# Introduction to Topological Data Analysis
# Lecture Notes FS 2023

Patrick Schnider <`patrick.schnider@inf.ethz.ch`>

July 13, 2023

## Preface

These lecture notes are designed to accompany a course “Introduction to topological data analysis” that I teach at the Department of Computer Science, ETH Zürich, since 2023. The course is intended for students with a background in computer science or data science. It requires knowledge of linear algebra, but does not assume any previous experience with topology.

The course can be roughly divided into four parts. In the first part, we go over the necessary mathematical foundations, in particular concepts from algebraic topology such as *homology*. In the second part, we study the persistent homology pipeline. In the third part, we discuss Reeb graphs and the Mapper algorithm. Finally, the fourth part contains other applications of topology in computer and data science, as well as applications of topological data analysis to other fields. At the end of each chapter there is a list of questions that students are expected to be able to answer in the oral exam.

In the current setting, the course runs over 14 weeks, with three hours of lectures and two hours of exercises each week. In addition, there are two sets of graded homeworks which students have to hand in spread over the course.

These notes are an extended version of the scribe notes written by Simon Weber from the first iteration of the course. We have tried our best to avoid mistakes, but experience tells that there will be many that escape our detection. So in case you notice some problem, please let me know, regardless of whether it is a minor typo or punctuation error, a glitch in formulation, or a hole in an argument. This way the issue can be fixed for the next edition and future readers profit from your findings.

I thank Anton Künzi and Simon Weber for their helpful contributions.

Patrick Schnider
Department of Computer Science, ETH Zürich
Andreasstrasse 5, CH-8050 Zürich, Switzerland
E-mail address: `patrick.schnider@inf.ethz.ch`

# Contents

# Chapter 1

# Mathematical Foundations

## 1.1 Topological Spaces

Topology, sometimes also called "rubber-sheet goemtetry", stems from the Greek words *tópos*, which means place or locality, and *lógos*, which means study. So, it can be roughly translated as the study of places and shapes. Indeed, as the name rubber-sheet geometry suggests, topology studies similar objects as geometry, but in a setting where properties are preserved under continuous deformations like stretching and twisting. In particular, these properties should be independent of metrics, but we would still like to have ways to describe proximity between points. We do this by looking at open neighborhoods of points. The core objects in topology are *topological spaces*, whose definition captures the system of open neighborhoods of the points in the space.

**Definition 1.1.** *A topological space* $(X, T)$ *is a set of points* $X$*, with a system* $T$ *of subsets of* $X$ *(called the topology on* $X$*), such that*

1. $\emptyset \in T$, $X \in T$.
2. *For every* $S \subseteq T$, $\bigcup S \in T$.
3. *For every* **finite** $S \subseteq T$, $\bigcap S \in T$.

*The sets in* $T$ *are called the* open *sets of* $X$.

For example, setting $X = \mathbb{R}^2$ and $T$ to be the collection of open subsets (in the geometric/calculus sense) of $\mathbb{R}^2$, we can check that $(X, T)$ is a topological space. A further example of a topological space is $(X, 2^X)$, where $2^X$ denotes the family of all subsets of X. This is called a *discrete topology.*

You might wonder, why do we consider infinite unions of open sets to be open, but restrict to finite intersections in Condition 3. This is such that the open sets of Euclidean space can actually be called open in the language of topology. If we allowed infinite intersections in Condition 3, a set $\{p\}$ consisting a single point $p \in \mathbb{R}^2$ would have to be considered to be open: it is the intersection of the infinite series of open balls of radius $1/n$ centered at $p$, for $n \in \mathbb{N}$.

In most applications in these lecture notes, we work with subspaces of the Euclidean space $\mathbb{R}^d$, so apart from open sets, we also know from calculus notions such as *closed sets, closure, interior* and *boundary.* These terms can be defined also for abstract topological spaces:

**Definition 1.2.** *A set* $Q \subseteq X$ *is called* closed*, if its complement* $X \setminus Q$ *is open. The* closure $\operatorname{cl} Q$ *is the smallest closed set containing* $Q$*. The* interior $\operatorname{int} Q$ *is the union of all open subsets of* $Q$*. The* boundary $\operatorname{bnd} Q$ *is the set minus its interior:* $\operatorname{bnd} Q = Q \setminus \operatorname{int} Q$.

Note that sets can be open and closed simultaneously: in every topological space $(X, T)$, $\emptyset$ and $X$ are such examples. In a discrete topology, every subset $S \subseteq X$ is open and closed.

**Exercise 1.3.** *Show that a finite union of closed sets is closed.*

So far we have only seen two topological spaces: Euclidean space, or any set with the (rather boring) discrete topology. In order to see the value in the abstractions we are doing, we would like to have more examples of topological spaces. In particular, it would be great if we had a way to get new topological spaces from known ones. In the following we discuss some ways to this, starting with taking intersections.

**Lemma 1.4.** *Let* $(X, T)$ *be some topological space, and* $Y \subseteq X$*. Then,* $U := \{A \cap Y \mid A \in T\}$ *is a topology on* $Y$*. We call this a* subspace topology*.*

*Proof.* We check the three conditions of a topology:

1. $\emptyset = \emptyset \cap Y$, therefore $\emptyset \in U$. Similarly, $Y = X \cap Y$, and thus $Y \in U$.
2. $\bigcup_{i \in I}(A_i \cap Y) = (\bigcup_{i \in I} A_i) \cap Y$, and thus $\bigcup_{i \in I}(A_i \cap Y) \in U$.
3. $\bigcap_{i=1}^{n}(A_i \cap Y) = (\bigcap_{i=1}^{n} A_i) \cap Y$, and thus $\bigcap_{i=1}^{n}(A_i \cap Y) \in U$.

□

Since we have seen that $\mathbb{R}^d$ is a topological space, this already tells us that all subsets of $\mathbb{R}^d$ are topological spaces.

Another way to get topological spaces is as a product of spaces. We will not discuss the details of this here, and refer the interested reader to any textbook on topology, such as the excellent book by Munkres [1].

**Fact 1.5.** *Let* $X, Y$ *be two topological spaces. Then,* $X \times Y$ *is a topological space, with the so-called* product topology*.*

The definition of topological spaces allows us to formally define concepts from geometry in a more abstract setting:

**Definition 1.6.** *A topological space* $(X, T)$ *is* disconnected*, if there are two disjoint non-empty open sets* $U, V \in T$*, such that* $X = U \cup V$*. A topological space is* connected*, if it is not disconnected.*

**Exercise 1.7.** *In this exercise, we will use topology to prove that the set of primes is infinite.*

*We define the sets* $S(a, b)$ *as follows:*

$$S(a, b) := \{an + b \mid n \in \mathbb{Z}\}, \ \forall a \in \mathbb{Z} \setminus \{0\}, b \in \mathbb{Z}$$

*We then say that a set* $U \subseteq \mathbb{Z}$ *is open, if and only if for all* $x \in U$*, there exists* $a \in \mathbb{Z}$ *such that* $S(a, x) \subseteq U$*. This is equivalent to saying that every open set* $U$ *is a union of zero or more (including infinitely many) sets* $S(a, b)$*.*

*(a) Show that this defines a topology on* $\mathbb{Z}$*.*

*(b) Let* $A \subset \mathbb{Z}$ *be finite and non-empty. Show that* $\mathbb{Z} \setminus A$ *cannot be closed.*

*(c) Show that* $S(a, b)$ *is both open and closed.*

*(d) Show that*

$$\bigcup_{p \text{ prime}} S(p, 0) = \mathbb{Z} \setminus \{-1, 1\}$$

*(e) Conclude that there are infinitely many primes.*

## 1.2 Metric Spaces

Recall that topological spaces should capture neighborhoods of points without using the notion of a distance. However, if we do have distances, we should still be able to use the framework of topological spaces, that is, topological spaces should be a generalization of spaces with distances.

**Definition 1.8.** *A* metric space $(X, d)$ *is a set* $X$ *of points and a distance function* $d : X \times X \to \mathbb{R}$ *satisfying*

1. $d(p, q) = 0$ *if and only if* $p = q$.
2. $d(p, q) = d(q, p)$, $\forall p, q \in X$. *(Symmetry)*
3. $d(p, q) \leqslant d(p, s) + d(s, q)$, $\forall p, q, s \in X$. *(Triangle inequality)*

Note that these three conditions imply that $d(p, q) \geqslant 0$ for all $p, q \in X$: If some distance $d(p, q)$ would be negative, we would have $0 = d(p, p) \leqslant d(p, q) + d(q, p) = 2 \cdot d(p, q) < 0$, a contradiction.

**Fact 1.9.** *Every metric space has a topology (the* metric space topology*) given by the open metric balls* $B(c, r) = \{p \in X \mid d(p, c) < r\}$ *and their unions.*

## 1.3 Maps between topological spaces

In most areas of mathematics, there are two things that are at the core of every theory: the studied objects, and maps between them. For example, in linear algebra we study vector spaces and the linear maps between then. Now that we have defined the objects of study, which are topological spaces, we want to look at the maps between them.

**Definition 1.10.** *A function* $f : X \to Y$ *is* continuous *if for every open set* $U \subseteq Y$*, its pre-image* $f^{-1}(U) \subseteq X$ *(the set of all elements* $x \in X$ *such that* $f(x) \in U$*) is open. Continuous functions are also called* maps*. If* $f$ *is injective, it is called an* embedding*.*

Let us give some examples:

- For $X \subseteq Y$, we write $X \hookrightarrow Y$ for the function $f(x) = x, \forall x \in X$. This function, which is also called the *inclusion map*, is continuous: $f^{-1}(U) = U \cap X$, which is open in the subspace topology on $X$.
- For a function $f : \mathbb{R} \to \mathbb{R}$, continuity agrees with the "$\epsilon$-$\delta$" definition of continuity from calculus.

**Exercise 1.11.** *A topological space* $(X, T)$ *is called* path-connected *if any two points* $x, y \in X$ *can be joined by a path, i.e., there exists a map* $f : [0, 1] \to X$ *of the segment* $[0, 1] \subset \mathbb{R}$ *onto* $X$ *such that* $f(0) = x$ *and* $f(1) = y$*. Prove that a path-connected space is connected.*

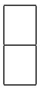
We can also use continuous functions to define when we consider two topological spaces to be "the same":

**Definition 1.12.** *A* homeomorphism *is a bijective map* $f : X \to Y$ *whose inverse is also continuous. Two topological spaces are* homeomorphic*, if there is a homeomorphism between them. We also write* $X \simeq Y$ *to say that* $X, Y$ *are homeomorphic.*

Let us again look at some examples:

- The boundary of a tetrahedron is homeomorphic to the sphere $S^2$. Idea: Take a point $c$ within the tetrahedron, and send each point $p$ to the point $f(p)$ on the ray from $c$ through $p$ such that $d(c, f(p)) = 1$.
- $I := (-1, 1)$ is homeomorphic to $\mathbb{R}$. The following map $f$ is a homeomorphism: $f : I \to \mathbb{R}$, $x \mapsto \frac{x}{1-|x|}$. Its inverse is $f^{-1} : \mathbb{R} \to I$, $y \mapsto \frac{y}{1+|y|}$.
- All knots (embeddings of the circle into $\mathbb{R}^3$) are homeomorphic. Thus, we cannot distinguish between knots using only homeomorphism.

**Exercise 1.13.** *Give an example of a map* $f : X \to Y$ *that is bijective but not a homeomorphism.*

**Exercise 1.14.** *Consider a grid of* $2$ *vertical line segments and* $k+2$ *horizontal segments, for some* $k \geqslant 0$. *For* $k = 1$, *this looks as follows:* 
*Now, we consider the problem of placing a point on each of the* $k+2$ *horizontal line segments, such that each of the* $k+4$ *total line segments contains at least one point.*

*(a) How could one define a topology on the set of all such point placements?*

*(b) Prove that this space is homeomorphic to* $S^k$.

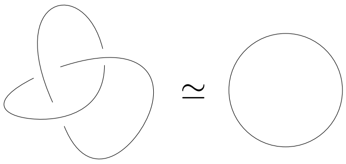


**Figure 1.1:** *Two knots.*

The example of the knots shows that in certain cases, maps and homeomorphism are not a good language to capture the relevant properties. In some cases, we want to look at an entire process of continuously deforming one object into another.

**Definition 1.15.** *An* isotopy *connecting* $X \subseteq \mathbb{R}^d$ *and* $Y \subseteq \mathbb{R}^d$ *is a continuous map* $\phi : X \times [0,1] \to \mathbb{R}^d$, *such that* $\phi(X,0) = X$, $\phi(X,1) = Y$, *and* $\forall t \in [0,1]$, $\phi(\cdot,t)$ *is a homeomorphism between* $X$ *and its image. Two spaces are called* isotopic, *if there is an isotopy connecting them.*

Some examples:

- Let $X \subset \mathbb{R}$ be the union of $0$, and $[1,2]$, and let $Y \subset \mathbb{R}$ be the union of $[0,1]$ and $2$. These spaces are homeomorphic ($X \simeq Y$), but not isotopic.
- The two knots from Figure 1.1 above are also not isotopic.
- Consider the two spaces in Figure 1.2. Do you think they are isotopic? Most people would probably argue that they are not, as in one of them the "handcuff" wraps around the "pole" once and in the other one twice. However, it turns out that the spaces are in fact isotopic. An isotopy is illustrated by the following video: https://www.youtube.com/watch?v=wDZx9B4TAXo

We can also study continuous deformations between maps:

**Definition 1.16.** *Let* $g, h$ *be maps* $X \to Y$. *A* homotopy *connecting* $g$ *and* $h$ *is a map* $H : X \times [0,1] \to Y$ *such that* $H(\cdot,0) = g$ *and* $H(\cdot,1) = h$. *In this case* $g$ *and* $h$ *are called* homotopic.

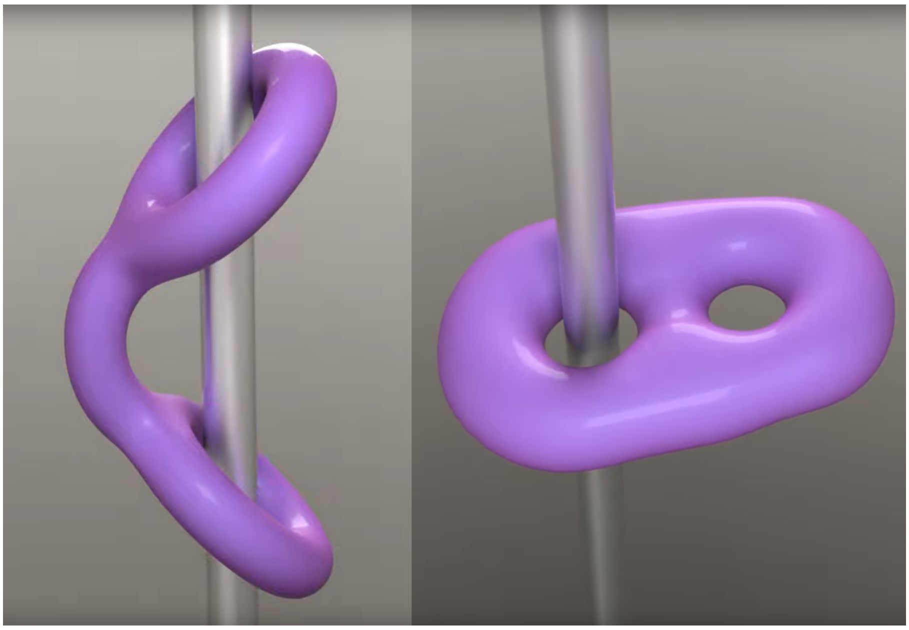


**Figure 1.2:** *Left: Both handcuffs are connected to an infinite pole. Right: Only one loop of the handcuffs is connected to the infinite pole. These spaces are isotopic.*

Some examples:

- The inclusion map $g : \mathbb{B}^3 \hookrightarrow \mathbb{R}^3$ (where $\mathbb{B}^3$ is the unit ball in $\mathbb{R}^3$), and $h : \mathbb{B}^3 \to \mathbb{R}^3$ which sends every point to the origin, are homotopic, as shown by the homotopy

$$H(x, t) = (1 - t)g(x).$$

- The identity function $g : S^1 \to S^1$ and $h : S^1 \to S^1$ which sends everything to a single point $p \in S^1$ are *not* homotopic.

The notion of homotopy allows us to define an equivalence on topological spaces that is weaker than homeomorphism. Intuitively, this notion says that two spaces are "the same" if they can be continuously transformed into each other not only by bending, twisting and stretching, but also by shrinking or blowing up parts of different dimensions.

**Definition 1.17.** *Two spaces* $X, Y$ *are* homotopy equivalent *if there exist maps* $g : X \to Y$ *and* $h : Y \to X$ *such that:*

- $h \circ g$ *is homotopic to* $\mathrm{id}_X$ *(the identity map* $x \mapsto x$*), and*
- $g \circ h$ *is homotopic to* $\mathrm{id}_Y$*.*

For example, the circle $S^1$ and $\mathbb{R}^2 \setminus \{0\}$ are homotopy equivalent. We pick $g$ as the inclusion map $S^1 \hookrightarrow \mathbb{R}^2 \setminus \{0\}$, and $h(x) := \frac{x}{|x|}$. We see that $h \circ g(x) = x$, i.e., $h \circ g = \mathrm{id}_{S^1}$.

Furthermore, $g \circ h(x) = h(x)$. Finally, $g \circ h$ and $\mathrm{id}_{\mathbb{R}^2 \setminus \{0\}}$ are homotopic as certified by the homotopy $H(x, t) := tx + (1 - t)h(x)$.

An important example of a homotopy equivalence are *deformation retracts*:

**Definition 1.18.** *Let* $A \subseteq X$. *A* deformation retract *of* $X$ *onto* $A$ *is a map* $R : X \times [0, 1] \to X$, *such that*

- $R(\cdot, 0) = \mathrm{id}_X$
- $R(x, 1) \in A, \forall x \in X$
- $R(a, t) = a, \forall a \in A, t \in [0, 1]$

*If such a deformation retract of* $X$ *onto* $A$ *exists, we also say that* $A$ is *a deformation retract of* $X$.

The intuition behind a deformation retract is to continuously shrink $X$ to $A$, while leaving $A$ fixed.

**Fact 1.19.** *If* $A$ *is a deformation retract of* $X$ *(there exists a deformation retract of* $X$ *onto* $A$*), then* $A$ *and* $X$ *are homotopy equivalent.*

Some examples:

- The circle $S^1$ is a deformation retract of $\mathbb{R}^2 \setminus \{0\}$: $R(x, t) = (1 - t)x + t \cdot \frac{x}{|x|}$.
- A punctured torus can be deformation retracted onto the symbol 8 where one of the two circles is rotated by 90°, as seen by the following video: https://www.youtube.com/watch?v=tz3QWrfPQj4

**Lemma 1.20.** *If* $X$ *and* $Y$ *are homeomorphic, they are also homotopy equivalent.*

*Proof.* Let $g : X \to Y$ be the homeomorphism, and $h := g^{-1}$ its inverse. Then $g \circ h = \mathrm{id}_Y$ and $h \circ g = \mathrm{id}_X$, and $\mathrm{id}$ is homotopic to itself. □

The following is a nice way to show that two spaces are homotopy equivalent:

**Fact 1.21.** $X, Y$ *are homotopy equivalent if and only if there exists a space* $Z$ *such that* $X$ *and* $Y$ *are deformation retracts of* $Z$.

An example of this fact can be found in Figure 1.3.

**Exercise 1.22.** *Sort the letters of the alphabet into equivalence classes under homotopy equivalence.*

**Exercise 1.23.** *Show that both a cylinder and a Möbius strip are homotopy equivalent to a circle.*

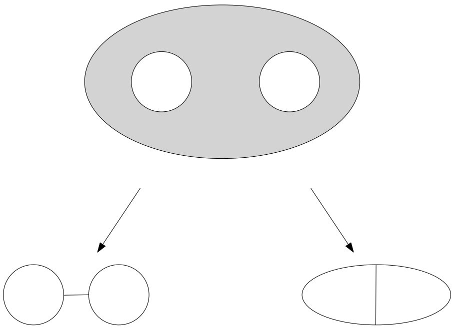
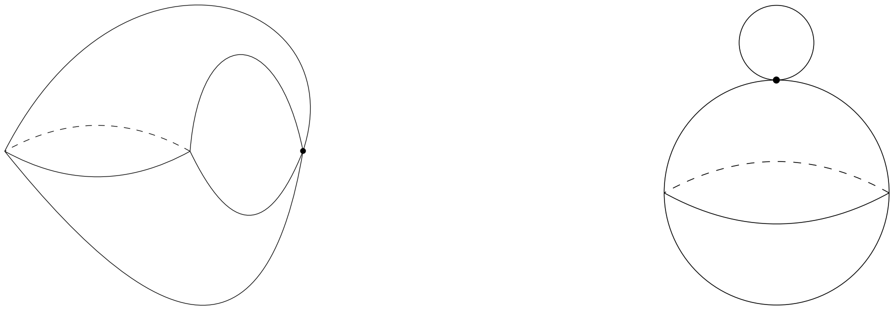


**Figure 1.3:** *The top space deformation retracts to both spaces below, showing that they are homotopy equivalent.*



**(a)** *The space* X. **(b)** *The space* Y.

**Figure 1.4:** *The spaces from Exercise 1.24.*

**Exercise 1.24.** *Let* X *be* $S^2$ *where the north pole and the south pole have been glued together, see Figure 1.4a. Let* Y *be* $S^2$ *with an* $S^1$ *attached at the north pole, see Figure 1.4b.*

*Give an informal argument that* X *and* Y *are homotopy equivalent. Bonus question: Are they also homeomorphic?*

We note that in general showing existence of a map with certain properties (e.g., a homeomorphism, isotopy, homotopy) is easy: just give a map and show that it satisfies the required properties. On the other hand, showing that such a map cannot exist is hard, as there are usually infinitely many candidate maps. The idea of algebraic topology is to construct invariants preserved by these maps. Then, we know that no map can exist between spaces on which these invariants differ. An example of such an invariant is the number of "holes" a space has, which we will formalize when we introduce the notion of *homology*.

## 1.4 Algebra

In this section we need the necessary background in algebra that is needed for the basics of homology theory. Just as for topology, we first introduce the objects of study, followed by the maps between them.

**Definition 1.25.** *A group* $(G, +)$ *is a set* $G$ *together with a binary operation "+" such that*

1. $\forall a, b \in G$*:* $a + b \in G$
2. $\forall a, b, c \in G$*:* $(a + b) + c = a + (b + c)$ *(Associativity)*
3. $\exists 0 \in G$*:* $a + 0 = 0 + a = a \; \forall a \in G$
4. $\forall a \in G \exists - a \in G$*:* $a + (-a) = 0$

$(G, +)$ *is* abelian *if we also have*

5. $\forall a, b \in G$*:* $a + b = b + a$ *(Commutativity)*

Examples:

- $(\mathbb{Z}, +)$ is a group (even an abelian one), but not $(\mathbb{N}, +)$.
- The moves of a Rubik's cube also form a group (with the operation being concatenation), but not an abelian one: let $L$ denote moving the left face clockwise, and let $U$ denote moving the upper face clockwise. Replacing "clockwise" by counterclockwise we get $-L$ and $-U$, respectively. Now, if the group was abelian, then $L + U - L - U$ should give the same configuration again, but if you do these moves on a Rubik's cube, you will see that the configuration has changed.

As groups can be very large, even infinite, it can be useful to have a concise way of writing them:

**Definition 1.26.** *Let* $(G, +)$ *be a group.*
*A subset* $A \subseteq G$ *is a* generator *if every element of* $G$ *can be written as a finite sum of elements of* $A$ *and their inverses.*
*A subset* $B \subseteq G$ *is a* basis *if every element of* $G$ *can be* uniquely *written as a finite sum of elements of* $B$ *and their inverses (ignoring trivial cancellations, i.e.,* $a + c + (-c) + (-b) = a + (-b)$*).*
*An abelian group that has a basis is called* free*.*

Examples:

- The six standard moves of the Rubik's cube (rotating the top, bottom, front, back, left, or right layer clockwise by 90°) are a generator for the Rubik's cube moves.
- $\{1\}$ is a basis of $(\mathbb{Z}, +)$.

**Exercise 1.27.** *A cyclic group is a group* $G$ *that contains an element* $g \in G$ *such that* $\{g\}$ *is a generator of* $G$*. Show that every cyclic group is abelian (commutative).*

**Exercise 1.28.** *Consider a Rubik's cube. Prove that no move (sequence of elementary moves)* $X$ *exists such that every Rubik's cube can be solved by repeatedly applying* $X$.

**Definition 1.29.** *For some group* $(G, +)$, $H \subseteq G$ *is a* subgroup, *if* $(H, +)$ *is also a group.*

For example, the even integers (including $0$) are a subgroup of $(\mathbb{Z}, +)$. Subgroups are important in group theory, as they can be used to partition a group into several parts:

**Definition 1.30.** *Let* $H \subseteq G$ *be a subgroup of* $(G, +)$, *and* $a \in G$.
*The* left coset $a + H$ *is the set* $a + H := \{a + b \mid b \in H\}$, *and the* right coset $H + a := \{b + a \mid b \in H\}$. *If* $G$ *is abelian,* $a + H = H + a$, *and they are simply called the* coset. *For* $G$ *abelian, the* quotient group of $G$ by $H$, *denoted by* $G/H$, *is the group on cosets* $\{a + H, a \in G\}$ *with the operation* $\oplus$ *defined as* $(a + H) \oplus (b + H) = (a + b) + H$, $\forall a, b \in G$.

Examples:

- Let $G = (\mathbb{Z}, +)$ and $H = n\mathbb{Z} = \{n \cdot a \mid a \in \mathbb{Z}\}$. Then, $G/H = \{0 + \mathbb{Z}, 1 + \mathbb{Z}, \ldots, (n-1) + \mathbb{Z}\}$ is the group usually referred to as $\mathbb{Z}_n$, the group of modular arithmetic modulo $n$.
- $\mathbb{R}/\mathbb{Z}$ is the circle group (the multiplicative group of all complex numbers of absolute value $1$). (You should try and convince yourself, why).

In order to compare groups with each other, we again want a notion of maps between groups, that behave well with the group structures:

**Definition 1.31.** *A map* $h : G \to H$ *between* $(G, +)$ *and* $(H, \star)$ *is a* homomorphism *if* $h(a + b) = h(a) \star h(b)$, $\forall a, b \in G$.
*A bijective homomorphism is called an* isomorphism, *and then we write* $G \cong H$ *and say that* $G$ *and* $H$ *are* isomorphic.

*kernel* $\ker h := \{a \in G \mid h(a) = 0\}$

*image* $\operatorname{im} h := \{b \in H \mid \exists a \in G$ *with* $h(a) = b\}$

*cokernel* $\operatorname{coker} h := H/\operatorname{im} h$

Note that we are assuming something in our definition of the cokernel: for the definition of a quotient group to apply, we need the divisor group to be a subgroup of the dividend group. Luckily, the following lemma says that $\operatorname{im} h$ is always a subgroup of $H$.

**Lemma 1.32.** $\ker h$ *and* $\operatorname{im} h$ *are subgroups of* $(G, +)$ *and* $(H, \star)$, *respectively.*

*Proof.* We first prove this for $\ker h$.

1. $a, b \in \ker h \Rightarrow h(a) = h(b) = 0$. By definition of homomorphism, $h(a+b) = h(a) \star h(b) = 0 \star 0 = 0$, and thus by definition of $\ker h$, $a + b \in \ker h$. We conclude that $\ker h$ is closed under $+$.
2. Associativity follows from associativity of $+$ in $G$, since $\ker h \subseteq G$.
3. $\forall a \in G :\ h(0) \star h(a) = h(0 + a) = h(a)$, and thus $h(0) = 0$, from which $0 \in \ker h$ follows.
4. Let $a \in \ker h$. Then, $0 = h(0) = h(a - a) = h(a) \star h(-a) = 0 \star h(-a) = h(-a)$, and thus $-a \in \ker h$.

The proof for $\operatorname{im} h$ is left as an exercise. □

**Exercise 1.33.** *Show that* $\operatorname{im} h$ *is a subgroup of* $H$.

**Exercise 1.34.** *For two Abelian groups* $(G, \star)$ *and* $(H, +)$, *let the set of all homomorphisms* $f : G \to H$ *be denoted by* $\operatorname{Hom}(G, H)$.

*(a) Show that for any groups* $G, H$, $(\operatorname{Hom}(G, H), \oplus)$, *where the operation* $+$ *is defined as*

$$(f \oplus g)(x) = f(x) + g(x), \forall x \in G,$$

*is also a group.*

*(b) Show* $\operatorname{Hom}(\mathbb{Z}_2^2, \mathbb{Z}_2) \cong \mathbb{Z}_2^2$, *i.e., the groups are isomorphic.*

As the example of the integers shows, a big motivation for the study of groups comes from number theory. However, in number theory we do not only have addition but also multiplication. This motivates the following definition:

**Definition 1.35.** $(R, +, \cdot)$ *is a* ring, *if*

1. $(R, +)$ *is an abelian group.*
2. $\forall a, b, c \in R$*:*
   $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ *and* *(Associativity of* $\cdot$*)*
   $a \cdot (b + c) = a \cdot b + a \cdot c,$
   $(b + c) \cdot a = b \cdot a + c \cdot a$ *(Distributivity)*
3. $\exists 1 \in R$, *such that* $a \cdot 1 = 1 \cdot a = a \ \forall a \in R$. *(Multiplicative identity)*

*If* $\cdot$ *is commutative, we say that* $R$ *is* commutative.

**Definition 1.36.** *A commutative ring in which every non-zero element has a multiplicative inverse (*$\forall a \in R \setminus \{0\}, \exists b \in R : a \cdot b = 1$*) is called a* field.

Another important area of algebra, which you already know, is linear algebra. Here, vectors can be added and subtracted. Further the field of real numbers are called *scalars* and they can be multiplied with vectors. So, we have very similar operations at hand. This motivates the following generalization of the concept of vector spaces.

**Definition 1.37.** *Given a ring* $(R, +, \cdot)$ *with multiplicative identity* $1$*, an* R-module $M$ *is an abelian group* $(M, \oplus)$ *with an operation* $\otimes : R \times M \to M$ *such that for all* $r, r' \in R$ *and* $x, y \in M$*, we have*

1. $r \otimes (x + y) = (r \otimes x) \oplus (r \otimes y)$
2. $(r + r') \otimes x = (r \otimes x) \oplus (r' \otimes x)$
3. $1 \otimes x = x$
4. $(r \cdot r') \otimes x = r \otimes (r' \otimes x)$

*If* $R$ *is a field, the* $R$*-module is called a* vector space*.*

In the literature, often the same symbol $(\cdot)$ is used for both operations $\cdot$ and $\otimes$, and $+$ for both $+$ in $R$ and $\oplus$ in $M$. For a vector space, this should feel quite normal, since for the vector space $\mathbb{R}^n$ (which is an $\mathbb{R}$-module), we also write $\cdot$ for multiplying scalars to both scalars and vectors, and $+$ for addition of both scalars and vectors.

Modules appear all over the place in homology theory. In some cases, in particular in all the ones we discuss in these lecture notes, the modules happen to be vector spaces. Thus, most of what we discuss in the following chapters could be phrased using only language from linear algebra. However, to be consisting with most of the existing literature, we will phrase most results slightly more general.

## Questions

1. *What is a topological space?* Give the formal definition and some examples.
2. *What is a continuous map between topological spaces? What is a homeomorphism?* State the definitions and give examples.
3. *What is a homotopy? What is a homotopy equivalence?* Give the formal definitions. Further, define deformation retracts and use them to give an alternative definition of homotopy equivalence.
4. *What are groups and the maps between them?* State the definitions and prove that the image and kernel are subgroups.

## References

[1] J.R. Munkres, *Topology*, Prentice Hall, Incorporated, 2000.

# Chapter 2

# Homology

In this chapter, we introduce *homology*, a fundamental concept in algebraic topology and, as the name suggests, a crucial element of the *persistent homology pipeline* in topological data analysis. Very informally, homology can be used to count the number of "holes" of a topological space, where holes can have any dimension. While you might have an intuition of what a 2-dimensional hole in a subspace of $\mathbb{R}^2$ might be, it is not at all clear what a 4-dimensional hole in some 7-dimensional space should be. The main idea of homology is to use algebra to talk about holes in an abstract setting.

## 2.1 Simplicial Complexes

In order to define homology, we restrict ourselves to special types of topological spaces, namely *simplicial complexes*. We will see that this covers most natural spaces. Furthermore, homology for simplicial complexes is sufficient for all classical applications in topological data analysis. We will briefly outline a more general definition later in the chapter.

**Definition 2.1.** *A* $k$*-simplex in* $\mathbb{R}^d$ *is the convex hull of* $k+1$ *affinely independent points in* $\mathbb{R}^d$*.*

A *face* of a simplex is the convex hull of a subset of its vertices. In particular, every face of a simplex is also a simplex. The empty set $\emptyset$ is also a face. The $(k-1)$-faces are called *facets*.

**Definition 2.2.** *A* geometric simplicial complex *is a family* $K$ *of simplices such that*

- *if* $\tau \in K$ *and* $\sigma$ *is a face of* $\tau$*, then* $\sigma \in K$*, and*
- *for* $\sigma, \tau \in K$*, their intersection* $\sigma \cap \tau$ *is a face of both.*

We say the *dimension* of a simplicial complex is the maximum dimension of any simplex, and the dimension of a $k$-simplex is $k$. In these lecture notes, and for applications

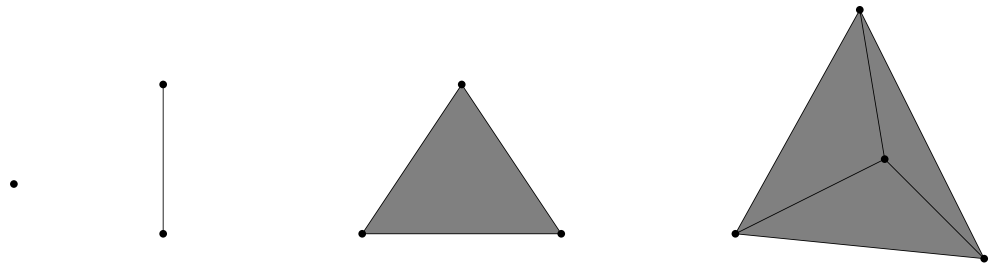


**Figure 2.1:** *Some examples of simplices: a point (0-dimensional), a line segment (1-dimensional), a triangle (2-dimensional) and a (filled) tetrahedron (3-dimensional).*

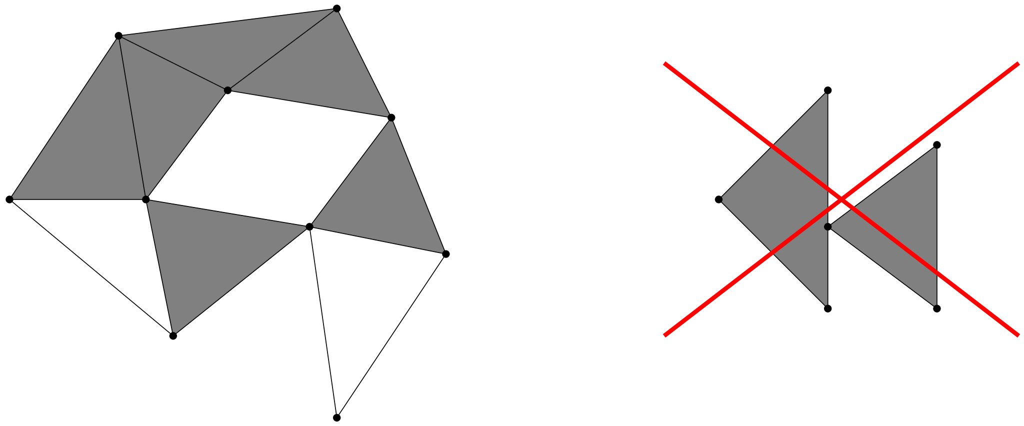


**Figure 2.2:** *The left is a simplicial complex. The right is not, as the intersection of the two triangles is not a face of both of them.*

in topological data analysis, we may assume that all simplicial complexes are *finite*, that is, consisting of finitely many simplices.

The way we defined them, simplicial complexes are geometric objects. However, we can also study them in a purely combinatorial setting.

**Definition 2.3.** *An* abstract simplicial complex $K$ *is a family of subsets of a vertex set* $V(K)$ *such that if* $\tau \in K$ *and* $\sigma \subseteq \tau$*, then* $\sigma \in K$*.*

A $k$-simplex here is a subset of $k+1$ elements, and thus again called $k$-dimensional. Note that 1-dimensional abstract simplicial complexes are exactly graphs: they are defined by a vertex set $V$ and a system of two-element subsets of $V$, called *edges*.

From every geometric simplicial complex we get an abstract simplicial complex by simply taking the set of points as the vertex set and adding the correct subset for every simplex. For the inverse direction, we have to talk about geometric realizations:

**Definition 2.4.** *A geometric simplicial complex* $K$ *is a geometric realization of some abstract simplicial complex* $K'$*, if there is an embedding* $e : V(K') \to \mathbb{R}^d$ *that takes* *every (abstract)* $k$*-simplex* $\{v_0, \dots, v_k\}$ *in* $K'$ *to the (geometric)* $k$*-simplex that is the convex hull of* $e(v_0), \dots, e(v_k)$*.*

Does every abstract simplicial complex have a geometric realization? For 1-dimensional complexes (graphs), we know that not all graphs admit a straight-line embedding in the plane, as only planar graphs admit any embedding, i.e., crossing-free drawing, in the plane. However, by placing the vertices in $\mathbb{R}^3$ in such a way that no four vertices lie on a common plane, we see that we can always find a geometric realization of a graph in $\mathbb{R}^3$. This generalizes to the following realization theorem:

**Theorem 2.5.** *Every* $k$*-dimensional simplicial complex has a geometric realization in* $\mathbb{R}^{2k+1}$*.*

*Proof.* Place the vertices as distinct points on the *moment curve* in $\mathbb{R}^{2k+1}$, which is the curve given by $f(t) = (t, t^2, \dots, t^{2k+1})$. This way, any $2k+2$ of the placed points are affinely independent. Thus, any two faces with disjoint vertex sets will not intersect in the realization, showing that the realization is indeed an embedding. $\square$

Since we now know that abstract and geometric simplicial complexes can be translated into one another, we will not make the distinction between them again and just use the word *simplicial complex* for both objects in the following. As a subset of Euclidean space, a simplicial complex thus also inherits the subspace topology from $\mathbb{R}^d$, which allows us to view simplicial complexes as topological spaces.

On the other hand, most topological spaces are not simplicial complexes by definition. For example, the 2-sphere $S^2$ is not a simplicial complex, as it is not defined by a vertex set and faces. However, the boundary of a tetrahedron is a simplicial complex, and it is homeomorphic to $S^2$, so if we want to work with $S^2$, from a topologist's point of view, we might as well work with the boundary of a tetrahedron instead. This motivates the following definition.

**Definition 2.6.** *A simplicial complex* $K$ *is a* triangulation *of a topological space* $X$*, if* $|K|$ *is homeomorphic to* $X$*.*

We say that a topological space $X$ is *triangulable* if it has a triangulation. Triangulable spaces are nice for as, as we can replace them by simplicial complexes without any loss of topological information. Unfortunately, not all topological spaces are triangulable, but in this course we will not deal with such spaces. Also note that a triangulable space has infinitely many triangulations, for example by subdividing simplices.

While triangulations give us simplicial complexes from topological spaces, we can also go the other way: many combinatorial structures naturally give rise to (abstract) simplicial complexes, which can in turn be interpreted as topological spaces. Thus, we can use the machinery of topological also for many combinatorial problems. This gives rise to a subfield of combinatorics called *topological combinatorics*, where the topology of simplicial complexes associated to combinatorial objects is studied. Let us give some examples of such simplicial complexes.

- As we have already discussed, a graph is exactly a 1-dimensional simplicial complex.
- Given a graph $G = (V, E)$, define a simplicial complex on $V$ by including a face $\{v_1, \ldots, v_k\}$ whenever these vertices form a clique in $G$. This is called the *clique complex* of $G$.
- For a poset $(P, \leqslant)$, the set of all chains of $P$ forms a simplicial complex, giving rise to the *order topology*.

Another example of high relevance for topological data analysis is the *nerve*:

**Definition 2.7.** *For a finite collection* $U$ *of sets, its* nerve $N(U)$ *is a simplicial complex on the vertex set* $U$ *that contains* $U_0, \ldots, U_k$ *as a* $k$*-simplex if and only if* $U_0 \cap \ldots \cap U_k \neq \emptyset$*.*

In many applications, the considered sets are subsets of some topological space. In this case, we often want the intersections to be "well-behaved".

**Definition 2.8.** *Let* $X$ *be a metric space, and* $U$ *a finite family of closed subsets of* $X$*. We call* $U$ *a* good cover*, if every non-empty intersection of sets in* $U$ *is contractible (i.e., homotopy equivalent to a point).*

Under these conditions on the sets, we get the following, very powerful theorem, which allows us to relate complicated spaces (unions of sets) with a much simpler simplicial complex, namely the nerve. For a proof of this we refer to any textbook on algebraic topology, for example the one by Hatcher [2].

**Theorem 2.9** (Nerve theorem). *If* $U$ *is a good cover, then* $|N(U)|$ *is homotopy equivalent to* $\bigcup U$*.*

The nerve theorem also holds if all the sets in $U$ are open with contractible intersections, but it may fail if some sets in $U$ are closed, and some open: We can have an open and a closed set which do not intersect, but whose union is connected.

Now that we have defined simplicial complexes, once again we want to study maps between them. The study of simplicial complexes and the maps between them, as we will define them, is called *combinatorial topology*.

**Definition 2.10.** *A map* $f : K_1 \to K_2$ *(which maps vertices of* $K_1$ *to vertices of* $K_2$*, also called a* vertex map*) is called* simplicial *if for every simplex* $\{v_0, \ldots, v_k\} \in K_1$*, we have that* $\{f(v_0), \ldots, f(v_k)\}$ *is a simplex in* $K_2$*.*

Recall that simplicial complexes are topological spaces, so there is also the notion of continuous maps between them. It can be shown that every simplicial map is continuous.

**Exercise 2.11.** *Let* $f : K_1 \to K_2$ *be a simplicial map. Show that* $f$ *is continuous.*

On the other hand, continuous maps are in general not even vertex maps and thus not simplicial. Thus, simplicial maps are more restrictive than continuous maps. However, the difference of the two concepts is smaller than one might think at first glance.

**Fact 2.12.** *Every continuous map* $f : |K_1| \to |K_2|$ *can be approximated arbitrarily closely by simplicial maps on appropriate subdivisions of* $K_1$ *and* $K_2$*.*

This shows that we can consider simplicial maps to be the analogue of continuous maps in the world of simplicial complexes. This begs the question whether other definitions from topology, such as homotopies or deformation retracts, have simplicial analogues. As we will see in the next few definitions, they do.

**Definition 2.13.** *Two simplicial maps* $f_1, f_2 : K_1 \to K_2$ *are* contiguous *if for every simplex* $\sigma \in K_1$ *we have that* $f_1(\sigma) \cup f_2(\sigma)$ *is a simplex in* $K_2$*.*

This is the simplicial analogue of two continuous maps being homotopic.

**Definition 2.14.** *A face of a simplicial complex is called* free*, if it is a non-maximal (not inclusion-maximal) and contained in a unique maximal face.*

Note that every face that is a superset of a free face is either a maximal face or also free.

**Definition 2.15.** *A* collapse *is the operation of removing all faces* $\gamma$ *that contain some fixed free face* $\tau$*. A simplicial complex is* collapsible *if there is a sequence of collapses leading to a point.*

A collapse can be written as a deformation retract. Thus, a simplicial complex that is collapsible is contractible, and we consider collapses to be the simplicial analogue of deformation retracts.

You might wonder whether every contractible simplicial complex is also collapsible. We will see that this not hold: A good counterexample for this is Bing's house with two rooms, see Figure 2.3. In any triangulation of it, there are no free faces: As a 2-dimensional space, there are only vertices, edges and triangles. We only have to check edges, since triangles are maximal, and vertices are part of edges which are never maximal. Every edge is incident to at least two triangles (there are no edges on the "boundary"), and thus they are not free. Since we have no free faces, it is not collapsible.

On the other hand, Bing's house is contractible: while there is no deformation retract from Bing's house to a point, both Bing's house and a point are deformation retracts of a 3-dimensional ball, and thus by Fact 1.21 the are homotopy equivalent. For a visual sketch of the deformation retract from a 3-dimensional ball to Bing's house, see Figure 2.4.

To summarize, the connection between simplicial complexes and topological spaces is that every simplicial complex defines a topological space, since we can consider a geometric embedding, and the underlying space of the embedding inherits the subspace topology

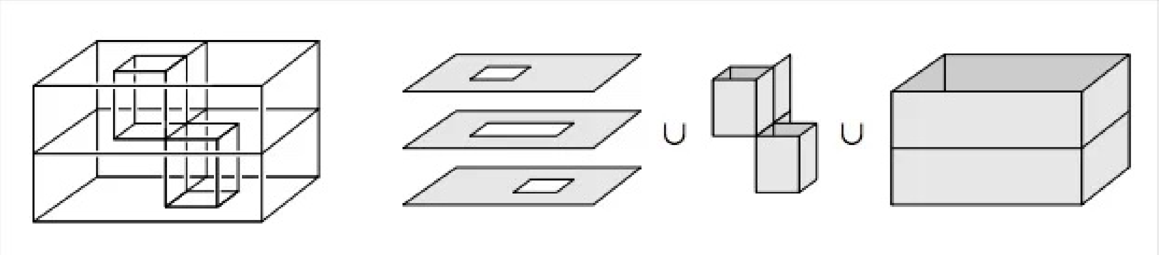


**Figure 2.3:** *Bing's house with two rooms. Image taken from [2].*

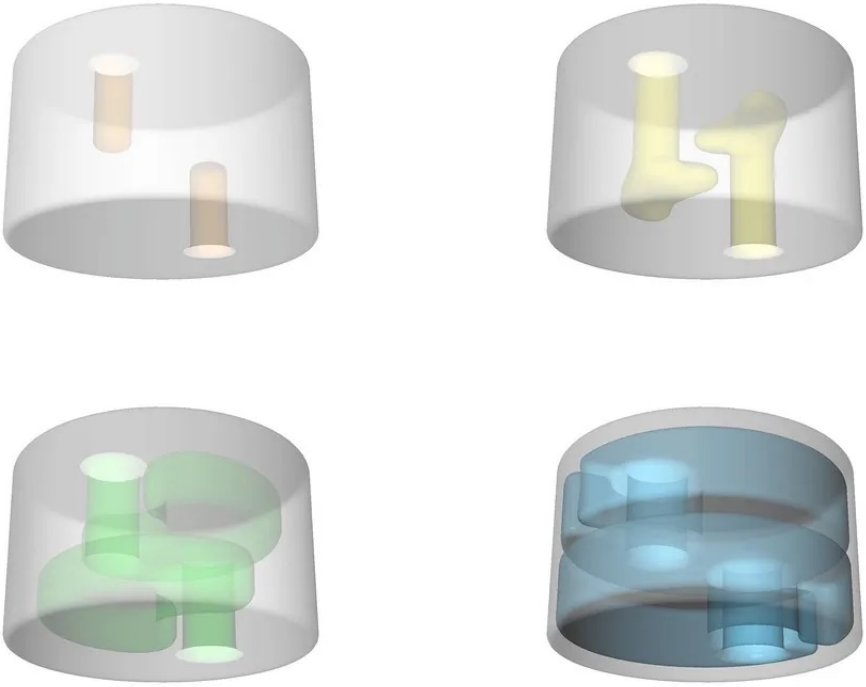


**Figure 2.4:** *A visual representation of the deformation retract from a 3-dimensional ball to Bing's house. Images taken from the blog Sketches of topology [1].*

from $\mathbb{R}^d$. On the other hand, some topological spaces (the triangulable ones) can be expressed by simplicial complexes. As for maps, every simplicial map is continuous. On the other hand, continuous maps between simplicial complexes can be approximated by simplicial maps between subdivisions of the simplicial complexes. A similar property holds between homotopic maps and contiguous maps, as well as between deformation retracts and collapses. In general, we can say that the terms in combinatorial topology are special cases of their "continuous" counterparts, and if we consider triangulable spaces, the continuous terms can be approximated in some way by their combinatorial counterparts, and can thus be considered to be equivalent.

Table 2.1 summarizes the equivalent words in “continuous topology” and in combinatorial topology on simplicial complexes.

| “continuous” topology | combinatorial topology |
|---|---|
| topological spaces | simplicial complexes |
| continuous maps | simplicial maps |
| homotopic maps | contiguous maps |
| deformation retracts | collapses |

**Table 2.1**: *Equivalent notions in “continuous” and combinatorial topology*

## 2.2 Homology

### 2.2.1 An intuitive view at holes

In this short section we will try to give a comically simplified sketch of the basic intuition of homology, in the hope that it might help understanding the next, more technical sections.

Recall that homology is intended as a tool to count holes in objects. We now know which objects we want to consider, namely simplicial complexes. Consider the following two simplicial complexes on 4 vertices $a, b, c$ and $d$. In the fist complex $K_1$, we include all four possible triangles $abc$, $abd$, $acd$ and $bcd$ as 2-faces (and thus also all six possible edges as 1-faces). In the second complex $K_2$, we again include all possible edges, but only three of the triangles, namely $abc$, $abd$ and $acd$. Thus, $K_1$ is homeomorphic to the boundary of a tetrahedron whereas $K_2$ can be viewed as a triangulation of four points in the plane, where the point $a$ lies inside the convex hull of the other three points. How many holes do these complexes have? The complex $K_1$ is homeomorphic to a 2-dimensional sphere, which is hollow, so we would like to say that it has a 2-dimensional hole, or *cavity*. On the other hand, the complex $K_2$ should not have any holes.

There is also another difference between the two complexes: the complex $K_2$ has a *boundary* consisting of the edges $bc$, $bd$ and $cd$. On the other hand, $K_1$ has no boundary, just as a sphere has no boundary. Indeed, for every *pure* simplicial complex, that is, a simplicial complex whose maximal faces all have the same dimension, we will define a notion of boundary capturing this intuition. For example, for a 1-dimensional simplicial complex, that is, a graph, the boundary will contain all the leaves (vertices of degree 1). Some complexes, like $K_1$, will have an empty boundary, and, in analogy to graphs without leaves, we call such complexes *cycles*. Under this viewpoint, our $d$-dimensional holes of a simplicial complex $K$ should be $(d-1)$-subcomplexes that are cycles, but that are not themselves boundaries of another $d$-dimensional subcomplex which would be “filling up” the hole.

In the following we will make this intuition precise by precisely defining the types of subcomplexes we consider, as well as the notions of boundaries and cycles, and how we can mathematically describe the cycles that are not boundaries.

### 2.2.2 Chains

Let K be a simplicial complex with $m_p$ p-simplices.

**Definition 2.16.** *A* p-chain c (in K) *is a formal sum*[1] *of* $p$*-simplices added with some coefficients from some ring* $R$*.*

$$c = \sum_{i=1}^{m_p} \alpha_i \sigma_i$$

*where* $\alpha_i \in R$ *and* $\sigma_i \in K$ *are* $p$*-simplices.*

Two p-chains $c = \sum \alpha_i \sigma_i$ and $c' = \sum \alpha_i' \sigma_i$ (both in K) can be added:

$$c + c' := \sum_{i=1}^{m_p} (\alpha_i + \alpha_i')\sigma_i$$

We write $C_p(K)$ for the set of all p-chains in K, called the $p$*-th chain group*. The following observation shows that this name makes sense:

**Observation 2.17.** $(C_p(K), +)$ *is an abelian group, it is free, and the* $p$*-simplices form a basis.*

*Proof.* To show that it is a group, we have

1. $\forall c_1, c_2 \in C_p(K)$, we have $c_1 + c_2 \in C_p(K)$
2. $\forall c_1, c_2, c_3 \in C_p(K)$, $(c_1 + c_2) + c_3 = \sum(\alpha_i^{(1)} + \alpha_i^{(2)})\sigma_i + \sum \alpha_i^{(3)}\sigma_i = \sum(\alpha_i^{(1)} + \alpha_i^{(2)} + \alpha_i^{(3)})\sigma_i = \sum \alpha_i^{(1)}\sigma_i + \sum(\alpha_i^{(2)} + \alpha_i^{(3)})\sigma_i = c_1 + (c_2 + c_3)$.
3. $0 = \sum 0\sigma_i \in C_p(K)$
4. $\forall c \in C_p(K)$ we have $-c = \sum(-\alpha_i\sigma_i) \in C_p(K)$ and $c + (-c) = \sum(\alpha_i - \alpha_i)\sigma_i = 0$

Commutativity follows from + being commutative, thus the group is abelian. The p-simplices clearly form a basis, since the set of chains is defined as the set of formal sums of these p-simplices. □

---

[1] A formal sum just means that we formally write a sum, but that there is no meaning behind the operation of adding the simplices.

**Observation 2.18.** *Equipped with the appropriate function* $\cdot : R \times C_p(K) \to C_p(K)$*,* $C_p(K)$ *is an* $R$*-module.*

The proof is similar and left as an exercise, but the statement should feel natural since every chain is simply described by a vector of $m_p$ elements of $R$, with addition being element-wise addition in $R$.

**Exercise 2.19.** *Prove Observation 2.18.*

From now on we will always work with the ring $R = \mathbb{Z}_2$, so in particular we have that $c + c = 0$. With this, we will define *homology over* $\mathbb{Z}_2$. Using some slightly more abstract definitions, all of the following can be extended to define homology over any ring $R$. For more on this, we refer to any textbook on algebraic topology, e.g. the one by Hatcher [2].

### 2.2.3 Boundary Maps

In this section we will make formal the notion of a boundary. It should be intuitively clear what the boundary of a simplex should be: just take the chain formed by its facets. Thanks to the group structure of the chain group, this intuition can now be algebraically extended to any chain.

More formally, let $\sigma = \{v_0, \ldots, v_p\}$ be a p-simplex. Then, $\delta_p(\sigma)$ is defined by

$$\{v_1, \ldots, v_p\} + \{v_0, v_2, \ldots, v_p\} + \ldots + \{v_0, \ldots, v_{p-1}\} = \sum_{i=0}^{p} \{v_0, \ldots, \hat{v_i}, \ldots, v_p\}$$

In the above notation, $\hat{v_i}$ denotes that the element $v_i$ is omitted from the set. Note that $\delta_p(\sigma)$ is a $(p-1)$-chain. For some examples, see Figure 2.5.

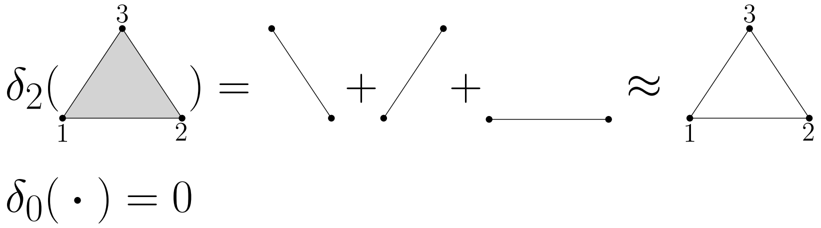
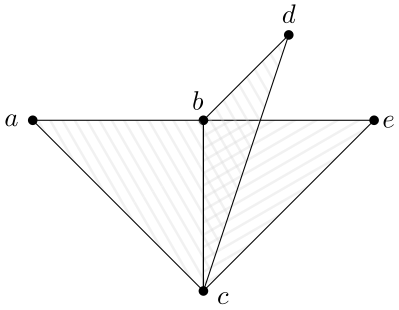


**Figure 2.5:** *The boundary chains of two different simplices.*

We have seen that $\delta_p$ is a map that sends a p-simplex to a $(p-1)$-chain. Extending this to chains, $\delta_p$ defines a *boundary operator* homomorphism:

$$\begin{aligned} \delta_p &: C_p(K) \to C_{p-1}(K) \\ c = \sum \alpha_i \sigma_i &\mapsto \delta_p(c) = \sum \alpha_i(\delta_p(\sigma_i)) \end{aligned}$$

Let us apply this definition to the following example. In a slight abuse of notation, we denote a face $\{a, b, c\}$ by $abc$.



$$\begin{aligned}\delta_2(abc + bcd) &= \delta_2(abc) + \delta_2(bcd)\\ &= (ab + bc + ac) + (bc + cd + bd)\\ &= ab + ac + cd + bd\end{aligned}$$

$$\begin{aligned}\delta_2(abc + bcd + bce) &= (ab + bc + ac) + (bc + cd + bd) + (bc + ce + be)\\ &= ab + bc + ac + cd + bd + ce + be\end{aligned}$$

Note that an edge is in the boundary of a chain of triangles exactly if it is contained in an odd number of triangles of the chain.

The following lemma states that the boundary of a boundary is empty. This formalizes the intuition that we want to consider the boundary of a simplex to be a cycle, where the interior of the simplex fills up the cavity given by its boundary.

**Lemma 2.20.** *For* $p > 0$, $\delta_{p-1} \circ \delta_p(c) = 0$, *for any* $p$-*chain* $c$.

In the example above, $\delta_1(ab+ac+cd+bd) = (a+b)+(a+c)+(c+d)+(b+d) = 0$.

*Proof.* It is enough to show this for simplices, as $\delta_{p-1} \circ \delta_p(c) = \delta_{p-1}(\sum \alpha_i(\delta_p(\sigma_i))) = \sum \alpha_i(\delta_{p-1} \circ \delta_p(\sigma_i))$.
For a p-simplex $\sigma$, every $(p-2)$-face is contained in exactly 2 $(p-1)$-faces, and does thus not appear in $\delta_{p-1} \circ \delta_p(\sigma)$. □

For a k-dimensional simplicial complex K, we get a sequence of homomorphisms, called the *chain complex*:

$$0 = C_{k+1}(K) \xrightarrow{\delta_{k+1}} C_k(K) \xrightarrow{\delta_k} C_{k-1}(K) \cdots C_2(K) \xrightarrow{\delta_2} C_1(K) \xrightarrow{\delta_1} C_0(K) \xrightarrow{\delta_0} C_{-1} = 0$$

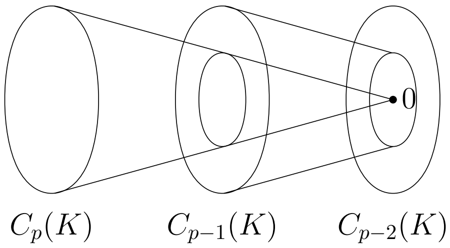


**Figure 2.6**: *A schematic illustration of a part of a chain complex.*

### 2.2.4 Cycle and boundary groups

As already mentioned, chains without boundaries are called cycles. These are the objects potentially giving rise to holes or cavities.

**Definition 2.21.** *A* $p$*-chain* $c$ *is a* $p$-cycle *if* $\delta(c) = 0$*.* $Z_p$ *is the* $p$-th cycle group*, consisting of all* $p$*-cycles.*

**Lemma 2.22.** $Z_p$ *is a group.*

*Proof.* $Z_p = \ker \delta_p$. (Recall that the kernel of a homomorphism is a subgroup of its domain.) □

So far we have only formally defined a boundary operator, but have not specified which chains we call boundaries. Of course, as already used implicitly before, the boundaries are the chains that are the result of applying the boundary operator.

**Definition 2.23.** *A* $p$*-chain* $c$ *is a* $p$-boundary *if* $\exists c' \in C_{p+1}$ *such that* $\delta(c') = c$*.* $B_p$ *is the* $p$-th boundary group*, consisting of all* $p$*-boundaries.*

**Lemma 2.24.** $B_p$ *is a group.*

*Proof.* $B_p = \operatorname{im} \delta_{p+1}$. □

**Fact 2.25.** $B_p \subseteq Z_p \subseteq C_p$*, and all of them are abelian and free.*

We will not prove this statement here, but to see that $B_p \subseteq Z_p$, recall that by Lemma 2.20 the boundary of a boundary is empty.

### 2.2.5 Homology Groups

We are now ready to formalize the notion of holes or cavities. Recall that intuitively, a hole is a cycle that is not a boundary, that is, not filled by something higher-dimensional. Using that all objects defined so far form abelian groups, we can phrase this in algebraic terms using quotient groups.

**Definition 2.26.** *The* $p$-th homology group $H_p(K; \mathbb{Z}_2)$ *is the quotient group* $Z_p(K)/B_p(K)$*.*

The $\mathbb{Z}_2$ in the notation emphasizes that this is homology over $\mathbb{Z}_2$. In general, we write $H_p(K; R)$ for homology over some ring $R$. As in these lecture notes we only work with homology over $\mathbb{Z}_2$, we will from now on just write $H_p(K)$ for homology over $\mathbb{Z}_2$.

In essence, in the homology group cycles that differ only by boundaries are equivalent. More formally, the coset $[c] = c + B_p$ is the *homology class* of $c$. We say that $c$ and $c'$ are *homologous*, if $[c] = [c']$, or equivalently $c \in c' + B_p$ or equivalently $c + c' \in B_p$. See Figure 2.7 for an example of homologous cycles, and Figure 2.9 for an example of the homology group of a small complex.

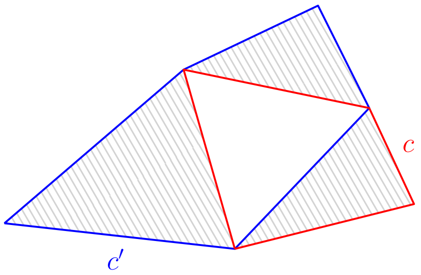


**Figure 2.7**: $c'$ *and* $c$ *are homologous cycles.*

**Exercise 2.27**. *Visualize the following simplicial complex* $K$*: 0-faces* $\{a, b, c, d, e\}$*, 1-faces* $\{ab, ac, ad, bc, bd, cd, ce, de\}$ *and 2-faces* $\{abc, abd, acd, bcd\}$*. For the dimensions 1 & 2, what are the cycle, boundary, and homology groups of* $K$*? Note: You can express the groups by their generators. You do not need to write out all the elements.*

**Exercise 2.28**. *Give an informal derivation for the homology groups of a torus (see Figure 2.8). Can you find a space with isomorphic homology that is not homeomorphic to the torus?*

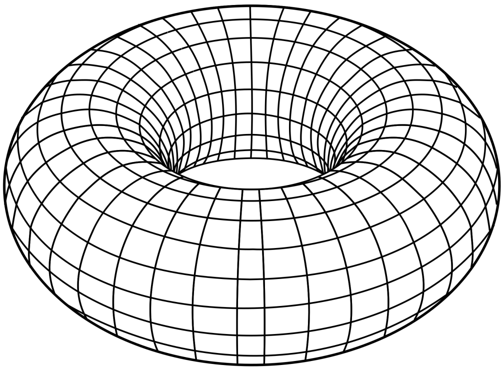


**Figure 2.8**: *A torus.*

**Exercise 2.29.** *For a simplicial complex* $\mathsf{K}$*, its cone* $\mathsf{CK}$ *is the complex with the same set of vertices plus one additional vertex* $z$*, and such that for all simplices in* $\mathsf{K}$ *we have*

$$\{a, b, c, \ldots\} \in \mathsf{K} \implies \{a, b, c, \ldots, z\} \in \mathsf{CK}$$

*(a) Visualize a cone operation. What does it intuitively do to a complex?*

*(b) Show that the homology of the cone* $\mathsf{CK}$ *is* $0$ *in all dimensions* $d > 0$*, for any* $\mathsf{K}$*.*

*(c) Bonus: What would happen (intuitively and to the homology) if we extended* $\mathsf{K}$ *in the same way as before, but with two points? (this is called the* suspension *of* $\mathsf{K}$*)*

Here are some nice properties of homology groups, that will be beneficial for us, but that we will not prove here.

**Fact 2.30.**

- $\mathsf{H}_p$ *is abelian and free.*
- $\mathsf{H}_p$ *is a* $\mathbb{Z}_2$*-vector space.*

**Remark 2.31.** *If we consider homology defined over other rings, e.g. over* $\mathbb{Z}$ *instead of* $\mathbb{Z}_2$*, the homology groups might not be free.*

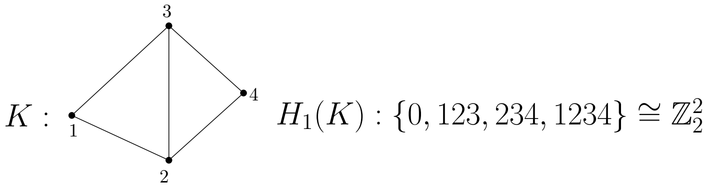


**Figure 2.9:** *The first homology group of a small complex.*

Recall that our original motivation was to count the number of holes. With homology as we defined it, we have the algebraic structure of a vector space where we can add holes. The number of distinct holes is now just the dimension of this vector space.

**Definition 2.32.** $\beta_p := \dim \mathsf{H}_p = \dim \mathsf{Z}_p - \dim \mathsf{B}_p$ *is the* p-th Betti number.

In the definition above, dim denotes the dimension of a vector space as you know it from Linear Algebra, i.e., $\dim \mathsf{H}_p$ is the number of elements in a basis of $\mathsf{H}_p$.

**Exercise 2.33.** *The* **Euler characteristic** *of a simplicial complex* $K$ *is defined as*

$$\chi = k_0 - k_1 + k_2 - \dots$$

*with* $k_i$ *denoting the number of i-dimensional simplices in* $K$*. Convince yourself that this is an invariant property for all triangulations of the same topological space* $X$*.*

Hint: *Show instead that* $\chi = \beta_0(K) - \beta_1(K) + \dots$*. The statement then follows by the fact that homeomorphic spaces have the same homology.*

**Exercise 2.34.** *Take any vector* $v = (a_0, \dots, a_d) \in \mathbb{N}^{d+1}$ *with* $a_0 > 0$*. Show that there exists a simplicial complex* $K_v$ *with that vector as its Betti numbers.*

### 2.2.6 Singular Homology

With our definition of homology for simplicial complexes, we get for free a notion of homology for many topological spaces, namely the triangulable ones: we can simply triangulate them and take the homology of the triangulation. But, so far, it seems like the structure of the homology group might differ depending on the choice of triangulation. The aim of this section is to sketch the tools that show that the homology of a triangulable space is independent of the chosen triangulation. The idea of singular homology is to remove the need for a fixed triangulation by looking at all possible simplices at once.

Let $X$ be a topological space, and let $\Delta^p$ be the standard p-simplex in $\mathbb{R}^{p+1}$. We want to consider all possible occurrences of this simplex in $X$.

**Definition 2.35.** *A* singular p-simplex *is a map* $\sigma : \Delta^p \to X$*.*

Note that in this definition we do not require $\sigma$ to be injective, thus it would even be possible to map the simplex to a single point.

We now define $C_p$ the same way as before, but now on the family of all singular p-simplices, which in general makes the group uncountably infinite. We also define $\delta_p$ as before, leading to $Z_p$ and $B_p$ now also being uncountably infinite. Similarly, $H_p(X) = Z_p(X)/B_p(X)$. The following relates singular homology and simplicial homology.

**Theorem 2.36.** *Let* $X$ *be a topological space,* $K$ *a triangulation of* $X$*. Then we have* $H_p(X) \cong H_p(K)$ *for all* $p \geqslant 0$*.*

As isomorphisms for vector spaces are an equivalence relation, we also get the desired independence of the triangulation.

**Corollary 2.37.** *Let* $K_1, K_2$ *be two distinct triangulations of* $X$*. Then,* $H_p(K_1) \cong H_p(K_2)$ *for all* $p \geqslant 0$*, that is, homology is independent of the chosen triangulation.*

For the remainder of these notes, we will only work with simplicial homology, but we often talk about the homology of a triangulable space without specifying a triangulation. The above corollary gives us the right to do this.

### 2.2.7 The 0-th homology group

In this section, we take a closer look at the 0-th homology group. Recall that the 0-simplices of a simplicial complex $K$ are simply its vertices. Since vertices do not have any boundaries, every vertex is a 0-cycle. The boundary of a 1-simplex simply consists of the two vertices which are connected by the edge. We can thus see that two vertices $v_1$ and $v_2$ are homologous if there is a path from $v_1$ to $v_2$, and the homology class $[v_1]$ is simply the connected component containing $v_1$.

**Observation 2.38.** $\beta_0(K)$ *is the number of connected components of* $K$.

Further, the 0-homology classes are the formal sums of connected components.

### 2.2.8 Homology of Spheres

We will now investigate the homology of the spheres $S^d$. Since we have seen in Section 2.2.6 that homology is independent from the chosen triangulation, let us fix some triangulation of the sphere $S^d$. A good candidate (for its simplicity) is the boundary of a simplex, that is, $S^d \simeq \delta(\Delta^{d+1})$, with the vertex set $V = \{v_0, \dots, v_{d+1}\}$.

$H_0(S^d)$: Let us first investigate $H_0(S^d)$. Since all vertices are connected, all vertices are homologous, and $H_0(S^d) = \langle [v] \rangle \cong \mathbb{Z}_2$.

$H_d(S^d)$: Now, let us check $H_d(S^d)$. We first compute $Z_d$: Obviously, the zero element is part of $Z_d$. Furthermore, the d-simplices are exactly the sets $\sigma_i = \{v_0, \dots, \hat{v_i}, \dots, v_{d+1}$. The sum $c$ of all these d-simplices must be a cycle, since every $d-1$-simplex occurs in exactly two d-simplices, thus the boundary of $c$ must be empty. Thus, $c \in Z_d$. We cannot have any other cycle, since for any other chain there must be some d-simplex for which we include one neighbor but not the other, thus this d-simplex would be part of the boundary. We conclude that $Z_d(S^d) = \langle c \rangle$.

Since $\delta(\Delta^{d+1})$ is a d-dimensional simplicial complex, and thus does not contain any $(d+1)$-simplices, $c$ cannot be a boundary. Since $B_d$ is a subgroup of $Z_d$, we thus get that $B_d(S^d)$ is the group containing only 0. Alternatively, we can also get this by noticing that $C_{d+1} = 0$, and $B_d = \operatorname{im} \delta_{d+1} = 0$.

We finally get $H_d(S^d) = Z_d / B_d = Z_d \cong \mathbb{Z}_2$.

$H_p(S^d)$: Finally, let us go to $H_p(S^d)$, for $0 < p < d$: Let $c = \sum \alpha_i \sigma_i$ be a p-cycle. We aim to show that $c$ is homologous to the 0-chain, i.e., that $[c] = 0$. Equivalently, we show that $c$ must be a boundary.

Let $\sigma = (v_{m_0}, \dots, v_{m_p})$ be any p-simplex in $c$ which does not include $v_0$. We will keep replacing such simplices by simplices which do contain $v_0$, until we have no more simplices not containing $v_0$.

Let $b$ be the $(p+1)$-simplex $(v_0, v_{m_0}, \dots, v_{m_p})$. Note that $b \in \delta(\Delta^{d+1})$ and thus $\delta(b)$ is a p-boundary. Also note that $\sigma$ is in $\delta(b)$. Furthermore, $\sigma$ is the only p-simplex in $\delta(b)$ which does not contain $v_0$. We now add $\delta(b)$ to $c$, to get $c' := c + \delta(b)$. Since we added a boundary, $[c] = [c']$ (i.e., $c$ and $c'$ are homologous). Furthermore, $c'$ contains one fewer p-simplex not containing $v_0$, when compared to $c$.

We repeat this process until we reach a cycle $c^*$ in which every p-simplex contains $v_0$. We now claim that $c^*$ must be the trivial cycle: Assume $c^*$ contains some p-simplex $a = (v_0, v_{a_1}, \dots, v_{a_p})$. Then, the $(p-1)$-simplex $a' = (v_{a_1}, \dots, v_{a_p})$ is part of $\delta(a)$. But, $a'$ cannot be part of the boundary of any other p-simplex in $c^*$, since the only p-simplex containing $a'$ as a face while also containing $v_0$ is $a$. Thus, to have an empty boundary, $c^*$ must be 0. We thus have $[c^*] = 0$, and by construction, $[c] = [c^*]$, therefore $[c] = 0$ as we aimed to prove.

We have proven that every cycle is homologous to 0, and we can conclude that for all $0 < p < d$, $H_p(S^d) = 0$.

By these arguments we conclude the following theorem:

**Theorem 2.39.** *For any $d > 0$, we have*

$$H_p(S^d) = \begin{cases} \mathbb{Z}_2 & p \in \{0, d\} \\ 0 & \textit{else.} \end{cases}$$

$$\beta_p(S^d) = \begin{cases} 1 & p \in \{0, d\} \\ 0 & \textit{else.} \end{cases}$$

### 2.2.9 Induced Homology

As in many mathematical theories, we are not only interested in single objects, but also in the maps between them. For simplicial complexes we have defined simplicial maps between them, and we now want to study the effect that simplicial maps have on the homology of a space.

Let $f : K_1 \to K_2$ be a simplicial map. This induces a *chain map*

$$\begin{aligned} &f_\# : C_p(K_1) \to C_p(K_2) \\ &c = \sum \alpha_i \sigma_i \mapsto f_\#(c) := \sum \alpha_i \tau_i, \text{ where } \tau_i = \begin{cases} f(\sigma_i) & \text{if } f(\sigma_i) \text{ is p-simplex in } K_2 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Note that $f(\sigma_i)$ is always a simplex in $K_2$ since $f$ is a simplicial map, but it could be a simplex of smaller dimension. This is why we have the condition in the above definition of $\tau_i$.

The following can be shown with a bit of work:

- $f_\# \circ \delta = \delta \circ f_\#$
- $f_\#(B_p(K_1)) \subseteq f_\#(Z_p(K_1))$

- $f_{\#}(Z_p(K_1)) \subseteq Z_p(K_2)$, $f_{\#}(B_p(K_1)) \subseteq B_p(K_2)$

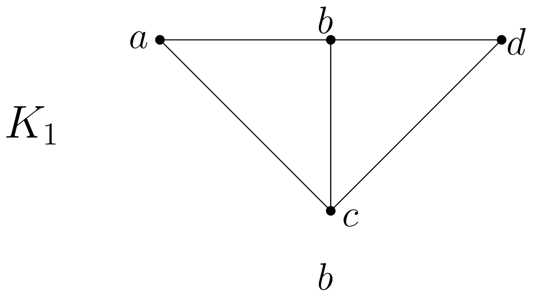
From this chain map $f_{\#}$, we now get a well-defined *induced homomorphism* between the homology groups of $K_1$ and $K_2$:

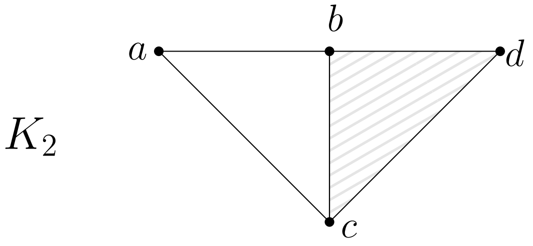
$$\begin{aligned} f_* &: H_p(K_1) \to H_p(K_2) \\ [c] = c + B_p &\mapsto f_{\#}(c) + B_p(K_2) = [f_{\#}(c)] \end{aligned}$$

**Fact 2.40.** *If* $H_p(K_1)$ *and* $H_p(K_2)$ *are vector spaces (as they are in e.g.* $\mathbb{Z}_2$*-homology, which is what we are using), then* $f_*$ *is a linear map.*

We also get the following *functorial property*, which we will not prove: if $f : X \to Y$, $g : Y \to Z$, then $(g \circ f)_* = g_* \circ f_*$.

Let us look at a small example:





We consider $f : K_1 \hookrightarrow K_2$ the inclusion map.

$$H_1(K_1) = \{0, [abc], [bcd], [abdc]\} \cong \mathbb{Z}_2^2$$

$$f_*(0) = 0,\ f_*([abc]) = [abc]$$

$$f_*([bcd]) = 0,\ f_*([abdc]) = [abc]$$

**Exercise 2.41.** *Let*

$$K_1 = \{\emptyset, a, b, c, d, e, ab, ac, bc, bd, cd, ce, de, abc\}$$

*and*

$$K_2 = \{\emptyset, w, x, y, z, wx, wy, xy, xz, yz\}.$$

*Consider the map* $f : K_1 \to K_2$ *induced by*

$$a \mapsto y, b \mapsto x, c \mapsto y, d \mapsto z, e \mapsto z.$$

*You can verify easily that* $f$ *is simplicial. Compute* $f_* : H_p(K_1) \to H_p(K_2)$ *for* $0 \leqslant p \leqslant 2$.

**Exercise 2.42.** *Which of the following four statements is true for every simplicial map* $f$*?*
*"If* $f$ *is* $\{$*injective, surjective*$\}$*, then* $f_*$ *is* $\{$*injective, surjective*$\}$*."*

The following fact has some very powerful consequences, as we will see.

**Fact 2.43.** *If* $f, g : K_1 \to K_2$ *are contiguous,* $f_* = g_*$*.*

Note that the definition of induced homology extends from simplicial maps to maps between any topological spaces. We will not state the exact definitions, but the following fact is the continuous analogue (remember that two simplicial maps being contiguous is analogous to two maps being homotopic) of the previous fact.

**Fact 2.44.** *If* $f, g : X \to Y$ *are homotopic,* $f_* = g_*$*.*

The following corollary is very useful to compute the homology of a space, as it gives us the option to relate it to the homology of a potentially simpler space.

**Corollary 2.45.** *If* $f : X \to Y$ *is a homotopy equivalence (i.e., there exists* $g : Y \to X$ *such that* $f \circ g$ *is homotopic to* $\mathrm{id}_Y$ *and* $g \circ f$ *is homotopic to* $\mathrm{id}_X$*), then* $f_*$ *is an isomorphism.*

In particular, if $Y$ is a deformation retract of $X$, then $H_p(Y)$ and $H_p(X)$ are isomorphic. As a special case of the above, we have that a contractible space has the same homology groups as a point.

**Corollary 2.46.** *If* $X$ *is contractible,* $H_p(X) = \begin{cases} \mathbb{Z}_2 & p = 0, \\ 0 & \text{otherwise.} \end{cases}$

**Exercise 2.47.**

*Consider the space you get when you glue together two points of a torus. What is the homology of this space?*

*Consider the space you get when you simultaneously pierce a balloon at* $n$ *distinct locations. What is the homology of this space?*

**Exercise 2.48.** *Let* $f, g : S^1 \to S^1$ *be continuous maps such that* $f(-x) = f(x)$ *and* $g(-x) = -g(x)$ *for all* $x \in S^1$*.*

*a) Convince yourself that* $f_* : H_1(S^1) \to H_1(S^1)$ *is trivial (maps everything to* $0$*) and that* $g_*$ *is an isomorphism.*

*b) Show that* $f$ *and* $g$ *are not homotopic.*

*c) Show that there is no map* $h : S^2 \to S^1$ *such that* $h(-x) = -h(x)$*.*

*d) Conclude that every map* $\phi : S^2 \to \mathbb{R}^2$ *with* $\phi(-x) = -\phi(x)$ *has a zero.*

*The statement you have proven in d) is equivalent to the 2-dimensional case of the famous Borsuk-Ulam theorem, which implies statements such as "at any time, there are two antipodal points on the earth with both the same temperature and pressure".*

### 2.2.10 Application: Brouwer fixed point theorem

In this section we finally collect the fruits of our hard work by using homology to give a relatively short proof of the famous fixed point theorem by Brouwer. Here, $B^d$ denotes the unit ball of dimension $d$.

**Theorem 2.49** (Brouwer fixed point theorem). *Let* $f : \mathbb{B}^d \to \mathbb{B}^d$ *be continuous. Then,* $f$ *has a fixed point, that is,* $\exists x \in \mathbb{B}^d$ *such that* $f(x) = x$.

This theorem has many fascinating implications:

- Take two sheets of paper lying on top of each other. Crumple the top sheet and set it back onto the other sheet. No matter how you crumpled the sheet, at least one point of the crumpled sheet lies exactly above its corresponding point in the bottom sheet.
- If you open a map of Switzerland in Switzerland, there is at least one point on the map which is at its exact position.
- If you take a cup of liquid and stir or slosh it, at least one atom ends up at its original position (but if you shake you might break continuity).
- The theorem also has many applications in mathematics and computer science, such as in fair divisions or for proving existence of Nash equilibria.

To prove Theorem 2.49, we first introduce the following definition and a helper lemma, which we only prove after proving Theorem 2.49 itself.

**Definition 2.50.** *A map* $r : X \to A \subseteq X$ *is a* retraction *if* $r(a) = a$, $\forall a \in A$.

**Lemma 2.51.** *There is no retraction* $r : \mathbb{B}^d \to S^{d-1}$.

*Proof of Theorem 2.49.* We prove the theorem by contradiction. For an illustration of the argument see Figure 2.10. Assume $f : \mathbb{B}^d \to \mathbb{B}^d$ has no fixed point. For each $x$, consider the ray $\overrightarrow{f(x)x}$ and let $r(x)$ be the intersection of this ray with $S^{d-1}$. Then, $r : \mathbb{B}^d \to S^{d-1}$ is continuous (which we do not prove here) and $r(s) = s \ \forall s \in S^{d-1}$, since no matter where $f(s)$ lies, $\overrightarrow{f(s)s}$ first intersects $S^{d-1}$ in $s$. Thus, $r$ is a retraction, which does not exist by Lemma 2.51. □

It remains to prove the helper lemma.

*Proof of Lemma 2.51.* Consider $i$, the inclusion map $S^{d-1} \hookrightarrow \mathbb{B}^d$, and a retraction $r : \mathbb{B}^d \to S^{d-1}$. By definition, we have $r \circ i = \mathrm{id}$. Let us look at the induced maps of $r$ and $i$ in the $(d-1)$-th homology of $S^{d-1}$ and $\mathbb{B}^d$. Recall that $H_{d-1}(S^{d-1}) \cong \mathbb{Z}_2$ and $H_{d-1}(\mathbb{B}^d) \cong 0$. We thus view $i_*$ as a homomorphism from $\mathbb{Z}_2$ to $0$, and $r_*$ as a homomorphism from $0$ to $\mathbb{Z}_2$. But since $r \circ i = \mathrm{id}$, we also have $r_* \circ i_* = \mathrm{id}$. We can combine this to reach a contradiction:

$$1 = \mathrm{id}(1) = r_* \circ i_*(1) = r_*(0) = 0$$

Thus, either $i$ or $r$ cannot exist, but since $i$ exists, $r$ cannot. □

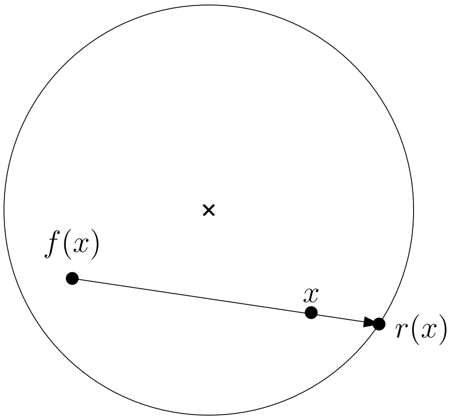


**Figure 2.10**: *If* f *has no fixed point, we get a retraction to the boundary.*

## Questions

5. *What is a simplicial complex?* Define geometric and abstract simplicial complexes and state and prove the realization theorem (Theorem 2.5).
6. *What are simplicial and contiguous maps?* State the definitions and discuss the connection to their counterparts in continuous topology.
7. *Is every contractible simplicial complex collapsible?* Define the notion of collapsibility and describe Bing's house with two rooms.
8. *What is simplicial homology?* Explain the intuition and give the formal definitions of chains, boundaries and cycles.
9. *Why is the homology of a triangulable space independent of the chosen triangulation?* Explain the idea of singular homology.
10. *What are the homology groups of a sphere?* State and prove the corresponding theorem (Theorem 2.39).
11. *How does a simplicial map between two simplicial complexes induce maps between their homology groups?* Define induced homomorphisms.
12. *What is the Brouwer fixed point theorem?* State, illustrate and prove the Brouwer fixed point theorem (Theorem 2.49).

## References

[1] Sketches of topology - Bing's house. `https://sketchesoftopology.wordpress.com/2010/03/25/bings-house/`, accessed: 2023-04-27.

[2] Allen Hatcher, *Algebraic topology*, Cambridge Univ. Press, Cambridge, 2000.

# Chapter 3

# Persistence

In the previous chapter, we have studied the homology of fixed simplicial complexes. In this chapter, we will look at simplicial complexes that vary over time. Let us start with a small example. Consider the following process of building up a triangle $abc$. At time $t_1$, we add the vertices $a$ and $b$ together with the edge $ab$. This gives birth to a single connected component. At time $t_2$ we add the vertex $c$, giving birth to a second connected component. At time $t_3$ we add the edge $ac$, connecting the two components. We can interpret this as the younger of the components dying. At time $t_4$ we add the final edge $bc$, which gives birth to a hole, that is, an element of the homology group $H_1$. Finally, at time $t_5$ we add the interior of the triangle, killing the hole born at $t_4$. We can summarize this process as follows: we have a connected component that was born at $t_1$ and survived the entire process, and a connected component that was born at $t_2$ that died again at $t_3$. Finally, we have a hole born at $t_4$ dying at $t_5$. Capturing this information of holes with their birth and death is the motivation of persistent homology.

Persistent homology can be applied to data analysis by defining (in a way that we will see soon) a process to build up a simplicial complex from point cloud data and computing the birth and death times of holes. Subtracting the birth time from the death time, we get the lifespan of a hole, and the underlying idea is that holes with a short lifetime are a byproduct of the process, whereas holes with a long lifespan convey information about the shape of the underlying data.

## 3.1 Filtrations

We start by a mathematical formulation of the process of building up a complex or, more general, a topological space. A *filtration* is a nested sequence of subspaces

$$\mathcal{F} : X_0 \subseteq X_1 \subseteq X_2 \subseteq \ldots \subseteq X_n = X.$$

For each $i \leqslant j$, we have the inclusion map $\iota_{i,j} : X_i \hookrightarrow X_j$. Given these functions $\iota$, we get induced maps in homology: $h_p^{i,j} = \iota_* : H_p(X_i) \to H_p(X_j)$. Filtrations are a very general object that appear naturally in many settings. Let us look at some important examples of filtrations.

- Given a function $f : X \to \mathbb{R}$, we can define the (uncountably infinite) *sublevel set filtration* $X_a = f^{-1}(-\infty, a]$.
- A *simplicial filtration* is a nested sequence of subcomplexes

  $$\mathcal{F} : K_0 \subseteq K_1 \subseteq \ldots \subseteq K_n = K.$$

  We call a simplicial filtration *simplex-wise*, if $K_i \setminus K_{i-1}$ is a single simplex (or empty).
- We call a function $f : K \to \mathbb{R}$ *simplex-wise monotone* if for every $\sigma \subseteq \tau$ we have $f(\sigma) \leqslant f(\tau)$. A simplex-wise monotone function guarantees us that the sublevel set filtration by f gives a proper simplicial filtration. Note that it does not necessarily guarantee us that the sublevel set filtration is simplex-wise (e.g., consider a function f that is not injective).
- We can also define a simplicial filtration by ordering our vertices $v_0, v_1, \ldots, v_n$. Then, let $K_i$ be the simplicial complex induced by the vertices $v_0, \ldots, v_i$. Then, we call the simplices $K_i \setminus K_{i-1}$ added when adding $v_i$ the *lower star* of $v_i$. Thus, this type of filtration is also called the *lower star filtration.*

The following simplicial filtration captures the process that is relevant for analyzing point cloud data.

**Definition 3.1.** *Let* $(M, d)$ *be a metric space. Let* $P$ *be a finite subset of* $M$*, and* $r > 0$ *a real number. The* Čech complex $\mathbb{C}^r(P)$ *is the nerve of the family of balls* $B(p, r) = \{x \in M | d(p, x) \leqslant r\}$ *for all* $p \in P$.

Since the balls $B(p, r)$ form a good cover, the nerve theorem tells us that the Čech complex is homotopy equivalent to the union of the balls.

By looking at the sequence of Čech complexes for increasing r, we get a simplicial filtration.

## 3.2 Persistent Homology

As we have seen, from a filtration $X_0 \subseteq X_1 \subseteq \ldots \subseteq X_n$ we get a sequence of homology groups with homomorphisms between them:

$$H_p(\mathcal{F}) : H_p(X_0) \to H_p(X_1) \to H_p(X_2) \to \ldots \to H_p(X_n).$$

Such an object is called *persistence module.*

**Definition 3.2.** *The* p-th persistent homology group $H_p^{i,j}$ *is defined by*

$$H_p^{i,j} := \operatorname{im} h_p^{i,j} = Z_p(K_i)/(B_p(K_j) \cap Z_p(K_i)).$$

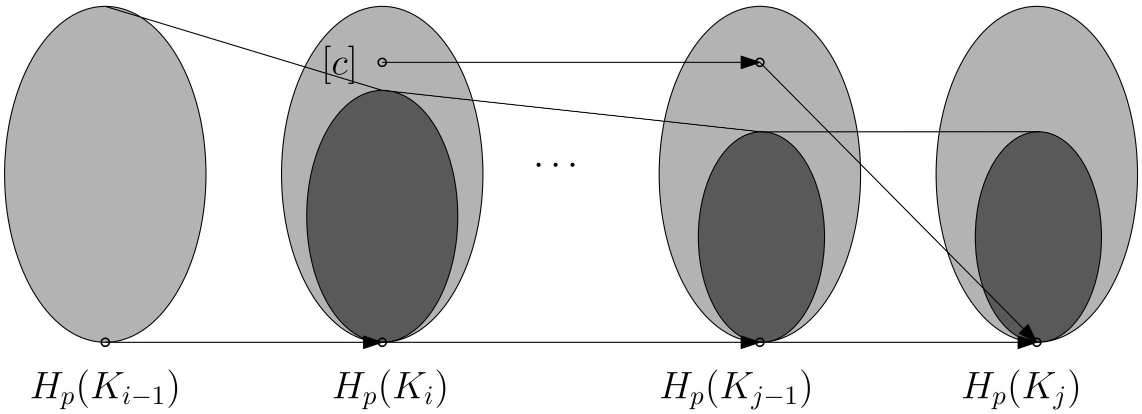


**Figure 3.1:** *An illustration of a class* $[c]$ *being born at* $K_i$ *and dying entering* $K_j$.

This definition characterizes the cycles that that are present already in $K_i$ and that are not boundaries even in $K_j$.

**Definition 3.3.** *The* p-th persistent Betti numbers $\beta_p^{i,j}$ *are the dimensions of the* $p$*-th persistent homology groups:* $\beta_p^{i,j} = \dim H_p^{i,j}$.

**Exercise 3.4.** *Let* $p \geqslant 1$. *For every* $n \geqslant 1$, *construct a filtration* $X_1 \subseteq X_2 \subseteq \ldots \subseteq X_n$ *such that*

- $H_p(X_k) \neq 0$ *for all* $k \in \{1, \ldots, n\}$ *and*
- $H_p^{i,j} = 0$ *for all* $i < j$.

We say that a p-homology class $[c]$ (a p-hole) is *born at* $K_i$ if $[c] \in H_p(K_i)$ but $[c] \in H_p^{i-1,i}$. Similarly, $[c]$ *dies entering* $K_j$, if $[c] \neq 0$ in $H_p(K_{j-1})$ but $h_p^{j-1,j}([c]) = 0$.

It is not always obvious which homology class dies. Consider the following filtration: $X_1$ consists of two points $a$ and $b$, and in $X_2$ the two points are connected by an edge. Let us look at $H_0$, that is, the connected components. We have that $H_0(X_1) \simeq \mathbb{Z}_2^2$, with the natural basis $\{[a], [b]\}$. On the other hand, in $X_2$ there is only a single connected component, and $[a] = [b]$. So a homology class is dying, but both our basis elements $[a]$ and $[b]$ survive. What is happening?

It turns out that we were not careful with our choice of basis: $H_0(X_1)$ can also be viewed as being generated by $[a]$ and $[a+b]$, and the class $[a+b]$ indeed dies going into $X_2$. In general, if two homology classes merge, they both do not die, but their sum does. There is a consistent choice of basis which allows us to only look at persistent homology in terms of basis elements, but we do not go into this at this point.

If we have a simplex-wise filtration, we can circumvent the above issue by sorting homology classes by the time where they were born, and when they merge, we just say the "younger one" dies. This can be seen as adapting the considered basis along the way.

*Persistence pairings* are another way around this issue. We add some final complex $K_{n+1}$ which has trivial homology (i.e., by adding all simplices that are not yet present).

Then, we aim to figure out how many holes get born at $K_i$ and die entering $K_j$. For this, we define

$$\mu_p^{i,j} = (\beta_p^{i,j-1} - \beta_p^{i,j}) - (\beta_p^{i-1,j-1} - \beta_p^{i-1,j}), \text{ for } i < j \leqslant n+1.$$

Here, the content of the left parenthesis denotes the number of holes born at or before $K_i$, which die entering $K_j$. Conversely, the right parenthesis denotes the number of holes born strictly before $K_i$, and die entering $K_j$. Thus, subtracting the two, gives the number of holes born exactly at $K_i$ and die entering $K_j$. Note that this conveys the information that we are interested in, but does not require choosing any basis.

The *persistence diagram* $\mathrm{Dgm}_p(\mathcal{F})$ is a birth-death diagram which contains a point for every pair $i, j$ for which $\mu_p^{i,j} > 0$. If we give each $K_i$ a timestamp $a_i$, the point is drawn at the coordinates $(a_i, a_j)$. We give each point multiplicity $\mu_p^{i,j}$. On the diagram we add points on the diagonal with infinite multiplicity, for some technical reasons that will become apparent later. We can also represent the same information by *barcodes*: For every $i, j$, we draw $\mu_p^{i,j}$ many intervals $[a_i, a_j]$. This is then called the p-th persistence barcode.

**Exercise 3.5.** *Consider the simplex-wise filtration induced by the order* $\sigma_1, \ldots, \sigma_N$ *on the simplices of a complex* $K$.

*When does the order*

$$\sigma_1, \ldots, \sigma_{k-1}, \sigma_{k+1}, \sigma_k, \sigma_{k+2}, \ldots, \sigma_N$$

*induce a simplex-wise filtration? When it does, describe the relation between the corresponding persistence diagrams.*

**Exercise 3.6.** *Give two filtrations* $X_1 \subseteq \ldots \subseteq X_n$ *and* $Y_1 \subseteq \ldots \subseteq Y_n$ *that have the same persistence diagrams but for which for any* $i \in \{1, \ldots, n\}$, $X_i$ *is not homotopy-equivalent to* $Y_i$.

## 3.3 Algorithms for persistent homology

### 3.3.1 Persistence pairing algorithm

We consider a simple-wise filtration. Consider some $j$, and let $p$ be the dimension of the simplex added in $K_j$, i.e., $K_j \setminus K_{j-1} = \sigma_j$ is a p-simplex. There are only two things that can happen when adding $\sigma_j$: Either, a new non-boundary p-cycle $c$ (a hole) is born. Then we say that $\sigma_j$ is a *creator*. It is also possible that adding $\sigma_j$, a $(p-1)$-cycle becomes a boundary, thus a hole dies. Then we say that $\sigma_j$ is a *destructor*. The fact that at exactly one of the two events happens is a consequence of the Euler characteristic, which is discussed in Exercise 2.33.

The persistence pairing algorithm pairs a destructor $\sigma$ with the youngest still unpaired creator within the cycle it destroys. To find this youngest unpaired creator, we look at

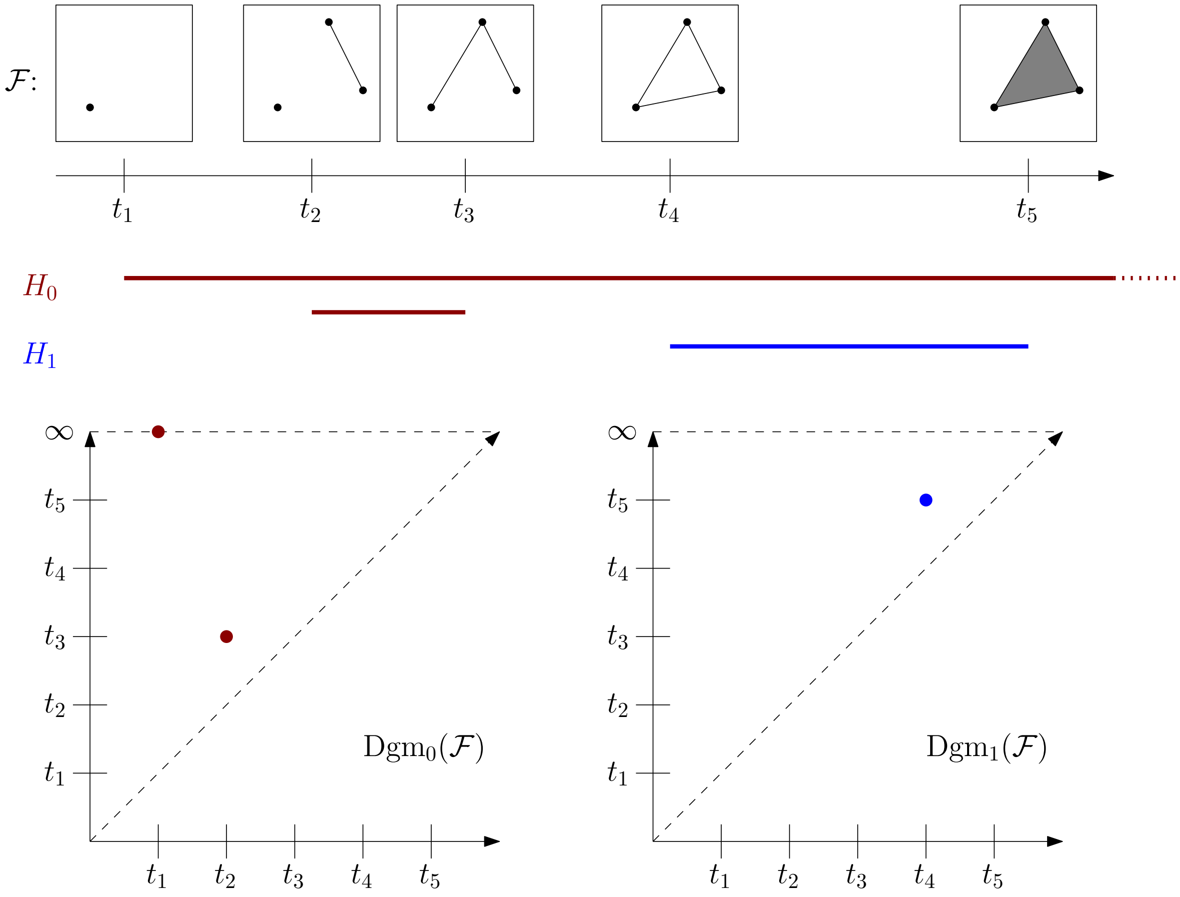


**Figure 3.2:** *An example of a filtration with the corresponding barcodes and persistence diagrams.*

the boundary of $\sigma$. We try pairing $\sigma$ to the youngest element $\rho$ of its boundary. If this element is already paired with some element $\tau$, we replace it by the sum of $\rho$ and the boundary of $\tau$. We now have a new set of candidate creators. We repeat this process until we found an unpaired creator we can pair to, or until we cannot continue (there are no more candidates). If we cannot pair $\sigma$ to anything, it must be a new creator. Whatever unpaired creators remain at the end of the algorithm are paired to an element $\infty$.

What is the runtime of this algorithm? Let $N$ be the total number of simplices in the final complex of our filtration. Whenever we add a simplex, and we replace a simplex by the boundary of its paired destructor, we add at most $O(N)$ simplices. We have to do this at most $O(N)$ times. Since we do this for each simplex, we get a runtime of $O(N^3)$. Surprisingly, this runtime is tight.

**Exercise 3.7.** *Let* $G$ *be a weighted connected graph, where all edge weights are pairwise distinct. Consider a filtration that first inserts all vertices (in some arbitrary order) and then inserts the edges one by one, ordered by increasing weight. What is the set of destructors?*

### 3.3.2 Matrix reduction algorithm

In practice, a different algorithm is used, which actually does the same but in the language of matrices. This is the *Matrix Reduction Algorithm.* Here, the filtration does not necessarily have to be simplex-wise. We write a large matrix, which is $N \times N$. Both rows and columns are labeled by the simplices, ordered by order of insertion. If several simplices are added at the same time, the ones of smaller dimension are put before the ones of larger dimension. We then insert a 1 at row $\sigma$ and column $\tau$, if $\sigma$ is part of the boundary of $\tau$. We now process the columns from left to right. For each column, we look at the lowest 1 in the column. If there is a 1 in the row towards the left of that column, add the column containing that other 1 to the current column (in $\mathbb{Z}_2$). At the end, empty columns correspond to creators (births). To find the death of a creator, look at its corresponding row, and find a pivot element in this row (a 1 which is the lowest 1 of its column). If there is no pivot element, this creator never dies, i.e., is unpaired.

Let us again look at the runtime. For each column ($O(N)$), we might have to add $O(N)$ times a column, and each addition takes $O(N)$. So again, we have $O(N^3)$ runtime. But, we can write the algorithm in a way such that it runs in $O(N^\omega)$, where $\omega$ is the matrix-multiplication exponent. In practice, it runs in essentially $O(N)$.

**Exercise 3.8.** *Consider the following simplicial complex, and the simplex-wise filtration which first inserts the vertices in the order* $a, b, c, d, e$*, and the rest of the simplices as specified by the numbering in Figure Figure 3.3.*

*Execute both the persistence pairing algorithm and matrix reduction algorithm on this filtration. What are the similarities and differences in the algorithms? To better see what happens, label the columns in the matrix by the sum of columns they currently represent.*

*Represent the results you obtained by a persistence diagram, and also by the persistence barcodes.*

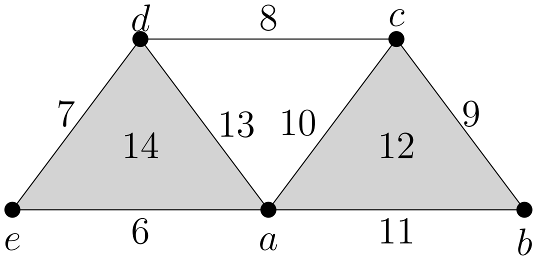


**Figure 3.3:** *The filtration for Exercise 3.8.*

**Exercise 3.9.** *A Union-Find data structure is a data structure that maintains disjoint sets dynamically. Given a ground set* $X$*, such a data structure maintains a family* $S$ *of disjoint subsets of* $X$*, where each subset is represented by the* smallest *element contained in it. It supports three operations:* `MakeSet(x)` *creates a new set* $\{x\}$*.* `FindSet(x)` *returns the representative (minimum) of the set in* $S$ *which contains* $x$ *(or "no" if* $x$ *is not contained in any set).* `Union(x, y)` *merges the sets containing* $x$ *and* $y$ *into a single one. All of these operations can be implemented in amortized* $\Theta(\alpha(n))$ *time, where* $\alpha$ *is the extremely slowly growing inverse Ackermann function and can be considered a constant for any real world application.*

*Consider a simplicial complex* $K$ *with its vertices ordered* $v_0, \dots, v_n$*, and consider its lower star filtration. Find an algorithm to compute the* $0$*-dimensional persistence diagram (i.e., the persistence pairings) of* $K$ *which makes use of a Union-Find data structure. How many Union-Find operations do you need to perform?*

## 3.4 Simplicial Complexes on Point Sets

In general, the data we wish to analyze will not come in the form of a simplicial filtration, so in order to use persistent homology we need to transform our data into one. Ideally, the way we do this should retain the underlying shape of the data, that we want to analyze. In this section we discuss several ways of constructing simplicial complexes from point cloud data or, more generally, finite metric spaces (i.e., a finite set of data points with given pairwise distances).

### 3.4.1 Čech and Vietoris-Rips complexes

**Definition 3.10.** *Given a metric space* $(M, d)$*, a finite point set* $P \subseteq M$*, and a real number radius* $r > 0$*, the* Čech complex $\mathbb{C}^r(P)$ *is defined as the nerve of the set of balls* $B(p, r) = \{x \in M \mid d(p, x) \leqslant r\}$ *for all* $p \in P$*.*

The Čech complex has the nice property that by the nerve theorem, it is homotopy equivalent to the union of the balls $B(p,r)$. In particular, for nice radii, it will capture the underlying shape. Sadly, checking whether a large number of balls have a common intersection can be computationally expensive. Further, the definition requires that the data points are embedded in a metric space. These two issues motivate the next definition.

**Definition 3.11.** *Given a finite metric space* $(P, d)$ *and a real number radius* $r > 0$, *the* Vietoris-Rips complex $\mathbb{VR}^r(P)$ *is defined as the simplicial complex containing a simplex* $\sigma$ *if and only if* $d(p,q) \leqslant 2r$ *for every pair* $p, q \in \sigma$.

Clearly, for finite subsets of metric spaces, by definition, the Čech complex and the Vietoris-Rips complex for the same radius and the same point set have the same set of 1-simplices (the same 1-skeleton). While the Čech complex then contains additional information about the common intersections of balls, the Vietoris-Rips complex is simply the clique complex of this 1-skeleton. This makes the Vietoris-Rips complex easier to compute. Furthermore, we make the following simple observation, showing that the Vietoris-Rips complex still captures shapes in the data:

**Observation 3.12.** $\mathbb{C}^r(P) \subseteq \mathbb{VR}^r(P) \subseteq \mathbb{C}^{2r}(P)$.

**Exercise 3.13.** *Prove Observation 3.12.*

**Exercise 3.14.** *Find a point set* $P \subset \mathbb{R}^2$ *and a radius* $r$ *such that its Vietoris-Rips complex has non-trivial 2-homology, i.e., such that* $H_2(\mathbb{VR}^r(P)) \not\cong 0$.
*Furthermore, is there a dimension* $k$ *such that* $H_{k'}(\mathbb{VR}^r(Q)) = 0$ *for all* $k' \geqslant k$, *all* $r > 0$, *and all point sets* $Q \subset \mathbb{R}^2$*?*

### 3.4.2 Delaunay and Alpha complexes

Recall that computing persistent homology takes $O(N^3)$ time, where $N$ is the size of the simplicial complex in the filtration. For large enough radii, both the Čech and the Vietoris-Rips complex become complete, and thus contain $2^n$ simplices. Thus, computing persistent homology using those complexes is computationally very expensive, which is why in many applications we would like to have sparser complexes. For data in $\mathbb{R}^d$ we can look at the so-called Delaunay triangulation, which only has complexity $O(n^{\lceil d/2 \rceil})$.

**Definition 3.15.** *Given a finite point set* $P \subset \mathbb{R}^d$, *a* Delaunay simplex *is a geometric simplex whose vertices are in* $P$ *and lie on the boundary of a ball whose interior contains no points of* $P$.

*A* Delaunay triangulation $\mathrm{Del}(P)$ *of* $P$ *is a geometric simplicial complex where every simplex is a Delaunay simplex and whose underlying space covers the convex hull of* $P$.

*Given a finite point set* $P \subset \mathbb{R}^d$, *the* extended Delaunay complex *is the simplicial complex where for every face* $\sigma$, *for* $d' \leqslant d$, *every* $d'$*-face of* $\sigma$ *is a Delaunay simplex.*

It is a well-known fact that for a point set in general position (no $d+2$ points lie on a common sphere), there is a unique Delaunay triangulation. Furthermore, in this case the extended Delaunay complex and the unique Delaunay triangulation coincide.

**Definition 3.16.** *Given a finite point set* $P \subset \mathbb{R}^d$*, the* Voronoi diagram *is the tessellation of* $\mathbb{R}^d$ *into the* Voronoi cells

$$V_p = \{x \in \mathbb{R}^d \mid d(x,p) \leqslant d(x,q) \forall q \in P\}$$

*for all* $p \in P$*.*

**Fact 3.17.** *The nerve of the Voronoi cells of* $P$ *is the extended Delaunay complex of* $P$*.*

**Exercise 3.18.** *Convince yourself that for a point set in* $\mathbb{R}^2$*, the nerve of the Voronoi diagram is the extended Delaunay complex. Furthermore, convince yourself that if the points are in general position (there are no three points that are collinear, and no four points that are cocircular), then there is a unique Delaunay triangulation.*

Based on the Delaunay triangulation, we define the *Alpha complex* by parameterizing using a radius as follows:

**Definition 3.19.** *Given a finite point set* $P \subset \mathbb{R}^d$ *in general position as well as a real number radius* $r > 0$*, the* Alpha complex $\mathrm{Del}^r(P)$ *consists of all simplices* $\sigma \in \mathrm{Del}(P)$ *for which the circumscribing ball of* $\sigma$ *has radius at most* $r$*.*

The following fact provides us with an alternative definition of the Alpha complex:

**Fact 3.20.** *The Alpha complex* $\mathrm{Del}^r(P)$ *is the nerve of the sets* $B(p,r) \cap V_p$ *for all* $p \in P$*.*

Since the Alpha complex is a subset of the Delaunay triangulation (and for large enough radius is equal to the Delaunay triangulation), it also has complexity $O(n^{\lceil d/2 \rceil})$. Further, the above fact together with the Nerve theorem implies that the Alpha complex $\mathrm{Del}^r(P)$ is homotopy equivalent to the Čech complex $\mathbb{C}^r(P)$.

**Exercise 3.21.** *Is the following true or false? Consider a point set* $P \subset \mathbb{R}^2$ *in general position and a radius* $r > 0$*. Then the Alpha complex (with radius* $r$*) is the intersection of the Čech complex (with radius* $r$*) with the Delaunay triangulation.*

### 3.4.3 Subsample Complexes

For many applications, the Alpha complex is still too large. It is further expensive to compute, as computing a Delaunay triangulation in $\mathbb{R}^d$ takes $O(n^{\lceil d/2 \rceil})$ time. Sparser complexes can be constructed by looking at subsamples of the data, and relating the rest of the data to these subsamples. In the following, we will discuss two examples of complexes based on this idea.

**Definition 3.22.** *Given a finite point set* $Q$ *and a point set* $P \supset Q$ *in some metric space, we say that a simplex* $\sigma \subseteq Q$ *is* weakly witnessed *by* $x \in P \setminus Q$*, if* $d(q, x) \leqslant d(p, x)$ *for every* $q \in \sigma$ *and* $p \in Q \setminus \sigma$.

Note that the set of weakly witnessed simplices is not downwards closed. We thus define a simplicial complex by requiring that all faces are weakly witnessed:

**Definition 3.23.** *The* Witness complex $\mathbb{W}(Q, P)$ *is the collection of simplices on* $Q$ *for which all faces are weakly witnessed by some point* $p \in P \setminus Q$.

Note that if we take the metric space $\mathbb{R}^d$ and we let $P$ be the whole $\mathbb{R}^d$, then $\mathbb{W}(Q, P) = \mathrm{Del}(Q)$, and by definition we thus get in general that $\mathbb{W}(Q, P) \subseteq \mathrm{Del}(Q)$.

To arrive at a filtration, we again have to introduce a parameter $r > 0$:

**Definition 3.24.** *Given a finite point set* $Q$ *and a point set* $P \supset Q$ *in some metric space as well as a real number radius* $r > 0$*, the* parameterized *Witness complex* $\mathbb{W}^r(Q, P)$ *is defined as follows:*
*An edge* $pq$ *is in* $\mathbb{W}^r(Q, P)$ *if it is weakly witnessed by* $x \in P \setminus Q$ *and* $d(p, x) \leqslant r$ *and* $d(q, x) \leqslant r$*. A simplex* $\sigma$ *is in* $\mathbb{W}^r(Q, P)$ *if all its edges are.*

The idea of this complex is that it should approximate the Vietoris-Rips complex on $P$. There are theoretical guarantees about this approximation for manifolds of dimension at most 2, but the parameterized witness complex may fail to capture the topology of manifolds in dimension 3 and above.

Note that from the definition it is not guaranteed that the parameterized Witness complex is a subcomplex of the Witness complex.

**Definition 3.25.** *Given two finite point sets* $Q, P$ *in* $\mathbb{R}^d$*, as well as a graph* $G(P)$ *with vertices in* $P$*, we define* $\nu : P \to Q$ *by sending each point in* $P$ *to its closest point in* $Q$*. The* graph induced complex $\mathbb{G}(Q, G(P))$ *contains a simplex* $\sigma = \{q_0, \ldots, q_k\} \subset Q$ *if and only if there is a clique* $\{p_0, \ldots, p_k\}$ *in* $G(P)$ *for which* $\nu(p_i) = q_i$.

We again parameterize this:

**Definition 3.26.** *Let* $G^r(P)$ *be the graph on* $P$ *where* $pq$ *is an edge if and only if* $d(p, q) \leqslant 2r$*. The* parameterized graph induced complex $\mathbb{G}^r(Q, P)$ *is defined as* $\mathbb{G}(Q, G^r(P))$.

This complex again has theoretical guarantees of approximating the Vietoris-Rips complex on $P \cup Q$.

**Exercise 3.27.** *Let* $P, Q$ *be point sets and* $G(P)$ *a graph with* $P$ *as its vertex set. Let* $\nu : P \to Q$ *be the map sending each point of* $P$ *to its closest point of* $Q$ *(assume that this closest point is always unique). Let* $C$ *be the clique complex of* $G(P)$ *(the complex which includes a simplex iff its corresponding vertices in* $G(P)$ *form a clique).*

*Show that* $\nu$ *extends to a simplicial map* $\bar{\nu} : C \to \mathbb{G}(Q, G(P))$*. Also show that any simplicial complex* $K$ *with* $V(K) = Q$ *for which* $\nu$ *has a simplicial extension must contain* $\mathbb{G}(Q, G(P))$.

## 3.5 Distance Metrics on Persistence Diagrams

In this section we will define some distance metrics that can be used to compare different persistence diagrams.

### 3.5.1 Bottleneck Distance

Let $\mathcal{F}, \mathcal{G}$ be two filtrations giving rise to persistence modules $H_p\mathcal{F}, H_p\mathcal{G}$. Let $\mathrm{Dgm}_p(\mathcal{F})$ and $\mathrm{Dgm}_p(\mathcal{G})$ be their corresponding persistence diagrams. These diagrams are the information we want to use to compare $\mathcal{F}$ and $\mathcal{G}$.

The general idea of the bottleneck distance is to pair up points of the two persistence diagrams, i.e., consider bijections between points of $\mathrm{Dgm}_p(\mathcal{F})$ and $\mathrm{Dgm}_p(\mathcal{G})$. Since we can only find bijections between sets of the same cardinality, we need the two diagrams to have the same number of points. This is where the definition of the persistence diagram comes in: recall that a persistence diagram includes every possible point on the diagonal with infinite multiplicity. Thus, both sets of points have the same (infinite) cardinality, and bijections between these sets are thus well-defined.

To measure the "quality" or "distance" of such a bijection, we use the $L_\infty$-norm:

**Definition 3.28.** *Let* $x = (x_1, x_2), y = (y_1, y_2)$ *be two points in* $\mathbb{R}^2$. *Then,*

$$||x - y||_\infty := \max(|x_1 - y_1|, |x_2 - y_2|),$$

*where we say that* $\infty - \infty = 0$ *for points with coordinates that are* $\infty$ *(i.e., points in persistence diagrams that correspond to holes that did not die).*

**Definition 3.29.** *Let* $\Pi = \{\pi : \mathrm{Dgm}_p(\mathcal{F}) \to \mathrm{Dgm}_p(\mathcal{G}) \mid \pi$ *is bijective*$\}$ *be the set of all bijections between* $\mathrm{Dgm}_p(\mathcal{F})$ *and* $\mathrm{Dgm}_p(\mathcal{G})$. *Then, the* Bottleneck distance *is defined as*

$$d_b(\mathrm{Dgm}_p(\mathcal{F}), \mathrm{Dgm}_p(\mathcal{G})) := \inf_{\pi \in \Pi} \sup_{x \in \mathrm{Dgm}_p(\mathcal{F})} ||x - \pi(x)||_\infty.$$

The Bottleneck distance thus minimizes the maximum $L_\infty$-norm of any pairing, over all pairings of points.

**Observation 3.30.** *The Bottleneck distance is a metric on the space of persistence diagrams with finitely many off-diagonal points.*

*Proof.* We check the three properties of metrics:

1. $d_b(X, Y) = 0$ if and only if $X = Y$ is simple to see, since if $X = Y$, every point can be matched to its copy, and if $X \neq Y$, there exists some point $p \in X \setminus Y \cup Y \setminus X$ which must be matched to some point with positive $L_\infty$-distance to $p$.

2. $d_b(X, Y) = d_b(Y, X)$ is clear by definition.

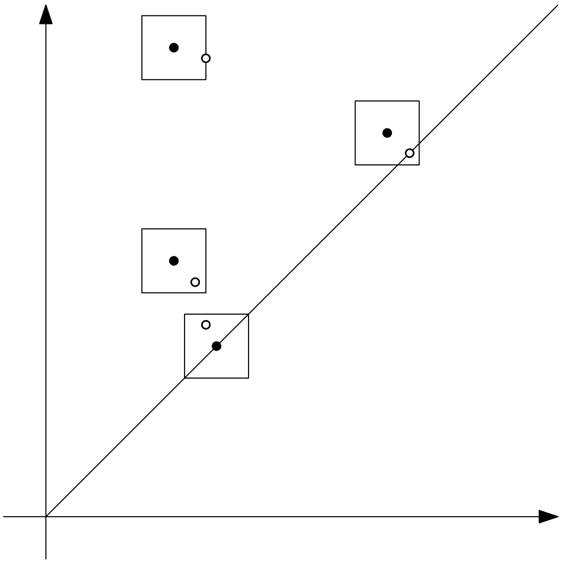


**Figure 3.4**: *An illustration of the idea of bottleneck distance.*

3. $d_b(X, Y) \leqslant d_b(X, Z) + d_b(Z, Y)$. Take a bijection $\pi_1$ witnessing $d_b(X, Z)$ and a bijection $\pi_2$ witnessing $d_b(Z, Y)$, and concatenate the two: $\pi := \pi_2 \circ \pi_1$ is a bijection $X \to Y$ where for every $x \in X$ we have $\|x - \pi(x)\|_\infty \leqslant \|x - \pi_1(x)\|_\infty + \|\pi_1(x) - \pi_2(x)\|_\infty$. Note that since $d_b$ is an infimum and not a minimum, there may not be $\pi_1$ and $\pi_2$ witnessing $d_b$. In this case, the same argument can be applied to the converging sequences of bijections witnessing $d_b$. □

**Exercise 3.31**. *Give an algorithm to compute the Bottleneck distance between two persistence diagrams. Your algorithm should be polynomial in* $n$*, where* $n$ *is the total number of off-diagonal points in the two persistence diagrams.*

Recall that simplex-wise monotone functions $f, g : K \to \mathbb{R}$ give rise to simplicial sublevel set filtrations $\mathcal{F}_f, \mathcal{F}_g$. We could now compare the persistence diagrams of these two filtrations using the Bottleneck distance, but we wish to define a metric directly between the two functions $f, g$:

**Definition 3.32** (infinity norm). *Let* $f, g : X \to \mathbb{R}$*. Then, the infinity norm of* $f - g$ *is defined as*

$$\|f - g\|_\infty := \sup_{x \in X} |f(x) - g(x)|.$$

The following theorem tells us that this infinity norm and the Bottleneck distance are closely related:

**Theorem 3.33** (Stability for simplicial filtrations). *Let* $f, g : K \to \mathbb{R}$ *be simplex-wise monotone functions. Then,* $\forall p \geqslant 0$ *we have* $d_b(\mathrm{Dgm}_p(\mathcal{F}_f), \mathrm{Dgm}_p(\mathcal{F}_g)) \leqslant \|f - g\|_\infty$.

*Proof.* Let $f_t := (1-t)f + tg$ for $t \in [0,1]$ be the linear interpolation between f and g. Note that $f_0 = f, f_1 = g$.

We first show that each $f_t$ is a simplex-wise monotone function. It is clearly simplex-wise, and we prove that it is also monotone: Let $\sigma \subseteq \tau$. Since f and g are monotone, we have $f(\sigma) \leqslant f(\tau)$ and $g(\sigma) \leqslant g(\tau)$. Thus,

$$f_t(\sigma) = (1-t)f(\sigma) + tg(\sigma) \leqslant (1-t)f(\tau) + tg(\tau) = f_t(\tau).$$

Let $p \geqslant 0$ be fixed. We now draw the family of persistence diagrams $\mathrm{Dgm}_p(\mathcal{F}_{f_t})$ as a multiset in $\mathbb{R}^2 \times [0,1]$. Each off-diagonal point of $X_t := \mathrm{Dgm}_p(\mathcal{F}_{f_t})$ is of the form $x(t) = (f_t(\sigma), f_t(\tau), t)$ for $\sigma$ being the creator and $\tau$ being the destructor. Note that the persistence pairings $(\sigma, \tau)$ may only change when the order of simplex insertion changes, which only happens finitely many times when going from $t = 0$ to $t = 1$. Let us call these values $0 = t_0 < t_1 < t_2 < \ldots < t_n < t_{n+1} = 1$. Without loss of generality, we assume that at each of these values $t_i$ exactly two simplices have the same value $f_{t_i}$.

Within each open interval $(t_i, t_{i+1})$ the pairings stay constant. Furthermore, every off-diagonal point $x(t)$ is a linear function of t in all three coordinates, meaning that it defines a line segment.

At $t_{i+1}$, if $x(t_{i+1})$ is an off-diagonal point whose creator and destructor are still paired after $t_{i+1}$, $x(t)$ continues in the same direction after $t_{i+1}$.

If on the other hand $x(t_{i+1})$ is an off-diagonal point whose creator and destructor get paired differently, recall by Exercise Sheet 5, Question 3, there are exactly two pairs that swap their creators or destructors, and these creators or destructors that are swapped must have the same value in $f_{t_{i+1}}$. In the persistence diagram, this means that two points vertically *or* horizontally of each other swap creators/destructors, and there is a unique continuing line segment for both of them.

Lastly, if $x(t_{i+1})$ is on the diagonal, this means that its previous creator and destructor now have the same value in $f_{t_{i+1}}$. There is no continuation for this point.

Every point thus moves along a polygonal path monotone in t. Every such path is called a *vine*, and the multiset of all vines is called a *vineyard*, see Figure 3.5 for an illustration. Based on this vineyard, we now wish to find a good matching giving an upper bound on the Bottleneck distance. We simply take the matching where we match the start point of every vine with its endpoint. To get a bound on the Bottleneck distance, we simply need to get a bound for the distance of each matched pair.

Between $t_i$ and $t_{i+1}$ we get for $\frac{\delta x(t)}{\delta t}$:

$$\frac{\delta}{\delta t}((1-t)(f(\sigma), f(\tau), t)) + t(g(\sigma), g(\tau), t)) = (g(\sigma) - f(\sigma), g(\tau) - f(\tau), 1)$$

Projecting $x(t_{i+1})$ and $x(t_i)$ to $\mathbb{R}^2$ we get two points $y_{i+1}, y_i$ such that

$$||y_{i+1} - y_i||_\infty = (t_{i+1} - t_i) \cdot \max(g(\sigma) - f(\sigma), g(\tau) - f(\tau) \leqslant (t_{i+1} - t_i) \cdot ||f - g||_\infty$$

Thus, since $||\cdot||_\infty$ is a norm and fulfills the triangle inequality, we also have that from $t = 0$ to $t = 1$, the point can move at most $||f - g||_\infty$. We thus have the desired bound on the Bottleneck distance. ☐

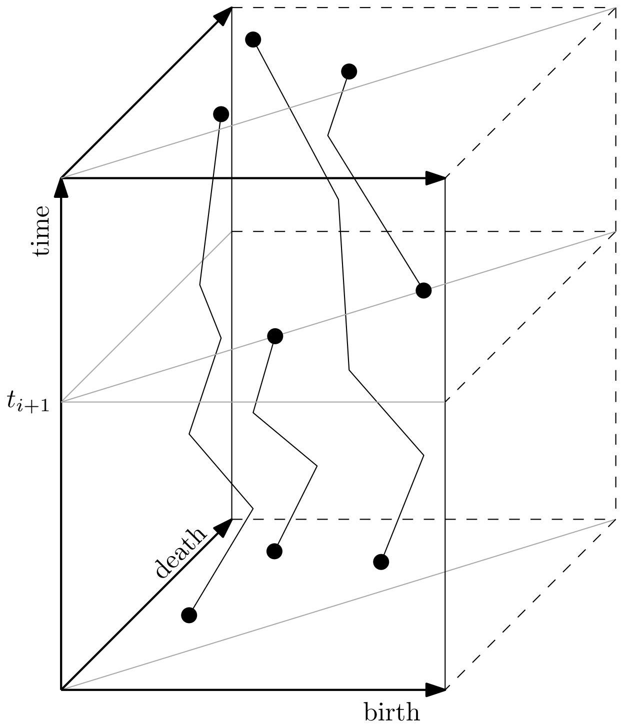


**Figure 3.5**: *The vineyards in the proof of Theorem 3.33.*

**Exercise 3.34**. *Show that Theorem 3.33 (Stability for simplicial filtrations) can be tight for all* $p \geqslant 0$ *and all values of* $\|f - g\|_\infty$.

We wish to generalize the stability result above to general topological spaces.

Consider some topological space X and a function $f : X \to \mathbb{R}$, which induces a sublevel set filtration for every $r \in \mathbb{R}$. We only want to consider *tame* functions: A function f is *tame* if all homology groups of sublevel sets have finite rank, and the homology groups only change at finitely many values, called *critical values*.

**Theorem 3.35**. *Let* X *be a triangulable topological space, and* $f, g : X \to \mathbb{R}$ *be two tame functions, then* $\forall p \geqslant 0$*, we have*

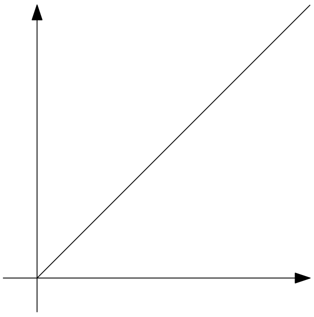
$$d_b(\mathrm{Dgm}_p(\mathcal{F}_f), \mathrm{Dgm}_p(\mathcal{F}_g)) \leqslant \|f - g\|_\infty.$$

We do not prove this theorem at this point, but with additional tools that we will develop in Section 3.6, the proof of this (and of Theorem 3.33) will follow quite easily.

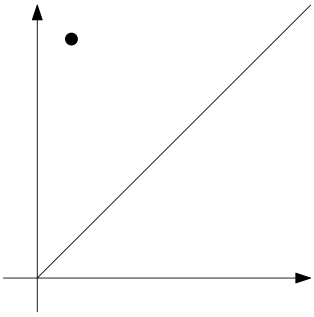
### 3.5.2 Wasserstein Distance

Consider the following three diagrams:

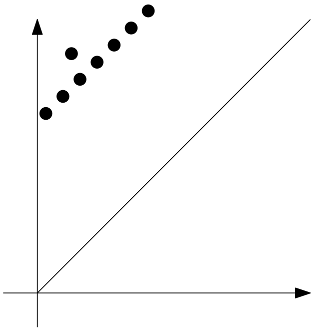
Which of $Y_1$ and $Y_2$ is X closer to? Intuitively, one clearly says $Y_1$: There are simply fewer features in $Y_1$ that are not present in X. In terms of Bottleneck distance, there is



$X$



$Y_1$



$Y_2$

only one reasonable matching between $X$ and $Y_1$, and also only one between $X$ and $Y_2$: We simply match each off-diagonal point with its closest point on the diagonal. Since we only look at the longest edge in this matching, the Bottleneck distance $d_b(X, Y_1) = d_b(X, Y_2)$.

We can get rid of this counter-intuitive behavior of the Bottleneck distance by using the Wasserstein distance.

**Definition 3.36** (Wasserstein distance). *For* $p \geqslant 0$*, and* $q \geqslant 1$*, the* $q$*-Wasserstein distance is defined as*

$$d_{W,q}(\mathrm{Dgm}_p(\mathcal{F}), \mathrm{Dgm}_p(\mathcal{G})) := \left[ \inf_{\pi \in \Pi} \Bigg( \sum_{x \in \mathrm{Dgm}_p(\mathcal{F})} (\|x - \pi(x)\|_\infty)^q \Bigg) \right]^{1/q}$$

Intuitively, we now consider the length of all edges in the matching induced by the bijection, as opposed to just the longest one, but the longer ones get more weight. Note that for $q = \infty$, we retrieve the bottleneck distance, that is, $d_{W,\infty} = d_b$.

We can see that the stability theorem we proved for Bottleneck distance does not hold for Wasserstein distance: consider two simplex-wise monotone functions $f$ and $g$ on a path, as illustrated in Figure 3.6. In both $f$ and $g$ the first vertex on the path is mapped to $1$ and the edges along the path are mapped to increasing odd numbers. In $f$ the remaining vertices along the path get mapped to increasing even numbers, and in $g$ to increasing odd numbers. In particular, $\|f - g\|_\infty = 1$. In the filtration defined by $f$, at every even step we add a vertex, creating a new connected component, which gets connected to the rest of the path at the next step. Thus, each vertex of the path will give an off-diagonal point in the 0-persistence diagram, where all of them except the first one have a lifespan of 1. On the other hand, in the filtration defined by $g$, we always add the new vertices and their connecting edge in the same step, thus the 0-persistence diagram only has a single off-diagonal point with infinite lifespan. In particular, we have that for arbitrarily long paths we get arbitrarily large Wasserstein distances between the diagrams for all $q < \infty$.

A similar counterexample can also be found for topological spaces. Consider the topological space $[0, 1]$ and the two functions depicted by the curves in Figure 3.7. Here

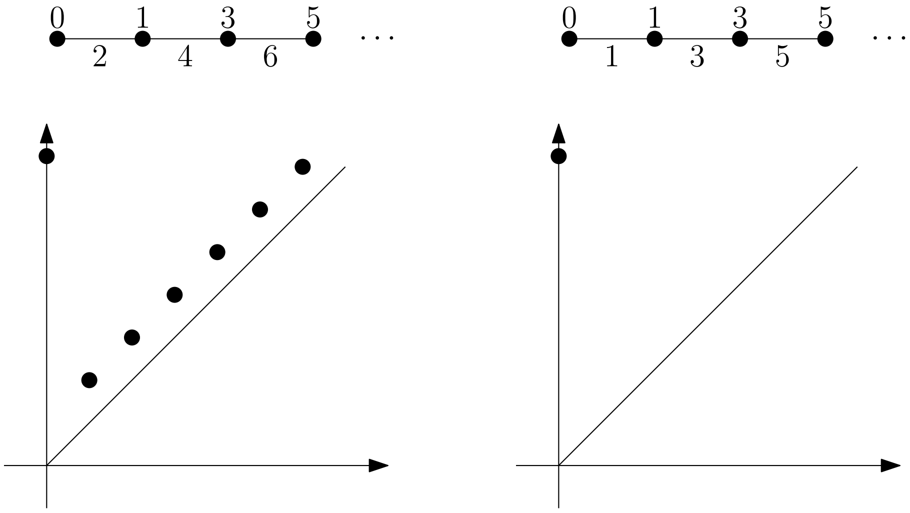


**Figure 3.6:** *Two simplex-wise monotone functions with bounded infinity norm whose persistence diagrams have unbounded Wasserstein distance.*

we again have that $\|f-g\|_\infty \leqslant \epsilon$, but the Wasserstein distance between the two diagrams can be made arbitrarily big.

To avoid these types of counterexamples, we only want to consider even nicer functions:

**Definition 3.37** (Lipschitz). *Let* $(X, d)$ *be a metric space. A function* $f : X \to \mathbb{R}$ *is* Lipschitz *if there exists a constant* $C$ *such that* $|f(x) - f(y)| \leqslant c \cdot d(x, y)$ *for all* $x, y \in X$.

For these functions we again get stability theorems, that we will not prove here.

**Theorem 3.38.** *Let* $X$ *be a triangulable, compact metric space. Let* $f, g : X \to \mathbb{R}$ *be Lipschitz functions. Then there exist constants* $C$ *and* $k$ *(that may only depend on* $X$ *and on the Lipschitz constants of* $f, g$*) such that for every* $p \geqslant 0$ *and every* $q \geqslant k$,

$$d_{W,q}(\mathrm{Dgm}_p(\mathcal{F}_f), \mathrm{Dgm}_p(\mathcal{F}_g)) \leqslant C \cdot \|f-g\|_\infty^{1-k/q}.$$

**Theorem 3.39.** *Let* $f, g : K \to \mathbb{R}$ *be simplex-wise monotone functions. Then for all* $p \geqslant 0$ *and all* $q \geqslant 1$,

$$d_{W,q}(\mathrm{Dgm}_p(\mathcal{F}_f), \mathrm{Dgm}_p(\mathcal{F}_g)) \leqslant \|f-g\|_q = \Big(\sum_{\sigma \in K} |f(\sigma) - g(\sigma)|^q\Big)^{1/q}.$$

## 3.6 Interleaving of persistence modules

### 3.6.1 Interleaving distance

Until now, we compared persistence diagrams. We will now introduce the interleaving distance, which instead compares persistence modules. Let us begin with a formal

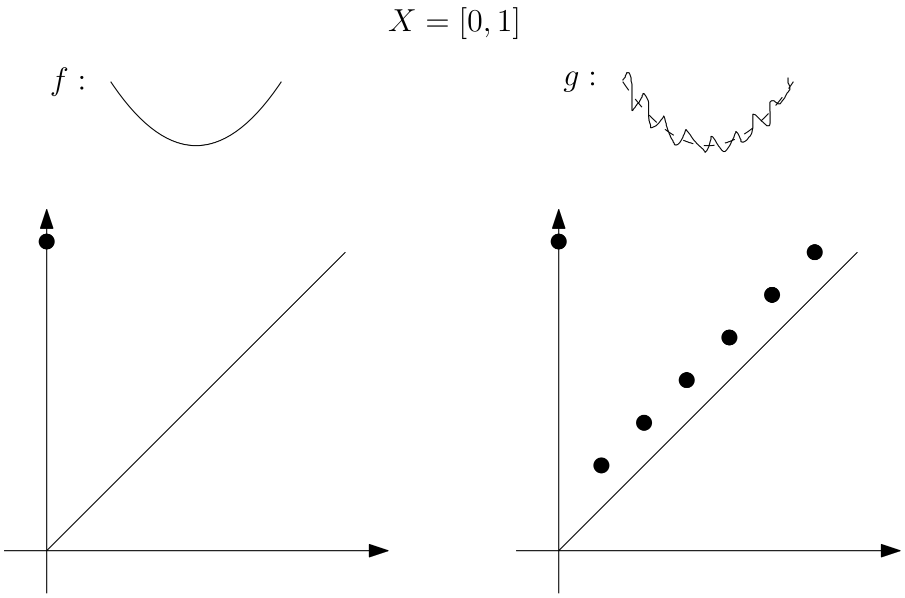


**Figure 3.7:** *Two functions* $[0,1] \to \mathbb{R}$ *with bounded infinity norm whose persistence diagrams have unbounded Wasserstein distance.*

definition of persistence modules.

**Definition 3.40.** *A* persistence module $\mathbb{V}$ *over* $\mathbb{R}$ *is a collection* $\mathbb{V} = \{V_a\}_{a\in\mathbb{R}}$ *of vector spaces* $V_a$ *together with linear maps* $v_{a,a'} : V_a \to V_{a'}$ *such that* $v_{a,a} = \mathrm{id}$ *and* $v_{b,c} \circ v_{a,b} = v_{a,c}$ *for all* $a \leqslant b \leqslant c$.

You already know a few examples of persistence modules, e.g., the persistent homology of sublevel set filtrations or of Čech or Vietoris-Rips complexes (here one simply defines $V_a = 0$ for $a < 0$).

We again want to consider maps between persistence modules, starting with a notion of isomorphism, telling us when two persistence modules are "the same".

**Definition 3.41.** *We say that two persistence modules* $\mathbb{U}$ *and* $\mathbb{V}$ *are* isomorphic *if there are isomorphisms* $f_a : U_a \to V_a$ *such that*

$$\begin{tikzcd}
U_a \arrow[r, "u_{a,a'}"] \arrow[d, leftrightarrow, "f_a"] & U_{a'} \arrow[d, leftrightarrow, "f_{a'}"] \\
V_a \arrow[r, "v_{a,a'}"] & V_{a'}
\end{tikzcd}$$

*commutes both ways, i.e.,* $f_{a'} \circ u_{a,a'} = v_{a,a'} \circ f_a$, *and* $u_{a,a'} \circ f_a^{-1} = f_{a'}^{-1} \circ v_{a,a'}$.

The basic idea of interleaving distance is to measure how close two persistence modules are to being isomorphic. For this, we allow ourselves some slack, in the sense that $U_a$ does not need to map to $V_a$, but it can map to $V_{a+\epsilon}$, as long as all the relevant maps still behave like they would for an isomorphism. We make this formal in the next definition.

**Definition 3.42** ($\epsilon$-interleaving persistence modules). *Let $\mathbb{U}$ and $\mathbb{V}$ be persistence modules over $\mathbb{R}$. We say that $\mathbb{U}$ and $\mathbb{V}$ are* $\epsilon$-interleaved *if there exist two families of maps, $\varphi_a : U_a \to V_{a+\epsilon}$ and $\psi_a : V_a \to U_{a+\epsilon}$ such that the following four diagrams are commutative:*

$$\begin{tikzcd}
U_a \arrow[rr, "u_{a,a'}"] \arrow[dr, "\varphi_a"] & & U_{a'} \arrow[dr, "\varphi_{a'}"] & \\
& V_{a+\epsilon} \arrow[rr, "v_{a+\epsilon,a'+\epsilon}"] & & V_{a'+\epsilon}
\end{tikzcd}$$

*and*

$$\begin{tikzcd}
& U_{a+\epsilon} \arrow[rr, "u_{a+\epsilon,a'+\epsilon}"] & & U_{a'+\epsilon} \\
V_a \arrow[ur, "\psi_a"] \arrow[rr, "v_{a,a'}"] & & V_{a'} \arrow[ur, "\psi_{a'}"] &
\end{tikzcd}$$

$$\begin{tikzcd}
U_a \arrow[rr, "u_{a,a+2\epsilon}"] \arrow[dr, "\varphi_a"] & & U_{a+2\epsilon} \\
& V_{a+\epsilon} \arrow[ur, "\psi_{a+\epsilon}"] &
\end{tikzcd}$$

*and*

$$\begin{tikzcd}
& U_{a+\epsilon} \arrow[dr, "\varphi_{a+\epsilon}"] & \\
V_a \arrow[ur, "\psi_a"] \arrow[rr, "v_{a,a+2\epsilon}"] & & V_{a+2\epsilon}
\end{tikzcd}$$

Note that if $\mathbb{U}$ and $\mathbb{V}$ are isomorphic, then they are 0-interleaved: the first type of diagrams (the *square diagrams*) are the commutative diagrams in the definition of isomorphic persistence modules and the the second type of diagrams (the *triangular diagrams*) collapse to two arrows that say that the maps $\varphi_a$ are isomorphisms with inverses $\psi_a$.

**Theorem 3.43.** *Assume $\mathbb{U}$ and $\mathbb{V}$ are $\epsilon$-interleaving. Let $\delta > \epsilon$. Then $\mathbb{U}$ and $\mathbb{V}$ are also $\delta$-interleaving.*

*Proof.* Given $\varphi'_a : U_a \to V_{a+\epsilon}$ we define $\varphi_a : U_a \to V_{a+\delta}$ simply as $\varphi_a := v_{a+\epsilon,a+\delta} \circ \varphi'_a$. Symmetrically, we define $\psi_a := u_{a+\epsilon,a+\delta} \circ \psi'_a$. To check that the correct diagrams commute, we only check the right of every pair of symmetric ones above. We have to distinguish two cases for the first diagram, $a + \delta < a' + \epsilon$ and $a + \delta > a' + \epsilon$.

For the first case, we get the following diagram:

$$\begin{tikzcd}
U_a \arrow[rrr] \arrow[dr] \arrow[drr, dashed] & & & U_{a'} \arrow[dr] \arrow[drr, dashed] & & \\
& V_{a+\epsilon} \arrow[r] & V_{a+\delta} \arrow[rr] & & V_{a'+\epsilon} \arrow[r] & V_{a'+\delta}
\end{tikzcd}$$

For the second case we get the diagram:

$$\begin{tikzcd}
U_a \arrow[rr] \arrow[dr] \arrow[drrrr, dashed] & & U_{a'} \arrow[dr] \arrow[drrr, dashed] & & & \\
& V_{a+\epsilon} \arrow[rr] & & V_{a'+\epsilon} \arrow[r] & V_{a+\delta} \arrow[r] & V_{a'+\delta}
\end{tikzcd}$$

And finally, for the triangular diagram we get:

$$\begin{tikzcd}
U_a \arrow[rr] \arrow[dr] \arrow[drr, dashed] & & U_{a+2\epsilon} \arrow[r] & U_{a+\delta+\epsilon} \arrow[r] & U_{a+2\delta} \\
& V_{a+\epsilon} \arrow[r] \arrow[ur] & V_{a+\delta} \arrow[ur] \arrow[urr, dashed] & &
\end{tikzcd}$$

One can now verify that in all of these diagrams the correct paths commute. □

Thus, the following definition makes sense:

**Definition 3.44** (Interleaving distance). $d_I(\mathbb{U}, \mathbb{V}) := \inf\{\epsilon \mid \mathbb{U}$ *and* $\mathbb{V}$ *are* $\epsilon$*-interleaved* $\}$.

**Exercise 3.45.** *Show that interleaving distance is a pseudo-metric for persistence modules (up to isomorphism), i.e., prove that (i) the interleaving distance between isomorphic persistence modules is* $0$*, (ii) the interleaving distance is non-negative, and (iii) the interleaving distance fulfills the triangle inequality.*

*Also show that it is* not *a metric by showing that there exist non-isomorphic persistence modules with interleaving distance* $0$*.*

**Exercise 3.46.** *Let* $W_1$ *and* $W_2$ *be two arbitrary vector spaces. Let* $\mathbb{U}$ *be the persistence module such that* $U_a = W_1$ *for* $a \in [w, x)$*, and* $U_a = 0$*, otherwise. For* $a, a' \in [w, x)$ *we have* $u_{a,a'}$ *being the identity map. For* $a < w$ *or* $a' \geqslant x$ *(or both), we have* $u_{a,a'}$ *being the zero map. Similarly, we define the persistence module* $\mathbb{V}$ *which is* $W_2$ *in* $a \in [y, z)$ *and* $0$ *otherwise.*

*Show that* $d_I(\mathbb{U}, \mathbb{V}) \leqslant \max(\frac{w-x}{2}, \frac{z-y}{2})$.

The underlying ideas that allowed us to define the interleaving distance of persistence modules can also be applied to filtrations.

**Definition 3.47** (Interleaving for Filtrations). *Let* $\mathcal{F}, \mathcal{G}$ *be filtrations over* $\mathbb{R}$*.* $\mathcal{F}$ *and* $\mathcal{G}$ *are* $\epsilon$*-interleaved if there exist maps* $\varphi_a : F_a \to G_{a+\epsilon}$ *and* $\psi_a : G_a \to F_{a+\epsilon}$ *such that the same type of diagrams commute up to homotopy, that is, for example* $\varphi_{a'} \circ \iota^F_{a,a'} \simeq \iota^G_{a+\epsilon,a'+\epsilon} \circ \varphi_a$ *are homotopic (contiguous).*

*We again define the interleaving distance (now between filtrations):*

$$d_I(\mathcal{F}, \mathcal{G}) = \inf\{\epsilon \mid \mathcal{F} \textit{ and } \mathcal{G} \textit{ are } \epsilon\textit{-interleaved} \}.$$

**Observation 3.48.** *For all* $p \geqslant 0$*,* $d_I(H_p\mathcal{F}, H_p\mathcal{G}) \leqslant d_I(\mathcal{F}, \mathcal{G})$.

The proof follows immediately from induced homology.

As a first application of interleaving distance, we can quantify how different the Čech and Vietoris-Rips filtrations are. Recall that for a point cloud $P$ and a radius $r$, we have the relationship between the Čech and Vietoris-Rips complexes as follows: $\mathbb{C}^r(P) \subseteq \mathbb{VR}^r(P) \subseteq \mathbb{C}^{2r}(P)$. Since this factor 2 is multiplicative, and we need an additive $\epsilon$ for interleaving, let us just take the logarithmic scale (base 2) for the radius, i.e., we define $\mathbb{C}^r_{\log} = \mathbb{C}^{2^r}$ and similarly $\mathbb{V}\mathbb{R}^r_{\log} = \mathbb{V}\mathbb{R}^{2^r}_{\log}$. Since $2^{(r+1)} = 2r$, we have $\mathbb{C}^r_{\log}(P) \subseteq \mathbb{V}\mathbb{R}^r_{\log}(P) \subseteq \mathbb{C}^{r+1}_{\log}(P)$.

We thus have the following inclusions:

$$\begin{tikzcd}
\mathbb{C}^r_{\log} \arrow[r] \arrow[dr] & \mathbb{C}^{r+1}_{\log} \arrow[r] \arrow[dr] & \mathbb{C}^{r+2}_{\log} \\
\mathbb{V}\mathbb{R}^r_{\log} \arrow[r] \arrow[ur] & \mathbb{V}\mathbb{R}^{r+1}_{\log} \arrow[r] \arrow[ur] & \mathbb{V}\mathbb{R}^{r+2}_{\log}
\end{tikzcd}$$

Since these are all inclusions, all relevant diagrams must commute, and thus we get that $d_I(\mathbb{C}_{\log}, \mathbb{V}\mathbb{R}_{\log}) \leqslant 1$.

### 3.6.2 Stability with respect to interleaving distance

The main motivation for interleaving distance is, that it can be used to prove stability result, at least under some tameness conditions.

**Definition 3.49.** *A persistence module* $\mathbb{V}$ *is* q-tame *if the linear maps have finite rank.*

Note that in this definition, the q is not a parameter, just a name. All persistence modules that show up in the context of persistent homology on point clouds are q-tame, so this condition is not restrictive.

**Theorem 3.50.** *If* $\mathbb{U}, \mathbb{V}$ *are q-tame persistence modules over* $\mathbb{R}$*, then*

$$d_b(\mathrm{Dgm}\mathbb{U}, \mathrm{Dgm}\mathbb{V}) = d_I(\mathbb{U}, \mathbb{V}).$$

Thus, for every interleaving one can find between two persistence modules or between filtrations, one immediately gets a bound on the Bottleneck distance. This is a very powerful result, and the proof of this is out of scope for these lecture notes. One direction of the proof however follows from a decomposition result of persistence modules, that we will discuss in Section 3.7. But first, we will look at some examples, how we can use Theorem 3.50 to prove stability theorems.

**Exercise 3.51.** *Prove Theorem 3.35.*

### 3.6.3 Stability for Čech Complexes

So far, we have only seen stability results comparing filtrations induced by different functions on a fixed space. However, in applications in data analysis, we consider complexes on point clouds, and two different point clouds might not have the same size, and thus the simplicial complexes on which we get filtrations are generally different. Using interleaving distance, we can however still give stability results. In this section, we will do this for Čech complexes.

Consider two point clouds $P, Q$ in the same metric space $X$. Let us first consider the really simple case, where $P = \{p\}$, and $Q = \{q\}$ with $d(p, q) = d$. Then, $B(p, r) \subseteq B(q, r + d)$. Now, how does this generalize to larger point sets? To get the same kind of behavior, we need that for every point in $P$, there exists some point in $Q$ with distance at most $d$. This motivates the following distance measure:

**Definition 3.52** (Hausdorff distance). *Let $A, B \subseteq X$ be compact sets. Then the Hausdorff distance between $A$ and $B$ is defined as*

$$d_H(A, B) := \max\{\max_{a\in A} d(a, B), \max_{b\in B} d(b, A)\}.$$

**Exercise 3.53.** *Show that Hausdorff distance is a distance measure.*

Let $d_H(P, Q) = d$. Then, $\bigcup_{p\in P} B(p, r) \subseteq \bigcup_{q\in Q} B(q, r + d)$. From this, we get the following lemma:

**Lemma 3.54.** *The (filtration given by) the Čech complexes of $P$ and $Q$ are $d$-interleaved.*

*Proof.*

$$\begin{tikzcd}
\mathbb{C}^r(P) \arrow[r] \arrow[d, leftrightarrow, "\simeq"] & \mathbb{C}^{r+d}(P) \arrow[r] \arrow[d, leftrightarrow, "\simeq"] & \mathbb{C}^{r+2d}(P) \arrow[d, leftrightarrow, "\simeq"] \\
\bigcup_{p\in P} B(p, r) \arrow[dr] & \bigcup_{p\in P} B(p, r+d) \arrow[dr] & \bigcup_{p\in P} B(p, r+2d) \\
\bigcup_{q\in Q} B(q, r) \arrow[ur] \arrow[d, leftrightarrow, "\simeq"'] & \bigcup_{q\in Q} B(q, r+d) \arrow[ur] \arrow[d, leftrightarrow, "\simeq"'] & \bigcup_{q\in Q} B(q, r+2d) \arrow[d, leftrightarrow, "\simeq"'] \\
\mathbb{C}^r(Q) \arrow[r] & \mathbb{C}^{r+d}(Q) \arrow[r] & \mathbb{C}^{r+2d}(Q)
\end{tikzcd}$$

The relevant diagrams commute up to homotopy, since we only chain together homotopies and inclusion maps. □

We can conclude the following

**Theorem 3.55.** $d_b(\mathrm{Dgm}_p(\mathbb{C}(P)), \mathrm{Dgm}_p(\mathbb{C}(Q))) \leqslant d_H(P, Q)$ *for all* $p \geqslant 0$.

*Proof.* By Theorem 3.50, Observation 3.48, and finally Lemma 3.54, we have

$$d_b(\ldots) = d_I(H_p\mathbb{C}(P), H_p\mathbb{C}(Q)) \leqslant d_I(\mathbb{C}(P), \mathbb{C}(Q)) \leqslant d_H(P, Q).$$

□

## 3.7 Interval decomposition of Persistence Modules

In this section, we again look at persistence modules, this time as algebraic structures. We consider persistence modules over $\mathbb{R}$ of vector spaces over some field $\mathbb{F}$. We start by looking at some special persistence modules, called *interval modules*.

**Definition 3.56.** *A interval module* $\mathbb{I}[b, d]$ *is an persistence module*

$$V_a = \begin{cases} \mathbb{F} & \text{if } a \in [b, d], \\ 0 & \text{otherwise.} \end{cases} \qquad \text{and} \qquad v_{a,a'} = \begin{cases} \mathrm{id} & b \leqslant a \leqslant a' \leqslant d, \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, we can define interval modules on open and clopen intervals, denoted by $\mathbb{I}(b, d)$, $\mathbb{I}(b, d]$, and $\mathbb{I}[b, d)$. We write $\mathbb{I}\langle b, d\rangle$ to include all four of these types.

For an interval module we can easily talk about birth and death as we did in persistent homology. If we have a persistent homology module that is (isomorphic to) an interval module, the birth and death correspond to the boundaries $b, d$ of the interval.

**Definition 3.57.** *A persistence module* $\mathbb{U}$ *is called* pointwise finite dimensional (p.f.d.) *if for all* $a \in \mathbb{R}$, $U_a$ *has finite dimension.*

Note that all p.f.d. persistence modules are also q-tame.

**Definition 3.58.** *Given two persistence modules* $\mathbb{U}, \mathbb{V}$, *we define their direct sum* $\mathbb{U} \oplus \mathbb{V}$ *by* $(U \oplus V)_a = U_a \oplus V_a$ *and* $(u \oplus v)_{a,a'} = u_{a,a'} \oplus v_{a,a'}$.

Here, the direct sum of maps just means applying the respective maps componentwise.

**Proposition 3.59.** *If* $\mathbb{U}_1, \mathbb{U}_2$ *are* $\epsilon$*-interleaved, and* $\mathbb{V}_1, \mathbb{V}_2$ *are* $\delta$*-interleaved, then* $\mathbb{U}_1 \oplus \mathbb{V}_1$ *and* $\mathbb{U}_2 \oplus \mathbb{V}_2$ *are* $\max\{\epsilon, \delta\}$*-interleaved.*

*Proof.* Without loss of generality, let $\epsilon \geqslant \delta$, so we need to show that they are $\epsilon$-interleaved. Recall that if two persistence modules are $\delta$-interleaved, they are also $\epsilon$-interleaved. Let $\varphi^u, \psi^u$ be (series of) functions showing that $\mathbb{U}_1, \mathbb{U}_2$ are $\epsilon$-interleaved. Similarly, let $\varphi^v, \psi^v$ be (series of) functions showing that $\mathbb{V}_1, \mathbb{V}_2$ are $\epsilon$-interleaved. Then, $\varphi^u \oplus \varphi^v, \psi^u \oplus \psi^v$ show that $\mathbb{U}_1 \oplus \mathbb{V}_1$ and $\mathbb{U}_2 \oplus \mathbb{V}_2$ are $\epsilon$-interleaved. □

If we now have a direct sum of interval modules, we can still nicely talk about birth and death: we just look at each interval module in isolation. The following theorem shows that surprisingly most persistence modules can be expressed as direct sums of interval modules.

**Theorem 3.60** (Structure theorem). *Any p.f.d. persistence module decomposes uniquely into interval modules, i.e., we have*

$$\mathbb{U} \cong \bigoplus_{i \in I} \mathbb{I}\langle b_i, d_i\rangle.$$

*The intervals* $\langle b_i, d_i\rangle$ *are exactly the barcodes if* $\mathbb{U}$ *is a persistent homology module.*

Note that unless we have some additional tame-ness condition on $\mathbb{U}$, I is not guaranteed to be finite.

Recall that when we talked about persistent homology, we said that there is some consistent global choice of basis for persistent homology groups. That is a consequence of the structure theorem. The structure theorem also allows us to prove one direction of Theorem 3.50, which we will do in the following.

**Proposition 3.61.** *Consider two interval modules* $\mathbb{I}_1 = \mathbb{I}\langle b_1, d_1\rangle$ *and* $\mathbb{I}_2 = \mathbb{I}\langle b_2, d_2\rangle$. *Then,* $d_I(\mathbb{I}_1, \mathbb{I}_2) = d_b(\mathrm{Dgm}\mathbb{I}_1, \mathrm{Dgm}\mathbb{I}_2)$.

*Proof.* To prove that $d_I(\mathbb{I}_1, \mathbb{I}_2) \geqslant d_b(\mathrm{Dgm}\mathbb{I}_1, \mathrm{Dgm}\mathbb{I}_2)$, we show that every upper bound on $d_I$ is also an upper bound on $d_b$: assume that we have maps $\varphi, \psi$ showing that the two modules are $\epsilon$-interleaved. Then, consider $\psi_{a+\epsilon} \circ \varphi_a = \nu^1_{a,a+2\epsilon}$, equality holding because $\varphi, \psi$ certify $\epsilon$-interleaving. Consider $a \in \langle b_1, d_1\rangle$.

**Case 1:** $\nu^1_{a,a+2\epsilon} = 0$ for all $a \in \langle b_1, d_1\rangle$. Then, $d_1 - b_1 < 2\epsilon$, and the (infinity-norm) distance of $(b_1, d_1)$ to the diagonal is less than $\epsilon$.

**Case 2:** $\nu^1_{a,a+2\epsilon} = \mathrm{id}$ for some $a \in \langle b_1, d_1\rangle$. Then, $d_1 - b_1 \geqslant 2\epsilon$. Furthermore, we have $\varphi_a(\mathbb{F}) = \mathbb{F}$ for all $a \in \langle b_1, d_1 - 2\epsilon\rangle$. So, for these $a$, we must also have $a + \epsilon \in \langle b_2, d_2\rangle$. This tells us that $\langle b_2, d_2\rangle$ must "cover" a large part of $\langle b_1, d_1\rangle$, namely we get $b_2 \leqslant b_1 + \epsilon$, and $d_2 \geqslant d_1 - \epsilon$. We can now see that $|b_2 - b_1| \leqslant \epsilon$ and $|d_2 - d_1| \leqslant \epsilon$: to violate this, $\langle b_2, d_2\rangle$ would have to be a larger interval than $\langle b_1, d_1\rangle$ (in particular, it would be longer than $2\epsilon$), and we could thus exchange their roles and get that $b_1 \leqslant b_2 + \epsilon$ and $d_1 \geqslant d_2 - \epsilon$. From this, we get that $d_\infty((b_1, d_1), (b_2, d_2)) \leqslant \epsilon$, and thus get the bound on $d_b$.

We now prove the other direction, $d_I(\mathbb{I}_1, \mathbb{I}_2) \leqslant d_b(\mathrm{Dgm}\mathbb{I}_1, \mathrm{Dgm}\mathbb{I}_2)$. To see this, we show that from every matching whose longest edge is $\epsilon$, we get an $\epsilon$-interleaving.

**Case 1:** The two off-diagonal points are matched to the diagonal. Then, we get that $d_i - b_i \leqslant 2\epsilon$ for both of them, and thus for all $\epsilon' > \epsilon$, $\mathbb{I}_1$ and $\mathbb{I}_2$ are $\epsilon'$-interleaved with $\varphi, \psi = 0$. Thus, $d_I \leqslant \epsilon$.

**Case 2:** The points are matched with each other. Then, $|b_2 - b_1| \leqslant \epsilon$ and $|d_2 - d_1| \leqslant \epsilon$. Taking $\varphi, \psi = \mathrm{id}$ we can see that $\mathbb{I}_1$ and $\mathbb{I}_2$ are $\epsilon$-interleaved. Thus, $d_I \leqslant \epsilon$. □

**Corollary 3.62.** *Let* $\mathbb{U}, \mathbb{V}$ *be p.f.d. persistence modules. Then,* $d_I(\mathbb{U}, \mathbb{V}) \leqslant d_b(\mathrm{Dgm}\mathbb{U}, \mathrm{Dgm}\mathbb{V})$.

*Proof.* We apply the structure theorem to write $\mathbb{U} = \bigoplus_{i\in I} \mathbb{I}\langle b_i, d_i\rangle \oplus \bigoplus_{j\in J} 0$ and $\mathbb{V} = \bigoplus_{j\in J} \mathbb{I}\langle b_j, d_j\rangle \oplus \bigoplus_{i\in I} 0$. From the Bottleneck matching we get a matching between parts making up $\mathbb{U}$ and $\mathbb{V}$. Since the Bottleneck distance is the maximum length of any edge, we have $d_b(\mathrm{Dgm}\mathbb{U}, \mathrm{Dgm}\mathbb{V}) \geqslant d_b(\mathrm{Dgm}\mathbb{I}_1, \mathrm{Dgm}\mathbb{I}_2) = d_I(\mathbb{I}_1, \mathbb{I}_2)$ for every two interval modules that were matched together, where we used Proposition 3.61. Finally, we use Proposition 3.59 to get the desired statement. □

## Questions

13. *What is a filtration?* State the definition and describe different ways how filtrations appear in topology and data analysis.
14. *What persistent homology?* State the formal definitions and give examples.
15. *How can persistent homology be computed?* Discuss the two algorithms described in Section 3.3.
16. *What are the Čech and Vietoris-Rips complexes?* Give the definitions, discuss their size and theoretical guarantees, and how they are related.
17. *What are the Delaunay and Alpha complexes?* Give the definitions, discuss their size and theoretical guarantees, and how they are related.
18. *What is the Witness complex?* State the Definition and describe how it relates to the non-sparse complexes.
19. *What is the Graph induced complex?* State the Definition and describe how it relates to the non-sparse complexes.
20. *How can we measure distances between persistence diagrams?* Discuss Bottleneck and Wasserstein distance.
21. *How stable are filtrations derived from simplex-wise monotone functions with respect to Bottleneck distance?* State, illustrate and prove the stability theorem (Theorem 3.33).
22. *How can we measure distances between persistence diagrams?* Define interleaving distance and discuss its relation to Botleneck distance.
23. *How stable are Čech complexes to perturbations of the underlying point set?* Define Hausdorff distance, state and prove the stability theorem for Čech complexes (Theorem 3.55).

# Chapter 4

# Reeb graphs and Mapper

In this chapter we look at another tool in topological data analysis, called *Mapper*. The underlying idea of Mapper has its roots in Morse theory, where Georges Reeb defined a graph to summarize a Morse function on a manifold. We first discuss these graphs, called *Reeb graphs*, and then how to mimic the ideas for the case where instead of a manifold we have point cloud data.

Before we dive into the mathematical details, a short remark about the pronunciation of the word "Reeb graph". Georges Reeb, after whom these graphs are named, was a French mathematician born in the German speaking region Alsace. Thus, he likely pronounced his name the German way, that is, with the "ee" spoken similar to the "ea" in "bear" (as opposed to "beer").

## 4.1 Reeb Graphs

The idea of Reeb graphs is that given some topological space $X$, and some function $f : X \to \mathbb{R}$, we consider the preimage of $f$ for some fixed value $a \in \mathbb{R}$. We place one point per connected path-component of the preimage. We do this for some values in $\mathbb{R}$, and connect the points corresponding to neighboring connected components in adjacent preimages. More formally,

**Definition 4.1.** *Let* $X$ *be some topological space, and* $f$ *a function* $f : X \to \mathbb{R}$*. Two points* $x, y$ *are called equivalent (*$x \sim y$*), iff* $f(x) = f(y) = \alpha$ *and* $x$ *and* $y$ *are in the same path-connected component of* $f^{-1}(\alpha)$*. The* Reeb graph $R_f$ *is the quotient space* $X/\sim$*.*

To make sure that nothing weird happens due to some things being infinite, we assume all of our functions to be levelset tame:

**Definition 4.2.** *A function* $f : X \to \mathbb{R}$ *is* levelset tame *if*

- *each levelset* $f^{-1}(\alpha)$ *has finitely many connected components, all of which are path-connected, and*

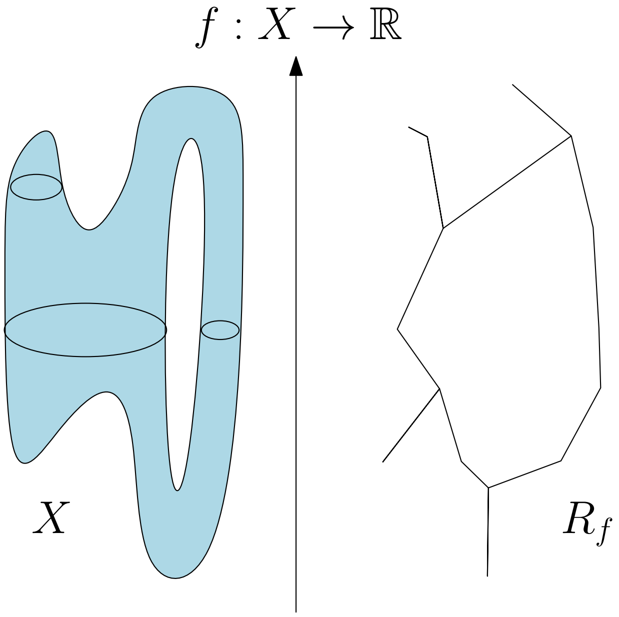


**Figure 4.1:** *An example of a Reeb graph*

- *the homology groups of the levelsets only change at finitely many critical values.*

The Reeb graph itself is just a (continuous) topological space. We call it a graph, since it is 1-dimensional. To arrive at a graph as we know it in combinatorics, we will need to discretize it. To discretize the Reeb graph, we need to define vertices and edges. There are many different possibilities of defining vertices and edges to discretize the Reeb graph, but we want to define some type of minimal one.

Let us look at the neighborhood of some point $p$ in the Reeb graph (as a topological space). We look at how many ways there exist to go from $p$ towards the direction of higher $f$-value (we call this number the up-degree $u$), and how many ways to go towards the direction of lower $f$-value (we call this the down-degree $l$). Depending on $u$ and $l$, we classify $p$ as in Table 4.1.

**Table 4.1:** *Classifications of points in the Reeb graph.*

| $u$ | $l$ | Classification |
|---|---|---|
| 1 | 1 | regular |
| 0 | $> 0$ | maximum |
| $> 0$ | 0 | minimum |
| $\geqslant 2$ | $l$ | up-fork |
| $u$ | $\geqslant 2$ | down-fork |

Note that a point can fall into multiple of these classes, for example it can be a maximum and a down-fork simultaneously, or an up-fork and a down-fork simultaneously. We call the minima, maxima, up-forks, and down-forks *critical points.* Our discretization places vertices at the critical points. Note that the graph we get through this process is not necessarily simple, we may have multi-edges.

**Exercise 4.3.** *Consider a double torus embedded in* $\mathbb{R}^3$*. You can imagine it as the result of taking the figure depicted in Figure 4.2 embedded in the plane* $x_3 = 0$*, replacing every point by a 3-dimensional ball with radius* $r < \min\{d/2, R/2\}$*, and taking the boundary of the union of these balls.*

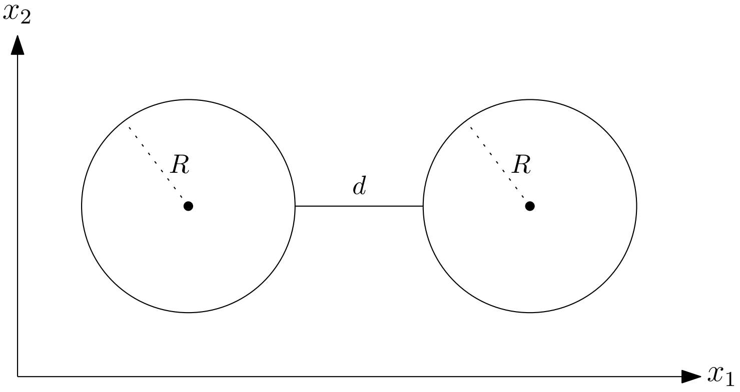


**Figure 4.2**: *The space blown up to a double torus in Exercise 4.3.*

*Draw the Reeb graph for the three functions* $f_1(x) = x_1$*,* $f_2(x) = x_2$*, and* $f_3(x) = x_3$*.*

We next consider merge trees and split trees, which are variants of the Reeb graph, where instead of levelsets, we look at sub-level sets or super-level sets.

**Definition 4.4.** *Let* $X$ *be some topological space, and* $f$ *a function* $f : X \to \mathbb{R}$*. We have* $x \sim_M y$ *for two points* $x, y$*, if and only if* $f(x) = f(y) = \alpha$ *and* $x$ *and* $y$ *are in the same connected component of* $f^{-1}((-\infty, \alpha])$*. The* merge tree $T_M$ *is the quotient space* $X/\sim_M$*.*

Note that in the merge tree, since we only increase the space under consideration, we never have a connected component that splits. We can only have new connected components appearing, and connected components merging. This also tells us that the Merge tree (or its discretization) is always a tree.

**Definition 4.5.** *Let* $X$ *be some topological space, and* $f$ *a function* $f : X \to \mathbb{R}$*. We have* $x \sim_S y$ *for two points* $x, y$*, if and only if* $f(x) = f(y) = \alpha$ *and* $x$ *and* $y$ *are in the same connected component of* $f^{-1}([\alpha, \infty))$*. The* split tree $T_S$ *is the quotient space* $X/\sim_S$*.*

In topological data analysis, we use computers, which cannot handle arbitrary topological spaces. We thus now look more at Reeb graphs in the context of simplicial complexes. We consider a simplicial complex $K$ and a function $f : |K| \to \mathbb{R}$, which is piece-wise linear (linear on each simplex). We observe that the Reeb graph then only depends on the 2-skeleton of $K$. This is the case since looking at a levelset is the same as cutting through the simplicial complex. When we cut through a simplex, we generally get a simplex of one dimension lower. In a simplicial complex, connectivity is completely determined by the 1-skeleton. Thus, before cutting, the 2-skeleton suffices. Furthermore, we can see that the critical points are images of the vertices of $K$. This happens since a connected component can only appear, disappear, split, or merge at some local maximum or minimum of the connected component. Since the function is linear, the maximum or minimum of every simplex is also attained at some vertex. We define the *augmented Reeb graph* of a simplicial complex with a PL-function, by just taking all the images of the vertices as our graph vertices.

How can we compute this augmented Reeb graph? We can do a discrete sweep (or scan) through the simplicial complex in the order given by $f$, only stopping at values $a$ such that $f(v) = a$ for some vertex $v$. In this sweep, we want to keep track of the connected components. The levelset $f^{-1}(\alpha)$ of the 2-skeleton of $K$ is just a graph $G_\alpha$: vertices and edges of $K$ induce vertices of $G_\alpha$, triangles induce edges. We can now go through our vertices in order, look at these graphs, and update the connected components. The runtime of this algorithm is given by the data structure used to manage the connected components. We want a data structure that can update the connected components under insertion and deletions of edges and vertices. There are such data structures that can do each update in amortized time $O(\log m)$, where $m$ is the size of the graph. The size of the graph is bounded by the sum $m$ of vertices, edges, and triangles in $K$. Each such feature appears at one point, and disappears at one point, and we thus have at most $2m$ insertions and deletions in total, giving an $O(m \log m)$ algorithm. We thus have the following theorem.

**Theorem 4.6.** *Given a 2-dimensional simplicial complex* $K$ *with* $m$ *faces and a piece-wise linear function* $f : |K| \to \mathbb{R}$ *on it, we can compute the augmented Reeb graph* $R_f$ *of* $K$ *with respect to* $f$ *in time* $O(m \log m)$.

**Exercise 4.7.** *Consider a simplicial complex* $K$ *and a PL (piece-wise linear) function* $f : |K| \to \mathbb{R}$. *What happens to the Reeb graph when you add one additional face to* $K$ *and extend* $f$ *accordingly?*

### 4.1.1 Homology of Reeb graphs

The Reeb graph of a topological space $X$ with respect to a function $f$ can be viewed as a summary of $X$ through the lens of $X$. The natural question is: how good of a summary is it? It is clear that in general we lose information, for example on the dimension of $X$, but we can still hope that some topological information is retained. In this section, we thus compare the homology of the Reeb graph to the homology of $X$. Since the Reeb graph $R_f$ is a graph (a 1-dimensional object), we have $H_p(R_f) = 0$ for $p \geqslant 2$, so any higher-dimensional homology gets lost. However, a graph still has homology in dimensions 0 and 1.

**Observation 4.8.** *For a levelset tame* $f : X \to \mathbb{R}$*, we have* $\beta_0(X) = \beta_0(R_f)$.

In other words, the Reeb graph captures the 0-homology of the input space $X$ perfectly, no matter which levelset tame function $f$ we use.

Sadly, the same does not hold for the 1-homology. Let us consider a torus, as in Figure 4.3. In general, it can be that the choice of function $f$ determines whether we capture a hole or not, consider e.g. a cylinder. Note that for the torus, it is actually the case that no matter which function $f$ we choose, we cannot capture its 1-homology (this is non-trivial to show).

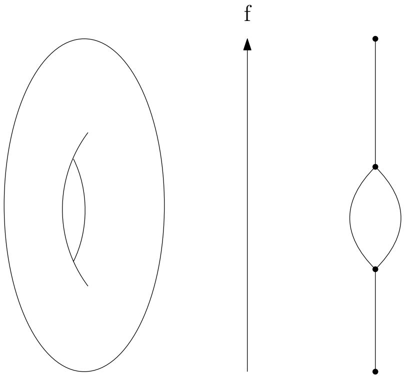


**Figure 4.3:** *The torus and its Reeb graph.*

On the other hand, we can see that every cycle in the Reeb graph is indeed also a cycle in the topological space $X$, and it cannot be filled in, so it is indeed a hole. Thus we also get the following observation:

**Observation 4.9.** *For a levelset tame* $f : X \to \mathbb{R}$*, we have* $\beta_1(X) \geqslant \beta_1(R_f)$.

Can we somehow formalize which holes we lose? To do this, we split up homology into "horizontal" and a "vertical" parts, where horizontal and vertical are of course relative to $f$.

**Definition 4.10.** *A* $p$*-th homology class* $h \in H_p(X)$ *is called* horizontal *if there is a finite set of values* $A = \{a_1, \dots, a_k\}$ *such that* $h$ *has a pre-image under the map* $H_p(\bigcup_{a \in A} X_a) \to H_p(X)$ *induced by inclusion, where* $X_a = f^{-1}(a)$.

This definition means that we need to be able to find a finite set of levelsets, such that we can find cycles contained in these levelsets, which are in the homology class $h$ in $H_p(X)$.

One now wonders whether the set of horizontal homology classes forms a group. Let this set be $\overline{H_p}(X)$. It turns out that it is indeed a group.

**Lemma 4.11.** $\overline{H_p}(X)$ *is a subgroup of* $H_p(X)$.

*Proof.* First, we see that the identity element $0$ is in $\overline{H_p}(X)$. We can take an arbitrary set $A$, and we can always map the $0$ element of $H_p(\bigcup_{a \in A} X_a)$ to $0$.

Next, we show that the set is closed under addition. Let $p, q \in \overline{H_p}(X)$, and we show that $p + q \in \overline{H_p}(X)$. $p$ has a pre-image in some levelset $A_p$, and $q$ has a pre-image in some levelset $A_q$. $p + q$ must have a pre-image in $A_p \cup A_q$.

Finally, we show that the inverse of every element is contained in the group, but since every element is self-inverse in $\mathbb{Z}_2$-homology, we get that for every element its inverse is also contained in $\overline{H_p}(X)$. □

Since the horizontal homology is a sub-group, we can now easily define *vertical homology* by taking quotient groups.

**Definition 4.12.** *The* vertical homology group *of* $X$ *with respect to* $f$ *is the group* $\overset{\vee}{H}_p(X) := H_p(X)/\overline{H_p}(X)$.

**Observation 4.13.** $\operatorname{rank}(H_p(X)) = \operatorname{rank}(\overline{H_p}(X)) + \operatorname{rank}(\overset{\vee}{H}_p(X))$.

**Fact 4.14.** *The surjection* $\phi : X \to R_f$ *induces an isomorphism* $\overset{\vee}{\Phi} : \overset{\vee}{H}_1(X) \to H_1(R_f)$.

In other words, when we go from a space $X$ to its Reeb graph, we keep the vertical homology classes, and lose the horizontal ones.

**Corollary 4.15.** *Given* $X$ *an orientable connected compact 2-manifold, and a Morse function* $f : X \to \mathbb{R}$, *then* $\beta_1(R_f) = \beta_1(X)/2$.

Here, a 2-manifold is a space that locally at every point looks like $\mathbb{R}^2$. Orientable means that there is an inside and an outside side. A Morse function is a "nice enough" function defined in terms of some derivatives, which we do not need to specify here.

**Exercise 4.16.** *(a) Consider a 2-dimensional geometric simplicial complex* $K$ *embedded in* $\mathbb{R}^2$. *Consider the function* $f(x) = x_1$. *Show that* $\beta_1(K) = \beta_1(R_f)$.

*(b) Find a geometric simplicial complex* $K$ *embedded in* $\mathbb{R}^2$ *and a map* $f : K \to \mathbb{R}$ *such that* $\beta_1(K) > \beta_1(R_f)$.

## 4.2 Distances for Reeb Graphs

In order to compare Reeb graphs to each other, we again want to define distance measures between them. We discuss two such measures here. The first one, called *interleaving distance*, is, not surprisingly, similar to the interleaving distance of persistence modules. The second one, called *functional distortion distance* is similar to the Gromov-Hausdorff distance for metric spaces.

### 4.2.1 Interleaving Distance

When do we want two Reeb graphs to be considered the same, and thus have distance 0? We definitely need that the graphs are isomorphic in the sense of graph isomorphism. But further than that, we also want that this graph isomorphism is "function preserving". In other words, the critical points should lie on the same function levels. The idea of the interleaving distance is to measure how far away from this we are. Thus, given two Reeb graphs $R_f, R_g$, "how much" is missing towards a "function preserving isomorphism"? Towards formalizing this idea, we need a few definitions.

Note that when we compare two Reeb graphs $R_f, R_g$, those can be Reeb graphs of *different spaces* with regards to *different functions*.

**Definition 4.17.** *An* $\epsilon$-thickening $X_\epsilon$ *of some topological space* $X$ *is given by* $X_\epsilon := X \times [-\epsilon, +\epsilon]$.

**Definition 4.18.** *For a Reeb graph* $R_f$ *consider a function* $f_\epsilon : (R_f)_\epsilon \to \mathbb{R}$ *such that*

$$(x, t) \mapsto f(x) + t.$$

*The* $\epsilon$-smoothing *of* $R_f$*, denoted by* $S_\epsilon(R_f)$ *is the Reeb graph of* $(R_f)_\epsilon$ *with regards to* $f_\epsilon$.

An example of these definitions can be seen in Figure 4.4. Note that when we say $(R_f)_\epsilon$, we mean an $\epsilon$-thickening of $R_f$, *not* a Reeb graph with regards to some function $f_\epsilon$. The $\epsilon$-smoothing $S_\epsilon(R_f)$ is then a Reeb graph with regards to the function $f_\epsilon$, but *of* $(R_f)_\epsilon$, and not of the original space $R_f$ is the Reeb graph of. Furthermore, when we write $f(x)$ for some $x \in R_f$, we mean that we extend $f$ to some function $f^* : R_f \to \mathbb{R}$ by defining $f^*(x) = f(f^{-}1(x))$. We will just call this function $f$ as well for simplicity.

**Definition 4.19.** *The function* $\iota : R_f \to S_\epsilon(R_f)$ *with* $x \mapsto [(x, 0)]$ *is the* quotiented inclusion map. *Here,* $[(x, 0)]$ *denotes the equivalence class, or the connected component that contains* $(x, 0)$ *in* $f_\epsilon^{-1}(f_\epsilon(x, 0))$.

Consider some function $\mu : R_f \to R_g$ which is function preserving, i.e., $f(x) = g(\mu(x))$ for all $x \in R_f$. A function-preserving map $\mu : R_f \to S_\epsilon(R_g)$ induces a function preserving map $\mu_\epsilon : S_\epsilon(R_f) \to S_{2\epsilon}(R_g)$ with $[x, t] \mapsto [\mu(x), t]$.

**Definition 4.20** (Reeb graph interleaving). *A Reeb graph interleaving is a pair of function preserving maps* $\varphi : R_f \to S_\epsilon(R_g)$, $\psi : R_g \to S_\epsilon(R_f)$ *are* $\epsilon$*-interleaved, if the following diagram commutes:*

$$\begin{tikzcd}
R_f \arrow[r, "\iota"] \arrow[dr, "\varphi" near start] & S_\epsilon(R_f) \arrow[r, "\iota_\epsilon"] \arrow[dr, "\varphi_\epsilon" near start] & S_{2\epsilon}(R_f) \\
R_g \arrow[r, "\iota"] \arrow[ur, "\psi" near end] & S_\epsilon(R_g) \arrow[r, "\iota_\epsilon"] \arrow[ur, "\psi_\epsilon" near end] & S_{2\epsilon}(R_g)
\end{tikzcd}$$

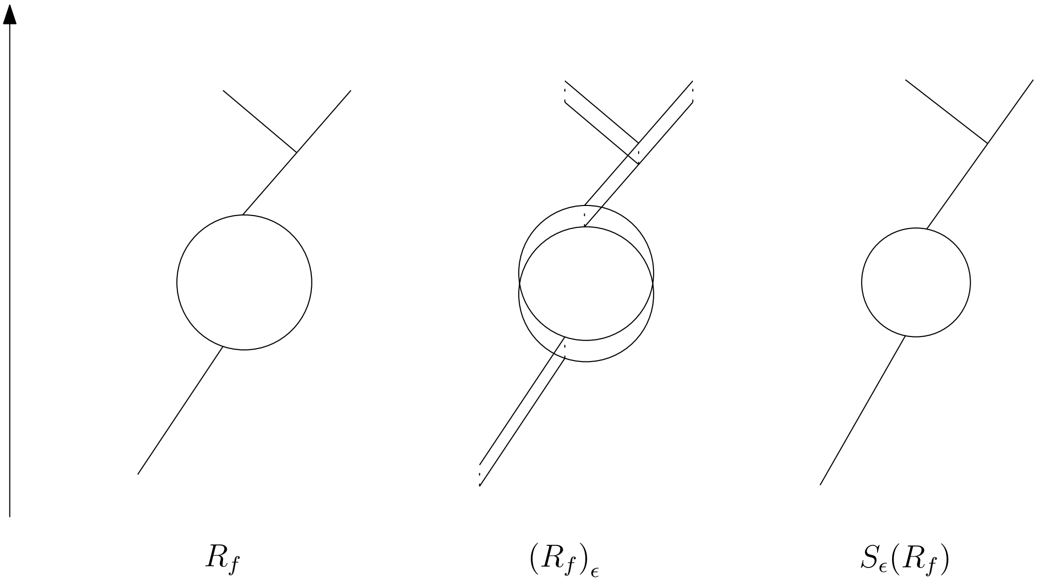


**Figure 4.4**: *A Reeb graph, its* $\epsilon$*-thickening, and its* $\epsilon$*-smoothing.*

Here, to understand why $\iota_\epsilon$ makes sense, we need the following fact, the proof of which is left as an exercise.

**Observation 4.21.** $S_\delta(S_\epsilon(R_f)) = S_{\delta+\epsilon}(R_f)$.

Note that by construction of $\iota, \iota_\epsilon$ and $\varphi_\epsilon$ (or $\psi_\epsilon$, respectively), the trapezoidal parts of this diagram commute trivially: $\varphi_\epsilon \circ \iota(x) = \varphi_\epsilon([x, 0]) = [\varphi(x), 0] = \iota_\epsilon \circ \varphi(x)$. Furthermore, note that for sufficiently large $\epsilon$, $S_\epsilon(R_f)$ is a union of segments, i.e., any two Reeb graphs of compact connected spaces are $\epsilon$-interleaved for some $\epsilon$. Lastly, if $R_f$ and $R_g$ are $\epsilon$-interleaved, then they are also $\delta$-interleaved for all $\delta \geqslant \epsilon$.

**Definition 4.22.** $d_I(R_f, R_g) = \inf\{\epsilon \mid R_f, R_g$ *are* $\epsilon$*-interleaved*$\}$.

We once again have a stability theorem, which we will not prove here.

**Theorem 4.23.** *For tame functions* $f, g : X \to \mathbb{R}$ *we have* $d_I(R_f, R_g) \leqslant \|f - g\|_\infty$.

### 4.2.2 Functional Distortion Distance

As mentioned above, the functional distortion distance in motivated by the Gromov-Hausdorff distance for metric spaces. Thus, the first step is to define a metric on a Reeb graph.

**Definition 4.24.** *Let* $R_f$ *be a Reeb graph of a space* $X$*, and* $u, v \in R_f$ *(in the same connected component), and let* $\pi$ *be a path from* $u$ *to* $v$*. We define the* height *of* $\pi$ *as* $\text{height}(\pi) = \max_{x\in\pi} f(x) - \min_{x\in\pi} f(x)$*. To turn this into a distance metric, we* *consider* $\Pi(u,v)$,*the set of all paths between* $u$ *and* $v$. *Then, the* function induced metric *on* $R_f$ *is defined as*

$$d_f(u,v) = \min_{\pi \in \Pi(u,v)} \text{height}(\pi).$$

In a sense, $d_f(u,v)$ is the "thickness" of the thinnest "slice" of the space $X$ in which $u$ and $v$ are connected.

**Definition 4.25** (Functional distortion distance). *Let* $R_f$ *and* $R_g$ *be two Reeb graphs. Let* $\Phi : R_f \to R_g$, $\Psi : R_g \to R_f$ *be continuous functions, but not necessarily function-preserving. Then, we define correspondence and distortion:*

$$C(\Phi,\Psi) = \{(x,y) \in R_f \times R_g \mid \Phi(x) = y \text{ or } x = \Psi(y)\}$$

$$D(\Phi,\Psi) = \sup_{(x,y),(x',y') \in C(\Phi,\Psi)} \frac{1}{2}|d_f(x,x') - d_g(y,y')|.$$

*And finally, we define the* functional distortion distance,

$$d_{FD}(R_f,R_g) = \inf_{\Phi,\Psi} \max\{D(\Phi,\Psi), \|f - (g \circ \Phi)\|_\infty, \|g - (f \circ \Psi)\|_\infty\}.$$

Also for this distance measure there is a stability theorem.

**Theorem 4.26.** *Let* $f, g : X \to \mathbb{R}$ *be tame functions. Then,* $d_{FD}(R_f,R_g) \leqslant \|f - g\|_\infty$.

We can also quantify the relation between the two discussed distances.

**Theorem 4.27.** $d_I(R_f,R_g) \leqslant d_{FD}(R_f,R_g) \leqslant 3d_I(R_f,R_g)$.

**Exercise 4.28.** *Consider a merge tree* $T$ *with regards to a function* $f$. *We define the* $a$*-shift* $x^a$ *for any* $x \in T$ *to be the unique "ancestor" of* $x$ *with function value* $f(x^a) = f(x) + a$.

*We now consider two merge trees;* $T_1$ *with regards to* $f$, *and* $T_2$ *with regards to* $g$. *We call* $T_1$ *and* $T_2$ $\epsilon$-compatible *if there exist maps* $\alpha : T_1 \to T_2$ *and* $\beta : T_2 \to T_1$ *such that we get the following commutativities:*

- $g(\alpha(x)) = f(x) + \epsilon$ *for all* $x \in T_1$
- $f(\beta(y)) = g(y) + \epsilon$ *for all* $y \in T_2$
- $\beta \circ \alpha(x) = x^{2\epsilon}$ *for all* $x \in T_1$
- $\alpha \circ \beta(y) = y^{2\epsilon}$ *for all* $y \in T_2$.

*The interleaving distance between merge trees* $d_I(T_1,T_2)$ *can now be defined as the infimum of all* $\epsilon$ *such that* $T_1$ *and* $T_2$ *are* $\epsilon$*-compatible. Show that* $d_I(T_1,T_2) = d_{FD}(T_1,T_2)$.

*Note: we technically only defined* $d_{FD}$ *for Reeb graphs. You can simply consider a merge tree to be the Reeb graph of itself (with regards to the same filter function).*

## 4.3 Mapper

### 4.3.1 An approximation of the Reeb graph

Reeb graphs loose a lot of information, since they at most retain some 1-dimensional holes, but no larger holes. To generalize Reeb graphs further, we start looking at neighborhoods instead of levelsets, which will then lead to the Mapper algorithm.

To begin, we consider the 1-dimensional case, and try to find an approximation of the Reeb graph. Instead of looking at pre-images of points, we will now look at pre-images of intervals. Let $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$ be an open cover of $\mathbb{R}$ (i.e., a collection of open sets whose union is $\mathbb{R}$). As always, we consider a function $f : X \to \mathbb{R}$. For each $f^{-1}(U_\alpha)$, we consider a partition into path-connected components, i.e., $f^{-1}(U_\alpha) = \bigcup_{\beta \in B_\alpha} V_\beta$. We then look at $f^*(\mathcal{U}) := \{V_\beta\}$, the set of all $V_\beta$ we get over all $\alpha$. Our object of interest is the nerve of this family, i.e., $N(f^*(\mathcal{U}))$.

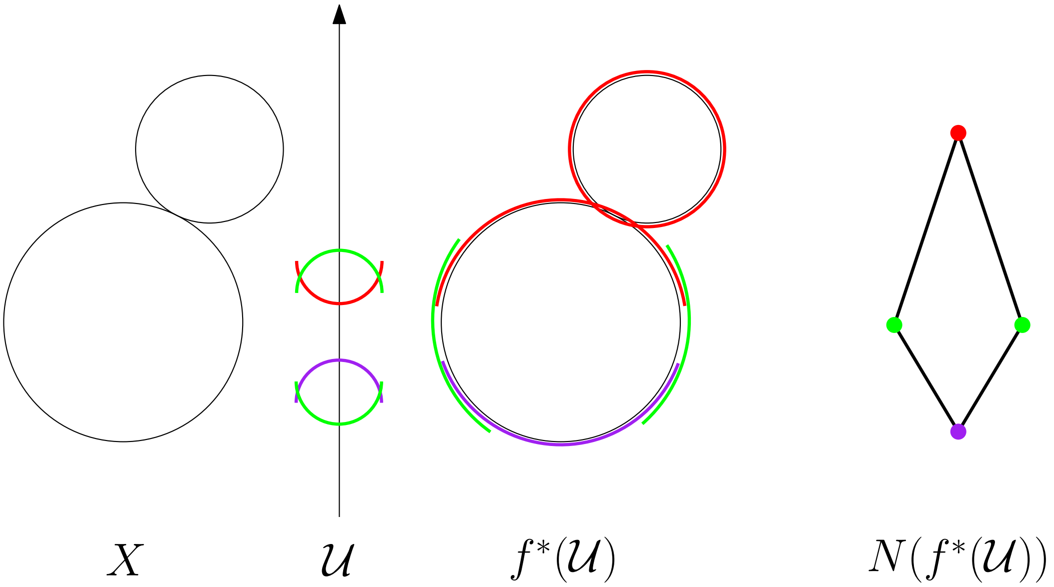


**Figure 4.5**: *A space* $X$*, an open cover* $\mathcal{U}$ *of* $\mathbb{R}$*, the family* $f^*(\mathcal{F})$*, and its nerve.*

If we take sufficiently nice functions, and sufficiently fine covers, then $N(f^*(\mathcal{U}))$ is isomorphic to $R_f$.

### 4.3.2 Topological Mapper

We can generalize this idea to maps to arbitrary spaces.

**Definition 4.29.** *Let* $X, Z$ *be topological spaces. Then we call* $f : X \to Z$ well-behaved *if for all open sets* $U \subseteq Z$*,* $f^{-1}(U)$ *has finitely many path-connected components.*

**Definition 4.30** (Mapper). *Let* $f : X \to Z$ *be well-behaved, and* $\mathcal{U}$ *be a (finite) open cover of* $Z$*. Then the* Mapper *is defined as* $M(\mathcal{U}, f) := N(f^*(\mathcal{U}))$.

As an example, we look at $X$ being the boundary of the 3-cube $[0,1]^3$. We then also look at $Z_1 = \mathbb{R}^2$ spanned by the $x$- and $y$-axis, with $f_1 : X \to Z_1$ being the projection onto this plane. Furthermore, we look at $Z_2 = \mathbb{R}$, spanned by just the $x$-axis, and $f_2 : X \to Z_2$ being again the projection.

We consider the open cover $\mathcal{U}_2$ of $Z_2$: $\{(-\infty, \frac{1}{3}), (0,1), (\frac{2}{3}, +\infty)\}$. For $Z_1$, we consider the cover $\mathcal{U}_1 := \mathcal{U}_2 \times \mathcal{U}_2$.

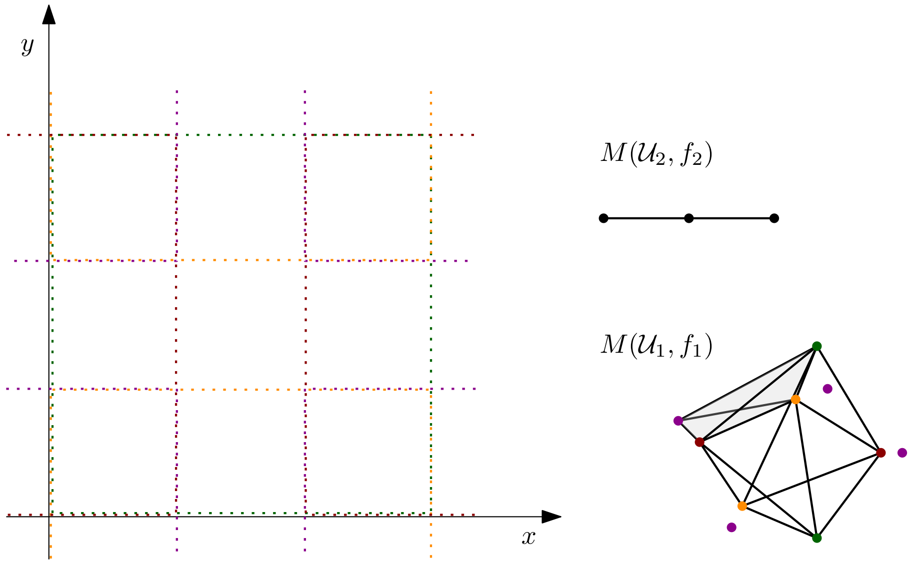


**Figure 4.6:** *The cover* $\mathcal{U}_\infty$*, and the two Mappers. The Mapper* $M(\mathcal{U}_1, f_1)$ *consists of an empty octahedron, with additional filled tetrahedra attached at the purple vertices. The whole space thus collapses to an octahedron.*

**Exercise 4.31.** *(a) Consider spaces* $X, Z$*, a filter function* $f : X \to Z$*, and an open cover* $\mathcal{U}$ *of* $Z$*. Show that if the* pullback cover $f^*(\mathcal{U})$ *is a* good cover *of* $X$*, then* $M(\mathcal{U}, f)$ *is homotopy equivalent to* $X$*.*

*(b) Give an example of spaces* $X, Z$*, a filter function* $f : X \to Z$*, and a* good cover $\mathcal{U}$ *of* $Z$*, such that* $M(\mathcal{U}, f)$ *is* not *homotopy equivalent to* $X$*.*

*(c) Give an example of spaces* $X, Z$*, a filter function* $f : X \to Z$*, and an open cover* $\mathcal{U}$ *of* $Z$ *such that the pullback cover* $f^*(\mathcal{U})$ *is* not *a good cover, but* $M(\mathcal{U}, f)$ *is still homotopy equivalent to* $X$*.*

### 4.3.3 Mapper for Point Clouds

We would like to apply the ideas of the topological Mapper to analyze the shape of data. However, once again we have the issue that data usually does not come in the form of a topological space, but as a set of data points with a notion of distance between them. The *Mapper algorithm* for point clouds adapts the ideas of the topological Mapper to this setting.

**Input:** In the most general setting, data comes as a finite metric space $(P, d_P)$, for example as points in $\mathbb{R}^d$ or as vertices of a graph. We also requires a cover $\mathcal{U}$ of a space $Z$, usually $Z = \mathbb{R}$, as input. Finally, we also need a filter function $f : P \to Z$ and a clustering algorithm (which might also require some input parameters).

**Algorithm:** Since at the moment we only have a discrete metric space, we do not really have the notion of connected components yet. For every $U \in \mathcal{U}$, we thus cluster the pre-image $f^{-1}(\mathcal{U})$ using some clustering algorithm, which we can also consider as an input. Now, we can just consider each cluster $C_i$ as a vertex of some simplicial complex $K$, and add a face $\{C_1, \dots, C_k\}$ to $K$ if these clusters (which are just point sets) have a common point.

**Output:** We output $K$, or even just its 1-skeleton.

As you can see, this algorithm requires a lot of input parameters. While this allows to encode previous knowledge of the data set (e.g. by choosing as filter function the distance to a known center of the data), it also makes the space of possible outputs very large. Picking the correct parameters is currently still an art form on its own, and there is currently significant research being done towards understanding the interplay between the parameters and statistical guarantees for certain good choices of parameters.

## 4.4 Multiscale Mapper

Motivated by the many tuneable parameters, we discuss here one idea to look at many values at once. The multiscale Mapper is a combination of the ideas of persistence and of Mapper. We here want to look at different covers.

**Definition 4.32.** *Let* $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$ *and* $\mathcal{F} = \{V_\beta\}_{\beta \in B}$ *be two covers of the same space* $X$. *A* map of covers *is a map* $\varphi : A \to B$ *such that for every* $\alpha \in A$*, we have* $U_\alpha \subseteq V_{\varphi(\alpha)}$.

**Proposition 4.33.** *If* $\varphi : \mathcal{U} \to \mathcal{V}$ *is a map of covers (with a slight abuse of notation), then the map* $N(\varphi) : N(\mathcal{U}) \to N(\mathcal{V})$ *given on the vertices by* $\varphi$ *is simplicial.*

*Proof.* Let $\sigma \in N(\mathcal{U})$. We need to show that the intersection $\bigcap_{\beta \in \varphi(\sigma)} V_\beta$ is non-empty.

$$\bigcap_{\beta \in \varphi(\sigma)} V_\beta = \bigcap_{\alpha \in \sigma} V_{\varphi(\alpha)} \supseteq \bigcap_{\alpha \in \sigma} U_\alpha \neq \emptyset$$

Thus, $\varphi(\sigma) \in N(\mathcal{V})$. □

**Proposition 4.34.** *Let* $f : X \to Z$ *be some map, and* $\mathcal{U}, \mathcal{V}$ *be covers of* $Z$*, with* $\varphi : \mathcal{U} \to \mathcal{V}$ *some map of covers. Then, there exists a map of covers* $f^*(\varphi) : f^*(\mathcal{U}) \to f^*(\mathcal{V})$.

Recall that $f^*(\mathcal{U})$ is the cover of $X$ consisting of the connected components of the pre-images of the sets of $\mathcal{U}$ under $f$.

*Proof.* For every $\alpha$, we have $U_\alpha \subseteq V_{\varphi(\alpha)} \implies f^{-1}(U_\alpha) \subseteq f^{-1}(V_{\varphi(\alpha)})$. We now need to go from these pre-images to their connected components. Since every connected component of $f^{-1}(U_\alpha)$ must lie in a unique connected component of $f^{-1}(V_{\varphi(\alpha)})$, our desired map of covers is given by exactly mapping to this connected component. □

If we have multiple maps of covers, $\mathcal{U} \xrightarrow{\varphi} \mathcal{V} \xrightarrow{\psi} \mathcal{W}$, we can concatenate the maps, and the $f^*$ function distributes: $f^*(\psi \circ \varphi) = f^*(\psi) \circ f^*(\varphi)$.

Let $\mathfrak{U} = \mathcal{U}_1 \xrightarrow{\varphi_1} \mathcal{U}_2 \xrightarrow{\varphi_2} \dots \xrightarrow{\varphi_{n-1}} \mathcal{U}_n$ be a sequence of covers of $Z$ with maps between them, which we call a *cover tower*. By applying $f^*$ we get a cover tower $f^*(\mathfrak{U})$ of $X$.

**Definition 4.35** (Multiscale Mapper). *Let* $f : X \to Z$, $\mathfrak{U}$ *a cover tower of* $Z$. *Then, the* Multiscale Mapper $MM(\mathfrak{U}, f)$ *is*

$$MM(\mathfrak{U}, f) := N(f^*(\mathfrak{U})) = \{N(f^*(\mathcal{U}_i)) \mid \mathcal{U}_i \in \mathfrak{U}\})$$

*together with the induced simplicial maps*

$$N(f^*(\varphi_i)) : N(f^*(\mathcal{U}_i)) \to N(f^*(\rangle + \infty)).$$

*Applying homology, we get the sequence homology groups with induced homomorphisms between them, i.e., a persistence module:*

$$H_p(N(f^*(\mathcal{U}_1))) \xrightarrow{N(f^*(\varphi_1))} \dots \xrightarrow{N(f^*(\varphi_{n-1}))} H_p(N(f^*(\mathcal{U}_n))).$$

*We can now view* $\mathrm{Dgm}_p MM(\mathfrak{U}, f)$ *as a topological summary of* $f$ *through the lens of* $\mathfrak{U}$.

As opposed to the normal Mapper, at first glance the Multiscale Mapper adds even more parameters. But a cover tower can be seen as a way of looking at a whole interval of covers. For example, we can get a cover tower by increasing the size of all intervals in an interval cover. The features of the data should show up as a robust feature that persists for a longer time over this process, while spurious features obtained from choosing "wrong" Mapper parameters should disappear quickly.

## Questions

24. *What is a Reeb graph?* State the definition and describe how we get the graph structure.
25. *How can we compute the augmented Reeb graph of a piece-wise linear function?* Define the augmented Reeb graph and explain the algorithm to compute it.
26. *How much of the homology of the underlying topological space is captured by the Reeb graph?* Explain vertical and horizontal homology.

27. *What is the interleaving distance for Reeb graphs?* Give the definitions and state the relevant stability theorems.
28. *What is the functional distortion distance for Reeb graphs?* Give the definitions and state the relevant stability theorems.
29. *What is the topological Mapper?* State the Definition and give an example.
30. *How can we use Mapper on point cloud data?* Explain the Mapper algorithm and describe the input parameters.
31. *How can we use Mapper on several covers at once?* Explain the Multiscale Mapper.

# Chapter 5

# Optimal Generators

In some applications, we are not only interested in the number of holes in our data, but we also want to look at specific holes, that is, we would like to have a representation of this hole in the data, or even a basis of the homology group. However, in a homology class, there are many homologous cycles. Furthermore, there are many different choices of homology classes which form a basis of the homology group. Thus, there are many different choices for cycles as bases of the homology group. How do we find good bases?

We define a weight function $w : K_p \to \mathbb{R}_{\geqslant 0}$ on the p-simplices, and the weight of a chain is simply the sum, i.e., $w(c) = \sum \alpha_i w(\sigma_i)$ for $c = \sum \alpha_i \sigma_i$. The weight of a set of cycles $\mathcal{C}$ is then the sum of weights of each cycle. We are now interested in cycles that have minimal weight in their homology class, or at bases with minimum total weight.

We look at this problem in two settings: first we look at the case where we are given a fixed simplicial complex and we want to find an optimal basis for the homology of this complex. This can be applied for example if the persistence diagram of a filtration gives us a range of values in which we expect the complex to nicely capture the shape of the data. We can then compute an optimal basis for the fixed complex for some value in this range.

In some applications, we might also want to take a closer look at single intervals in the persistence barcode, that is, understand a hole that is born at time $b$ and dies at time $d$ (for example, to decide whether it is corresponds to a feature in the data or is just a consequence of the process). This brings us to the second setting we look at in this chapter, where we want to find an optimal representative of a persistent homology class.

## 5.1 Optimal basis of a fixed complex

**Definition 5.1.** *A set $\mathcal{C}$ of cycles is an* optimal basis *for $H_p(K)$ if it is a basis and there is no other basis $\mathcal{C}'$ with $w(\mathcal{C}') < w(\mathcal{C})$.*

How can we compute such an optimal basis?

In a first step, we are going to compute a set of cycles $\mathcal{C}$ which contains an optimal basis. Then, we sort the cycles by increasing weight, and pick the first cycle to be part of our basis $B$. Then, we simply iterate through our cycles and add a cycle $c_i$ to our basis if it cannot be written as a linear combination of our current basis. Finally, if $c_1$ is a boundary, we return the $B \setminus \{c_1\}$, and otherwise we return $B$.

Assuming that we can do all these steps, it follows from a more general framework in *matroid theory* that the computed basis is indeed optimal.

**Exercise 5.2.** *A* matroid *is given by a collection $\mathcal{I}$ of subsets of some universe $U$, such that*

1. *$\emptyset \in \mathcal{I}$, and if some set $L$ is in $\mathcal{I}$, all $L' \subseteq L$ are also in $\mathcal{I}$.*
2. *If some $L, L'$ are in $\mathcal{I}$, and $|L'| = |L| + 1$, then there exists an element $f \in L' \setminus L$, such that $L \cup \{f\} \in \mathcal{I}$.*

*The sets in $\mathcal{I}$ are also called the* independent sets *of the matroid. The inclusion-maximal sets in $\mathcal{I}$ are called* bases.

- *(a) Show that for $U$ being any finite set of vectors in some vector space, the family $\mathcal{I}$ of subsets of $U$ corresponding to linearly independent vectors forms the family of independent sets of a matroid.*
- *(b) Show that for any graph $G = (V, E)$, the family $\mathcal{I}$ of subsets of $E$ corresponding to forests in $G$ forms the family of independent sets of a matroid.*
- *(c) Consider a matroid on a universe $U$ with a weight function $w : U \to \mathbb{R}$. Consider the following greedy algorithm: begin with $L = \emptyset$, and consecutively add the lowest-weight element $e \notin L$ such that $L \cup \{e\}$ remains an independent set, until reaching a basis. Show that this greedy algorithm finds a minimum-weight basis.*

For the first step of the above algorithm, we need to be able to compute our beginning set $\mathcal{C}$. Furthermore, we need to be able to check linear independence.

From now on, we will focus on computing a basis for $H_1(K)$. Without loss of generality, we say that $K$ is 2-dimensional, with $n$ triangles, $O(n)$ edges and vertices. To compute $\mathcal{C}$, we begin with $\mathcal{C} = \emptyset$. For all vertices $v$, we compute the shortest path tree $T_v$ rooted at $v$. We can do this for example with Dijkstra's algorithm. For every edge $e$ that is not in $T_v$, we add the unique cycle in $T_v \cup \{e\}$ to $\mathcal{C}$. This can be implemented in $O(n^2 \log n)$, and yields a set of cycles with $|\mathcal{C}| \in O(n^2)$. But, we need to prove that it is indeed a set which contains an optimal basis.

**Lemma 5.3.** *$\mathcal{C}$ as computed by the algorithm above contains an optimal basis.*

*Proof.* Let $\mathcal{C}^*$ be an optimal basis, and towards a contradiction, let $c$ be a cycle contained in $\mathcal{C}^* \setminus \mathcal{C}$. As the weights are non-negative, we can assume that $c$ is simple, i.e., no edge is used multiple times.

Let $v$ be a vertex in $c$, and let $T_v$ be the corresponding shortest path tree. There must be an edge $e = \{u, w\}$ in $c$, which is not in $T_v$, since $T_v$ is a tree. Let $\Pi_{v,u}$ and $\Pi_{v,w}$ be the shortest paths from $v$ to $u, w$ respectively. These paths must be contained in $T_v$. Let us similarly consider $\Pi'_{v,u}$ and $\Pi'_{v,w}$ to be the (shortest) paths from $v$ to $u, w$ in $c$. We know that not both $\Pi'_{v,u} = \Pi_{v,u}$ and $\Pi'_{v,w} = \Pi_{v,w}$, so w.l.o.g. assume that $\Pi'_{v,u} \neq \Pi_{v,u}$.

We now define the cycle $c_1 = \{\Pi'_{v,w}, e, \Pi_{v,u}\}$ and $c_2 = \{\Pi_{v,u}, \Pi'_{v,u}\}$. We can now see that as we work in $\mathbb{Z}_2$, $c = c_1 + c_2$. Furthermore, we have $w(c_1) \leqslant w(c)$, since $\Pi_{v,u}$ is a shortest path (in $K$), while $\Pi'_{v,u}$ is not necessarily shortest. The same also holds for $c_2$: $w(c_2) \leqslant w(c)$ since $\{\Pi'_{v,w}, e\}$ can not be shorter than $\Pi_{v,u}$.

Let us now consider the homology classes of $c_1$ and $c_2$. If both $[c_1]$ and $[c_2]$ were dependent on $\mathcal{C}^* \setminus \{c\}$, then so would $[c]$, since $c = c_1 + c_2$. Then, $\mathcal{C}^*$ would not be a basis. Thus, at least one of $[c_1]$ and $[c_2]$ has to be dependent of $\mathcal{C}^* \setminus \{c\}$. Let us consider first that $c_1$ is independent. Then, we could replace $c$ by $c_1$ in $\mathcal{C}^*$ and get a basis which is at least as good as $\mathcal{C}^*$. We can repeat the argument for that basis with $v'$, the common ancestor of $\Pi_{v,u}$ and $\Pi'_{v,w}$. If $c_2$ is independent, we replace $c$ by $c_2$ in $\mathcal{C}^*$ and repeat the argument with $v'$ the common ancestor of $\Pi_{v,u}, \Pi'_{v,u}$ and $e$ an edge incident to $u$.

At the end, we get a basis $\mathcal{C}'$ with $w(\mathcal{C}') \leqslant w(\mathcal{C}^*)$ with $\mathcal{C}' \subseteq \mathcal{C}$. □

So, we have finished the first step of our algorithm. It remains to figure out how to check independence. For this, we introduce *annotations.*

**Definition 5.4.** *An* annotation of p-simplices *is a function* $a : K^p \to \mathbb{Z} - 2^g$ *giving each* $p$*-simplex a binary vector of size* $g$*. This extends to chains by sums. An annotation must fulfill:*

- $g = \beta_p(K)$
- $a(z_1) = a(z_2)$ *iff* $[z_1] = [z_2]$.

Given an annotation, we can now clearly check linear independence of cycles by simply checking linear independence of a set of vectors, for which we have existing tools such as Gaussian elimination.

**Proposition 5.5.** *In every simplicial complex* $K$ *and for every* $p \geqslant 0$*, there exists an annotation of* $p$*-simplices, and can also be computed.*

*Proof.* (Sketch for $p = 1$) We can compute a spanning forest $T$, and let $m$ be the number of remaining edges. We initialize annotations of length $m$, and set $a(e) = 0$ for every edge in the spanning forest $T$. For every remaining edge $e_i$, we set $a_j(e_i) = 1$ if and only if $j = i$, and $0$ otherwise.

For every triangle $t$, if the annotation of its boundary $\delta t$ is not $0$, we find a non-zero entry $b_u$ in $a(\delta t)$ and add $a(\delta t)$ to every edge with $a_u(e) = 1$, and we delete the $u$-th entry from all annotations. One can show that this yields a valid annotation, and it can be implemented in $O(n^3)$, and more clever implementations work in $O(n^\omega)$. □

To check independence more efficiently, we add auxiliary annotations also to vertices in a shortest path tree $T_v$ rooted at $v$. We give $v$ the annotation $0$, and for a vertex $x$ that is the child of $y$, we set $a(x) := a(y) + a(e_{xy})$. For every cycle defined by the non-tree edge $e = uw$, we now have $a(c_e) = a(u) + a(w) + a(e)$. So, we never actually have to compute an explicit representation of a cycle by its edges, we only need to store its weight, the shortest path trees with the auxiliary annotations, and the non-tree edge $e$. Note that the auxiliary annotations can be computed in $O(gn)$ for the whole tree, thus in $O(gn^2)$ for all trees.

Finally, we have to check independence. Given an $(n \times m)$ matrix $M$, we can find the lexicographic leftmost set of independent columns in time $O(\max(n, m)^{\omega})$. Instead of naively doing this $n^2$ times (once for every cycle), we group our cycles of $\mathcal{C}$ into groups $A_i$ of size $g$, and compute the leftmost set for $[B|A_i]$, and thus we get $O(n^2 g^{\omega-1})$ runtime for this step.

To summarize, computing $\mathcal{C}$ takes $O(n^2 \log n)$, sorting the $O(n^2)$ cycles also takes $O(n^2 \log n)$, and for checking linear independence we need $O(n^{\omega})$ for the annotations of the edges, $O(gn^2)$ for the auxiliary annotations, and $O(n^2 g^{\omega-1})$ for the block-wise linear independence checking. Overall, we thus get a runtime of $O(n^{\omega} + n^2 g^{\omega-1})$.

**Theorem 5.6.** *Given a 2-dimensional simplicial complex* $K$ *with* $n$ *faces and a weight function* $w$ *on its edges, we can compute an optimal basis of* $H_1(K)$ *in time* $O(n^{\omega} + n^2 g^{\omega-1})$.

## 5.2 Persistent cycles

In the persistent setting, given a filtration $\mathcal{F}$ and an interval $[b, d]$, can we find an optimal persistent $p$-cycle $c$ that is born at $b$ and dies at $d$.

Sadly, this problem is already known to be NP-hard for $d < \infty$ and $p \geqslant 1$. However, if we assume that $K$ is a weak $(p+1)$-pseudomanifold, i.e., a simplicial complex in which each $p$-simplex is a face of at most 2 $(p+1)$-simplices, then there exists a polynomial-time algorithm, which we will describe in this section.

If we consider cycles that live until $\infty$, we can solve the problem in polynomial time for $p = 1$, but it is NP-hard for $p \geqslant 2$. Here, the assumption of $K$ being a weak $(p+1)$-pseudomanifold does not save us. However, if we further assume that the complex can be embedded in $\mathbb{R}^{p+1}$, then it is again polynomial.

To solve the problem for $d < \infty$ in a weak $(p+1)$-pseudomanifold, we consider undirected flow networks: We have a graph, where every edge has a capacity in $[0, \infty]$, some sources, and some sinks, and we want to find the maximum flow we can send from the sources to the sinks without sending too much flow through any edge. Recall that if we consider a cut which separates the sources from the sinks, the capacity of this cut is an upper bound on the value of the maximum flow. Furthermore, if we consider the minimum such cut, its capacity is equal to the value of the maximum flow. This can be solved in polynomial time.

We can build a dual graph $G$, by placing a vertex into every $(p+1)$-simplex and adding an edge whenever they share a p-simplex. We furthermore add a dummy vertex which gets connected to all vertices which only have one neighbor. We are going to make the vertex belonging to the $(p+1)$-simplex which is the destructor of our desired cycle the source. Furthermore, we make the dummy vertex as well as all vertices belonging to $(p+1)$-simplices added after the destructor into sinks. Edges added at or before the birth are getting the capacity equal to their weight, while all other edges get capacity $\infty$. Then, it turns out that the p-simplices belonging to the edges in a minimum cut separating the sources from the sinks are an optimal persistent cycle.

**Exercise 5.7.** *Consider a simplex-wise filtration on a simplicial complex that is a weak* $(p+1)$*-pseudomanifold, and consider some interval* $[b, d]$ *(for* $d < \infty$*) such that there exists a* $p$*-cycle born at* $b$ *and dying at* $d$*. We look at the dual graph* $G$ *with source and sinks defined as in the lecture. Consider a cut with finite capacity that separates the source from the sinks. Let* $c$ *be the chain corresponding to the* $p$*-simplices dual to the edges going over this cut. Show that* $c$ *is a* $p$*-cycle born at* $b$ *and dying at* $d$*, and show that its weight is equal to the capacity of the cut.*

This exercise proves one direction of the correctness of the algorithm described above. The other direction is similar. We get the following result.

**Theorem 5.8.** *Given a a simplex-wise filtration on a simplicial complex that is a weak* $(p+1)$*-pseudomanifold and an interval* $[b, d]$ *(for* $d < \infty$*), we can compute an optimal* $p$*-cycle born at* $b$ *and dying at* $d$ *in polynomial time.*

For details, we refer to Chapter 5 in the book of Dey and Wang [1].

## Questions

32. *How can we compute an optimal basis given a set of cycles that contain one?* Explain the algorithm described in Section 5.1. Further, explain annotations and how they can be used to check linear independence.
33. *How can we compute a set of 1-cycles that contain an optimal basis of* $H_1$*?* Describe the algorithm to do this and prove its correctness.
34. *How can we compute an optimal persistent cycle?* Explain the algorithm described in Section 5.2.

## References

[1] Tamal Krishna Dey and Yusu Wang, *Computational topology for data analysis*, Cambridge University Press, 2022.