(-2) = 3, f(-1) = 0, f(0) = 1, f(1) = 0, f(2) = 2, f(3) = 4, f(4) = 3$ , and  $f(5) = 5$ . Fix  $\epsilon = 1$ , and run the ToMATo algorithm for  $\delta \in \{1, 2, 3, 4\}$ .
5. Install Gudhi and play around with ToMATo.

## 12 Reeb Graphs and Mapper

The goal of this section is to discuss the basic properties of Reeb graphs, and to see how the Mapper graph can be seen as a coarsened Reeb graph.

### 12.1 Reeb Graphs

We define an  **$\mathbb{R}$ -space** to be a pair  $(X, f)$  where  $X$  is a topological space and  $f: X \rightarrow \mathbb{R}$  is a continuous function. Associated to  $(X, f)$  we let  $R(X, f)$  denote the topological space  $X/\sim$  where  $\sim$  is the equivalence relation generated by  $x \sim y$  if  $f(x) = f(y) = t$  and  $x$  and  $y$  belong to the same path component of the fiber  $f^{-1}(t)$ . We define the **Reeb graph** of  $(X, f)$  to be the  $\mathbb{R}$ -space  $(R(X, f), \hat{f})$  where  $\hat{f}: R(X, f) \rightarrow \mathbb{R}$  given by  $\hat{f}([x]) = f(x)$ . The fact that  $\hat{f}$  is well-defined and continuous follows from the following well-known lemma.

**Lemma 12.1.** *Let  $X/\sim$  be the quotient of a topological space  $X$  by an equivalence relation  $\sim$ . Any continuous function  $f: X \rightarrow Y$  which is constant on equivalence classes induces a continuous function  $X/\sim \rightarrow Y$  with respect to the quotient topology.*

*Remark 12.2.* By replacing  $\mathbb{R}$  with  $\mathbb{R}^d$  we can associate a **Reeb space** to a function  $f: X \rightarrow \mathbb{R}^d$  precisely as we did for the Reeb graph. The resulting spaces are typically quite complicated.

**Example 12.3.** Fig. 40 shows a genus 2 surface  $X$  together with a function  $f: X \rightarrow \mathbb{R}$ , and the associated Reeb graph. Note that the vertical circular features are preserved.

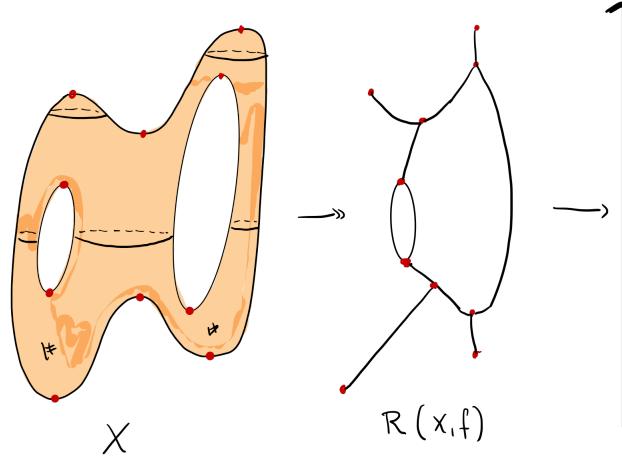


Figure 40: An  $\mathbb{R}$ -space and the associated Reeb graph.

The  $\mathbb{R}$ -spaces  $(X, f)$  and  $(Y, g)$  are **isomorphic** if there exists a homeomorphism  $\phi: X \rightarrow Y$  such that  $g \circ \phi = f$ . For topological spaces in general, the space  $R(X, f)$  need not be a graph, but it will be so for sufficiently tame  $\mathbb{R}$ -spaces. The following definition is taken from [15].

**Definition 12.4.** An  $\mathbb{R}$ -space  $(X, f)$  is **constructible** if  $(X, f)$  is isomorphic to a pair  $(Y, g)$  constructed in the following way. Assuming the following data:

1. A finite set of increasingly ordered *critical values*  $\{c_0, c_1, \dots, c_n\}$ .

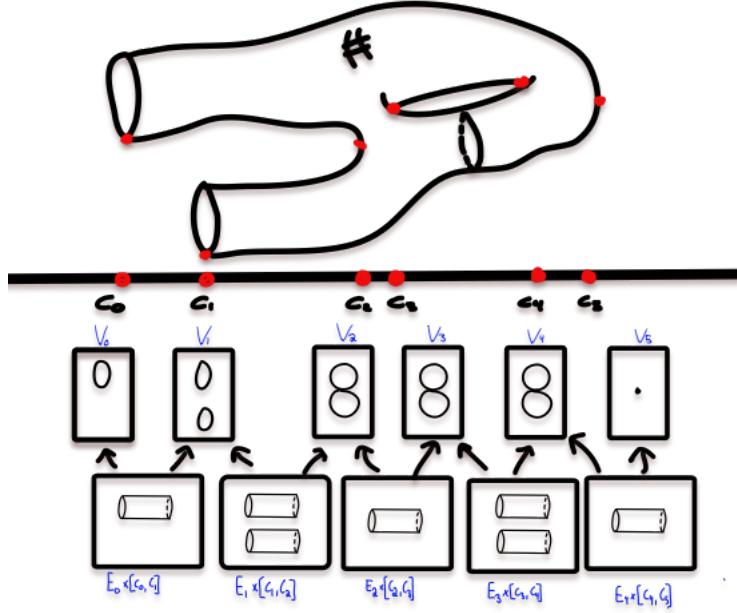


Figure 41: Gluing together the components of a constructible  $\mathbb{R}$ -space.

2. For  $0 \leq i \leq n$ , a locally path-connected and compact space  $V_i$ .
3. For  $0 \leq i \leq n-1$ , a locally path-connected and compact space  $E_i$ , together with continuous maps  $l_i: E_i \rightarrow V_i$  and  $r_i: E_i \rightarrow V_{i+1}$ .

The topological space  $Y$  is the disjoint union of the spaces  $V_i \times \{c_i\}$  and  $E_i \times [c_i, c_{i+1}]$  relative to the relations  $(l_i(x), c_i) \sim (x, c_i)$  and  $(r_i(x), c_{i+1}) \sim (x, c_{i+1})$  for all  $i$  and  $x \in E_i$ . The map  $g: Y \rightarrow \mathbb{R}$  is projection onto the second factor.

**Example 12.5.** Examples of constructible  $\mathbb{R}$ -spaces include 1)  $M$  is a compact manifold and  $f$  is a Morse function, and 2)  $X$  is a compact polyhedron and  $f$  is a piecewise-linear map. In the latter case, assuming the function is injective on the vertices, the Reeb graph can be computed in time  $O(m \log m)$  where  $m$  denotes the total number of vertices, edges and triangles in  $X$ .

This process of construction a space is illustrated in Fig. 41. One sees from the definition of a constructible space that the homotopy type of the fibers  $f^{-1}(a)$  is constant between critical values. Indeed, we have the following commutative diagram where the vertical maps are homotopy equivalences,

$$\begin{array}{ccccc}
 E_{i-1} & \xrightarrow{r_{i-1}} & V_i & \xleftarrow{l_i} & E_i \\
 \downarrow \simeq & & \downarrow \simeq & & \downarrow \simeq \\
 f^{-1}(c_{i-1}, c_i) & \longrightarrow & f^{-1}(c_{i-1}, c_{i+1}) & \longleftarrow & f^{-1}(c_i, c_{i+1})
 \end{array} . \quad (19)$$

It is also not hard to see that  $R(X, f)$  is a graph with a vertex for each critical value  $c_i$  and edges connecting neighboring critical values.

A natural question to consider is how much of the information is preserved under such a simplification. Clearly, a graph is an intrinsically one-dimensional object, so three-dimensional structures such as cavities must be lost. We have the following preliminary lemma.

**Lemma 12.6.** Let  $(X, f)$  be a constructible  $\mathbb{R}$ -space and let  $\pi: X \rightarrow \mathbb{R}(X, f)$  denote the quotient map. Then,  $(\pi)_*: H_0(X) \rightarrow H_0(R(X, f))$  is an isomorphism, and  $(\pi)_*: H_1(X) \rightarrow H_1(R(X, f))$  is an epimorphism.

We leave it as an exercise to verify these two claims. The salient point is that all connectivity information is preserved (by construction) and every loop in the graph represents a true topological feature in the original space. In particular, if  $X$  is connected and simply-connected, then  $R(X, f)$  must be a tree regardless of the function  $f$ . When considering *levelset zigzag persistence*, we shall give a precise description of the homological properties preserved by the Reeb graph.

## 12.2 Morse Functions

Reeb graphs were originally introduced in the context of Morse theory to study the evolution level sets of real-valued functions. And this for good reason: when  $f: M \rightarrow \mathbb{R}$  is a Morse function on a two-dimensional compact and oriented surface, then  $M$  can be recovered up to homeomorphism from its Reeb graph. Note that Morse functions are dense in the space of smooth real-valued functions on  $M$ . Following [17], we will now prove this result, making use of a few elementary properties of Morse functions; see e.g. [24, Chapter 1] for a concise introduction. We need the following facts:

- F1:** The vertices of  $R(M, f)$  correspond to critical points of  $f$ , i.e., points such that  $df(p) = 0$ .
- F2:** Each vertex in  $R(M, f)$  has degree 1 (minimum or maximum) or degree 3 (saddle point). This is a consequence of the *Morse lemma* that states that in a suitable coordinate system around the critical point  $p$  (i.e. a vertex in the Reeb graph), one has that  $f(x) = f(p) - x_1^2 - x_2^2$  (maximum),  $f(x) = f(p) \mp x_1^2 \pm x_2^2$  (saddle), or  $f(x) = f(p) + x_1^2 + x_2^2$  (minimum).
- F3:** Let  $C_0$ ,  $C_1$  and  $C_2$  denote the total number of vertices corresponding to minima, saddle points and maxima, respectively. Then  $C_2 - C_1 + C_0 = \beta_2(M) - \beta_1(M) + \beta_0(M) = \chi(M)$ . This follows from the *Morse inequalities*.
- F4:** The Euler-characteristic of a surface of genus  $g$  is  $2 - 2g$ .

The reader is encouraged to verify these properties in Fig. 40.

**Theorem 12.7.** *The number of loops in  $R(M, f)$  equals the genus  $g$  of  $M$ .*

*Proof.* Let  $n_i$  denote the number of vertices of degree  $i$ , and let  $m$  and  $n$  denote the total number of edges and vertices, respectively. From **F1**, we know that  $n = n_1 + n_3$ , where  $n_1 = C_0 + C_2$  and  $n_3 = C_1$ . Furthermore, a simple edge-counting gives the relation  $m = \frac{1}{2}(n_1 + 3n_3)$ .

From the Euler-Poincare theorem (Theorem 2.15),  $\beta_0(R(M, f)) - \beta_1(R(M, f)) = n - m$ . Using the fact that the graph is connected, i.e.,  $\beta_0(R(M, f)) = 1$ , we get

$$\beta_1(R(M, f)) = 1 + m - n = 1 - \frac{1}{2}(n_1 - n_3) = 1 - \frac{1}{2}(C_0 + C_2 - C_1) = 1 - \frac{1}{2}\chi(M) = g.$$

□

A similar argument can be applied to non-orientable surfaces: the Reeb graph of a non-orientable surface of genus  $g$  has at most  $g/2$  loops.

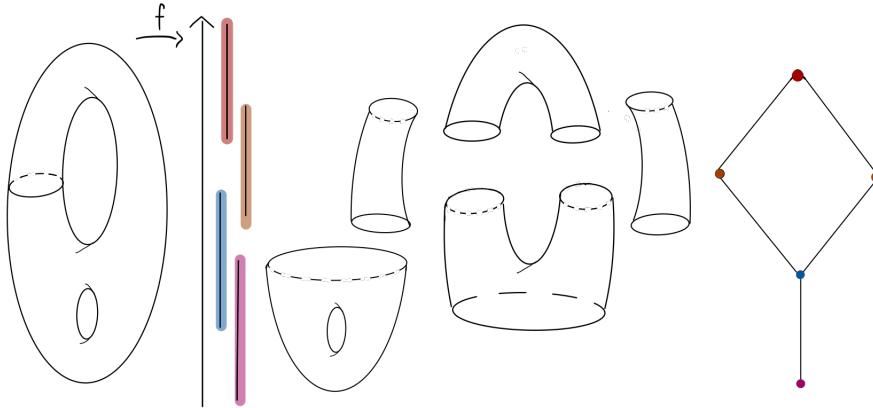


Figure 42: The mapper graph associated to a function  $f: X \rightarrow \mathbb{R}$ .

### 12.3 The Continuous Mapper

An **open cover** of a topological space  $X$  is a collection  $\mathcal{U} = \{U_i\}_{i \in I}$  of open sets such that  $\cup_{i \in I} U_i = X$ . Associated to a finite cover there is a simplicial complex  $N(\mathcal{U})$  called the **nerve** of  $\mathcal{U}$  containing simplices  $\{\sigma \subseteq I : \cap_{i \in \sigma} U_i \neq \emptyset\}$ . Assuming a continuous function  $f: X \rightarrow \mathbb{R}^d$ , and a finite open cover  $\mathcal{U}$  of  $\mathbb{R}^d$ , we get a **pullback cover** of  $X$  defined by

$$f^*\mathcal{U} = \{f^{-1}(U_i) : U_i \in \mathcal{U}\}.$$

We can further refine the pullback cover  $f^*\mathcal{U}$  by splitting every open set  $f^{-1}(U_i)$  into its path components. We denote the resulting refined cover by  $\widehat{f^*\mathcal{U}}$ . The **(continuous) mapper** associated to the pair  $(X, f)$  is the simplicial complex  $M(X, f) := N(\widehat{f^*\mathcal{U}})$ .

**Example 12.8.** Fig. 42 shows a continuous map on a surface and the resulting Mapper graph.

While the resulting simplicial complex of the previous example turned out to be a graph, that need not be the case in general. We do however have the following result:

**Lemma 12.9.** *The dimension of  $M(X, f)$  is bounded by the dimension of  $N(\mathcal{U})$ .*

*Proof.* If  $V$  and  $V'$  are different path-connected components of some  $f^{-1}(U_i)$  then clearly  $V \cap V' = \emptyset$ . Let  $V_i$  be a path-component of  $f^{-1}(U_i)$  and assume that  $\cap_{i=1}^{m+1} V_i$  is non-empty. By the previous assumption we must have that each  $U_i$  is distinct, and furthermore

$$\emptyset \neq f\left(\bigcap_{i=1}^{m+1} V_i\right) \subseteq \bigcap_{i=1}^{m+1} U_i.$$

We conclude that if  $M(X, f)$  contains an  $m$ -simplex, then so does  $N(\mathcal{U})$ .  $\square$

We define the **resolution** of a cover to be  $\text{res}(\mathcal{U}) = \sup\{\text{diam}(U_i) : U_i \in \mathcal{U}\}$ . One can think of  $M(X, f)$  as approaching  $R(X, f)$  as the resolution tends to 0 (this can be made precise): rather than probing the connectivity information over every fiber, we only consider the connectivity information over preimages of (small) open sets. It is not surprising that as long as the fibers are "locally constant", the the Mapper graph should resemble the Reeb graph for covers of sufficiently fine resolution. The next example illustrates this.

**Example 12.10.** Consider the Reeb graph and two mapper graphs shown in Fig. 43. The first cover is rather coarse and fails to detect any of the two loops, whereas the cover to the right captures the larger loop but an even smaller resolution would be needed to detect both loops.

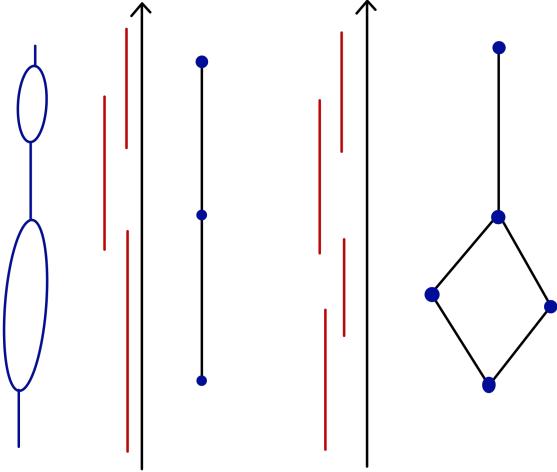


Figure 43: Left: the Reeb graph of a pair  $(X, f)$ . Middle: the Mapper graph for a coarse cover. Right: The Mapper graph for a cover of finer resolution.

## 12.4 Mapper in Practice

Mapper was first introduced in [27] and works as follows. Given a finite set of data  $P$  together with a metric (or dissimilarity measure) between data points, the user is typically left to decide the filter function  $f: P \rightarrow \mathbb{R}^d$  and the cover of  $\mathbb{R}^d$ . The outputted Mapper graph (or, simplicial complex if  $d > 1$ ) will be highly sensitive to these parameters. Furthermore, the computation of path components (clusters) of pre-images will too depend on one or more parameters. Assuming for the moment that  $f$  and  $\mathcal{U}$  are provided, and that we have a way of clustering data, then the Mapper algorithm is as follows:

---

**Algorithm 2:** The Mapper algorithm.

---

```

Data: A data set  $P$  with a pairwise distance measure  $d$ , a function  $f: P \rightarrow \mathbb{R}^d$  and a cover  $\mathcal{U}$  of  $\mathbb{R}^d$ .
Result: The mapper graph associated to  $(P, d)$ .
Let  $K = \emptyset$ .
for  $U \in \mathcal{U}$  do
    Decompose  $f^{-1}(U)$  into clusters  $C_{U,1}, \dots, C_{U,k_U}$ .
    Add a vertex  $v_{U,i}$  to  $K$  for every cluster  $C_{U,i}$ .
end
if  $C_{U_1,j_1} \cap \dots \cap C_{U_m,j_m}$  then
    Add the simplex  $\{v_{U_1,j_1}, \dots, v_{U_m,j_m}\}$  to  $K$ .
end
return  $K$ .
```

---

**Example 12.11.** The Mapper algorithm applied to two different data sets are shown in Fig. 45 and Fig. 46.

Some remarks on the parameters in question:

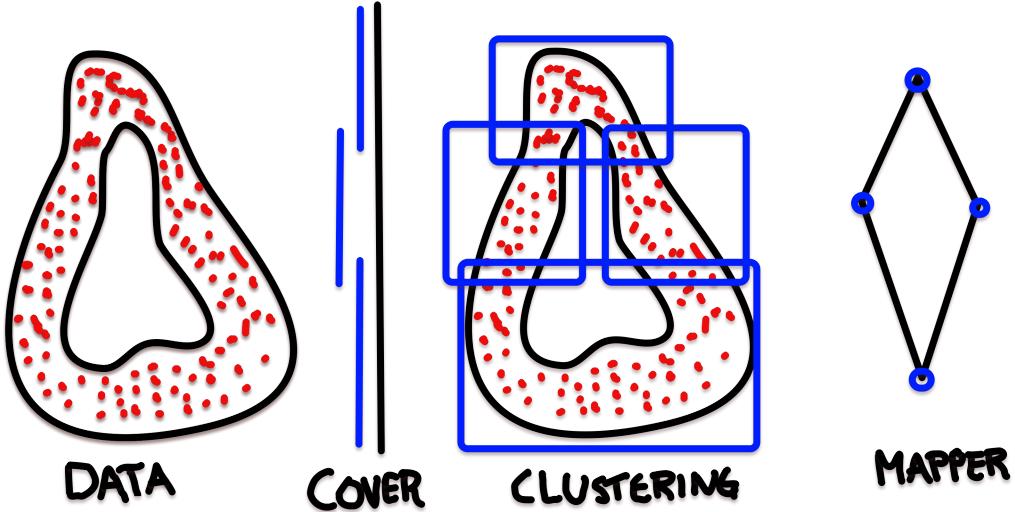


Figure 44: Mapper from a finite data set.

**The filter function** Examples of typical functions encountered in the literature.

- For data in  $\mathbb{R}^d$  one can project down onto a low dimensional subspace, e.g. by means of principal component analysis or by a non-linear dimension reduction algorithm such as ISOMAP.
- Eccentricity measures, in an intuitive sense, how far from the "center" of the data a given point is. That is,

$$E_r(p) = \left( \frac{\sum_{q \in P} d(p, q)^r}{N} \right)^{1/r}$$

where  $1 \leq r \leq \infty$ .

- Density estimates such as Gaussian kernel,

$$\delta_\epsilon(p) = \sum_{q \in P} \exp(-d(p, q)^2/\epsilon),$$

where  $\epsilon > 0$ .

**The cover** For covers of  $\mathbb{R}$  it is customary to let  $\mathcal{U}$  consist of regularly spaced intervals of equal resolution, and such that neighbouring cover elements overlap with a percentage below 50%. For covers of  $\mathbb{R}^d$  one typically considers hypercubes in the analogue way. The resulting Mapper will in general not be graphs but for ease of visualization it is customary to merely output the 1-skeleton. Note that the output of the algorithm is very sensitive to the chosen cover.

**Clustering** The most common choices are  $kNN$  clustering, DBSCAN, and the  $\epsilon$ -neighborhood graph. Note that one can approach the clustering problem in two ways: one can either cluster all the data immediately and then computing connected components of preimages can be done by means of standard graph algorithms. Another, more local, approach is to cluster each preimage separately. A local computation may be a better option when dealing with very large data.

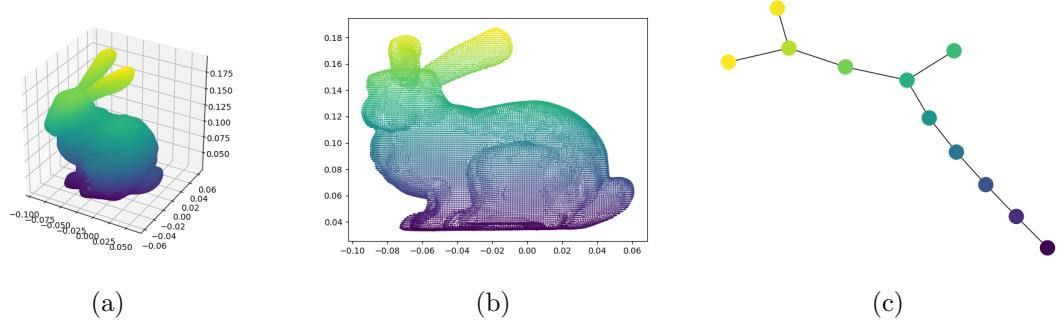


Figure 45: (a) and (b) show the data set  $P$  (the Stanford bunny) and the function  $f: P \rightarrow \mathbb{R}$  given by height. (c) is the associated Mapper graph as outputted by Gudhi using the settings: min\_points\_per\_node=0, clustering=None, N=100, beta=0.1, C=10, filter\_bnds=None, resolutions=None, gains=None, verbose=verbose.

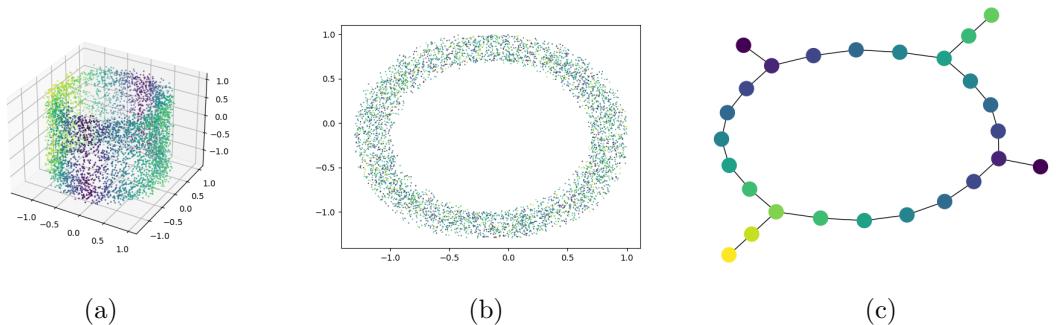
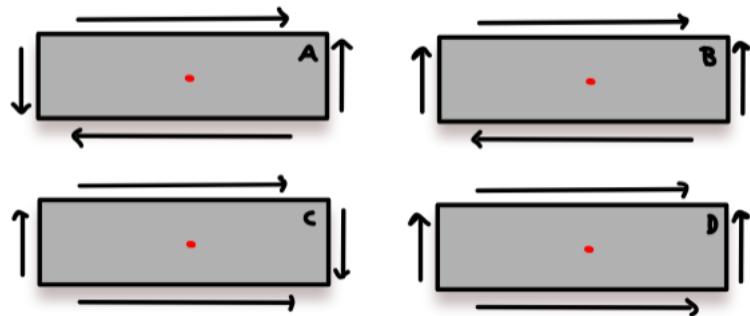


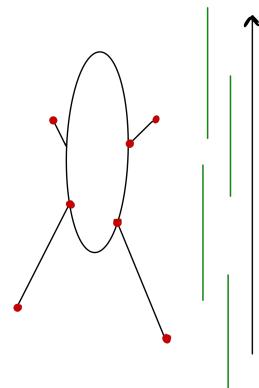
Figure 46: (a) and (b) show the projection onto the three first and two last coordinates of a noisy sample  $P$  of the 2-torus  $\mathbb{S}^1 \times \mathbb{S}^1 \subset \mathbb{R}^4$ , where the function is  $f: P \rightarrow \mathbb{R}$  is given by  $f(x_1, x_2, x_3, x_4) = |x_1|$ . (c) is the associated Mapper graph as outputted by Gudhi using the settings: min\_points\_per\_node=0, clustering=None, N=100, beta=0.1, C=10, filter\_bnds=None, resolutions=10, gains=None, verbose=verbose.

## 12.5 Exercises

1. Prove Lemma 12.6.
2. (a) For a non-orientable closed two-manifold  $M$  of genus  $g$ , one has that  $\chi(M) = 2 - g$ . If  $f: M \rightarrow \mathbb{R}$  is a Morse function, then one can show that  $R(M, f)$  has no vertex of degree greater than 3. Prove that the number of loops in  $R(M, f)$  is given by  $\frac{1}{2}(g - n_2)$  where  $n_2$  is the number of vertices of degree 2 in  $R(M, f)$ .
- (b) Let  $f: \mathbb{RP}^2 \rightarrow \mathbb{R}$  be a Morse function. Show that  $R(\mathbb{RP}^2, f)$  contains no loops.
- (c) Consider the following familiar surfaces, each equipped with the Morse function given by the distance to the central point. Draw the corresponding Reeb graphs. Discuss the results in light of what we know from (a) and (b).



3. What is the continuous Mapper associated to the following cover?



## 13 Zigzag Persistent Homology

Thus far we have focused on applying homology to a filtration of a topological space. In this section we shall see that considering posets for which the underlying diagram is not linearly oriented can lead to richer invariants.

**Example 13.1.** Zigzag persistence is a natural tool to explore the topology of time-varying point clouds, e.g., for studying the movement of bird flocks, and dynamic graphs, i.e., graphs where edges can appear and disappear. As an example of the former, consider a data point to be a continuous function  $p_i: [0, 1] \rightarrow \mathbb{R}^d$ . For each time  $t \in [0, 1]$  one obtains a point cloud  $P(t) = \{p_1(t), \dots, p_n(t)\}$  and the corresponding Vietoris–Rips complex at a scale  $\epsilon$ ,  $\text{VR}_\epsilon(P(t))$ . We shall assume that there is a finite set of values  $0 = c_0 < c_1 < \dots < c_m = 1$  such that  $\text{VR}_\epsilon(P(t))$  is constant for  $t \in [c_i, c_{i+1}]$ . To study the evolution of the homology of  $P(t)$ , we apply homology to the following zigzag of simplicial complexes:

$$\begin{array}{ccccccc} V_\epsilon(P(c_0)) & & V_\epsilon(P(c_1)) & & V_\epsilon(P(c_{m-1})) & & V_\epsilon(P(c_m)) \\ \searrow & \swarrow & \searrow & \cdots & \searrow & \swarrow & \searrow \\ V_\epsilon(P(c_0)) \cup V_\epsilon(P(c_1)) & & & & V_\epsilon(P(c_{m-1})) \cup V_\epsilon(P(c_m)) & & \end{array}$$

### 13.1 Zigzag Persistence Modules

A **zigzag poset** on  $n$  vertices is a partial order of the form

$$\overset{1}{\bullet} \leftrightarrow \overset{2}{\bullet} \leftrightarrow \cdots \leftrightarrow \overset{n-1}{\bullet} \leftrightarrow \overset{n}{\bullet}$$

where  $\leftrightarrow$  denotes that the arrow can be either  $\leftarrow$  or  $\rightarrow$ . A **zigzag persistence module** is a  $P$ -module where  $P$  is any zigzag poset. Note that if  $P$  has  $n$  vertices and all the arrows point in the same direction, then a zigzag persistence module is nothing more than an  $[n]$ -module as considered in Section 6.

For a zigzag poset  $P$  on  $n$  vertices, let  $[a, b]$  denote the restriction of  $P$  to the vertices  $i \in [a, b]$ . The associated **interval module**  $I^{[a, b]}$  is the  $P$ -module defined by

$$I_i^{[a, b]} = \begin{cases} \mathbf{k} & \text{if } i \in [a, b] \\ 0 & \text{otherwise} \end{cases},$$

together with the identity morphism  $\text{id}: I_i^{[a, b]} \rightarrow I_j^{[a, b]}$  whenever  $i, j \in [a, b]$  and  $i$  and  $j$  are comparable.

**Example 13.2.** If  $n = 4$ , and the arrows alternate, then  $I^{[2, 3]}$  is the persistence module

$$0 \rightarrow \mathbf{k} \xleftarrow{1} \mathbf{k} \rightarrow 0.$$

We state the following result without proof. A proof can be obtained by a variation of the proof of Theorem 6.16.

**Theorem 13.3.** Let  $V$  be a zigzag module on  $n$  vertices such that  $\dim V_p < \infty$  for all  $p \in [n]$ . Then

$$V \cong \bigoplus_{[a, b] \in B(V)} I^{[a, b]},$$

where  $B(V)$  is a unique multiset of intervals in  $[n]$  called the **barcode** of  $V$ .

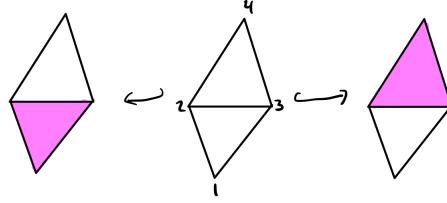


Figure 47: A zigzag of simplicial complexes:  $K \hookleftarrow K \cap K' \hookrightarrow K'$ .

While the situation to a large extent mimics that of standard persistent homology, the barcodes in zigzag persistence can be more refined as the following example illustrates.

**Example 13.4.** Let  $K$  and  $K'$  be the simplicial complexes shown in Fig. 47, and let  $K \cap K'$  denote the simplex-wise intersection. This defines a zigzag of simplicial complexes

$$K \hookleftarrow K \cap K' \hookrightarrow K',$$

and a corresponding zigzag persistence module:

$$\mathbb{Z}_2 \cong H_1(K) \hookleftarrow H_1(K \cap K') \cong \mathbb{Z}_2 \oplus \mathbb{Z}_2 \hookrightarrow H_1(K') \cong \mathbb{Z}_2.$$

Note that the 1-cycle  $\{\{1, 2\}, +\{1, 3\} + \{2, 3\} + \{3, 4\}\}$  is non-trivial in homology in all three complexes, and one may therefore be inclined to think that there is a bar in the barcode spanning all three vertices. That is however not the case, as the following choice of a basis for the middle vector space shows

$$\{\{1, 2\} + \{1, 3\} + \{2, 3\}, \{2, 3\} + \{2, 4\} + \{3, 4\}\}$$

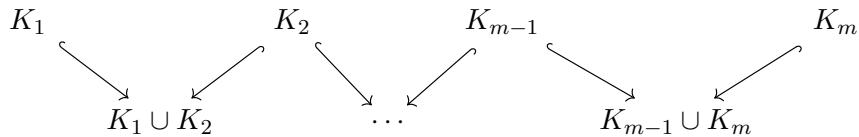
for  $H_1(K \cap K')$ . Indeed, this gives the decomposition

$$\mathbb{Z}_2 \xleftarrow{[1, 0]} \mathbb{Z}_2 \oplus \mathbb{Z}_2 \xrightarrow{[0, 1]} \mathbb{Z}_2 = \left( \mathbb{Z}_2 \xleftarrow{1} \mathbb{Z}_2 \rightarrow 0 \right) \oplus \left( 0 \leftarrow \mathbb{Z}_2 \xrightarrow{1} \mathbb{Z}_2 \right).$$

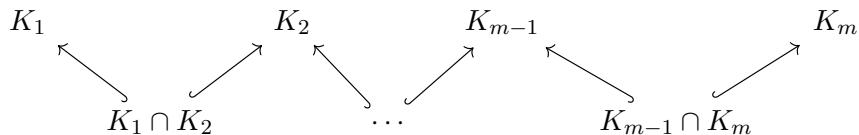
The barcode thus consists of the intervals  $\{[1, 2], [2, 3]\}$ .

## 13.2 The diamond principle

Given a sequence of simplicial complexes  $\{K_i\}_{i=1}^m$ , there are two natural ways of linking  $K_i$ 's. The union zigzag  $K_{\cup}$



and the intersection zigzag  $K_{\cap}$



For instance, replacing the middle simplicial complex in Example 13.4 by the union of the two outer complexes yields a barcode  $\{[1, 1], [3, 3]\}$  in degree 1 homology. The question is thus if there is a method to obtain the one barcode from the other. That turns out to be the case if one dispenses with intervals supported only at a intersection or a union.

**Theorem 13.5** (The strong diamond principle [9]). *Consider the following diagram of simplicial complexes and simplicial maps where the four middle maps are inclusions.*

$$\begin{array}{ccccccc}
 & & A \cup B & & & & \\
 & \nearrow & & \swarrow & & & \\
 K_1 & \longleftrightarrow & \cdots & \longleftrightarrow & K_{l-2} & \longleftrightarrow & A \\
 & & \downarrow & & \downarrow & & \\
 & & A \cap B & & B & \longleftrightarrow & K_{l+2} \\
 & & \uparrow & & \uparrow & & \\
 & & A & & B & \longleftrightarrow & \cdots \longleftrightarrow K_m
 \end{array}$$

Let  $K^+$  and  $K^-$  denote the zigzags passing through the union and intersection, respectively. Then there is the following correspondence of intervals in the barcodes:

$$[l, l] \in B(H_{p+1}(K^+)) \leftrightarrow [l, l] \in B(H_p(K^-)),$$

In the remaining cases, the matching preserves homological dimension:

$$\begin{aligned}
 [b, l] \in B(H_p(K^*)) &\leftrightarrow [b, l-1] \in B(H_p(K^\dagger)), & \text{if } b < l, \\
 [b, d] \in B(H_p(K^*)) &\leftrightarrow [l+1, d] \in B(H_p(K^\dagger)), & \text{if } d > l, \\
 [b, d] \in B(H_p(K^*)) &\leftrightarrow [b, d] \in B(H_p(K^\dagger)), & \text{for all other cases.}
 \end{aligned}$$

where  $*, \dagger \in \{+, -\}$  and  $* \neq \dagger$ .

By iteratively applying the previous theorem we see that  $B(H_p(K_\cup))$  is determined by  $B(H_p(K_\cap)) \cup B(H_{p-1}(K_\cap))$ , and  $B(H_p(K_\cap))$  is determined by  $B(H_p(K_\cup)) \cup B(H_{p+1}(K_\cup))$ . It is left as an exercise to spell out precisely how to transform an interval in one barcode to an interval in the other (possibly shifted by a degree in homology).

### 13.3 Levelset Zigzag Persistent Homology

The persistent homology of an  $\mathbb{R}$ -space  $f: X \rightarrow \mathbb{R}$  studies the evolution of the homology of the sublevel sets. An alternative approach is to study how the homology persists across the levelsets  $f^{-1}(t)$  as the parameter  $t$  sweeps over the real line. In the following we shall assume that the pair  $(X, f)$  is a constructible  $\mathbb{R}$ -space. By labelling the critical values  $c_1 < c_2 < \dots < c_n$ , we obtained the following zigzag of topological spaces

$$\begin{array}{ccccccc}
 f^{-1}(-\infty, c_2) & & f^{-1}(c_1, c_3) & & f^{-1}(c_{n-2}, c_n) & & f^{-1}(c_{n-1}, \infty) \\
 \swarrow & & \searrow & & \swarrow & & \searrow \\
 f^{-1}(c_1, c_2) & & \dots & & f^{-1}(c_{n-1}, c_n) & &
 \end{array}$$

The interest in this diagram comes from the fact that

- The fibers are constant (up to homeomorphism) between critical values. That is,  $f^{-1}(c_i, c_{i+1})$  is homeomorphic to  $f^{-1}(s) \times (c_i, c_{i+1})$  with  $f$  being the projection onto the second component and  $s$  is any point in  $c_i < s < c_{i+1}$ .

- $f^{-1}(c_{i-1}, c_{i+1})$  deformation retracts onto  $f^{-1}(c_i)$ .

Hence, the topological evolution of the fibers is completely captured up to homotopy by the above zigzag; applying  $H_p$  to the above zigzag yields the **levelset zigzag persistence module in dimension  $p$  (associated to  $f$ )**, and the corresponding barcode  $\text{ZZ}_p(f)$ . The intervals in  $\text{ZZ}_p(f)$  are naturally given in terms of the critical values. To see this, assume that we have a bar born at  $H_p(f^{-1}(c_i, c_{i+2})) \cong H_p(f^{-1}(c_i))$  that lives up to (and including)  $H_p(f^{-1}(c_j, c_{j+1}))$ . This represents a feature that is alive at all fibers  $f^{-1}(s)$  for  $s \in [c_{i+1}, c_{j+1}]$ , and the corresponding interval in  $\text{ZZ}_p(f)$  is thus  $[c_i, c_{j+1}] \in \text{ZZ}_p(f)$ . Likewise we have intervals  $(c_i, c_j)$ ,  $[c_i, c_j]$ , and  $(c_i, c_j]$ .

**Example 13.6.** Consider the constructible  $\mathbb{R}$ -space shown in Fig. 48. Proceeding in  $H_0$  with the "obvious basis" we get the following zigzag of vector spaces

$$\begin{array}{cccccccc} \mathbb{Z}_2 \oplus \mathbb{Z}_2 & & \mathbb{Z}_2 & & \mathbb{Z}_2 & & \mathbb{Z}_2 & & \mathbb{Z}_2 \\ \swarrow I_2 & & \nearrow [1,1] & & \swarrow 1 & & \nearrow [1,1] & & \swarrow 1 \\ \mathbb{Z}_2 \oplus \mathbb{Z}_2 & & \mathbb{Z}_2 & & \mathbb{Z}_2 & & \mathbb{Z}_2 \oplus \mathbb{Z}_2 & & \mathbb{Z}_2 \end{array}$$

For each copy of  $\mathbb{Z}_2 \oplus \mathbb{Z}_2$  above, we replace the basis  $\{v_1, v_2\}$  with the basis  $\{v_1, v_1 + v_2\}$ . The zigzag then diagonalizes as

$$\begin{array}{cccccccc} \mathbb{Z}_2 \oplus \mathbb{Z}_2 & & \mathbb{Z}_2 & & \mathbb{Z}_2 & & \mathbb{Z}_2 & & \mathbb{Z}_2 \\ \swarrow I_2 & & \nearrow [1,0] & & \swarrow 1 & & \nearrow [1,0] & & \swarrow 1 \\ \mathbb{Z}_2 \oplus \mathbb{Z}_2 & & \mathbb{Z}_2 & & \mathbb{Z}_2 & & \mathbb{Z}_2 \oplus \mathbb{Z}_2 & & \mathbb{Z}_2 \end{array}$$

which is the direct sum of the following three zigzag modules:

$$\begin{array}{ccccccccc} \mathbb{Z}_2 & & \mathbb{Z}_2 & & \mathbb{Z}_2 & & \mathbb{Z}_2 & & \mathbb{Z}_2 \\ \swarrow 1 & & \nearrow 1 & & \swarrow 1 & & \nearrow 1 & & \swarrow 1 \\ \mathbb{Z}_2 & & \mathbb{Z}_2 & & \mathbb{Z}_2 & & \mathbb{Z}_2 & & \mathbb{Z}_2 \\ \mathbb{Z}_2 & & \mathbb{Z}_2 & & \mathbb{Z}_2 & & \mathbb{Z}_2 & & \mathbb{Z}_2 \\ \swarrow 1 & & \nearrow 0 & & \swarrow 0 & & \nearrow 0 & & \swarrow 0 \\ 0 & & 0 & & 0 & & 0 & & 0 \\ \swarrow 0 & & \nearrow 0 & & \swarrow 0 & & \nearrow 0 & & \swarrow 0 \\ 0 & & 0 & & 0 & & 0 & & 0 \end{array}$$

We conclude that

$$\text{ZZ}_0(f) = \{[c_1, c_5], [c_1, c_2), (c_3, c_4)\}.$$

And for  $H_1$  (exercise)

$$\text{ZZ}_1(f) = \{[c_1, c_5), [c_1, c_2], [c_3, c_4]\}.$$

Let  $\text{PH}_p(f)$  denote the barcode of the sublevel persistence module  $M^f$  in homology dimension  $p$  (see Section 9.1.1). A simple computation yields

$$\text{PH}_0(f) = \{[c_1, \infty), [c_1, c_2)\} \quad \text{and} \quad \text{PH}_1(f) = \{[c_1, \infty), [c_1, c_5), [c_3, \infty), [c_4, \infty)\}.$$

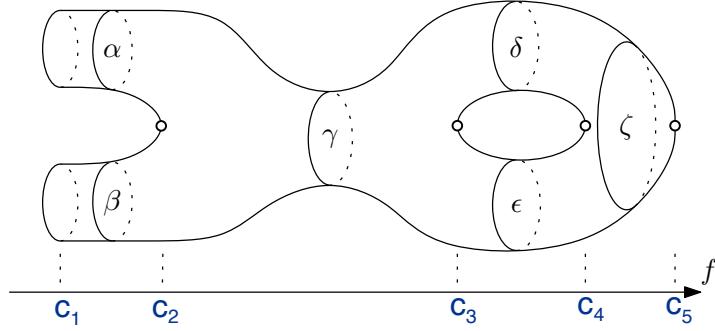


Figure 48: A topological space projected onto the horizontal axis. Copied from [10].

Comparing with the above computations, there seems to be a close connection between the bars in levelset and sublevel persistent homology. We now state a theorem that makes this connection precise.

**Theorem 13.7** ([10]). *Assume that  $(X, f)$  is a constructible  $\mathbb{R}$ -space, and that  $\dim H_p(f^{-1}(t)) < \infty$  for all  $t$ . Then the sublevel barcode  $\text{PH}_p(f)$  is given as the following union:*

$$\{[c_i, c_j) : [c_i, c_j] \in \text{ZZ}_p(f)\} \cup \{[c_i, \infty) : [c_i, c_j] \in \text{ZZ}_p(f)\} \cup \{[c_j, \infty) : (c_i, c_j) \in \text{ZZ}_{p-1}(f)\}.$$

Hence, levelset persistent homology is a finer invariant than sublevel persistence. Moreover, and as illustrated by the previous example, the types of endpoints carry concrete information about the topological feature: closed-closed and open-open intervals correspond to global features of  $X$ , i.e. features that are present irrespective of the function (but potentially with different birth and death times). The half-open intervals on the other hand can be perturbed away.

We end this section by stating an important theorem.

**Theorem 13.8.** *Assume that  $(X, f)$  and  $(X, g)$  are constructible  $\mathbb{R}$ -spaces, such that*

$$\dim H_p(f^{-1}(t)), \dim H_p(g^{-1}(t)) < \infty$$

for all  $t \in \mathbb{R}$ . Then,

$$d_B(\text{ZZ}_p(f), \text{ZZ}_p(g)) \leq \|f - g\|_\infty.$$

In fact, here we can replace the standard bottleneck distance with a more refined version that only allows matchings of intervals with the same type of endpoints, i.e. closed-closed must be matched to closed-closed and so forth. Much more can be said on this topic, but we will not pursue it further.

### 13.3.1 Reeb Graphs

Let  $(X, f)$  be a constructible  $\mathbb{R}$ -space. Then, as an immediate consequence of the definition of  $(R(X, f), \hat{f})$ , we have that  $H_0(f^{-1}(t)) \cong H_0(\hat{f}^{-1}(t))$ , and that  $H_i(\hat{f}^{-1}(t)) = 0$  for  $i > 0$ . In particular, the homological information preserved by the Reeb graph is precisely that of the 0-th levelset persistence barcode:

$$\text{ZZ}_0(f) = \text{ZZ}_0(\hat{f}).$$

Furthermore, these intervals have a very clear geometric interpretation as the following example illustrates.

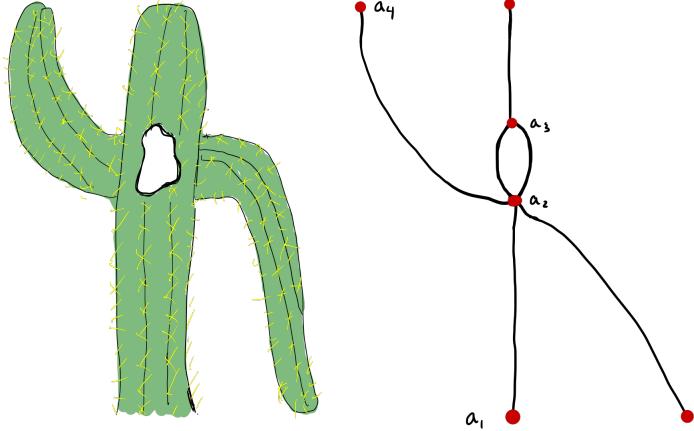


Figure 49: A Reeb graph.

**Example 13.9.** Consider the Reeb graph shown in Fig. 49. It has 4 critical points  $a_1 < a_2 < a_3 < a_4$ , yielding the following zigzag of spaces

$$\begin{array}{ccccccc} f^{-1}(-\infty, a_2] & & f^{-1}(a_1, a_3) & & f^{-1}(a_2, a_4) & & f^{-1}(a_3, \infty) \\ \swarrow & & \searrow & & \swarrow & & \searrow \\ f^{-1}(a_1, a_2) & & f^{-1}(a_2, a_3) & & f^{-1}(a_3, a_4) & & \end{array}$$

Applying  $H_0$  gives the following zigzag of vector spaces:

$$\begin{array}{ccccccc} \mathbb{Z}_2 \oplus \mathbb{Z}_2 & & \mathbb{Z}_2 & & \mathbb{Z}_2 \oplus \mathbb{Z}_2 & & \mathbb{Z}_2 \oplus \mathbb{Z}_2 \\ \swarrow I_2 & & \nearrow [1,1] & & \swarrow I_2 & & \nearrow I_2 \\ \mathbb{Z}_2 \oplus \mathbb{Z}_2 & & \mathbb{Z}_2 & & \mathbb{Z}_2 \oplus \mathbb{Z}_2 & & \mathbb{Z}_2 \oplus \mathbb{Z}_2 \\ & & \swarrow [1,1,1] & & \nearrow \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} & & \swarrow I_2 \\ & & \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2 & & & & \end{array}$$

The associated barcode is

$$\text{ZZ}_0(\hat{f}) = \{[a_1, a_4], [a_1, a_2), (a_2, a_4], (a_2, a_3)\},$$

with the interpretation

- $[a_1, a_4]$ : the global connected component.
- $[a_1, a_2)$ : the downward facing branch.
- $(a_2, a_4]$ : the upward facing branch.
- $(a_2, a_3)$ : the loop.

This interpretation of different types of intervals as particular graph theoretic features holds for Reeb graphs in general. While different (=non-isomorphic) Reeb graphs may exhibit the same barcode, the associated bottleneck distance still provides an efficiently computable similarity measure. This contrasts the more discriminative distances - such as an interleaving distance for Reeb graphs - which are generally NP-hard to compute.

### 13.4 Computing the Zigzag Barcode

In this section we will consider a high-level algorithm for computing the barcode of a zigzag persistence module coming from applying homology to a zigzag of simplicial complexes. Before turning to the algorithm, let us make some general observations for zigzag persistence modules.

Consider a zigzag persistence module  $V$  of the form

$$V_1 \xleftarrow{f_1} V_2 \xleftarrow{f_2} \cdots \xleftarrow{f_{n-2}} V_{n-1} \xleftarrow{f_{n-1}} V_n.$$

If  $f_i: V_i \rightarrow V_{i+1}$ , then we say that  $f_i$  is **forward**, and if  $f_i: V_{i+1} \rightarrow V_i$ , then we say that  $f_i$  is **backward**.

**Definition 13.10.** A **representative sequence** in  $V$  is an  $n$ -tuple

$$v = (v_1, v_2, \dots, v_n) \in V_1 \times V_2 \times \cdots \times V_n$$

such that:

1. The index  $b(v) = \min_i v_i \neq 0$ , satisfies either of the following:  $b(v) = 1$ ,  $f_{b(v)-1}$  is forward, or  $f_{b(v)-1}$  is backward and  $f_{b(v)-1}(v_{b(v)}) = 0$ . In addition,  $v_i = 0$  for all  $i < b(v)$ .
2. The index  $d(v) = \max_i v_i \neq 0$ , satisfies either of the following:  $d(v) = n$ ,  $f_{d(v)}$  is backward, or  $f_{d(v)}$  is forward and  $f_{d(v)}(v_{d(v)}) = 0$ . In addition,  $v_i = 0$  for all  $i > d(v)$ .
3. For all  $b(v) \leq i \leq d(v)$ ,  $v_i \neq 0$ .
4. If  $b(v) \leq i < d$  and  $f_i$  is forward, then  $f_i(v_i) = v_{i+1}$ .
5. If  $b(v) < i \leq d$  and  $f_i$  is backward, then  $f_i(v_i) = v_{i-1}$ .

For a representative sequence  $v$ , we get a submodule  $\langle v \rangle \subseteq V$ , where  $\langle v \rangle_i = \text{Span}(\{v_i\}) \subseteq V_i$ . Furthermore, observe that  $\langle v \rangle \cong I^{[b(v), d(v)]}$ .

**Definition 13.11.** A finite collection  $\{v^j\}_{j \in J}$  of representative sequences in  $V$  is a **basis** for  $V$  if  $\{v_i^j : v_i^j \neq 0\}_{j \in J}$  is a basis for  $V_i$  for all  $i$ .

The following is left as an exercise.

**Proposition 13.12.** If  $\{v^j\}_{j \in J}$  is a basis for  $V$ , then

$$V = \bigoplus_{j \in J} \langle v^j \rangle \cong \bigoplus_{j \in J} I^{[b(v_j), d(v_j)]}.$$

In particular,  $B(V) = \{[b(v_j), d(v_j)] : j \in J\}$ .

### 13.5 An Algorithm for Zigzag Persistent Homology

In this section we assume that  $K$  is a zigzag of simplicial complexes

$$K := K_1 \xleftarrow{h_1} K_2 \xleftarrow{h_2} \cdots \xleftarrow{h_{n-2}} K_{n-1} \xleftarrow{h_{n-1}} K_n,$$

where each  $h_i$  is an inclusion, and any two consecutive simplicial complexes differ by exactly one simplex. In other words, if  $h_i$  is forward, then  $K_{i+1} = K_i \cup \{\sigma\}$ , and if  $h_i$  is backward then  $K_i = K_{i+1} \cup \{\sigma\}$ . This assumption implies that the induced map  $(h_i)_*: H_p(K_i) \leftrightarrow H_p(K_{i+1})$

is either injective, surjective or an isomorphism, and that the dimension between consecutive vector spaces differ by at most 1.

Let  $H_p(K)$  denote the zigzag module

$$H_p(K_1) \xrightarrow{(h_1)_*} H_p(K_2) \xrightarrow{(h_2)_*} \cdots \xrightarrow{(h_{n-2})_*} H_p(K_{n-1}) \xrightarrow{(h_{n-1})_*} H_p(K_n).$$

We define a **representative cycle** in  $H_p(K)$  to be an  $n$ -tuple

$$c = (c_1, \dots, c_n) \in C_p(K_1) \times \cdots \times C_p(K_n)$$

such that  $0 \neq [c_i] \in H_p(K_i)$  if and only if  $c_i \neq 0$ , and such that  $[c] = ([c_1], \dots, [c_n])$  is a representative sequence in  $H_p(K)$ .

Our goal is to inductively construct a family of representative cycles  $\{c^j\}_{j \in J}$  such that  $\{[c^j]\}_{j \in J}$  is a basis for  $H_p(K)$ . The base case  $k = 1$  is trivial. Let  $\Sigma^i = \{c^j\}_{j \in J}$  denote such a family for  $H_p(K)$  restricted to  $\{1, 2, \dots, i\}$ .

We shall assume that the elements of  $J$  are ordered such that  $b([c^j]) < b([c^{j'}])$  if  $j < j'$ . We say that  $b([c^j])$  is a **forward birth** if  $b([c^j]) = 1$  or  $f_{b([c^j])-1}$  is forward. Symmetrically,  $b([c^j])$  is a **backward birth** if  $b([c^j]) > 1$  and  $f_{b([c^j])-1}$  is backward.

We extend  $\Sigma^i$  to  $\Sigma^{i+1}$  by means of a case-by-case analysis.

1. If  $(h_i)_*$  is forward and an isomorphism, define  $\Sigma^{i+1} = \{(c_1^j, \dots, c_i^j, c_i^j) : c^j \in \Sigma^i\}$ .
2. If  $(h_i)_*$  is backward and an isomorphism, define  $\Sigma^{i+1} = \{(c_1^j, \dots, c_i^j, c_i^j) : c^j \in \Sigma^i\}$ .
3. If  $(h_i)_*$  is forward and injective, let

$$\Sigma^{i+1} = \{(c_1^j, \dots, c_i^j, c_i^j) : c^j \in \Sigma^i\} \cup (0, 0, \dots, 0, c')\},$$

where  $c'$  is any  $p$ -cycle containing the added  $p$ -simplex  $\sigma$ .

4. If  $f_i$  is backward and surjective, let

$$\Sigma^{i+1} = \{(c_1^j, \dots, c_i^j, c_i^j) : c^j \in \Sigma^i\} \cup (0, 0, \dots, 0, \partial_{p+1}(\sigma)\},$$

where  $\sigma$  is the  $(p+1)$ -simplex such that  $K_i = K_{i+1} \cup \{\sigma\}$ .

5. If  $(h_i)_*$  is forward and surjective, then there exists a set  $A \subseteq J$  such that  $\sum_{j \in A} [c_i^j] = 0$ , and where each  $[c_i^j] \neq 0$ . If  $b([c^j])$  is a backward birth for all  $j \in A$ , let  $\lambda$  be the lowest index in  $A$ . If not, let  $\lambda$  be the largest index in  $A$  such that  $b([c^\lambda])$  is a forward birth. Define

$$\Sigma^{i+1} = \{(c_1^j, \dots, c_i^j, c_i^j) : j \neq \lambda, c^j \in \Sigma^i\} \cup \{c'\}$$

where  $c'_x = 0$  for  $x < b(v^\lambda)$  and for  $x = i+1$ , and  $c'_x = \sum_{j \in A} c_x^j$  otherwise.

6. If  $(h_i)_*$  is backward and injective, let  $A \subseteq J$  be the set of indices such that  $c_i^j$  contains  $\sigma$ , where  $\sigma$  is the  $p$ -simplex such that  $K_i = K_{i+1} \cup \{\sigma\}$ . If  $b([c^j])$  is a forward birth index for all  $j \in A$ , then let  $\lambda$  be the smallest index in  $A$ . Otherwise, let  $\lambda$  be the largest index in  $A$  such that  $b([c^\lambda])$  is a backward birth. Define

$$\Sigma^{i+1} = \{(c_1^j, \dots, c_i^j, c_i^j) : j \notin A, c^j \in \Sigma^i\} \cup \{(c_1^\lambda, \dots, c_i^\lambda, 0)\} \cup \{(c^j)' : \lambda \neq j \in A\},$$

where  $(c^j)'_x = c_x^j + c_x^\lambda$  for  $b([c^\lambda]) \leq x \leq i$ ,  $(c^j)'_{i+1} = (c^j)'_i$ , and  $(c^j)'_x = c_x^j$  otherwise.

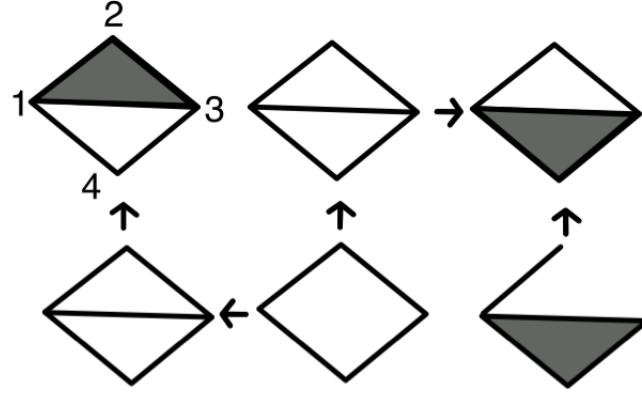


Figure 50: See Example 13.14.

*Remark 13.13.* The presentation here follows [16] and is an adaptation of the algorithm given in [23].

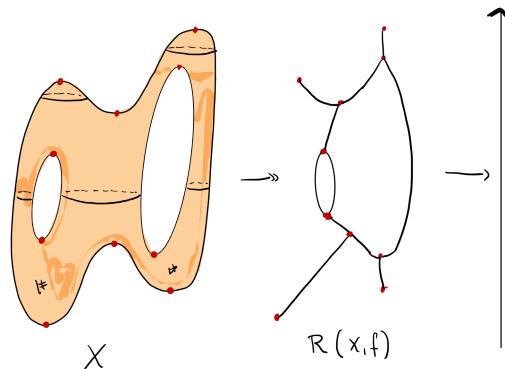
**Example 13.14.** Let us use the algorithm to compute the barcode in dimension 1 of the zigzag filtration shown in Fig. 50. Working inductively we get:

$$\begin{aligned}
 \Sigma^1 &= \{(13 + 14 + 34)\} \\
 \Sigma^2 &= \{(13 + 14 + 34, 13 + 14 + 34), (0, 12 + 13 + 23)\} \\
 \Sigma^3 &= \{(13 + 14 + 34, 14 + 34 + 12 + 23, 14 + 34 + 12 + 23), (0, 12 + 13 + 23, 0)\} \\
 \Sigma^4 &= \{(13 + 14 + 34, 14 + 34 + 12 + 23, 14 + 34 + 12 + 23, 14 + 34 + 12 + 23), \\
 &\quad (0, 12 + 13 + 23, 0, 0), (0, 0, 0, 12 + 13 + 23)\} \\
 \Sigma^5 &= \{(13 + 14 + 34, 14 + 34 + 12 + 23, 14 + 34 + 12 + 23, 14 + 34 + 12 + 23), \\
 &\quad (0, 12 + 13 + 23, 0, 0, 0), (0, 0, 0, 12 + 13 + 23, 0)\} \\
 \Sigma^6 &= \{(13 + 14 + 34, 14 + 34 + 12 + 23, 14 + 34 + 12 + 23, 14 + 34 + 12 + 23, 0), \\
 &\quad (0, 12 + 13 + 23, 0, 0, 0, 0)(0, 0, 0, 12 + 13 + 23, 0, 0)\}
 \end{aligned}$$

From  $\Sigma^6$  we see that the barcode is  $\{[1, 5], [2, 2], [4, 4]\}$ .

### 13.6 Exercises

1. Give an example of two Reeb graphs with the same barcode in dimension 0 levelset persistent homology that are not isomorphic as constructible  $\mathbb{R}$ -spaces.
2. Compute  $ZZ_1(f)$  for the function  $f: X \rightarrow \mathbb{R}$  in Fig. 48.
3. Use the algorithm in Section 13.5 to compute the barcode of Example 13.4.
4. Prove Proposition 13.12.
5. Let  $f: X \rightarrow \mathbb{R}$  be as in the figure below. Compute  $ZZ_0(f)$  and interpret the results geometrically.
6. For an interval  $[b, d] \in B(H_p(K_{\cup}))$  determine the corresponding interval in  $B(H_p(K_{\cap})) \cup B(H_{p-1}(K_{\cap}))$ .



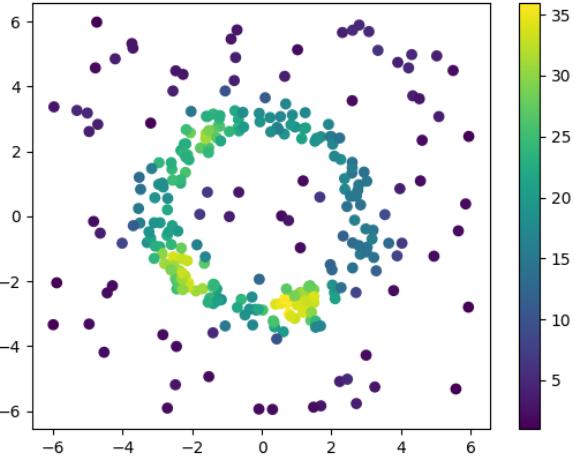


Figure 51: A noisy circle.

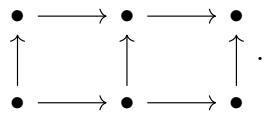
## 14 Multiparameter Persistent Homology

Persistent homology is stable with respect to perturbation, but at the same time highly sensitive to noise and outliers. Take for instance the noisy sample of a circle depicted in Fig. 51: its associated barcode in homology dimension 1 contains no long bars. One could attempt to rectify that by restricting to data points with a local density estimate greater than a predefined threshold  $\delta$ . However, the resulting barcode would be highly sensitive to the choice of  $\delta$ : chosen too large and there won't be enough points to cover a circle, chosen too small and the data will be too noisy. Another approach would be to consider all possible density thresholds at once and construct a  $P$ -module where  $P$  is a grid. That is the idea behind **multiparameter persistent homology**. Parts of the exposition in this section follows a recent introduction to multiparameter persistence [7].

### 14.1 Multiparameter Persistence

By a **multiparameter persistence module** we mean a  $P$ -module  $M$  where  $P = T_1 \times \cdots \times T_n$  and  $T_i \subseteq \mathbb{R}$ . Here  $P$  has the partial order  $(p_1, \dots, p_n) \leq (q_1, \dots, q_n)$  iff  $p_i \leq q_i$  for all  $1 \leq i \leq n$ . Our primary focus will be with the three cases:  $P = \mathbb{R}^2$ ,  $P = \mathbb{Z}^n$ , and  $P$  finite.

**Example 14.1.** An example of finite poset is  $P = [3] \times [2]$ :

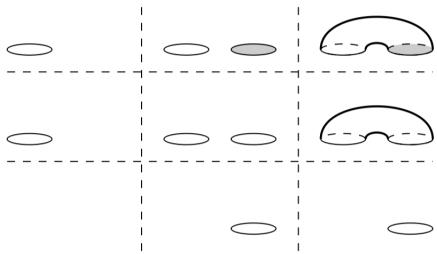


Multiparameter persistence modules typically arise from applying homology to a *multifiltration* of a topological space. As an example, let  $(Q, d)$  be a finite metric space, and let  $f: Q \rightarrow \mathbb{R}$  be any function. This defines an  $\mathbb{R}^2$ -module  $M$  by

$$M_{(x,y)} = H_p(\text{VR}_x(f^{-1}(-\infty, y]))$$

and where the map  $M_{(x,y)} \rightarrow M_{(x,y')}$  is induced in homology by the inclusion

$$\text{VR}_x(f^{-1}(-\infty, y)) \hookrightarrow \text{VR}_{x'}(f^{-1}(-\infty, y')).$$



(a)

$$\begin{array}{ccccccc}
 \mathbb{Z}_2 & \xrightarrow{\equiv} & \mathbb{Z}_2 & \longrightarrow & 0 \\
 \uparrow & & [1,0] \uparrow & & \uparrow \\
 \mathbb{Z}_2 & \xrightarrow{[1,0]^T} & \mathbb{Z}_2^2 & \xrightarrow{[1,1]} & \mathbb{Z}_2 \\
 \uparrow & & [0,1]^T \uparrow & & \uparrow \\
 0 & \longrightarrow & \mathbb{Z}_2 & \xrightarrow{\equiv} & \mathbb{Z}_2
 \end{array}$$

(b)

Figure 52: Applying  $H_1(-)$  to the bifiltration in (a) yields the indecomposable bipersistence module in (b).

The 2-parameter filtration  $\text{VR}_-(f^{-1}(-\infty, -])$  of  $\text{VR}_\infty(Q)$  is called the **sublevel Vietoris–Rips bifiltration**. In practice, we consider a discretized version of  $M$  obtained by choosing a finite set of distance thresholds  $A = \{\epsilon_1, \dots, \epsilon_n\}$  and function thresholds  $B = \{\delta_1, \dots, \delta_m\}$ . The restriction of  $M$  to  $A \times B$  yields an  $[n] \times [m]$ -module.

**Example 14.2.** Let  $Q = \{q_1, q_2, q_3\}$  with pairwise distances  $d(q_2, q_3) < d(q_1, q_3) = d(q_1, q_2)$  and  $f(q_1) = f(q_2) < f(q_3)$ , as depicted in Fig. 53. Let  $A = \{0, d(q_2, q_3), d(q_1, q_2)\}$  and  $B = \{f(q_1), f(q_3)\}$ . The resulting  $[3] \times [2]$ -module in dimension 0 homology is:

$$\begin{array}{ccccc}
 \text{span}\{q_1, q_2, q_3\} & \xrightarrow{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}} & \text{span}\{q_1, q_2\} & \xrightarrow{[1,1]} & \text{span}\{q_1\} \\
 \uparrow \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} & & \uparrow \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & & \uparrow 1 \\
 \text{span}\{q_1, q_2\} & \xrightarrow{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}} & \text{span}\{q_1, q_2\} & \xrightarrow{[1,1]} & \text{span}\{q_1\}
 \end{array}$$

By an appropriate change of basis this diagram transforms to:

$$\begin{array}{ccccc}
 \text{span}\{q_1, q_1 + q_2, q_2 + q_3\} & \xrightarrow{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}} & \text{span}\{q_1, q_1 + q_2\} & \xrightarrow{[1,0]} & \text{span}\{q_1\} \\
 \uparrow \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} & & \uparrow \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & & \uparrow 1 \\
 \text{span}\{q_1, q_1 + q_2\} & \xrightarrow{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}} & \text{span}\{q_1, q_1 + q_2\} & \xrightarrow{[1,0]} & \text{span}\{q_1\}
 \end{array}$$

which is isomorphic the direct sum of the following three modules:

$$\begin{array}{ccccccc}
 \mathbb{Z}_2 & \xrightarrow{1} & \mathbb{Z}_2 & \xrightarrow{1} & \mathbb{Z}_2 & , & \mathbb{Z}_2 & \xrightarrow{1} & \mathbb{Z}_2 & \longrightarrow & 0 & , & \mathbb{Z}_2 & \longrightarrow & 0 & \longrightarrow & 0 \\
 1 \uparrow & & 1 \uparrow & & 1 \uparrow & , & 1 \uparrow & & 1 \uparrow & & \uparrow & , & \uparrow & & \uparrow & & \uparrow \\
 \mathbb{Z}_2 & \xrightarrow{1} & \mathbb{Z}_2 & \xrightarrow{1} & \mathbb{Z}_2 & & \mathbb{Z}_2 & \xrightarrow{1} & \mathbb{Z}_2 & \longrightarrow & 0 & & 0 & \longrightarrow & 0 & \longrightarrow & 0
 \end{array}$$

With the previous example fresh in mind one may be tempted to conjecture that one can decompose multiparameter persistence modules into simple components as in Theorem 13.3.

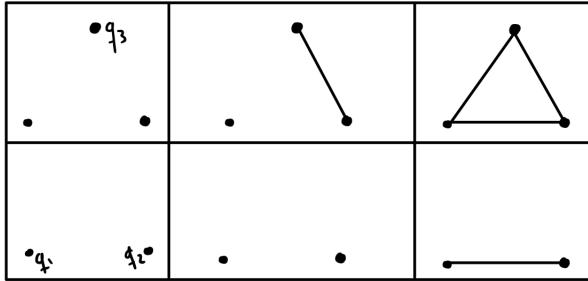


Figure 53: Bifiltration

That turns out to be very far from the truth: one can realize any persistence module over a finite grid by applying  $H_1(-)$  to a cellular or simplicial sublevel filtration. As an example, Fig. 52a illustrates how one can realize the *indecomposable* persistence module in Fig. 52b. In particular, the indecomposable persistence modules are no longer completely identified with a simple region in space, and it is unclear how such modules can be used in data analysis. Furthermore, while it follows from Lemma 6.12 that any  $[n] \times [m]$ -module decomposes into a direct sum of indecomposable modules, one does not even have a reasonable way of parametrizing all the indecomposable representations which could potentially arise<sup>7</sup>.

## 14.2 The Fibered Barcode and RIVET

Rather than seeking to understand the full complexity of multiparameter persistence modules, one can study them through the lens of restrictions to totally ordered subposets. Specifically, assuming that  $M$  is an  $\mathbb{R}^2$ -indexed persistence module, we can restrict  $M$  to an arbitrary straight line  $L$  in the plane with non-negative slope. The restricted module  $M|_L$  has a well-defined barcode by means of Theorem 6.19, assuming that the vector space dimensions are finite. As  $L$  ranges over all valid lines we can gain valuable insight into the underlying 2-parameter persistence module  $M$ .

**Definition 14.3.** Let  $M$  be an  $\mathbb{R}^2$ -module such that  $\dim M_{(x,y)} < \infty$  for all  $(x,y)$ . The **fibered barcode** of  $M$  is the map that sends a line  $L$  of non-negative slope to  $B(M|_L)$ .

*Remark 14.4.* This definition extends trivially to  $\mathbb{R}^n$ .

RIVET (the Rank Invariant Visualization and Exploration Tool) [29] allows for the visualization and analysis of two-parameter persistence modules, and arguably its main feature is the fast computation of the fibered barcode. The algorithm is rather involved and requires a discussion of commutative algebra beyond the scope of this course.

**Example 14.5.** Let  $Q$  be the points in Fig. 51 and let  $f: Q \rightarrow \mathbb{R}$  be the function which for each point  $q$  counts the number of data points within unit distance. Associated to this data RIVET considers the  $\mathbb{R}^2$ -indexed persistence module  $M$  given by

$$M_{(s,t)} = \text{VR}_t(f^{-1}[s, \infty)).$$

Be aware that the density filtration takes place in the *horizontal* direction, and that the scale parameter in RIVET - as in Ripser - is twice the one in the definition of the Vietoris–Rips complex used in these notes. Consider the subset of  $Q$  admitting a local density estimate at least 18, as seen in Fig. 54c. Its associated barcode in the Rips filtration is displayed along

---

<sup>7</sup>A so-called *wild* quiver.

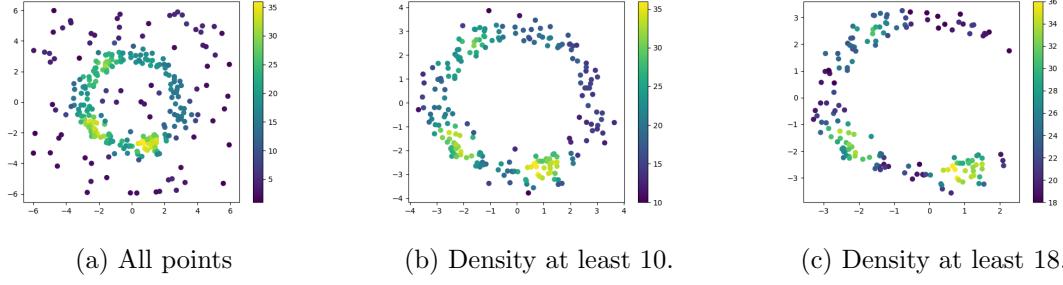


Figure 54: See Example 14.5.

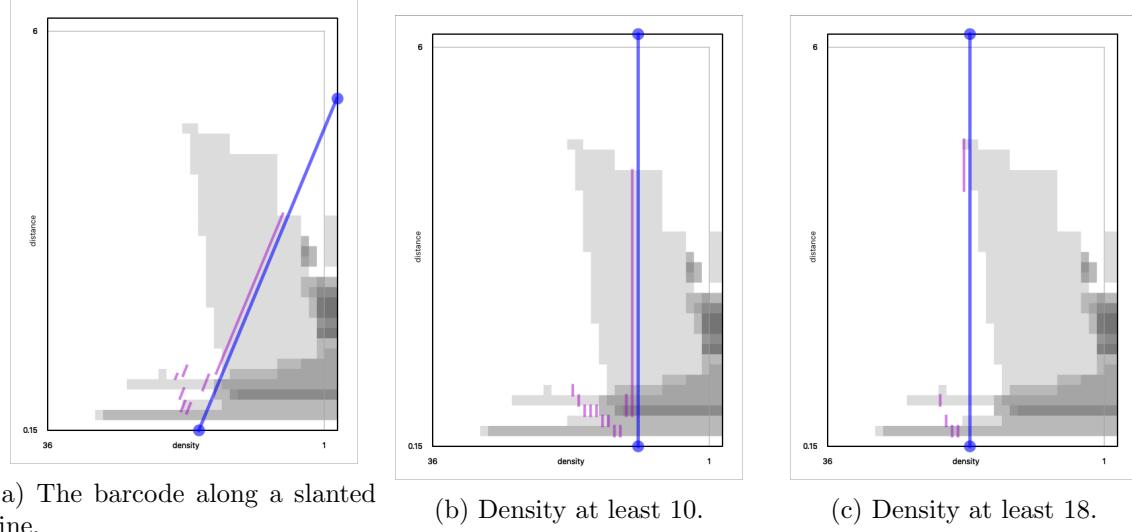


Figure 55: See Example 14.5.

the thick blue line in Fig. 55c. We see that there is a short bar generated at a fairly large scale corresponding to the scale at which the right-end of the point cloud connects. For a much cleaner barcode we consider all the points with local density estimate at least 10, as shown in Fig. 54b together with its barcode in Fig. 55b. It is often more fruitful to consider the barcode along a slanted line; see Fig. 55a. RIVET updates the barcode in real-time as the user modifies the slope and intercept.

### 14.3 Distances

With the failure of a barcode-like decomposition for multiparameter persistence module the question of how to compare modules arises. We now give an introduction to two well-studied distances in the context of multiparameter persistent homology.

**The interleaving distance** Generalizing the definition of an interleaving in 9.11, define the  **$\epsilon$ -shift** of an  $\mathbb{R}^n$ -modules  $M$  to be the persistence module  $M^\epsilon$  defined by  $M_p^\epsilon = M_{p+(\epsilon, \dots, \epsilon)}$  and  $M^\epsilon(p \leq p') = M(p+(\epsilon, \dots, \epsilon)) \leq p'+(\epsilon, \dots, \epsilon))$  for all  $p \leq p' \in \mathbb{R}^2$ . For a morphism  $f: M \rightarrow N$ , we get an  $\epsilon$ -shifted morphism  $f^\epsilon: M^\epsilon \rightarrow N^\epsilon$  defined by  $f_p^\epsilon = f_{p+(\epsilon, \dots, \epsilon)}$ . Let  $\eta_M^\epsilon: M \rightarrow M^\epsilon$  be the morphism whose restriction to each  $M_p$  is the internal morphism  $M(p \leq p + (\epsilon, \dots, \epsilon))$ .

**Definition 14.6.** Given  $\epsilon \in [0, \infty)$ , an  **$\epsilon$ -interleaving** between  $M$  and  $N$  is a pair of morphisms

$\psi : M \rightarrow N^\epsilon$  and  $\varphi : N \rightarrow M^\epsilon$  such that  $\varphi^\epsilon \circ \psi = \eta_M^{2\epsilon}$  and  $\psi^\epsilon \circ \varphi = \eta_N^{2\epsilon}$ . We say that  $M$  and  $N$  are  $\epsilon$ -interleaved.

Michael Lesnick[21] showed that the interleaving distance is the *most discriminative and stable* distance measure on multiparameter persistence modules<sup>8</sup>. Unfortunately, it turns out that its computation is NP-hard.

**Theorem 14.7** ([3]). *For  $n = 2$ , it is NP-hard to approximate the interleaving distance within a factor of 3.*

The proof proceeds by a reduction from the following problem which is subsequently shown to be NP-complete: let  $A$  and  $B$  be  $n \times n$ -matrices and assume that a pre-assigned subset of the indices of  $A$  and  $B$  are set 0. Is it possible to assign values to the remaining entries such that the resulting product  $AB$  is the identity matrix  $I_n$ ?

**Example 14.8.** Let

$$A = \begin{bmatrix} * & * & * \\ * & 0 & * \\ * & * & 0 \end{bmatrix} \quad B = \begin{bmatrix} * & * & * \\ * & * & 0 \\ * & 0 & * \end{bmatrix}.$$

In this case the decision problem has a positive answer:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} -1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix} = I_3.$$

**The matching distance** Theorem 14.7 motivates the search for a more computable surrogate for the interleaving distance. The matching distance [11] has emerged as a popular choice. It is defined by restricting the modules to (suitably parameterized) affine lines with positive slope, and taking bottleneck distances:

**Definition 14.9.** The **matching distance** between  $\mathbb{R}^n$ -indexed modules  $M$  and  $N$ , satisfying  $\dim M_p < \infty$  and  $\dim N_p < \infty$  for all  $p$ , is given by:

$$d_{\text{match}}(M, N) = \sup_L d_B(B(M|_L), B(N|_L)),$$

where  $L : \mathbb{R} \rightarrow \mathbb{R}^n$  ranges over lines of the form  $L(t) = vt + b$  where  $v \in [1, \infty)^n$  and  $b \in \mathbb{R}^n$ .

The following is an immediate consequence of Lesnick's universality result.

**Proposition 14.10.** *For  $\mathbb{R}^n$ -modules  $M$  and  $N$  satisfying  $\dim M_p < \infty$  and  $\dim N_p < \infty$  for all  $p$ ,*

$$d_{\text{match}}(M, N) \leq d_I(M, N).$$

While the matching distance is less sensitive than the interleaving distance, it can be computed exactly in polynomial time for bipersistence modules. Here the input size is given as the number of simplices in the bifiltration.

---

<sup>8</sup>Assuming the field is  $\mathbb{Z}_p$  or  $\mathbb{Q}$ .

## 14.4 The Rank Invariant

The rank invariant  $\text{Rk } M$  of a persistence module  $M$  is the function  $(s, t) \mapsto \text{Rk}(M_s \rightarrow M_t)$ , where  $s \leq t$ . In standard persistent homology, the rank invariant determines the barcode via inclusion-exclusion formula given in Eq. (3). Conversely, the rank-invariant can clearly be deduced from the barcode. Thus, the barcode and the rank invariant determine each other. For zigzag (and thus multiparameter persistence) two modules can be non-isomorphic but still have the same rank invariant.

**Example 14.11.** The following two zigzag modules are non-isomorphic but they have the same rank invariant:

$$\left( \mathbb{Z}_2 \xleftarrow{1} \mathbb{Z}_2 \xrightarrow{0} \mathbb{Z}_2 \right) \oplus \left( 0 \xleftarrow{0} \mathbb{Z}_2 \xrightarrow{0} 0 \right) \quad \left( \mathbb{Z}_2 \xleftarrow{1} \mathbb{Z}_2 \xrightarrow{0} 0 \right) \oplus \left( 0 \xleftarrow{0} \mathbb{Z}_2 \xrightarrow{1} \mathbb{Z}_2 \right).$$

Though the rank invariant is incomplete outside of the 1-parameter setting, one may hope that the correspondence between barcodes and rank invariants can be extended to the multi-parameter setting.

**Definition 14.12.** A multiset  $B$  of subsets of  $\mathbb{R}^n$  is a **good barcode** of  $M$  if for all  $x \leq y \in \mathbb{R}^n$  we have

$$\text{Rk}(M_x \rightarrow M_y) = |\{S \in B : x, y \in S\}|,$$

i.e., the rank of the map  $M_x \rightarrow M_y$  is the number of elements of  $B$  containing both  $x$  and  $y$ .

Given how barcodes of 1-parameter persistence modules are usually interpreted and used in TDA, Definition 14.12 is quite natural. However, the next example shows that a good barcode of  $M$  need not exist.

**Example 14.13.** A simple argument by contradiction shows that the bipersistence module  $M$  in Fig. 52 does not have a good barcode. If  $B$  is a good barcode of  $M$ , then since all the maps in the diagram have rank one,  $B$  must include an interval containing the whole poset. This contradicts the fact that  $\text{Rk}(M_{(1,0)} \rightarrow M_{(1,2)}) = 0$ .

As we will now explain, this is indeed possible if one considers *signed barcodes*.

## 14.5 Signed Barcodes

**Definition 14.14.** An **interval** in a poset  $P$  is a non-empty subset  $I$  of  $P$  satisfying the two following conditions:

1. If  $s, t \in I$  and  $s \leq u \leq t$ , then  $u \in I$ ,
2. If  $s, t \in I$ , then there are  $s = u_0, \dots, u_m = t \in I$  such that  $u_i$  and  $u_{i+1}$  are comparable for all  $0 \leq i < m$ .

Generalizing our discussion in Section 6, we associate a  $P$ -module to an interval. Here  $\mathbf{k}$  is any field.

**Definition 14.15.** For an interval  $J$  in a poset  $P$ , the *interval module*  $I^J$  is defined by

$$(I^J)_x = \begin{cases} \mathbf{k} & \text{if } x \in J, \\ 0 & \text{otherwise,} \end{cases} \quad (I^J)_{x,y} = \begin{cases} \text{id}_{\mathbf{k}} & \text{if } x \leq y \in J, \\ 0 & \text{otherwise.} \end{cases}$$

One can check that  $I^J$  is indecomposable. If  $M \cong \bigoplus J I^J$ , then we say that  $M$  is **interval-decomposable**. The rank invariant of  $M$  in Example 14.13 can be expressed as the *difference* between the rank invariants of two interval-decomposable modules:

$$\text{Rk} \left( \begin{array}{ccccc} & & & & 0 \\ & k & \xrightarrow{\text{id}} & k & \longrightarrow \\ & \downarrow \text{id} & & \downarrow [1,0] & \\ k & \xrightarrow{[1]} & k^2 & \xrightarrow{[1,1]} & k \\ \downarrow \text{id} & & \downarrow [0] & & \downarrow \text{id} \\ 0 & \longrightarrow & k & \xrightarrow{\text{id}} & k \end{array} \right) = \text{Rk} \left( \begin{array}{c} \text{Diagram 1} \\ \oplus \\ \text{Diagram 2} \end{array} \right) - \text{Rk} \left( \begin{array}{c} \text{Diagram 3} \end{array} \right) \quad (20)$$

Such a decomposition need not be unique:

$$\begin{aligned} \text{Rk} \left( \begin{array}{ccccc} & & & & 0 \\ & k & \xrightarrow{\text{id}} & k & \longrightarrow \\ & \downarrow \text{id} & & \downarrow [1,0] & \\ k & \xrightarrow{[1]} & k^2 & \xrightarrow{[1,1]} & k \\ \downarrow \text{id} & & \downarrow [0] & & \downarrow \text{id} \\ 0 & \longrightarrow & k & \xrightarrow{\text{id}} & k \end{array} \right) &= \text{Rk} \left( \begin{array}{c} \text{Diagram 1} \\ \oplus \\ \text{Diagram 4} \\ \oplus \\ \text{Diagram 5} \\ \oplus \\ \text{Diagram 6} \end{array} \right) \\ &\quad - \text{Rk} \left( \begin{array}{c} \text{Diagram 7} \\ \oplus \\ \text{Diagram 8} \end{array} \right) \end{aligned} \quad (21)$$

It turns out, however, that the decomposition is unique if we restrict our attention to rectangles.

**Definition 14.16.** A *rectangle* in  $P$  is a non-empty subset of the form  $[s, t] = \{u : s \leq u \leq t\}$ .

**Theorem 14.17.** Let  $P = T_1 \times \cdots \times T_n$ , where each  $T_i \subseteq \mathbb{R}$  is finite. For any  $P$ -module  $M$ , satisfying  $\dim M_p < \infty$  for all  $p$ , there exists a pair  $(\mathcal{R}, \mathcal{S})$  of finite multisets of rectangles in  $P$  such that

$$\text{Rk } M = \text{Rk} \left( \bigoplus_{R \in \mathcal{R}} I^R \right) - \text{Rk} \left( \bigoplus_{S \in \mathcal{S}} I^S \right) \quad (22)$$

and if  $(\mathcal{R}', \mathcal{S}')$  is any other such pair satisfying Eq. (22), then  $\mathcal{R} \subseteq \mathcal{R}'$  and  $\mathcal{S} \subseteq \mathcal{S}'$ .

We shall refer to the pair  $(\mathcal{R}, \mathcal{S})$  as the **signed barcode** of  $\text{Rk } M$ . Note that if  $d = 1$ , then by Theorem 6.19 and the uniqueness of the signed barcode, we must have that  $\mathcal{S} = \emptyset$ . We shall prove Theorem 14.17 by first proving the following closely related statement.

**Theorem 14.18.** Let  $M$  and  $P$  be as in Theorem 14.17. Then  $M$  can be expressed uniquely as a  $\mathbb{Z}$ -linear combination of the rank functions  $\{\text{Rk}(I^R) : R \text{ is a rectangle in } P\}$ . That is, there exist unique integers  $\alpha_R$  such that

$$\text{Rk } M = \sum_{\text{Rectangles } R} \alpha_R \text{Rk}(I^R).$$

*Proof.* For a rectangle  $R = [s, t]$ , let  $v(R) = \text{Rk}(M_s \rightarrow M_t)$ , and define  $U(R) = \{R' \supseteq R : v(R') > 0\}$ . Observe that if  $v(R) > 0$  and  $R \neq R' \in U(R)$ , then  $|U(R)| > |U(R')|$ .

We shall inductively define the integers  $\alpha_R$ . If  $v(R) = 0$ , then let  $\alpha_R = 0$ . Continuing inductively on the size of  $|U(R)|$ , define

$$\alpha_R = v(R) - \sum_{R' \in U(R)} \alpha_{R'}.$$

Note that this is well-defined. Now, let  $s \leq t$ , and observe that for a rectangle  $R = [s', t']$ ,  $\text{Rk}(I^R)(s, t) = 1$  if  $s, t \in [s', t']$ , and 0 otherwise. In particular,

$$\sum_R \alpha_R \text{Rk}(I^R)(s, t) = \sum_{R \supseteq [s, t]} \alpha_R \text{Rk}(I^R)(s, t) = \sum_{R \in U([s, t])} \alpha_R \text{Rk}(I^R)(s, t) = \sum_{R \in U([s, t])} \alpha_R = v(R).$$

To see that this unique, assume that

$$\sum_R \alpha_R \text{Rk}(I^R) = \sum_R \beta_R \text{Rk}(I^R)$$

where  $\alpha_R \neq \beta_R$  for at least one  $R$ . This yields a  $\mathbb{Z}$ -linear combination

$$\sum_R \gamma_R \text{Rk}(I^R) = 0$$

where at least one  $\gamma_R \neq 0$ . Choose a maximal  $[s, t]$  such that  $\gamma_{[s, t]} \neq 0$ . Then,

$$0 = \sum_R \gamma_R \text{Rk}(I^R)(s, t) = \gamma_{[s, t]},$$

contradicting that  $\gamma_{[s, t]} \neq 0$ .  $\square$

Theorem 14.17 now follows by letting  $\mathcal{R} = \{R : \alpha_R > 0\}$  and  $\mathcal{S} = \{R : \alpha_R < 0\}$  in the decomposition

$$\text{Rk } M = \sum_{\text{Rectangles } R} \alpha_R \text{Rk}(I^R).$$

**Computation** The signed barcode of  $M$  indexed over a finite grid  $\prod_{i=1}^d [1, n_i]$  can be computed by means of a simple inclusion-exclusion formula. Let  $\alpha_{[s, t]}$  denote the multiplicity of the rectangle

$$[s, t] = \{u \in \mathbb{R}^d \mid s \leq u \leq t\},$$

in the signed barcode  $(\mathcal{R}, \mathcal{S})$ . That is, if  $\alpha_{[s, t]} > 0$  then  $[s, t]$  appears with multiplicity  $\alpha_{[s, t]}$  in  $\mathcal{R}$ , and if  $\alpha_{[s, t]} < 0$  then  $[s, t]$  appears with multiplicity  $-\alpha_{[s, t]}$  in  $\mathcal{S}$ . Then we have:

$$\alpha_{[s, t]} = \sum_{\substack{s' \leq s \\ \|s' - s\|_\infty \leq 1}} \sum_{\substack{t' \geq t \\ \|t' - t\|_\infty \leq 1}} (-1)^{\|s' - s\|_1 + \|t' - t\|_1} \text{Rk } M(s', t'). \quad (23)$$

*Remark 14.19.* The case  $d = 1$  gives the well-known inclusion-exclusion formula relating the persistence diagram of a one-parameter persistence module to its rank invariant.

A simple inspection of the formula in Eq. (23) reveals that computation is bounded in time  $O(2^{2d} \prod_{i=1}^d n_i^2)$ , assuming constant-time access to the ranks  $\text{Rk } M(s', t')$  and constant-time arithmetic operations.

**Visualization** Visualizing the signed barcode of a bipersistence module by directly plotting the rectangles can be messy for examples of realistic size, since many rectangles may overlap. To obtain a cleaner visualization, one can represent each rectangle via a line segment (“bar”) connecting its infimum to its supremum, as illustrated in Fig. 56 (b). Such visualizations of the signed barcodes of larger persistence modules can be found in [8].

$$\text{Rk} \left( \begin{array}{c} \text{gray rectangles} \\ \mathbf{k} \end{array} \right) = \text{Rk} \left( \begin{array}{c} \text{blue rectangles} \\ \mathbf{k}^4 \quad \mathbf{k}^3 \quad \mathbf{k}^2 \quad \mathbf{k} \end{array} \right) - \text{Rk} \left( \begin{array}{c} \text{red rectangles} \\ \mathbf{k}^3 \quad \mathbf{k}^2 \quad \mathbf{k} \end{array} \right)$$

(a)

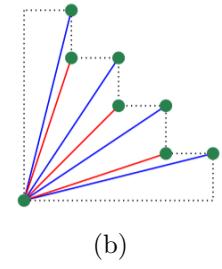


Figure 56: (a) shows the rectangles in a rank decomposition, and (b) shows the same decomposition using line segments.

## 14.6 Exercises

Show that Eq. (23) for  $d = 1$  reduces to the standard formula for computing the barcode in one-parameter persistence.

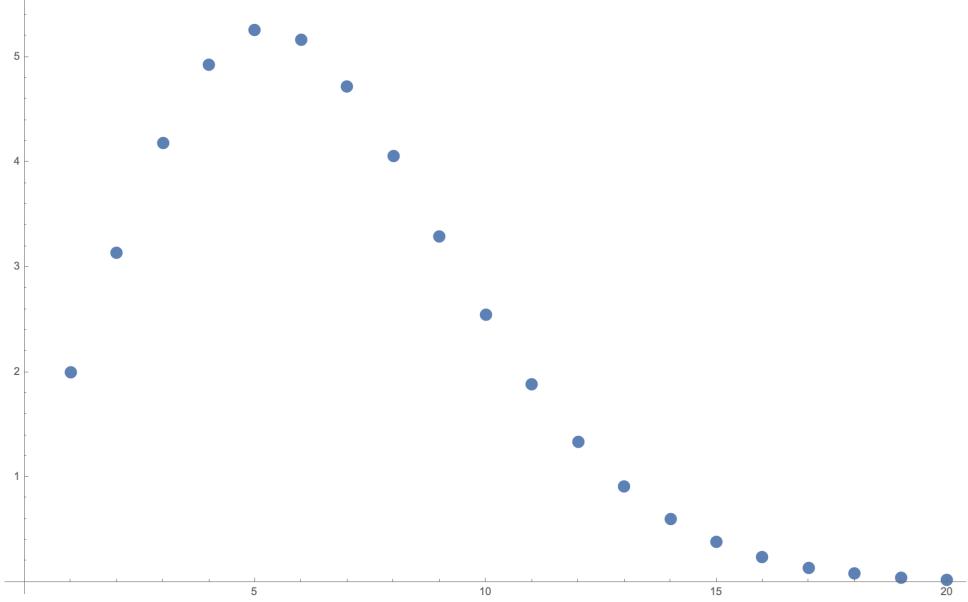


Figure 57: The volume of the unit  $n$ -ball for  $1 \leq n \leq 20$ .

## 15 Non-linear dimensionality reduction

Working with data in high-dimensional space presents several challenges, commonly referred to as the "curse of dimensionality." One significant issue is that data becomes extremely sparse as the number of dimensions increases. This is reflected in the fact that the volume of the unit  $n$ -ball  $\mathbb{D}^n$  goes to 0 as  $n$  increases; see Fig. 57. Because of this sparseness, the amount of data needed for a 'sufficiently' dense sampling grows exponentially with the number of dimensions rendering (topological) inference useless. However, data tends to have additional structure and sit near a, not necessarily linear, low-dimensional subspace. The goal of low-dimensional embedding techniques is to represent the data with respect to coordinates of this lower-dimensional space. Such dimension reductions are essential in uncovering structure in data, as strong (topological) signals can otherwise be totally dominated by a small amounts of noise in each component. The signal can be recovered by first passing to a lower-dimensional space as the following example illustrates. The goal of this section is to explain the mathematics behind *ISOMAP* (Section 15.1) and *UMAP* (Section 15.2).

**Example 15.1.** Consider the 500 points on a torus knot in  $\mathbb{R}^3$  in Figure Fig. 58a and the corresponding persistence diagram in Fig. 58b. As expected, there is a single significant feature in degree 1 homology. Now, let us embed the same point cloud as the three first coordinates in  $\mathbb{R}^{50000}$ , and perturb each coordinate of each point by a random number in  $[-0.2, 0.2]$ . The corresponding persistence diagram is shown in Fig. 59b. We observe that the signal is dominated by the noise. However, by applying *principal component analysis (PCA)* (Section 15.1.1), we can embed the data in  $\mathbb{R}^3$  and recover the circular structure; see Fig. 59a. Figure Fig. 60 shows the embedding of the same point cloud into  $\mathbb{R}^2$  using PCA, ISOMAP, and UMAP.

### 15.1 ISOMAP

We shall assume that we are given  $n$  data points in  $\mathbb{R}^d$  stored as a  $n \times d$  matrix  $A$ . In order to explain how ISOMAP works, we need to briefly discuss PCA and MDS.

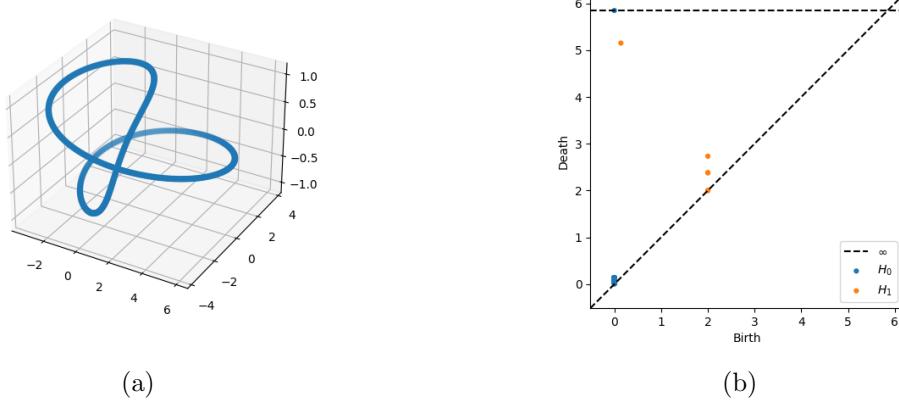


Figure 58: See Example 15.1. (a) A torus knot in  $\mathbb{R}^3$ . (b) The persistence diagram of the associated Rips filtration.

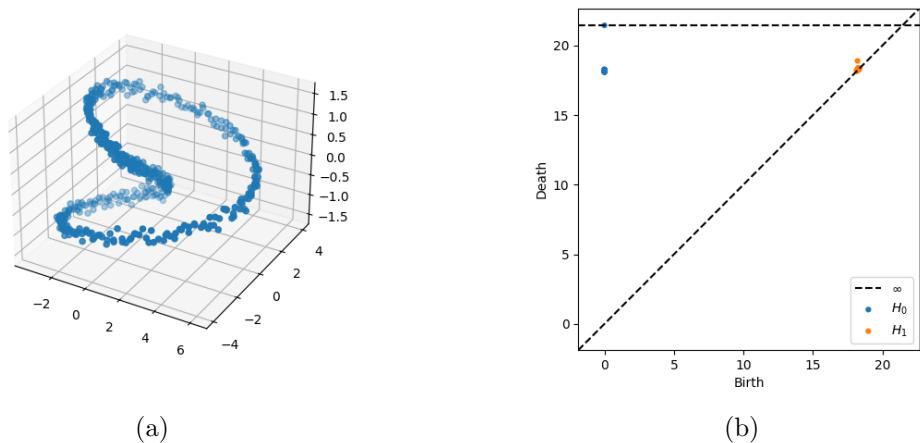


Figure 59: See Example 15.1. (a) The projection of the point cloud onto the three first principal components. (b) The persistence diagram of the original point cloud in  $\mathbb{R}^{50000}$ .

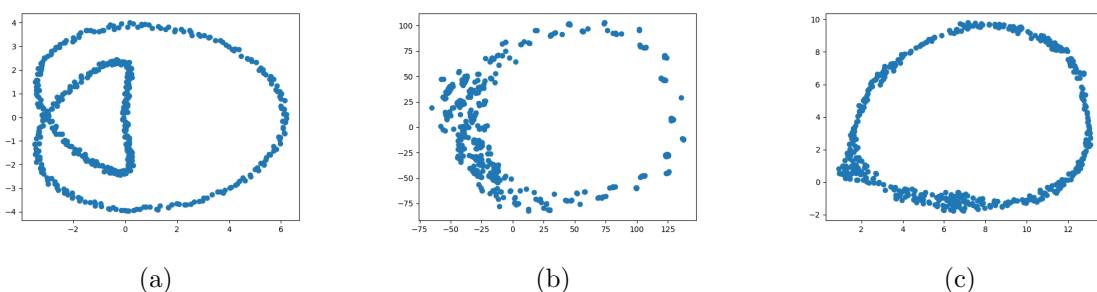


Figure 60: See Example 15.1. (a) PCA. (b) ISOMAP. (c) UMAP.

### 15.1.1 Principal component analysis

Principal component analysis is a *linear* dimensionality reduction technique. For a fixed  $k$ , PCA projects the data onto the  $k$ -dimensional subspace of  $\mathbb{R}^d$  which best approximates the data.

Recall that any real-valued matrix admits a *singular value decomposition* (SVD) and so

$$A = U\Sigma V^T$$

where  $U$  is an  $n \times n$  orthogonal matrix,  $\Sigma$  is a  $n \times d$  diagonal matrix, and  $V$  is a  $d \times d$  orthogonal matrix. Moreover, the entries of  $\Sigma$  are non-negative and satisfy  $\Sigma_{1,1} \geq \Sigma_{2,2} \geq \dots$ . We let  $\sigma_i = \Sigma_{i,i}$ . For an  $n \times d$  matrix  $B$ , we shall let  $B[k]$  denote the restriction of  $B$  to its first  $k$  columns, and we define the 2-norm  $B$  to be

$$\|B\|_2 = \sqrt{\sum_{(1,1) \leq (i,j) \leq (n,d)} B_{i,j}^2}.$$

**Theorem 15.2.** *Let  $A$  and  $B$  be  $n \times d$  matrices. If the rank of  $B$  is at most  $k$ , then*

$$\|A - A_k\|_2 \leq \|A - B\|_2.$$

Furthermore,

$$\|A - A_k\|_2^2 = \sum_{i>k} \sigma_i^2.$$

Now, assume that the columns of  $B[k]$  is an orthonormal basis for a  $k$ -dimensional subspace  $W$  of  $\mathbb{R}^n$ . Then  $AB[k]$  has rank at most  $k$ , and the orthogonal projection of the data points in  $A$  onto  $W$  is given by  $AB[k]$ . Observe that if  $\pi_W(p)$  denotes the orthogonal projection of  $p$  onto  $W$ , then

$$\text{Err}(P, W)^2 := \sum_{p \in P} \|p - \pi_W(p)\|_2^2 = \|A - AB[k]\|_2^2.$$

To find a best possible  $k$ -linear approximation to the data in  $A$ , we would thus like to find a  $k$ -subspace  $O_k$  such that  $\text{Err}(P, O_k)^2$  is minimal. Observing that  $A[k] = AV[k]$ , it follows from Theorem 15.2 that  $\text{Err}(P, O_k)$  is minimized when  $O_k$  is the  $k$ -subspace by the first  $k$  columns in  $V$ .

If the data were to be located on a  $k$ -hyperplane not passing through the origin, then a  $k+1$ -dimensional subspace would be needed to adequately capture the variance in the data. For that reason, as a preprocessing step, the data is first centralized by subtracting the mean in each coordinate. Then the projected data is obtained by computing  $AV[k]$  for some user-defined  $k$ . Note that  $\text{Err}(P, O_k)^2 = \sum_{i>k} \sigma_i^2$  and therefore one chooses  $k$  such the singular values  $\sigma_i$  for  $i > k$  are negligible. Often (but not always) there is some  $i$  such that  $\sigma_i \gg \sigma_{i+1}$ .

### 15.1.2 Multidimensional scaling

We shall now see that a low-dimensional embedding closely related to that of PCA can be obtained from the pairwise distances between the data points. Following the notation from Section 15.1.1, and using that  $A = U\Sigma V^T$ , we get  $AA^T = U\Sigma U^T$  and  $U$  are therefore the eigenvectors of  $AA^T$ . Furthermore, observe that  $(AA^T)_{i,j} = p_i \cdot p_j$ . In conclusion: if we can compute all the pairwise inner products, then we can do PCA. An elementary calculation gives

$$(p_i \cdot p_j)^2 = \frac{1}{2}(-\|p_i - p_j\|_2^2 + \|p_i\|_2^2 + \|p_j\|_2^2).$$

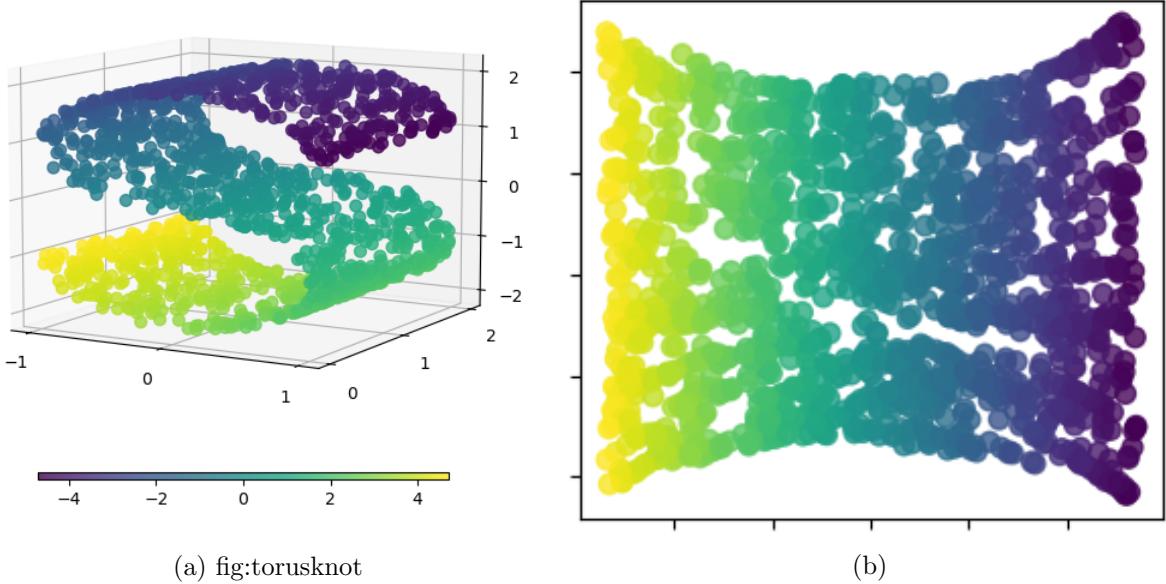


Figure 61: (a) The data lies on an S-shape in  $\mathbb{R}^3$  (b) The ISOMAP projection to  $\mathbb{R}^2$  for suitable parameters.

By assumption,  $\|p_i - p_j\|_2^2$  is given for all pairs  $i$  and  $j$ . But what do we do about the remaining norms? If we assume that some arbitrary point, say  $p_0$ , sits at the origin, then  $\|p_i\|_2^2 = \|p_i - p_0\|_2^2$ . This gives us sufficient data to compute  $AA^T$ , and therefore the eigenvectors  $U$  and the singular values  $\Sigma$ . We can now project the data onto  $V[k]$  as in Section 15.1.1. This approach to embedding a set of data points in Euclidean space is called *multidimensional scaling* (MDS).

### 15.1.3 The ISOMAP algorithm

ISOMAP assumes that the data lies on the isometric embedding  $\phi(C) \subset \mathbb{R}^d$  of some convex set  $C \subset \mathbb{R}^k$ . The standard example of this is the "swiss roll", or an S-shape, see Fig. 61. The important point is that the intrinsic distance  $d_C(p, q)$  between the points  $p$  and  $q$  on  $\phi(C)$  equals the Euclidean distance between  $\phi^{-1}(p)$  and  $\phi^{-1}(q)$ . We can therefore apply MDS to our estimate for  $d_C$  and obtain an embedding in  $\mathbb{R}^k$ . This leaves us with the following algorithm:

1. Estimate the intrinsic distance between any two points  $p_i$  and  $p_j$  to obtain a matrix of distances  $D_{i,j} \sim d_C(p_i, p_j)$ .
2. Apply MDS to  $D$  to get an embedding of the data points in  $\mathbb{R}^l$  for some user-defined  $l$ .

The estimation of the intrinsic distance can be done by forming the  $m$ -th nearest neighbor graph for some fixed  $m \geq 1$ . The weight associated to the edge is then the distance between  $p_i$  and  $p_j$  in the euclidean metric on  $\mathbb{R}^d$ . The intrinsic distance from  $p_i$  to  $p_k$  is then approximated by the shortest path in the graph, which can be calculated with standard graph-theoretic algorithms, e.g., Dijkstra's algorithm.

## 15.2 UMAP

*Uniform Manifold Approximation and Projection for Dimension Reduction* (UMAP) is a more recent approach to non-linear dimensionality reduction which has become immensely popular. At a high level, the algorithm associates to a set of data a weighted simplicial complex with

weights in  $[0, 1]$ ; here the weight is to be interpreted as the probability of that simplex being present. Then, a low-dimensional embedding of the data is sought for which the associated simplicial complex has similar probabilities. For the sake of efficiency, only the 1-skeletons of the simplicial complexes are considered.

### 15.2.1 Step 1: the local fuzzy graph

Fix some integer  $k \geq 1$ , and for each point  $p \in P \subset \mathbb{R}^d$ , let  $N_k(p)$  denote the  $k$ -th nearest neighbor graph of  $p$ . We shall let  $d: \mathbb{R}^d \times \mathbb{R}^d$  denote the distance on  $\mathbb{R}^d$ , e.g., the Euclidean distance. Our first goal is to associate to every  $p$  a *fuzzy* graph with vertex set  $P$ , i.e., a symmetric function  $w_p: P \times P \rightarrow [0, 1]$ . To do this, we let  $\rho_p := \min_{q \neq p} d(p, q)$ , i.e., the distance to  $p$ 's nearest neighbor, and define

$$d_p(p, q) = \begin{cases} d(p, q) - \rho_p & \text{if } \{p, q\} \in N_k(p) \\ \infty & \text{otherwise.} \end{cases}.$$

Set

$$w_p(q, r) = \exp(-d_p(q, r)/\sigma_p)$$

where  $\sigma_p$  is such that

$$\sum_{(p, q) \in N_p(k)} w_p(p, q) = \log_2(k).$$

The intuition behind this (I think) is that  $\sum_{(p, q) \in N_p(k)} w_p(p, q)$  represents the expected number of points connected to  $p$ , and it is assumed that the sampling is such that all points are expected to have the same degree. That is, no point is more connected than any other point and while average distances can vary locally from one point to another, this is just an artifact of a uniform sampling on a Riemannian manifold which has been deformed under a (not necessarily isometric) embedding into  $\mathbb{R}^d$ . That the expected number should be  $\log_2(k)$  is a choice; it seems reasonable that it should increase slowly with  $k$ .

### 15.2.2 Step 2: gluing fuzzy graphs

For each  $p \in P$  we now have a weight function  $w_p$  and we would like to glue them all together into a single weight function  $w: P \times P \rightarrow [0, 1]$ . Keeping the intuition that the weights represent probabilities, we define  $w(p, q)$  to be the probability that the edge  $\{p, q\}$  appears in at least one of the fuzzy graphs. In other words,

$$w(q, r) = 1 - \prod_{p \in P} (1 - w_p(q, r)) = 1 - (1 - w_q(q, r))(1 - w_r(q, r)) = w_q(q, r) + w_r(q, r) - w_q(q, r)w_r(q, r).$$

And what about higher-dimensional simplices? Taking a cue from Rips complexes, we let

$$w(p_0, \dots, p_n) = \min_{1 \leq i, j \leq n} w(p_i, p_j),$$

i.e., the probability of the *least* probable edge. A simplicial complex  $K$  with a function  $f: K \rightarrow [0, 1]$  such that  $f(\sigma) \geq f(\tau)$  if  $\sigma \supseteq \tau$  is called a *fuzzy simplicial complex* and  $f$  is a *fuzzy map*.

### 15.2.3 Step 3: finding a low-dimensional embedding

Given a simplicial complex  $K$  and two fuzzy maps  $f, g: K \rightarrow [0, 1]$ , one can compare them as probability spaces using Kullback-Leibner (KL) divergence or other similarity measures. The authors of UMAP make a small modification to KL divergence, and introduce the following measure of *cross entropy* between  $f$  and  $g$ :

$$C(f, g) = \sum_{\sigma \in K} \left( f(\sigma) \log \left( \frac{f(\sigma)}{g(\sigma)} \right) + (1 - f(\sigma)) \log \left( \frac{1 - f(\sigma)}{1 - g(\sigma)} \right) \right).$$

Here interpretation is that  $0 \cdot \log(0/a) = 0$  for all  $a$ , and  $a \cdot \log(a/0) = +\infty$  if  $a \neq 0$ . In practice when comparing discrete probability distributions, 0 is typically replaced with some small positive number to avoid division by 0.

Returning to our construction in Section 15.2.2,  $w$  plays the role of  $f$ , and we seek a low-dimensional embedding of the data such that the associated probability distribution  $v$  minimizes  $C(w, v)$ . Elementary algebra gives

$$\begin{aligned} C(w, v) &= \sum_{\sigma \in K} \left( w(\sigma) \log \left( \frac{w(\sigma)}{v(\sigma)} \right) + (1 - w(\sigma)) \log \left( \frac{1 - w(\sigma)}{1 - v(\sigma)} \right) \right) \\ &= \sum_{\sigma \in K} (w(\sigma) \log(w(\sigma)) + (1 - w(\sigma)) \log(1 - w(\sigma))) \\ &\quad - \sum_{\sigma \in K} (w(\sigma) \log(v(\sigma)) + (1 - w(\sigma)) \log(1 - v(\sigma))) \end{aligned}$$

Since the first term only depends on  $w$ , it suffices to minimize the second term. UMAP initializes the points in  $\mathbb{R}^k$  using a *spectral embedding* and then moves the points around such that

$$- \sum_{\sigma \in K} (w(\sigma) \log(v(\sigma)) + (1 - w(\sigma)) \log(1 - v(\sigma)))$$

hits a local minimum. The question that remains to be answered is how  $v$  ought to be defined. When we defined  $w$ , we assumed that the local geometry of the embedded manifold is different from that of the ambient Euclidean space. For the embedding however, the assumption is that the geometry is precisely that of the ambient space. Furthermore, UMAP requires you to set a *min\_dist* hyper-parameter, which should reflect that points at a distance smaller than *min\_dist* are connected with probability 1. For two points  $(p, q) \in \mathbb{R}^l$ , set

$$v(p, q) = \begin{cases} 1 & \text{if } d(p, q) \leq \text{min\_dist} \\ \exp(-(d(p, q) - \text{min\_dist})) & \text{otherwise.} \end{cases}$$

This is then extended to higher-dimensional simplices by taking the minimum over edges as before. However, for the sake of efficiency, UMAP only works with the 1-skeleton of  $K$ . To find a local minimum for  $C(w, v)$  UMAP applies stochastic gradient descent, and a necessary step is to replace  $v$  with a smooth approximation  $\psi$ . This smooth approximation will also ensure that one does not run into issues with  $\log(0)$ . Specifically,

$$\psi(p, q) = \left( 1 + a(d(p, q)^2)^b \right)^{-1}$$

where  $a$  and  $b$  are chosen by non-linear least squares fitting against  $v$ .

### 15.3 Exercises

1. Show that any  $n \times n$  distance matrix can be realized by  $n$  points in  $\mathbb{R}^n$  with the Euclidean distance.

## References

- [1] Ulrich Bauer. Ripser: efficient computation of vietoris-rips persistence barcodes. *arXiv preprint arXiv:1908.02518*, 2019.
- [2] Håvard Bakke Bjerkevik. Stability of higher-dimensional interval decomposable persistence modules. *arXiv preprint arXiv:1609.02086*, 2016.
- [3] Håvard Bakke Bjerkevik, Magnus Bakke Botnan, and Michael Kerber. Computing the interleaving distance is np-hard. *Foundations of Computational Mathematics*, pages 1–35, 2019.
- [4] Håvard Bakke Bjerkevik and Michael Lesnick.  $l_p$ -distances on multiparameter persistence modules. *arXiv preprint arXiv:2106.13589*, 2021.
- [5] Jean-Daniel Boissonnat, Frédéric Chazal, and Mariette Yvinec. *Geometric and topological inference*, volume 57. Cambridge University Press, 2018.
- [6] Magnus Botnan and William Crawley-Boevey. Decomposition of persistence modules. *Proceedings of the American Mathematical Society*, 148(11):4581–4596, 2020.
- [7] Magnus Bakke Botnan and Michael Lesnick. An introduction to multiparameter persistence. *arXiv preprint arXiv:2203.14289*, 2022.
- [8] Magnus Bakke Botnan, Steffen Oppermann, and Steve Oudot. Signed barcodes for multiparameter persistence via rank decompositions and rank-exact resolutions. *arXiv preprint arXiv:2107.06800*, 2021.
- [9] Gunnar Carlsson and Vin De Silva. Zigzag persistence. *Foundations of computational mathematics*, 10:367–405, 2010.
- [10] Gunnar Carlsson, Vin De Silva, and Dmitriy Morozov. Zigzag persistent homology and real-valued functions. In *Proceedings of the twenty-fifth annual symposium on Computational geometry*, pages 247–256, 2009.
- [11] Andrea Cerri, Barbara Di Fabio, Massimo Ferri, Patrizio Frosini, and Claudia Landi. Betti numbers in multidimensional persistent homology are stable functions. *Mathematical Methods in the Applied Sciences*, 36(12):1543–1557, 2013.
- [12] Frédéric Chazal, Leonidas J Guibas, Steve Y Oudot, and Primoz Skraba. Persistence-based clustering in riemannian manifolds. *Journal of the ACM (JACM)*, 60(6):1–38, 2013.
- [13] Vin De Silva and Robert Ghrist. Coordinate-free coverage in sensor networks with controlled boundaries via homology. *The International Journal of Robotics Research*, 25(12):1205–1222, 2006.
- [14] Vin De Silva and Robert Ghrist. Homological sensor networks. *Notices of the American mathematical society*, 54(1), 2007.
- [15] Vin De Silva, Elizabeth Munch, and Amit Patel. Categorified reeb graphs. *Discrete & Computational Geometry*, 55(4):854–906, 2016.
- [16] Tamal Krishna Dey and Yusu Wang. *Computational Topology for Data Analysis*. Cambridge University Press, 2022.

- [17] Herbert Edelsbrunner and John Harer. *Computational topology: an introduction*. American Mathematical Soc., 2010.
- [18] Michael Kerber, Dmitriy Morozov, and Arnur Nigmetov. Geometry helps to compare persistence diagrams, 2017.
- [19] Jon Kleinberg. An impossibility theorem for clustering. *Advances in neural information processing systems*, 15, 2002.
- [20] Dmitry Kozlov. *Combinatorial algebraic topology*, volume 21. Springer Science & Business Media, 2008.
- [21] Michael Lesnick. The theory of the interleaving distance on multidimensional persistence modules. *Foundations of Computational Mathematics*, 15(3):613–650, 2015.
- [22] Jie Liang, Herbert Edelsbrunner, Ping Fu, Pamidighantam V Sudhakar, and Shankar Subramaniam. Analytical shape computation of macromolecules: I. molecular area and volume through alpha shape. *Proteins: Structure, Function, and Bioinformatics*, 33(1):1–17, 1998.
- [23] Clément Maria and Steve Y Oudot. Zigzag persistence via reflections and transpositions. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 181–199. SIAM, 2014.
- [24] Yukio Matsumoto. *An introduction to Morse theory*, volume 208. American Mathematical Soc., 2002.
- [25] James R Munkres. *Elements of algebraic topology*. CRC Press, 2018.
- [26] Steve Y Oudot. *Persistence theory: from quiver representations to data analysis*, volume 209. American Mathematical Society Providence, 2015.
- [27] Gurjeet Singh, Facundo Mémoli, and Gunnar E Carlsson. Topological methods for the analysis of high dimensional data sets and 3d object recognition. In *SPBG*, pages 91–100, 2007.
- [28] Primoz Skraba and Katharine Turner. Wasserstein stability for persistence diagrams. *arXiv preprint arXiv:2006.16824*, 2020.
- [29] The RIVET Developers. Rivet. <https://github.com/rivetTDA/rivet>, 2020.
- [30] Christopher Tralie, Nathaniel Saul, and Rann Bar-On. Ripser.py: A lean persistent homology library for python. *The Journal of Open Source Software*, 3(29):925, Sep 2018.