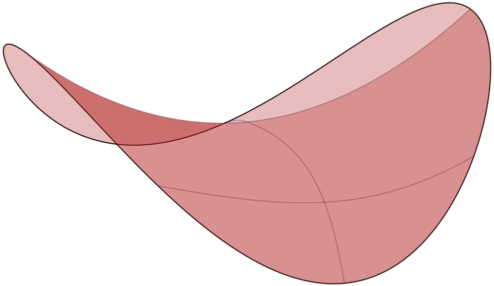
Introduction to Persistent Homology

Luciano Melodia

Data manifolds

Underlying spaces



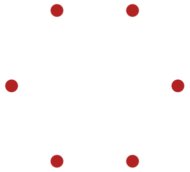
What if this **manifold** describes a set of points close to perfection?
How can we determine a suitable **topology** for a given set of points?

Overview

Part I: **Motivation**

Data manifolds

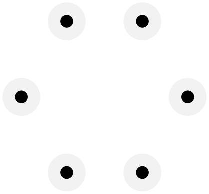
The structure of data



There exists at least a **topological manifold** underlying the data.
A **dataset** is a set of points embedded in some $\mathbb{R}^n$.

Data manifolds

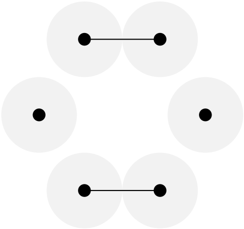
The structure of data



How can we make the **topological structure** of the underlying space visible?
We span a **simplicial complex** using closed balls starting with radius $r = 0.25$.

Data manifolds

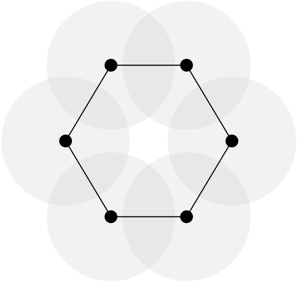
The structure of data



No additional **simplex** is created, thus we enlarge the radius.
This **simplicial complex** is spanned with an $r = 0.5$.

Data manifolds

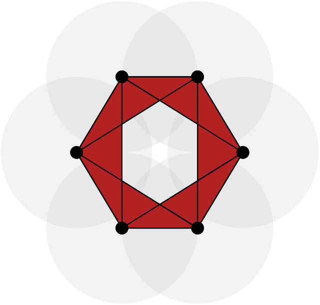
The structure of data



Whenever two of the **closed balls** touch, an edge is created.
This **simplicial complex** is spanned with an $r = 0.85$.
This simplicial complex is called 1-skeleton or $K^{(1)}$, with $\dim \sigma \leq 1$.

Data manifolds

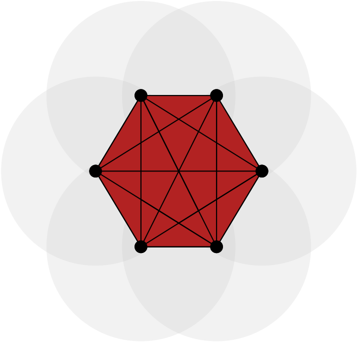
The structure of data



Whenever three or more of the **closed balls** touch, a 2-**simplex** is created.
This **simplicial complex** is spanned with an $r = 1.0$.

Data manifolds

The structure of data



When no more additional **simplices** appear, we have a **triangulation**.
This **triangulation** does not embed into $\mathbb{R}^2$, but into $\mathbb{R}^3$.

Part II: **Simplicial complexes**

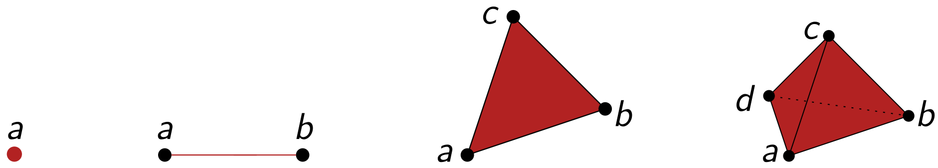
Simplices

Building blocks

Given a set $X = \{x_0, \cdots, x_k\} \subset \mathbb{R}^d$ of $k+1$ points that **do not lie on a hyperplane with dimension less than** d, the k-dimensional simplex v spanned by X is the **set of convex combinations**

$$\sum_{i=0}^{k} \lambda_i x_i \quad \text{with} \quad \sum_{i=0}^{k} \lambda_i = 1 \quad \text{and} \quad \lambda_i \geq 0. \tag{1}$$

Simplices
Examples



0-simplex point | 1-simplex segment | 2-simplex triangle | 3-simplex tetrahedron

The coefficients λ_i are chosen from the **vectorspace** $\mathbb{Z}_2 := \mathbb{Z}/2\mathbb{Z}$, such that we can **neglect the orientation** of the simplices. This is done for illustrative purposes, but can also be used for highly efficient computations.

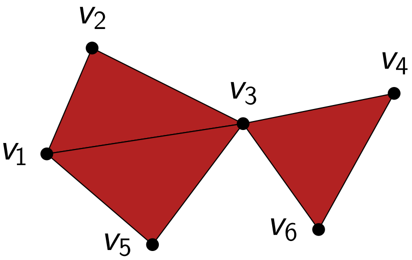
Abstract simplicial complexes
Definition

Let V be a finite set of vertices $\{v_1, \cdots, v_n\}$. Let K be a set of subsets of V, such that all subsets of elements from K also belong to K. Then we call the elements of K **faces** and K an **abstract simplicial complex**.

The dimension of K is the maximum dimension of any of its simplices. The underlying space is denoted by $|K|$. It is the union of its simplices together with the topology inherited from $\mathbb{R}^d$, with d being the dimension of K.

Abstract simplicial complexes

Example



$$V = \{v_1, v_2, v_3, v_4, v_5, v_6\}.$$

$$K = \{\{v_1\}, \cdots, \{v_6\}, \{\{v_1\}, \{v_2\}\}, \cdots, \{\{v_4\}, \{v_6\}\},$$
$$\{\{v_1\}, \{v_2\}, \{v_3\}\}, \cdots, \{\{v_3\}, \{v_4\}, \{v_6\}\}\}.$$

Geometric simplicial complexes

The geometric realization theorem

An **abstract simplicial complex** of dimension d has a **geometric realization** in $\mathbb{R}^{2d+1}$.

Geometric simplicial complexes

Proof of the geometric realization theorem

Let $f : K^{(0)} \to \mathbb{R}^{2d+1}$ be injective, whose image is a **set of points in general position** ($2d+2$ or fewer points are affinely independent). Let σ and σ_0 be simplices in K with $k = \dim \sigma$ and $k_0 = \dim \sigma_0$. Their union has size

$$
\begin{aligned}
\text{card}\,(\sigma \cup \sigma_0) &= \text{card}\,\sigma + \text{card}\,\sigma_0 - \text{card}\,(\sigma \cap \sigma_0) \qquad (2)\\
&\leq k + k_0 + 2 \leq 2d + 2. \qquad (3)
\end{aligned}
$$

Geometric simplicial complexes

Proof of the geometric realization theorem

The points in $\sigma \cup \sigma_0$ are affinely independent.
$\implies$ Every **convex combination** x **of points** in $\sigma \cup \sigma_0$ is unique.

Hence $x \in \tau = \text{conv } f(\sigma)$ as well $x \in \tau_0 = \text{conv } f(\sigma_0)$,
iff x is a convex combination of $\sigma \cap \sigma_0$.
$\implies$ The intersection of τ and τ_0 is either empty or within $f(\sigma \cap \sigma_0)$.

Geometric simplicial complexes
Simplicial maps

A **vertex map** is a function $\varphi : K^{(0)} \longrightarrow L^{(0)}$ with the property that the vertices of every simplex in K map to vertices of a simplex in L. Then φ can be extended to a continuous map $f : |K| \longrightarrow |L|$ defined by

$$f(x) = \sum_{i=0}^{n} \lambda_i(x)\varphi(v_i). \tag{4}$$

This map is the **simplicial map** induced by φ.

Part III: **Filtrations**

Filtered (simplicial) complexes

Brief motivation

We believe that different data comes from different **topological spaces**. How can we distinguish these using the data?

*We don't want to distinguish data only by their **holes** of their **triangulation**, therefore we introduce a **magnifying glass** which allows us to capture the **structure of a topological space** in different granularity.*

Filtered (simplicial) complexes
Definition

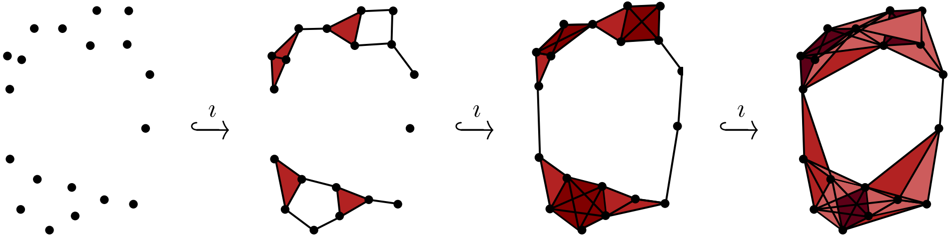
A **filtration** is a nested sequence of complexes K_i, which induce an ordering of the sublevel complexes. These complexes, together with the inclusion $K_i \hookrightarrow K_j$ for $0 \leq i \leq j \leq n$ is called a filtration and denoted by $\mathbb{K}$:

$$\mathbb{K} : \quad \emptyset = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_n = K.$$

The inclusion on the filtration induces a homomorphism of groups $f_k^{i,j} : H_k(K_i) \to H_k(K_j)$, in this case H_k are the k-th homology groups.

Filtered (simplicial) complexes

Example



Vastly used complexes to create a filtration on a point set X are:

Čech complex: $(x_0, x_1, \cdots, x_k) \in \text{Čech}_r(X) \iff \bigcap_{i=0}^{k} B_r(x_i) \neq \emptyset.$

Vietoris-Rips complex: $(x_0, x_1, \cdots, x_k) \in \text{Rips}_r(X) \iff \|x_i - x_j\| \leq r.$

Part IV: **Homology groups**

Groups

Recall

A **set** G together with a map $+ : G \times G \longrightarrow G$, $(x, y) \mapsto x + y$ and an element $e = e_G$ are called **group**, iff

1. $(xy)z = x(yz)$ for all $x, y, z \in G$,
2. $xe = ex = x$ for all $x \in G$,
3. for all $x \in G$ there exists $x^{-1} \in G$ such that $xx^{-1} = x^{-1}x = e$.

If the group is abelian, which means commutative in every element under the operation, one denotes the group operation as $+$.

Homomorphisms of groups

Recall

A **homomorphism of groups** is a map $\varphi : G \longrightarrow H$, with G, H being groups, such that

$$\varphi(e_G) = e_H,$$
$$\varphi(g_1 \star g_2) = \varphi(g_1) \circ \varphi(g_2).$$

Chain groups

Definition

The **kth chain group** of a **simplicial complex** K is $(C_k(K), +)$, the **free commutative group** on the (oriented) k-simplices.

An element of $C_k(K)$ is called k-chain, $\sum_i \lambda_i \sigma_i, \lambda_i \in \mathbb{Z}_2, \sigma_i \in K$.

Boundary homomorphism

Definition

Let K be a simplicial complex and $\sigma \in K$, $\sigma = [v_0, v_1, \cdots, v_k]$. The **boundary homomorphism** $\partial_k : C_k(K) \to C_{k-1}(K)$ is

$$\partial_k \sigma = \sum_i (-1)^i [v_0, v_1, \cdots, \hat{v}_i, \cdots, v_n]. \tag{5}$$

Boundary homomorphism

Fundamental lemma

The composition $\partial_{k-1} \circ \partial_k$ is zero.
We have

$$\partial_k(\sigma) = \sum_i (-1)^i \sigma|[v_0, \cdots, \hat{v}_i, \cdots, v_k], \tag{6}$$

and hence

$$\partial_{k-1}\partial_k(\sigma) = \sum_{j<i} (-1)^i (-1)^j \sigma|[v_0, \cdots, \hat{v}_i, \cdots, \hat{v}_j, \cdots, v_k] \tag{7}$$

$$+ \sum_{j>i} (-1)^i (-1)^{j-1} \sigma|[v_0, \cdots, \hat{v}_i, \cdots, \hat{v}_j, \cdots, v_k]. \tag{8}$$

The latter two summations cancel, as the second sum becomes the negative of the first after switching i and j.

Cycle group

The ***k*th cycle group** Z_k is defined as

$$Z_k = \ker \partial_k \tag{9}$$
$$= \{c \in C_k \mid \partial_k c = 0\}. \tag{10}$$

An element of this group is called k-**cycle**.

Boundary group

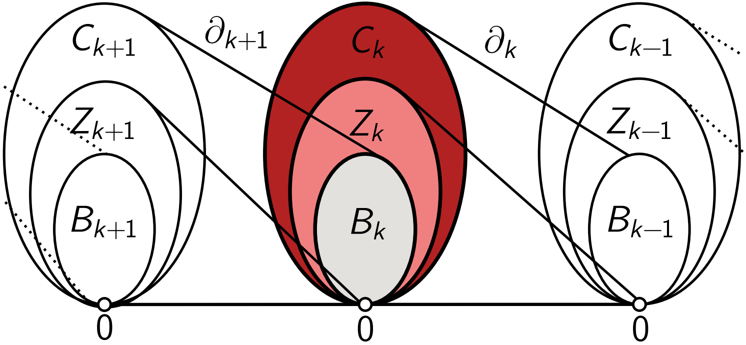
The ***k*th boundary group** B_k is defined as

$$\begin{aligned} B_k &= \operatorname{im} \partial_{k+1} && (11) \\ &= \{c \in C_k \mid \exists d \in C_{k+1} : c = \partial_{k+1} d\}. && (12) \end{aligned}$$

An element of this group is called k-**boundary**.

Normal subgroup relation

Following Edelsbrunner, Zomorodian and many more



A subgroup B of a group Z is normal with respect to Z, iff $zbz^{-1} \in B$. Thus, if it is invariant under conjugation. Normal subgroups are important. Only they can construct quotient groups.

Chain complex

$$0 \xrightarrow{\partial_{k+1}} C_k \xrightarrow{\partial_k} C_{k-1} \xrightarrow{\partial_{k-1}} \cdots \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_1} 0 \tag{13}$$

Homology groups
Definition

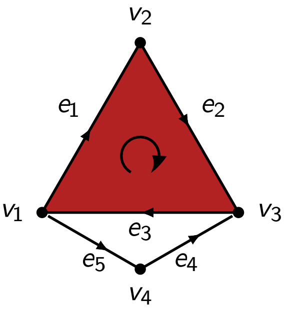
The kth **homology group** of a simplicial complex is defined as

$$H_k(K) = \ker \partial_k C_k(K) / \mathrm{im}\ \partial_{k+1} C_{k+1}(X). \tag{14}$$

Intuitively, the kernel of the boundary homomorphism of the kth chain group gives all k-cycles, thus the cycle group, from which we quotient out all elements of the kth boundary group, i.e. $H_k(K) = Z_k(K)/B_k(K)$.

Technical details: Both subgroups, the cycle and boundary group, are normal because our chain groups are abelian.

Homology groups
Example



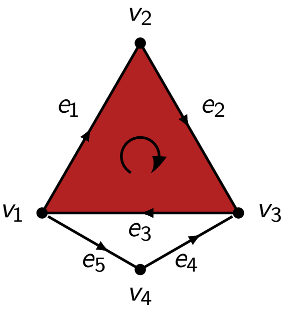
What we want to compute:

$$
\begin{aligned}
H_0(K) &= \ker(\partial_0(C_0(K)))/\mathrm{im}\,(\partial_1(C_1(K))) \\
&= C_0(K)/\mathrm{im}(\partial_1(C_1(K)))
\end{aligned}
$$

Applying the boundary operator:

$$
\begin{aligned}
\partial_1(c \in C_1(K)) &= \partial_1(\lambda_1 e_1 + \lambda_2 e_2 + \lambda_3 e_3 + \lambda_4 e_4 + \lambda_5 e_5) \\
&= \lambda_1(\partial_1(e_1)) + \lambda_2(\partial_1(e_2)) + \lambda_3(\partial_1(e_3)) \\
&\quad + \lambda_4(\partial_1(e_4)) + \lambda_5(\partial_1(e_5)) \\
&= \lambda_1(v_2 - v_1) + \lambda_2(v_3 - v_2) + \lambda_3(v_1 - v_3) \\
&\quad + \lambda_4(v_3 - v_4) + \lambda_5(v_4 - v_1)
\end{aligned}
$$

Homology groups
Example



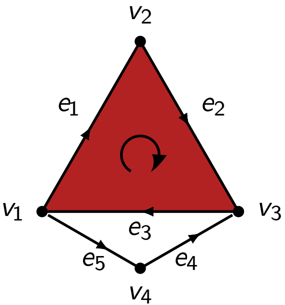
What we have computed:

$$\partial_1(c \in C_1(K)) = \lambda_1(v_2 - v_1) + \lambda_2(v_3 - v_2) + \lambda_3(v_1 - v_3) + \lambda_4(v_3 - v_4) + \lambda_5(v_4 - v_1)$$

We bring this into matrix form:

$$M_{\partial_1(c)} = \begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{array} \begin{array}{c} \begin{array}{ccccc} v_1 & v_2 & v_3 & v_4 & v_5 \end{array} \\ \begin{pmatrix} -1 & 0 & 1 & 0 & -1 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

Homology groups
Example



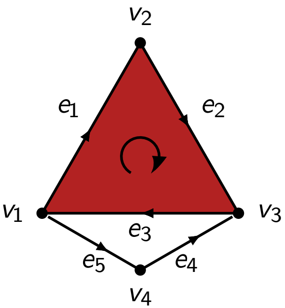
What we have computed:

$$M_{\partial_1(c)} = \begin{pmatrix} -1 & 0 & 1 & 0 & -1 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

We conclude that:

$C_1(K) \simeq \mathbb{Z}^5$.
What about $M_{\partial_1(c)}$?

Homology groups
Example



After some Gaussian elimination:

$$M_{\partial_1(c)} = \begin{pmatrix} 1 & 0 & -1 & 0 & 1 \\ 0 & 1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Thus $\mathrm{rank}(M_{\partial_1(c)}) = 3$ and $M_{\partial_1(c)} \simeq \mathbb{Z}^3$.
Therefore

$$H_0(K) = C_0(K)/B_0(K) \simeq \mathbb{Z}^4/M_{\partial_1(c)} \simeq \mathbb{Z}^4/\mathbb{Z}^3 \simeq \mathbb{Z}. \quad (15)$$

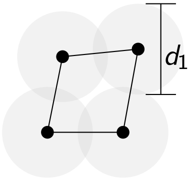
Betti numbers

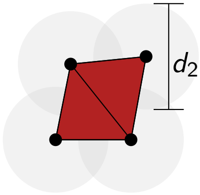
The kth **Betti** number β_k is the rank of the kth homology group $H_k(K)$ of the topological space K.

We'll track the betti numbers to track the amount of holes along the filtration. Other properties of the vector space could also be helpful.

Part V: **Persistent homology**

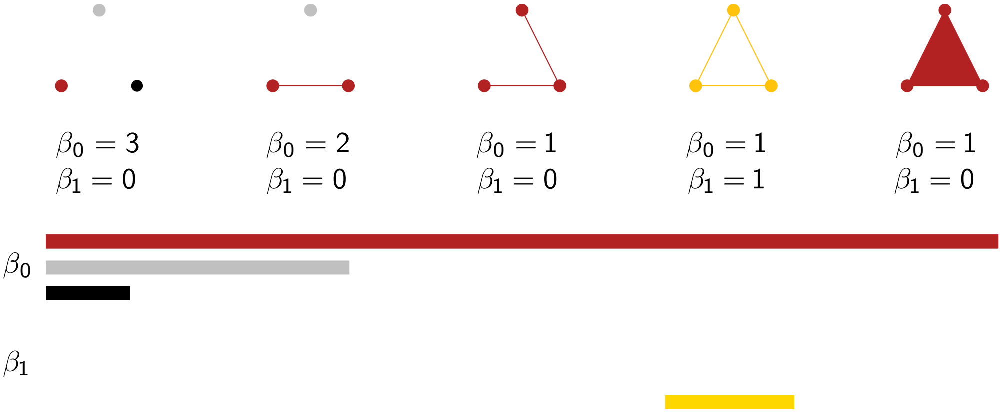
What is persistence?





Each hole, corresponding to a **representative** of a particular **homology group**, appears at a particular value of d and disappears at another value of d. We can represent the **persistence** of such a representative by a pair (d_1, d_2).

Persistence barcodes



Takeaway message

Persistent homology is a new way to analyse high dimensional data.

Persistent homology uncovers highly **nonlinear features**.

It has connections to other **homology / homotopy theories**.

It has connections to **homological algebra**.

It is used along many mathematical disciplines.

Algorithms compute in $\mathcal{O}(n^3)$ with respect to the number of simplices.

More Questions?
Drop me a line: **luciano.melodia@fau.de**,
and add me on GitHub: **karhunenloeve**.

Thank you for your attention!