Extracting Geometrical Features From Data

Topological Data Analysis

Ulderico Fugacci

CNR - IMATI

Topological Data Analysis

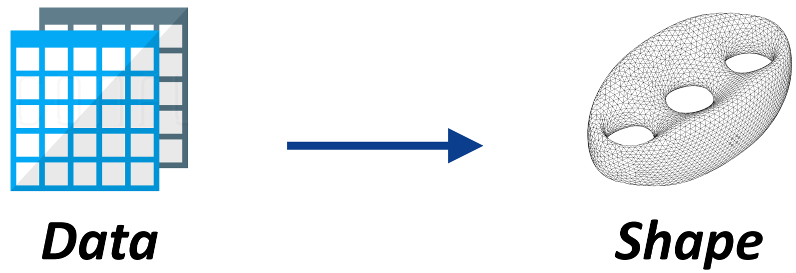
Outline:

Simplicial Complexes

Complexes & Data

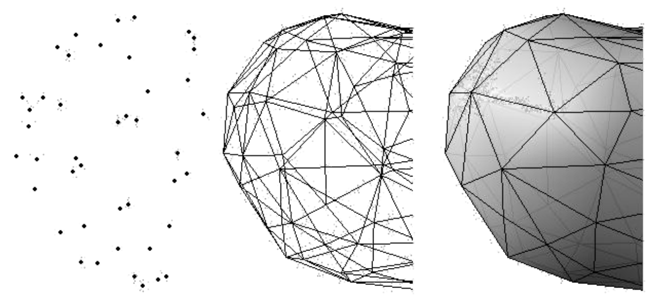
Goal:

We want to associate a topological structure to a given dataset



Due to the nature of data and to our computational ambitions, datasets will be represented by ***"discrete"*** structures

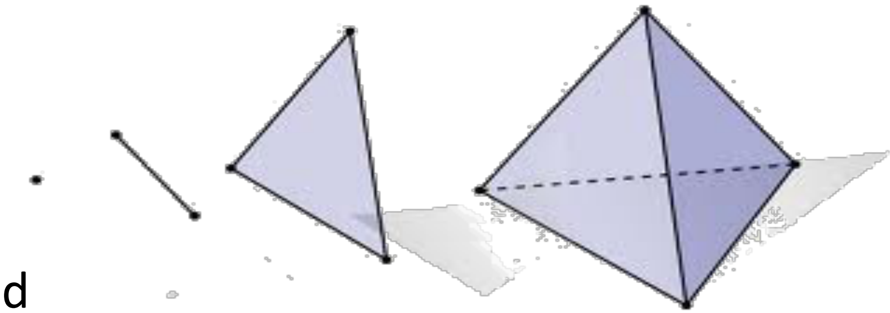
Among various possibilities, ***simplicial complexes*** represent the most suitable choice



In fact, simplicial complexes are able to deal with data:

- of ***large size*** (e.g. consisting of a huge number of samples)
- of ***high dimension*** (e.g. involving a large number of variables or parameters)
- ***unorganized*** (e.g. not arranged in a regular grid)

Simplicial Complexes



Definitions:

A set V := { v_0, v_1, ..., v_k } of points in $\mathbb{R}^n$ is called
geometrically independent if vectors $v_1 - v_0$, ..., $v_k - v_0$ are ***linearly independent*** over $\mathbb{R}$

E.g. two distinct points, three non-collinear points, four non-coplanar points

The ***k-simplex*** $\sigma = v_0 v_1 \dots v_k$ spanned by a geometrically independent set V = {v_0,v_1,...,v_k} of in $\mathbb{R}^n$ is the ***convex hull*** of V, i.e. the set of all points $x \in \mathbb{R}^n$ such that

$$x = \sum_{i=0}^{k} t_i v_i \quad \text{where} \quad \sum_{i=0}^{k} t_i = 1 \quad \text{and } t_i \geq 0 \text{ for all } i$$

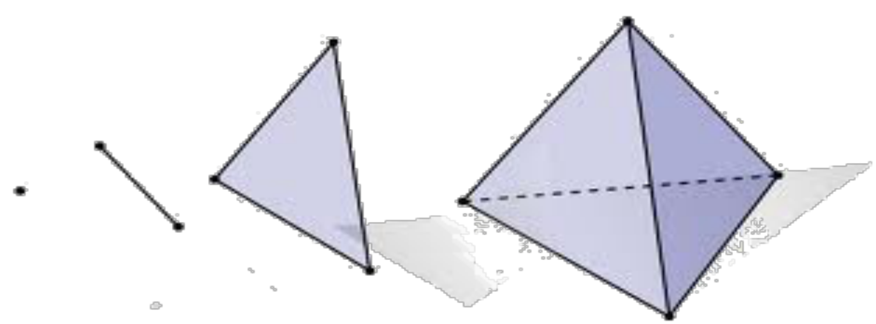
The numbers t_i are uniquely determined by x and are called ***barycentric coordinates*** of x

E.g. a 0-simplex is a vertex, a 1-simplex is an edge, a 2-simplex is a triangle, a 3-simplex is a tetrahedron

Simplicial Complexes

Definitions:

- The points $v_0, v_1, \ldots, v_k$ spanning a k-simplex σ are called the ***vertices*** of σ
- k is called the ***dimension*** of σ and denoted as dim(σ)
- Any simplex τ spanned by a non-empty subset of V is called a ***face*** of σ
- Conversely, σ is called a ***coface*** of τ

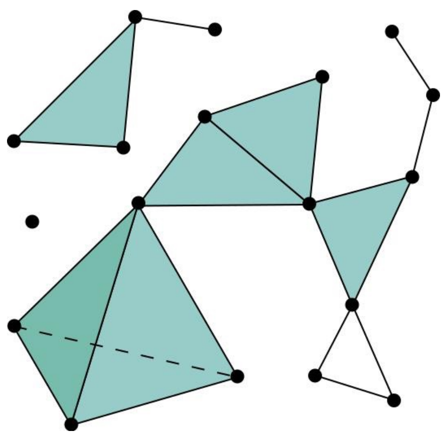


Simplicial Complexes

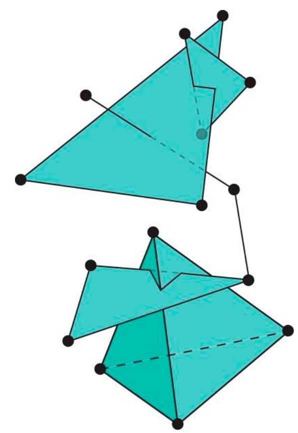
Definition:

A *(geometric)* ***simplicial complex*** K in $\mathbb{R}^n$ is a collection of simplices in $\mathbb{R}^n$ such that

- *Every face of a simplex of K is in K*
- *The non-empty intersection of any two simplices of K is a face of each of them*



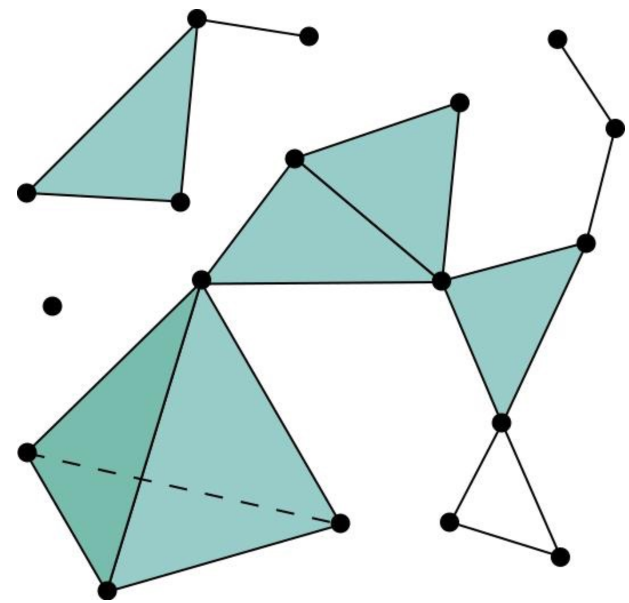
simplicial complex



non-simplicial complex

Simplicial Complexes

Definitions:



Given a (geometric) simplicial complex K in $\mathbb{R}^n$,

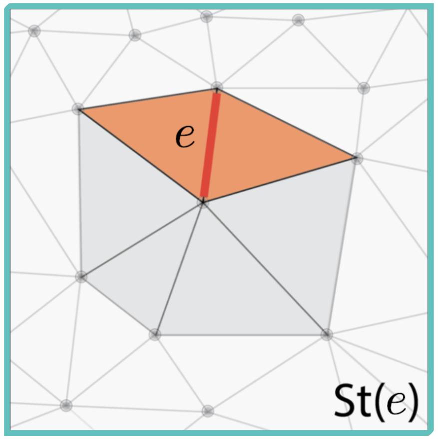
- The ***dimension*** of a simplicial complex K in $\mathbb{R}^n$, denoted as dim(K), is the supremum of the dimensions of the simplices of K
- A simplex σ of K such that dim(σ) = dim(K) is called ***maximal***
- A simplex σ of K which is not a proper face of any simplex of K is called ***top***
- A subcollection of K that is itself a simplicial complex is called a ***subcomplex*** of K

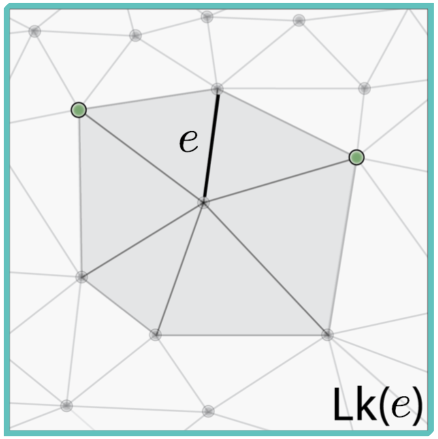
Simplicial Complexes

Definitions:

Given a simplex σ of a (geometric) simplicial complex K in $\mathbb{R}^n$,

- The ***star*** of σ is the set ***St(σ)*** of the cofaces of σ
- The ***link*** of σ is the set ***Lk(σ)*** of the faces of the simplices in St(σ) such that do not intersect σ





Simplicial Complexes

Given a (geometric) simplicial complex K in $\mathbb{R}^n$,

its ***polytope*** |K| is the subset of $\mathbb{R}^n$ defined as the union of the simplices of K

The polytope |K| can be endowed with ***two possible topologies*** T_1 and T_2:

- ***T_1***: A subset F of |K| is a closed set of $(|K|, T_1)$ if and only if $F \cap \sigma$ is a closed set of (σ, T_σ) for each σ in K where T_σ is the subspace topology induced on σ by $\mathbb{E}^n$
- ***T_2***: The subspace topology induced on |K| by $\mathbb{E}^n$

In general, the two topologies T_1, T_2 are ***different***, but

Proposition: *If K is a **finite** simplicial complex, $T_1 = T_2$*

From now on, if not differently specified, we consider only *finite* simplicial complexes

Simplicial Complexes

Proposition:

Given a simplicial complex K and a topological space (X, T), a function f from (|K|, T_1) to (X, T) is ***continuous*** *if and only if $f|_\sigma$ is continuous for each* $\boldsymbol{\sigma} \in K$

Definition:

Given two simplicial complexes K and K',

- A function f: K $\longrightarrow$ K' is called a ***simplicial map*** if for every simplex $\boldsymbol{\sigma} = v_0 v_1 \dots v_k$ in K, $f(\boldsymbol{\sigma}) = f(v_0)f(v_1)\dots f(v_k)$ is a simplex in K'
- The restriction f_V of f to the set of vertices V of K is called the ***vertex map*** of f

Simplicial Complexes

Definition:

An ***abstract simplicial complex*** K on a set V is a collection of finite non-empty subsets of V, called ***simplices***, such that if $\sigma \in K$ and $\tau \subseteq \sigma$, then $\tau \in K$

Analogously to the case of a geometric simplicial complex,

- The elements of V are called ***vertices*** of K
- The ***dimension*** of a simplex σ is one less than the number of its elements
- The supremum of the dimensions of the simplices in K is called ***dimension*** of K
- Each non-empty subset τ of a simplex $\sigma \in K$ is called a ***face*** of σ and σ is called a ***coface*** of τ

The notions of geometric simplicial complex and abstract simplicial complex are equivalent. *More properly, it is always possible,*

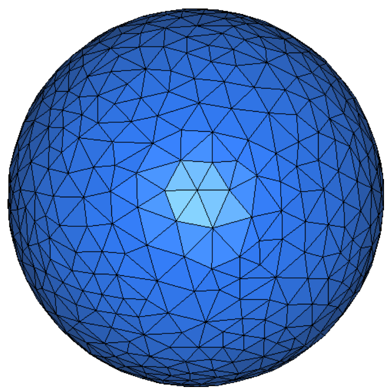
- *Given an abstract simplicial complex, to endow it with a* ***geometric realization***
- *Given a geometric simplicial complex, to* ***forget its geometry*** *thus obtaining an abstract simplicial complex*

Simplicial Complexes

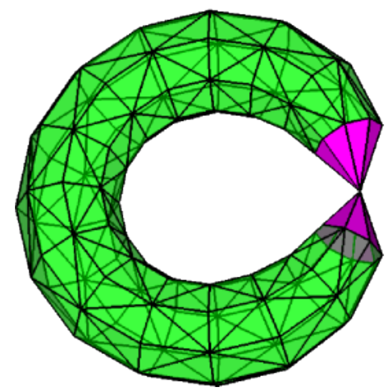
Definition: A simplicial complex K is called

- ***n-manifold [with boundary]*** if its polytope |K| is a (topological) n-manifold [with boundary]
- ***Combinatorial n-manifold [with boundary]*** if, for every vertex v, the link Lk(v) is homeomorphic to the (n − 1)-sphere S^{n-1} [or to the (n − 1)-disk $D^{n-1} := \{x \in \mathbb{R}^{n-1} : |x| \leq 1\}$]

combinatorial manifold



non-combinatorial manifold



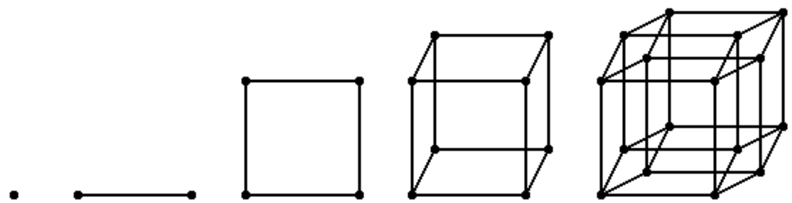
Proposition:

If K is a combinatorial n-manifold [with boundary], then K is a n-manifold [with boundary]

The converse is: ***True for n ≤ 3*** ***Open for n = 4*** ***False for n > 4***

Regular Grids

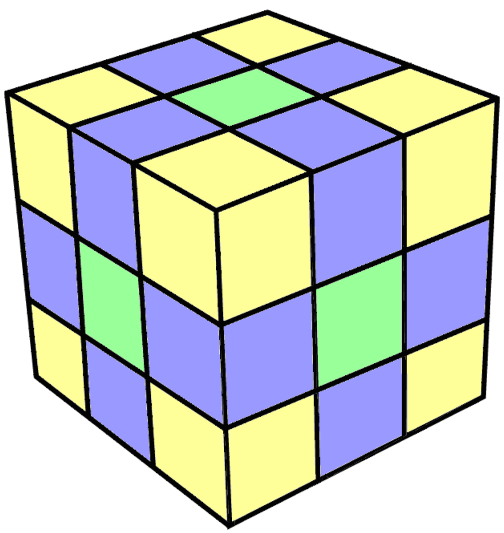
Hyper-Cube:



A ***k*-hyper-cube η** is the ***Cartesian product of k closed intervals*** of equal length

Regular Grids:

A ***regular grid H*** is a (finite) collection of hyper-cubes such that:



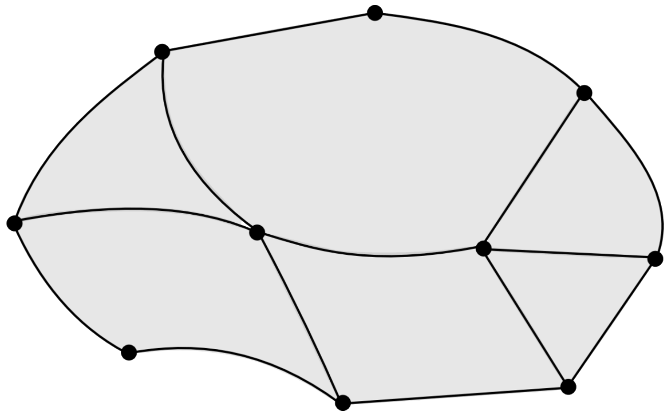
- *Each face of a hyper-cube of H is in H*
- *Each non-empty intersection of two hyper-cubes in H is a face of both*
- *The domain of H is a hyper-cube*

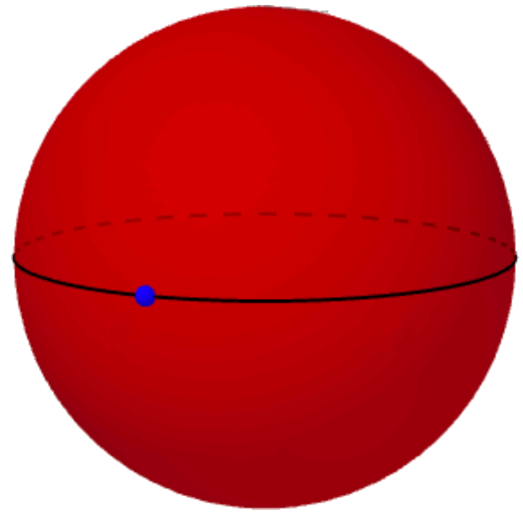
Cell Complexes

Intuitively:

Similarly to simplicial complexes and regular grids,

A ***cell complex*** Γ is a collection of cells ***"suitably glued together"***





Where a ***k-cell*** is a topological space homeomorphic to the ***k-dimensional open disk $i(D^k)$***

Bibliography

Some References:

- ***Simplicial Complexes:***
  - J. R. Munkres. ***Elements of algebraic topology***. CRC Press, 1984.

Simplicial Homology

Simplicial Homology

Given a topological space *X*, the ***homology of X*** is a ***topological invariant***

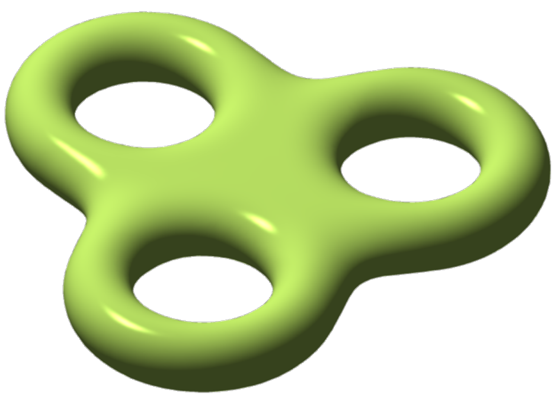
intuition ↑

detecting the "holes" of X

capturing the independent non-bounding cycles of X

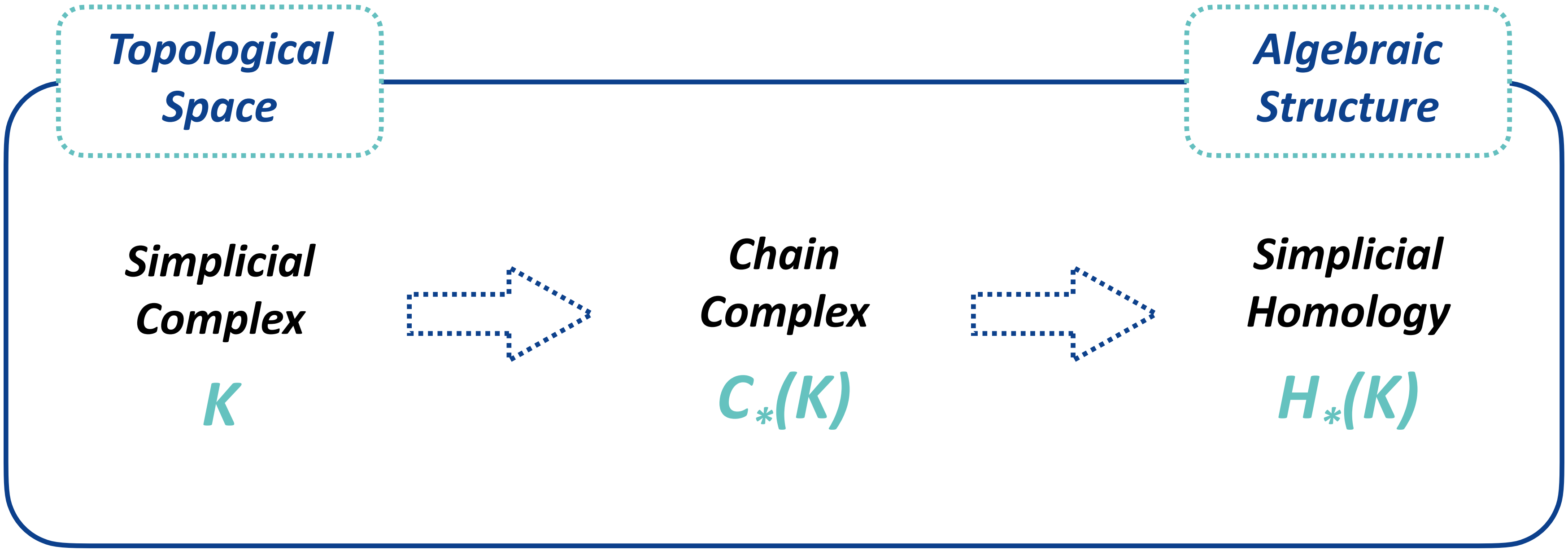
measuring how far the chain complex associated with X is from being exact

formalism ↓



$$\longmapsto H_k(X;\mathbb{Z}) \cong \begin{cases} \mathbb{Z} & \text{for } k=0 \\ \mathbb{Z}^6 & \text{for } k=1 \\ \mathbb{Z} & \text{for } k=2 \\ 0 & \text{otherwise} \end{cases}$$

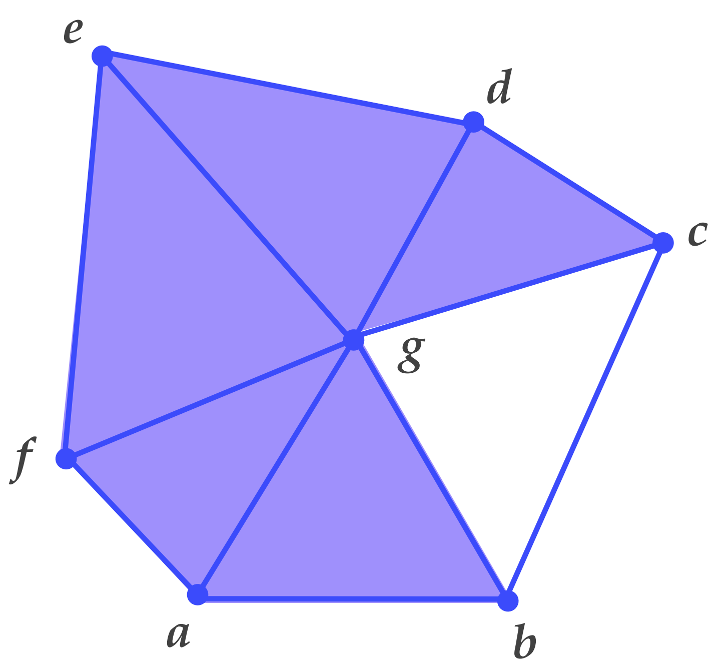
Simplicial Homology



Simplicial Homology

Given a simplicial complex K,

- a ***k-chain*** is a formal sum *(with $\mathbb{Z}_2$ coefficients)* of k-simplices of K



Examples:

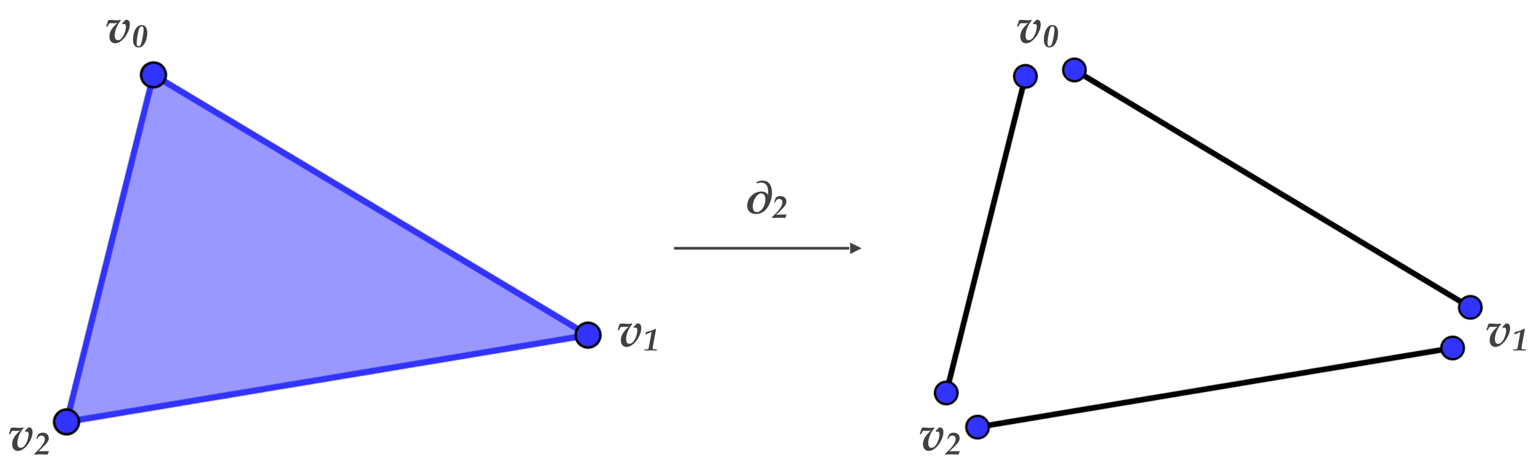
- $a + b + e$ is a 0-chain
- $fg + dg + de + eg$ is a 1-chain
- $abg + afg$ is a 2-chain

Simplicial Homology

The ***chain complex $C_*(K)$ associated with K*** consists of:

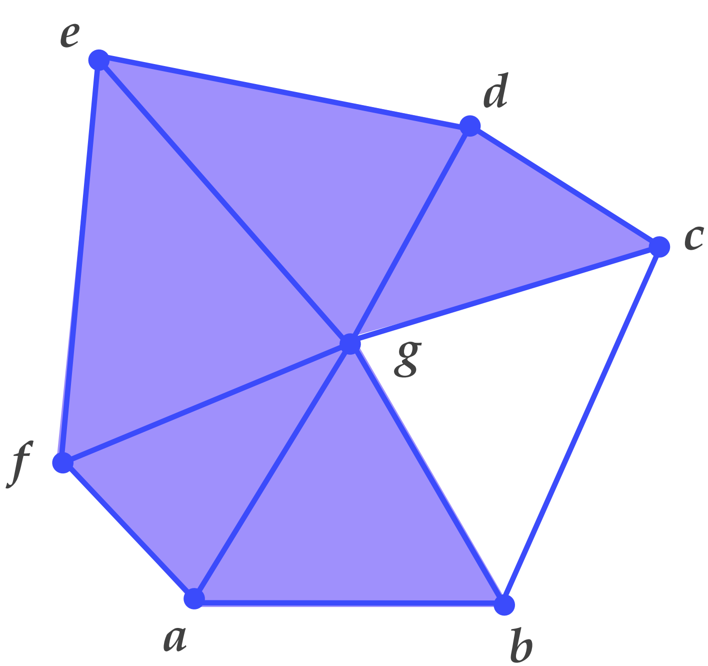
- a collection $\{C_k(K)\}_{k\in\mathbb{Z}}$ of vector spaces where $C_k(K)$ is the ***group of the k-chains*** of K
- a collection $\{\partial_k\}_{k\in\mathbb{Z}}$ of linear maps where the ***boundary map*** $\partial_k : C_k(K) \longrightarrow C_{k-1}(K)$ is defined by

$$\partial_k(v_0 \cdots v_k) := \sum_{i=0}^{k} v_0 \cdots \hat{v}_i \cdots v_k$$



Simplicial Homology

Examples:



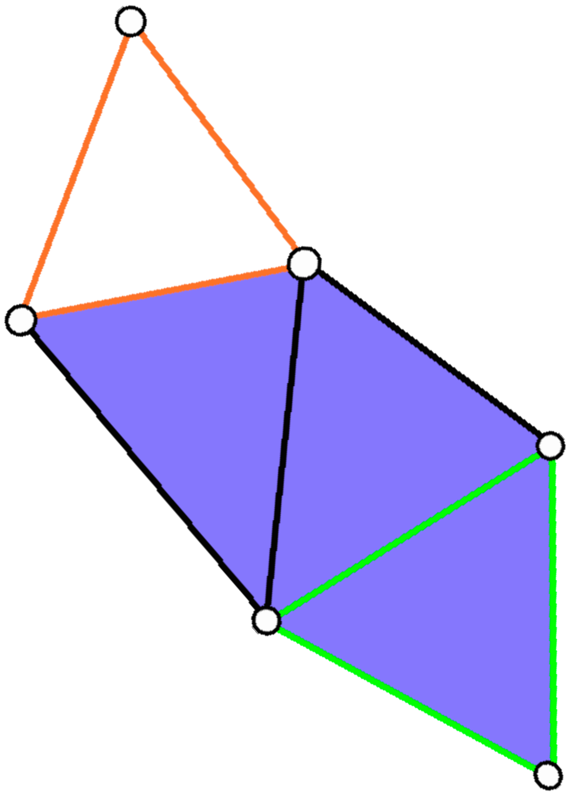
- $\partial_1(ab) = a + b$
- $\partial_1(ab + bc) = a + 2b + c = a + c$
- $\partial_2(afg + efg) = af + ag + 2fg + ef + eg =$
  $= af + ag + ef + eg$
- $\partial_1(af + ag + ef + eg) =$
  $= 2a + 2f + 2g + 2e = 0$

Simplicial Homology

Properties:

- For $k < 0$ or $k > dim(K)$, $C_k(K)$ is the ***null group***
- For $k \leq 0$ or $k > dim(K)$, ∂_k is the ***null map***
- For any $k \in \mathbb{Z}$, $\partial_k \circ \partial_{k+1} = 0$
- For any $k \in \mathbb{Z}$, $Im(\partial_{k+1}) \subseteq Ker(\partial_k)$

Simplicial Homology



Definition:

A k-chain c is called:

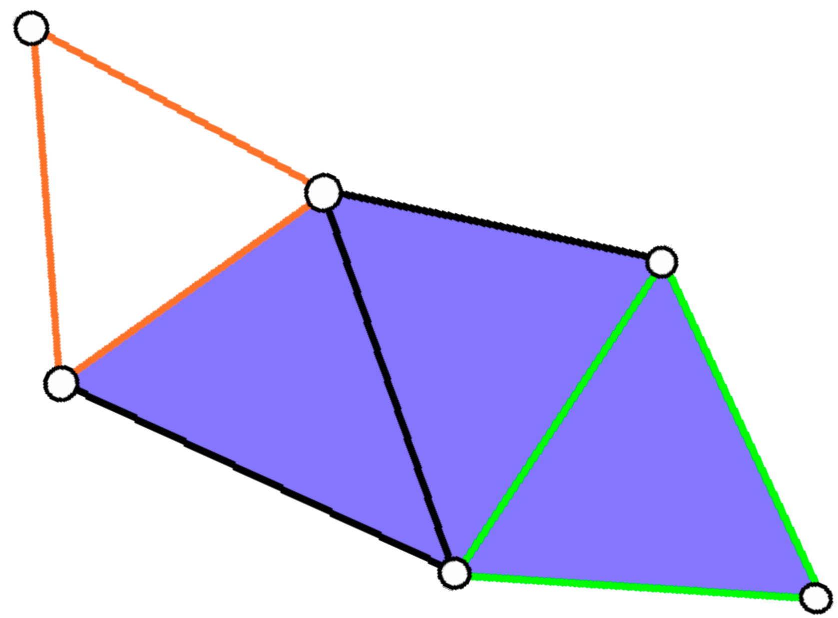
- ***k-cycle*** if $c \in Ker(\partial_k)$
- ***k-boundary*** if $c \in Im(\partial_{k+1})$

Each k-boundary is a k-cycle

Simplicial Homology

Given a simplicial complex K, the ***k-homology group*** ***$H_k(K)$*** of K is defined as

$$H_k(K) := Z_k(K)/B_k(K)$$

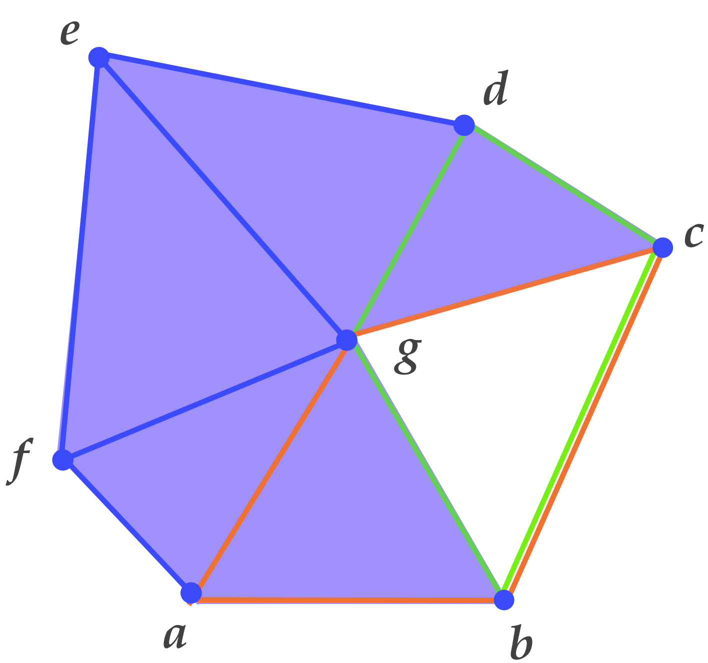


where:

- ***$Z_k(K)$*** is the ***group of k-cycles*** of K
- ***$B_k(K)$*** is the ***group of k-boundaries*** of K

Simplicial Homology

$H_k(K)$ partitions the k-cycles into equivalence classes called ***homology classes***



Definition:

Two k-cycles are said ***homologous*** if they belong to the same homology class or, equivalently, ***if their difference is a k-boundary***

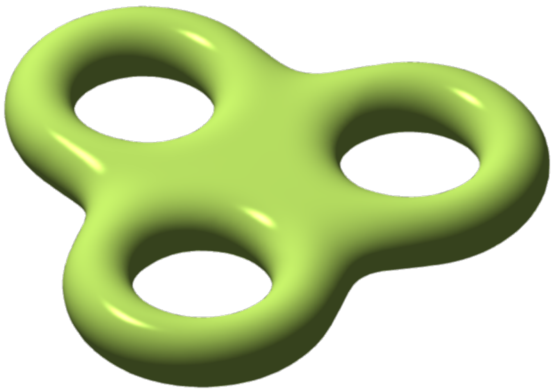
ab+ag+bc+cg is homologous to bc+bg+cd+dg

Simplicial Homology

Theorem:

Each homology group can be expressed as

$$H_k(K) \cong (\mathbb{Z}_2)^{\beta_k}$$



$$H_k(K) \cong \begin{cases} \mathbb{Z}_2 & \text{for } k = 0 \\ (\mathbb{Z}_2)^6 & \text{for } k = 1 \\ \mathbb{Z}_2 & \text{for } k = 2 \end{cases}$$

β_k is called the ***k^{th} Betti number*** of K

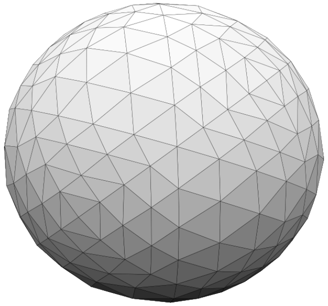
Simplicial Homology

Examples:

- ***point P***

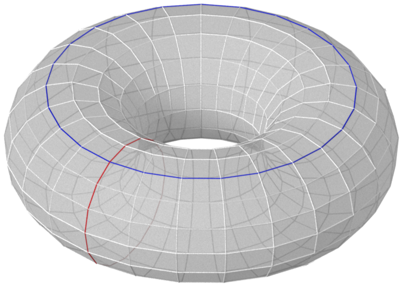
$$\beta_k(P) = \begin{cases} 1 & \text{for } k = 0 \\ 0 & \text{for } k > 0 \end{cases}$$

- ***n-dimensional sphere S^n***



$$\beta_k(S^n) = \begin{cases} 1 & \text{for } k = 0 \\ 0 & \text{for } 0 < k < n \\ 1 & \text{for } k = n \\ 0 & \text{for } k > n \end{cases}$$

- ***torus T***



$$\beta_k(T) = \begin{cases} 1 & \text{for } k = 0 \\ 2 & \text{for } k = 1 \\ 1 & \text{for } k = 2 \\ 0 & \text{for } k > 2 \end{cases}$$

Simplicial Homology

Homology groups can be defined ***in a more general way*** by choosing coefficients in $\mathbb{Z}$

Theorem:

Each homology group can be expressed as

$$H_k(K;\mathbb{Z}) \cong \mathbb{Z}^{\beta_k}\langle c_1,\dots,c_{\beta_k}\rangle \oplus \mathbb{Z}_{\lambda_1}\langle c'_1\rangle \oplus \cdots \oplus \mathbb{Z}_{\lambda_{p_k}}\langle c'_{p_k}\rangle$$

with $\lambda_{i+1} \mid \lambda_i$

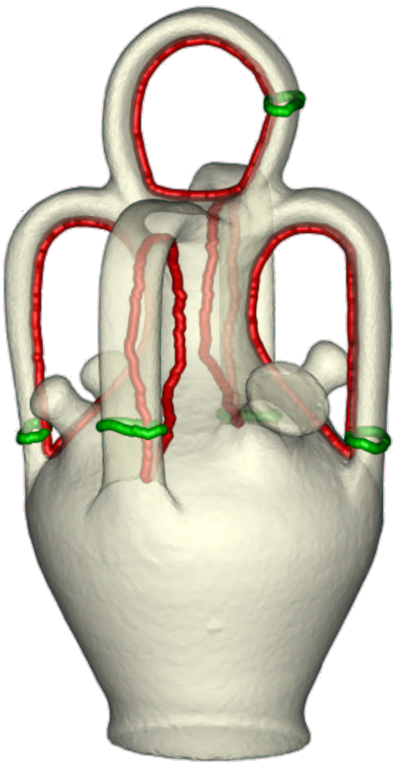


Image from [Dey et al. 2008]

We call:

- β_k, the ***kth Betti number*** of K
- $\lambda_1,\dots,\lambda_{p_k}$, the ***torsion coefficients*** of K
- $c_1,\dots,c_{\beta_k},c'_1,\dots,c'_{p_k}$, the ***homology generators*** of K

Simplicial Homology

Working with coefficients in $\mathbb{Z}$:

*Up to isomorphism, the **Betti numbers** and the **torsion coefficients** of K*
***completely characterize** the **homology groups** of K*

Working with coefficients in a field $\mathbb{F}$:

*Up to isomorphism, the **Betti numbers** of K*
***completely characterize** the **homology groups** of K*

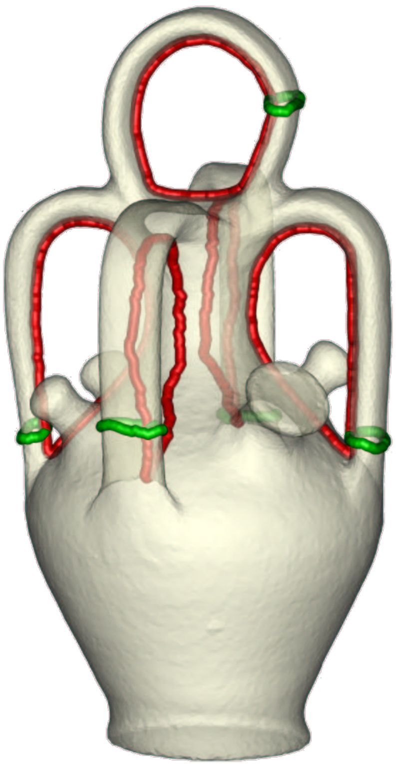
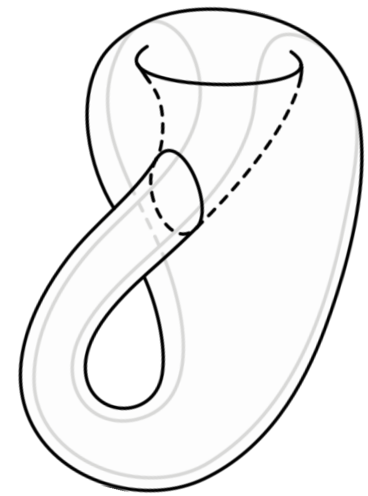


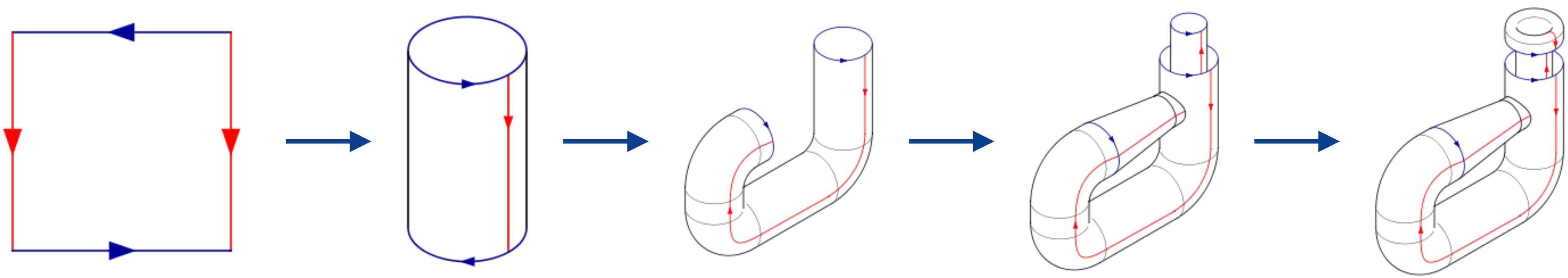
Image from [Dey et al. 2008]

Simplicial Homology

Example:

*The **Klein bottle K** is a non-orientable 2-dimensional manifold embeddable in $\mathbb{R}^4$ which can be built from a unit square by the following construction*





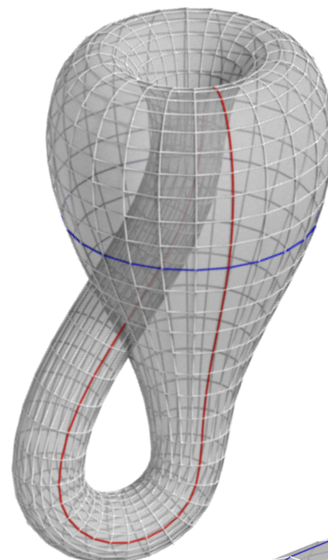
Simplicial Homology

Example:

By considering $\mathbb{Z}$ as coefficient group,

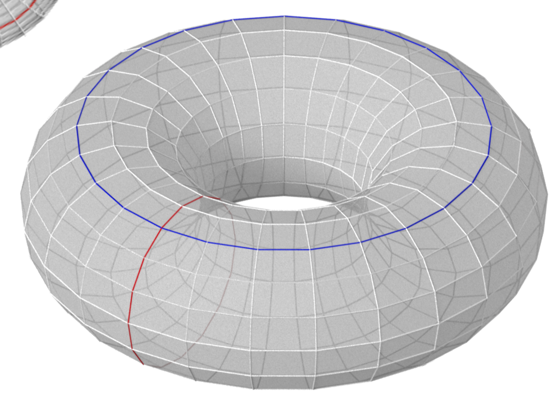
K has the following homology groups

$$H_k(K;\mathbb{Z}) \cong \begin{cases} \mathbb{Z} & \text{for } k = 0 \\ \mathbb{Z} \oplus \mathbb{Z}_2 & \text{for } k = 1 \\ 0 & \text{for } k \geq 2 \end{cases}$$



So, it can be distinguished from a torus T

$$H_k(T;\mathbb{Z}) \cong \begin{cases} \mathbb{Z} & \text{for } k = 0 \\ \mathbb{Z}^2 & \text{for } k = 1 \\ \mathbb{Z} & \text{for } k = 2 \\ 0 & \text{for } k > 2 \end{cases}$$

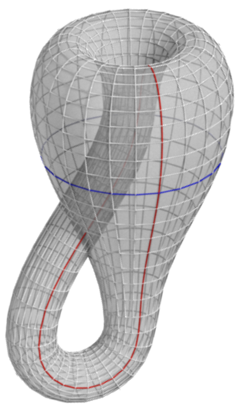


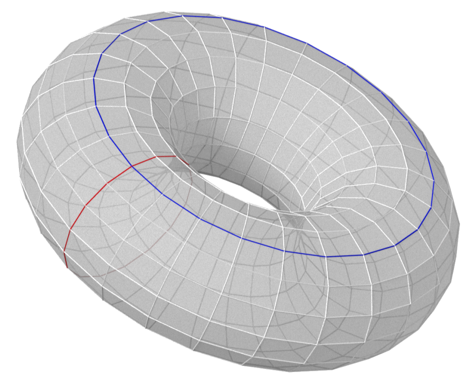
Simplicial Homology

Example:

By considering $\mathbb{Z}_2$ as coefficient group,

the Klein bottle K and the torus T have isomorphic homology groups





$$H_k(K;\mathbb{Z}_2) \cong \left\{ \begin{array}{ll} \mathbb{Z}_2 & \text{for } k=0 \\ \mathbb{Z}_2 \oplus \mathbb{Z}_2 & \text{for } k=1 \\ \mathbb{Z}_2 & \text{for } k=2 \\ 0 & \text{for } k>2 \end{array} \right\} \cong H_k(T;\mathbb{Z}_2)$$

Bibliography

Some References:

- ***Simplicial Homology:***

  - J. R. Munkres. ***Elements of algebraic topology***. CRC Press, 1984.