Introduction to simplicial homology

Ana Romero
University of La Rioja (Spain)

Introduction to Topological Data Analysis course
8th February 2024

Lecture 0: Introduction to simplicial homology
Necessary tools to work in Topological Data Analysis

Definitions and ideas used for more than 100 years, from the beginning of algebraic topology

I will present the ideas from scratch and very slowly, so that they can be understood by the whole audience of the course (not only mathematicians)

Contents

Part I:

Simplicial complexes

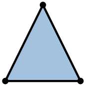
Simplicial complexes

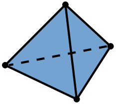
Simplicial complexes are a model for (topological) spaces made of simplices

Simplices can be seen as “triangles” in different dimensions

0-simplex: a point (called a vertex) •

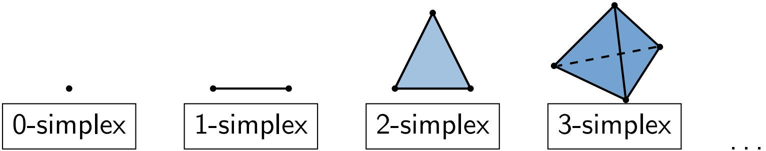
1-simplex: a line segment (called an edge) •——•

2-simplex: a triangle  (filled)

3-simplex: a tetrahedron  (solid)

...

Simplices



A k-simplex σ is made by $(k+1)$ vertices and all the space between them (the convex hull). The number k is the **dimension** of the simplex

The boundary of a k-simplex is made of k+1 simplices of dimension $k-1$, which are the $(k-1)$-**faces**

There also faces of dimension 0, 1, 2, $\ldots$, made of the (convex hull) of subsets of the vertices of σ

Faces of a 2-simplex:

- 1-faces: edges
- 0-faces: vertices

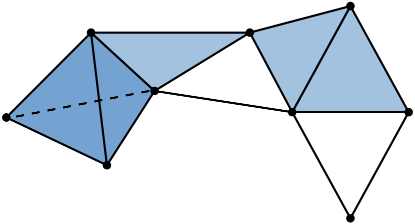
Faces of a 3-simplex:

- 2-faces: triangles
- 1-faces: edges
- 0-faces: vertices

Simplicial complexes

A **simplicial complex** is a set of simplices glued by their faces

Example:

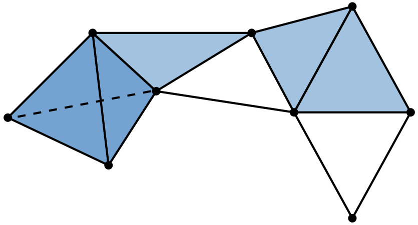


A **simplicial complex** is a set of simplices such that the intersection of any two is either

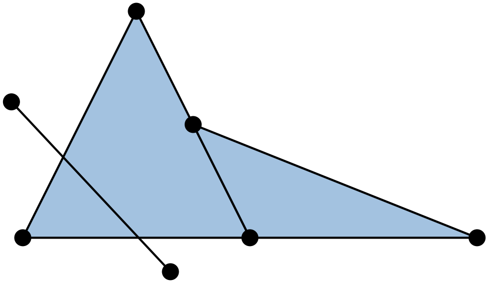
(i) empty, or

(ii) a single simplex, that is a face of both simplices

Example:



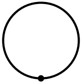
Not everything is allowed

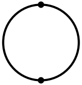


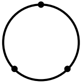
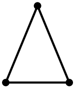
It is not a simplicial complex

Examples of simplicial complexes

How can we construct a circle?

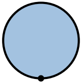
 is not valid

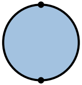
 is not valid

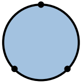
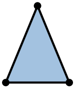
 is a simplicial complex $\equiv$ 

Examples of simplicial complexes

How can we construct a disk?

 is not valid

 is not valid

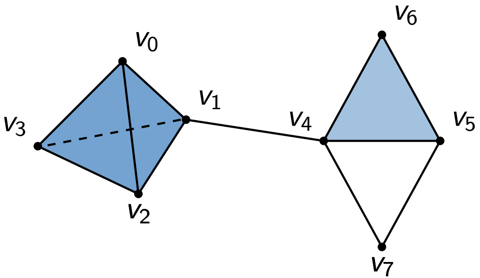
 is a simplicial complex $\equiv$ 

Abstract simplicial complexes

An **abstract simplicial complex** is a non-empty family of sets (called **simplices**) that is closed under taking subsets, i.e., every non-empty subset of a set in the family is also in the family:

A family (of sets) X is an abstract simplicial complex if for every set $Y_1 \in X$ and $Y_2 \subset Y_1$ and $Y_2 \neq \emptyset$ then $Y_2 \in X$.

Example:



$$X = [[v_0], [v_1], [v_2], [v_3], [v_4], [v_5], [v_6], [v_7], [v_0, v_1], [v_0, v_2], [v_0, v_3], [v_1, v_2], [v_1, v_3], [v_2, v_3], [v_1, v_4], [v_4, v_5], [v_4, v_6], [v_4, v_7], [v_5, v_6], [v_5, v_7], [v_0, v_1, v_2], [v_0, v_1, v_3], [v_0, v_2, v_3], [v_1, v_2, v_3], [v_4, v_5, v_6], [v_0, v_1, v_2, v_3]]$$

Abstract simplicial complexes

The **faces** of a simplex σ are other simplices τ such that the vertices of τ are also vertices of σ

Faces of the 3-simplex $[v_0, v_1, v_2, v_3]$:

- 2-faces: triangles $[v_1, v_2, v_3], [v_0, v_2, v_3], [v_0, v_1, v_3], [v_0, v_1, v_2]$
- 1-faces: edges $[v_2, v_3], [v_1, v_3], [v_1, v_2], [v_0, v_3], [v_0, v2], [v_0, v_1]$
- 0-faces: vertices $[v_0], [v_1], [v_2], [v_3]$

We denote the faces of dimension $k-1$ of a k-simplex as:

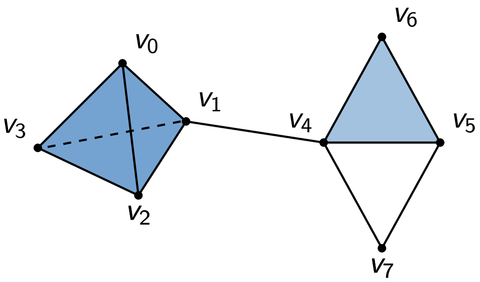
$$\partial_i([a_0, a_1, \ldots, a_k]) = [a_0, a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_k] = [a_0, a_1, \ldots, \hat{a}_i, \ldots, a_k]$$

(supposing that we have an order for the vertices)

The **facets** of an abstract simplicial complex are the maximal simplices, that is, simplices that are not face of another simplex

Abstract simplicial complexes

In the example



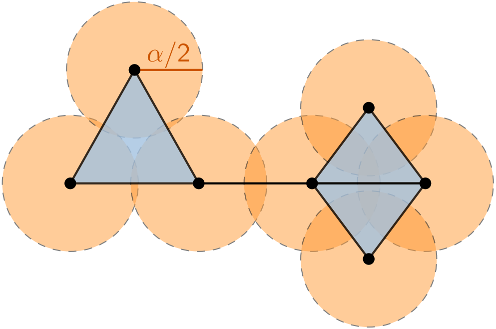
the facets are $[v_0, v_1, v_2, v_3]$, $[v_4, v_5, v_6]$, $[v_1, v_4]$, $[v_4, v_7]$ and $[v_5, v_7]$

An abstract simplicial complex can be geometrically realized into a simplicial complex. A simplicial complex can also be thought of as an abstract simplicial complex by giving names to the vertices. In this lecture, we will talk indistinctly about simplicial complexes and abstract simplicial complexes

Simplicial complexes from data clouds

Given a set of points P in a metric space (usually $\mathbb{R}^n$) and a real number $\alpha \geq 0$:

- The **Vietoris-Rips** complex $VR_\alpha(X)$ is the set of simplices $[x_0, ..., x_k]$ with $x_i \in P$ such that $d_X(x_i, x_j) \leq \alpha$ for all (i, j).

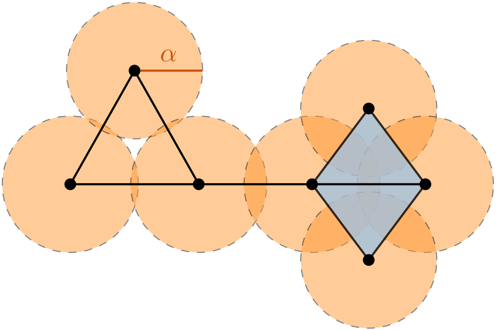


It follows from the definition that this is an abstract simplicial complex. If we change the parameter α, the VR complex can change by adding or removing some simplices.

Simplicial complexes from data clouds

Given a set of points P in a metric space (usually $\mathbb{R}^n$) and a real number $\alpha \geq 0$:

- The **Čech complex** $Cech_\alpha(X)$ is defined as the set of simplices $[x_0, ..., x_k]$ with $x_i \in P$ such that the $k+1$ closed balls $B(x_i, \alpha)$ have a non-empty intersection.



It follows from the definition that this is an abstract simplicial complex. If we change the parameter α, the Čech complex can change by adding or removing some simplices.

Simplicial complexes from data clouds

Notice that these two complexes are related:

$$VR_{\alpha}(X) \subseteq Cech_{\alpha}(X) \subseteq VR_{2\alpha}(X)$$

The following relations will be important in the next lectures:

$$VR_{\alpha}(X) \subseteq VR_{\alpha'}(X) \quad \text{for every } \alpha \leq \alpha'$$

$$Cech_{\alpha}(X) \subseteq Cech_{\alpha'}(X) \quad \text{for every } \alpha \leq \alpha'$$

Euler characteristic

The **Euler Characteristic** is a **topological invariant**, that is, a number that we assign to a shape, or a simplicial complex in our case, to learn something about its global structure.

Euler's initial observation: for any three-dimensional convex polyhedron, the number of vertices minus the number of edges plus the number of faces is always equal to 2. This number is called the Euler characteristic

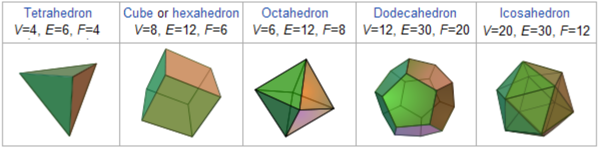


Image obtained from `https://mathstrek.blog/2013/12/02/from-euler-characteristics-to-cohomology-i/`

However, this is not true for other solids or shapes that are not convex (for instance because they have holes).

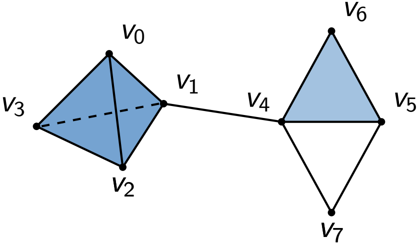
Euler characteristic

The Euler characteristic of a simplicial complex is defined as the alternating sum of the number of *n*-simplices:

$$k_0 - k_1 + k_2 - k_3 + \cdots$$

where k_i is the number of i-dimensional simplices.

Euler characteristic of our example: 8 - 12 + 5 - 1 = 0



As we will see later, the Euler characteristic can also be determined by using the notion of homology

Simplicial sets

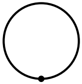
A **simplicial set** K consists of:

- a set K_n for every $n \geq 0$;
- for every pair of integers (i, n) such that $0 \leq i \leq n$, **face** and **degeneracy** operators $\partial_i : K_n \to K_{n-1}$ and $s_i : K_n \to K_{n+1}$ that satisfy the **simplicial identities:**

$$\begin{aligned}
\partial_i \partial_j &= \partial_{j-1} \partial_i && \text{if } i < j \\
s_i s_j &= s_{j+1} s_i && \text{if } i \leq j \\
\partial_i s_j &= s_{j-1} \partial_i && \text{if } i < j \\
\partial_i s_j &= \mathrm{Id} && \text{if } i = j, j+1 \\
\partial_i s_j &= s_j \partial_{i-1} && \text{if } i > j+1
\end{aligned}$$

It is a generalization of (abstract) simplicial complex

This construction allows us to represent spaces with less simplices. For instance, a circle can be represented as:



This is not a simplicial complex, but it is a simplicial set.

In general, we can represent the sphere S^n with a simplicial set with only two simplices: one vertex and one simplex of dimension n.

Part II:

Homology

Groups

A **group** is a set G with a binary operation $\cdot$ that combines two elements to yield another one, such that:

1. The set is **closed** under the operation: if $a, b \in G$ then $a \cdot b \in G$.
2. The operation is **associative**: $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for $a, b, c \in G$.
3. The set has an **identity element** under the operation that is also an element of the set: there exists $e \in G$ such that $e \cdot a = a \cdot e = a$ for all $a \in G$.
4. Every element of the set has an **inverse** under the operation that is also an element of the set: for all $a \in G$, there exists $a^{-1} \in G$ such that $a \cdot a^{-1} = e = a^{-1} \cdot a$.

The operation is not required to be commutative, that is, in general $a \cdot b$ is not equal to $b \cdot a$. If the operation is commutative, then the group is called **abelian**

Groups

Examples (and non-examples) of groups:

- The set of integers $\mathbb{Z}$ with the addition is a group
- The set of integers under subtraction is not a group
- The set of natural numbers under addition is not a group
- The set of linear combinations of a set of n generators $< a_1, \ldots, a_n >$ is a group with the formal addition of the coefficients. Example of element of the group: $3 * a_1 - 2 * a_2 + 7 * a_3$. Example of addition of two elements:
  $(3 * a_1 - 2 * a_2 + 7 * a_3) + (-1 * a_1 + 1 * a_3) = 2 * a_1 - 2 * a_2 + 8 * a_3$.
  This group is called the free abelian group generated by $< a_1, \ldots, a_n >$ and can be seen as $\mathbb{Z}^n$.

  It is also possible to work with coefficients over $\mathbb{Z}_2 \equiv \mathbb{Z}/2\mathbb{Z} = \{0, 1\}$

Chain complexes

Let A and B groups. A function $f : A \to B$ is called a **homomorphism** of groups if it is compatible with the group operation: $f(a \cdot b) = f(a) \cdot f(b)$ for all $a, b \in A$

A **chain complex** C_* is a sequence of groups and homomorphisms of groups

$$\cdots \leftarrow C_{n-1} \xleftarrow{d_n} C_n \xleftarrow{d_{n+1}} C_{n+1} \leftarrow \cdots$$

such that $d_n \circ d_{n+1} = 0$ for all $n \geq 0$

The homomorphisms d_n are called **differential or boundary maps**

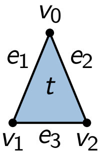
Chain complex of a simplicial complex

We consider a simplicial complex X where the simplices are oriented (for instance, choosing an order for the vertices).

We define a **chain complex** $C_*(X) = (C_n(X), d_n)$, where:

- $C_n(X)$ is the free abelian group generated by the set of n-simplices,

Example:



$C_0(X)$	$C_1(X)$	$C_2(X)$	$0 \cdots$
$\mathbb{Z}^3$	$\mathbb{Z}^3$	$\mathbb{Z}$	
$< v_0, v_1, v_2 >$	$< e_1, e_2, e_3 >$	$< t >$	

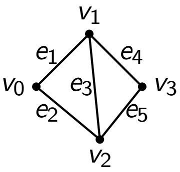
Chain complex of a simplicial complex

We consider a simplicial complex X where the simplices are oriented (for instance, choosing an order for the vertices).

We define a **chain complex** $C_*(X) = (C_n(X), d_n)$, where:

- $C_n(X)$ is the free abelian group generated by the set of n-simplices,

Example:



$C_0(X)$	$C_1(X)$	$0 \cdots$
$\mathbb{Z}^4$	$\mathbb{Z}^5$	
$< v_0, v_1, v_2, v_3 >$	$< e_1, e_2, e_3, e_4, e_5 >$	

Chain complex of a simplicial complex

We consider a simplicial complex X where the simplices are oriented (for instance, choosing an order for the vertices).

We define a **chain complex** $C_*(X) = (C_n(X), d_n)$, where:

- $C_n(X)$ is the free abelian group generated by the set of n-simplices,
- and the differential or boundary map $d_n : C_n(X) \to C_{n-1}(X)$ is defined on the generators as the alternating sum $d_n = \sum_{i=0}^{n} (-1)^i \partial_i$.

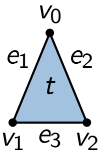
We recall $\partial_i([a_0, a_1, \dots, a_k]) = [a_0, a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_k]$
$= [a_0, a_1, \dots, \hat{a}_i, \dots, a_k]$

With coefficients over $\mathbb{Z}_2 \equiv \mathbb{Z}/2\mathbb{Z}$, the signs of the differential map can be ignored

Chain complex of a simplicial complex

The differential or boundary map $d_n : C_n(X) \to C_{n-1}(X)$ is defined on the generators as the alternating sum $d_n = \sum_{i=0}^{n} (-1)^i \partial_i$.

Example:



$$
\begin{array}{ccccccc}
C_0(X) & \xleftarrow{d_1} & C_1(X) & \xleftarrow{d_2} & C_2(X) & \xleftarrow{d_3} & 0 \\
\mathbb{Z}^3 & & \mathbb{Z}^3 & & \mathbb{Z} & & \\
< v_0, v_1, v_2 > & & < e_1, e_2, e_3 > & & < t > & &
\end{array}
$$

$$d_1(e_1) = d_1([v_0, v_1]) = \partial_0(e_1) - \partial_1(e_1) = v_1 - v_0$$

$$d_1(e_2) = d_1([v_0, v_2]) = v_2 - v_0; \quad d_1(e_3) = d_1([v_1, v_2]) = v_2 - v_1$$

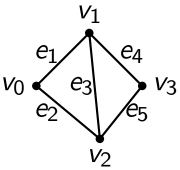
$$d_2(t) = d_2([v_0, v_1, v_2]) = \partial_0(t) - \partial_1(t) + \partial_2(t) = e_3 - e_2 + e_1$$

$$d_1 = \begin{bmatrix} -1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \end{bmatrix}; \quad d_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}; \quad d_3 = 0$$

Chain complex of a simplicial complex

The differential or boundary map $d_n : C_n(X) \to C_{n-1}(X)$ is defined on the generators as the alternating sum $d_n = \sum_{i=0}^{n} (-1)^i \partial_i$.

Example:



$$\begin{array}{ccccc} C_0(X) & \xleftarrow{d_1} & C_1(X) & \xleftarrow{d_2} & 0 \\ \mathbb{Z}^4 & & \mathbb{Z}^5 & & \\ < v_0, v_1, v_2, v_3 > & & < e_1, e_2, e_3, e_4, e_5 > & & \end{array}$$

$$d_1(e_1) = v_1 - v_0; \quad d_1(e_2) = v_2 - v_0; \quad d_1(e_3) = v_2 - v_1;$$

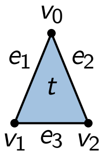
$$d_1(e_4) = d_1([v_1, v_3]) = v_3 - v_1; \quad d_1(e_5) = d_1([v_2, v_3]) = v_3 - v_2$$

$$d_1 = \begin{bmatrix} -1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}; \quad d_2 = 0$$

Chain complex of a simplicial complex

We observe that $d_n \circ d_{n+1} = 0$:

For instance, in the disk:



$$
\begin{aligned}
d_2(t) &= d_2([v_0, v_1, v_2]) = e_3 - e_2 + e_1 \\
d_1(e_3 - e_2 + e_1) &= d_1(e_3) - d_1(e_2) + d_1(e_1) \\
&= v_2 - v_1 - v_2 + v_0 + v_1 - v_0 = 0
\end{aligned}
$$

Kernels, images and quotients

The **kernel** of a homomorphism of groups $f : A \to B$ is the set of all elements that are mapped to the zero element:

$$\mathrm{Ker}(f) = \{a \in A | f(a) = 0\} \subseteq A$$

The **image** of f is the set of the outputs:

$$\mathrm{Im}(f) = \{f(a) | a \in A\} \subseteq B$$

Let G be an abelian group and N a subgroup of G. Then the **quotient group** is defined as

$$G/N = \{gN | g \in G\}.$$

Intuitively, G/N consists of all elements in G that are not in N.

In a chain complex

$$C_* : \qquad \cdots \leftarrow C_{n-1} \xleftarrow{d_n} C_n \xleftarrow{d_{n+1}} C_{n+1} \leftarrow \cdots$$

Cycle group: $Z_n = \mathrm{Ker}(d_n) \subseteq C_n$

Boundary group: $B_n = \mathrm{Im}(d_{n+1}) \subseteq C_n$

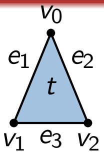
Since $d_n \circ d_{n+1} = 0$, one has $B_n \subseteq Z_n$. Then, the ***n*-th homology group of** C_* is defined as the quotient group:

$$H_n(C) = Z_n / B_n = \mathrm{Ker}(d_n) / \mathrm{Im}(d_{n+1})$$

When working with coefficients over a field (for instance, $\mathbb{Z}_2$), these groups are vector spaces and their ranks are called the **Betti numbers** of the chain complex (or the simplicial complex):

$$\beta_n = \mathrm{rank}(H_n(C))$$

Homology



$$C_0(X) \xleftarrow{d_1} C_1(X) \xleftarrow{d_2} C_2(X)$$

$$\mathbb{Z}^3 \qquad \mathbb{Z}^3 \qquad \mathbb{Z}$$

$$< v_0, v_1, v_2 > \qquad < e_1, e_2, e_3 > \qquad < t >$$

$$d_1(e_1) = v_1 - v_0; \quad d_1(e_2) = v_2 - v_0; \quad d_1(e_3) = v_2 - v_1$$

$$d_2(t) = e_3 - e_2 + e_1$$

$$\mathrm{Ker}(d_0) = C_0(X) =< v_0, v_1, v_2 >= \mathbb{Z}^3$$

$$\mathrm{Im}(d_1) =< v_1 - v_0, v_2 - v_0, v_2 - v_1 >=< v_1 - v_0, v_2 - v_0 >= \mathbb{Z}^2$$

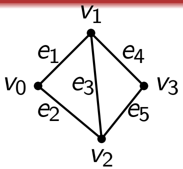
$$H_0(X) =< v_0, v_1, v_2 > / < v_1 - v_0, v_2 - v_0 >= \mathbb{Z}$$

$$\mathrm{Ker}(d_1) =< e_3 - e_2 + e_1 >= \mathbb{Z}; \quad \mathrm{Im}(d_2) =< e_3 - e_2 + e_1 >= \mathbb{Z}$$

$$H_1(X) =< e_3 - e_2 + e_1 > / < e_3 - e_2 + e_1 >= 0$$

With coefficients over $\mathbb{Z}_2 \equiv \mathbb{Z}/2\mathbb{Z}$, we obtain $H_0(X) = \mathbb{Z}/2\mathbb{Z}$ and $H_1(X) = 0$. Hence, $\beta_0 = \mathrm{rank}(H_0(X)) = 1$ and $\beta_1 = \mathrm{rank}(H_1(X)) = 0$

Homology



$$C_0(X) \xleftarrow{d_1} C_1(X)$$

$$\mathbb{Z}^4 \qquad \mathbb{Z}^5$$

$$< v_0, v_1, v_2, v_3 > \qquad < e_1, e_2, e_3, e_4, e_5 >$$

$$d_1(e_1) = v_1 - v_0; \quad d_1(e_2) = v_2 - v_0; \quad d_1(e_3) = v_2 - v_1;$$
$$d_1(e_4) = v_3 - v_1; \quad d_1(e_5) = v_3 - v_2$$

$$\mathrm{Ker}(d_0) =< v_0, v_1, v_2, v_3 >= \mathbb{Z}^4$$
$$\mathrm{Im}(d_1) =< v_1 - v_0, v_2 - v_0, v_2 - v_1, v_3 - v_1, v_3 - v_2 >$$
$$=< v_1 - v_0, v_2 - v_0, v_3 - v_1 >= \mathbb{Z}^3; \quad H_0(X) = \mathbb{Z}$$
$$\mathrm{Ker}(d_1) =< e_3 - e_2 + e_1, e_5 - e_4 + e_3 >; \quad \mathrm{Im}(d_2) = 0$$
$$H_1(X) =< e_3 - e_2 + e_1, e_5 - e_4 + e_3 > /0 = \mathbb{Z}^2$$

With coefficients over $\mathbb{Z}_2 \equiv \mathbb{Z}/2\mathbb{Z}$, we obtain $H_0(X) = \mathbb{Z}/2\mathbb{Z}$ and $H_1(X) = (\mathbb{Z}/2\mathbb{Z})^2$. Hence, $\beta_0 = \mathrm{rank}(H_0(X)) = 1$ and $\beta_1 = \mathrm{rank}(H_1(X)) = 2$

Homology computation

Homology groups (and Betti numbers) can be determined by means of operations on the differential matrices

A technique that can be used is called **Smith Normal Form**, that is a particular type of diagonalization of matrices (similar to Gaussian elimination)

When working over a field, the Betti numbers can be determined by computing the ranks of Smith Normal Forms of each matrix:

- $\text{rank}(Z_n)$ is the number of zero columns in the SNF of d_n
- $\text{rank}(B_n)$ is the number of non-zero rows in the SNF of d_{n+1}
- β_n is the difference of these ranks

What does homology (and Betti numbers) measure?

- H_0 : number of connected components
- H_1 : number of cycles
- H_2 : number of voids
- H_n : number of n-dimensional holes

Euler characteristic (revisited)

We have seen that the Euler characteristic of a simplicial complex is defined as the alternating sum of the number of n-simplices:

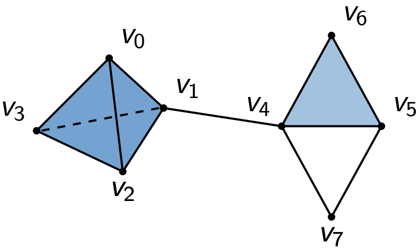
$$k_0 - k_1 + k_2 - k_3 + \cdots$$

where k_i is the number of i-dimensional simplices.

It can also be determined as:

$$\beta_0 - \beta_1 + \beta_2 - \beta_3 + \cdots$$

Euler characteristic of our example: 8 - 12 + 5 - 1 = 1 -1 = 0



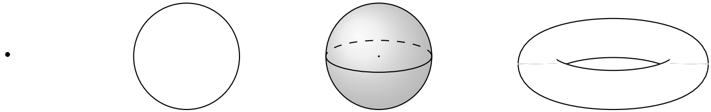
Part III:

Introduction to TDA

Topological data analysis and Betti numbers

Topological Data Analysis (TDA) consists in applying techniques from topology and algebra to the analysis of data, studying the **shape** of the data

This can be done by using homology groups and Betti numbers



Space	β_0	β_1	β_2
Point	1	0	0
Circle	1	1	0
Sphere	1	0	1
Torus	1	2	1

Persistent homology

From point clouds (or data sets), we can first construct one (or several) simplicial complexes by means of the Vietoris-Rips or Čech constructions, and then compute their Betti numbers

Problem of this approach: in the VR or Čech complexes, we can have many small cycles or voids that are not representative of the data. To solve this problem, the notion of **persistent homology** is used

Idea of persistent homology: construct the VR or Čech complexes for different parameters α and **track** the Betti numbers in each of them

Thank you!