EMAp summer course,

26^{th} January ~ 10^{th} February, 2021

Topological Data Analysis with Persistent Homology

Raphaël Tinarrage

raphael.tinarrage@fgv.br

Last update: 10^{th} February, 2021

Abstract. This course is intended for a 3^{rd} year graduate student with no background on topology. The present document is a collection of notes for each lesson.

Course webpage. Various information (schedule, homework) are gathered on https://raphaeltinarrage.github.io/EMAp.html.

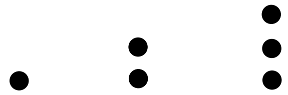
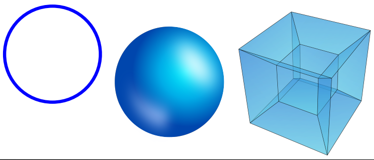
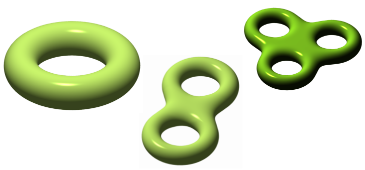
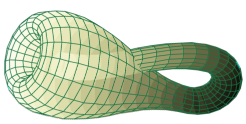
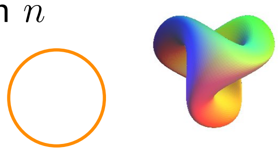
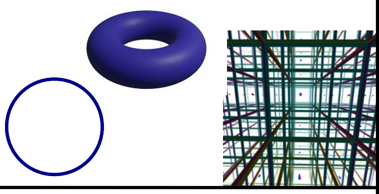
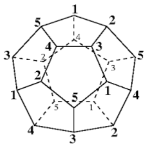
Numerical experiments. Python notebooks containing illustrations can be found at https://github.com/raphaeltinarrage/EMAp.
Before the first tutorial (4^{th} lesson), you should be able to run the following notebook: https://github.com/raphaeltinarrage/EMAp/blob/main/Tutorial0.ipynb.

Homework. Exercises with a vertical segment next to them are your homework. Here is the first one:

> **Exercise 0.** Send me an email answering the following questions:
>
> - Do you understand English well?
> - Have you ever studied topology?
> - Have you ever coded? In which language?
> - Any remarks?

Warning. I took some shortcuts in the exposition of persistent homology. Notably: we won't study basic general topology notion that are worth it (adherence, compactness, path-connectedness). We will not study singular homology, but rather define the homology of topological spaces via the simplicial homology of triangulations, and only with coefficients in the finite field $\mathbb{Z}/2\mathbb{Z}$. Concerning persistent homology, we will restrict ourselves to simplicial filtrations.

Contents

Topological space	Homology groups over $\mathbb{Z}/2\mathbb{Z}$ (only nonzero)	Homology groups over $\mathbb{Z}/3\mathbb{Z}$ (only nonzero)	Homology groups over $\mathbb{Z}$ (only nonzero)
n points 	$H_0 = (\mathbb{Z}/2\mathbb{Z})^n$	$H_0 = (\mathbb{Z}/3\mathbb{Z})^n$	$H_0 = \mathbb{Z}^n$
Bouquet of n circles 	$H_0 = \mathbb{Z}/2\mathbb{Z}$ $H_1 = (\mathbb{Z}/2\mathbb{Z})^n$	$H_0 = \mathbb{Z}/3\mathbb{Z}$ $H_1 = (\mathbb{Z}/3\mathbb{Z})^n$	$H_0 = \mathbb{Z}$ $H_1 = \mathbb{Z}^n$
Sphere of dim n 	$H_0 = \mathbb{Z}/2\mathbb{Z}$ $H_n = \mathbb{Z}/2\mathbb{Z}$	$H_0 = \mathbb{Z}/3\mathbb{Z}$ $H_n = \mathbb{Z}/3\mathbb{Z}$	$H_0 = \mathbb{Z}$ $H_n = \mathbb{Z}$
g-holed torus 	$H_0 = \mathbb{Z}/2\mathbb{Z}$ $H_1 = (\mathbb{Z}/2\mathbb{Z})^{2g}$ $H_2 = (\mathbb{Z}/2\mathbb{Z})^g$	$H_0 = \mathbb{Z}/3\mathbb{Z}$ $H_1 = (\mathbb{Z}/3\mathbb{Z})^{2g}$ $H_2 = (\mathbb{Z}/3\mathbb{Z})^g$	$H_0 = \mathbb{Z}$ $H_1 = \mathbb{Z}^{2g}$ $H_2 = \mathbb{Z}^g$
Klein bottle 	$H_0 = \mathbb{Z}/2\mathbb{Z}$ $H_1 = (\mathbb{Z}/2\mathbb{Z})^2$ $H_2 = \mathbb{Z}/2\mathbb{Z}$	$H_0 = \mathbb{Z}/3\mathbb{Z}$ $H_1 = \mathbb{Z}/3\mathbb{Z}$	$H_0 = \mathbb{Z}$ $H_1 = \mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$
Projective space of dim n 	$\forall i \leq n,$ $H_i = \mathbb{Z}/2\mathbb{Z}$	$H_0 = \mathbb{Z}/3\mathbb{Z}$ $H_n = \mathbb{Z}/3\mathbb{Z}$ if n odd	$H_0 = \mathbb{Z}$ $H_n = \mathbb{Z}$ if n odd $H_i = \mathbb{Z}/2\mathbb{Z}$, $0 < i < n$, i odd
Torus of dim n 	$\forall i \leq n,$ $H_i = (\mathbb{Z}/2\mathbb{Z})^{\binom{n}{i}}$	$\forall i \leq n,$ $H_i = (\mathbb{Z}/3\mathbb{Z})^{\binom{n}{i}}$	$\forall i \leq n,$ $H_i = \mathbb{Z}^{\binom{n}{i}}$
Poincaré dodecahedral space 	$H_0 = \mathbb{Z}/2\mathbb{Z}$ $H_3 = \mathbb{Z}/2\mathbb{Z}$	$H_0 = \mathbb{Z}/3\mathbb{Z}$ $H_3 = \mathbb{Z}/3\mathbb{Z}$	$H_0 = \mathbb{Z}$ $H_3 = \mathbb{Z}$

1 General topology

1.1 Topological spaces

Topological spaces are abstractions of the concept of 'shape' or 'geometric object'.

Definition 1.1. A *topological space* is a pair $(X, \mathcal{T})$ where X is a set and $\mathcal{T}$ is a collection of subsets of X such that:

- $\emptyset \in \mathcal{T}$ and $X \in \mathcal{T}$,
- for every infinite collection $\{O_\alpha\}_{\alpha \in A} \subset \mathcal{T}$, we have $\bigcup_{\alpha \in A} O_\alpha \in \mathcal{T}$,
- for every finite collection $\{O_i\}_{1 \leq i \leq n} \subset \mathcal{T}$, we have $\bigcap_{1 \leq i \leq n} O_i \in \mathcal{T}$.

The set $\mathcal{T}$ is called a *topology* on X. The elements of $\mathcal{T}$ are called the *open sets*. In other words, the previous definition says that:

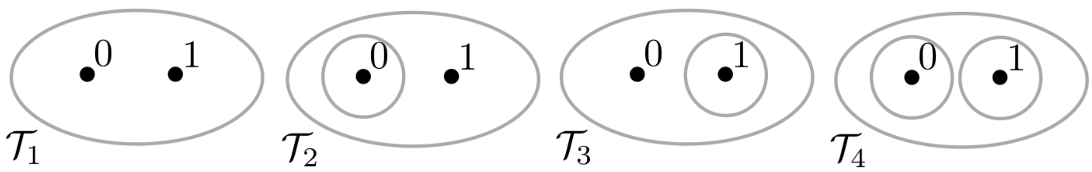
- the empty set is an open set, the set X itself is an open set,
- an infinite union of open sets is an open set,
- a finite intersection of open sets is an open set.

Note that the following is also true: an finite union of open sets is an open set.

Example 1.2. Let $X = \{0\}$ be a set with one element. There exists only one topology on X: $\mathcal{T} = \{\emptyset, \{0\}\}$.

Example 1.3. Let $X = \{0, 1\}$ be a set with two elements. There exists only four different topologies on X:

- $\mathcal{T}_1 = \{\emptyset, \{0, 1\}\}$,
- $\mathcal{T}_2 = \{\emptyset, \{0\}, \{0, 1\}\}$,
- $\mathcal{T}_3 = \{\emptyset, \{1\}, \{0, 1\}\}$,
- $\mathcal{T}_4 = \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$.



Example 1.4. Let $X = \{0, 1, 2\}$ be a set with three elements. The set

$$\mathcal{T} = \{\emptyset\}$$

is not a topology on X because the whole set $X = \{1, 2, 3\}$ does not belong to $\mathcal{T}$. Likewise, the set

$$\mathcal{T} = \{\emptyset, \{0\}, \{1\}, \{0, 1, 2\}\}$$

is not a topology on X because the finite union $\{0\} \cup \{1\} = \{0, 1\}$ does not belong to $\mathcal{T}$.

Exercise 1. Let $X = \{0, 1, 2\}$ be a set with three elements. What are the different topologies that X admits?
Hint: There are 29 of them.

Exercise 2. Let $\mathbb{Z}$ be the set of integers. Consider the *cofinite topology* $\mathcal{T}$ on $\mathbb{Z}$, defined as follows: a subset $O \subset \mathbb{Z}$ is an open set if and only if $O = \emptyset$ or ${}^cO$ is finite. Here, ${}^cO = \{x \in \mathbb{Z}, x \notin O\}$ represents the complementary of O in $\mathbb{Z}$.

1. Show that $\mathcal{T}$ is a topology on $\mathbb{Z}$.
2. Exhibit an sequence of open sets $\{O_n\}_{n\in\mathbb{N}} \subset \mathcal{T}$ such that $\bigcap\limits_{n\in\mathbb{N}} O_n$ is not an open set.

Conclusion: In general, in a given topology, an infinite intersection of open sets may not be open.
To meditate: However, if X is finite, every infinite intersection of open sets is an open set. Indeed, any topology on X must be finite, hence every infinite intersection of open sets must actually be a finite intersection.

Example 1.5. The set
$$\mathcal{T} = \big\{\emptyset, \mathbb{R}\big\} \cup \big\{[0, a], a > 0\big\}$$
is not a topology on $\mathbb{R}$. Indeed, the following union of open sets is not an open set:
$$\bigcup_{a>0} [0, a] = [0, +\infty).$$

Another fondamental object of topological spaces is the following:

> **Definition 1.6.** Let $(X, \mathcal{T})$ be a topological space. For every open set $O \in \mathcal{T}$, its complementary ${}^cO = \{x \in X, x \notin O\}$ is called a *closed set.*

We can deduce the following fact: **a subset $P \subset X$ is closed if and only if ${}^cP$ is open**. Indeed, a set P is closed if there exists an open set O such that $P = {}^cO$. Using the relation ${}^c({}^cO) = O$, we obtain ${}^cP = O$.

Proposition 1.7. *We have:*

- *the sets $\emptyset$ and X are closed sets,*
- *for every infinite collection $\{P_\alpha\}_{\alpha\in A}$ of closed set, $\bigcap\limits_{\alpha\in A} P_\alpha$ is a closed set,*
- *for every finite collection $\{P_i\}_{1\le i\le n}$ of closed sets, $\bigcup\limits_{1\le i\le n} P_i$ is a closed set.*

Proof. Proof of first point: The set $\emptyset$ is closed because ${}^c\emptyset = X$ is open. The set X is closed because ${}^cX = \emptyset$ is open.

Proof of second point: If $\{P_\alpha\}_{\alpha\in A}$ is an infinite collection of closed set, then for every $\alpha \in A$, ${}^cP_\alpha$ is open. Now, we use the relation
$${}^c\left(\bigcap_{\alpha\in A} P_\alpha\right) = \bigcup_{\alpha\in A} {}^cP_\alpha.$$

This is a union of open sets, hence it is open. Hence $\bigcap_{\alpha \in A} P_\alpha$ is closed.

Proof of third point: If $\{P_i\}_{1 \leq i \leq n}$ is a finite collection of closed set, then for every $i \in [\![1, n]\!]$, ${}^c P_i$ is open. Now, we use the relation

$$^c\left(\bigcup_{1 \leq i \leq n} P_i\right) = \bigcap_{1 \leq i \leq n} {}^c P_i.$$

This is a *finite* intersection of open sets, hence it is open. Hence $\bigcup_{1 \leq i \leq n} P_i$ is closed. □

1.2 Topology of $\mathbb{R}^n$

The study of general topological spaces is wild. In this course, we will mainly consider topological spaces that are sub-spaces of the spaces $\mathbb{R}^n$, $n \geq 0$. On $\mathbb{R}^n$, we will always consider the *Euclidean topology.*

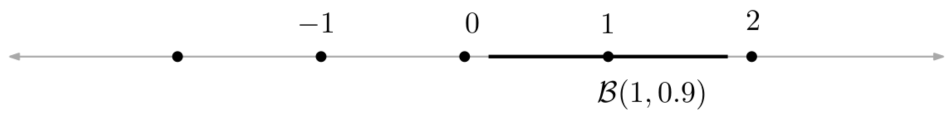
In order to define this topology, we will use open balls. Remind that the Euclidean metric on $\mathbb{R}^n$ is defined for all $x = (x_1, ..., x_n) \in \mathbb{R}^n$ as:

$$\|x\| = \sqrt{x_1^2 + ... + x_n^2}.$$

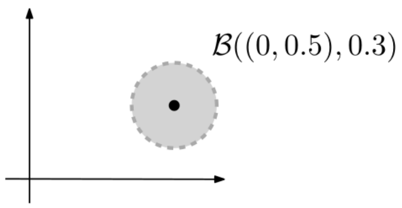
Definition 1.8. Let $x \in \mathbb{R}^n$ and $r > 0$. The *open ball* of center x and radius r, denoted $\mathcal{B}(x, r)$, is defined as:

$$\mathcal{B}(x, r) = \{y \in \mathbb{R}^n, \|x - y\| < r\}.$$

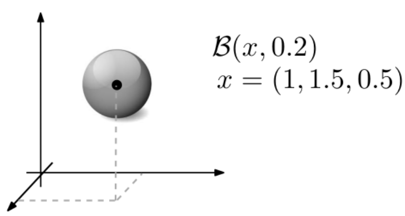
In $\mathbb{R}$:



In $\mathbb{R}^2$:

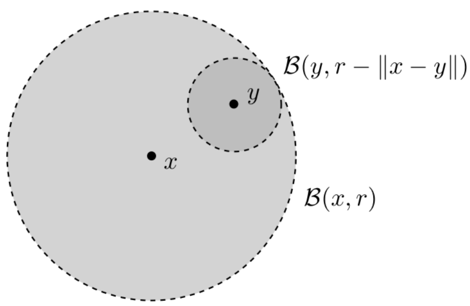


In $\mathbb{R}^3$:



Exercise 3. Let $x \in \mathbb{R}^n$, and $r > 0$. Let $y \in \mathcal{B}(x, r)$. Show that

$$\mathcal{B}(y, \|x - y\|) \subset \mathcal{B}(x, r - \|x - y\|).$$



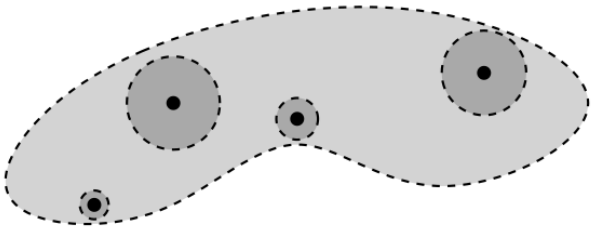
Exercise 4. Let $x, y \in \mathbb{R}^n$, and $r = \|x - y\|$. Show that

$$\mathcal{B}\left(\frac{x+y}{2}, \frac{r}{2}\right) \subset \mathcal{B}(x, r) \cap \mathcal{B}(y, r).$$

Now we can define the Euclidean topology on $\mathbb{R}^n$.

Definition 1.9. Let $A \subset \mathbb{R}$ be a subset. Let $x \in A$. We say that A is *open around* x if there exists $r > 0$ such that $\mathcal{B}(x, r) \subset A$. We say that A is *open* if for every $x \in A$, A is open around x.

We denote the set of such open sets by $\mathcal{T}_{\mathbb{R}^n}$.



Proposition 1.10. *$\mathcal{T}_{\mathbb{R}^n}$ is a topology on $\mathbb{R}^n$.*

Proof. We have to check the three axioms of a topological space.
First axiom (the empty set and the set X are open sets).
The set $\emptyset$ is clearly open according to the definition of $\mathcal{T}_{\mathbb{R}^n}$ (indeed, $\emptyset$ contains no point.) The set $\mathbb{R}^n$ also is open: for every $x \in \mathbb{R}^n$, the ball $\mathcal{B}(x, 1)$ is a subset of $\mathbb{R}^n$.

Second axiom (an infinite union of open sets is an open set).
Let $\{O_\alpha\}_{\alpha \in A} \subset \mathcal{T}_{\mathbb{R}^n}$ be a infinite collection of open sets, and define $O = \bigcup_{\alpha \in A} O_\alpha$.
Let $x \in O$. There exists an $\alpha \in A$ such that $x \in O_\alpha$. Since O_α is open, it is open around x, i.e., there exists $r > 0$ such that $\mathcal{B}(x, r) \subset O_\alpha$.

We deduce that $\mathcal{B}(x, r) \subset O$, and that O is open around x. Since this it true for any $x \in O$, we proved that O is open.

Third axiom (a finite intersection of open sets is an open set).
Consider a finite collection $\{O_i\}_{1 \leq i \leq n} \subset \mathcal{T}_{\mathbb{R}^n}$, and define $O = \bigcap_{1 \leq i \leq n} O_i$.
Let $x \in O$. For every $i \in [\![1, n]\!]$, we have $x \in O_i$. Since O_i is open, it is open around x, i.e., there exists $r_i > 0$ such that $\mathcal{B}(x, r_i) \subset O_i$. Define $r_{\min} = \min\{r_1, ...r_n\}$. For every $i \in [\![1, n]\!]$, we have $\mathcal{B}(x, r_{\min}) \subset O_i$.

We deduce that $\mathcal{B}(x, r_{\min}) \subset O$, and that O is open around x. Since this it true for any $x \in O$, we proved that O is open. $\square$

Exercise 5. Show that the open balls $\mathcal{B}(x, r)$ of $\mathbb{R}^n$ are open sets (with respect to the Euclidean topology).
Hint: You may use Exercise 3.

Exercise 6. Consider $X = \mathbb{R}$ endowed with the Euclidean topology. Are the following sets open? Are they closed?

1. $[0, 1]$,
2. $[0, 1)$,
3. $(-\infty, 1)$,
4. the singletons $\{x\}$, $x \in \mathbb{R}$,
5. the rationnals $\mathbb{Q}$.

1.3 Topology of subsets of $\mathbb{R}^n$

Definition 1.11. Let $(X, \mathcal{T})$ be a topological space, and $Y \subset X$ a subset. We define the *subspace topology on* Y as the following set:

$$\mathcal{T}_{|Y} = \{O \cap Y, O \in \mathcal{T}\}.$$

Proposition 1.12. *The set* $\mathcal{T}_{|Y}$ *is a topology on* Y.

Proof. We have to check the three axioms of a topological space.
First axiom (the empty set and the set X are open sets).
The set $\emptyset$ is clearly open for $\mathcal{T}_{|Y}$ because it can be written $\emptyset \cap Y$. The set Y also is open for $\mathcal{T}_{|Y}$ because it can be written $X \cap Y$, and X is open for $\mathcal{T}$.

Second axiom (an infinite union of open sets is an open set).
Let $\{O_\alpha\}_{\alpha \in A} \subset \mathcal{T}_{|Y}$ be a infinite collection of open sets, and define $O = \bigcup_{\alpha \in A} O_\alpha$. By definition of $\mathcal{T}_{|Y}$, for every $\alpha \in A$, there exists O'_α such that $O_\alpha = O'_\alpha \cap Y$. Define $O' = \bigcup_{\alpha \in A} O'_\alpha$. It is an open set for $\mathcal{T}$. We have

$$O = \bigcup_{\alpha \in A} O_\alpha = \bigcup_{\alpha \in A} O'_\alpha \cap Y = \left(\bigcup_{\alpha \in A} O'_\alpha \right) \cap Y = O' \cap Y.$$

Hence $O \in \mathcal{T}_{|Y}$.

Third axiom (a finite intersection of open sets is an open set). Consider a finite collection $\{O_i\}_{1 \leq i \leq n} \subset \mathcal{T}_{\mathbb{R}^n}$, and define $O = \bigcap_{1 \leq i \leq n} O_i$. Just as before, for every $i \in [\![1, n]\!]$, there exists O'_i such that $O_i = O'_i \cap Y$. Define $O' = \bigcup_{1 \leq i \leq n} O'_i$. It is an open set for $\mathcal{T}$. We have

$$O = \bigcap_{1 \leq i \leq n} O_\alpha = \bigcap_{1 \leq i \leq n} O'_\alpha \cap Y = \left(\bigcap_{1 \leq i \leq n} O'_\alpha \right) \cap Y = O' \cap Y.$$

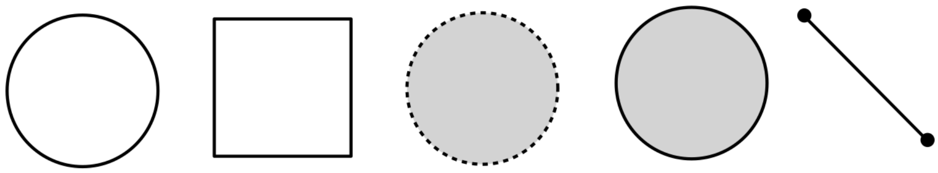
Hence $O \in \mathcal{T}_{|Y}$. ∎

Thanks to the subspace topology, any subset of $\mathbb{R}^n$ inherits a particular topology. This is the only topology we will consider on subsets of $\mathbb{R}^n$.

Among the subsets of $\mathbb{R}^n$ that we will consider, let us list:

- the unit sphere $\mathbb{S}_{n-1} = \{x \in \mathbb{R}^n, \|x\| = 1\}$
- the unit cube $\mathcal{C}_{n-1} = \{x = (x_1, ..., x_n) \in \mathbb{R}^n, \max(|x_1|, ..., |x_n|) = 1\}$
- the open balls $\mathcal{B}\,(x, r) = \{y \in \mathbb{R}^n, \|x - y\| < r\}$
- the closed balls $\overline{\mathcal{B}}\,(x, r) = \{y \in \mathbb{R}^n, \|x - y\| \leq r\}$
- the standard simplex

$$\Delta_{n-1} = \{x = (x_1, ..., x_n) \in \mathbb{R}^n, x_1, ..., x_n \geq 0 \text{ and } x_1 + ... + x_n = 1\}$$



1.4 Continuous maps

The topologist's point of view allows to define the notion of continuity in great generality. In this subsection, we consider two topological spaces $(X, \mathcal{T})$ and $(Y, \mathcal{U})$.

Definition 1.13. Let $f : X \to Y$ be a map. We say that f is *continuous* if for every $O \in \mathcal{U}$, the preimage $f^{-1}(O) = \{x \in X, f(x) \in O\}$ is in $\mathcal{T}$.

In other words, a map is continuous if **the preimage of any open set is an open set**. As shown in the following example, the continuity of a map depends on the topologies that are given to X and Y.

Example 1.14. Let $X = Y = \{0, 1\}$ and $f : \{0, 1\} \to \{0, 1\}$ be the identity map, that is, $f(0) = 0$ and $f(1) = 1$. Let

$$\mathcal{T} = \{\emptyset, \{0, 1\}\} \qquad \text{and} \qquad \mathcal{U} = \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}.$$

The map f, seen as a map between the topological spaces $(X, \mathcal{T})$ and $(Y, \mathcal{U})$, is not continuous. Indeed, $\{0\}$ is an open set of $(Y, \mathcal{U})$, but $f^{-1}(\{0\}) = \{0\}$ is not an open set of $(X, \mathcal{T})$.

However, seen as a map between the topological spaces $(X, \mathcal{U})$ and $(Y, \mathcal{U})$, f is continuous. In particular, $f^{-1}(\{0\}) = \{0\}$ is an open set of $(X, \mathcal{U})$.

Remark 1.15. According to the previous Example, we should not say

$$f : X \to Y \textit{ is continuous},$$

without specifying the topologies on X and Y. We should say

$$f\colon (X,\mathcal{T}) \to (Y,\mathcal{U}) \text{ is continuous.}$$

However, when it will be clear what topologies we are considering, and when there will be no risk of confusion, we will use the first sentence.

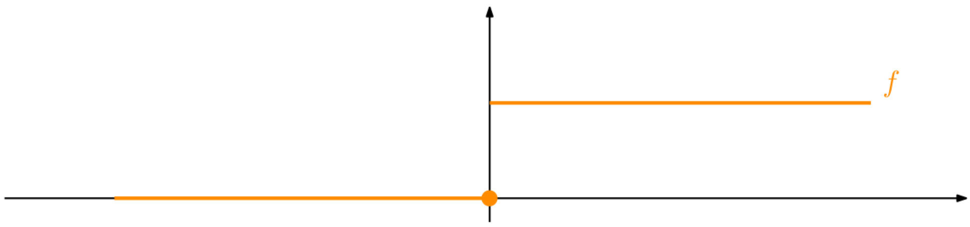
Continuity can also be stated in terms of closed sets:

Proposition 1.16. *A map is continuous if and only if the preimage of closed sets are closed sets.*

Exercise 7. Prove Proposition 1.16.
Hint: For any subset $A \subset Y$, show that $f^{-1}({}^{c}A) = {}^{c}(f^{-1}(A))$.

Example 1.17. Let $X = Y = \mathbb{R}$, endowed with the Euclidean topology. Let $f\colon \mathbb{R} \to \mathbb{R}$ be defined as $f(x) = 0$ for all $x \leq 0$, and $f(x) = 1$ for all $x > 0$.

The set $\{0\}$ is closed, but $f^{-1}(\{0\}) = (-\infty, 0)$ is not. Hence f is not continuous.



Proposition 1.18. *Let $(X,\mathcal{T})$, $(Y,\mathcal{U})$ and $(Z,\mathcal{V})$ be three topological spaces, and $f\colon X \to Y$, $g\colon Y \to Z$ two continuous maps. The composition $g \circ f$, defined as*

$$\begin{aligned} g \circ f\colon X &\longrightarrow Z \\ x &\longmapsto g(f(x)) \end{aligned}$$

is a continuous map.

In other words, we say that the composition of two continuous maps is a continuous map.

Proof. Let $O \in \mathcal{V}$ be an open set of Z. We have to show that $(g \circ f)^{-1}(O)$ is in $\mathcal{T}$. First, note that $(g \circ f)^{-1}(O) = f^{-1}\big(g^{-1}(O)\big)$. Since g is continuous, the set $g^{-1}(O)$ is in $\mathcal{U}$, i.e., it is an open set of Y. But since f is continuous, its preimage $f^{-1}\big(g^{-1}(O)\big)$ also is an open set (of X).

Since this is true for any open set $O \in \mathcal{V}$, we deduce that $g \circ f$ is continuous. $\square$

Link with the usual ϵ-δ calculus. We now investigate what continuity means between the Euclidean spaces $\mathbb{R}^n$. Consider a continuous map $f\colon \mathbb{R}^n \to \mathbb{R}^m$. Let $\epsilon > 0$.

We have seen that the open ball $\mathcal{B}(f(x), \epsilon)$ is an open set of $\mathbb{R}^m$. By continuity of f, the preimage $f^{-1}(\mathcal{B}(f(x), \epsilon))$ is an open set.

Note that x belongs to $f^{-1}(\mathcal{B}(f(x), \epsilon))$. By definition of the Euclidean topology, we have that:

$$f^{-1}(\mathcal{B}(f(x), \epsilon)) \text{ is open around } x.$$

In other words, there exists a $\eta > 0$ such that

$$\mathcal{B}(x, \eta) \subset f^{-1}(\mathcal{B}(f(x), \epsilon)).$$

This is equivalent to

$$\forall y \in \mathcal{B}(x, \eta),\ f(y) \in \mathcal{B}(f(x), \epsilon).$$

We deduce that, for all $y \in \mathbb{R}^n$,

$$\|x - y\| < \eta \implies \|f(x) - f(y)\| < \epsilon.$$

We recognize **the usual definition of continuity**.

Proposition 1.19. *A map $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is continuous if and only if, for every $x \in \mathbb{R}^n$ and $\epsilon > 0$, there exists $\eta > 0$ such that for all $y \in \mathbb{R}^n$,*

$$\|x - y\| < \eta \implies \|f(x) - f(y)\| < \epsilon.$$

Remark 1.20. As a consequence, what you already know about continuity still applies here.

Moralidade

A topologia geral contém todo o ϵ-δ cálculo, e muito mais.

The following proposition will be useful to study maps between subsets of $\mathbb{R}^n$:

Proposition 1.21. *Let f be a continuous map between $(X, \mathcal{T})$ and $(Y, \mathcal{U})$. Consider a subset $A \subset X$, and endow it with the subspace topology $\mathcal{T}_{|A}$. The induced map*

$$f_{|A}\colon (A, \mathcal{T}_{|A}) \to (Y, \mathcal{U})$$

is continuous. Moreover, for any subset $B \subset Y$ such that $f(A) \subset B$, the induced map

$$f_{|A,B}\colon (A, \mathcal{T}_{|A}) \to (B, \mathcal{U}_{|B})$$

also is continuous.

Proof. We will only prove the second statement. For every open set $O \in \mathcal{U}_{|B}$, let us show that $(f_{|A,B})^{-1}(O)$ is in $\mathcal{T}_{|A}$. By definition of $\mathcal{U}_{|B}$, there exists $O' \in \mathcal{U}$ such that $O = O' \cap B$. Now, we have

$$(f_{|A,B})^{-1}(O) = (f_{|A,B})^{-1}(O' \cap B) = (f_{|A,B})^{-1}(O') \cap (f_{|A,B})^{-1}(B).$$

Because of the assumption $f(A) \subset B$, we have $(f_{|A,B})^{-1}(B) = A$, and we deduce

$$(f_{|A,B})^{-1}(O) = (f_{|A,B})^{-1}(O') \cap A.$$

Since f is continuous, the preimage $(f_{|A,B})^{-1}(O')$ is in $\mathcal{T}$, hence the intersection $(f_{|A,B})^{-1}(O') \cap A$ is in $\mathcal{T}_{|A}$. $\square$

Example 1.22. For any $\lambda > 0$ and $v \in \mathbb{R}^n$, we already know that the following map is continuous:

$$\begin{aligned} f\colon \mathbb{R}^n &\longrightarrow \mathbb{R}^n \\ x &\longmapsto \lambda x + v \end{aligned}$$

As a consequence, the restricted map $f_{|\mathcal{B}(0,1),\mathcal{B}(v,\lambda)}\colon \mathcal{B}(0,1) \to \mathcal{B}(v,\lambda)$, seen between subspaces of $\mathbb{R}^n$ endowed with the subspace topology, is continuous.

2 Homeomorphisms

2.1 Definition

Definition 2.1. Let $(X,\mathcal{T})$ and $(Y,\mathcal{U})$ be two topological spaces, and $f\colon X \to Y$ a map. We say that f is a *homeomorphism* if

- f is a bijection,
- $f\colon X \to Y$ is continuous,
- $f^{-1}\colon Y \to X$ is continuous.

If there exists such a homeomorphism, we say that the two topological spaces are *homeomorphic*.

Remark 2.2. In practice, finding the inverse f^{-1} of f consists in finding a map $g\colon Y \to X$ such that

$$g \circ f = \mathrm{id} \qquad \text{and} \qquad f \circ g = \mathrm{id}.$$

In this case, g is the inverse of f.

Example 2.3. Consider the following circles of $\mathbb{R}^2$:

$$\begin{aligned} \mathbb{S}(0,1) &= \{x \in \mathbb{R}^2, \|x\| = 1\}, \\ \mathbb{S}(0,2) &= \{x \in \mathbb{R}^2, \|x\| = 2\} \end{aligned}$$

and the map

$$\begin{aligned} f\colon \mathbb{S}(0,1) &\longrightarrow \mathbb{S}(0,2) \\ x &\longmapsto 2x \end{aligned}$$

It is, bijective, and its inverse $f^{-1}\colon x \mapsto \frac{1}{2}x$ also is continuous. Hence f is a homeomorphism.

Example 2.4. Still in $\mathbb{R}^2$, consider a circle and a square:

$$\mathbb{S}\,(0,1) = \{x \in \mathbb{R}^2, \|X\| = 1\},$$
$$\mathcal{C} = \left\{(x_1, x_2) \in \mathbb{R}^2,\ \max(|x_1|, |x_2|) = 1\right\}.$$

Let $f\colon \mathbb{S}\,(0,1) \to \mathcal{C}$ be the map

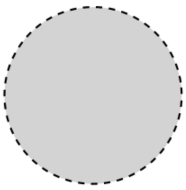
$$f\colon (x_1, x_2) \mapsto \frac{1}{\max(|x_1|, |x_2|)}(x_1, x_2).$$

It is continuous. More over, it admits the following inverse *(check that this is true)*:

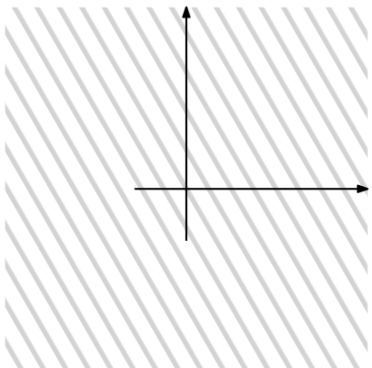
$$f^{-1}\colon x \mapsto \frac{1}{\sqrt{x_1^2 + x_2^2}}(x_1, x_2).$$

This map is continuous, hence f is a homeomorphism.

Exercise 8. Show that the topological spaces $\mathbb{R}^n$ and $\mathcal{B}\,(0,1) \subset \mathbb{R}^n$ are homeomorphic.



$\simeq$



Hint: Consider the map $f\colon x \mapsto \frac{\|x\|}{(\|x\|+1)^2}x$.
Better hint: Consider the map $f\colon x \mapsto \frac{1}{\|x\|+1}x$.

Exercise 9. Show that $\mathcal{B}\,(x,r)$ and $\mathcal{B}\,(y,s)$ are homeomorphic.

Exercise 10. Show that $\mathbb{S}\,(0,1)$, the unit circle of $\mathbb{R}^2$, is homeomorphic to the ellipse

$$\mathcal{S}(a,b) = \left\{(x_1, x_2) \in \mathbb{R}^2, \left(\frac{x_1}{a}\right)^2 + \left(\frac{x_2}{b}\right)^2 = 1\right\}$$

for any $a, b > 0$.

Example 2.5. Let $\mathbb{S}(0,1)$ denote the unit circle of $\mathbb{R}^2$, and consider the map

$$\begin{aligned} f\colon [0,2\pi) &\longrightarrow \mathbb{S}(0,1) \\ \theta &\longmapsto (\cos(\theta), \sin(\theta)) \end{aligned}$$

It is continuous, and admits the following inverse:

$$\begin{aligned} g\colon \mathbb{S}(0,1) &\longrightarrow [0,2\pi) \\ (x_1,x_2) &\longmapsto \arctan\left(\frac{x_2}{x_1}\right) \end{aligned}$$

This comes from the relation $\theta = \arctan\left(\frac{\sin(\theta)}{\cos(\theta)}\right)$ for all $\theta \in [0,2\pi)$.

The map g is **not** continuous. Indeed, $[0,\pi)$ is an open subset of $[0,2\pi)$, but $g^{-1}([0,\pi))$ is not an open subset of $\mathbb{S}(0,1)$ (it is not open around $g^{-1}(0) = (1,0)$).

We will see in Example 2.16 that there exists no homeomorphism between $[0,2\pi)$ and $\mathbb{S}(0,1)$.

Homeomorphism is an equivalence relation. Let us write $X \simeq Y$ if the two topological spaces X and Y are homeomorphic, i.e., if there exists a homeomorphism $f\colon X \to Y$. It is clear that, for any X, we have

$$X \simeq X.$$

Moreover, we have *(mental exercise)*:

$$X \simeq Y \iff Y \simeq X.$$

We also have a third property, stated in the following proposition:

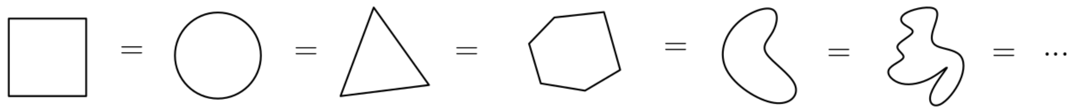
Proposition 2.6. *If three topological spaces X, Y, Z are such that X is homeomorphic to Y and Y is homeomorphic to Z, then X is homeomorphic to Z. In other words,*

$$X \simeq Y \text{ and } Y \simeq Z \implies X \simeq Z.$$

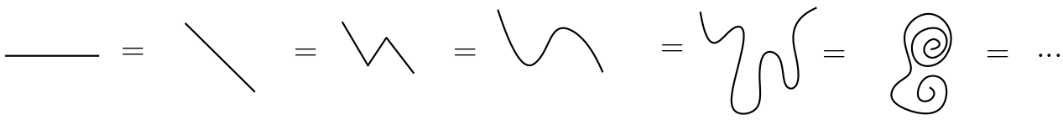
Proof. Supppose that X, Y are homeomorphic, and Y, Z too. This means that we have homeomorphisms $f\colon X \to Y$ and $g\colon Y \to Z$. Consider the map $g \circ f\colon X \to Z$. It is continuous (by Proposition 1.18) bijective (composition of bijective maps) and its inverse $f^{-1} \circ g^{-1}\colon Z \to X$ is also continuous (by Proposition 1.18 too). Hence $g \circ f$ is a homeomorphism, and the spaces X, Z are homeomorphic. $\square$

The three previous properties are called respectively *reflexivity*, *symmetry* and *transitivity*. Hence **being homeomorphic** is what we call an **equivalence relation**. It allows to classify topological spaces in classes (called *classes of homeomorphism equivalence*):

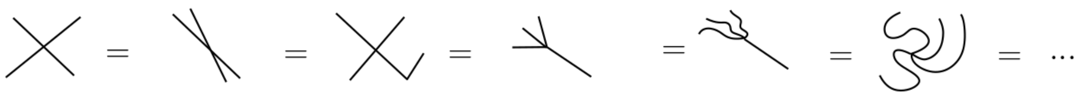
- the class of circles:



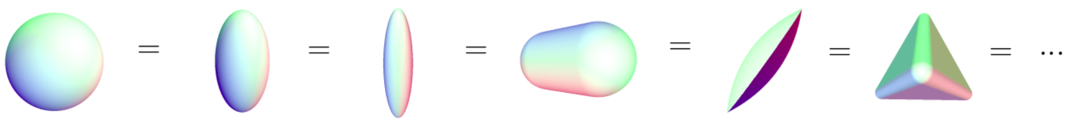
- the class of intervals:



- the class of crosses:



- the class of spheres of dimension 2:



- the class of torii, the class of Klein bottles, etc...

Moralidade
Para um·a topólogisto·a, dois espaços topológicos homeomórficos são o mesmo.

In general, it may be complicated to determine whether two topological spaces are homeomorphic. To answer this problem, we will use the notion of *invariant*. An invariant is a property, a characteristic, that is shared by all the topological space of a same class. Our first example will be connectedness.

2.2 Connected components

Definition 2.7. Let $(X, \mathcal{T})$ be a topological space. We say that X is *connected* if for every open sets $O, O' \in \mathcal{T}$ such that $O \cap O' = \emptyset$ (i.e., they are disjoint), we have

$$X = O \cup O' \implies O = \emptyset \text{ or } O' = \emptyset.$$

In other words, a connected topological space cannot be divided into two non-empty disjoint open sets.

One shows that a connected topological space cannot be divided into two non-empty disjoint **closed** sets.

Example 2.8. The subset $X = [0,1] \cup [2,3]$ of $\mathbb{R}$, endowed with the subspace topology, is not connected. Indeed, its subsets $[0,1]$ and $[2,3]$ are open disjoint non-empty sets that covers X.

We will accept the following result without proving it:

Proposition 2.9. *The balls of $\mathbb{R}^n$ are connected. More generally, any convex set is connected.*

If a space is not connected, we can consider its connected components. Let $x \in X$. The connected component of x is defined as the largest subset of X that is connected. The set of connected components of X forms a partition of X into **open** sets. Moreover, if there are only finitely many connected components, they are also **closed**.

Definition 2.10. Let $(X, \mathcal{T})$ be a topological space. Suppose that there exists a collection of n **non-empty**, **disjoint** and **connected** open sets $(O_1, ..., O_n)$ such that

$$\bigcup_{1 \leq i \leq n} O_i = X.$$

Then we say that X admits n connected components.

Remark 2.11. One shows that if there exists a collection of n **non-empty** and **disjoint** sets $(O_1, ..., O_n)$ such that

$$\bigcup_{1 \leq i \leq n} O_i = X,$$

then X admits at least n connected components.

Example 2.12. Consider the subset $X = \{0,1,2,3,4,5,6,7,8,9,10\}$ of $\mathbb{R}$. Each of its subsets $\{i\}$, $i \in X$, are open. They are all non-empty, connected and disjoint. Hence X admits ten connected components.

Lemma 2.13. *Let $f : X \to Y$ be a continuous map and O a connected component of X. Then $f(O) \subset Y$ is connected.*

Proof. Denote $O' = f(O)$. We will apply the definition of a connected topological space.

Suppose that there exists two disjoint open sets A, A' of Y such that $O' = A \cup A'$. The preimages $f^{-1}(A)$ and $f^{-1}(A')$ are disjoint open sets of X. Moreover,

$$O \subset f^{-1}(O') = f^{-1}(A \cup A') = f^{-1}(A) \cup f^{-1}(A').$$

Since O is connected, we deduce that $f^{-1}(A) = \emptyset$ or $f^{-1}(A') = \emptyset$. Therefore, $A = \emptyset$ or $A' = \emptyset$. This shows that O' is connected. $\square$

2.3 Connectedness as an invariant

Proposition 2.14. *Two homeomorphic topological spaces admit the same number of connected components.*

Proof. Let $f\colon X \to Y$ be a homeomorphism. Let n be the number of connected components of Y, and m the number of X. Let us show that $m = n$.

Suppose that Y admits n connected components. We can write $Y = \bigcup_{1 \leq i \leq n} O_i$ where the O_i are disjoint non-empty connected sets. Also, we have seen that the O_i are open. For all $i \in [\![1, n]\!]$, define $O'_i = f^{-1}(O_i)$. We have:

- for all $i \in [\![1, n]\!]$ $O'_i = f^{-1}(O_i)$ is open (because f is continuous),
- $X = \bigcup_{1 \leq i \leq n} O'_i$ (because f is a map)
- for all $i, j \in [\![1, n]\!]$ with $i \neq j$, $O'_i \cap O'_j = f^{-1}(O_i) \cap f^{-1}(O_j) = f^{-1}(O_i \cap O_j) = \emptyset$
- for all $i \in [\![1, n]\!]$, $O'_i = f^{-1}(O_i) \neq \emptyset$ (because f is a bijection).

Hence X can be covered by n disjoint non-empty open sets. Using Remark 2.11, we deduce that X admits at least n connected components.

Now, suppose that X admits m connected components. Using the same reasoning, one shows that Y admits at least m connected components. Hence we have $n \geq m \geq n$, that is, $n = m$. □

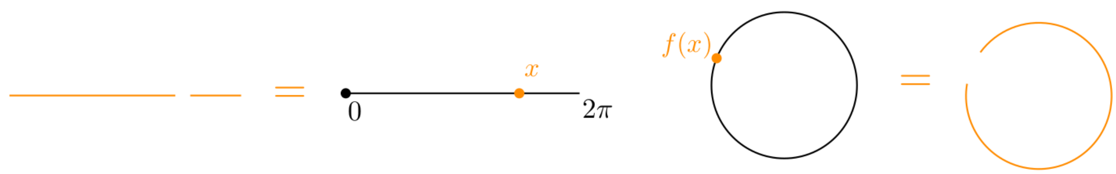
Example 2.15. The subsets $[0, 1]$ and $[0, 1] \cup [2, 3]$ of $\mathbb{R}$ are not homeomorphic. Indeed, the first one has one connected component, and the second one two.

Example 2.16. The interval $[0, 2\pi)$ and the circle $\mathbb{S}(0, 1) \subset \mathbb{R}^2$ are not homeomorphic. We will prove this by contradiction. Suppose that they are homeomorphic. By definition, this means that there exists a map $f\colon [0, 2\pi) \to \mathbb{S}(0, 1)$ which is continuous, inversible, and with continuous inverse.

Let $x \in [0, 2\pi)$ such that $x \neq 0$. Consider the subsets $[0, 2\pi) \setminus \{x\} \subset [0, 2\pi)$ and $\mathbb{S}(0, 1) \setminus \{f(x)\} \subset \mathbb{S}(0, 1)$, and the induced map

$$g\colon [0, 2\pi) \setminus \{x\} \to \mathbb{S}(0, 1) \setminus \{f(x)\}.$$

The map g is a homeomorphism. Moreover, it is clear that $[0, 2\pi) \setminus \{x\}$ has two connected components, and $\mathbb{S}(0, 1) \setminus \{f(x)\}$ only one. This contradicts Proposition 2.14.



Example 2.17. $\mathbb{R}$ and $\mathbb{R}^2$ are not homeomorphic. Just as before, we will prove this by contradiction. Suppose that there exists a homeomorphism $f\colon \mathbb{R} \to \mathbb{R}^2$. Choose any $x \in \mathbb{R}$. The induced map
$$g\colon \mathbb{R} \setminus \{x\} \to \mathbb{R}^2 \setminus \{f(x)\}$$
is still a homeomorphism, but $\mathbb{R} \setminus \{x\}$ has two connected components, while $\mathbb{R}^2 \setminus \{f(x)\}$ has one. This is a contradiction.

The same reasoning shows that $\mathbb{R}$ and $\mathbb{R}^n$ are not homeomorphic either.

Remark 2.18. More generally, the *invariance of domain* is a theorem that says that for every integers m, n such that $m \neq n$, the spaces $\mathbb{R}^n$ and $\mathbb{R}^m$ are not homeomorphic. We will need much more sophisticated tools to prove that (homology of spheres).

Exercise 11. Show that $[0, 1)$ and $(0, 1)$ are not homeomorphic.
Hint: Use the strategy of Examples 2.16 or 2.17.

Remark 2.19. The number of connected components is an example of a topological invariant: if two topological spaces are homeomorphic, they must admit the same number of connected components.

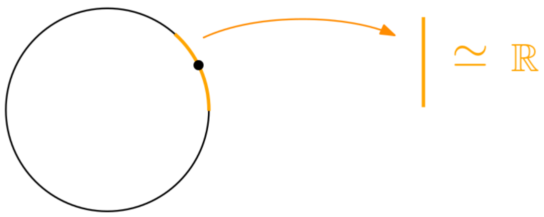
The previous examples show the general morale of a topological invariant: to prove that two spaces are not homeomorphic, prove that their invariant (here, the number of connected components) differ.

2.4 Dimension

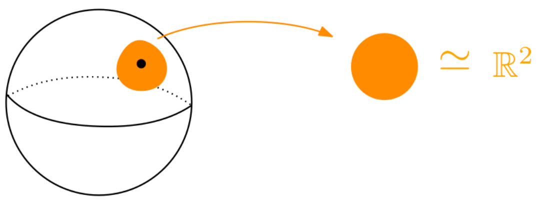
Definition 2.20. Let $(X, \mathcal{T})$ be a topological space, and $n \geq 0$. We say that it *has dimension* n if the following is true: for every $x \in X$, there exists an open set O such that $x \in O$, and a homeomorphism $O \to \mathbb{R}^n$.

In other words, a topological space of dimension n is a topological space that locally looks like the Euclidean space $\mathbb{R}^n$. For instance, one shows that

- the open intervals $(a, b) \subset \mathbb{R}$ have dimension 1,
- the circle $\mathbb{S}_1 \subset \mathbb{R}^2$ has dimension 1,



- more generally, the spheres $\mathbb{S}(v, r) \subset \mathbb{R}^n$ have dimension $n - 1$,



- the open balls $\mathcal{B}(v,r) \subset \mathbb{R}^n$ have dimension n,
- the Euclidean space $\mathbb{R}^n$ itself has dimension n.

Remark 2.21. For this definition to make sense, we have to make sure that the topological spaces $\mathbb{R}^n$, $n \geq 0$, are all not-homeomorphic. Otherwise, a topological space could have several dimensions. As we said earlier, this result, the *invariance of domain*, will be proved later.

Proposition 2.22. *Let X, Y be two homeomorphic topological spaces. If X has dimension n, then Y also has dimension n.*

Proof. Let n be the dimension of X, and consider a homeomorphism $g\colon Y \to X$.

Let $y \in Y$, and $x = g(y)$. Since x has dimension n, there exists an open set O of X, with $x \in O$, and a homeomorphism $h\colon O \to \mathbb{R}^n$.

Define $O' = g^{-1}(O)$. It is an open set of Y, with $y \in O'$. Moreover, the map $h \circ g\colon O' \to \mathbb{R}^n$ is a homeomorphism.

This being true for every $y \in Y$, we deduce that Y has dimension n. □

We can read the previous proposition as follows: dimension is an invariant of homeomorphic spaces. As before, we can use it to show that two spaces are not homeomorphic.

Example 2.23. The unit circle $\mathbb{S}_1 \subset \mathbb{R}^2$ and the unit sphere $\mathbb{S}_2 \subset \mathbb{R}^3$ are not homeomorphic. Indeed, the first one has dimension 1, and the second one dimension 2.

Moralidade
Uma invariante é uma quantidade compartilhada
por todos os espaços topológicos idênticos.

3 Homotopies

3.1 Homotopy equivalence between maps

Definition 3.1. Let $(X,\mathcal{T})$ and $(Y,\mathcal{U})$ be two topological spaces, and $f,g\colon X \to Y$ two continuous maps. A *homotopy* between f and g is a map $F\colon X \times [0,1] \to Y$ such that:

- $F(\cdot,0)$ is equal to f,
- $F(\cdot,1)$ is equal to g,
- $F\colon X \times [0,1] \to Y$ is continuous.

If such a homotopy exists, we say that the maps f and g are *homotopic*.

Remark 3.2. For any $t \in [0,1]$, the notation $F(\cdot,t)$ refers to the map

$$\begin{aligned} F(\cdot,t)\colon X &\longrightarrow Y \\ x &\longmapsto F(x,t) \end{aligned}$$

Remark 3.3. Before asking for $F\colon X \times [0,1] \to Y$ to be continuous, we have to give $X \times [0,1]$ a topology. The topology we choose is the *product topology*.

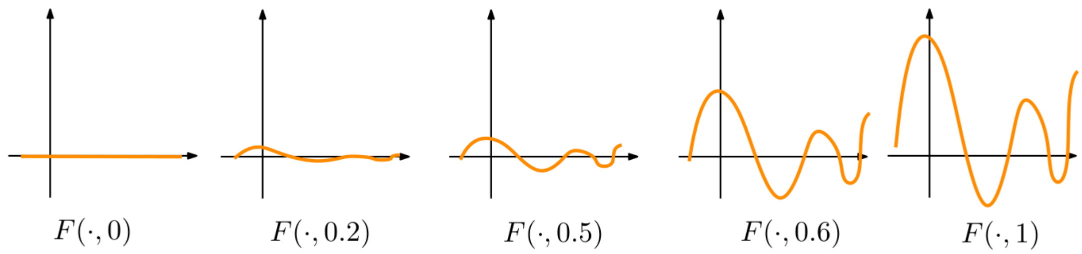
Consider the topological space $(X, \mathcal{T})$, and endow $[0,1]$ with the subspace topology of $\mathbb{R}$, denoted $\mathcal{T}_{|[0,1]}$. The product topology on $X \times [0,1]$, denoted $\mathcal{T} \otimes \mathcal{T}_{|[0,1]}$, is defined as follows: a set $O \subset X \times [0,1]$ is open if and only if it can be written as a union

$$\bigcup_{\alpha \in A} O_\alpha \times O'_\alpha$$

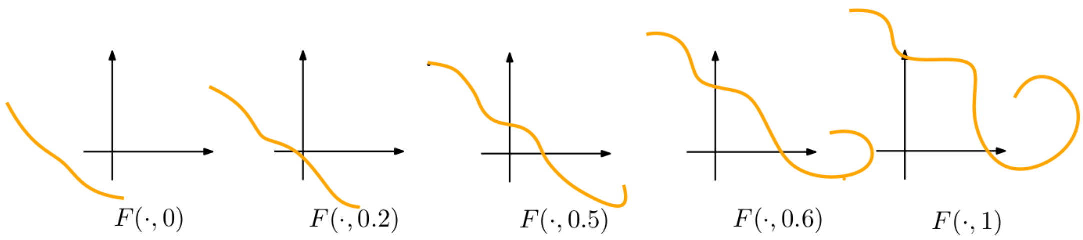
where every O_α is an open set of X and O'_α is an open set of $[0,1]$.

When $(X, \mathcal{T})$ is a subspace of $\mathbb{R}^n$ endowed with the subspace topology, we can describe the product topology in a different way. The product $X \times [0,1]$ can be seen as a subset of $\mathbb{R}^{n+1}$, and one shows that the product topology $\mathcal{T} \otimes \mathcal{T}_{|[0,1]}$ is equal to the subspace topology $\mathcal{T}_{|X \times [0,1]}$.

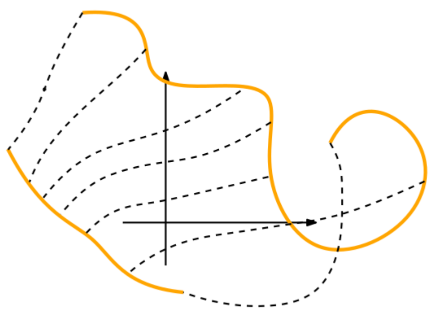
We may represent graphically a homotopy $F\colon \mathbb{R} \times [0,1] \to \mathbb{R}$ by plotting it for each value of $t \in [0,1]$:



This is an example for $F\colon [0,1] \times [0,1] \to \mathbb{R}^2$:



Sometimes we prefer to plot the deformation:



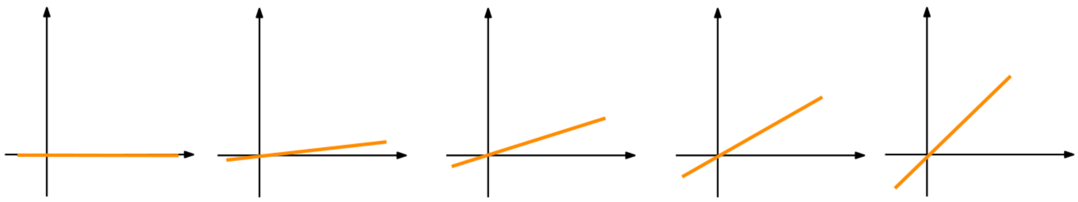
Example 3.4. Let $X = Y = [-1, 1]$ endowed with the Euclidean topology, and consider the maps $f, g \colon X \to Y$ defined as

$$f \colon x \mapsto 0$$
$$g \colon x \mapsto x$$

Let us prove that they are homotopic. Consider the map

$$\begin{aligned} F \colon X \times [0,1] &\longrightarrow Y \\ (x,t) &\longmapsto tx \end{aligned}$$

We see that $F(\cdot, 0) \colon x \mapsto 0$ is equal to f, and $F(\cdot, 1) \colon x \mapsto x$ is equal to g. Moreover, F is continuous. Hence, F is an homotopy between f and g. Thus these two maps are homotopic.



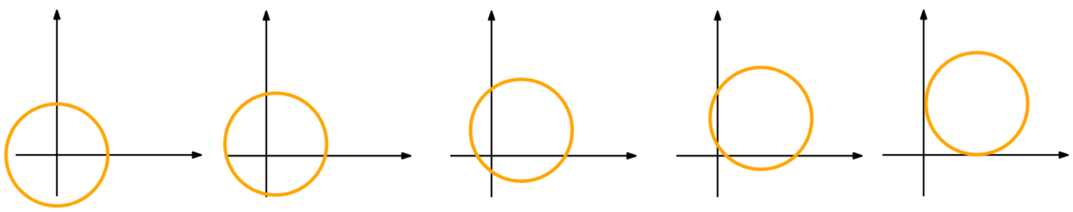
$F(\cdot, 0)$ $F(\cdot, 0.2)$ $F(\cdot, 0.5)$ $F(\cdot, 0.6)$ $F(\cdot, 1)$

Example 3.5. The following map

$$\begin{aligned} F \colon \mathbb{S}_1 \times [0,1] &\longrightarrow \mathbb{R}^2 \\ \theta &\longmapsto (\cos(\theta) + t, \sin(\theta) + t) \end{aligned}$$

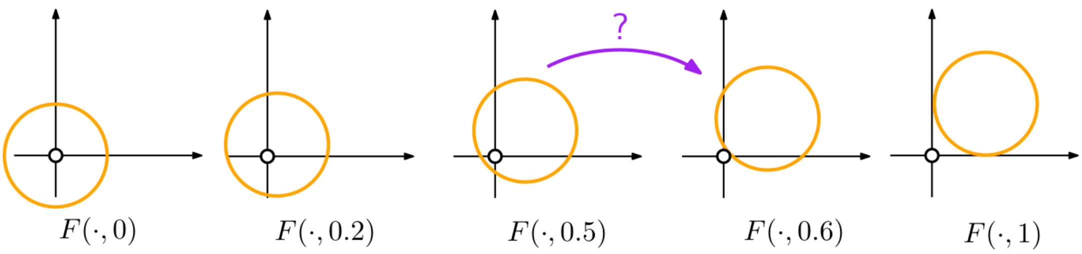
is a homotopy between the maps

$$f \colon \theta \mapsto (\cos(\theta), \sin(\theta)) \qquad \text{and} \qquad g \colon \theta \mapsto (\cos(\theta) + 1, \sin(\theta) + 1)$$



$F(\cdot, 0)$ $F(\cdot, 0.2)$ $F(\cdot, 0.5)$ $F(\cdot, 0.6)$ $F(\cdot, 1)$

Example 3.6. Between $\mathbb{S}_1$ and $\mathbb{R}^2 \setminus \{(0,0)\}$, the plane without the origin, there is no homotopy between the maps f and g of the previous example. Indeed, the homotopy F would pass through the point $(0,0)$ at some point, which is impossible.



We have to wait for the next lessons to prove formally that such a homotopy does not exist.

From a homotopic point a view, a trivial map is a map that is homotopic to a constant map. For instance, the identity map of Example 3.4 is homotopic to the constant map $x \mapsto 0$. More generally, we have:

Proposition 3.7. *Let $f\colon X \to \mathbb{R}^n$ be a continuous map. Then f is homotopic to a constant map.*

Proof. Consider the continuous application

$$\begin{aligned} F\colon X \times [0,1] &\longrightarrow \mathbb{R}^n \\ x &\longmapsto tf(x) \end{aligned}$$

We have that $F(\cdot, 1) = f$, and $F(\cdot, 0)\colon x \mapsto 0$ is a constant map. $\square$

Proposition 3.8. *Let $f\colon \mathbb{R}^n \to X$ be a continuous map. Then f is homotopic to a constant map.*

Exercise 12. Prove the previous proposition.

As a consequence, the theory of maps with domain or codomain $\mathbb{R}^n$ is trivial from a homotopy equivalence perspective. For instance, *knot theory*, the theory that studies maps $\mathbb{S}_1 \to \mathbb{R}^3$, does not exist for us.

However, when the domain and codomain are not Euclidean spaces, as in Example 3.6, many non-homotopic maps may exist.

Moralidade
Para um·a topólogisto·a, duas aplicações homotópicas são a mesma coisa.

Exercise 13. Let $f\colon\ \mathbb{S}_1 \to \mathbb{S}_2$ be a continuous map which is not surjective. Prove that it is homotopic to a constant map.
Hint: Let $x_0 \in \mathbb{S}_2$ be such that $x_0 \notin f(\mathbb{S}_1)$. Find a homotopy between f and the constant map $g\colon x \mapsto -x_0$.
More complicated question: Is every continuous map $f\colon\ \mathbb{S}_1 \to \mathbb{S}_2$ homotopic to a constant map?

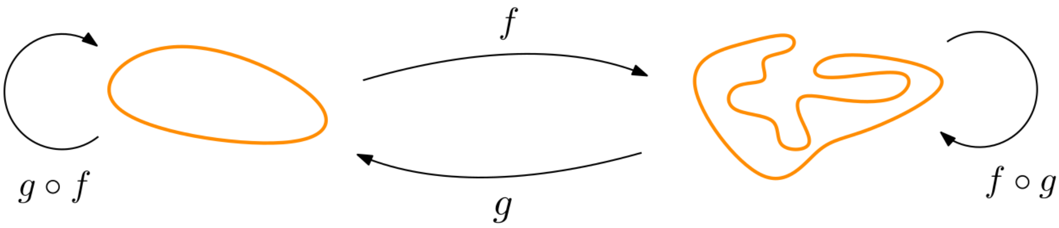
Exercise 14. Show that being homotopic is a *transitive* relation between maps: for every triplet of maps $f, g, h\colon X \to Y$, if f, g are homotopic and g, h are homotopic, then f, h are homotopic.

3.2 Homotopy equivalence between topological spaces

Definition 3.9. Let $(X, \mathcal{T})$ and $(Y, \mathcal{U})$ be two topological spaces. A *homotopy equivalence* between X and Y is a pair of continuous maps $f: X \to Y$ and $g: Y \to X$ such that:

- $g \circ f: X \to X$ is homotopic to the identity map $\mathrm{id}: X \to X$,
- $f \circ g: Y \to Y$ is homotopic to the identity map $\mathrm{id}: Y \to Y$.

If such a homotopy equivalence exists, we say that X and Y are *homotopy equivalent*.



Determining whether two topological spaces are homotopy equivalent may be difficult. When one is a subset of the other, we have a handy tool:

Definition 3.10. Let $(X, \mathcal{T})$ be a topological space and $Y \subset X$ a subset, endowed with the subspace topology $\mathcal{T}_{|Y}$. A *retraction* is a continuous map $r: X \to X$ such that $\forall x \in X, r(x) \in Y$ and $\forall y \in Y, r(y) = y$.

A *deformation retraction* is a homotopy $F: X \times [0, 1] \to Y$ between the identity map $\mathrm{id}: X \to X$ and a retraction $r: X \to X$.

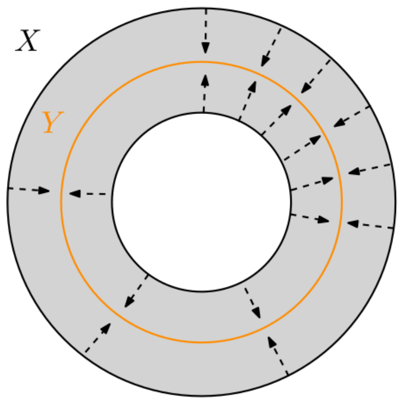
Proposition 3.11. *If a deformation retraction exists, then X and Y are homotopy equivalent.*

Proof. Let $r: X \to X$ denote the retraction, and consider the inclusion map $i: Y \to X$. Note that, since $\forall x \in X, r(x) \in Y$, we can see the retraction r as a map $r: X \to Y$. Let us prove that r, i is a homotopy equivalence.

First, let us prove that $i \circ r: X \to X$ is homotopic to the identity map $\mathrm{id}: X \to X$. This is clear because $i \circ r = r$, and r is homotopic to the identity by definition of a deformation retraction.

Second, let us prove that $r \circ i: Y \to Y$ is homotopic to the identity map $\mathrm{id}: Y \to Y$. This is obvious because $r \circ i = \mathrm{id}$ by definition of a retraction. □

Example 3.12. The circle and the annulus are homotopy equivalent. Indeed, the circle can be seen as a subset of the annulus, and we have a deformation retraction:



Example 3.13. The letter O and the letter Q are homotopy equivalent. Indeed, O can be seen as a subset of Q, and Q deform retracts on it.

Example 3.14. For any $n \geq 1$, the Euclidean space $\mathbb{R}^n$ is homotopy equivalent to the point $\{0\} \subset \mathbb{R}^n$. To prove this, consider the retraction

$$\begin{aligned} r\colon \mathbb{R}^n &\longrightarrow \{0\} \\ x &\longmapsto 0 \end{aligned}$$

It is homotopic to the identity $\mathrm{id}\colon \mathbb{R}^n \to \mathbb{R}^n$ via the deformation retraction

$$\begin{aligned} F\colon \mathbb{R}^n \times [0,1] &\longrightarrow \mathbb{R}^n \\ x &\longmapsto (1-t)x \end{aligned}$$

Indeed, we have $F(\cdot, 0) = \mathrm{id}$ and $F(\cdot, 1) = r$.

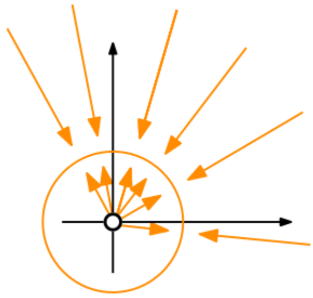
Example 3.15. For any $n \geq 1$, the Euclidean space without origin, $\mathbb{R}^n \setminus \{0\}$, is homotopy equivalent to the sphere $\mathbb{S}(0,1) \subset \mathbb{R}^n$. To prove this, consider the retraction

$$\begin{aligned} r\colon \mathbb{R}^n \setminus \{0\} &\longrightarrow \mathbb{S}(0,1) \\ x &\longmapsto \frac{x}{\|x\|} \end{aligned}$$

It is homotopic to the identity $\mathrm{id}\colon \mathbb{R}^n \setminus \{0\} \to \mathbb{R}^n \setminus \{0\}$ via the deformation retraction

$$\begin{aligned} F\colon (\mathbb{R}^n \setminus \{0\}) \times [0,1] &\longrightarrow \mathbb{R}^n \setminus \{0\} \\ x &\longmapsto \left(1 - t + \frac{t}{\|x\|}\right) x \end{aligned}$$

Indeed, we have $F(\cdot, 0) = \mathrm{id}$ and $F(\cdot, 1) = r$.

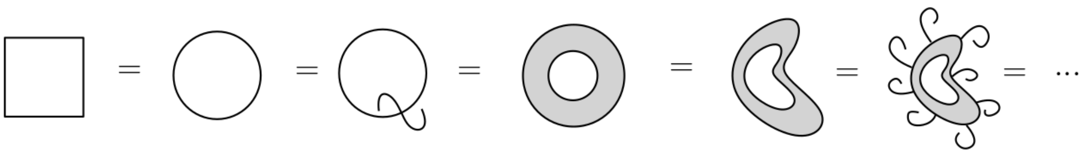


Remark 3.16. Let us denote $X \approx Y$ if the two topological spaces X and Y are homotopy equivalent. Just as for homeomorphic spaces, being homotopy equivalent is an *equivalence relation*. That is:

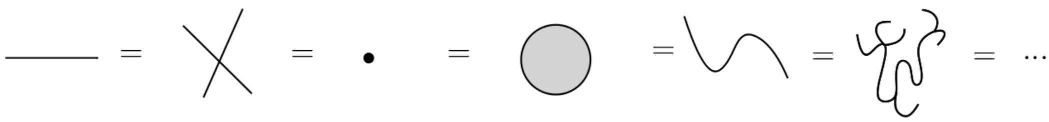
- *(Reflexivity)* $X \approx X$
- *(Symmetry)* $X \approx Y \implies Y \approx X$.
- *(Transitivity)* $X \approx Y$ and $Y \approx Z \implies X \approx Z$.

We can classify topological spaces according to this relation, and obtain *classes of homotopy equivalence*:

- the class of circles:



- the class of points:



- the class of spheres, the class of torii, the class of Klein bottles, etc...

Moralidade
Para um·a topólogisto·a, dois espaços topológicos homotópico-equivalentes são o mesmo.

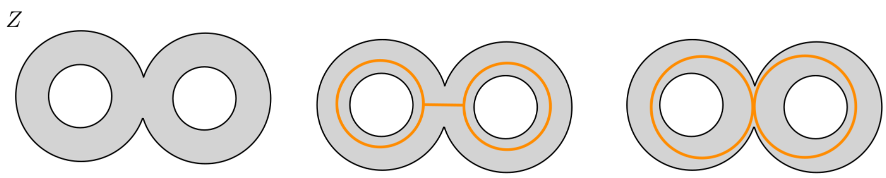
Exercise 15. Show that being homotopy equivalent is an equivalence relation (reflexive, symmetric and transitive).
Hint: You can use Exercise 14.

Remark 3.17. A method to show that two topological spaces X, Y are homotopy equivalent: find a third space Z that contains X, Y and such that there exist a deformation retraction from Z to X and from Z to Y.

If this is the case, we have $X \approx Z$ and $Y \approx Z$, and by using symmetry and transitivity, we deduce $X \approx Y$.

For instance, consider the two following subspaces of $\mathbb{R}^2$:

They are not included one in another. However, the following space contains them, and we see that it deform retracts on both X and Y.



> **Exercise 16.** Classify the letters of the alphabet into homotopy equivalence classes.

3.3 Link with homeomorphic spaces

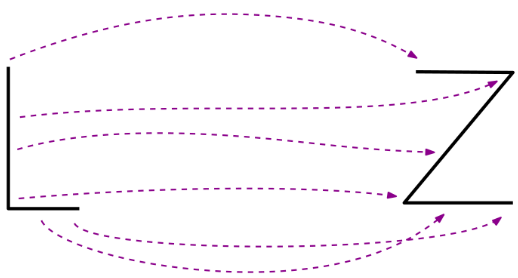
We have studied in the previous lesson another equivalence relation: the relation of homeomorphism. It turns out that it is stronger than the homotopy equivalence relation:

Proposition 3.18. *Let X, Y be two topological spaces. If they are homeomorphic, then they are homotopy equivalent. In other words:*

$$X \simeq Y \implies X \approx Y.$$

As a consequence, in order to prove that two spaces are homotopy equivalent, it is enough to show that they are homeomorphic. However, this strategy does not always work: some spaces are homotopy equivalent but not homeomorphic. This is the case for $\mathbb{R}^n$ and $\{0\}$ for instance.

Example 3.19. The letter L and the letter Z are homeomorphic via the following homeomorphism. Hence they are homotopy equivalent.



3.4 Topological invariants

We now investigate how the invariants *connected components* and *dimension* behave with respect to the homotopy equivalence.

The following result should be compared with Proposition 2.14:

Proposition 3.20. *Two homotopy equivalent topological spaces admit the same number of connected components.*

Proof. Let X, Y be two topological spaces, and $f\colon X \to Y, g\colon Y \to X$ a homotopy equivalence. We will show that f induces a bijection between the connected components of X and Y.

Let $F\colon X \times [0,1] \to X$ be a homotopy between $g \circ f$ and $\mathrm{id}\colon X \to X$. Let $x \in X$, and O the connected component of x. The space $O \times [0,1]$ is connected. Hence its image $F(O \times [0,1]) \subset X$ is connected too (this is Lemma 2.13).

Moreover, $O = F(O \times \{1\}) \subset F(O \times [0,1])$. Hence $F(O \times [0,1])$ is a connected subset of X that contains O, and we deduce that $O = F(O \times [0,1])$. Last, notice that

$$g \circ f(O) = F(O \times \{0\}) \subset F(O \times [0,1]) = O.$$

We can now conclude from the relation $g \circ f(O) \subset O$. Suppose that X admits n connected components $O_1, ..., O_n$, and that Y admits m of them. By contradiction, suppose that $m < n$. This implies that we have two components O_i, O_j such that $f(O_i)$ and $f(O_j)$ are included in the same connected component O' of Y. Hence $g \circ f(O_i)$ and $g \circ f(O_j)$ are included in a common connected component of X. This is absurd because $g \circ f(O_i) \subset O_i$ and $g \circ f(O_j) \subset O_j$.

By exchanging the roles of X and Y in the whole reasonning, we obtain that $m > n$ also is absurd. We deduce that $m = n$. □

In other words, number of connected components is an invariant of homotopy equivalence. As for homeomorphic equivalence, this allows to show that two spaces are not equivalent.

Example 3.21. For any $n, m \geq 0$ such that $n \neq m$, the subspaces $\{1, ..., n\}$ and $\{1, ..., m\}$ of $\mathbb{R}$ are not homotopy equivalent. Indeed, the first one admits n connected components, and the second one m components.

On the other hand, dimension is **not** an invariant of homotopy equivalence. Indeed, some homotopy equivalent spaces have different dimensions. This is the case, for instance, with all the Euclidean spaces $\mathbb{R}^n$, $n \geq 0$. They are all homotopy equivalent by Example 3.14, but all with different dimensions ($\mathbb{R}^n$ has dimension n).

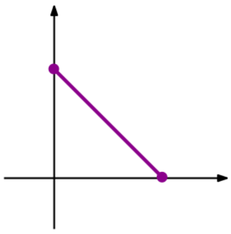
4 Simplicial complexes

4.1 Definition

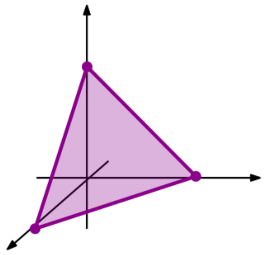
Topological spaces, such as subsets of $\mathbb{R}^n$, may be difficult to deal with on a computer. In order to describe them nicely, we may try to decompose them into simpler pieces. The pieces we shall consider are the *standard simplices*. We recall that the standard simplex of dimension n is the following subset of $\mathbb{R}^{n+1}$

$$\Delta_n = \{x = (x_1, ..., x_{n+1}) \in \mathbb{R}^{n+1},\ x_1, ..., x_{n+1} \geq 0 \text{ and } x_1 + ... + x_{n+1} = 1\}.$$

Δ_0



Δ_1



Δ_2

Remark 4.1. For any collection of points $a_1, ..., a_k \in \mathbb{R}^n$, we define their *convex hull* as:

$$\operatorname{conv}(\{a_1 ... a_k\}) = \left\{ \sum_{1 \leq i \leq k} t_i a_i, \quad t_1 + ... + t_k = 1, \quad t_1, ..., t_k \geq 0 \right\}.$$

Therefore we can say that Δ_n is the convex hull of the vectors $e_1, ..., e_{n+1} \in \mathbb{R}^{n+1}$, where

$$e_i = (0, ..., 1, 0, ..., 0) \qquad (i^{\text{th}} \text{ coordinate } 1, \text{ the other ones } 0).$$

Note that the simplex Δ_n is described by $n+1$ vertices. Let us keep this geometric picture in mind in what follows.

Definition 4.2. Let V be a set (called the set of *vertices*). A *simplicial complex* over V is a set K of subsets of V (called the *simplices*) such that, for every $\sigma \in K$ and every non-empty $\tau \subset \sigma$, we have $\tau \in K$.

By convention, when talking about simplices, we write them with square brackets instead of curly brackets. For instance, the simplex $\{0, 1\}$ will be denoted $[0, 1]$

If $\sigma \in K$ is a simplex, its non-empty subsets $\tau \subset \sigma$ are called *faces* of σ, and σ is called a *coface* of τ. For instance, $[0, 1]$ is a face of $[0, 1, 2]$, and $[0, 1, 2]$ is a coface of $[0, 1]$.

Example 4.3. Let $V = \{0, 1, 2\}$ and

$$K = \{[0], [1], [2], [0, 1], [1, 2], [2, 0]\}.$$

This is a simplicial complex.

Example 4.4. Let $V = \{0, 1, 2\}$ and

$$K = \{[0], [1], [2], [0, 1], [1, 2], [0, 1, 2]\}.$$

This is not a simplicial complex. Indeed, the simplex $[0, 1, 2]$ admits a face $[2, 0]$ that is not included in V.

If σ is a simplex, its dimension is defined as $|\sigma| - 1$ (cardinality of σ minus 1). If K is a simplicial complex, its dimension is defined as the maximal dimension of its simplices.

Example 4.5. Let $V = \{0, 1, 2, 3\}$ and

$$K = \{[0], [1], [2], [3], [0, 1], [1, 2], [2, 3], [3, 0], [0, 2], [1, 3], [0, 1, 2], [0, 1, 3], [0, 2, 3], [1, 2, 3]\}.$$

It a simplicial complex of dimension 2.

Example 4.6. Let $V = \{0, 1, 2, 3\}$ and

$$K = \{[0], [1], [2], [3], [0,1], [1,2], [2,3], [3,0],$$
$$[0,1,2], [0,1,3], [0,2,3], [1,2,3], [0,1,2,3]\}.$$

It a simplicial complex of dimension 3.

At the moment, a simplicial complex has no topology. It is a purely combinatorial object. However, in order to represent it, we can draw it as follows : put the points V in the plane or the space, and for each simplex of K, fill the convex hull of its vertices. For instance, the simplicial complexes of Examples 4.3 and 4.5 looks:

Remark 4.7. For reasons that will be clearer later, when drawing a simplicial complex, the simplices must not cross each other. However, it is not always possible to draw a simplicial complex in the plane (or space) this way.

As an example, the bipartite graph $K_{3,3}$ is a simplicial complex of dimension 1 (a *graph*) that cannot be drawn in the plane without crossing itself.

4.2 Topology

In this section, we will give simplicial complexes a topology. There are two ways of doing that: by embedding the simplicial complex in a Euclidean space $\mathbb{R}^n$ for n large enough, or via the gluing construction. We shall consider the first one.

Definition 4.8. Let K be a simplicial complex, with vertex $V = [\![1, ..., n]\!]$. In $\mathbb{R}^{n+1}$, consider, for every $i \in [\![0, n]\!]$, the vector $e_i = (0, ..., 1, 0, ..., 0)$ (i^{th} coordinate 1, the other ones 0). Let $|K|$ be the subset of $\mathbb{R}^{n+1}$ defined as:

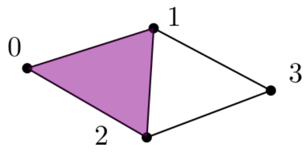
$$|K| = \bigcup_{\sigma \in K} \operatorname{conv}\left(\{e_j, j \in \sigma\}\right)$$

where conv represent the convex hull (see Remark 4.1).

Endowed with the subspace topology, $(|K|, \mathcal{T}_{||K|})$ is a topological space, that we call the *topological realization of* K.

Remark 4.9. There exists another definition of topological realization, via *quotient topology.* Basically, it consists in giving each simplices a topology (namely, the subspace topology of the standard simplex), and in gluing all these simplices together.

Remark 4.10. If a simplicial complex can be drawn in the plane (or space) without crossing itself, then its topological realization simply is the subspace topology. This is the case for $K = \{[0], [1], [2], [3], [0,1], [1,2], [2,0], [1,3], [2,3], [0,1,2]\}$.



Definition 4.11. Let $(X, \mathcal{T})$ be a topological space. A *triangulation* of X is a simplicial complex K such that its topological realization $(|K|, \mathcal{T}_{||K|})$ is homeomorphic to $(X, \mathcal{T})$.

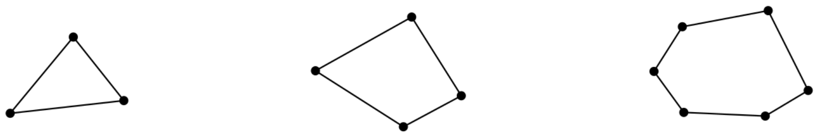
Example 4.12. The following simplicial complex, as in Example 4.3, is a triangulation of the circle:

$$K = \{[0], [1], [2], [0, 1], [1, 2], [2, 0]\}.$$

Example 4.13. The following simplicial complex, as in Example 4.5, is a triangulation of the sphere:

$$K = \{[0], [1], [2], [3], [0, 1], [1, 2], [2, 3], [3, 0], [0, 2], [1, 3], [0, 1, 2], [0, 1, 3], [0, 2, 3], [1, 2, 3]\}.$$

Given a topological space, it is not always possible to triangulate it. However, when it is, there exists many different triangulations. For instance, all the following simplicial complexes are triangulations of the circle.



Exercise 17. Give a triangulation of the cylinder.

4.3 Euler characteristic

Until here, we defined two invariants of topological space: number of connected components (homotopy type invariant), and dimension (homeomorphic invariant). We will now define one suited for simplicial complexes.

Definition 4.14. Let K be a simplicial complex of dimension n. Its *Euler characteristic* is the integer

$$\chi(K) = \sum_{0 \leq i \leq n} (-1)^i \cdot (\text{number of simplices of dimension } \ i).$$

Example 4.15. The simplicial complex of Example 4.3 has Euler characteristic

$$\chi(K) = 3 - 3 = 0.$$

Exercise 18. What are the Euler characteristics of Examples 4.5 and 4.5? What is the Euler characteristic of the icosahedron?

Exercise 19. Let K be a simplicial complex (with vertex set V). A *sub-complex* of K is a set $M \subset K$ that is a simplicial complex. Suppose that there exists two sub-complexes M and N of K such that $K = M \cup N$. Show the *inclusion-exclusion principle:*

$$\chi(K) = \chi(M) + \chi(N) - \chi(M \cap N).$$

Now, let $(X, \mathcal{T})$ be a topological space, and K a triangulation of it. We would like to define the Euler characteristic of X to be equal to the Euler characteristic of K:

$$\chi(X) = \chi(K).$$

Is it well-defined? In other words, if K' is another triangulation of X, is it true that

$$\chi(K) = \chi(K')?$$

It turns out that **this is true**, but we won't be able to prove it in this summer course.

> **Definition 4.16.** The Euler characteristic of a topological space is the Euler characteristic of any triangulation of it.

Here is a key fact: the Euler characteristic is a topological invariant.

Proposition 4.17. *If X and Y are two homotopy equivalent topological spaces, then $\chi(X) = \chi(Y)$.*

> **Exercise 20.** What is the Euler characteristic of a sphere of dimension 1? 2? 3?
> *Hint:* First, find a triangulation of the sphere $\mathbb{S}_n \subset \mathbb{R}^{n+1}$. It can be triangulated with $n+2$ simplices of dimension n.

Exercise 21. Using the previous exercise, show that $\mathbb{R}^3$ and $\mathbb{R}^4$ are not homeomorphic.
Hint: By contradiction, suppose that they are. Using Example 3.15, deduce that the unit sphere $\mathbb{S}_2 \subset \mathbb{R}^3$ and $\mathbb{S}_3 \subset \mathbb{R}^4$ are homotopy equivalent. Conclude with Proposition 4.17 and Exercise 20.

4.4 Python tutorial

Notebook available at
https://github.com/raphaeltinarrage/EMAp/blob/main/Tutorial1.ipynb.

In order to deal with simplicial complexes, we use the Gudhi library. We shall also use the libraries Matplotlib and Networkx (for plotting). Make sure to download the latest version!

Our code starts with

```
import gudhi
import numpy as np
import networkx as nx
```

We define a simplicial complex in GUDHI via

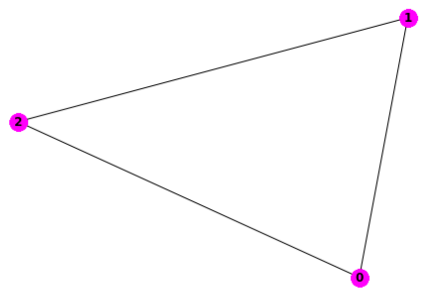
```
simpcomplex = gudhi.SimplexTree()

# We add the vertices
simpcomplex.insert([0])
simpcomplex.insert([1])
simpcomplex.insert([2])

# We add the edges
simpcomplex.insert([0,1])
simpcomplex.insert([1,2])
simpcomplex.insert([2,0])
```

The simplicial complex `simpcomplex` being created, we can use the functions

- `PrintSimplices(simpcomplex)` to print a list of its simplices:

```
The simplicial complex contains the following simplices:
Dimension 0: [0], [1], [2]
Dimension 1: [0, 1], [0, 2], [1, 2]
```

- `DrawSimplicialComplex(simpcomplex)` to output a visual representation of the simplicial complex (only its vertices and edges):



- `NumberOfConnectedComponents(simpcomplex)` to give its connected components:

```
The simplicial complex admits 1 connected component(s).
```

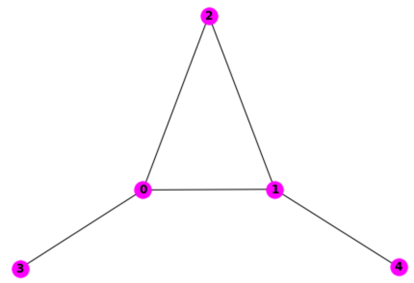
- `EulerCharacteristic(simpcomplex)` to give its Euler characteristic:

```
The simplicial complex has Euler characteristic equal to 0.
```

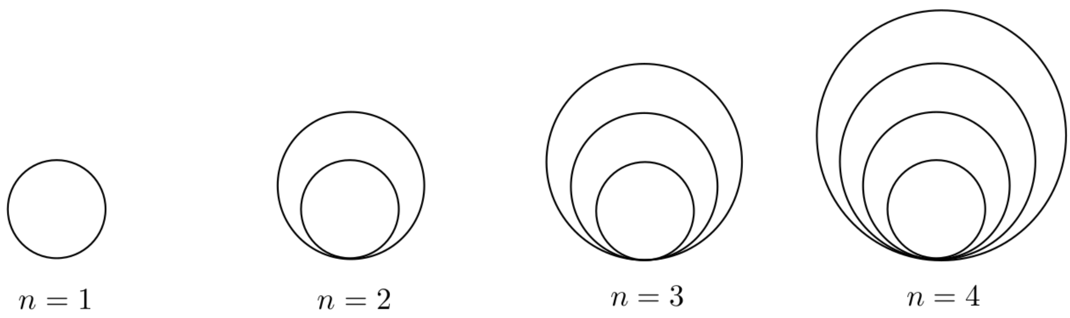
Exercise 22. Build triangulations of the letters of the alphabet, and compute their Euler characteristic.

Given two letters that are homotopy equivalent, is it true thar their Euler characteristic are equal? Given two letters that are not homotopy equivalent, is it true that their Euler characteristic are different? (see Exercise 16)

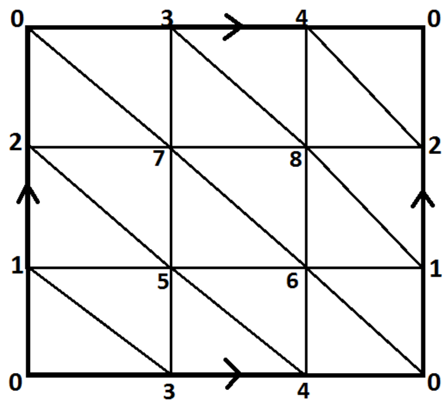
Hint: For instance, the following is a triangulation of A:



Exercise 23. For every n, triangulate the bouquet of n circles (see below). Compute their Euler characteristic.



Exercise 24. Implement the following triangulation of the torus:



Compute its Euler characteristic.

Exercise 25. Consider the following dataset of 30 points $x_0, ..., x_{29}$ in $\mathbb{R}^2$:

```
{0: [0.29409772548731694, 0.6646626625013836],
 1: [0.01625840776679577, 0.1676405753593595],
 2: [0.15988905150272759, 0.6411323760808338],
 3: [0.9073191075894482, -0.16417982219713312],
 4: [-0.18661467838673884, 0.31618948583046413],
 5: [-0.3664040542098381, 0.9098590694955988],
 6: [-0.43753448716144905, -0.8820102274699417],
```

```
7: [0.4096730199915961, -0.23801426675264126],
8: [0.5903822103474676, -0.7285102954232894],
9: [0.9133851839417766, -0.6606557328320093],
10: [-0.15516122940597588, 0.7565411235103017],
11: [-0.38626186295039866, -0.3662321656058476],
12: [0.005209710070218199, 0.27655964872153116],
13: [0.670078068894711, -0.00932202688834849],
14: [-0.011268465716772091, 0.24340880308017376],
15: [-0.6441978411451603, -0.9672635759413206],
16: [-0.2841794022401025, -0.6734801188906114],
17: [-0.15473260248990717, -0.1365357396855129],
18: [0.7177096105982121, 0.9378197891592468],
19: [-0.4677068504994166, 0.1533930130294956],
20: [-0.32379909116817096, 0.9694800649768063],
21: [-0.2886940472879451, -0.039544695812395725],
22: [-0.5900701743351606, 0.8350804500575086],
23: [0.14931959728335853, 0.869106793774487],
24: [-0.14500672678238824, -0.3170082291070364],
25: [0.07324547392476122, 0.6653572287065117],
26: [-0.662990048258566, 0.1908198608241125],
27: [-0.25641262456436276, -0.9844196180941553],
28: [-0.5105685407819842, -0.4236604017060557],
29: [0.6792549581008038, -0.026215820387260003]}
```

Write a function that takes as an input a parameter $r \geq 0$, and returns the simplicial complex $\mathcal{G}(r)$ defined as follows:

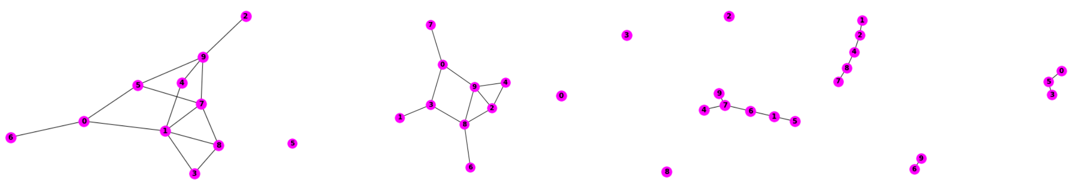
- the vertices of $\mathcal{G}(r)$ are the points $x_0, ..., x_{29}$,
- for all $i, j \in \llbracket 0, 29 \rrbracket$ with $i \neq j$, the edge $[i, j]$ belongs to $\mathcal{G}(r)$ if and only if $\|x_i - x_j\| \leq r$.

Compute the number of connected components of $\mathcal{G}(r)$ for several values of r. What do you observe?

Exercise 26. A *Erdős–Rényi random graph* $\mathcal{G}(n,p)$ is a simplicial complex obtained as follows:

- add n vertices $1, ..., n$,
- for every $a, b \in \llbracket 1, n \rrbracket$, add the edge $[a, b]$ to the complex with probability p.

Builds a function that, given n and p, outputs a simplicial complex $\mathcal{G}(n,p)$. Observe the influence of p on the number of connected components of $\mathcal{G}(10,p)$ and $\mathcal{G}(100,p)$.



Hint: If V is a list, `itertools.combinations(V,2)` can be used to generate all the non-ordered pairs `[a,b]` in `V` (from package `itertools`).
The command `random.random()` can be used to generate a random number between 0 and 1, and `random.random()<p` is `True` with probability `p` (from package `random`).

5 Homological algebra

This subsection is devoted to defining a powerful invariant in algebraic topology, called *homology*. We will restrict to the case of *simplicial homology* over the *finite fied* $\mathbb{Z}/2\mathbb{Z}$.

5.1 Reminder on $\mathbb{Z}/2\mathbb{Z}$-vector spaces

We review some basic notions of algebra: groups and vector spaces.

Groups. We recall that a *group* $(G, +)$ is a set G endowed with an operation

$$\begin{aligned} G \times G &\longrightarrow G \\ (g, h) &\longmapsto g + h \end{aligned}$$

such that:

- (associativity) $\forall a, b, c \in G$, $(a + b) + c = a + (b + c)$,
- (identity) $\exists\ 0 \in G, \forall a \in G, a + 0 = 0 + a = a$,
- (inverse) $\forall a \in G, \exists b \in G, a + b = b + a = 0$.

Moreover, we say that G is *commutative* if $\forall a, b \in G, a + b = b + a$. In this course, the only groups we consider will be commutative and finite.

A *subgroup* of $(G, +)$ is a subset $H \subset G$ such that

$$\forall a, b \in H, a + b \in H.$$

If H is a subgroup of G, the operation $+ : G \times G \to G$ restricts to an operation $+ : H \times H \to H$, making H a group on its own.

Suppose that G is commutative, and that H is a subgroup of H. We define the following equivalence relation on G: for all $a, b \in G$,

$$a \sim b \iff a - b \in H.$$

Denote by G/H the quotient set of G under this relation. For any $a \in G$, one shows that the equivalence class of a is equal to

$$a + H = \{a + h, h \in H\}.$$

Let $a_0 = 0, a_1, ..., a_n$ be a choice of representants of equivalence classes of the relation $\sim$. The quotient set can be written as

$$G/H = \big\{0 + H, a_1 + H, ..., a_n + H\big\}.$$

One defines a group structure $\oplus$ on G/H as follows: for any $i, j \in [\![0, n]\!]$,

$$(a_i + H) \oplus (a_j + H) = (a_i + a_j) + H.$$

The group $(G/H, \oplus)$ is called the *quotient group*.

Consider two groups $(G, +)$ and $(H, +)$ (for simplicity, we denote the operations with the same symbol $+$). An *morphism* between them is an application $f : G \to H$ such that

$$\forall a, b \in G, f(a + b) = f(a) + f(b).$$

If f is a bijection, it is called an *isomorphism*.

If $f : G \to H$ is a morphism, the *image* of f is defined as

$$\mathrm{Im}(f) = \{f(a), a \in G\}.$$

One shows that it is a subgroup of H. The *kernel* of f is defined as

$$\mathrm{Ker}(f) = \{a \in G, f(a) = 0\}.$$

One shows that it is a subgroup of G. The first isomorphism theorem states that the quotient group $G/\mathrm{Ker}(f)$ is isomorphic to the subgroup $\mathrm{Im}(f)$. More explicitely, an isomorphism $G/\mathrm{Ker}(f) \to \mathrm{Im}(f)$ is given by

$$a + \mathrm{Ker}(f) \longmapsto f(a).$$

The group $\mathbb{Z}/2\mathbb{Z}$. Consider the group $(\mathbb{Z}, +)$. It admits a subgroup $2\mathbb{Z} = \{2n, n \in \mathbb{Z}\}$. The equivalence relation $\sim$ admits two equivalence classes:

$$2\mathbb{Z} = \{2n, n \in \mathbb{Z}\} \qquad \text{and} \qquad 1 + 2\mathbb{Z} = \{1 + 2n, n \in \mathbb{Z}\}.$$

The quotient group can be seen as the group $\mathbb{Z}/2\mathbb{Z} = \{0, 1\}$ with the operations

$$\begin{gathered} 0 + 0 = 0 \\ 0 + 1 = 1 \\ 1 + 0 = 1 \\ 1 + 1 = 0 \end{gathered}$$

The group $(\mathbb{Z}/2\mathbb{Z}, +)$ is the only group with two elements. Note that it can also be given a *field* structure, via the operation

$$\begin{gathered} 0 \times 0 = 0 \\ 0 \times 1 = 0 \\ 1 \times 0 = 0 \\ 1 \times 1 = 1 \end{gathered}$$

For any $n \geq 1$, the *product group* $((\mathbb{Z}/2\mathbb{Z})^n, +)$ is the group whose underlying set is

$$(\mathbb{Z}/2\mathbb{Z})^n = \{(\epsilon_1, ..., \epsilon_n), \epsilon_1, ..., \epsilon_n \in \mathbb{Z}/2\mathbb{Z}\}$$

and whose operation is defined as

$$(\epsilon_1, ..., \epsilon_n) + (\epsilon'_1, ..., \epsilon'_n) = (\epsilon_1 + \epsilon'_1, ..., \epsilon_n + \epsilon'_n).$$

Note that the set $(\mathbb{Z}/2\mathbb{Z})^n$ has 2^n elements.

Vector spaces. Let $(\mathbb{F}, +, \times)$ be a field. We recall that a vector space over $\mathbb{F}$ is a group $(V, +)$ endowed with an operation

$$\begin{aligned} \mathbb{F} \times V &\longrightarrow V \\ (\lambda, v) &\longmapsto \lambda \cdot v \end{aligned}$$

such that

- (compatibility of multiplication) $\forall \lambda, \mu \in \mathbb{F}, \forall v \in V, \lambda \cdot (\mu \cdot v) = (\lambda \times \mu) \cdot v,$
- (identity) $\forall v \in V, 1 \cdot v = v$ where 1 denotes the unit of $\mathbb{F}$,
- (scalar distributivity) $\forall \mu, \nu \in \mathbb{F}, \forall v \in V,\ (\lambda + \nu) \cdot v = \lambda \cdot v + \nu \cdot v,$
- (vector distributivity) $\forall \mu \in \mathbb{F}, \forall v, w \in V,\ \lambda \cdot (u + v) = \lambda \cdot v + \nu \cdot v.$

When there is no risk of confusion, we will write λv instead of $\lambda \cdot v$.

Let $\{v_1, ..., v_n\}$ be a collection of elements of V. We say that it is *free* if

$$\forall \lambda_1, ..., \lambda_n \in \mathbb{F}, \sum_{1 \leq i \leq n} \lambda_i v_i = 0 \implies \lambda_1 = ... = \lambda_n = 0.$$

We say that it is *spans* V if

$$\forall v \in V, \exists \lambda_1, ..., \lambda_n \in \mathbb{F}, \sum_{1 \leq i \leq n} \lambda_i v_i = v.$$

If the collection $\{v_1, ..., v_n\}$ is free and spans V, we say that it is a *basis*. One shows that $\{v_1, ..., v_n\}$ is a basis if and only if

$$\forall v \in V, \exists! \lambda_1, ..., \lambda_n \in \mathbb{F}, \sum_{1 \leq i \leq n} \lambda_i v_i = v.$$

A *linear subspace* of $(V, +, \cdot)$ is a subset $W \subset V$ such that

$$\forall u, v \in W, u + v \in W \qquad \text{and} \qquad \forall v \in W, \forall \lambda \in \mathbb{F}, \lambda v \in W.$$

Just as for groups, we can define an equivalence relation $\sim$ on V, and a *quotient vector space* V/W. The quotient has dimension $\dim V/W = \dim V - \dim W$.

Let $(V, +, \cdot)$ and $(W, +, \cdot)$ be two vector spaces. A linear map is a map $f \colon V \to W$ such that

$$\forall u, v \in V, f(u + v) = f(u) + f(v) \qquad \text{and} \qquad \forall v \in V, \forall \lambda \in \mathbb{F}, f(\lambda v) = \lambda \cdot f(v).$$

If f is a bijection, it is called an *isomorphism*, and we say that V and W are *isomorphic*. If $(V, +, \cdot)$ is a vector space of dimension n, one shows that it is isomorphic to the product vector space $\mathbb{F}^n$.

Structure of $\mathbb{Z}/2\mathbb{Z}$-vector spaces. Not all groups $(V, +)$ can be given a $\mathbb{Z}/2\mathbb{Z}$-vector space structure. The following statement gives precisely when they can:

Proposition 5.1. *Le $(V, +)$ be a commutative group. It can be given a $\mathbb{Z}/2\mathbb{Z}$-vector space structure if and only if $\forall v \in V, v + v = 0$.*

Proof. Suppose that $(V, +, \cdot)$ is a $\mathbb{Z}/2\mathbb{Z}$-vector space. For all $v \in V$, we have

$$0 = 0 \cdot v = (1 + 1) \cdot v = v + v,$$

which shows an application. In the other direction, if $\forall v \in V, v + v = 0$, then we can define a vector space structure on $(V, +)$ as follows: for all $v \in V$,

$$\begin{aligned} 0 \cdot v &= 0 \\ 1 \cdot v &= v \end{aligned}$$

One verifies the axioms of a vector space. $\square$

Applying the usual theory of vector spaces, we obtain the following proposition:

Proposition 5.2. *Let $(V, +, \cdot)$ be a finite $\mathbb{Z}/2\mathbb{Z}$-vector space. Then there exists $n \geq 0$ such that V has cardinal 2^n, and $(V, +, \cdot)$ is isomorphic to the vector space $(\mathbb{Z}/2\mathbb{Z})^n$.*

Exercise 27. Let V be a $\mathbb{Z}/2\mathbb{Z}$-vector space, and W a linear subspace. Using Proposition 5.2, prove that

$$\dim V/W = \dim V - \dim W.$$

Exercise 28. Let $(G, +)$ be a group, potentially non-commutative. Prove that

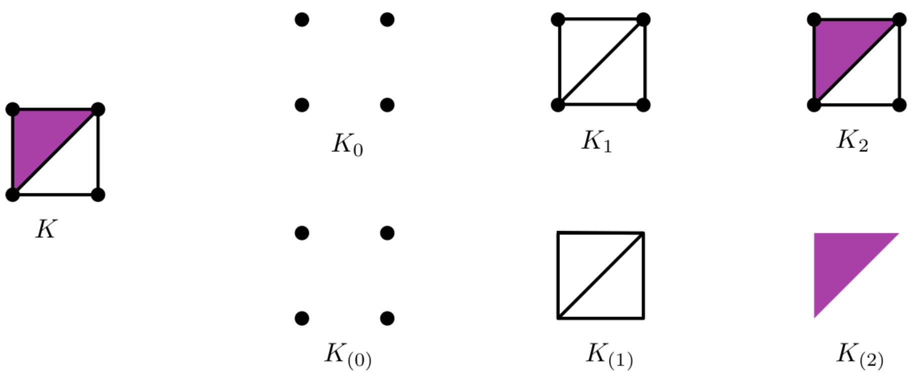
$$\forall g \in G, g + g = 0 \implies G \text{ is commutative.}$$

5.2 Chains, cycles and boundaries

Let K be a simplicial complex. For any $n \geq 0$, define the sets

$$\begin{aligned} K_n &= \{\sigma \in K, \dim(\sigma) \leq n\} \\ K_{(n)} &= \{\sigma \in K, \dim(\sigma) = n\}. \end{aligned}$$

The first set is a simplicial complex, called the *n-skeleton* of K. The second one is not a simplicial complex in general, and has no name.



Chains. Let $n \geq 0$. The *n-chains* of K is the set $C_n(K)$ whose elements are the formal sums

$$\sum_{\sigma \in K_{(n)}} \epsilon_\sigma \cdot \sigma \quad \text{where} \quad \forall \sigma \in K_{(n)},\ \epsilon_\sigma \in \mathbb{Z}/2\mathbb{Z}.$$

We can give $C_n(K)$ a group structure via

$$\sum_{\sigma \in K_{(n)}} \epsilon_\sigma \cdot \sigma + \sum_{\sigma \in K_{(n)}} \eta_\sigma \cdot \sigma = \sum_{\sigma \in K_{(n)}} (\epsilon_\sigma + \eta_\sigma) \cdot \sigma.$$

Moreover, $C_n(K)$ can be given a $\mathbb{Z}/2\mathbb{Z}$-vector space structure. To see this, observe that for any element of $C_n(K)$,

$$\sum_{\sigma \in K_{(n)}} \epsilon_\sigma \cdot \sigma + \sum_{\sigma \in K_{(n)}} \epsilon_\sigma \cdot \sigma = \sum_{\sigma \in K_{(n)}} (\epsilon_\sigma + \epsilon_\sigma) \cdot \sigma = \sum_{\sigma \in K_{(n)}} 0 \cdot \sigma = 0$$

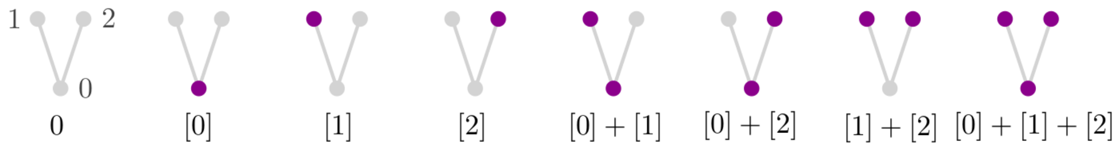
the second equality follows from $0+0 = 1+1 = 0$ in $\mathbb{Z}/2\mathbb{Z}$. We conclude with Proposition 5.1.

Example 5.3. Consider the simplicial complex

$$K = \{[0], [1], [2], [0, 1], [0, 2]\}.$$

The 0-chains $C_0(K)$ consists in 8 elements:

$$C_0(K) = \{0,\ [0],\ [1],\ [2],\ [0] + [1],\ [0] + [2],\ [1] + [2],\ [0] + [1] + [2]\}$$

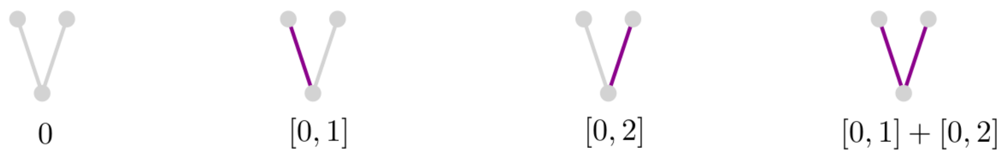


As an example, in $C_0(K)$, we have

$$([0] + [1]) + ([0] + [2]) = [0] + [0] + [1] + [2] = [1] + [2].$$

Besides, the 1-chains $C_1(K)$ consists in 4 elements:

$$C_1(K) = \{0,\ [0, 1],\ [0, 2],\ [0, 1] + [0, 2]\}.$$



Remark 5.4. The group $C_n(K)$ can be seen a the group of maps $K_{(n)} \to \mathbb{Z}/2\mathbb{Z}$, endowed with the addition operation. For instance, the chain $[0] + [1]$ would correspond to the map $f\colon K_{(0)} \to \mathbb{Z}/2\mathbb{Z}$ defined as

$$f([0]) = 1, \qquad f([1]) = 1, \qquad f([2]) = 0,$$

and the chain $[0]+[2]$ the map $g\colon K_{(0)} \to \mathbb{Z}/2\mathbb{Z}$ defined as

$$f([0]) = 1, \qquad f([1]) = 0, \qquad f([2]) = 1.$$

Their sum is the map $f+g$ defined as

$$(f+g)([0]) = 1+1 = 0, \qquad (f+g)([1]) = 1+0 = 1, \qquad (f+g)([2]) = 0+1 = 1.$$

Remark 5.5. The group $C_n(K)$ can also be seen as the set $\mathcal{P}(K_{(n)})$ of subsets of K_n, endowed with the symmetric difference operation, defined as $A\Delta B = (A\cup B)\setminus(A\cap B)$. For instance, the chain $[0]+[1]$ would correspond to the subset $\{[0],[1]\}$, and the chain $[0]+[2]$ to $\{[0],[2]\}$. Their sum is the subset

$$\{[0],[1]\}\ \Delta\ \{[0],[2]\} = \{[1],[2]\}.$$

Boundary operator. Let $n \geq 1$, and $\sigma = [x_0, ..., x_n] \in K_{(n)}$ a simplex of dimension n. We define its *boundary* as the following element of $C_{n-1}(K)$:

$$\partial_n \sigma = \sum_{\substack{\tau \subset \sigma \\ |\tau| = |\sigma|-1}} \tau$$

where $|\tau|$ denotes the cardinal of τ. We can extend the operator ∂_n as a linear map $\partial_n \colon C_n(K) \to C_{n-1}(K)$ as follows: for any element of $C_n(K)$,

$$\partial_n \sum_{\sigma \in K_{(n)}} \epsilon_\sigma \cdot \sigma = \sum_{\sigma \in K_{(n)}} \epsilon_\sigma \cdot \partial_n \sigma.$$

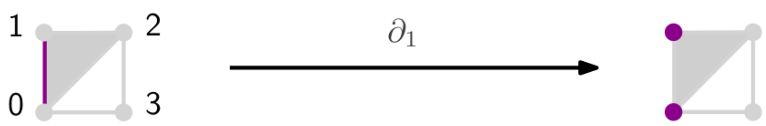
Besides, for $n=0$, we define the boundary operator ∂_0 as the zero map $C_0(K) \to \{0\}$, i.e., for all $c \in C_0(K), \partial_0(c) = 0$. In what follows, we denote $C_{-1}(K) = \{0\}$.

Example 5.6. Consider the simplicial complex

$$K = \{[0],[1],[2],[3],[0,1],[0,2],[1,2],[1,3],[2,3],[0,1,2]\}.$$

The simplex $[0,1]$ has the faces $[0]$ and $[1]$. Hence

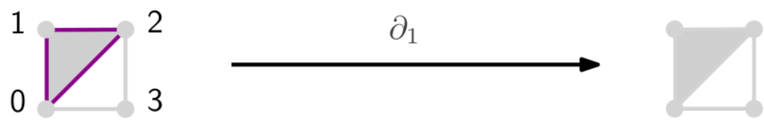
$$\partial_1[0,1] = [0]+[1].$$



Similarly, the boundary of the 1-chain $[0,1]+[1,2]+[2,0]$ is

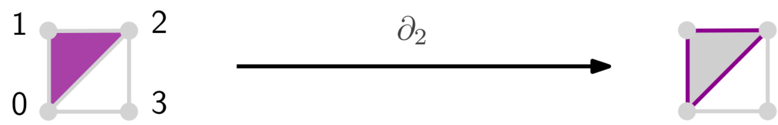
$$\begin{aligned} \partial_1\big([0,1]+[1,2]+[2,0]\big) &= \partial_1[0,1] + \partial_1[0,2] + \partial_1[2,0] \\ &= [0]+[1]+[0]+[2]+[2]+[0] \\ &= 0 \end{aligned}$$

since $[0]+[0] = [1]+[1] = [2]+[2] = 0$ in $C_0(K)$.



The simplex $[0,1,2]$ has the faces $[0,1]$ and $[1,2]$ and $[2,0]$. Hence

$$\partial_2[0,1,2] = [0,1] + [1,2] + [2,0].$$



Boundary and cycles. Let $n \geq 0$. We have a triplet of vector spaces

$$C_{n+1}(K) \xrightarrow{\partial n+1} C_n(K) \xrightarrow{\partial n} C_{n-1}(K).$$

The maps ∂_{n+1} and ∂_n are linear maps, and we can consider their kernel and image (see reminder in Subsection 5.1). We define:

- The n-*cycles*: $Z_n(K) = \mathrm{Ker}(\partial_n)$,
- The n-*boundaries*: $B_n(K) = \mathrm{Im}(\partial_{n+1})$.

We say that two chains $c, c' \in C_n(K)$ are *homologous* if there exists $b \in B_n(K)$ such that $c = c' + b$. In other words, two chains are homologous if they are equal up to a boundary.

Example 5.7. Consider the simplicial complex of Example 5.6. The set of cycles $Z_1(K)$ consists in the chains

$$0, \quad [0,1]+[1,2]+[0,2], \quad [0,2]+[2,3]+[0,3] \quad \text{and} \quad [0,1]+[1,2]+[2,3]+[0,3].$$

The only boundaries $B_1(K)$ is given by

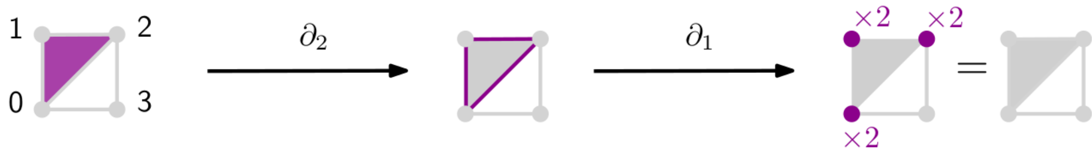
$$\partial_2(0) = 0 \quad \text{and} \quad \partial_2([0,1,2]) = [0,1]+[0,2]+[1,2].$$

We see that the chains $[0,2]+[2,3]+[0,3]$ and $[0,1]+[1,2]+[2,3]+[0,3]$ are homologous. Indeed,

$$[0,2]+[2,3]+[0,3] = [0,1]+[1,2]+[2,3]+[0,3]+[0,1]+[0,2]+[1,2].$$

Here is a key property of the boundary operator:

Lemma 5.8. *For any $n \geq 0$, for any $c \in C_n(K)$, we have $\partial_{n-1} \circ \partial_n(c) = 0$.*
In other words, the map $\partial_{n-1} \circ \partial_n \colon C_n(K) \to C_{n-2}(K)$ is zero.



Proof. Suppose that $n \geq 2$, the result being trivial otherwise. Since the boundary operators are linear, it is enough to prove that $\partial_{n-1} \circ \partial_n(\sigma) = 0$ for all simplex $\sigma \in K_{(n)}$. By definition,

$$\partial_n(\sigma) = \sum_{\substack{\tau \subset \sigma \\ |\tau|=|\sigma|-1}} \tau,$$

and

$$\partial_{n-1} \circ \partial_n(\sigma) = \sum_{\substack{\tau \subset \sigma \\ |\tau|=|\sigma|-1}} \partial_{n-1}(\tau) = \sum_{\substack{\tau \subset \sigma \\ |\tau|=|\sigma|-1}} \sum_{\substack{\nu \subset \tau \\ |\nu|=|\tau|-1}} \nu$$

We can write this last sum as

$$\sum_{\substack{\tau \subset \sigma \\ |\tau|=|\sigma|-1}} \sum_{\substack{\nu \subset \tau \\ |\nu|=|\tau|-1}} \nu = \sum_{\substack{\nu \subset \sigma \\ |\nu|=|\sigma|-2}} \alpha_\nu \nu$$

where $\alpha_\nu = \{\tau \subset \sigma, |\tau| = |\sigma| - 1, \nu \subset \tau\}$. It is easy to see that for every ν such that $|\nu| = |\tau| - 2$, we have $\alpha_\nu = 2 = 0$. $\square$

Corollary 5.9. *We have $B_n(K) \subset Z_n(K)$. In other words, any boundary is a cycle.*

Proof. Let $b \in B_n(K)$ be a boundary. By definition, there exists $c \in C_{n+1}(K)$ such that $b = \partial_{n+1}(c)$. Using Lemma 5.8, we obtain

$$\partial_n(b) = \partial_n \partial_{n+1}(c) = 0,$$

hence $b \in Z_n(K)$. $\square$

5.3 Homology groups

In the previous subsection, we have defined a sequence of vector spaces, connected by linear maps

$$\dots \xrightarrow{\partial_{n+2}} C_{n+1}(K) \xrightarrow{\partial_{n+1}} C_n(K) \xrightarrow{\partial_n} C_{n-1}(K) \xrightarrow{\partial_{n-1}} \dots$$

and for every $n \geq 0$, we have defined the cycles and the boundaries $Z_n(K)$ and $B_n(K)$. According to Corollary 5.9, $B_n(K)$ is a linear subspace of $Z_n(K)$. We can consider the corresponding quotient vector space:

Definition 5.10. The n^{th} *homology group* of K is $H_n(K) = Z_n(K)/B_n(K)$.

Since $H_n(K)$ is a quotient of $\mathbb{Z}/2\mathbb{Z}$-vector spaces, it is a $\mathbb{Z}/2\mathbb{Z}$-vector space. According to Proposition 5.2, it is isomorphic to $(\mathbb{Z}/2\mathbb{Z})^d$, where $d = \dim H_n(K)$. We also have $|H_n(K)| = 2^d$. By applying Exercise 27, we obtain the relation

$$\dim H_n(K) = \dim B_n(K) - \dim Z_n(K).$$

Example 5.11. We consider the simplicial complex of Example 5.7. As we have seen, $Z_1(K)$ has cardinal 4, and $B_1(K)$ cardinal 2. We deduce that $\dim Z_1(K) = 2$, $\dim B_1(K) = 1$, and

$$\dim H_1(K) = 2 - 1 = 1.$$

In other words, we have an isomorphism $H_1(K) \simeq \mathbb{Z}/2\mathbb{Z}$.

Definition 5.12. Let K be a simplicial complex and $n \geq 0$. Its n^{th} *Betti number* is the integer $\beta_n(K) = \dim H_n(K)$.

Exercise 29. Compute the Betti numbers $\beta_0(K)$, $\beta_1(K)$ and $\beta_2(K)$ of the following simplicial complex:

$$K = \{[0], [1], [2], [3], [0, 1], [1, 2], [2, 3], [3, 0]\}.$$

Exercise 30. Compute the Betti numbers $\beta_0(K)$, $\beta_1(K)$ and $\beta_2(K)$ of the following simplicial complex:

$$K = \{[0], [1], [2], [3], [0, 1], [1, 2], [2, 3], [3, 0], [0, 2], [1, 3], [0, 1, 2], [0, 1, 3], [0, 2, 3], [1, 2, 3]\}.$$

Moralidade
Os grupos de homologia são os espaços vetoriais
dos circulos módulo os limites.

5.4 Homology groups of topological spaces

Just as we did for the Euler characteristic, we will define the homology groups of topological spaces via triangulations of it.

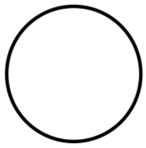
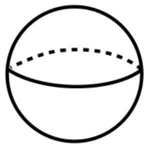
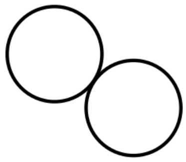
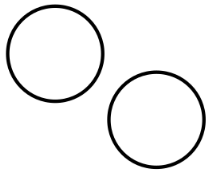
Definition 5.13. The homology groups of a topological space are the homology groups of any triangulation of it. We define their Betti numbers similarly.

For this definition to make sense, we have to make sure that the homology groups are an invariant of *homeomorphism equivalence*. We can prove an even stronger result: homology groups are an invariant of *homotopy equivalence*. We will admit this statement.

Proposition 5.14. *If X and Y are two homotopy equivalent topological spaces, then for any $n \geq 0$ we have isomorphic homology groups $H_n(X) \simeq H_n(Y)$. As a consequence, $\beta_n(X) = \beta_n(Y)$.*

Remark 5.15. Again, the previous definition suffers from the fact that all topological spaces are not triangulable. However, there exists a definition of homology that is better suited for topological spaces in many ways. It is called *singular homology*, but it is beyond the scope of this summer course.

To close this section, we give some examples of homology groups:

X				
$H_0(X)$	$\mathbb{Z}/2\mathbb{Z}$	$\mathbb{Z}/2\mathbb{Z}$	$\mathbb{Z}/2\mathbb{Z}$	$(\mathbb{Z}/2\mathbb{Z})^2$
$\beta_0(X)$	1	1	1	2
$H_1(X)$	$\mathbb{Z}/2\mathbb{Z}$	0	$(\mathbb{Z}/2\mathbb{Z})^2$	$(\mathbb{Z}/2\mathbb{Z})^2$
$\beta_1(X)$	1	0	2	2
$H_2(X)$	0	$\mathbb{Z}/2\mathbb{Z}$	0	0
$\beta_2(X)$	0	1	0	0

6 Incremental algorithm

In this section, we present the first algorithm of persistent homology, as in [ELZ00].

6.1 Incremental algorithm à la main

We start by presenting a version of the incremental algorithm that can be applied by hand. In Subsection 6.3 we will present a matrix version of the algorithm.

Let K be a simplicial complex with n simplices. Choose a total order of the simplices

$$\sigma^1 < \sigma^2 < ... < \sigma^n$$

such that

$$\forall \sigma, \tau \in K,\ \tau \subsetneq \sigma \implies \tau < \sigma.$$

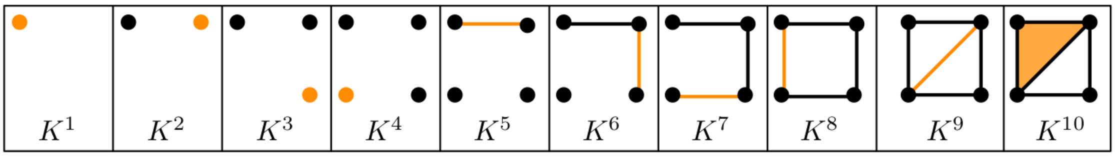
In other words, a face of a simplex is lower than the simplex itself. For every $i \leq n$, consider the simplicial complex

$$K^i = \{\sigma^1, ..., \sigma^i\}.$$

We have the relation $K^{i+1} = K^i \cup \left\{\sigma^{i+1}\right\}$. They form an inscreasing sequence of simplicial complexes

$$K^1 \subset K^2 \subset ... \subset K^n,$$

with $K^n = K$.



We will compute the homology groups of K^i incrementally. To do so, we need the following notion:

> **Definition 6.1.** Let $i \in [\![1, n]\!]$, and $d = \dim(\sigma_i)$. The simplex σ^i is *positive* if there exists a cycle $c \in Z_d(K^i)$ that contains σ_i. In other words, there exist $c = \sum_{\sigma \in K^i_{(n)}} \epsilon_\sigma \cdot \sigma \in C_n(K^i)$ such that $\epsilon_{\sigma^i} = 1$ and $\partial_n(c) = 0$. Otherwise, σ^i is *negative*.

For instance:

- $\sigma^1 \in K^1$ is **positive** because it is included in the cycle $c = \sigma^1$ (indeed, $\partial_0(\sigma^1) = 0$).
- $\sigma^2 \in K^2$ is **positive** because it is included in the cycle $c = \sigma^2$ (indeed, $\partial_0(\sigma^2) = 0$).
- $\sigma^5 \in K^5$ is **negative** because it is not included in a cycle $Z_1(K^5)$. Indeed, $C_1(K^5)$ only contains 0 and σ^5, and $\partial_1(\sigma^5) = \sigma^1 + \sigma^2 \neq 0$.
- $\sigma^8 \in K^8$ is **positive** because it is included in the cycle $c = \sigma^5 + \sigma^6 + \sigma^7 + \sigma^8$ (indeed, $\partial_1(c) = 2\sigma^1 + 2\sigma^2 + 2\sigma^3 + 2\sigma^4 = 0$).

Note that, by adding σ^i in the simplicial complex, the only groups that may change are $Z_d(K^i)$ and $B_{d-1}(K^i)$, with $d = \dim(\sigma^i)$. The following lemmas state precisely what happens.

Lemma 6.2. *Let $d = \dim(\sigma^i)$. If σ^i is positive, then $\beta_d(K^i) = \beta_d(K^{i-1}) + 1$, and for all $d' \neq d$, $\beta_{d'}(K^i) = \beta_{d'}(K^{i-1})$.*

Proof. Since σ^i is positive, there is a cycle $c \in Z_d(K^i)$ that contains σ^i. This cycle is not included in $Z_d(K^{i-1})$ (we just added σ^i). As a consequence,

$$\dim Z_d(K^i) = \dim Z_d(K^{i-1}) + 1.$$

Besides, $\partial_i(\sigma^i) = \partial_i(c) + \partial_i(\sigma^i) = \partial_i(c + \sigma^i)$, and $c + \sigma^i$ is a chain of K^{i-1}. Hence

$$\dim B_{d-1}(K^i) = \dim B_{d-1}(K^{i-1}).$$

We conclude by using the relation $\beta_d(K^i) = \dim Z_d(K^i) - \dim B_d(K^i)$. □

Lemma 6.3. *Let $d = \dim(\sigma^i)$. If σ^i is negative, then $\beta_{d-1}(K^i) = \beta_{d-1}(K^{i-1}) - 1$, and for all $d' \neq d - 1$, $\beta_{d'}(K^i) = \beta_{d'}(K^{i-1})$.*

Proof. We start by proving the following fact: $\partial_d(\sigma^i)$ is not a boundary of K^{i-1}. Otherwise, we would have $\partial_d(\sigma^i) = \partial_d(c)$ with $c \in C_d(K^{i-1})$, i.e. $\partial_d(\sigma^i + c) = 0$. Hence $\sigma^i + c$ would be a cycle of K^i that contains c, contradicting the negativity of σ^i. As a consequence,

$$\dim B_{d-1}(K^i) = \dim B_{d-1}(K^{i-1}) + 1.$$

Moreover, since σ^i is negative, we have

$$\dim Z_d(K^i) = \dim Z_d(K^{i-1}).$$

We conclude by using the relation $\beta_d(K^i) = \dim Z_d(K^i) - \dim B_d(K^i)$. □

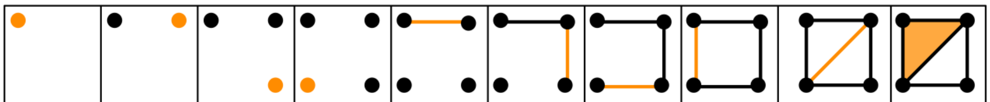
We derive the following algorithm:

Algorithm 1: Incremental algorithm for homology

Input: an increasing sequence of simplicial complexes $K^1 \subset \cdots \subset K^n = K$
Output: the Betti numbers $\beta_0(K), ...\beta_d(K)$
$\beta_0 \leftarrow 0, ..., \beta_d \leftarrow 0$;
for $i \leftarrow 1$ **to** n **do**
  $d = \dim(\sigma^i)$;
  if σ^i *is positive* **then**
    $\beta_k(K^i) \leftarrow \beta_k(K^i) + 1$;
  else if $d > 0$ **then**
    $\beta_{k-1}(K^i) \leftarrow \beta_{k-1}(K^{i-1}) - 1$;

Of course, there remains the problem of determining automatically whether the simplex is positive. We will propose a solution in Subsection 6.3.

We now apply the algorithm to our simplicial complex. The output is $\beta_0(K) = 1$ and $\beta_1(K) = 1$.



	K^1	K^2	K^3	K^4	K^5	K^6	K^7	K^8	K^9	K^{10}
Dimension	0	0	0	0	1	1	1	1	1	2
Positivity	$+$	$+$	$+$	$+$	$-$	$-$	$-$	$+$	$+$	$-$
$\beta_0(K^i)$	1	2	3	4	3	2	1	1	1	1
$\beta_1(K^i)$	0	0	0	0	0	0	0	1	2	1

Exercise 31. Compute again the Betti numbers of the simplicial complexes of Exercises 29 and 30, using the incremental algorithm.

6.2 Applications

Number of connected components. We link the notion of connectedness with the homology groups.

Proposition 6.4. *Let X be a (triangulable) topological space. Then its 0^{th} Betti number, $\beta_0(X)$, is equal to the number of connected components of X.*

Proof. First, a definition: say that a simplicial complex L is *combinatorially connected* of for every vertex v, w of L, there exists a sequence of edges that connects v and w:

$$[v, v_1], \quad [v_1, v_2], \quad [v_2, v_3], \quad ..., \quad [v_n, w].$$

Let m be the number of connected components X, and let K be triangulation of X. We accept the following equivalent statement: there exists m **disjoint**, **non-empty** and **combinatorially connected** simplicial sub-complex $L_1, ..., L_m$ of K such that

$$K = \bigcup_{1 \leq i \leq m} L_i.$$

Now, let T be a spanning forest of K, that is, a union of spanning trees. One shows that admits m combinatorially connected components.

Consider an ordering of the simplices of K that begins with an ordering of T. We apply the incremental algorithm. First, each vertex increases β_0 by 1. Next, since T is a tree, all its edges are negative simplices (T has no cycles), and hence decrease β_0. We know that each tree of the forest contains $k-1$ edges, where k is the number of vertices of the corresponding component. At that point of the algorithm, when all T is added, β_0 is equal to m.

Now, since T is a spanning tree, each other edges of K is positive, hence β_0 does not change. Similarly, the other simplices of K do not change β_0. We deduce the result. $\square$

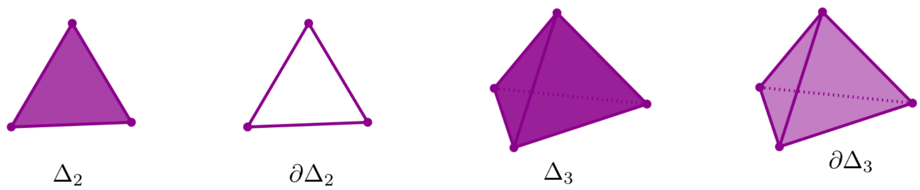
Homology of spheres. Let us compute the homology of spheres. For any $n \geq 1$, consider the vertex set $V = \{0, \ldots, n\}$, and the simplicial complex

$$\Delta_n = \{S \subset V, S \neq \emptyset\}.$$

We call it the *simplicial standard n-simplex*. Define its boundary as

$$\partial\Delta_n = \Delta_n \setminus V.$$

One shows that $\partial\Delta_n$ is a triangulation of the $(n-1)$-sphere $\mathbb{S}_{n-1} \subset \mathbb{R}^n$.



Exercise 32. Prove that $\partial\Delta_n$ is a triangulation of the $(n-1)$-sphere.

As a consequence, for all $i \geq 0$, we have $H_i(\mathbb{S}_n) = H_i(\partial\Delta_{n+1})$. We will use this simplicial complex to compute these homology groups.

Proposition 6.5. *The Betti numbers of $\mathbb{S}_n$ are:*

- $\beta_i(\mathbb{S}_n) = 1$ *for* $i = 0, n$,
- $\beta_i(\mathbb{S}_n) = 0$ *else.*

Proof. Consider the *simplicial standard n-simplex* Δ_n. It is homotopy equivalent to a point (its topological realization, as in Definition 4.8, deform retracts on any point of it). Hence Δ_n has the same Betti numbers as the point:

- $\beta_1(\Delta_n) = 1$,
- $\beta_i(\Delta_n) = 0$ for $i > 0$.

Now, if we run the incremental algorithm for homology on Δ_n, but stopping before adding the n-simplex V, we would obtain the Betti numbers of $\partial\Delta_n$. Also, note that the n-simplex is negative. Hence

- $\beta_{n-1}(\partial\Delta_n) = \beta_{n-1}(\Delta_n) + 1$,
- $\beta_i(\partial\Delta_n) = \beta_i(\Delta_n)$ for $i \neq n-1$.

We deduce the result. $\square$

From the homology of the spheres, one deduces the theorem of Invariance of Domain.

Theorem 6.6. *For every integers m, n such that $m \neq n$, the spaces $\mathbb{R}^n$ and $\mathbb{R}^m$ are not homeomorphic.*

Proof. Let m, n such that $m \neq n$. By contradiction, suppose that $\mathbb{R}^n$ and $\mathbb{R}^m$ are homeomorphic via f. Let 0 denote the origin of $\mathbb{R}^n$. By restriction, we get a homeomorphism

$$\mathbb{R}^n \setminus \{0\} \to \mathbb{R}^m \setminus \{f(0)\}.$$

We deduce the following weaker statement: $\mathbb{R}^n \setminus \{0\}$ and $\mathbb{R}^m \setminus \{f(0)\}$ are homotopic equivalent. Now, using Example 3.15, we deduce that the sphere $\mathbb{S}_{n-1}$ and $\mathbb{S}_{m-1}$ are homotopic equivalent. Hence, according to Proposition 5.14, they must admit the same homology groups. This contradict Proposition 6.5. $\square$

Euler characteristic. Finally, we prove that the Euler characteristic is an information already included in the homology groups.

Proposition 6.7. *Let X be a (triangulable) topological space. Then its Euler characteristic is equal to*

$$\chi(X) = \sum_{0\leq i\leq n} (-1)^i \cdot \beta_i(X)$$

where n is the maximal integer such that $\beta_i(X) \neq 0$.

Proof. Let K be a triangulation of X. By definition, we have $\chi(X) = \chi(K)$ and

$$\chi(K) = \sum_{0\leq i\leq n} (-1)^i \cdot (\text{number of simplices of dimension } i).$$

Now, pick an ordering $K^1 \subset \cdots \subset K^n = K$ of K, with $K^i = K^{i-1} \cup \{\sigma^i\}$ for all $2 \leq i \leq n$. We will apply the incremental algorithm. By induction, let us show that, for all $1 \leq m \leq n$,

$$\sum_{0\leq i\leq n} (-1)^i \cdot \beta_i(K^m) = \sum_{0\leq i\leq n} (-1)^i \cdot (\text{number of simplices of dimension } i \text{ of } K^m). \quad (1)$$

For $m = 1$, σ^m is a 0-simplex, and the equality reads $1 = 1$. Now, suppose that the equality is true for $1 \leq m < n$, and consider the simplex σ^{m+1}. Let $d = \dim \sigma^{m+1}$. The right-hand side of Equation (1) is increased by $(-1)^d$.

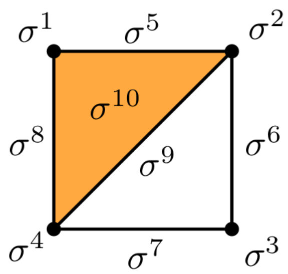
If σ^{m+1} is positive, then $\beta_d(K^{m+1}) = \beta_d(K^m) + 1$, hence the left-hand side of Equation (1) is increased by $(-1)^d$. Otherwise, it is negative, and $\beta_{d-1}(K^{m+1}) = \beta_{d-1}(K^m) - 1$, hence the left-hand side of Equation (1) is increased by $-(-1)^{d-1} = (-1)^d$. We deduce the result by induction. $\square$

6.3 Matrix algorithm

The only thing missing to apply Algorithm 1 is to determine whether a simplex is positive or negative. It turns out that this problem can be conveniently solved by using a matrix representation of the simplicial complex.

Let K be a simplicial complex, and $\sigma^1 < \sigma^2 < \cdots < \sigma^n$ and ordering of its simplices, as in Subsection 6.1. Define the *boundary matrix* of K, denoted Δ, as follows: Δ is a $n \times n$ matrix, whose (i,j)-entry (i^{th} row, j^{th} column is)

$$\begin{aligned} \Delta_{i,j} &= 1 \text{ if } \sigma^i \text{ is a face of } \sigma^j \text{ and } |\sigma^i| = |\sigma^j| - 1 \\ & \quad 0 \text{ else.} \end{aligned}$$



$$\begin{array}{c} \\ \sigma^1 \\ \sigma^2 \\ \sigma^3 \\ \sigma^4 \\ \sigma^5 \\ \sigma^6 \\ \sigma^7 \\ \sigma^8 \\ \sigma^9 \\ \sigma^{10} \end{array} \begin{array}{c} \begin{matrix} \sigma^1 & \sigma^2 & \sigma^3 & \sigma^4 & \sigma^5 & \sigma^6 & \sigma^7 & \sigma^8 & \sigma^9 & \sigma^{10} \end{matrix} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

By adding columns one to the others, we create chains. If we were able to reduce a column to zero, then we found a cycle.

$$\begin{array}{c} \\ \sigma^1 \\ \sigma^2 \\ \sigma^3 \\ \sigma^4 \\ \sigma^5 \\ \sigma^6 \\ \sigma^7 \\ \sigma^8 \\ \sigma^9 \\ \sigma^{10} \end{array} \begin{array}{c} \begin{matrix} \sigma^1 & \sigma^2 & \sigma^3 & \sigma^4 & \sigma^5 & \sigma^6 & \sigma^7 & \sigma^8 & \sigma^9 & \sigma^{10} \end{matrix} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

$$\partial_1(\sigma^6) = \sigma^2 + \sigma^3$$

$$\begin{array}{c} \\ \sigma^1 \\ \sigma^2 \\ \sigma^3 \\ \sigma^4 \\ \sigma^5 \\ \sigma^6 \\ \sigma^7 \\ \sigma^8 \\ \sigma^9 \\ \sigma^{10} \end{array} \begin{array}{c} \begin{matrix} \sigma^1 & \sigma^2 & \sigma^3 & \sigma^4 & \sigma^5 & \sigma^6 & \sigma^7 & \sigma^5+\sigma^6+\sigma^7+\sigma^8 & \sigma^9 & \sigma^{10} \end{matrix} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

$$\partial_1(\sigma^5 + \sigma^6 + \sigma^7 + \sigma^8) = 0$$

The process of reducing columns to zero is called *Gauss reduction*. For any $j \in [\![1, n]\!]$, define

$$\delta(j) = \max\{i \in [\![1, n]\!], \Delta_{i,j} \neq 0\}.$$

If $\Delta_{i,j} = 0$ for all j, then $\delta(j)$ is *undefined*. We say that the boundary matrix Δ is *reduced* if the map δ is injective on its domain of definition. The following algorithm allows to compute a reduced matrix.

Algorithm 2: Reduction of the boundary matrix

Input: a boundary matrix Δ

Output: a reduced matrix $\widetilde{\Delta}$

for $j \leftarrow 1$ **to** n **do**

  while *there exists* $i < j$ *with* $\delta(i) = \delta(j)$ **do**

    add column i to column j;

We give the first iterations of the algorithm:

$$
\begin{array}{c}
\\ \sigma^1 \\ \sigma^2 \\ \sigma^3 \\ \sigma^4 \\ \sigma^5 \\ \sigma^6 \\ \sigma^7 \\ \sigma^8 \\ \sigma^9 \\ \sigma^{10}
\end{array}
\begin{array}{c}
\begin{array}{cccccccccc}
\sigma^1 & \sigma^2 & \sigma^3 & \sigma^4 & \sigma^5 & \sigma^6 & \sigma^7 & \sigma^8 & \sigma^9 & \sigma^{10}
\end{array} \\
\begin{pmatrix}
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
\end{array}
$$

$$
\begin{array}{c}
\\ \sigma^1 \\ \sigma^2 \\ \sigma^3 \\ \sigma^4 \\ \sigma^5 \\ \sigma^6 \\ \sigma^7 \\ \sigma^8 \\ \sigma^9 \\ \sigma^{10}
\end{array}
\begin{array}{c}
\begin{array}{cccccccccc}
\sigma^1 & \sigma^2 & \sigma^3 & \sigma^4 & \sigma^5 & \sigma^6 & \sigma^7 & \sigma^8+\sigma^7 & \sigma^9 & \sigma^{10}
\end{array} \\
\begin{pmatrix}
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
\end{array}
$$

$$
\begin{array}{c}
\\ \sigma^1 \\ \sigma^2 \\ \sigma^3 \\ \sigma^4 \\ \sigma^5 \\ \sigma^6 \\ \sigma^7 \\ \sigma^8 \\ \sigma^9 \\ \sigma^{10}
\end{array}
\begin{array}{c}
\begin{array}{cccccccccc}
\sigma^1 & \sigma^2 & \sigma^3 & \sigma^4 & \sigma^5 & \sigma^6 & \sigma^7 & \sigma^8+\sigma^7 & \sigma^9 & \sigma^{10}
\end{array} \\
\begin{pmatrix}
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
\end{array}
$$

$$
\begin{array}{c}
\\ \sigma^1 \\ \sigma^2 \\ \sigma^3 \\ \sigma^4 \\ \sigma^5 \\ \sigma^6 \\ \sigma^7 \\ \sigma^8 \\ \sigma^9 \\ \sigma^{10}
\end{array}
\begin{array}{c}
\begin{array}{cccccccccc}
\sigma^1 & \sigma^2 & \sigma^3 & \sigma^4 & \sigma^5 & \sigma^6 & \sigma^7 & \sigma^8+\sigma^7+\sigma^6 & \sigma^9 & \sigma^{10}
\end{array} \\
\begin{pmatrix}
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
\end{array}
$$

$$\begin{array}{cc} & \begin{matrix} \sigma^1 & \sigma^2 & \sigma^3 & \sigma^4 & \sigma^5 & \sigma^6 & \sigma^7 & \sigma^8+\sigma^7+\sigma^6 & \sigma^9 & \sigma^{10} \end{matrix} \\ \begin{matrix} \sigma^1 \\ \sigma^2 \\ \sigma^3 \\ \sigma^4 \\ \sigma^5 \\ \sigma^6 \\ \sigma^7 \\ \sigma^8 \\ \sigma^9 \\ \sigma^{10} \end{matrix} & \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

$$\begin{array}{cc} & \begin{matrix} \sigma^1 & \sigma^2 & \sigma^3 & \sigma^4 & \sigma^5 & \sigma^6 & \sigma^7 & \sigma^8+\sigma^7+\sigma^6+\sigma^5 & \sigma^9 & \sigma^{10} \end{matrix} \\ \begin{matrix} \sigma^1 \\ \sigma^2 \\ \sigma^3 \\ \sigma^4 \\ \sigma^5 \\ \sigma^6 \\ \sigma^7 \\ \sigma^8 \\ \sigma^9 \\ \sigma^{10} \end{matrix} & \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

...

$$\begin{array}{cc} & \begin{matrix} \sigma^1 & \sigma^2 & \sigma^3 & \sigma^4 & \sigma^5 & \sigma^6 & \sigma^7 & \sigma^5+\sigma^6+\sigma^7+\sigma^8 & \sigma^9+\sigma^6+\sigma^7 & \sigma^{10} \end{matrix} \\ \begin{matrix} \sigma^1 \\ \sigma^2 \\ \sigma^3 \\ \sigma^4 \\ \sigma^5 \\ \sigma^6 \\ \sigma^7 \\ \sigma^8 \\ \sigma^9 \\ \sigma^{10} \end{matrix} & \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

Exercise 33. Show that the algorithm stops after a finite number of steps.

At the end of the algorithm, we can read the positivity of the simplices:

Lemma 6.8. *Suppose that the boundary matrix is reduced. Let $j \in [\![1, n]\!]$. If $\delta(j)$ is defined, then the simplex σ^j is negative. Otherwise, it is positive.*

Proof. Indeed, at the end of the algorithm, $\delta(j)$ is undefined if and only if σ^i is included in a cycle of K^i, that is, if σ^i if positive. $\square$

As a consequence, we can read on the reduced boundary matrix the positivity of the simplices. Combined with Algorithm 1, we are able to compute the Betti numbers of any simplicial complex.

For instance, we read on the following matrix that the only positive simplices are σ^1, σ^2, σ^3, σ^4, σ^8 and σ^9.

$$
\begin{array}{cc}
 & \begin{array}{cccccccccc} \sigma^1 & \sigma^2 & \sigma^3 & \sigma^4 & \sigma^5 & \sigma^6 & \sigma^7 & \sigma^5+\sigma^6+\sigma^7+\sigma^8 & \sigma^9+\sigma^6+\sigma^7 & \sigma^{10} \end{array} \\
\begin{array}{c} \sigma^1 \\ \sigma^2 \\ \sigma^3 \\ \sigma^4 \\ \sigma^5 \\ \sigma^6 \\ \sigma^7 \\ \sigma^8 \\ \sigma^9 \\ \sigma^{10} \end{array} &
\begin{pmatrix}
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix} \\
 & \begin{array}{cccccccccc} \sigma^1 & \sigma^2 & \sigma^3 & \sigma^4 & \sigma^5 & \sigma^6 & \sigma^7 & \sigma^8 & \sigma^9 & \sigma^{10} \\ + & + & + & + & - & - & - & + & + & - \end{array}
\end{array}
$$

Remark 6.9. Algorithm 2, also called the *standard algorithm for reduction of the boundary matrix*, is the one developped first, in the paper [ELZ00]. Then, many other algorithms have been proposed to reduce the boundary matrix. See the review [OPT+17].

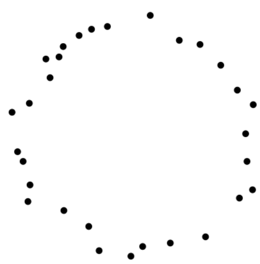
> **Exercise 34.** Apply Algorithm 2 to solve Exercise 31.

7 Topological inference

7.1 Thickenings

In real life, we are often given datasets that are subsets of the Euclidean space: $X \subset \mathbb{R}^n$. Of course, X is finite. And it has no interesting topology.

The Betti numbers of $X \subset \mathbb{R}^2$ on the right are $\beta_0(X) = 30$ and $\beta_1(X) = 0$. It is a *discrete* topological space.



However, in Topological Data Analysis, we think of X as being a sample of an underlying continuous object, $\mathcal{M} \subset \mathbb{R}^n$. Understanding the topology of $\mathcal{M}$ would give us interesting insights about our dataset.

X $\qquad\qquad$ $\mathcal{M}$

With Persistent Homology, we aim at understanding the topology of $\mathcal{M}$ via its homology. The problem of homology inference can be stated as follows:

Let $\mathcal{M} \subset \mathbb{R}^n$ be a bounded subset.
Suppose that we are given a finite sample $X \subset \mathcal{M}$.
Estimate the homology groups of $\mathcal{M}$ from X.

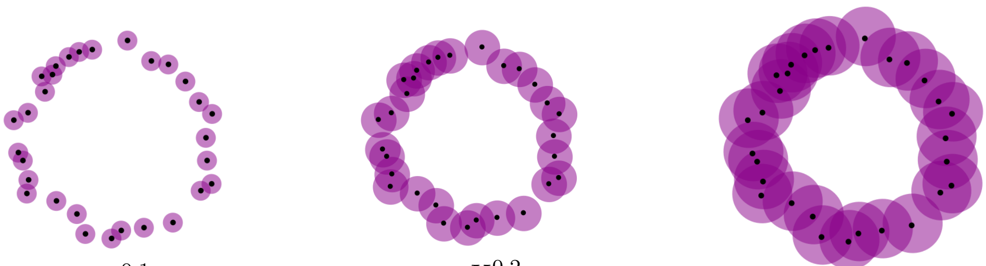
Unfortunately, the homology of X is very far from the homology of $\mathcal{M}$. However, if it is sampled close enough to $\mathcal{M}$, there exists a construction that allows to recover the homotopy type of $\mathcal{M}$ from X, hence its homology groups as well. This construction consists in *thickening* X.

For every $t \geq 0$, the *t-thickening* of the set X, denoted X^t, is the set of points of the ambient space with distance at most t from X:

$$X^t = \{y \in \mathbb{R}^n, \exists x \in X, \|x - y\| \leq t\}.$$

Equivalently, X^t can be seen as a union of closed balls centered around every point of X:

$$X^t = \bigcup_{x \in X} \overline{\mathcal{B}}(x, t).$$



$X^{0.1}$ $\qquad\qquad$ $X^{0.2}$ $\qquad\qquad$ $X^{0.3}$

Observe that the last figure is a thickening which has the homotopy type of a circle: $X^t \approx \mathcal{M}$ (it deform retracts on it). If we are able to select such a t, then we have access to the homology groups of $\mathcal{M}$:

$$\forall i \geq 0, H_i(\mathcal{M}) \simeq H_i(X).$$

We are in front of two questions:

1. How to select a t such that $X^t \approx \mathcal{M}$?
2. How to compute the homology groups of X^t?

We will give a partial answer to Question 1 in this subsection, and to Question 2 in the next one. First, we need to define some geometric quantities.

Hausdorff distance. Let X be any subset of $\mathbb{R}^n$. The function *distance to* X is the map

$$\begin{aligned} \operatorname{dist}(\cdot, X) : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ y &\longmapsto \operatorname{dist}(y, X) = \inf\{\|y - x\|, x \in X\} \end{aligned}$$

A *projection* of $y \in \mathbb{R}^n$ on X is a point $x \in X$ which attains this infimum. If such a point x exists and is *unique*, we denote it $\operatorname{proj}(y, X)$.

Exercise 35. Prove that, when X is closed, a projection always exists.
To meditate: In general, is the projection unique?

If Y is another subset of $\mathbb{R}^n$, we define the distance from Y to X as

$$\mathrm{d_H}(X; Y) = \sup\{\operatorname{dist}(y, X), y \in Y\}.$$

Also, we define the *Hausdorff distance* between X and Y as

$$d_{\mathrm{H}}(X, Y) = \max\{\mathrm{d_H}(X; Y), \mathrm{d_H}(Y; X)\}.$$

If X and Y are bounded, their Hausdorff distance is finite. We can also write the Hausdorff distance as follows:

$$\begin{aligned} d_{\mathrm{H}}(X, Y) &= \max\left\{\sup_{y\in Y} \operatorname{dist}(y, X),\ \sup_{x\in X} \operatorname{dist}(x, Y)\right\} \\ &= \max\left\{\sup_{y\in Y} \inf_{x\in X} \|x - y\|,\ \sup_{x\in X} \inf_{y\in Y} \|x - y\|\right\}. \end{aligned}$$

Exercise 36. Let $\|\cdot\|_\infty$ be the sup norm of function $f\colon \mathbb{R}^n \to \mathbb{R}^m$: $\|f\|_\infty = \sup_{x\in\mathbb{R}^n} \|f(x)\|$. Prove that $d_{\mathrm{H}}(X, Y) = \|\operatorname{dist}(\cdot, X) - \operatorname{dist}(\cdot, Y)\|_\infty$.

Exercise 37. Let X, Y be two closed and bounded subsets of $\mathbb{R}^n$. Show that for every $t \geq 0$, the thickenings satisfy

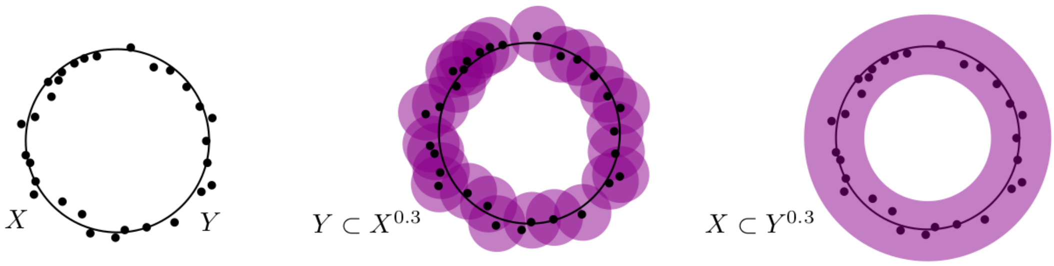
$$d_{\mathrm{H}}(X^t, Y^t) \leq d_{\mathrm{H}}(X, Y).$$

Give an example for which $d_{\mathrm{H}}(X^t, Y^t) < d_{\mathrm{H}}(X, Y)$.

Exercise 38. Show that the Hausdorff distance is equal to

$$\inf\left\{t \geq 0, X \subset Y^t \text{ and } Y \subset X^t\right\}.$$

In other words, the Hausdorff distance is the smallest value of t such that Y is included in the t-thickening of X, and vice versa.

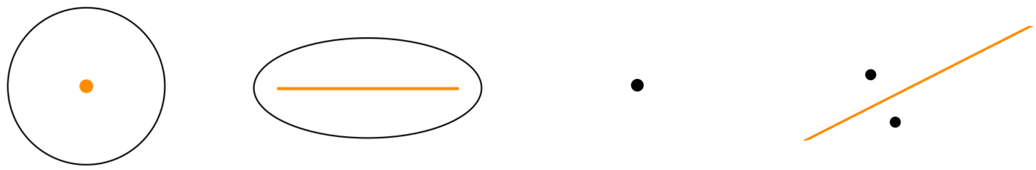


Medial axis. Let X be any subset of $\mathbb{R}^n$. The *medial axis* of X is the subset $\mathrm{med}\,(X) \subset \mathbb{R}^n$ which consists of points $y \in \mathbb{R}^n$ that admit at least two projections on X:

$$\mathrm{med}\,(X) = \left\{ y \in \mathbb{R}^n, \exists x, x' \in X, x \neq x', \|y - x\| = \|y - x'\| = \mathrm{dist}\,(y, X) \right\}.$$

Example 7.1. In $\mathbb{R}^2$,

- the medial axis of a circle is its center,
- the medial axis of an ellipse is an interval,
- the medial axis of a point is the emptyset,
- the medial axis of two distinct points is their bisector.



Exercise 39. Compute the reach of the following subsets of $\mathbb{R}^2$:

- the set $\{(0, n), n \in \mathbb{Z}\}$,
- the segment $\{(t, 0), t \in [0, 1]\}$,
- the unit circle with origin $\mathbb{S}_1 \cup \{(0, 0)\}$,
- the square $\{(x, y) \in \mathbb{R}^2, \max\{|x|, |y|\} = 1\}$,
- *(more difficult)* the ellipse $\left\{(x_1, x_2) \in \mathbb{R}^2, \left(\frac{x_1}{a}\right)^2 + \left(\frac{x_2}{b}\right)^2 = 1\right\}$, $a, b > 0$.

Reach. Now, we define the *reach* of X as its proximity from its medial axis:

$$\begin{aligned} \mathrm{reach}(X) &= \inf \left\{ \mathrm{dist}\,(y, X), y \in \mathrm{med}\,(X) \right\} \\ &= \inf \left\{ \|x - y\|, x \in X, y \in \mathrm{med}\,(X) \right\}. \end{aligned}$$

One shows that the reach of X is equal to the supremum of $t \geq 0$ such that the thickening X^t does not intersect $\mathrm{med}\,(X)$. In other words,

$$\mathrm{reach}(X) = \sup\{t \geq 0, X^t \cap \mathrm{med}\,(X) = \emptyset\}$$

Exercise 40. Compute the reaches of the subsets of Exercise 39.

Suppose that X is closed and that $\mathrm{reach}(X)$ is positive. Then for every $t \in [0, \mathrm{reach}(X))$, the thickening X^t deform retracts onto X. A homotopy is given by the following map:

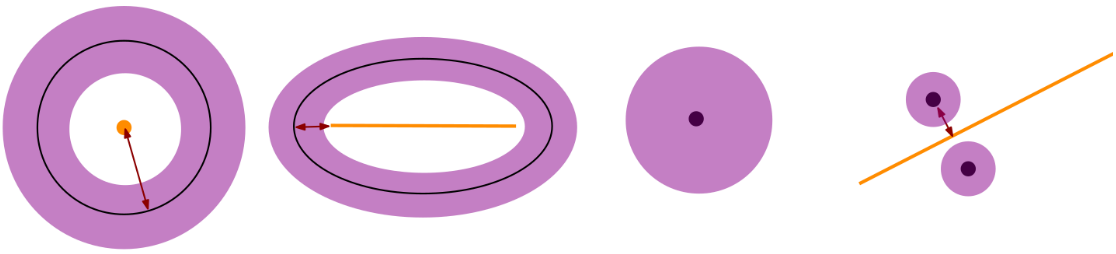
$$\begin{aligned} X^t \times [0,1] &\longrightarrow X^t \\ (x,t) &\longmapsto (1-t)x + t \cdot \mathrm{proj}\,(x, X)\,. \end{aligned}$$

The projection map $\mathrm{proj}\,(\cdot, X) : X^t \to X$ is well defined since, for every $t < \mathrm{reach}(X)$, the points $x \in X^t$ admits a unique projection. We deduce the following proposition:

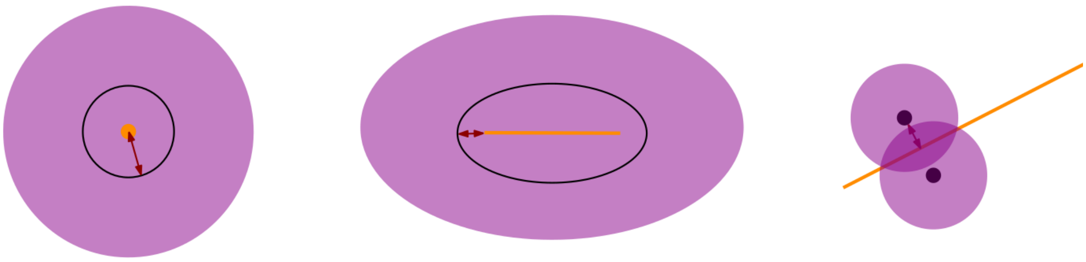
Proposition 7.2. *For every $t \in [0, \mathrm{reach}(X))$, the spaces X and X^t are homotopy equivalent.*

Hence the reach acts as a threshold below which the thickenings have the same homotopy type as the original subset. Over this value, the homotopy type may change. Note however that the converse is not true: we may have $X \approx X^t$, even with $t \geq \mathrm{reach}(X)$.

As an example, here are some thickenings of Example 7.1 that have the same homotopy type than the initial subset:



At some point, they become homotopy equivalent to a point:



Back to our problem: given a finite subset X that samples an underlying object $\mathcal{M}$, can we find a t such that X^t is homotopy equivalent to $\mathcal{M}$? We give two such results, proven in 2009 and 2008. The key conditions are the following: the reach of $\mathcal{M}$ has to be large enough, and the Hausdorff distance $d_{\mathrm{H}}(X, \mathcal{M})$ small enough.

Theorem 7.3 (Corollary of [CCSL09, Theorem 4.6, case $\mu = 1$])**.** *Let X and $\mathcal{M}$ be subsets of $\mathbb{R}^n$. Suppose that $\mathcal{M}$ has positive reach, and that $d_{\mathrm{H}}(X, \mathcal{M}) \leq \frac{1}{17}\mathrm{reach}(\mathcal{M})$. Then X^t and $\mathcal{M}$ are homotopic equivalent, provided that*

$$t \in [4 d_{\mathrm{H}}(X, \mathcal{M}), \mathrm{reach}(\mathcal{M}) - 3 d_{\mathrm{H}}(X, \mathcal{M}))\,.$$

Theorem 7.4 ([NSW08, Proposition 3.1]). *Let X and $\mathcal{M}$ be subsets of $\mathbb{R}^n$, with $\mathcal{M}$ a submanifold, and X a finite subset of $\mathcal{M}$. Suppose that $\mathcal{M}$ has positive reach. Then X^t and $\mathcal{M}$ are homotopic equivalent, provided that*

$$t \in \left[2d_{\mathrm{H}}(X, \mathcal{M}), \sqrt{\frac{3}{5}}\mathrm{reach}(\mathcal{M})\right).$$

Remark 7.5. In practice, these theorems do not directly solve our Question 1. Indeed, they give formulas for the values of t that we are looking for, but the formulas depends on quantities that we do not know ($d_{\mathrm{H}}(X, \mathcal{M})$ and $\mathrm{reach}(\mathcal{M})$). We have to wait for Persistent Homology to obtain a satisfactory solution to this problem.

7.2 Čech complex

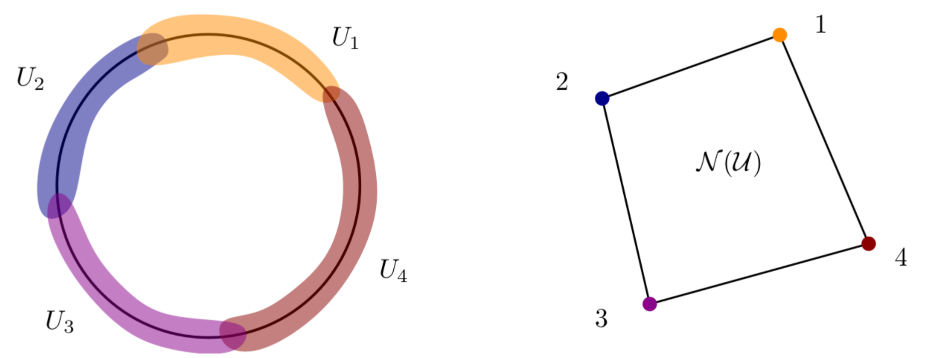
Remember Question 2: if $X \in \mathbb{R}^n$ is a finite subset and $t \geq 0$, we want to compute the homology groups $H_0(X^t), \dots, H_n(X^t)$. To do so, we have to find a triangulation of X^t, that is, a simplicial complex K homeomorphic to X. Actually, we're going to look for something a little weaker: we want a simplicial complex K that is homotopy equivalent to X.

Remark 7.6. I am a bit annoyed that such an object doesn't have a name in mathematics... I propose the following definition (be careful, nobody uses that): a simplicial complex K is a *weak triangulation* of a topological space X if $|K|$ (the topological realization of K) and X are homotopy equivalent.

It turns out that it is easy to represent the thickenings X^t as simplicial complexes, via the notion of covers. In general, if X is any topological space, a *cover* of X is a collection $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$ of subsets $U_i \subset X$ such that

$$\bigcup_{1 \leq i \leq N} U_i = X.$$

Definition 7.7. Let X be a topological space, and $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$ a cover of X The *nerve* of $\mathcal{U}$ is the simplicial complex with vertex set $\{1, ..., N\}$ and whose m-simplices are the subsets $\{i_1, ..., i_m\} \subset \{1, ..., N\}$ such that $\bigcap_{k=0}^{m} U_{i_k} \neq \emptyset$. It is denoted $\mathcal{N}(\mathcal{U})$.

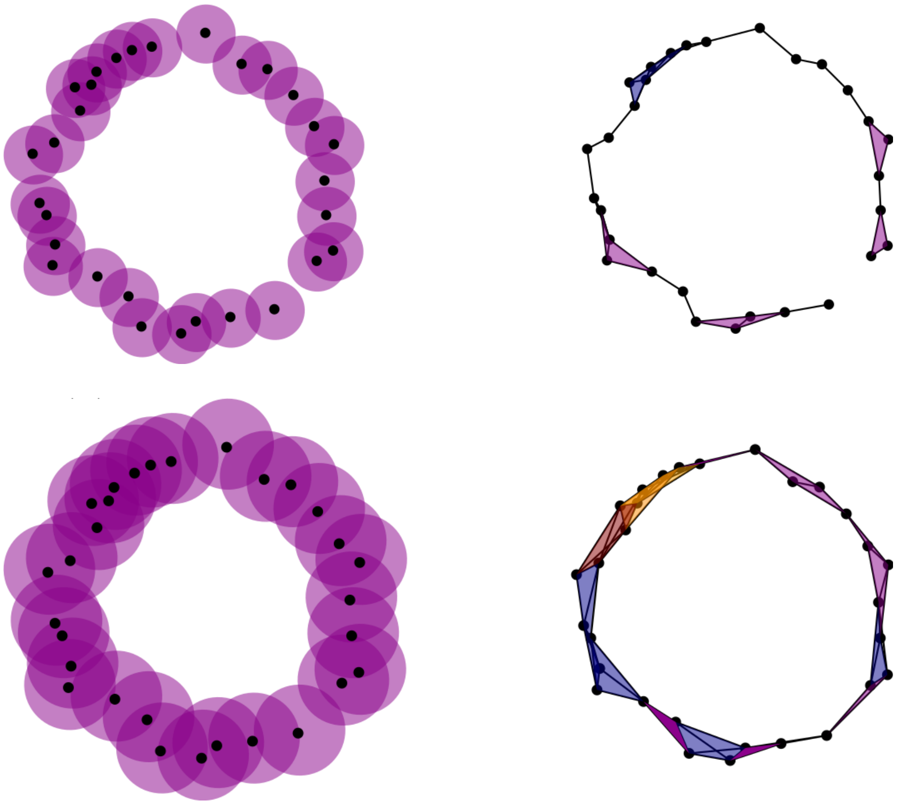


Now let X be a finite subset of $\mathbb{R}^n$, and $t \geq 0$. Consider the collection

$$\mathcal{V}^t = \left\{\overline{\mathcal{B}}\left(x, t\right), x \in X\right\}.$$

This is a cover of the thickening X^t, and each components are closed balls. Consequently, we can consider its nerve $\mathcal{N}(\mathcal{V}^t)$. The following theorem states that it has the homotopy type of X^t.

Theorem 7.8 ([BCY18, Theorem 2.9])**.** *Suppose that Y is a subset of $\mathbb{R}^n$. Consider a cover $\mathcal{U} = \{U_i, 1 \leq i \leq N\}$ of Y such that each of the U_i are balls (or more generally, closed and convex). Then $\mathcal{N}(\mathcal{U})$ is homotopy equivalent to Y.*



Definition 7.9. Let $t \geq 0$ and consider the collection $\mathcal{V}^t = \left\{\overline{\mathcal{B}}(x,t), x \in X\right\}$. Its nerve is denoted $\check{\mathrm{C}}\mathrm{ech}^t(X)$ and is called the *Čech complex of X at time t*.

The homotopy equivalence between $\check{\mathrm{C}}\mathrm{ech}^t(X)$ and X implies that all the homology groups $H_i(\check{\mathrm{C}}\mathrm{ech}^t(X))$ and $H_i(X^t)$ are isomorphic. Therefore, we are able to compute the homology of the thickening X^t.

Remark 7.10. We gave here a convenient version of the Nerve Theorem, but it actually holds with weaker hypothesis. See for instance [Hat02, Corollary 4G.3] (available online at `https://pi.math.cornell.edu/~hatcher/AT/AT.pdf`).

7.3 Rips complex

Let $X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^n$ be finite, let $t \geq 0$ and consider the t-thickening

$$X^t = \bigcup_{x \in X} \overline{\mathcal{B}}(x,t).$$

By definition, its nerve, $\check{\mathrm{C}}\mathrm{ech}^t(X)$, the Čech complex at time t, is a simplicial complex on the vertices $\{1, \dots, N\}$ whose simplices are the subsets $\{i_1, ..., i_m\}$ such that

$$\bigcap_{1 \leq k \leq m} \overline{\mathcal{B}}\left(x_{i_k}, t\right) \neq \emptyset.$$

Therefore, computing the Čech complex relies on the following geometric predicate:

Given m closed balls of $\mathbb{R}^n$, do they intersect?

This problem is known as the *smallest circle problem*. It can can be solved in $O(m)$ time, where m is the number of points, but with a constant depending on the dimension of the ambiant space $\mathbb{R}^n$. However, in practice, we prefer a more simple version of the Čech complex, that does not require this predicate.

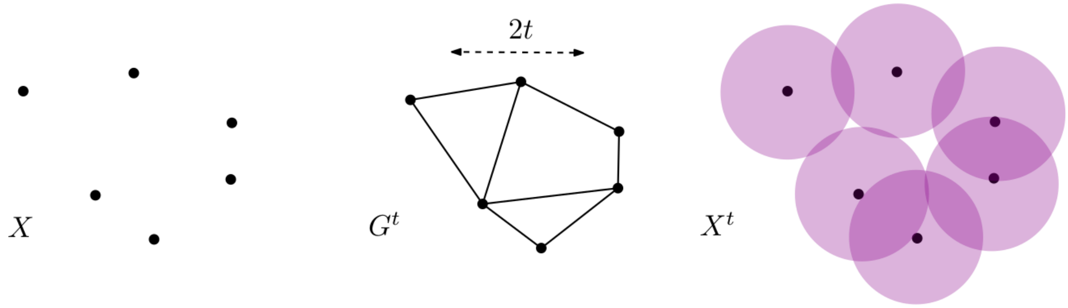
We will need the following notions: Let G be a graph. We call a *clique* of G a set of vertices $v_1, ..., v_m$ such that for every $i, j \in [\![1, m]\!]$ with $i \neq j$, the edge $[v_1, v_j]$ belongs to G. In other words, the subgraph G restricted to the vertices $v_1, ..., v_m$ is *complete*.

Given a graph G, the corresponding *clique complex* is the simplicial complex whose vertices are the vertices of G, and whose simplices are the sets of vertices of the cliques of G. Some authors also call it the *expansion* of G.

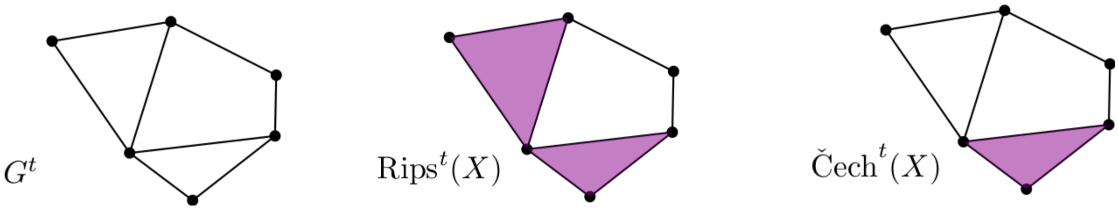
Exercise 41. Verify that the clique complex of a graph is a simplicial complex. If the graph contains n vertices, give an upper bound on the number of simplices of the clique complex.

Let us get back to the set $X = \{x_1, \dots, x_N\}$. Let $t \geq 0$. Consider the graph G^t whose vertex set is $\{1, \dots, N\}$, and whose edges are the pairs (i, j) such that $\|x_i - x_j\| \leq 2t$. Alternatively, G^t can be seen as the 1-skeleton of the Čech complex $\check{\mathrm{C}}\mathrm{ech}^t(X)$

$$G^t = \left(\check{\mathrm{C}}\mathrm{ech}^t(X)\right)_1$$



Definition 7.11. The *Rips complex of X at time t* is the clique complex of the graph G^t defined above. We denote it $\mathrm{Rips}^t(X)$.



Note that the Rips complex may not be homotopy equivalent to the corresponding Čech complex. However, they are not linked by the following relation:

Proposition 7.12. *Let $X \subset \mathbb{R}^n$ be a finite subset. For every $t \geq 0$, we have*

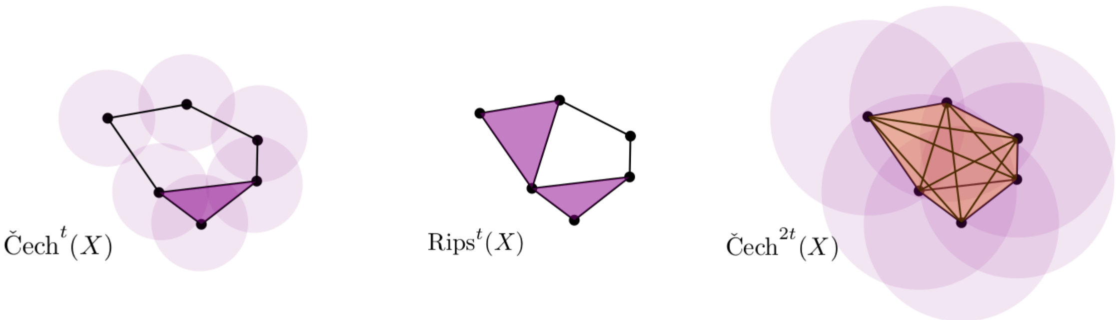
$$\check{\mathrm{C}}\mathrm{ech}^t(X) \subset \mathrm{Rips}^t(X) \subset \check{\mathrm{C}}\mathrm{ech}^{2t}(X).$$

Proof. Let $t \geq 0$. The first inclusion follows from the fact that $\mathrm{Rips}^t(X)$ is the clique complex of $\check{\mathrm{C}}\mathrm{ech}^t(X)$.

To prove the second one, choose a simplex $\sigma \in \mathrm{Rips}^t(X)$. Let us prove that $\sigma \in \check{\mathrm{C}}\mathrm{ech}^{2t}(X)$. Let $x \in \sigma$ be any vertex. Note that $\forall y \in \sigma$, we have $\|x - y\| \leq 2t$ by definition of the Rips complex. Therefore

$$x \in \bigcap_{y \in \sigma} \overline{\mathcal{B}}\left(y, 2t\right).$$

The intersection being non-empty, we deduce $\sigma \in \check{\mathrm{C}}\mathrm{ech}^{2t}(X)$. $\square$



Exercise 42. Improve the previous proposition as follows: $\check{\mathrm{C}}\mathrm{ech}^t(X) \subset \mathrm{Rips}^t(X) \subset \check{\mathrm{C}}\mathrm{ech}^{ct}(X)$, where $c = \sqrt{\frac{2n}{n+1}}$.
Warning: Not easy to prove. This is Theorem 2.5 of [DSG07].

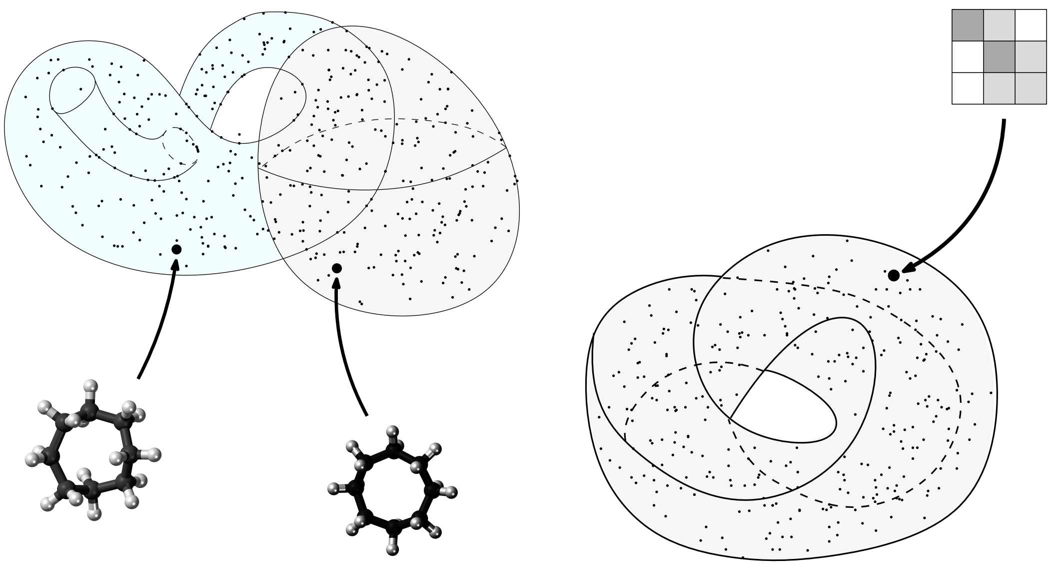
8 Datasets have topology

8.1 Some examples

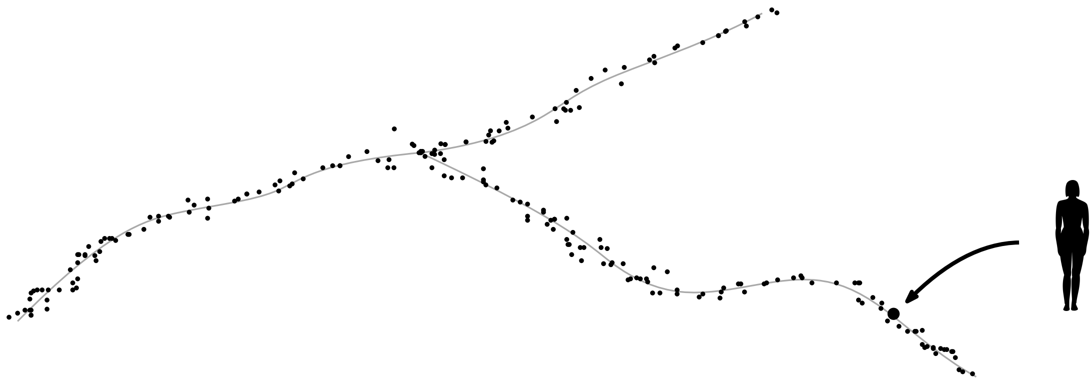
We give three datasets where interesting topology appears.

Cyclo-octane molecules. The first one comes from chemistry [MTCW10], where is studied the cyclo-octane molecule C_8H_{16}. The configuration of such a molecule can be represented by 72 variables—the 3D coordinates of each of its 24 atoms—, or equivalently, by a point in $\mathbb{R}^{72}$. By analyzing many of these molecules, the authors obtain a point cloud in $\mathbb{R}^{72}$. In this large dimensional space, it turns out that the point cloud lies on an object of much smaller dimension, namely, the union of a sphere and a Klein bottle, intersecting in two rings. These two components correspond to distinct spatial arrangements of the molecule: crown conformation in the sphere, and boat-chair conformation in the Klein bottle. The behavior of molecules lying in the intersection is still an open question.

Natural images. A second example comes from image processing [CIDSZ08]. From a large collection of natural images, the authors extract 3×3 patches. Since it consists of 9 pixels, each of these patches can be seen as a 9-dimensional vector, and the whole set as a point cloud in $\mathbb{R}^9$. It appears that this dataset concentrates near an object that has the homology of a Klein bottle. In a second step, the authors show that a significant part of the points (60%) are well approximated by an embedding of the Klein bottle in $\mathbb{R}^9$. This discovery has led to the construction of Klein-bottle-based image analysis methods [PC14, CG20].

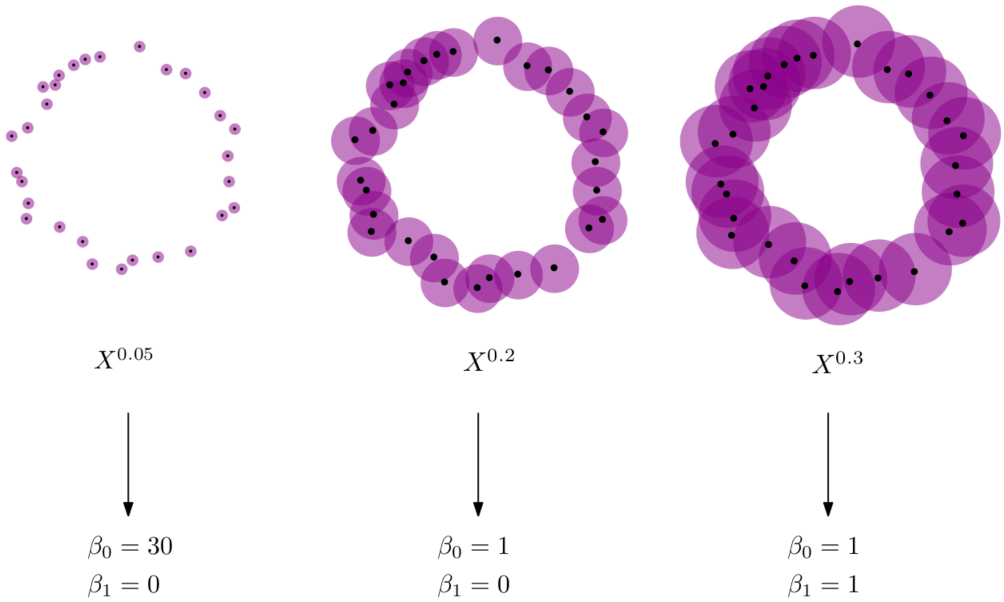


Breast cancer. We give a last example, taken from biomedicine [NLC11]. Tissues of patients infected with breast cancer has been analyzed, resulting in 262 genomic variables per patients. Gathering these data yields a point cloud in $\mathbb{R}^{262}$. In a different context from the two previous examples, the analysis here consists in reducing the dimension of the dataset, while not changing its topology too much. More precisely, one is looking for its 1-dimensional structure, known as the Reeb graph. This is performed in practice with the so-called MAPPER algorithm [SMC07]. The result is a graph, which turned out to be composed of three distinct branches. Taking advantage of this structure, the authors discovered an unexpected subset of patients: they exhibit a 100% survival, and no metastasis. They correspond to a unique molecular signature, that yields to the designation of a new type of breast cancer (*c-MYB*$^+$).



8.2 Betti curves

We introduce a new tool before the tutorial. Let X be a subset of $\mathbb{R}^n$, and consider its thickenings X^t, $t \geq 0$. We have studied in the last section the problem of computing the homology of these thickenings, via the complexes $\check{\mathrm{C}}\mathrm{ech}^t(X)$ and $\mathrm{Rips}^t(X)$.



In practice, we may want to understand the evolution of this homology, when t grows. This is the aim of the following definition.

Definition 8.1. Let $X \subset \mathbb{R}^n$ and $i \geq 0$. The i^{th} *Betti curve* of X is the map

$$\begin{aligned} \beta_i(t) \colon \mathbb{R}^+ &\longrightarrow \mathbb{N} \\ t &\longmapsto \beta_i(X^t) \end{aligned}$$

As a consequence of the nerve theorem, the map $t \mapsto \beta_i(t)$ is equal to $t \mapsto \beta_i(\check{\mathrm{C}}\mathrm{ech}^t(X))$. In practice, we may use the following map, called the i^{th} *Betti curve of the Rips complex of* X:

$$\begin{aligned} \beta_i^{\mathrm{Rips}}(t) \colon \mathbb{R}^+ &\longrightarrow \mathbb{N} \\ t &\longmapsto \beta_i(\mathrm{Rips}^t(X)) \end{aligned}$$

Exercise 43. Show that $t \mapsto \beta_0(t)$ is non-increasing. Show that $t \mapsto \beta_0^{\mathrm{Rips}}(t)$ is also non-increasing.

8.3 Python tutorial

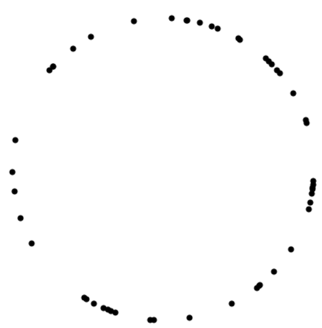
Notebook available at
https://github.com/raphaeltinarrage/EMAp/blob/main/Tutorial2.ipynb.

Our code starts with

```
import gudhi
import numpy as np
import matplotlib.pyplot as plt
```

Let us generate a dataset and plot it.

```
N = 50      # number of points
X = SampleOnCircle(N)

# plotting the point cloud
fig = plt.figure(figsize=(8,8)); ax = fig.add_subplot(1,1,1)
plt.scatter(X[:,0],X[:,1], c='black', s=50)
plt.axis('equal'); plt.axis('off'); plt.show()
```



We now build its Rips complex at time t. First, we add the simplices of the underlying graph (vertices and edges), then we compute its clique complex, via the function `expansion`. We have to give `expansion` a parameter: the maximal dimension of the simplices to add in the complex. If we want to compute n-homology accurately, we have to insert the simplices up to dimension $n+1$, hence we use `expansion(n+1)`.

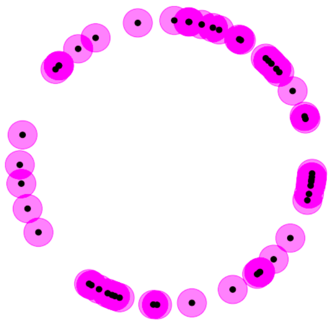
```
t = 0.1

# adding the 0-simplices (vertices)
st = gudhi.SimplexTree()
for i in range(N):
    st.insert([i])
```

```
# adding the 1-simplices (edges)
for i in range(N):
    for j in range(i):
        if np.linalg.norm(X[i,:]-X[j,:])<=2*t:
            st.insert([i,j])
# computing the clique complex (up to the 2-simplices)
st.expansion(2)
```

Now we can read the Betti numbers with `BettiNumbers(st)`:

```
The Betti numbers are:
Beta_0 = 11
Beta_1 = 0
```

When X is a subset of the plane, we can visualize the thickenings thanks to the function `PlotThickening`.



With GUDHI, we can also obtain the Rips complex at time t from a point cloud X as follows:

```
t = 0.1
rips = gudhi.RipsComplex(X, max_edge_length = 2*t)
st = rips.create_simplex_tree(max_dimension=2)
```

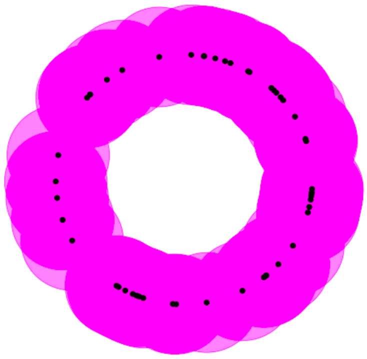
We repeat the experiment with another value of t:

```
t = 0.4

# adding the 0-simplices (vertices)
st = gudhi.SimplexTree()
for i in range(N):
    st.insert([i])
# adding the 1-simplices (edges)
for i in range(N):
    for j in range(i):
        if np.linalg.norm(X[i,:]-X[j,:])<=2*t:
            st.insert([i,j])

# computing the clique complex (up to the 2-simplices)
st.expansion(2)
```

```
BettiNumbers(st)

PlotThickening(X,t)
```

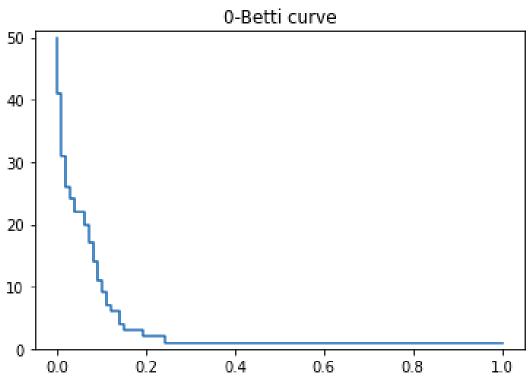
We obtain the following result:

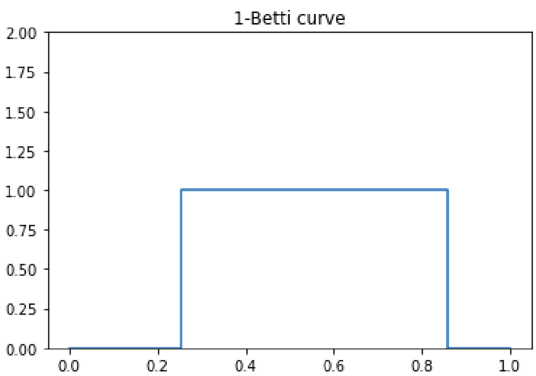
```
The Betti numbers are:
Beta_0 = 1
Beta_1 = 1
```



Now we compute the Betti curves of the point cloud $X \subset \mathbb{R}^2$, with the function `GetBettiCurvesFromPointCloud`.

```
I = np.linspace(0,1,100)
Betti_curves = GetBettiCurvesFromPointCloud(X, I)

plt.figure()
plt.step(I, Betti_curves[0])
plt.ylim(0, max(Betti_curves[0])+1)
plt.title('0-Betti curve')
plt.show()

plt.figure()
plt.step(I, Betti_curves[1])
plt.ylim(0, max(Betti_curves[1])+1)
plt.title('1-Betti curve')
plt.show()
```

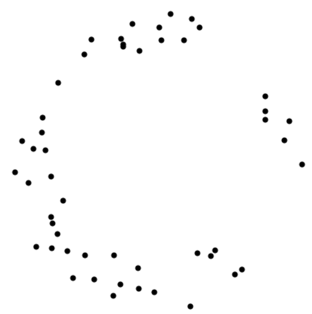




We observe that there exists an interval $I \subset [0, 1]$ on which the thickenings have the homology of a circle.

Finally, we repeat the experiment with another point cloud: a *noisy* sample of the circle.

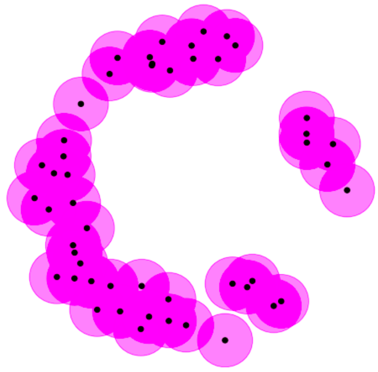
```
N = 50         # number of points
sd = 0.1       # standard deviation of the noise
X = SampleOnCircle(N, sd)

# plotting the point cloud
fig = plt.figure(figsize=(8,8)); ax = fig.add_subplot(1, 1, 1)
plt.scatter(X[:,0],X[:,1], c='black', s=50)
plt.axis('equal'); plt.axis('off'); plt.show()
```



```
# adding the 0-simplices (vertices)
st = gudhi.SimplexTree()
for i in range(N):
    st.insert([i])

# adding the 1-simplices (edges)
for i in range(N):
    for j in range(i):
        if np.linalg.norm(X[i,:]-X[j,:])<=2*t:
            st.insert([i,j])

# computing the clique complex (up to the 2-simplices)
st.expansion(2)
```

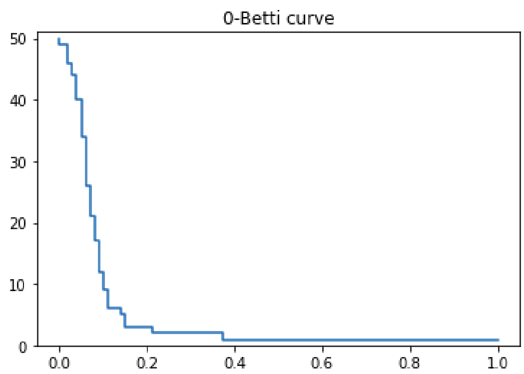
With the functions `BettiNumbers(st)` and `PlotThickening(X,t)`, we obtain

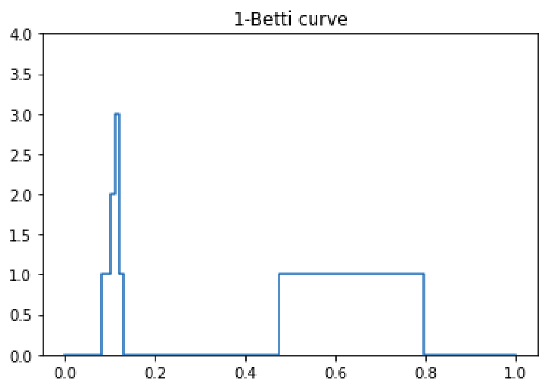
```
The Betti numbers are:
Beta_0 = 3
Beta_1 = 0
```



Now we compute the Betti curves of X. Observe that the 1-Betti curve is less regular than in the previous example.

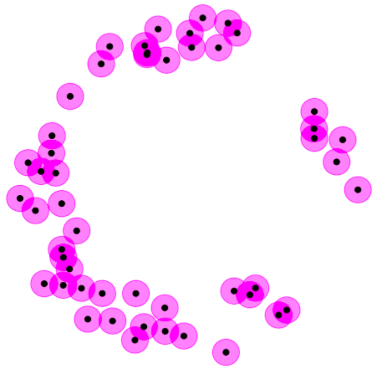
```
I = np.linspace(0,1,100)
Betti_curves = GetBettiCurvesFromPointCloud(X, I)

plt.figure()
plt.step(I, Betti_curves[0])
plt.ylim(0, max(Betti_curves[0])+1)
plt.title('0-Betti curve')
plt.show()

plt.figure()
plt.step(I, Betti_curves[1])
plt.ylim(0, max(Betti_curves[1])+1)
plt.title('1-Betti curve')
plt.show()
```





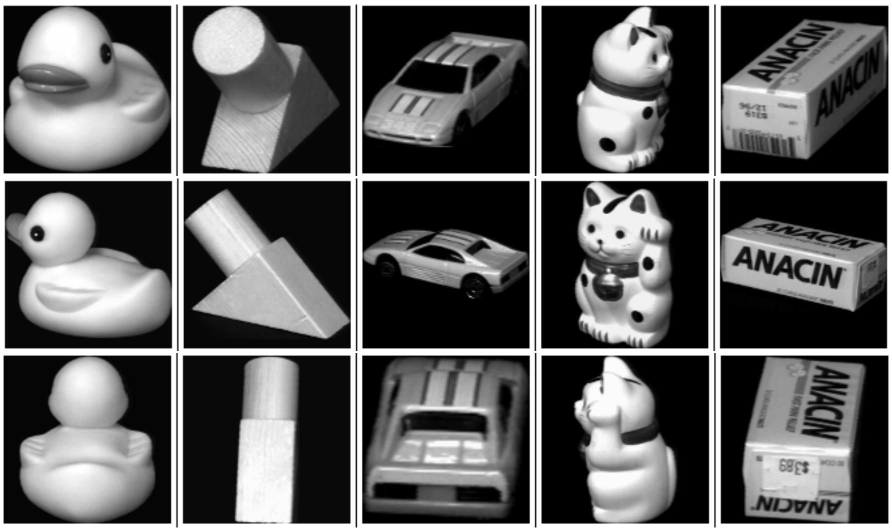
There are values of t for which $\beta_1(X^t)$ is nonzero, yet X^t is not a circle. This phenomenon is called *topological noise*.

```
t_noise = I[np.where(Betti_curves[1]>0)][0]
                  # first appartition of topological noise
PlotThickening(X,t=t_noise)
```

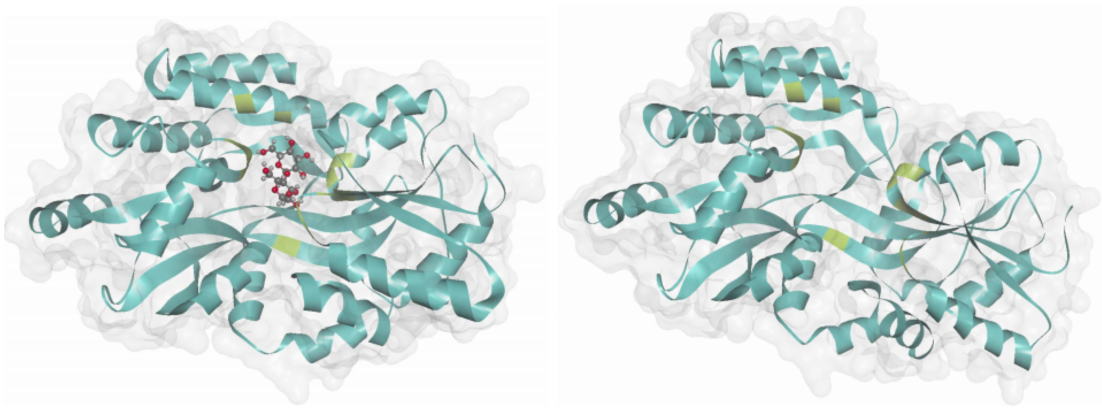


Exercise 44. In the notebook is given a subset of $\mathbb{R}^4$ of 200 elements. It has been sampled on a famous 2-dimensional object. Compute the Betti curves of its Rips complex on $[0, 1]$. Can you recognize which surface it is?

Exercise 45. In the notebook is given a collection of images from `https://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php`. It consists of 20 objects, for each of which 72 pictures have been taken. Each image has 128×128 pixels. Embed each collection of 72 images in $\mathbb{R}^{128\times 128}$, and compute the Betti curves of the corresponding Rips complex.



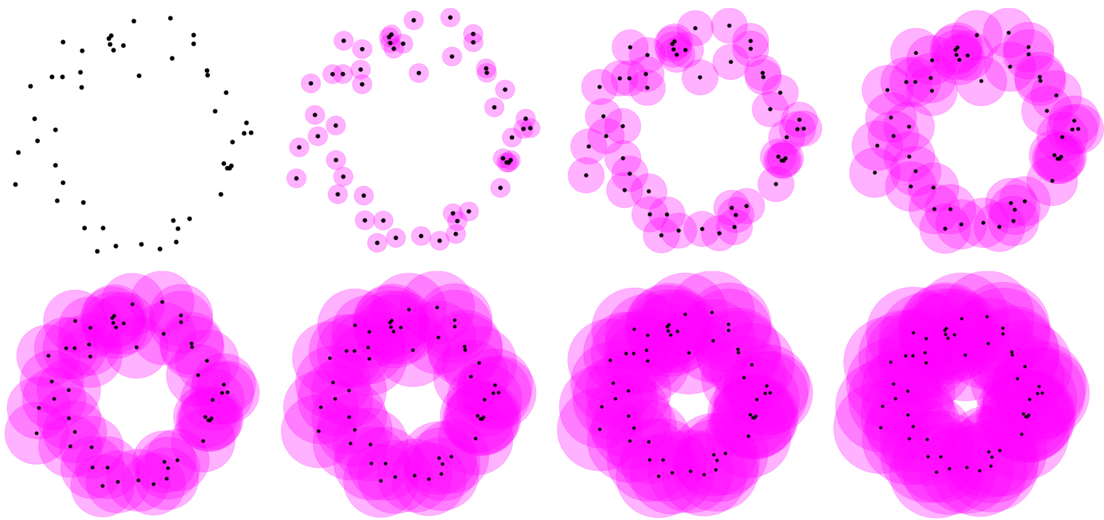
Exercise 46. We are given the data of [KNBNH16], where the authors study the *maltose-binding protein* (MBP). Such a protein can be grouped into 370 components, called *amino acid residues*. There is two types of MBPs: *open* and *closed*. The goal is to identify these types from topological properties of the proteins.



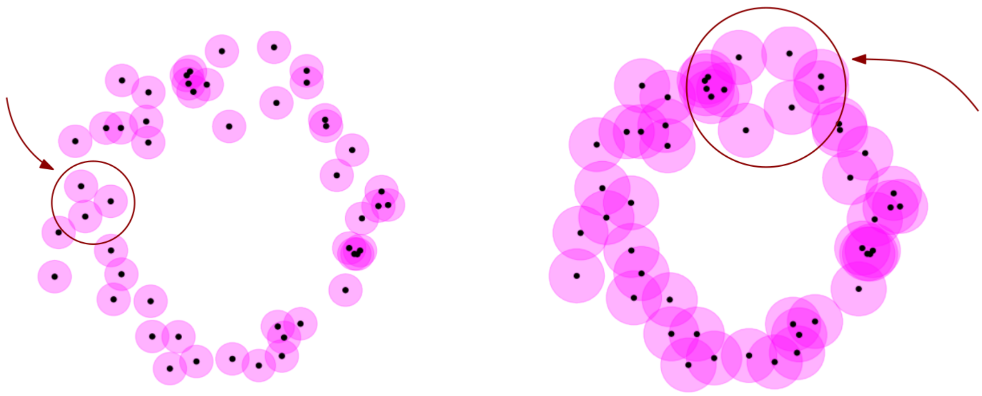
The daset consists in 14 correlation matrices, each matrix representing correlations between the 370 components of a protein. Transform the matrices of correlations into matrices of distances, via the formula $D_{i,j} = 1 - |C_{i,j}|$. Then, compute the 1-Betti curves of the Rips complex for each of these matrices of distances. Compare the Betti curves of the different proteins. Do you recognize two different types of proteins (*open* and *closed*)?

9 Decomposition of persistence modules

Let $X \subset \mathbb{R}^n$ be a finite point cloud, that we suppose close to an underlying object of interest, $\mathcal{M}$. In order to answer the problem of homological inference, we considered the thickenings X^t of X, $t \geq 0$. Some of them have the homotopy type of $\mathcal{M}$, and we can compute their homology via the complexes $\check{\mathrm{C}}\mathrm{ech}^t(X)$ and $\mathrm{Rips}^t(X)$.



However, in practice, the choice of t is critical. The Betti curves $t \mapsto \beta_i(X^t)$ are not continuous (they are integer-valued). We have to deal with *topological noise*, that is, small cycles that appear in the thickenings, but who diseappear fast.



In order to infer the homology of $\mathcal{M}$, we should should sort out the true underlying homology from the topological noise. Instead of selecting these thickenings, we will look at them all at once, and then to retrieve the homology groups of $\mathcal{M}$ from this collection. Persistent homology allows to make these ideas rigorous.

9.1 Functoriality of homology

Until here, I kept secret a fundamental property of homology: its functoriality. Homology does not only transforms topological spaces (into vector spaces), but also continuous maps (into linear maps). We will study this property from a simplicial viewpoint. We need the following definition, which should be seen as a simplicial version of the notion of continuous maps.

Definition 9.1. Let K and L be two simplicial complexes, and V_K, V_L their set of vertices. A *simplicial map* between K and L is a map $f: V_K \to V_L$ such that

$$\forall \sigma \in K, f(\sigma) \in L.$$

When there is no risk of confusion, we may denote a simplicial map $f: K \to L$ instead of $f: V_K \to V_L$.

Example 9.2. Let $K = \{[0], [1], [0, 1]\}$, $L = \{[0], [1], [2], [0, 1], [0, 2], [1, 2]\}$ and $f: \{0, 1\} \to \{0, 1, 2\}$ defined as $f(0) = 0$ and $f(1) = 1$. It is simplicial since $f([0, 1]) = [0, 1]$ is a simplex of L.

Example 9.3. Let $K = \{[0], [1], [2], [0, 1], [0, 2], [1, 2]\}$, $L = \{[0], [1], [2], [0, 1], [0, 2]\}$ and $f: \{0, 1, 2\} \to \{0, 1, 2\}$ defined as $f(0) = 0$, $f(1) = 1$ and $f(2) = 2$. It is not simplicial since $f([1, 2]) = [1, 2]$ is not a simplex of L.

Example 9.4. Let $X \subset \mathbb{R}^n$ and $s, t \geq 0$ such that $s \leq t$. Consider the Čech complexes $\check{\mathrm{C}}\mathrm{ech}^s(X)$ and $\check{\mathrm{C}}\mathrm{ech}^t(X)$. The inclusion map $i\colon \check{\mathrm{C}}\mathrm{ech}^s(X) \to \check{\mathrm{C}}\mathrm{ech}^t(X)$ is a simplicial map. Indeed, the sequence of simplicial complexes $\left(\check{\mathrm{C}}\mathrm{ech}^t(X)\right)_{t \geq 0}$ is non-decreasing.

Remark 9.5. Simplicial maps allows to encode continuous maps combinatorially, just as simplicial complexes do with topological spaces. How exaclty one goes from a simplicial map to a continuous map, and vice versa?

If one starts with a simplicial map $f\colon K \to L$, it always induces a continous map $|f|: |K| \to |L|$ between topological realizations of simplicial complexes, called the *topological realization* of f (to do so, one uses *barycentric coordinates*). Now, given a continuous map $g\colon |K| \to |L|$, it is not clear how to deduce a simplicial map $f\colon K \to L$. The problem of *simplicial approximation* consists in finding a simplicial map $f\colon K \to L$ with topological realization $|f|: |K| \to |L|$ homotopy equivalent to g. One solves this problem by applying *barycentric subdivisions*.

Let $f\colon K \to L$ be a simplicial map. Let $n \geq 0$, and consider the groups of chains of K and L:

$$C_n(K) = \left\{ \sum_{\sigma \in K_{(n)}} \epsilon_\sigma \cdot \sigma, \ \forall \sigma \in K_{(n)}, \ \epsilon_\sigma \in \mathbb{Z}/2\mathbb{Z} \right\}$$

$$C_n(L) = \left\{ \sum_{\sigma \in K_{(n)}} \epsilon_\sigma \cdot \sigma, \ \forall \sigma \in L_{(n)}, \ \epsilon_\sigma \in \mathbb{Z}/2\mathbb{Z} \right\}$$

We define a linear map $f_n\colon C_n(K) \to C_n(L)$ as follows (definition on the simplices, and extended by linearity):

$$f_n\colon \sigma \longmapsto \begin{array}{ll} f(\sigma) & \text{if} \ \ \dim(f(\sigma)) = n, \\ 0 & \text{else.} \end{array}$$

We obtain a diagram of chain complexes

$$\begin{tikzcd}
{} \arrow[r, dashed] & C_3(K) \arrow[r, "\partial_3"] \arrow[d, "f_3"] & C_2(K) \arrow[r, "\partial_2"] \arrow[d, "f_2"] & C_1(K) \arrow[r, "\partial_1"] \arrow[d, "f_1"] & C_0(K) \arrow[r, "\partial_0"] \arrow[d, "f_0"] & \{0\} \arrow[d, "f_{-1}"] \\
{} \arrow[r, dashed] & C_3(L) \arrow[r, "\partial_3"] & C_2(L) \arrow[r, "\partial_2"] & C_1(L) \arrow[r, "\partial_1"] & C_0(L) \arrow[r, "\partial_0"] & \{0\}
\end{tikzcd}$$

where we define $f_{-1} = 0$. The following lemma states that this diagram **commutes**, or is a **commutative diagram**: when composing maps in the diagram, all the directed paths with the same start and endpoints lead to the same result.

Lemma 9.6. *For every $n \geq 0$, we have $\partial_n \circ f_n = f_{n-1} \circ \partial_n$.*

Proof. Let $\sigma \in K_{(n)}$. We have the equalities

$$\partial_n \circ f_n(\sigma) = \sum_{\substack{\mu \subset f(\sigma) \\ |\mu| = |\sigma| - 1}} \mu$$
$$f_{n-1} \circ \partial_n(\sigma) = \sum_{\substack{\tau \subset \sigma \\ |\tau| = |\sigma| - 1}} f_n(\tau)$$

Let us show that these expressions coincide. We should distinguish three cases: $|f(\sigma)| = |\sigma|$ (i.e. f is injective on σ), $|f(\sigma)| < |\sigma| - 1$ or $|f(\sigma)| = |\sigma| - 1$. Proving the first case is straightforward. The second one too, since then $f_n(\sigma) = 0$, and each $f_n(\tau) = 0$.

Only the last case requires some work. Suppose that $|f(\sigma)| = |\sigma| - 1$. We have $f_n(\sigma) = 0$. Denote a, b the vertices of σ such that $f(a) = f(b)$. In the second sum, all the simplices $\tau \subset \sigma$ satisfy $|f(\tau)| = |\tau| - 1$, hence $f_n(\tau) = 0$, except for the simplex τ_a that does not contain a, and the smplex τ_b that does not contain b. We deduce

$$f_{n-1} \circ \partial_n(\sigma) = f_n(\tau_a) + f_n(\tau_b) = 0$$

since $f_n(\tau_a) = f_n(\tau_b)$. $\square$

The following proposition describes how f_n acts on the cycles and the boundaries.

Proposition 9.7. *For every $c \in Z_n(K)$, we have $f_n(c) \in Z_n(L)$. For every $c \in B_n(K)$, we also have $f_n(c) \in B_n(L)$.*

In other words, the image of a cycle is a cycle, and the image of a boundary is a boundary.

Proof. First, let $c \in Z_n(K)$. By using Lemma 9.6, we get

$$\partial_n \circ f_n(c) = f_{n-1} \circ \partial_n(c) = f_{n-1}(0) = 0,$$

hence $f_n(c) \in Z_n(L)$. Secondly, let $c \in B_n(K)$, and write $c = \partial_{n+1}(c')$ with $c' \in C_{n+1}(K)$. Still using Lemma 9.6, we get

$$f_n(c) = f_n \circ \partial_{n+1}(c') = \partial_{n+1} \circ f_{n+1}(c'),$$

hence $f_n(c) \in B_n(L)$. $\square$

Remind that $B_n(K) \subset Z_n(K)$ and $B_n(L) \subset Z_n(L)$. The previous proposition, together with the following exercise, shows that the map f_n induces a linear map between quotient vector spaces:

$$(f_n)_*\colon Z_n(K)/B_n(K) \longrightarrow Z_n(L)/B_n(L).$$

By definition of the homology groups, we have defined a map

$$(f_n)_*\colon H_n(K) \longrightarrow H_n(L).$$

It is called the *induced map in homology*. When there will be no risk of confusion, we may denote f_* instead of $(f_n)_*$.

$$\begin{tikzcd}
\cdots & H_3(K) \arrow[d, "(f_3)*"] & H_2(K) \arrow[d, "(f_2)*"] & H_1(K) \arrow[d, "(f_1)*"] & H_0(K) \arrow[d, "(f_0)*"] \\
\cdots & H_3(L) & H_2(L) & H_1(L) & H_0(L)
\end{tikzcd}$$

Explicitely, the map $(f_n)*$ can be described as follows (the following formula is to be read modulo boundaries):

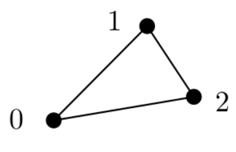
$$c = \textstyle\sum_{\sigma \in K_{(n)}} \epsilon_\sigma \cdot \sigma \longmapsto \sum_{\sigma \in K_{(n)}} \epsilon_\sigma \cdot f_n(\sigma).$$

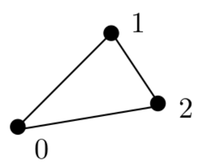
Exercise 47. Consider a linear map $f\colon V \to W$ between vector spaces. Suppose that there exists linear subspaces $A \subset V$ and $B \subset W$ such that $f(V) \subset W$. Show that one can define a map $f_*\colon V/A \to W/B$ as follows: to any equivalence class $v + A$ of V/A, let $f_*(v + A) = f(v) + B$.
Hint: You have to show that the map is well-defined, and linear. Well-defined means that for any equivalence class $v + A$, for any other representative $w \in v + A$, the equivalence class $f(w) + B$ is equal to $f(v) + B$.

Example 9.8. Consider the simplicial complexes $K = L = \{[0], [1], [2], [0, 1], [0, 2], [1, 2]\}$. The inclusion $i\colon K \to L$ induces the identity between 0^{th} homology groups and between 1^{st} homology groups.

$$\begin{aligned}(i_0)_*\colon H_0(K) \simeq \mathbb{Z}/2\mathbb{Z} &\longrightarrow H_0(L) \simeq \mathbb{Z}/2\mathbb{Z}\\ 1 &\longmapsto 1\end{aligned}$$

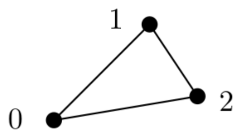
$$\begin{aligned}(i_1)_*\colon H_1(K) \simeq \mathbb{Z}/2\mathbb{Z} &\longrightarrow H_1(L) \simeq \mathbb{Z}/2\mathbb{Z}\\ 1 &\longmapsto 1\end{aligned}$$

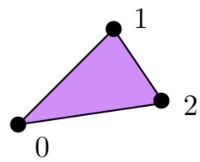




Example 9.9. Consider the simplicial complexes $K = \{[0], [1], [2], [0,1], [0,2], [1,2]\}$ and $L = \{[0], [1], [2], [0,1], [0,2], [1,2], [0,1,2]\}$. The inclusion $i\colon K \to L$ induces the identity between 0^{th} homology groups, and the zero map between 1^{st} homology groups:

$$\begin{aligned} (i_1)_*\colon H_1(K) \simeq \mathbb{Z}/2\mathbb{Z} &\longrightarrow H_1(L) \simeq \{0\} \\ 1 &\longmapsto 0 \end{aligned}$$

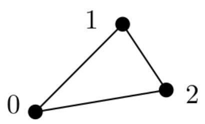


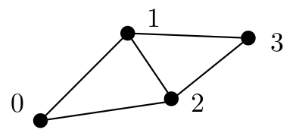


Example 9.10. Consider the simplicial complexes $K = \{[0], [1], [2], [0,1], [0,2], [1,2]\}$ and $L = \{[0], [1], [2], [3], [0,1], [0,2], [1,2], [1,3], [2,3]\}$. The homology group $H_1(L)$ is isomorphic to the vector space $(\mathbb{Z}/2\mathbb{Z})^2$ by identifying $[0,1] + [0,2] + [1,2] \mapsto (1,0)$ and $[1,2] + [2,3] + [1,3] \mapsto (0,1)$. The inclusion $i\colon K \to L$ induces the following map between 1^{st} homology groups:

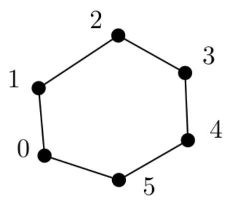
$$\begin{aligned} (i_1)_*\colon H_1(K) \simeq \mathbb{Z}/2\mathbb{Z} &\longrightarrow H_1(L) \simeq (\mathbb{Z}/2\mathbb{Z})^2 \\ 1 &\longmapsto (1,0) \end{aligned}$$

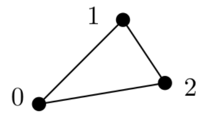
It can be represented as the matrix $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$.





Exercise 48. Let $K = \{[0], [1], [2], [3], [4], [5], [0,1], [1,2], [2,3], [3,4], [4,5], [5,0]\}$ and $L = \{[0], [1], [2], [0,1], [1,2], [2,0]\}$. Consider the simplical map $f\colon i \mapsto i$ modulo 3. Show that the induced map $(f_1)_*$ is zero.





Hint: Remember that we work over $\mathbb{Z}/2Z$.

Remark 9.11. Given a continuous map $f\colon X \to Y$ between topological spaces, one also defines induced map $(f_n)_*\colon H_n(X) \to H_n(Y)$ between homology groups. In order to do so, one uses the notion of *singular homology*, that we do not study in this summer course.

Finally, we state the most important result of this subection: homology is functorial.

Proposition 9.12. *Let K, L, M be three simplicial complexes, and consider two simplicial maps $f\colon K \to L$ and $g\colon L \to M$. For any $n \geq 0$, the induced map $((g \circ f)_n)_*\colon H_n(K) \to H_n(M)$ and $(g_n)_* \circ (f_n)_*\colon H_n(K) \to H_n(M)$ are equal.*

$$\begin{tikzcd} K \arrow[r, "f"'] \arrow[rr, bend left, "g\circ f"] & L \arrow[r, "g"'] & M, \end{tikzcd} \qquad \begin{tikzcd} H_n(K) \arrow[r, "f_*"'] \arrow[rr, bend left, "(g\circ f)_*"] & H_n(L) \arrow[r, "g_*"'] & H_n(M). \end{tikzcd}$$

Proof. Let $\sigma \in K_{(n)}$. The image $(g \circ f)_n(\sigma)$ is $(g \circ f)(\sigma)$ if $g \circ f$ is injective on σ, and 0 else. Let us distinguish the two cases.

If $g \circ f$ is injective on σ, then f is injective on σ *and* g is injective on $f(\sigma)$, hence $g_n \circ f_n(\sigma) = g \circ f(\sigma)$, and we deduce the result.

If $g \circ f$ is not injective on σ, then f is not injective on σ *or* g is not injective on $f(\sigma)$, hence $g_n \circ f_n(\sigma) = 0$, and we deduce the result. $\square$

Exercise 49. Fill the empty spaces (▩) in the following proof of *Brouwer's fixed point theorem.*

Let $f\colon \overline{\mathcal{B}}(0,1) \to \overline{\mathcal{B}}(0,1)$ be a continuous map, where $\overline{\mathcal{B}}(0,1)$ denotes the closed unit ball of $\mathbb{R}^n$. Let us show that f admits a fixed point (i.e., an element $x \in \overline{\mathcal{B}}(0,1)$ such that $f(x) = x$).

By contradiction, suppose that it is not the case. We can build an application $F\colon \overline{\mathcal{B}}(0,1) \to \mathbb{S}(0,1)$, where $\mathbb{S}(0,1) \subset \mathbb{R}^n$ is the unit sphere, such that F restricted to $\mathbb{S}(0,1)$ is the identity. To do so, define $F(x)$ as the first intersection point between the half-line $[x, f(x))$ and $\mathbb{S}(0,1)$.

Denote the inclusion $i\colon \mathbb{S}(0,1) \to \overline{\mathcal{B}}(0,1)$. We have that $F \circ i\colon \mathbb{S}(0,1) \to \mathbb{S}(0,1)$ is the identity. By functoriality of homology, we obtain, for all $i \geq 0$, the commutative diagrams

$$\begin{tikzcd} \mathbb{S}(0,1) \arrow[r, "i"'] \arrow[rr, bend left, "\mathrm{id}"] & \overline{\mathcal{B}}(0,1) \arrow[r, "F"'] & \mathbb{S}(0,1), \end{tikzcd} \quad \begin{tikzcd} H_i(\mathbb{S}(0,1)) \arrow[r, "i_*"'] \arrow[rr, bend left, "▩"] & H_i(\overline{\mathcal{B}}(0,1)) \arrow[r, "F_*"'] & H_i(\mathbb{S}(0,1)). \end{tikzcd}$$

But choosing $i = n-1$, we have $H_i(\mathbb{S}(0,1)) \simeq$ ▩, $H_i(\overline{\mathcal{B}}(0,1)) \simeq$ ▩, and the following diagram cannot commute:

$$\begin{tikzcd} ▩ \arrow[r] \arrow[rr, bend left, "▩"] & 0 \arrow[r] & ▩. \end{tikzcd}$$

9.2 Persistence modules

Let $X \subset \mathbb{R}^n$. The collection of its thickenings is an non-decreasing sequence of subsets

$$\ldots \subset X^{t_1} \subset X^{t_2} \subset X^{t_3} \subset \ldots$$

By considering the corresponding Čech complexes, we obtain an non-decreasing sequence of simplicial complexes

$$\ldots \subset \check{\mathrm{C}}\mathrm{ech}^{t_1}(X) \subset \check{\mathrm{C}}\mathrm{ech}^{t_2}(X) \subset \check{\mathrm{C}}\mathrm{ech}^{t_3}(X) \subset \ldots$$

Let us denote i_s^t the inclusion map corresponding to $\check{\mathrm{C}}\mathrm{ech}^s(X) \subset \check{\mathrm{C}}\mathrm{ech}^t(X)$. We can write

$$
\dashrightarrow \check{\mathrm{C}}\mathrm{ech}^{t_1}(X) \xrightarrow{i_{t_1}^{t_2}} \check{\mathrm{C}}\mathrm{ech}^{t_2}(X) \xrightarrow{i_{t_2}^{t_3}} \check{\mathrm{C}}\mathrm{ech}^{t_3}(X) \dashrightarrow
$$

Applying the i^{th} homology functor yields a diagram of vector spaces

$$
\dashrightarrow H_i(\check{\mathrm{C}}\mathrm{ech}^{t_1}(X)) \xrightarrow{\left(i_{t_1}^{t_2}\right)_*} H_i(\check{\mathrm{C}}\mathrm{ech}^{t_2}(X)) \xrightarrow{\left(i_{t_2}^{t_3}\right)_*} H_i(\check{\mathrm{C}}\mathrm{ech}^{t_3}(X)) \dashrightarrow
$$

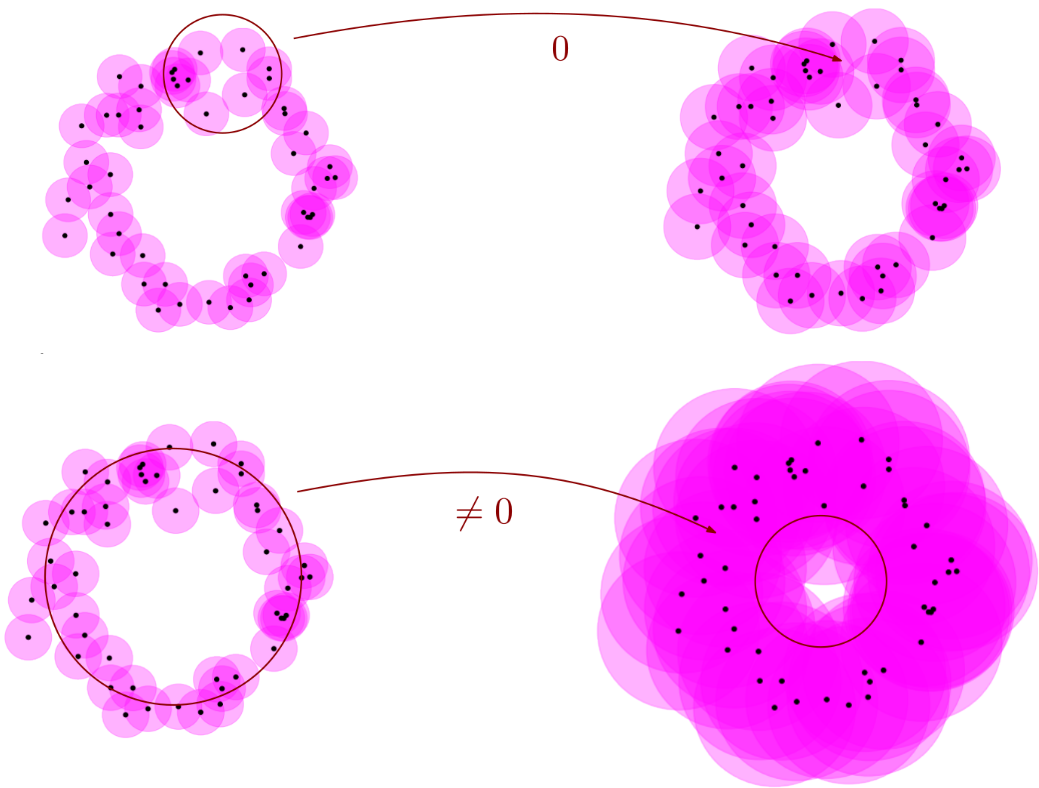
where the maps $(i_s^t)_*$ are those induced in homology by the inclusions i_s^t. In this sequence, we are able to read how long a cycle lives, or *persists*. Let $i \geq 0$, $t_0 \geq 0$ and consider a cycle $c \in H_i(\check{\mathrm{C}}\mathrm{ech}^{t_0}(X))$. Its *death time* is

$$
\sup \left\{t \geq t_0, \left(i_{t_0}^t\right)(c) \neq 0\right\}
$$

and its *birth time* is

$$
\inf \left\{t \geq t_0, \left(i_t^{t_0}\right)^{-1}(\{c\}) \neq \emptyset\right\}.
$$

We measure the *persistence* of c as the difference between its death time and its birth time. As a rule of thumb, we expect that cycles with large persistence correspond to important topological features of the dataset, and that cycles with short persistence corresponds to topological noise.



Definition 9.13. A *persistence module* $\mathbb{V}$ *over* $\mathbb{R}^+$ *with coefficients in* $\mathbb{Z}/2\mathbb{Z}$ is a pair $(\mathbb{V}, \mathbb{v})$ where $\mathbb{V} = (V^t)_{t\in\mathbb{R}^+}$ is a family of $\mathbb{Z}/2\mathbb{Z}$-vector spaces, and $\mathbb{v} = (v_s^t \colon V^s \to V^t)_{s\leq t\in\mathbb{R}^+}$ a family of linear maps such that:

- for every $t \in \mathbb{R}^+$, $v_t^t \colon V^t \to V^t$ is the identity map,
- for every $r, s, t \in \mathbb{R}^+$ such that $r \leq s \leq t$, we have $v_s^t \circ v_r^s = v_r^t$.

When the context is clear, we may denote $\mathbb{V}$ instead of $(\mathbb{V}, \mathbb{v})$.

Remark 9.14. In the language of categories, a persistence module is a functor

$$\mathbb{V} \colon (T, \leq) \to k\text{-Mod},$$

where $(T, \leq)$ is the category associated to an ordered set T, and k-Mod is the category of k-vector spaces. More precisely, the category $(T, \leq)$ has objects being the elements of T, and has an arrow $x \to y$ for every $x, y \in T$ such that $x \leq y$. This point of view is useful to generalize the notion of persistence modules we present in this subsection. For instance,

- [CZ09] defines a *multi-parameter persistence module* as a functor $(\mathbb{R}^n, \leq) \to k$-Mod, where $\leq$ denotes the usual partial order on $\mathbb{R}^n$,
- [CDS10] defines a *zigzag module* as a functor $Q \to k$-Mod, where Q is a quiver of type A_n,
- [BGO19] defines a *persistence comodule* as a contravariant functor $(\mathbb{R}, \leq) \to k$-Mod.
- [BCB20] defines a *(generalized) persistence module* as a functor $\mathcal{C} \to k$-Mod, where $\mathcal{C}$ is any small category.

In practice, one builds persistence modules from *filtrations.* A family of subsets $\mathbb{X} = (X^t)_{t \in \mathbb{R}^+}$ of E is a *filtration* if it is non-decreasing for the inclusion, i.e. for any $s, t \in \mathbb{R}^+$, if $s \leq t$ then $X^s \subseteq X^t$. In this course, we will consider filtrations of simplicial complexes, that is, non-decreasing families of simplicial complexes $\mathbb{S} = (S^t)_{t \in \mathbb{R}^+}$.

By applying the i^{th} homology functor to a filtration, we obtain a persistence module $\mathbb{V}[\mathbb{S}] = \big(H_i(S^t)\big)_{t \in \mathbb{R}^+}$, with maps $\big((i_s^t)_* \colon H_i(S^s) \to H_i(S^t)\big)_{s \leq t}$ induced by the inclusions. This is pictured by the two following diagrams.

$$\cdots\cdots \dashrightarrow S^{t_1} \xhookrightarrow{\;i_{t_1}^{t_2}\;} S^{t_2} \xhookrightarrow{\;i_{t_2}^{t_3}\;} S^{t_3} \xhookrightarrow{\;i_{t_3}^{t_4}\;} S^{t_4} \hookrightarrow \cdots\cdots$$

$$\cdots\cdots \dashrightarrow H_i(S^{t_1}) \xrightarrow{\left(i_{t_1}^{t_2}\right)_*} H_i(S^{t_2}) \xrightarrow{\left(i_{t_2}^{t_3}\right)_*} H_i(S^{t_3}) \xrightarrow{\left(i_{t_3}^{t_4}\right)_*} H_i(S^{t_4}) \cdots\cdots$$

The persistence module $\mathbb{V}[\mathbb{S}]$ is called the persistence module *associated to* the filtration $\mathbb{S}$.

Moralidade

Em homologia persistente, em vez de escolher um valor de raio r, tomamos todos eles ao mesmo tempo.

9.3 Decomposition

We introduce a few notions, in order to decompose persistence modules into smaller pieces.

Definition 9.15. An *isomorphism* between two persistence modules $\mathbb{V}$ and $\mathbb{W}$ is a family of isomorphisms of vector spaces $\phi = (\phi_t \colon V^t \to W^t)_{t\in\mathbb{R}^+}$ such that the following diagram commutes for every $s \leq t \in \mathbb{R}^+$:

$$\begin{tikzcd} V^s \arrow[r, "v_s^t"] \arrow[d, "\phi_s"] & V^t \arrow[d, "\phi_t"] \\ W^s \arrow[r, "w_s^t"] & W^t \end{tikzcd}$$

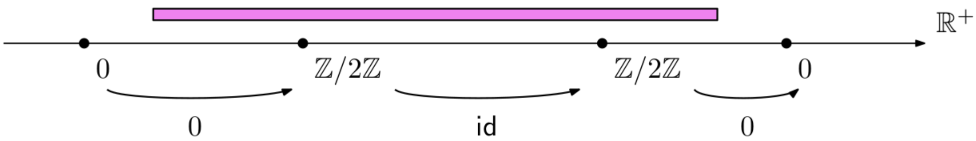
Decomposability. Let $(\mathbb{V}, \mathbb{v})$ and $(\mathbb{W}, \mathbb{w})$ be two persistence modules. Their *sum* is the persistence module $\mathbb{V} \oplus \mathbb{W}$ defined with the vector spaces $(V \oplus W)^t = V^t \oplus W^t$ and the linear maps

$$(v \oplus w)_s^t \colon (x, y) \in (V \oplus W)^s \longmapsto (v_s^t(x), w_s^t(y)) \in (V \oplus W)^t.$$

A persistence module $\mathbb{U}$ is *indecomposable* if for every pair of persistence modules $\mathbb{V}$ and $\mathbb{W}$ such that $\mathbb{U}$ is isomorphic to the sum $\mathbb{V} \oplus \mathbb{W}$, then one of the summands has to be a trivial persistence module, that is, equal to zero for every $t \in \mathbb{R}^+$. Otherwise, $\mathbb{U}$ is said *decomposable.*

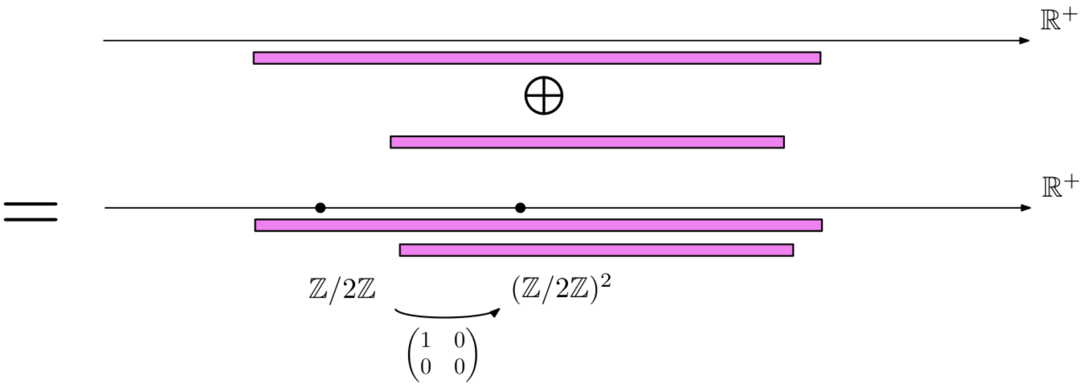
Interval modules. Let $I \subset \mathbb{R}^+$ be an interval, that is, a non-empty convex set. Intervals have the form $[a, b]$, $(a, b]$, $[a, b)$ or (a, b), with $a, b \in \mathbb{R}^+$ such that $a \leq b$, and potentially $a = -\infty$ or $b = +\infty$. The *interval module* associated to I is the persistence module $\mathbb{B}[I]$ with vector spaces $\mathbb{B}^t[I]$ and linear maps $v_s^t \colon \mathbb{B}^s[I] \to \mathbb{B}^t[I]$ defined as

$$\mathbb{B}^t[I] = \begin{cases} \mathbb{Z}/2\mathbb{Z} & \text{if } t \in I, \\ 0 & \text{otherwise,} \end{cases} \qquad \text{and} \qquad v_s^t = \begin{cases} \text{id} & \text{if } s, t \in I, \\ 0 & \text{otherwise.} \end{cases}$$

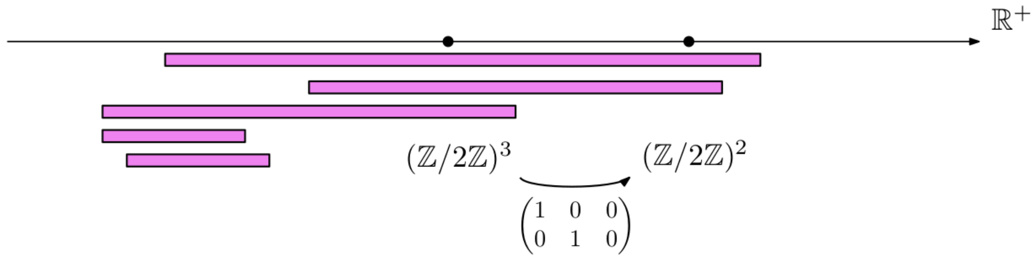


Exercise 50. Show that the interval modules are indecomposable.

In what follows, we will sum interval modules.



We may end up with a lot of them.



Persistence barcodes and persistence diagrams. A persistence module $\mathbb{V}$ *decomposes into interval module* if there exists a set $\{\mathbb{B}_i, i \in \mathcal{I}\}$ of interval modules such that $\mathbb{V}$ is isomorphic to the sum $\bigoplus_{i\in\mathcal{I}} \mathbb{B}_i$. In other words, there exists a multiset $\mathcal{I}$ of intervals of T such that

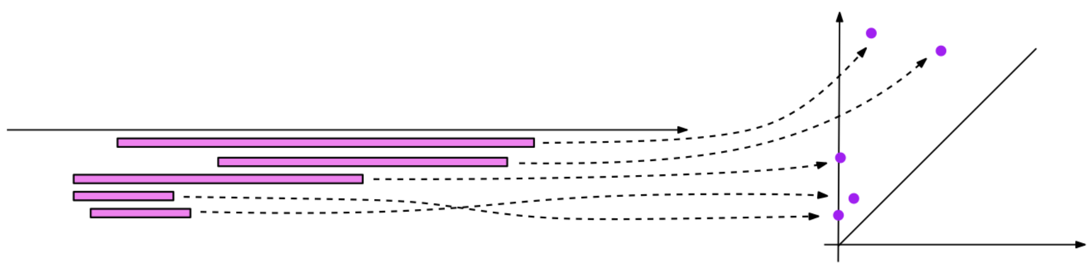
$$\mathbb{V} \simeq \bigoplus_{I\in\mathcal{I}} \mathbb{B}[I].$$

Multiset means that $\mathcal{I}$ may contain several copies of the same interval I. Such a module is said *decomposable into interval modules*, or simply *decomposable* when the context is clear.

Theorem 9.16 (Consequence of [Azu50, Theorem 1])**.** *If a persistence module decomposes into interval modules, then the multiset $\mathcal{I}$ of intervals is unique.*

In this case, $\mathcal{I}$ is called the *persistence barcode* of $\mathbb{V}$, or simply *barcode*. It is written Barcode$(\mathbb{V})$.

Let $\mathbb{V}$ be a decomposable persistence module and Barcode$(\mathbb{V})$ its barcode. For every $[a,b]$, $(a,b]$, $[a,b)$ or (a,b) in Barcode$(\mathbb{V})$, with potentially $a = -\infty$ or $b = +\infty$, consider the point (a,b) of $\mathbb{R}^2$. The collection of all such points is a multiset, that we call the *persistence diagram* of $\mathbb{V}$. It is denoted Diagram$(\mathbb{V})$.



Decomposition of pointwise finite-dimensional modules. A persistence module $\mathbb{V}$ is said *pointwise finite dimensional* if every vector space V^t has finite dimension. We have:

Theorem 9.17 ([CB15, Theorem 2.1])**.** *Every pointwise finite-dimensional persistence module decomposes into interval modules.*

Proof. We prove the result in a simpler case: when the persistence module is finite-dimensional *and* has finitely many terms. The proof comes from [AZ05].

We can write our persistence module as

$$V^1 \xrightarrow{v_1^2} V^2 \xrightarrow{v_2^3} V^3 \xrightarrow{v_3^4} V^4 \dashrightarrow \dots \dashrightarrow V^n$$

Consider the vector space $\mathcal{V} = \bigotimes_{1\leq i\leq n} V^i = V^1 \times \cdots \times V^n$. Let $\mathbb{Z}/2\mathbb{Z}[x]$ denote the space of polynomials with coefficients in $\mathbb{Z}/2\mathbb{Z}$. We give $\mathcal{V}$ an action of $\mathbb{Z}/2\mathbb{Z}[x]$ via

$$x \cdot (a^1, a^2, ..., a^n) = (0, v_1^2(a^1), v_2^3(a^2), ..., v_{n-1}^n(a^{n-1})).$$

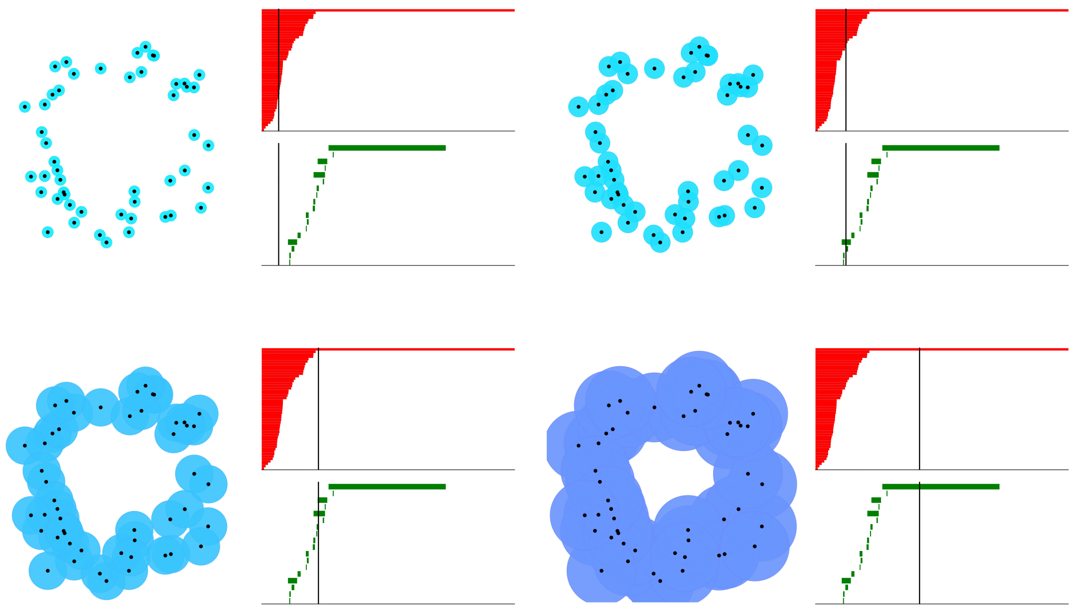
Hence $\mathcal{V}$ can be seen as a finitely generated module over the principal ideal domain $\mathbb{Z}/2\mathbb{Z}[x]$. By classification, $\mathcal{V}$ is isomorphic to a sum

$$\mathcal{V} \simeq \bigoplus_{i\in I} \mathbb{Z}/2\mathbb{Z}[x]/x^i \cdot \mathbb{Z}/2\mathbb{Z}[x].$$

Now, we can identify the components $\mathbb{Z}/2\mathbb{Z}[x]/x^i \cdot \mathbb{Z}/2\mathbb{Z}[x]$ with bars of the barcode of length i. Actually, to obtain precisely the birth time and death time of the bars, we should have applied the classification of finitely generated *graded* modules. $\square$

An example of a persistence module that does not decompose into interval modules can be found in [CdSGO16, Theorem 1.4 (3)]. This theorem does not hold for generalized definitions of persistence modules, where the notion of interval modules may not even be well-defined. Although, some weaker results exist [CB15, Theorem 1.1].

Let $X \subset \mathbb{R}^n$ be a finite point cloud, let $i \geq 0$, and consider the Čech persistence module $\left(H_i(\check{\mathrm{C}}\mathrm{ech}^t(X))\right)_{t\geq 0}$, or the Rips persistence module $\left(H_i(\mathrm{Rips}^t(X))\right)_{t\geq 0}$. On the barcode, we do not only read the homology of the thickenings at each step, we also see how they persist.



Exercise 51. Let $\mathcal{M}$ be the unit circle of $\mathbb{R}^2$, and $X \subset \mathbb{R}^2$ a finite subset. Denote the Hausdorff distance $\epsilon = d_{\mathrm{H}}(X, \mathcal{M})$. Suppose that ϵ is small enough. Let $\mathbb{U}$ denote the persistence module of the 1^{st} homology of the Čech filtration of X. Using Theorem 7.3, shows that there exists an interval I on which $\mathbb{U}$ is constant and equal to $\mathbb{Z}/2\mathbb{Z}$. Deduce the existence of a bar in the barcode, and give a lower bound on its persistence. Do the same with Theorem 7.4.

9.4 Persistent homology algorithm

It turns out that computing barcodes of a simplicial filtrations in already contained in the incremental algorithm for homology, as in Section 6.

The Čech or the Rips filtration define an increasing sequence of simplices

$$\ldots \subset \text{Čech}^{t_1}(X) \subset \text{Čech}^{t_2}(X) \subset \text{Čech}^{t_3}(X) \subset \ldots$$

Turn it consistently into an ordering of the simplices, as in Subsection 6.1, by inserting the simplices by order of apparition in the filtration.

$$\sigma^1 < \sigma^2 < \ldots < \sigma^n$$

In this subsection, we will denote $t(\sigma)$ the time of apparition of the simplex σ in the filtration. The total order on the simplices must satisfy

$$t(\sigma^i) < t(\sigma^j) \text{ for all } i < j.$$

In practice several simplices may appear at the same time. If this occurs, choose an order on the simplices such that $t(\sigma^i) = t(\sigma^j)$. Then, build the boundary matrix, and compute a Gauss reduction Δ.

$$\begin{array}{c} \\ \sigma^1 \\ \sigma^2 \\ \sigma^3 \\ \sigma^4 \\ \sigma^5 \\ \sigma^6 \\ \sigma^7 \\ \sigma^8 \\ \sigma^9 \\ \sigma^{10} \end{array}
\begin{array}{c}
\begin{array}{cccccccccc} \sigma^1 & \sigma^2 & \sigma^3 & \sigma^4 & \sigma^5 & \sigma^6 & \sigma^7 & \sigma^8 & \sigma^9 & \sigma^{10} \end{array} \\
\begin{pmatrix}
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
\end{array}
\qquad
\begin{array}{c} \\ \sigma^1 \\ \sigma^2 \\ \sigma^3 \\ \sigma^4 \\ \sigma^5 \\ \sigma^6 \\ \sigma^7 \\ \sigma^8 \\ \sigma^9 \\ \sigma^{10} \end{array}
\begin{array}{c}
\begin{array}{cccccccccc} \sigma^1 & \sigma^2 & \sigma^3 & \sigma^4 & \sigma^5 & \sigma^6 & \sigma^7 & \sigma^5+\sigma^6+\sigma^7+\sigma^8 & \sigma^9+\sigma^6+\sigma^7 & \sigma^{10} \end{array} \\
\begin{pmatrix}
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
\end{array}$$

Remember that we defined a map δ as follows: for any $j \in [\![1, n]\!]$,

$$\delta(j) = \max\{i \in [\![1, n]\!], \Delta_{i,j} \neq 0\},$$

and $\Delta_{i,j} = 0$ for all j, then $\delta(j)$ is *undefined.* Now, for all j such that $\delta(j)$ is defined, consider the pair of simplices

$$\left(\sigma^{\delta(j)}, \sigma^j\right).$$

Also, for all i such that $\forall j, \delta(j) \neq i$, we set

$$(\sigma^i, +\infty)\,.$$

The pairs of simplices (σ, τ) are called *persistence pairs.*

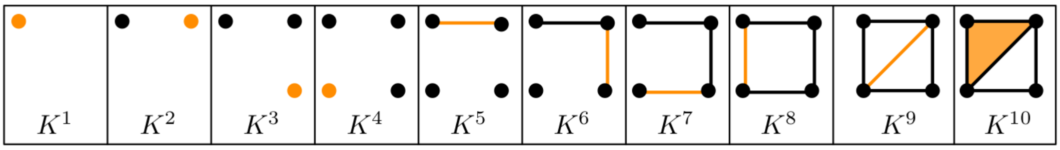
Proposition 9.18. *The n^{th} barcode of the filtration consists in the intervals*

$$\left\{ [t(\sigma), t(\tau)] \;\; \textit{for all persistence pairs} \;\; (\sigma, \tau) \;\; \textit{such that} \;\; \dim(\sigma) = n \;\; \textit{and} \;\; t(\sigma) \neq t(\tau) \right\}.$$

Proof. We shall show that the algorithm allows to define, for all $i, j \geq 0$, a basis $\mathcal{B}_i^j$ of $H_i(K^j)$, such that one passes from $\mathcal{B}_i^j$ to $\mathcal{B}_i^{j+1}$ by adding or removing a chain. As a consequence, we obtain an isomorphism between the persistence module and a sum of interval modules given by $\mathcal{I}$.

We build the basis as follows: for every $j \geq 0$, consider the simplex σ^j and its dimension $i = \dim(\sigma^j)$. If σ^j is positive, then we add the corresponding cycle to the basis $\mathcal{B}_i^{j-1}$. If it is negative, then there exists a simplex σ^k, with $k < j$, such that $\delta(k) = j$. We remove the cycle corresponding to σ^k to the basis $\mathcal{B}_{i-1}^{j-1}$. $\square$

Exercise 52. Compute the barcode of the filtration of Subsection 6.1:

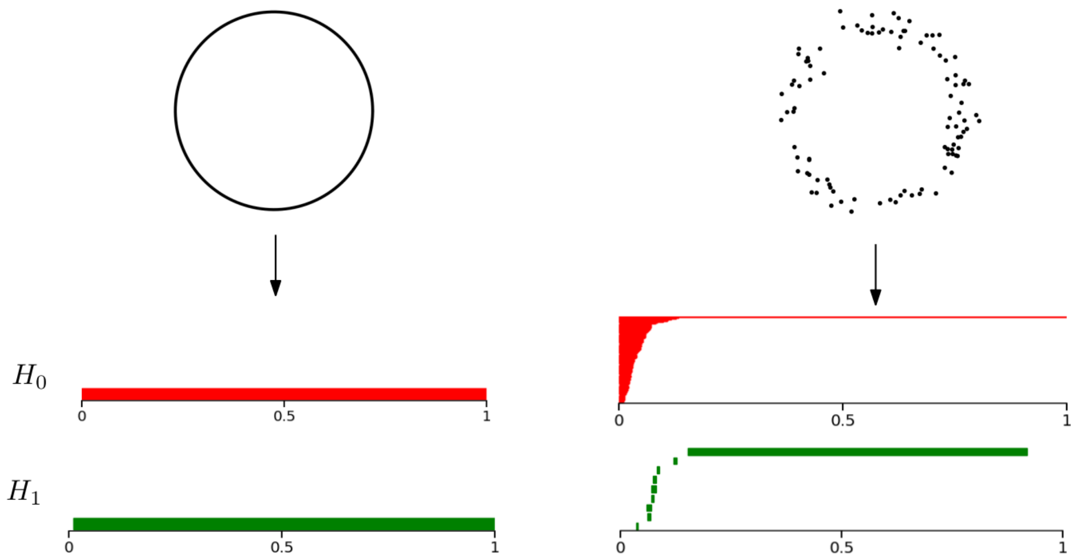


with the following filtration values: $t(\sigma) = 0$ for the vertices, $t(\sigma) = \frac{1}{2}$ for the edges of the square, and $t(\sigma) = \frac{\sqrt{2}}{2}$ for the diagonal edge and the triangle.

10 Stability of persistence modules

For the last theoretical lesson of this summer course, we will study the most important result of persistent homology theory: its stability. Such a result is crucial when applying persistent homology in practice.

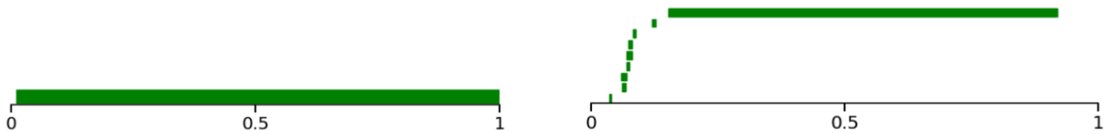
As an illustration, below are displayed the persistence barcodes of the Čech filtrations of the unit circle $\mathbb{S}_1 \subset \mathbb{R}^2$ and a finite sample X of it. We observe that the diagrams look close to each other.



10.1 Bottleneck distance and interleaving distance

Bottleneck distance. It is a distance between barcodes. In general, it is only an *extended* distance, meaning that it can take the value $+\infty$.

Consider two barcodes P and Q, that is, multisets of intervals $\{(a_i, b_i), i \in \mathcal{I}\}$ of $\overline{\mathbb{R}^+}^2$ such that $a_i \leq b_i$ for all $i \in \mathcal{I}$. Here, $\overline{\mathbb{R}^+}$ represent the extended real line $\mathbb{R}^+ \cup \{+\infty\}$.



A *partial matching* between the barcodes is a subset $M \subset P \times Q$ such that

- for every $p \in P$, there exists at most one $q \in Q$ such that $(p, q) \in M$,
- for every $q \in Q$, there exists at most one $p \in P$ such that $(p, q) \in M$.

The points $p \in P$ (resp. $q \in Q$) such that there exists $q \in Q$ (resp. $p \in P$) with $(p, q) \in M$ are said *matched* by M. If a point $p \in P$ (resp. $q \in Q$) is not matched by M, we consider that it is matched with the singleton $\overline{p} = \left[\frac{p_1+p_2}{2}, \frac{p_1+p_2}{2}\right]$ (resp. $\overline{q} = \left[\frac{q_1+q_2}{2}, \frac{q_1+q_2}{2}\right]$). The *cost* of a matched pair (p, q) (resp. $(p, \overline{p})$, resp. $(\overline{q}, q)$) is the sup norm $\|p - q\|_\infty = \sup\{|p_1 - q_1|, |p_2 - q_2|\}$ (resp. $\|p - \overline{p}\|_\infty$, resp. $\|\overline{q} - q\|_\infty$). The *cost* of the partial matching M, denoted $\operatorname{cost}(M)$, is the supremum of all such costs.

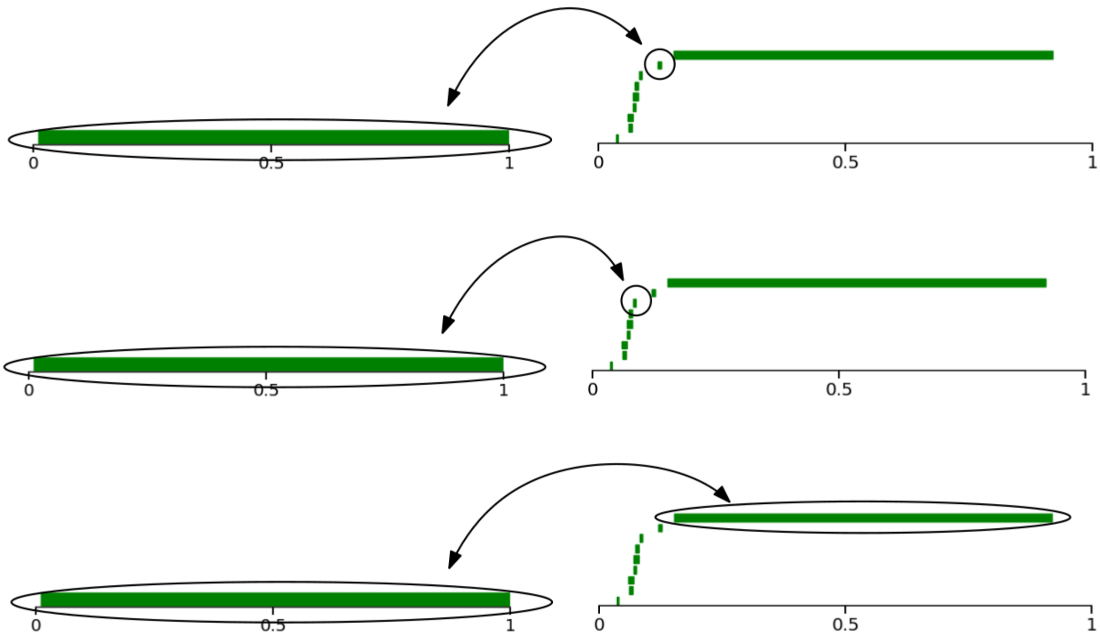
Definition 10.1. The *bottleck distance* between two barcodes P and Q is defined as the infimum of costs over all the partial matchings:

$$\mathrm{d_b}(P, Q) = \inf\{\operatorname{cost}(M),\ M \text{ is a partial matching between } P \text{ and } Q\}.$$

If $\mathbb{U}$ and $\mathbb{V}$ are two decomposable persistence modules, we define their *bottleneck* *distance* as

$$\mathrm{d_b}\left(\mathbb{U}, \mathbb{V}\right) = \mathrm{d_b}\left(\mathrm{Diagram}\left(\mathbb{U}\right), \mathrm{Diagram}\left(\mathbb{V}\right)\right).$$

Among the three following matchings, the last one is optimal:



Example 10.2. Let $a, a', b, b' \in \mathbb{R}^+$ such that $a \leq b$ and $a' \leq b'$. Consider the barcodes $P = \{[a, b]\}$ and $Q = \{[a', b']\}$.

First, consider the empty matching $M = \emptyset$. The intervals are matched to their projection, and the cost is

$$\left|(a, b) - \left(\frac{a+b}{2}, \frac{a+b}{2}\right)\right|_\infty = \frac{b-a}{2}$$
$$\left|(a', b') - \left(\frac{a'+b'}{2}, \frac{a'+b'}{2}\right)\right|_\infty = \frac{b'-a'}{2}$$

The total cost is $\mathrm{cost}(M) = \max\left\{\frac{b-a}{2}, \frac{b'-a'}{2}\right\}$. Next, consider the matching $M' = \left\{\left((a, b), (a', b')\right)\right\}$. The intervals are matched together, and the cost of the pair is

$$\left|(a, b) - (a', b')\right|_\infty = \max\{|a - a'|, |b - b'|\}.$$

which is also $\mathrm{cost}(M')$. These are the only two partial matchings, and we deduce the bottleneck distance

$$\mathrm{d_b}\left(P, Q\right) = \min\left\{\max\left\{\frac{b-a}{2}, \frac{b'-a'}{2}\right\}, \max\{|a - a'|, |b - b'|\}\right\}.$$

Example 10.3. Let $a, a', b, b' \in \mathbb{R}^+$ such that $a \leq b$ and $a' \leq b'$. Consider the interval-modules $\mathbb{B}[a, b]$ and $\mathbb{B}[a', b']$. Their barcodes are the sets P and Q of the previous example, from which we deduce

$$\mathrm{d_b}\left(\mathbb{B}[a, b], \mathbb{B}[a', b']\right) = \min\left\{\max\left\{\frac{b-a}{2}, \frac{b'-a'}{2}\right\}, \max\{|a - a'|, |b - b'|\}\right\}.$$

Interleaving distance. We now define an algebraic-flavored distance. Consider two persistence modules $\mathbb{V}$ and $\mathbb{W}$:

$$\begin{tikzcd} {} \arrow[r, dashed] & V^{t_1} \arrow[r, "v_{t_1}^{t_2}"] & V^{t_2} \arrow[r, "v_{t_2}^{t_3}"] & V^{t_3} \arrow[r, "v_{t_3}^{t_4}"] & V^{t_4} \arrow[r, dashed, dash] & {} \\ {} \arrow[r, dashed] & W^{t_1} \arrow[r, "w_{t_1}^{t_2}"] & W^{t_2} \arrow[r, "w_{t_2}^{t_3}"] & W^{t_3} \arrow[r, "w_{t_3}^{t_4}"] & W^{t_4} \arrow[r, dashed, dash] & {} \end{tikzcd}$$

Given $\epsilon \geq 0$, an ϵ-*morphism* between $\mathbb{V}$ and $\mathbb{W}$ is a family of linear maps $\phi = (\phi_t \colon V^t \to W^{t+\epsilon})_{t\in\mathbb{R}^+}$ such that the following diagram commutes for every $s \leq t \in \mathbb{R}^+$:

$$\begin{tikzcd} V^s \arrow[r, "v_s^t"] \arrow[d, "\phi_s"] & V^t \arrow[d, "\phi_t"] \\ W^{s+\epsilon} \arrow[r, "w_{s+\epsilon}^{t+\epsilon}"] & W^{t+\epsilon} \end{tikzcd}$$

If $\epsilon = 0$ and each ϕ_t is an isomorphism, we recover the notion of *isomorphism of persistence modules* (Definition 9.15).

An ϵ-*interleaving* between $\mathbb{V}$ and $\mathbb{W}$ is a pair of ϵ-morphisms $(\phi_t \colon V^t \to W^{t+\epsilon})_{t\in\mathbb{R}^+}$ and $(\psi_t \colon W^t \to V^{t+\epsilon})_{t\in\mathbb{R}^+}$ such that the following diagrams commute for every $t \in \mathbb{R}^+$:

$$\begin{tikzcd} V^t \arrow[rr, "v_t^{t+2\epsilon}"] \arrow[dr, "\phi_t"'] & & V^{t+2\epsilon} \\ & W^{t+\epsilon} \arrow[ur, "\psi_{t+\epsilon}"'] & \end{tikzcd} \qquad \begin{tikzcd} & V^{t+\epsilon} \arrow[dr, "\phi_{t+\epsilon}"] & \\ W^t \arrow[ur, "\psi_t"] \arrow[rr, "w_t^{t+2\epsilon}"] & & W^{t+2\epsilon} \end{tikzcd}$$

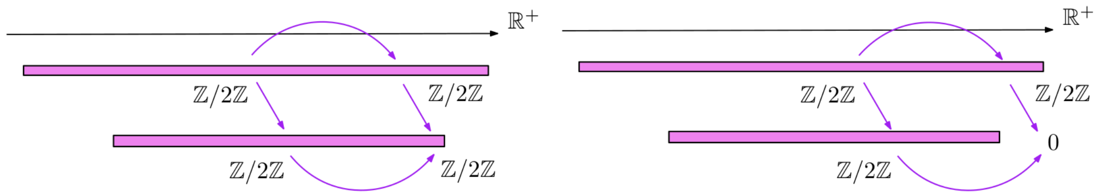
Definition 10.4. The *interleaving distance* between two persistence modules $\mathbb{V}$ and $\mathbb{W}$ is defined as

$$\mathrm{d_i}\left(\mathbb{V}, \mathbb{W}\right) = \inf\{\epsilon \geq 0, \mathbb{V} \text{ and } \mathbb{W} \text{ are } \epsilon\text{-interleaved}\}.$$

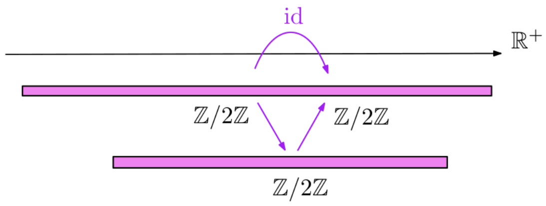
Example 10.5. Let $a, a', b, b' \in \mathbb{R}^+$ such that $a \leq b$ and $a' \leq b'$. Consider the interval-modules $\mathbb{B}[a, b]$ and $\mathbb{B}[a', b']$. Let us find an ϵ-interleaving (ϕ, ψ). By definition of an ϵ-morphism, we see that ϕ must be:

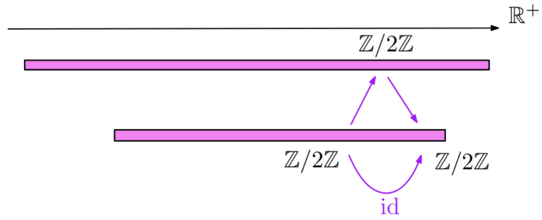
- always the zero map, or
- always nonzero when V^t and $W^{t+\epsilon}$ are nonzero.

We deduce a similar statement for ψ.



Now, by definition of an ϵ-interleaving, we deduce that $\psi_{t+\epsilon} \circ \phi_t$ must be nonzero when $[t, t + \epsilon] \subset [a, b]$. Similarly, $\phi_{t+\epsilon} \circ \psi_t$ must be nonzero when $[t, t + \epsilon] \subset [a', b']$.





Gathering these results, we see that an ϵ-interleaving exists if and only if

- $|a-b| \leq 2\epsilon$ and $|a'-b'| \leq 2\epsilon$
- or $|a-a'| \leq \epsilon$ and $|b-b'| \leq \epsilon$.

We deduce the interleaving distance:

$$\mathrm{d_i}\left(\mathbb{B}[a,b], \mathbb{B}[a',b']\right) = \min\left\{\max\left\{\frac{b-a}{2}, \frac{b'-a'}{2}\right\}, \max\{|a-a'|, |b-b'|\}\right\}$$

Note that this is the same formula as in Example 10.3.

10.2 Isometry theorem

At this point, the category of interval-decomposable modules is endowed with two notions of distance: the interleaving distance and the bottleneck distance. The are actually equal:

Theorem 10.6 ([CdSGO16, Theorem 4.11])**.** *If the persistence modules $\mathbb{U}$ and $\mathbb{V}$ are interval-decomposable, then $\mathrm{d_i}(\mathbb{U}, \mathbb{V}) = \mathrm{d_b}(\mathbb{U}, \mathbb{V})$.*

This result falls into two parts: *stability*, $\mathrm{d_i}(\mathbb{U}, \mathbb{V}) \geq \mathrm{d_b}(\mathbb{U}, \mathbb{V})$, and *converse stability*, $\mathrm{d_i}(\mathbb{U}, \mathbb{V}) \leq \mathrm{d_b}(\mathbb{U}, \mathbb{V})$. The second one is easier to prove:

Proof (of converse stability): Let us write the decomposition of the persistence modules in intervals:

$$\mathbb{V} \simeq \bigoplus_{I \in \mathcal{I}} \mathbb{B}[I] \qquad\qquad \mathbb{W} \simeq \bigoplus_{J \in \mathcal{J}} \mathbb{B}[J]$$

Suppose that we have a ϵ-partial matching $M \subset \mathcal{I} \times \mathcal{J}$. This gives a matching of some intervals (I, J), where $I = (a, b)$ and $J = (a', b')$, such that $|a - a'| \leq \epsilon$ and $|b - b'| \leq \epsilon$. According to our study in Example 10.5, we can build an ϵ-interleaving between $\mathbb{B}[I]$ and $\mathbb{B}[J]$, that we denote $(\phi_{(I,J)}, \psi_{(I,J)})$. Some intervals I (resp. J) are not matched, in which case their length is not greater than 2ϵ, and we can build an ϵ-interleaving with the zero persistence module. We denote this interleaving $(\phi_{(I,0)}, \psi_{(I,0)})$ (resp. $(\phi_{(0,J)}, \psi_{(0,J)})$).

Now, let us consider the sums of all these linear maps:

$$\overline{\phi} = \bigoplus_{(I,J) \text{ matched}} \phi_{(I,J)} \oplus \bigoplus_{I \text{ not matched}} \phi_{(I,0)}$$

$$\overline{\psi} = \bigoplus_{(I,J) \text{ matched}} \psi_{(I,J)} \oplus \bigoplus_{J \text{ not matched}} \phi_{(0,J)}$$

One verifies that $(\overline{\phi}, \overline{\psi})$ is an ϵ-interleaving. These considerations being true for any partial matching M, we deduce $\mathrm{d_i}(\mathbb{U}, \mathbb{V}) \leq \mathrm{d_b}(\mathbb{U}, \mathbb{V})$. $\square$

The stability part is less simple to prove. One way of tackling the problem consists in using the interpolation lemma:

Lemma 10.7 ([CdSGO16, Lemma 3.4]). *If $\mathbb{U}$ and $\mathbb{V}$ are δ-interleaved, then there exists a family of persistence modules $(\mathbb{U}_t)_{t\in[0,\delta]}$ such that $\mathbb{U}_0 = \mathbb{U}$, $\mathbb{U}_\delta = \mathbb{V}$ and $\mathrm{d_i}\left(\mathbb{U}_s, \mathbb{U}_t\right) \leq |s-t|$ for every $s, t \in [0, \delta]$.*

The theorem then follows from the box lemma [CdSGO16, Lemma 4.22] and a compacity argument. Another proof of the stability theorem is given in [BL13], which has the advantage of building an explicit partial matching from an interleaving.

10.3 Stability theorem

We present of particular case of the stability theorem, although in its full generality it states the stability of persistence modules of sublevel-set filtrations. It is a direct consequence of the isometry theorem stated in the previous section:

Theorem 10.8 ([CSEH07]). *Let X and Y be two subsets of $\mathbb{R}^n$. Consider their Čech (resp. Rips) filtrations, and the corresponding i^{th} homology persistence modules, $\mathbb{U}$ and $\mathbb{V}$. Suppose that they are interval-decomposables. Then $\mathrm{d_b}\left(\mathbb{U}, \mathbb{V}\right) \leq d_{\mathrm{H}}(X, Y)$ (Hausdorff distance).*

Proof. Let $\epsilon = d_{\mathrm{H}}(X, Y)$. We have seen that the thickenings satisfy $X \subset Y^\epsilon$ and $Y \subset X^\epsilon$. By using Exercise 37, we even have that $X^t \subset Y^{t+\epsilon}$ and $Y^t \subset X^{t+\epsilon}$ for all $t \geq 0$. By denoting j and k these inclusions, we have a commutative diagram

$$\begin{tikzcd}
{} \arrow[r, dashed] & X^t \arrow[rr, hook] \arrow[dr, "j_t"'] & & X^{t+2\epsilon} \arrow[rr, hook] \arrow[dr, "j_{t+2\epsilon}"'] & & X^{t+4\epsilon} \arrow[dr, "j_{t+4\epsilon}"'] \arrow[rr, hook, dashed, no head] & & {} \\
{} \arrow[rr, dashed] & & Y^{t+\epsilon} \arrow[rr, hook] \arrow[ur, "k_{t+\epsilon}"] & & Y^{t+3\epsilon} \arrow[rr, hook] \arrow[ur, "k_{t+3\epsilon}"] & & Y^{t+5\epsilon} \arrow[r, hook, dashed, no head] & {}
\end{tikzcd}$$

This also gives inclusions between Čech complexes (see the warning below). We could have chosen Rips complexes here.

$$\begin{tikzcd}
\check{\mathrm{C}}\mathrm{ech}^t(X) \arrow[rr, hook] \arrow[dr, "j_t"'] & & \check{\mathrm{C}}\mathrm{ech}^{t+2\epsilon}(X) \arrow[rr, hook] \arrow[dr, "j_{t+2\epsilon}"'] & & \check{\mathrm{C}}\mathrm{ech}^{t+4\epsilon}(X) \arrow[dr, "j_{t+4\epsilon}"'] \arrow[rr, hook, dashed, no head] & & {} \\
{} \arrow[r, dashed] & \check{\mathrm{C}}\mathrm{ech}^{t+\epsilon}(Y) \arrow[rr, hook] \arrow[ur, "k_{t+\epsilon}"] & & \check{\mathrm{C}}\mathrm{ech}^{t+3\epsilon}(Y) \arrow[rr, hook] \arrow[ur, "k_{t+3\epsilon}"] & & \check{\mathrm{C}}\mathrm{ech}^{t+5\epsilon}(Y) &
\end{tikzcd}$$

Now, we apply the i^{th} homology functor. We still obtain a commutative diagram:

$$\begin{tikzcd}
H_i(\check{\mathrm{C}}\mathrm{ech}^t(X)) \arrow[rr] \arrow[dr, "(j_t)_*"'] & & H_i(\check{\mathrm{C}}\mathrm{ech}^{t+2\epsilon}(X)) \arrow[rr] \arrow[dr, "(j_{t+2\epsilon})_*"'] & & H_i(\check{\mathrm{C}}\mathrm{ech}^{t+4\epsilon}(X)) \arrow[dr, "(j_{t+4\epsilon})_*"'] \arrow[rr, dashed, no head] & & {} \\
{} \arrow[r, dashed] & H_i(\check{\mathrm{C}}\mathrm{ech}^{t+\epsilon}(Y)) \arrow[rr] \arrow[ur, "(k_{t+\epsilon})_*"] & & H_i(\check{\mathrm{C}}\mathrm{ech}^{t+3\epsilon}(Y)) \arrow[rr] \arrow[ur, "(k_{t+3\epsilon})_*"] & & H_i(\check{\mathrm{C}}\mathrm{ech}^{t+5\epsilon}(Y)) &
\end{tikzcd}$$

On the first row, we recognize the persistence module corresponding to X, and on the second row the persistence module corresponding to Y. Observe that the maps (j, k) form an ϵ-interleaving between them. Hence their interleaving distance is bounded by ϵ. By invoking the isometry theorem, we obtain the result. $\square$

Warning: We actually cheated a little bit in the previous proof. The maps $j_t \colon X^t \to Y^{t+\epsilon}$ and $k_t \colon Y^t \to X^{t+\epsilon}$ may not induce well defined simplicial maps $j_t \colon \check{\mathrm{C}}\mathrm{ech}^t(X) \to \check{\mathrm{C}}\mathrm{ech}^{t+\epsilon}(Y)$ and $k_t \colon \check{\mathrm{C}}\mathrm{ech}^t(Y) \to \check{\mathrm{C}}\mathrm{ech}^{t+\epsilon}(X)$. In order to prove the statement properly, we should have used *singular homology*, a theory of homology more suited to deal with topological spaces.

> **Exercise 53.** Let $\mathcal{M}$ be the unit circle of $\mathbb{R}^2$, and $X \subset \mathbb{R}^2$ a finite subset. Denote the Hausdorff distance $\epsilon = d_{\mathrm{H}}(X, \mathcal{M})$. Suppose that ϵ is small enough. Let $\mathbb{U}$ denote the persistence module of the 1st homology of the Čech filtration of X. Using the stability theorem, deduce the existence of a bar in the barcode, and give a lower bound on its persistence. Compare your result with Exercise 51.

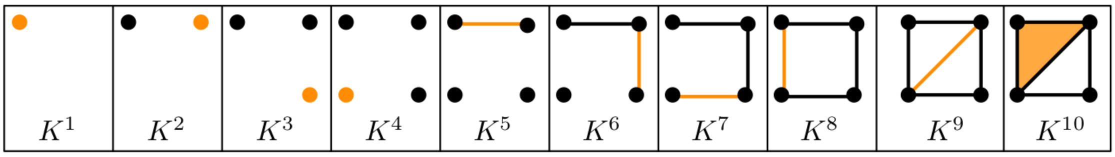
11 Python tutorial

Notebook available at
https://github.com/raphaeltinarrage/EMAp/blob/main/Tutorial3.ipynb.

Our code starts with

```
import gudhi
import numpy as np
import matplotlib.pyplot as plt
```

First, we build the filtration of Exercise 52.




```
st = gudhi.SimplexTree()

# insert vertices
st.insert([0]) # the default filtration value is 0
st.insert([1])
st.insert([2])
st.insert([3])

# insert edges
st.insert([0,1], 1/2) # we give the filtration value 1/2
st.insert([1,2], 1/2)
st.insert([2,3], 1/2)
st.insert([3,0], 1/2)
```

```
#insert diagonal and triangle
st.insert([1,3], np.sqrt(2)/2)
st.insert([0,1,3], np.sqrt(2)/2)
```

Then we compute the barcodes of the filtration.

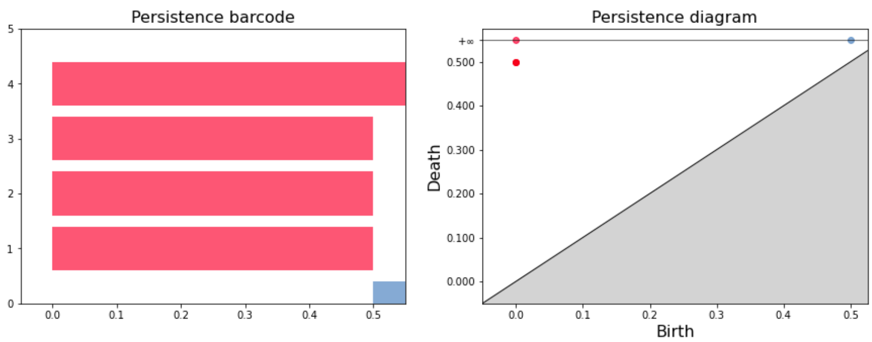
```
barcode = st.persistence(homology_coeff_field = 2)
print(barcode)
```

We obtain the following result. Each element $(d,(a,b))$ corresponds to the interval (a,b) in the d^{th} homology group H_d.

```
[(1, (0.5, inf)), (0, (0.0, inf)), (0, (0.0, 0.5)),
(0, (0.0, 0.5)), (0, (0.0, 0.5))]
```

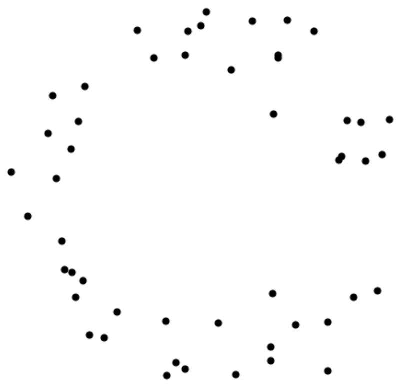
We can plot the barcodes and the persistence diagram. H_0 is represented in red, and H_1 in blue:

```
fig = plt.figure(figsize=(15,5))
ax1 = fig.add_subplot(1,2,1); ax2 = fig.add_subplot(1,2,2)

gudhi.plot_persistence_barcode(barcode, axes = ax1)
gudhi.plot_persistence_diagram(barcode, axes = ax2)
```



To continue, let us consider a noisy sample of the circle.

```
X = SampleOnCircle(N=50, sd=0.15)

# plotting the point cloud
fig = plt.figure(figsize=(8,8)); ax = fig.add_subplot(1, 1, 1)
plt.scatter(X[:,0],X[:,1], c='black', s=50)
plt.axis('equal'); plt.axis('off'); plt.show()
```



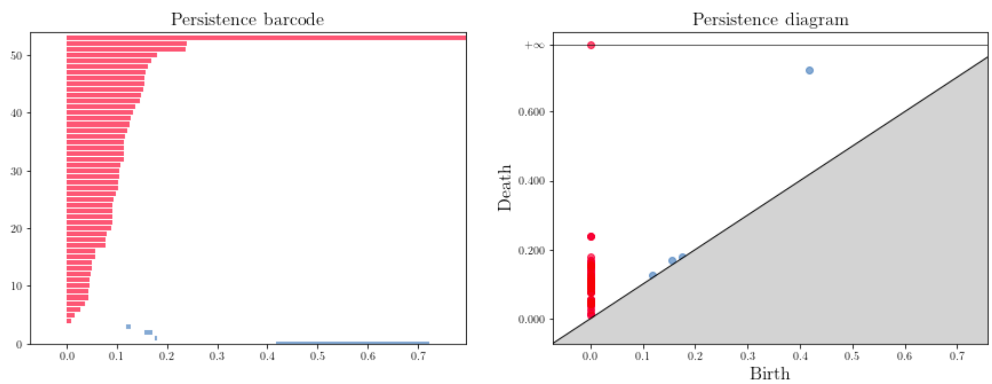
We create its Rips filtration, specifying its maximal edge length, and its maximal simplex dimension. In order to compute homology up to dimension d accurately, we have to insert the simplices up to dimension $d+1$.

Note that GUDHI uses another definition of the Rips complex at time t than us (it considers that it is the Rips complex at time $2t$). In order to correct this difference, we give the point cloud X divided by 2.

```
rips = gudhi.RipsComplex(points = X/2, max_edge_length = 1)
st = rips.create_simplex_tree(max_dimension=2)
        # we add the simplices up to dimension 2
```

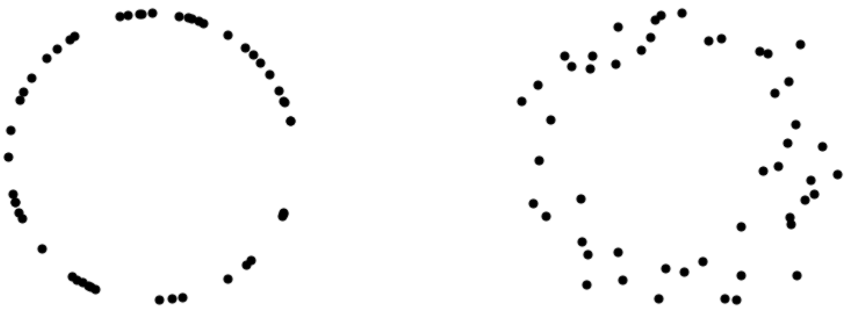
Now we can compute the barcodes.

```
barcodes = st.persistence(homology_coeff_field = 2)

fig = plt.figure(figsize=(15,5))
ax1 = fig.add_subplot(1,2,1); ax2 = fig.add_subplot(1,2,2)

gudhi.plot_persistence_barcode(barcodes, axes = ax1)
gudhi.plot_persistence_diagram(barcodes, axes = ax2)
```



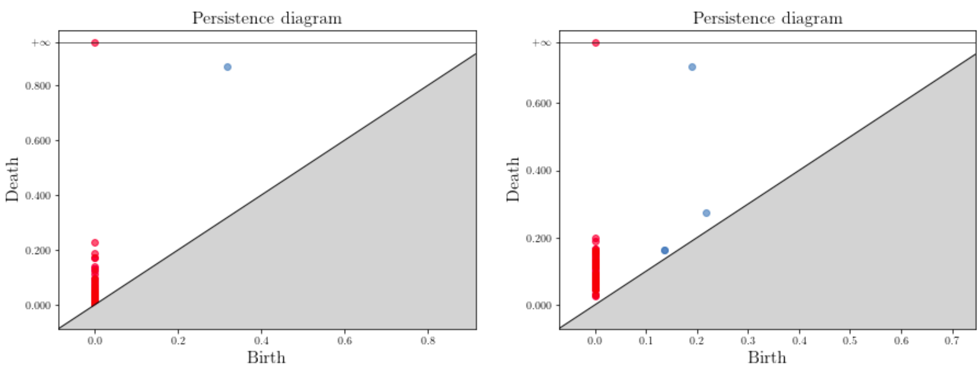
Finally, let us verify the stability theorem on an example. We generate two datasets in $\mathbb{R}^2$.

```
X = SampleOnCircle(N=50, sd=0)
Y = SampleOnCircle(N=50, sd=0.15)

# plotting the point cloud
fig = plt.figure(figsize=(15,5))
ax1 = fig.add_subplot(1,2,1); ax2 = fig.add_subplot(1,2,2)
ax1.scatter(X[:,0],X[:,1], c='black', s=50);
ax2.scatter(Y[:,0],Y[:,1], c='black', s=50);
```



We compute the barcodes of their Rips filtrations:

```
RipsX = gudhi.RipsComplex(points = X/2, max_edge_length = 2)
stX = RipsX.create_simplex_tree(max_dimension=2)
barcodeX = stX.persistence(homology_coeff_field = 2)
RipsY = gudhi.RipsComplex(points = X/2, max_edge_length = 2)
stY = RipsY.create_simplex_tree(max_dimension=2)
barcodeY = stY.persistence(homology_coeff_field = 2)
fig = plt.figure(figsize=(15,5))
ax1 = fig.add_subplot(1,2,1); ax2 = fig.add_subplot(1,2,2)
gudhi.plot_persistence_diagram(barcodeX, axes = ax1)
gudhi.plot_persistence_diagram(barcodeY, axes = ax2)
```



We compute the bottleneck distances, and observe that they are lower than the Hausdorff distance between X and Y.

```
barcodeX_0 = stX.persistence_intervals_in_dimension(0)
barcodeY_0 = stY.persistence_intervals_in_dimension(0)
bottleneck_distance_0 = gudhi.bottleneck_distance(barcodeX_0, barcodeY_0)
print('Bottleneck distance between H_0: '+repr(bottleneck_distance_0))

barcodeX_1 = stX.persistence_intervals_in_dimension(1)
barcodeY_1 = stY.persistence_intervals_in_dimension(1)
bottleneck_distance_1 = gudhi.bottleneck_distance(barcodeX_1, barcodeY_1)
print('Bottleneck distance between H_1: '+repr(bottleneck_distance_1))

from scipy.spatial.distance import directed_hausdorff
Hausdorff = max(directed_hausdorff(X, Y)[0], directed_hausdorff(Y, X)[0])
print('Hausdorff distance: '+repr(Hausdorff))
```

We obtain:

```
Bottleneck distance between H_0: 0.06381195771835965
Bottleneck distance between H_1: 0.16048548662007345
Hausdorff distance: 0.40380876493233253
```

Exercise 54. The *cyclo-octane molecules* dataset has been presented in Section 8. The authors simulated many molecules, resulting in a point cloud in $\mathbb{R}^{72}$. However, the dataset given in the notebook only contains the positions of the *carbon* atoms, hence we have a point cloud in $\mathbb{R}^{24}$.

Compute the barcodes of the Rips filtration of this dataset, up to dimension 3, and with a maximal edge length of 0.3.

Exercise 55. A flute and a clarinet have been recorded playing the note A. The recordings have been transformed into an array of length 39000 (flute) and length 96000 (clarinet). They last approximately 1 and 2 seconds (rate 44100 Hz).

For each of the two instruments,

- extract some samples of 500 points
- embbed them into $\mathbb{R}^2$ via time delay embedding
- compute the H^1-barcodes of their Rips filtration
- compute the number of cycles with persistence greater than 0.03, call them the *cycles with large persistence*
- compute the mean number of *cycles with large persistence* over all the samples (of a given instrument)

Indications:

- compute a time delay embedding with the function `TimeDelayEmbedding` given in the notebook, with `edim = 2` and `delay = 2`

- compute the Rips complex with `max_edge_length = 0.2`

Exercise 56. The walk of three persons A, B and C has been recorded using the accelerometer sensor of a smartphone in their pocket, giving rise to 3 multivariate time series in $\mathbb{R}^3$: each time series represents the 3 coordinates of the acceleration of the corresponding person in a coordinate system attached to the sensor (warning: as the smartphone was carried in a possibly different position for each person, these time series cannot be compared coordinates by coordinates). Using a sliding window, each series has been split in a list of 100 time series made of 200 consecutive points, that are now stored in `data_A`, `data_B` and `data_C`.

For each person,

- for each time series of 100 points, compute a time delay embedding with the function `TimeDelayEmbedding`, with `edim = 2` and `delay = 3`
- compute the H^1-barcode of its Rips filtration

Then, compute the bottleneck distance between each two pairs of barcodes. This gives you a distance matrix. Last, use multidimensional scaling to represent them in $\mathbb{R}^3$. Can you identify the three persons?

References

[AZ05] Gunnar Carlsson Afra Zomorodian. Computing persistent homology. *Discrete and Computational Geometry*, 2005. `https://geometry.stanford.edu/papers/zc-cph-05/zc-cph-05.pdf`.

[Azu50] Gorô Azumaya. Corrections and supplementaries to my paper concerning Krull-Remak-Schmidt's theorem. *Nagoya Mathematical Journal*, 1:117–124, 1950.

[BCB20] Magnus Botnan and William Crawley-Boevey. Decomposition of persistence modules. *Proceedings of the American Mathematical Society*, 148(11):4581–4596, 2020. `https://arxiv.org/abs/1811.08946`.

[BCY18] Jean-Daniel Boissonnat, Frédéric Chazal, and Mariette Yvinec. *Geometric and topological inference*, volume 57. Cambridge University Press, 2018. `https://hal.inria.fr/hal-01615863/`.

[BGO19] Nicolas Berkouk, Grégory Ginot, and Steve Oudot. Level-sets persistence and sheaf theory. Preprint. `https://arxiv.org/abs/1907.09759`, 2019.

[BL13] Ulrich Bauer and Michael Lesnick. Induced matchings and the algebraic stability of persistence barcodes. Preprint. `https://arxiv.org/abs/1311.3681`, 2013.

[CB15] William Crawley-Boevey. Decomposition of pointwise finite-dimensional persistence modules. *Journal of Algebra and its Applications*, 14(05):1550066, 2015. `https://arxiv.org/abs/1210.0819`.

[CCSL09] Frédéric Chazal, David Cohen-Steiner, and André Lieutier. A sampling theory for compact sets in Euclidean space. *Discrete & Computational Geometry*, 41(3):461–479, 2009. `https://link.springer.com/article/10.1007/s00454-009-9144-8`.

[CDS10] Gunnar Carlsson and Vin De Silva. Zigzag persistence. *Foundations of computational mathematics*, 10(4):367–405, 2010. `https://link.springer.com/article/10.1007/s10208-010-9066-0`.

[CdSGO16] Frédéric Chazal, Vin de Silva, Marc Glisse, and Steve Oudot. *The Structure and Stability of Persistence Modules*. SpringerBriefs in Mathematics, 2016. `https://arxiv.org/abs/1207.3674`.

[CG20] Gunnar Carlsson and Rickard Brüel Gabrielsson. Topological approaches to deep learning. In *Topological Data Analysis*, pages 119–146. Springer, 2020. `https://arxiv.org/abs/1811.01122`.

[CIDSZ08] Gunnar Carlsson, Tigran Ishkhanov, Vin De Silva, and Afra Zomorodian. On the local behavior of spaces of natural images. *International journal of computer vision*, 76(1):1–12, 2008. `https://www.researchgate.net/publication/220659347_On_the_Local_Behavior_of_Spaces_of_Natural_Images`.

[CSEH07] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete & computational geometry*, 37(1):103–120, 2007. https://link.springer.com/article/10.1007/s00454-006-1276-5.

[CZ09] Gunnar Carlsson and Afra Zomorodian. The theory of multidimensional persistence. *Discrete & Computational Geometry*, 42(1):71–93, 2009. https://link.springer.com/article/10.1007/s00454-009-9176-0.

[DSG07] Vin De Silva and Robert Ghrist. Coverage in sensor networks via persistent homology. *Algebraic & Geometric Topology*, 7(1):339–358, 2007. https://projecteuclid.org/euclid.agt/1513796672.

[ELZ00] Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological persistence and simplification. In *Proceedings 41st annual symposium on foundations of computer science*, pages 454–463. IEEE, 2000. http://www.math.uchicago.edu/ shmuel/AAT-readings/Data

[Hat02] Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 2002. https://pi.math.cornell.edu/~hatcher/AT/AT.pdf.

[KNBNH16] Violeta Kovacev-Nikolic, Peter Bubenik, Dragan Nikolić, and Giseon Heo. Using persistent homology and dynamical distances to analyze protein binding. *Statistical applications in genetics and molecular biology*, 15(1):19–38, 2016. https://pubmed.ncbi.nlm.nih.gov/26812805/.

[MTCW10] Shawn Martin, Aidan Thompson, Evangelos A Coutsias, and Jean-Paul Watson. Topology of cyclo-octane energy landscape. *The journal of chemical physics*, 132(23):234115, 2010. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3188624/.

[NLC11] Monica Nicolau, Arnold J Levine, and Gunnar Carlsson. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proceedings of the National Academy of Sciences*, 108(17):7265–7270, 2011. https://www.pnas.org/content/108/17/7265.

[NSW08] Partha Niyogi, Stephen Smale, and Shmuel Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete & Computational Geometry*, 39(1-3):419–441, 2008. http://people.cs.uchicago.edu/~niyogi/papersps/NiySmaWeiHom.pdf.

[OPT+17] Nina Otter, Mason A Porter, Ulrike Tillmann, Peter Grindrod, and Heather A Harrington. A roadmap for the computation of persistent homology. *EPJ Data Science*, 6:1–38, 2017. https://arxiv.org/abs/1506.08903.

[PC14] Jose A Perea and Gunnar Carlsson. A Klein-bottle-based dictionary for texture representation. *International journal of computer vision*, 107(1):75–97, 2014. https://fds.duke.edu/db/attachment/2638.

[SMC07] Gurjeet Singh, Facundo Mémoli, and Gunnar E Carlsson. Topological methods for the analysis of high dimensional data sets and 3d object recognition. *SPBG*, 91:100, 2007. `https://research.math.osu.edu/tgda/mapperPBG.pdf`.