# Topological Data Analysis

Albert Ruiz Cirera

Universitat Autònoma de Barcelona

Modelling for Science and Engineering

# Introduction

UAB Universitat Autònoma de Barcelona

**Question**

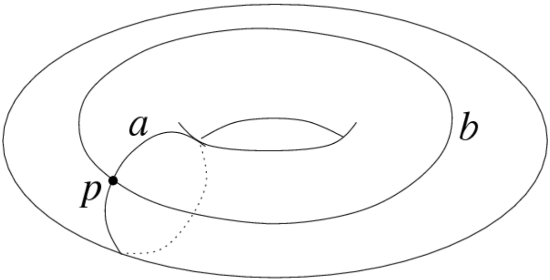
How can we describe the shape of an object?

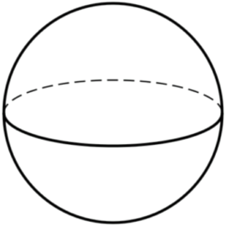
Describing it can be tedious and most of the information of this description can be useless.

**Topology** studies the shape in a very general way: it doesn't worry about continuous deformation, which means that it doesn't care about distances and angles.
We can stretch, crump and bend an object, but we cannot tear it nor glue it (to another object).

# A coffee cup is like a donut

# A donut is not like an orange



Any path in the sphere can be deformed continuously to a point, while paths *a* anb *b* in this torus can not.

# Topology

UAB Universitat Autònoma de Barcelona

## Invariants

In topology we assign to each geometric object an algebraic object in a "continuous way":

In the geometric object, continuous means "deformation", and we can do it.

In the algebraic object, continuous means constant.

So, if we can deform one object $X$ to another $Y$, we must assign the same algebraic object to both of them.

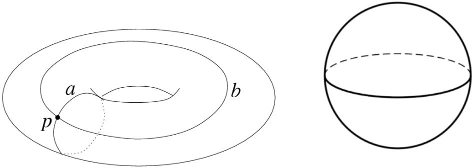
## Example

The algebraic object can be a number:

- If we have a subset of $\mathbb{R}^n$, count the number of pieces.
- If we have a manifold, consider the dimension.

# Donut $\neq$ orange

Neither the dimension (equals 2), nor the number of pieces (equals 1) distinguishes a torus from a sphere, but there are several algebraic invariants which distinguish them. Here we will see the Betti numbers.




$\beta = (1, 2, 1)$ $\qquad$ $\beta = (1, 0, 1)$

# How can we relate topology and data?

Universitat Autònoma de Barcelona

Gunnar Carlsson: "Data has shape."

- Foundational article: G. Carlsson, *Topology and data*, Bulletin of the AMS 46 (2009) 255–308.
- Ayasdi Inc. `http://www.ayasdi.com`

# Current projects (I)

Universitat Autònoma de Barcelona

## Ayasdi Inc

- With Leicester Univ., "Airway pathological heterogeneity in asthma: Visualization of disease microclusters using topological data analysis".
- With Univ. Lille, "When remote sensing meets topological data analysis"; "Exploring hyperspectral imaging data sets with topological data analysis".
- With Karolinska Inst., "Mass Cytometry and Topological Data Analysis Reveal Immune Parameters Associated with Complications after Allogeneic Stem Cell Transplantation".
- "A novel Approach to Identifying a Neuroimaging Biomarker for Patients With Serious Mental Illness".
- ...

# Current projects (II)

UAB Universitat Autònoma de Barcelona

## Blue Brain Project (EPFL)

- "A Topological Representation of Branching Neuronal Morphologies. ".doi.org/10.1007/s12021-017-9341-1.
- "Cliques of Neurons Bound into Cavities Provide a Missing Link between Structure and Function". doi.org/10.3389/fncom.2017.00048.
- "Rich cell-type-specific network topology in neocortical microcircuitry". doi:10.1038/nn.4576.
- "Quantifying Topological Invariants of Neuronal Morphologies". http://arxiv.org.abs/1603.08432
- . . .

# What can we expect from topology?

Universitat Autònoma de Barcelona

## Good things

1. Topology is coordinate independent, multidimensional and deformation free.
2. If there is any hidden non-trivial shape, topology will detect it.

## Problem

The shape is topologically trivial for almost all the data sets.

## Challenges

Solve the following problems:

1. Topology works with complete spaces, not isolated points.
2. Theorems in topology do not care about the computations.

# Algebraic Topology approach

Universitat Autònoma de Barcelona

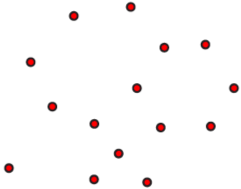
## Table of contents of a classical homology course

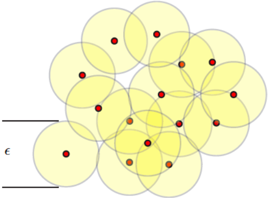
- Work with polyhedron: abstract polyhedron $K$ and its geometric realization $|K|$.
- Define homology of a polyhedron $K$.
- Develop algebraic tools to be able to compute homology groups of polyhedron.
- Associate to a topological space $X$ a (huge) polyhedron structure which just depends on the topology of $X$. Provide tools to compute this homology.
- Prove that the homology of a polyhedron $K$ is equal to the homology of its geometric realization $|K|$ as a topological space.

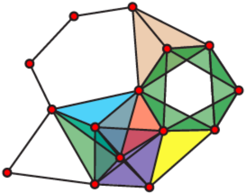
# Topological Data Analysis approach

## TDA

- Associate a polyhedron $K_\epsilon$ to a data set $D$ (will depend on a parameter $\epsilon > 0$).
- Compute the homology of $K_\epsilon$.





# Introduction to topology

UAB Universitat Autònoma de Barcelona

## Starting point

The starting point is a set $X$ where we define what an open subset $U \subset X$ is. The open sets must fulfill three axioms:

1. The empty set $\emptyset$ and $X$ are open subsets.
2. Any union of open subsets must be again an open subset.
3. Any finite intersection of open subsets must be again an open subset.

## Definition

A topological space $X$ is a set $X$ where the concept of open subset fulfilling the 3 axions above is defined.
If necessary, we write $(X, \tau)$, where $\tau$ is the family of open subsets in $X$.

## Definition

If $X$ is a topological space, we say that a subset $C \subset X$ is closed if $X \setminus C$ is an open subset.

# Examples

Universitat Autònoma de Barcelona

### Euclidean topology in $\mathbb{R}^n$

Consider the set $\mathbb{R}^n$ and define that $U \subset \mathbb{R}^n$ is an open subset if and only if for all $x \in U$ there is $\varepsilon > 0$ such that $B(x, \varepsilon) \subset U$, where:

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

- This is a particular case of a topology induced by a metric.
- Changing the metric may change (or not) the induced topology.

### More examples: consider $X$ a set

1. Trivial topology: the only open subsets are $\emptyset$ and $X$. If $X$ has more than one element, this topology is not induced by a metric. this is the most coarse topology on $X$.
2. Discrete topology: define that all subsets of $X$ are open sets. This is the finest topology you can consider on $X$.

# Generators of a topology

UAB Universitat Autònoma de Barcelona

### Definition

Given $\mathcal{A}$, a family of subsets of $X$, we can define the topology generated by $\mathcal{A}$ as the coarsest topology on $X$ such that the elements on $A$ are open subsets.

Equivalently, $U$ is an open subset if it is the empty set, the total or union of finite intersections of elements in $\mathcal{A}$.

Equivalently, $U$ is an open subset if it is the empty set, the total or for all $x \in U$ there is an element $A \in \mathcal{A}$ such that $x \in A \subset U$.

### Example

- $\mathcal{A} = \{B(x, \varepsilon)\}_{x \in \mathbb{R}^n, \varepsilon > 0}$ as defined before generates the Euclidean topology in $\mathbb{R}^n$.
- $\mathcal{A} = \emptyset$ generates the trivial topology on any $X$.
- $\mathcal{A} = \{\{x\}\}_{x \in X}$ generates the discrete topology on any $X$.

# Continuous maps

UAB Universitat Autònoma de Barcelona

### Definition

If $X$ and $Y$ are topological spaces, we say that $f\colon X \to Y$ (a map as sets) is continuous if for all $V$ open subset of $Y$, $f^{-1}(V)$ is an open subset of $X$.

### Definition

$f\colon X \to Y$, a map between topological spaces, is a homeomorphism if $f$ is bijective, continuous and $f^{-1}$ is also continuous.
We say that $X$ is homeomorphic to $Y$ (and write $X \cong Y$) is there exists a homeomorphism $f\colon X \to Y$.

### Exercise

1. Prove that "be homeomorphic to" is an equivalence relation.
2. Give an example for $f\colon X \to Y$ bijective, continuous and such that $f^{-1}$ is not continuous.

# Properties of continuous maps

- The identity map from one topological space to itself is continuous.
- The composition of continuous maps is continuous.
- The constant map is always continuous.
- When considering the Euclidean topology in $\mathbb{R}^n$, the definition of continuous map is the same as the usual one in calculus.
- If the source space has the discrete topology, all set maps are continuous.
- If the target space has de trivial topology, all set maps are continuous.

# Topological properties

UAB Universitat Autònoma de Barcelona

### Homeomorphism

A topological property of a topological space $X$ is a property which can be defined from its open subsets. Any homeomorphism $f\colon X \to Y$ satisfies the property that $U \subset X$ is open if and only if $f(U) \subset Y$ is open. This implies, that any topological property of $X$ is transmitted to $Y$ by $f$.

### Example

If we consider $\mathbb{R}$ with the Euclidean topology:

- $[0,1] \cong [a,b]$ for all $a < b \in \mathbb{R}$.
- $(0,1) \cong (a,b)$ for all $a < b \in \mathbb{R}$.
- $(0,1) \cong \mathbb{R}$.

# Comparing topologies

Given a set $X$, we can consider more than one topology on $X$: $(X, \tau_1)$, $(X, \tau_2)$, getting different topological paces.

## Definition

We say that $\tau_1$ is finer than $\tau_2$ (or $\tau_2$ is coarser than $\tau_1$) if all the open subsets in $\tau_2$ are open also in $\tau_1$.
Equivalently, $\tau_2 \subset \tau_1$.
Equivalently, $\mathrm{Id}\colon (X, \tau_1) \to (X, \tau_2)$ is continuous.

## Example

$$(\mathbb{R}^n, \{\emptyset, \mathbb{R}^n\}) \subset (\mathbb{R}^n, \mathrm{Eucl}) \subset (\mathbb{R}^n, \mathrm{Discrete})$$

# Construction of new topologies

UAB Universitat Autònoma de Barcelona

## Universal constructions

Now we can consider a sets map $f\colon X \to Y$:

- Initial topology: we assume that $Y$ is a topological space. We define the initial topology on $X$ over $f$ as the coarsest topology on $X$ such that $f$ is continuous.
- Final topology: we assume that $X$ is a topological space. We define the final topology on $Y$ over $f$ as the finest topology on $Y$ such that $f$ is continuous.

# Examples



Assume that $X$ and $Y$ are topological spaces.

1. Subspace: If $A \subset X$, we can consider the initial topology on $A$ induced by the inclusion $i\colon A \to X$. In other words $V \subset A$ is open if and only if there is an open subset $U \subset X$ such that $V = A \cap U$.
2. Quotient: If we define an equivalence relation on $X$: $\sim$, we can define a topology on $X/\sim$ as the final topology of the map $X \to X/\sim$.
3. Product: If we consider the set $X \times Y$, and the two projections $\pi_X\colon X \times Y \to X$, $\pi_Y\colon X \times Y \to Y$, we define a topology on $X \times Y$ as the coarser one such that $\pi_X$ and $\pi_Y$ are continuous.

**Exercise**

We can define the topology in $\mathbb{R}^n$ as the product topology on $\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}$. Prove that the Euclidean topology in $\mathbb{R}^n$ is the product topology of $n$ times the Euclidean topology in $\mathbb{R}$.

# Topological properties: compactness

UAB Universitat Autònoma de Barcelona

## Compactness

We say that a $\{U_\alpha\}_{\alpha \in A}$, a family of open subsets of $X$, is an open cover of $X$ if for all $x \in X$, there is $U_{\alpha_x}$ such that $x \in U_{\alpha_x}$.
We say that $X$ is compact if and only if any open cover of $X$ can be reduced to a finite open subcover.

## Theorem

*A subset $A$ of $\mathbb{R}^n$ with the Euclidean topology is compact if and only if $A$ is closed and bounded.*

## Example (Subspaces of $\mathbb{R}^n$ with the Euclidean topology)

- $[0,1]$ is compact.
- $S^n = \{x \in \mathbb{R}^{n+1} \mid ||x|| = 1\}$ is compact.
- $\mathbb{R}$ is not compact, and then $(0,1)$ isn't neither.

# Topological properties: connected

UAB Universitat Autònoma de Barcelona

## Connected

We say that $X$ is connected if and only if when we write $X = U \cup V$, $U$ and $V$ open sets such that $U \cap V = \emptyset$, then either $U = \emptyset$ or $V = \emptyset$. Equivalently, $X$ is connected if it is not the disjoint union of two non-empty open sets.

## Example (Subspaces of $\mathbb{R}^n$ with the Euclidean topology)

- $[0, 1]$ is connected.
- $S^n = \{x \in \mathbb{R}^{n+1} \mid ||x|| = 1\}$ is connected if $n \geq 1$.
- $S^0 = \{-1, 1\}$ and $\mathbb{Z} \subset \mathbb{R}$ are not connected.
- $\mathbb{Q} \subset \mathbb{R}$ is not connected.

## Exercise

Prove that $\mathbb{Q} \not\cong \mathbb{Z}$ as a subspaces of $\mathbb{R}$ with the Euclidean topology.

# Topological properties: Hausdorff

Universitat Autònoma de Barcelona

## Hausdorff

We say that $X$ is Hausdorff if and only for any $x \neq y \in X$, there are open subsets $U_x$ and $U_y$ such that $x \in U_x$, $y \in U_y$ and $U_x \cap U_y = \emptyset$.

## Properties

- Any metric space is Hausdorff.
- In particular, $\mathbb{R}^n$ with the Euclidean topology.
- This property is preserved by subspaces: so any subspace of $\mathbb{R}^n$ with the Euclidean topology is Hausdorff.

## Example (of non Hausdorff)

Any $X$ with at least two elements with the trivial topology.

# Recall of (finite dimensional) vector spaces

We work with coefficients over a field $F$, which mainly will be either $F = \mathbb{R}$ or $F = \mathbb{F}_2$. Moreover, all vector spaces that we will consider here are finite dimensional.

## Definition ($n$-dimensional $F$-vector space)

Recall that a $F$-vector space $E$ with basis $e_1, \ldots, e_n$ is the set of sums $w = \lambda_1 e_1 + \cdots + \lambda_n e_n$, with $\lambda_i \in F$, which we usually write as an $n$-tuple $(\lambda_1, \ldots, \lambda_n)$ with the following properties:

1. We can sum vectors $w_1 + w_2$ (coordinate-wise),
2. we can multiply vectors by scalars $\lambda w$ (multiply all coordinates),
3. sum is commutative and associative: $w_1 + w_2 = w_2 + w_1$, $(w_1 + w_2) + w_3 = w_1 + (w_2 + w_3)$ $(\forall w_1, w_2, w_3 \in E)$,
4. there is a zero vector 0: $0 + w = w$ $(\forall w \in E)$,
5. works well with scalars: $0w = 0$, $1w = w$ and $(\lambda_1\lambda_2)w = \lambda_1(\lambda_2 w)$ $(\forall w \in E)$.

# Dimension and subspaces

**Definition**

The dimension of a vector space $E$ is the number of elements of a basis.

**Definition**

If $E$ is a vector space, a subset $V \subset E$ is a vector subspace if satisfies:

- $v_1 + v_2 \in V$, for all $v_1, v_2 \in V$.
- $\lambda v \in V$ for all $v \in V$ and $\lambda \in F$.

# Linear maps



## Definition

If $E$, $G$ are $\mathbb{R}$-vector spaces, a linear map $f\colon E \to G$ is a map such that: $f(w_1 + w_2) = f(w_1) + f(w_2)$ and $f(\lambda w) = \lambda f(w)$.

## Remark

If $e_1, \dots, e_n$ is a basis for $E$, and $g_1, \dots, g_m$ is a basis for $G$, we can characterize any linear map $f\colon E \to G$ with a $m \times n$-matrix $M_f$ (which depends on the basis).

If $f(u_j) = \sum_i a_{ij} v_i$, then:

$$M_f = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{pmatrix}$$

# Kernel and image

Consider $f\colon E \to G$ a linear map and fix basis on $E$ and $G$, in such a way that we can consider $M_f$.

**Definition**

We define the kernel of $f$ as the vectors $e \in E$ such that $f(e) = 0$.
We define the image of $f$ as the vectors $g \in G$ such that there exists $e \in E$ such that $f(e) = g$.

**Proposition**

- *The kernel of $f$ is a vector subspace of $E$.*
- *The image of $f$ is a subspace of $G$.*
- $\mathrm{Rank}(M_f) = \dim(\mathrm{Im}(f))$.
- *There is the relation:*

$$\dim(E) = \dim(\mathrm{Ker}(f)) + \dim(\mathrm{Im}(f)).$$

# Properties of linear maps and subspaces

UAB Universitat Autònoma de Barcelona

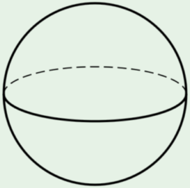
- The zero map is linear.
- The identity map from one vector space to itself is linear.
- The composition of linear maps is linear.
- $f(0) = 0$ for all $f$ a linear map.
- A linear map $f$ is injective if and only if $\ker(f) = \{0\}$.
- If $E$ is a subspace of a finite dimensional vector space $G$ such that $\dim(E) = \dim(G)$, then $E = G$.

# Introduction to combinatorial topology

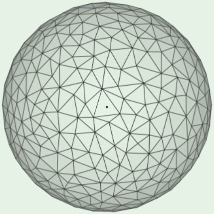
This section will provide tools to study some topological spaces combinatorially, replacing the original one by a triangulated one.

## Example

Study the properties of a 2-dimensional sphere using a triangulation (a vertices, edges and triangles structure):



$\cong$



## AIM

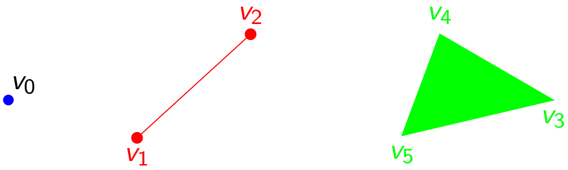
Get information from the original space $X$ from the triangulation $T_X$.

# $n$-Simplex

## Definition

Given $n+1$ points $(v_0, v_1, \ldots, v_n)$ in $\mathbb{R}^m$ such that $v_1 - v_0, \ldots, v_n - v_0$ are linearly independent, we define the *$n$-simplex* determined by these points as:

$$\Delta = [v_0, \ldots, v_n] = \left\{ \sum_{i=0}^{n} \lambda_i v_i \mid 0 \leq \lambda_i \leq 1, \sum_{i=0}^{n} \lambda_i = 1 \right\}$$




$[v_0]$ $[v_1, v_2]$ $[v_3, v_4, v_5]$

# Properties of a $n$-simplex

UAB Universitat Autònoma de Barcelona

## Proposition

*A n-simplex is a compact, connected and Hausdorff space.*

## Remark

There are $n$-simplices for any integer $n > 0$: If $e_i$ is the vector of $\mathbb{R}^{n+1}$ defined as all coordinates 0 but the $i$-th position, which is 1, then $[e_1, e_2, \dots, e_{n+1}]$ determines an $n$-dimensional simplex (called standard $n$-simplex).

# Faces of a $n$-simplex

UAB Universitat Autònoma de Barcelona

## Definition

Given a $n$-simplex $[v_0, \dots, v_n]$, a $k$-face is a $k$-simplex defined by $k+1$ points in $\{v_0, \dots, v_n\}$.

## Example

The 2-simplex $[v_0, v_1, v_2]$ has as faces:

- 0-faces (vertices): $[v_0]$, $[v_1]$ and $[v_2]$.
- 1-faces (edges): $[v_0, v_1]$, $[v_0, v_2]$ and $[v_1, v_2]$.
- 2-face: $[v_0, v_1, v_2]$.

## Convention: vertices are ordered

The vertices of a face spanned by a subset of the vertices of a simplex will always be ordered according to their order in the larger simplex.

# Polyhedron (or complex)

### Definition

A (finite) polyhedron $K$ is a (finite) union of simplices:

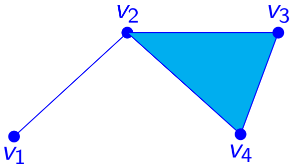
$$K = \bigcup_{1 \leq i \leq m} \Delta_i$$

such that:

- The faces of any $\Delta_i$ are in $K$.
- The intersection of two simplices of $K$: $\Delta_i \cap \Delta_j$ is either empty or a face of both simplices (so also in $K$).

De dimension of $K$ is the biggest $n$ such that there is a $n$-simplex in $K$.

$[v_1]$,$[v_2]$,$[v_3]$,$[v_4]$,
$[v_1, v_2]$, $[v_2, v_3]$,$[v_3, v_4]$,$[v_2, v_4]$,
$[v_2, v_3, v_4]$

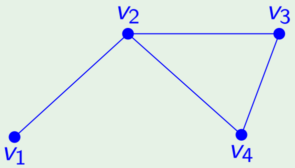


# Examples of polyhedron

Universitat Autònoma de Barcelona

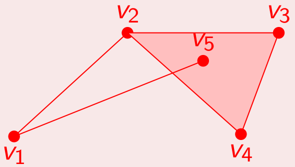
## Example

$[v_1],[v_2],[v_3],[v_4],$
$[v_1, v_2], [v_2, v_3],[v_3, v_4],[v_2, v_4]$




## Non-example

$[v_1],[v_2],[v_3],[v_4],[v_5]$
$[v_1, v_2], [v_2, v_3],[v_3, v_4],[v_2, v_4],[v_1, v_5],$
$[v_2, v_3, v_4]$



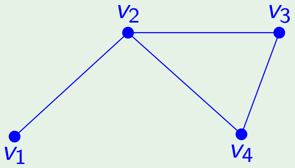
# Aim

Universitat Autònoma de Barcelona

Detect the hole which distinguish the following two examples:

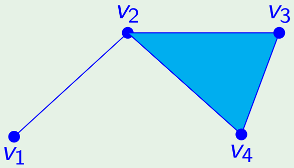
## Example (1)

$[v_1],[v_2],[v_3],[v_4],$
$[v_1, v_2],[v_2, v_3],[v_3, v_4],[v_2, v_4]$




## Example (2)

$[v_1],[v_2],[v_3],[v_4],$
$[v_1, v_2],[v_2, v_3],[v_3, v_4],[v_2, v_4],$
$[v_2, v_3, v_4]$



# Conventions

UAB Universitat Autònoma de Barcelona

- To describe a polyhedron, it is enough to list the maximal faces.
  For example, to describe Example 1, it is enough to talk about the polyhedron which contains $[v_1, v_2]$, $[v_2, v_3]$, $[v_3, v_4]$ and $[v_2, v_4]$.
  To describe Example 2, we just need to say the polyhedron which contains $[v_1, v_2]$ and $[v_2, v_3, v_4]$.
- We can think a simplex $[v_0, \dots, v_n]$ as a $n$-dimensional polyhedron with maximal face $[v_0, \dots, v_n]$.
  For example $[v_0, v_1, v_2]$, as a polyhedron, contains the simplices $[v_0]$, $[v_1]$, $[v_2]$, $[v_0, v_1]$, $[v_1, v_2]$, $[v_0, v_2]$, and $[v_0, v_1, v_2]$.
- A map of polyhedron $f\colon K \to L$ is a map which sends vertices to vertices and any point of any simplex $x = \lambda_0 v_0 + \cdots + \lambda_i v_i$ to $f(x) = \lambda_0 f(v_0) + \cdots + \lambda_i f(v_i)$.
- The identity map $\mathrm{Id}\colon K \to K$ is a map of polyhedron, and the composition of two maps of polyhedron is a map of polyhedron.

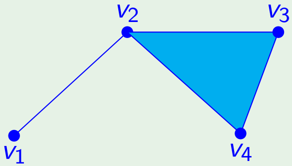
# Vector spaces associated to a polyhedron

## Definition

Given a polyhedron $K$, define $C_*(K)$ as the vector space with basis each simplex of $K$. We consider $C_i(K)$ the vector subspace with basis the simplices of dimension $i$.

## Example

$[v_1]$,$[v_2]$,$[v_3]$,$[v_4]$,
$[v_1, v_2]$, $[v_2, v_3]$,$[v_3, v_4]$,$[v_2, v_4]$,
$[v_2, v_3, v_4]$




$C_0$ has basis $[v_1]$,$[v_2]$,$[v_3]$,$[v_4]$ (so, dimension 4),
$C_1$ has basis $[v_1, v_2]$, $[v_2, v_3]$,$[v_3, v_4]$,$[v_2, v_4]$ (so, dimension 4) and
$C_2$ has basis $[v_2, v_3, v_4]$ (so, dimension 1).

# Boundary maps

UAB Universitat Autònoma de Barcelona

### Definition

Consider $C_i(K)$ the vector space with basis the $i$-dimensional simplices of $K$. We define the boundary map $\partial_i$ from $C_i(K)$ to $C_{i-1}(K)$ as the linear map such that to each element of the basis is defined as:

$$\partial_i([v_0, \ldots, v_i]) = \sum_{k=0}^{i} (-1)^k [v_0, v_1, \ldots, \hat{v}_k, \ldots, v_i]$$

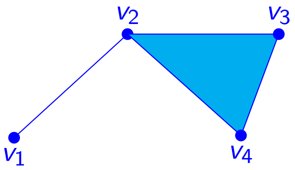
where $\hat{v}_k$ means that we remove that position.
By definition $\partial_0([v_i]) = 0$.

### Definition

- We define the $i$-boundaries of $K$ as the image of $\partial_{i+1}$: $\mathrm{Im}(\partial_{i+1})$.
- We define the $i$-cycles of $K$ as the kernel $\partial_i$: $\mathrm{Ker}(\partial_i)$.

# Example

$[v_1], [v_2], [v_3], [v_4],$
$[v_1, v_2], [v_2, v_3], [v_3, v_4], [v_2, v_4],$
$[v_2, v_3, v_4]$




**Example ($[v_3, v_4] - [v_2, v_4] + [v_2, v_3]$ is a boundary)**

$$\partial_2([v_2, v_3, v_4]) = [v_3, v_4] - [v_2, v_4] + [v_2, v_3]$$

**Example ($[v_3, v_4] - [v_2, v_4] + [v_2, v_3]$ is a cycle)**

$$\partial_1([v_3, v_4] - [v_2, v_4] + [v_2, v_3]) = [v_4] - [v_3] - ([v_4] - [v_2]) + [v_3] - [v_2] = 0$$

# Chain complexes

UAB Universitat Autònoma de Barcelona

## Definition

In general, a sequence of vector spaces $C_i$ and linear maps $\partial_i \colon C_i \to C_{i-1}$ such that $\partial_{i-1} \circ \partial_i = 0$ is called a chain complex.

## Definition

A morphism of chain complexes $f_* \colon C_* \to D_*$ is a sequence of linear maps $f_i \colon C_i \to D_i$ such that $\delta_i \circ f_i = \partial_i \circ f_{i-1}$, where $\partial_i$ and $\delta_i$ are the boundari maps for $C_i$ and $D_i$ respectively.

$$\begin{tikzcd}
\dots \arrow[r] & C_{i+1} \arrow[r, "\partial_{i+1}"] \arrow[d, "f_{i+1}"] & C_i \arrow[r, "\partial_i"] \arrow[d, "f_i"] & C_{i-1} \arrow[r, "\partial_{i-1}"] \arrow[d, "f_{i-1}"] & \cdots \\
\dots \arrow[r] & D_{i+1} \arrow[r, "\delta_{i+1}"] & D_i \arrow[r, "\delta_i"] & D_{i-1} \arrow[r, "\delta_{i-1}"] & \cdots
\end{tikzcd}$$

The identity map is a morphism of chain complexes and the composition of morphisms of chain complexes is a morphism of chain complexes.

# Homology groups

## Proposition

*Consider $K$ a polyhedron, $C_i(K)$ the vector space generated by the simplices of dimension $i$, and $\partial_i \colon C_i(K) \to C_{i-1}(K)$ the $i$-th boundary map. Then*

$$\partial_i \circ \partial_{i+1} = 0\,.$$

So, we get that this implies that $\mathrm{Im}(\partial_{i+1}) \subset \mathrm{Ker}(\partial_i)$.

## Definition

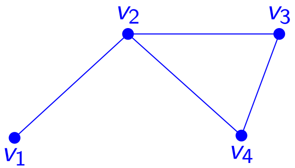
Consider $K$ a polyhedron, $C_i(K)$ the vector space generated by the simplices of dimension $i$, and $\partial_i \colon C_i(K) \to C_{i-1}(K)$ the $i$-th boundary map. Then we define the $i$-th homology group of $K$ as:

$$H_i(K) = \mathrm{Ker}(\partial_i) / \mathrm{Im}(\partial_{i+1})\,.$$

# Example 1

Consider $K$ as the following polyhedron:

$[v_1]$,$[v_2]$,$[v_3]$,$[v_4]$,
$[v_1, v_2]$, $[v_2, v_3]$,$[v_3, v_4]$,$[v_2, v_4]$




The boundary maps, in these basis are:

$$\partial_0 = \begin{pmatrix} 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\partial_1 = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

We can see that $\mathrm{Rank}(\partial_1) = 3$.

# Example 1

### $H_0(K)$

$$H_0(K) = \operatorname{Ker}(\partial_0)/\operatorname{Im}(\partial_1) = F^4/F^3 \cong F$$

Where $\operatorname{Ker}(\partial_0) = \langle [v_0], [v_1], [v_2], [v_3], [v_4] \rangle \cong F^4$ and $\operatorname{Im}(\partial_1) = \langle [v_1] - [v_0], [v_2] - [v_1], [v_3] - [v_2] \rangle \cong F^3$.

### $H_1(K)$

$$H_1(K) = \operatorname{Ker}(\partial_1)/\operatorname{Im}(\partial_2) = F$$

As $\operatorname{Ker}(\partial_1) = \langle [v_3, v_4] - [v_2, v_4] + [v_2, v_3] \rangle \cong F$ and $\partial_2 = 0$.

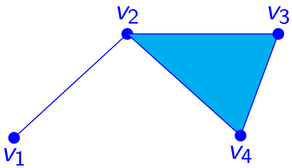
### $H_i(K)$, $i \geq 2$

As there are not simplices of dimension $i \geq 2$, we get that $H_i(K) = 0$ when $i \geq 2$.

# Example 2

Consider $K$ as the following polyhedron:

$[v_1],[v_2],[v_3],[v_4],$
$[v_1, v_2],[v_2, v_3],[v_3, v_4],[v_2, v_4],$
$[v_2, v_3, v_4]$




The boundary maps, in these basis are: $\partial_0$ and $\partial_1$ as in Example 1.

$$\partial_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ -1 \end{pmatrix}$$

We can see that $\mathrm{Rank}(\partial_2) = 1$.

# Example 2

## $H_0(K)$ (as Example 1)

$$H_0(K) = \operatorname{Ker}(\partial_0)/\operatorname{Im}(\partial_1) = F^4/F^3 \cong F$$

Where $\operatorname{Ker}(\partial_0) = \langle [v_0], [v_1], [v_2], [v_3], [v_4] \rangle \cong F^4$ and
$\operatorname{Im}(\partial_1) = \langle [v_1] - [v_0], [v_2] - [v_1], [v_3] - [v_2] \rangle \cong F^3$.

## $H_1(K)$

$$H_1(K) = \operatorname{Ker}(\partial_1)/\operatorname{Im}(\partial_2) = F/F \cong 0$$

As $\operatorname{Ker}(\partial_1) = \langle [v_3, v_4] - [v_2, v_4] + [v_2, v_3] \rangle \cong F$ and
$\operatorname{Im}(\partial_2) = \langle [v_3, v_4] - [v_2, v_4] + [v_2, v_3] \rangle \cong F$.

## $H_2(K)$

$$H_2(K) = \operatorname{Ker}(\partial_2)/\operatorname{Im}(\partial_3) = 0 \text{ as } \operatorname{Ker}(\partial_2) = 0.$$

# Induced map in chain complexes

Universitat Autònoma de Barcelona

## Definition

If we have $f\colon K \to L$ an injective map of polyhedrons such that preserve the order of the vertices, then we define the induced map of chain complexes $C_*(f)\colon C_*(K) \to C_*(L)$ as the linear map such that

$$C_i(f)([v_0, \dots, v_i]) = [f(v_0), \dots, f(v_i)]\,.$$

## Remark

This definition can be extended to:

- Non injective maps: sending to 0 all the simplices $[v_0, \dots, v_i]$ such that exists $k \neq l$ with $f(v_k) = f(v_l)$,
- and also to $f$ which do not preserve the order: change the sign as many times as necessary to get the correct order by transpositions: $[v_i, v_j] = -[v_j, v_i]$.

# Induced map in chain complexes

**Proposition**

*If $f\colon K \to L$ a map of polyhedrons, then $C_*(f)\colon C_*(K) \to C_*(L)$ is a morphism of chain complexes.*

**Properties**

- The identity map $\mathrm{Id}\colon K \to K$ induces the identity map $C_*(\mathrm{Id})\colon C_*(K) \to C_*(K)$. In other words:

$$C_*(\mathrm{Id}) = \mathrm{Id}\ .$$

- The composition of maps induced by $f\colon K \to L$ and $g\colon L \to M$ is the same as the induced map by the composition $g \circ f$. In other words:

$$C_*(g) \circ C_*(f) = C_*(g \circ f)\,.$$

# Induced map in homology

### Theorem

*If $f_*\colon C_* \to D_*$ is a morphism of chain complexes, then:*

- $f_i(\mathrm{Ker}(\partial_i)) \subset \mathrm{Ker}(\delta_i)$.
- *If* $x \in \mathrm{Im}(\partial_{i+1})$, *then* $f_i(x) \in \mathrm{Im}(\delta_{i+1})$.
- *Then,* $f_*$ *induces a map in homology groups:* $H_i(f_*)\colon H_i(C) \to H_i(D)$ *where* $H_i(f_*)([\sigma]) = [f_i(\sigma)]$.

### Properties

- The identity map $\mathrm{Id}_*\colon C_* \to C_*$ induces the identity map: $H_i(\mathrm{Id}_*) = \mathrm{Id}_i$.
- The composition of maps induced by $f_*\colon C_* \to D_*$ and $g\colon D_* \to E_*$ is the same as the induced map by the composition $g \circ f$.

$$H_i(g_*) \circ H_i(f_*) = H_i(g_* \circ f_*)\,.$$

# Example

Universitat Autònoma de Barcelona

## Example

If we consider $K$=Example 1 and $L$=Example 2 above, and $f$ the induced map such that at the level of vertices sends $f(v_i) = v_i$.
Then, $f$ induces:

- $f_0 = \mathrm{Id}_0 \colon C_0(K) \to C_0(L)$, with $C_0(K) \cong C_0(L) \cong F^4$.
- $f_1 = \mathrm{Id}_1 \colon C_1(K) \to C_1(L)$, with $C_1(K) \cong C_1(L) \cong F^4$.
- $f_2 = 0 \colon C_2(K) \to C_2(L)$, with $C_2(K) = 0$ and $C_2(L) \cong F$.
- $f_i = 0 \colon C_i(K) \to C_i(L)$, with $C_i(K) = C_i(L) = 0$, $\forall i \geq 3$.
- $H_0(f) = \mathrm{Id} \colon H_0(K) \to H_0(L)$, with $H_0(K) \cong H_0(L) \cong F$.
- $H_1(f) = 0 \colon H_1(K) \to H_1(L)$, with $H_1(K) \cong F$ and $H_1(L) = 0$.
- $H_2(f) = 0 \colon H_2(K) \to H_2(L)$, with $H_2(K) = H_2(L) = 0$.
- $H_i(f) = 0 \colon H_i(K) \to H_i(L)$, with $H_i(K) = H_i(L) = 0$, $\forall i \geq 3$.

# Betti numbers

UAB
Universitat Autònoma de Barcelona

**Definition**

We define the *$i$-th Betti number of $K$* as:

$$\beta_i(K) = \dim(H_i(K))\,.$$

**Example**

1. In Example 1 we had $\beta_0(K) = \beta_1(K) = 1$ and $\beta_i(K) = 0$ if $i \geq 2$.
2. In Example 2 we had $\beta_0(K) = 1$ and $\beta_i(K) = 0$ if $i \geq 1$.

# Computation of $\beta_0$

## Equivalence relation

We say that to vertices $[v_i]$ and $[v_j]$ are related, $[v_i] \sim [v_j]$, if $i = j$ or there are vertices $[w_0], \ldots, [w_k]$ such that $v_i = w_0$, $v_j = w_k$ and $[w_i, w_{i+1}]$ is an edge of $K$ for all $i$.
We can check that this is an equivalence relation.

## Proposition

*The first Betti number of $K$, $\beta_0(K)$ is the number of classes under this equivalence relation.*

# Euler-Poincaré Characteristic

Universitat Autònoma de Barcelona

## Definition

Given a (finite) polyhedron $K$ we define the Euler-Poincaré characteristic of $K$:

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

where $c_i(K) = \dim(C_i(K))$, so $c_i(K)$ is the number of $i$-dimensional simplices in $K$.

## Example

1. The Euler-Poincaré characteristic of Example 1 is $4 - 4 = 0$.
2. The Euler-Poincaré characteristic of Example 2 is $4 - 4 + 1 = 1$.

# More examples

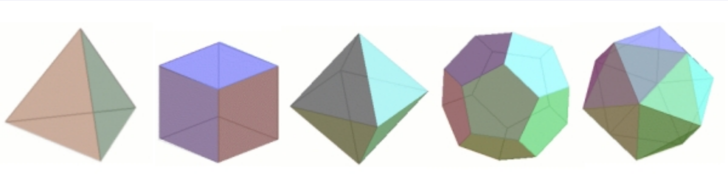
UAB
Universitat Autònoma de Barcelona

### Exercise

Compute the Euler-Characteristic of the polyhedron with the unique maximal face $[v_0, v_1, \ldots, v_n]$.

### Exercise

Compute the Euler-Poincaré characteristic of the 2-dimensional Platonic solids (non filled):



# Euler-Poincaré characteristic and homology

**Theorem**

*If $K$ is a finite polyhedron, then:*

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

## Proof

Recall the notation on $K$: $c_i(K)$ is the number of $i$-simplices in $K$, so, $c_i(K) = \dim(C_i(K))$. Then:

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

Use now that $\dim(C_i(K)) = \dim(\mathrm{Ker}(\partial_i)) + \dim(\mathrm{Im}(\partial_i))$, so $\chi(K)$ is the alternate sum of $\dim(\mathrm{Ker}(\partial_i))$ + the alternate sum of $\dim(\mathrm{Im}(\partial_i))$. In both cases, when $i$ is even, we consider the positive sign, and negative when $i$ is odd.

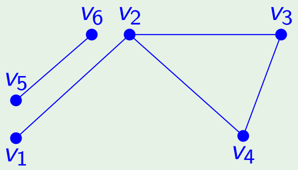
Now, $\beta_0(K) = \dim(H_0(K)) = \dim(C_0(K)) - \dim(\mathrm{Im}(\partial_1))$, and, for $i > 0$: $\beta_i(K) = \dim(H_i(K)) = \dim(\mathrm{Ker}(\partial_i)) - \dim(\mathrm{Im}(\partial_{i+1}))$. So, the alternate sum of the $\beta_i$ is the alternating sum of $\dim(\mathrm{Ker}(\partial_i))$ (positive when $i$ is even and negative when $i$ is odd) and $\dim(\mathrm{Im}(\partial_{i+1}))$ (negative when $i$ is even and positive when $i$ is odd). But the shift in the index $\dim(\mathrm{Im}(\partial_{i+1}))$ means that $\dim(\mathrm{Im}(\partial_i))$ is added as positive when $i$ is even and negative when $i$ is odd.

# Examples

UAB Universitat Autònoma de Barcelona

## Example

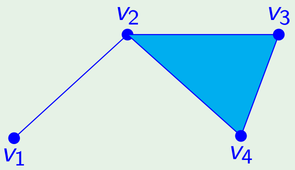
$[v_1],[v_2],[v_3],[v_4],[v_5],[v_6],$
$[v_1,v_2],[v_2,v_3],[v_3,v_4],[v_2,v_4],[v_5,v_6]$




$$\beta_0 = 2,\ \beta_1 = 1,\ \beta_k = 0 \text{ if } k \geq 2.$$

## Example

$[v_1],[v_2],[v_3],[v_4],$
$[v_1,v_2],[v_2,v_3],[v_3,v_4],[v_2,v_4],$
$[v_2,v_3,v_4]$




$$\beta_0 = 1,\ \beta_k = 0 \text{ if } k \geq 1.$$

# From points to polyhedron

There is more than one way to get a polyhedron from a points cloud in $\mathbb{R}^n$, which we divide into two families:

### Distance between points

These methods associate to a points set a polyhedron taking into account the distance between points. So, the starting point is a (symmetric) matrix containing all the distances between points and construct a polyhedron structure from it.
Here we will see the Čech and Vietoris-Rips constructions.

### Using auxiliary functions

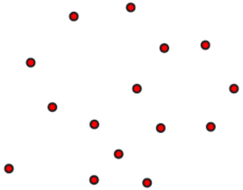
These methods construct a function using $X$ from the ambient space $f_X : \mathbb{R}^n \to \mathbb{R}^k$ and replaces the point set by the preimage of a cover of the image of $f_X$.

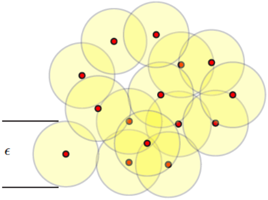
# Čech polyhedron

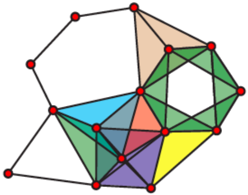
## Čech polyhedron

Consider a data set $X$ as points and fix a real number $\epsilon > 0$. Cover the points set with balls of diameter $\epsilon$ centered at each point and consider the polyhedron $C(X, \epsilon)$:

- The vertices are the points.
- $k$ vertices determine a simplex if and only if the corresponding centered balls have non-empty intersection.





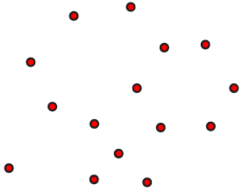



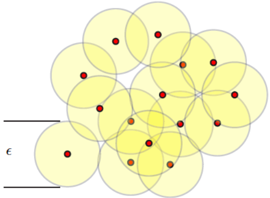
# Vietoris-Rips polyhedron

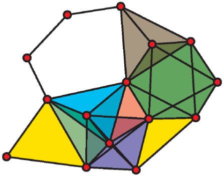
## Vietoris-Rips polyhedron

Consider a data set $X$ as points and fix a real number $\epsilon > 0$. Cover the points set with balls of diameter $\epsilon$ centered at each point and consider the polyhedron $VR(X, \epsilon)$:

- The vertices are the points.
- $k$ vertices determine a simplex if and only if for all combinations of two of these vertices, the corresponding two centered balls have non-empty intersection.





## Properties

### Remarks

- For $\epsilon$ near to zero, we will have $\beta_0 =$ number of points and all other Betti number equals 0.
- For $\epsilon$ very big, we will have $\beta_0 = 1$ and all other Betti number equals 0.
- Any hole which modifies the Betti numbers will survive in a period $[\epsilon_1, \epsilon_2]$ (one probably different period for each hole).

### Proposition

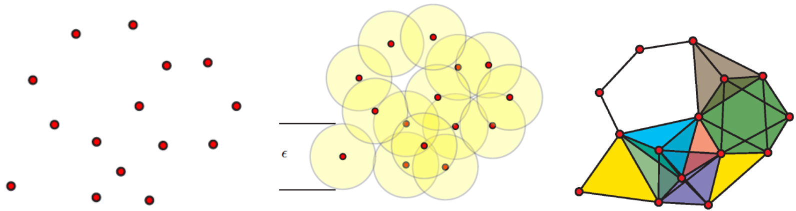
*Given a points set $X$ and $\epsilon > 0$, we have the inclusions:*

$$C(X,\epsilon) \subset VR(X,\epsilon) \subset C(X,2\epsilon)\,.$$

# Čech vs Vietoris-Rips

## Čech vs Vietoris-Rips

- Vietoris-Rips is better for computations: Čech polyhedron needs to check lots of intersections, while Vietoris-Rips can be deduced from the vertices and edges structure.
- The Vietoris-Rips polyhedron (usually) does not correspond to the subspace determined by the balls centered in the points cloud in the ambient space.



# Good and bad properties of these methods

## Good things

- The polyhedron is directly computed from the points and can be deduced from a distance matrix (easy to compute).
- The Čech complex of $X \subset R^n$ is actually a representation of a subspace of $\mathbb{R}^n$.

## Bad things

- If $X \subset \mathbb{R}^n$ has $N$ points, the polyhedron can have up to $2^N$ faces (so, will be huge).
- Do not use that (probably) $n$ is very small comparing to $N$: it will deal and use a lot of computations with simplices in dimension $k > n$ which, in this case, can not generate any new elements in homology.
- A little bit of noise removes possible important holes.

# Data reduction: witness polyhedron

Consider $X$ a points set and $L \subset X$ a significantly smaller subset.

### Definition

In the notation above, $\Delta \leq L$. We say that $x \in X$ is a weak witness for $\Delta$ with respect to $L$ is an only if $||x - a|| \leq ||x - b||$ for all $a \in \Delta$ and for all $b \in L \setminus \Delta$. We define the weak witness complex ox $X$ with respect to $L$ the one having as vertex set given by $L$, and $\Delta \subset L$ is a simplex if and only if has a weak witness in $X$.

### Variant of this definition

Consider $\epsilon > 0$ and replace $||x - a|| \leq ||x - b||$ by $||x - a|| \leq ||x - b|| + \epsilon$, defining weak $\epsilon$-witness for $\Delta$ with respect to $L$

### Lazy witness

The lazy witness complex is the complex which uses the vertices and edges of the witness complex and defines a simplex $\Delta \subset L$ to be in the complex if and only if all possible edges are in the witness complex.

# Using auxiliary functions

Universitat Autònoma de Barcelona

Assume we are given a points set $X \subset \mathbb{R}^n$.

## Procedure

- Consider a bounding box $B$ and triangulate it to get a polyhedron structure $K$. This can be done in different ways. For example:
  - A rectangular bounding box: $B = [a_1, b_1] \times \cdots \times [a_n, b_n]$ for $X$, so $X \subset B \subset \mathbb{R}^n$.
  - Divide it in a $k_1 \times k_2 \times \cdots \times k_n$ regular grid (vertices).
  - Construct the edges between consecutive vertices, and triangles, ..., getting a polyhedron structure $K$.
- Consider a function $f_X \colon B \to \mathbb{R}$ which depends on $X$. Usually, $f_X(x)$ is higher (or lower) if $x$ is surrounded by many points of $X$.
- For each $r \in R$, consider the subpolyhedron of $K$ generated by the vertices in $f_X^{-1}((-\infty, r))$ (sublevel) or $f_X^{-1}((r, \infty))$ (superlevel).

# Auxiliary functions

Universitat Autònoma de Barcelona

We need a function $f_X$, where $X = \{x_1, \ldots, x_N\} \leq \mathbb{R}^n$ is the set of points we want to analyze.

### Gaussian Kernel Density Estimator (sublevel)

$$KDE(x) = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{h^n}\frac{1}{\sqrt{2\pi}} e^{-\frac{||x-x_i||^2}{h^2\sqrt{2\pi}}},$$

where $h > 0$ is a fixed parameter.

### Distance to measure (superlevel)

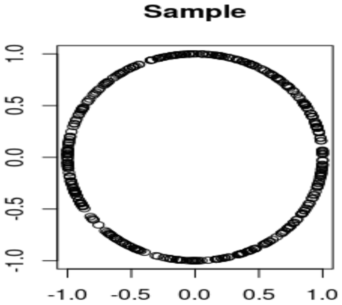
$$DTM(x) = \frac{1}{k}\sum_{x_i \in N_k(x)} ||x_i - x||^2,$$

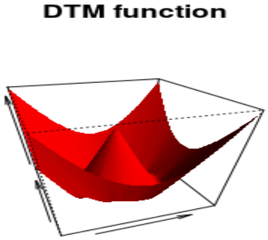
where $k$ is a positive integer and $N_k(x)$ is the subset of $X$ with the $k$ nearest points of $x$.

# Example Distance To Measure

Universitat Autònoma de Barcelona

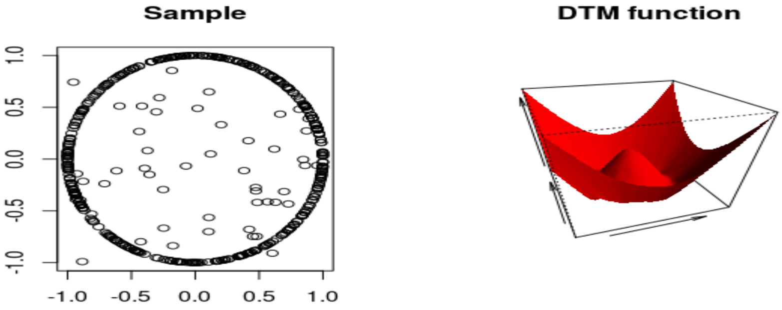
## Example (DTM with 400 points in a circle and 40 nearest points)



# Example Distance To Measure

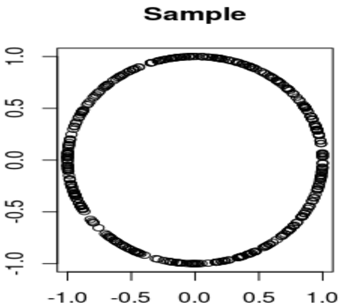
Example (DTM with 400 points in a circle with noise and 40 nearest points)

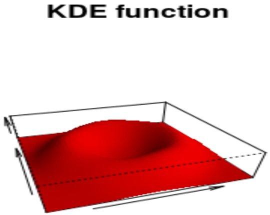


# Example Kernel Density Estimator

UAB Universitat Autònoma de Barcelona

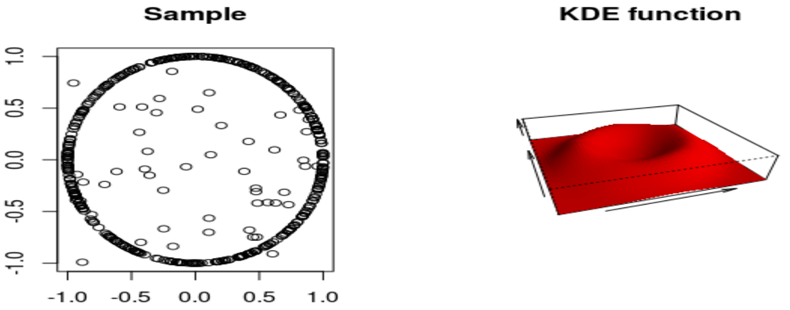
## Example (KDE with 400 points in a circle and $h = 0.3$)



# Example Kernel Density Estimator

## Example (KDE with 400 points in a circle with noise and $h = 0.3$)



# Good and bad things

UAB
Universitat Autònoma de Barcelona

## Good

- $X$ is only used to define $f_X$, so, the number of computations may not increase that much with the size of $X$.
- The dimension of the simplices is determined by the ambient space $\mathbb{R}^n$.

## Care

- We have to choose a function $f_X$.
- The computations increase adding points to the grid.

# Filtrations

When considering a point set $X$, $\epsilon < \epsilon'$ and the corresponding polyhedrons $K_\epsilon = K(X, \epsilon)$ and $K_{\epsilon'} = K(X, \epsilon')$ of any of the previous constructions. We get the inclusion:

$$\iota_{\epsilon,\epsilon'} \colon K_\epsilon \to K_{\epsilon'} \,.$$

This inclusion induces a linear map in homology:

$$H_*(\iota_{\epsilon,\epsilon'}) \colon H_*(K_\epsilon) \to H_*(K_{\epsilon'}) \,.$$

**Definition**

We say that a non-zero homology class $z \in H_i(K_\epsilon)$ survives in the interval $[\epsilon, \epsilon']$ if $H_i(\iota_{\epsilon,\epsilon'})(z) \neq 0$.

So, any class $z \neq 0$ is born at $\epsilon$ and (possibly) a dies at $\epsilon'$.

# Persistence diagrams

## Persistence diagram of a filtration

A persistence diagram associated to a filtration is a 2-dimensional graphic with the birth $\epsilon$ in the $x$ axis and the death $\epsilon'$ in the $y$ axis. Each point $(x, y)$ (which will be over the diagonal) will represent a class which is born at time $x$ and vanishes at time $y$.

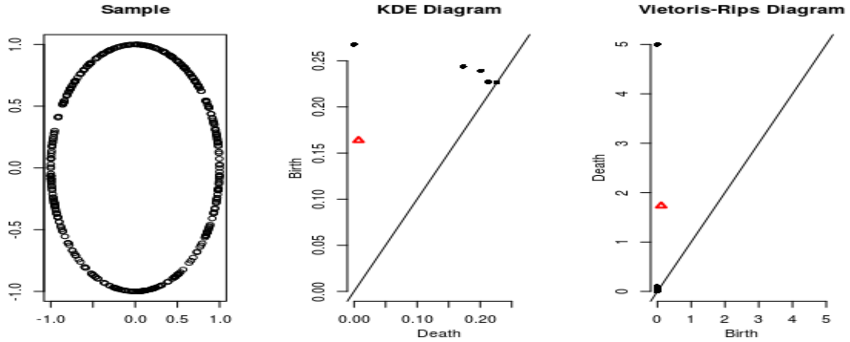
## Choosing the options

We will have to choose which of the previous filtrations do we use, which implies that we will need to add some parameters:

- If we choose Čech or Vietoris-Rips filtrations we will have to tell till which dimension we want to compute the homology.
- If we choose methods using the DTM or KDE functions we will have to choose the bounding box and the grid we want to work with.

# Persistence diagram of a circle

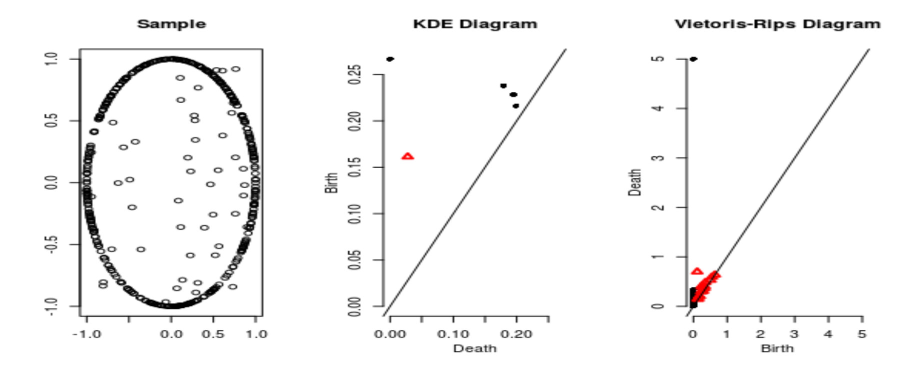
## Example (400 points in a circle)




Both diagrams contain the information, but VR took longer to be computed.

# Persistence diagram of a circle with noise

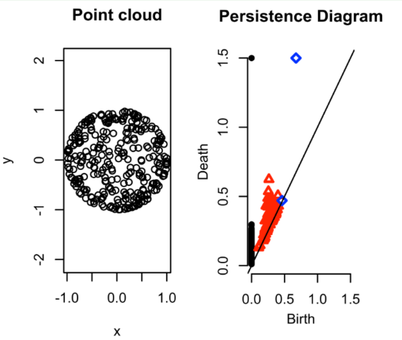
Example (400 points in a circle with noise)




KDE diagram contain the information, while VR is not so clear. Moreover VR took longer to be computed.

# Persistence diagrams: more examples

## Example (2d-sphere)




## Example (2d-torus)