[Otter et al] Sec. 5 "Computation of PH for data"

Data $\xrightarrow{\text{choice of complexes}}$ filtered complex $\longrightarrow$ Barcodes $\xrightarrow[\text{structural Thm}]{}$ Interpretation

Two features of TDA

① computable via linear algebra (finite dim. rep.)

② stable with respect to perturbations in the measurements of data (robust to scaling)

After we saw the minimal example using VR-complex, we can now consider more general a situation and formalisms.

An abstract definition of simplicial homology without referring to underlying $\mathbb{R}^d$ is:

Def. A simplicial complex is a collection K of nonempty subsets of a set K_0 s.t. $\{v\} \in K$ for all $v \in K_0$ and $\tau \subset \sigma$, $\sigma \in K$ guarantees $\tau \in K$ (filtering.)

Another idea: define probability structure over filtering: HOW?

The elements of K are called simplexes (simplicial complex is defined first.).

Additionally we say that a simplex has dimension p OR a p-simplex if it has a cardinality of $(p+1)$

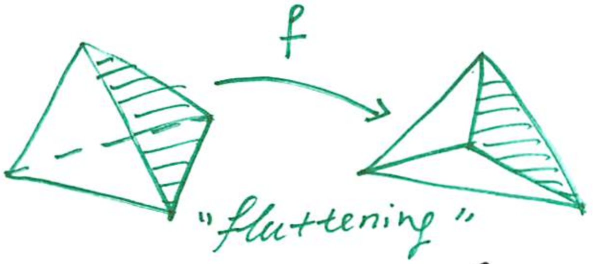
$K_p :=$ the collection of p-simplices.

The k-skeleton of K is the union of the sets K_p for all $p = 0, 1, 2 \ldots k$

If τ, σ are simplices s.t. $\tau \subset \sigma$, then we call τ a face of σ; AND τ is a face of codimension $k' := \dim \sigma - \dim \tau$.

The dimension of (abstruct) simplicial complex K is defined to be the maximum of the dimension of its simplices.

A map / mapping well-defined between simplicial complexes K_1, K_2 $f: K_1 \to K_2$, $\sigma \mapsto f(\sigma) \in K_2$



For finite simplicial complex K, it's always possible (although non-trivially). to embed it into $\mathbb{R}^N$

Now we define the Boundary mapping without geometric arguements:

$$d_p : C_p(K) \longrightarrow C_{p-1}(K)$$

$$\sigma \mapsto \sum_{\tau \subset \sigma, \tau \in K_{p-1}} \tau$$

where $C_p(K)$ is the $\mathbb{F}_2$ (two element field. it's unique and isomorphic to $\mathbb{Z}/2\mathbb{Z}$). vector space with basis given by the p-simplices of K

$\mathbb{F}_2 +$	0	1
0	0	1
1	1	0

addition

$\mathbb{F}_2 \cdot$	0	1
0	0	0
1	0	1

multiplication

Coro 1 $d_p \circ d_{p+1} = 0$

Coro 2 d_p is linear.

The p-th homotogy of a simplicial complex K is the quotient vector space

$$H_p(K) := \operatorname{Ker} d_p / \operatorname{Im} d_{p+1}.$$

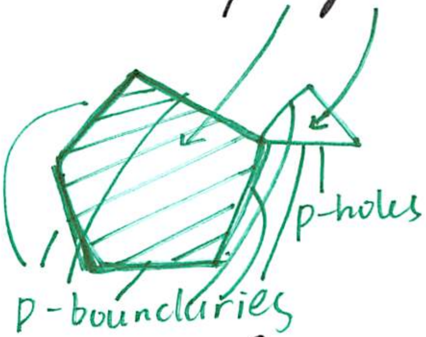
and its dimension is defined as

$$\beta_p(K) := \dim H_p(K) = \dim \operatorname{Ker} d_p - \dim \operatorname{Im} d_{p+1}$$

$\beta_p(K)$ is called the p-th Betti number of K.

$\mathrm{Im}\, d_{p+1} \leftarrow$ p-boundaries] compare with
$\mathrm{Ker}\, d_p \leftarrow$ p-cycles] their geometric constructions.

The p-cycles that are NOT p-boundaries are p-holes. and $\beta_p(K)$ is the number of p-holes (topological invariant, BUT with a direct geometric intuition and subject to simple calculation.)

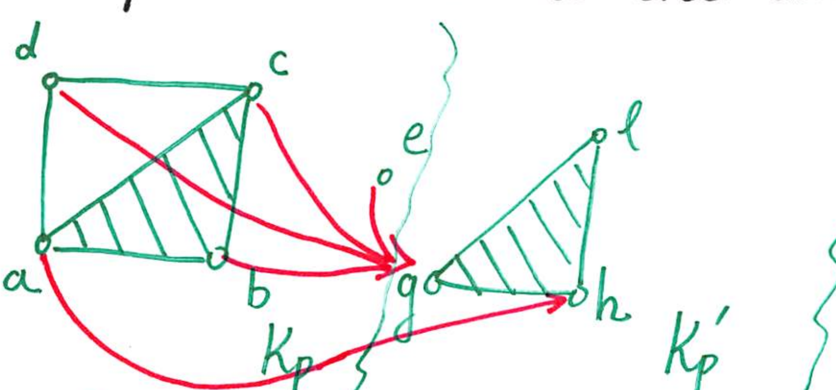


The functoriality of $H_{\mathbb{F}_2}$ functor allows us to compute induced mapping for $f: K \to K'$

$f_p: C_p(K) \to C_p(K')$ over $\mathbb{F}_2$

$$\sum_{\sigma \in K_p} c_\sigma \cdot \sigma \mapsto \sum_{\substack{\sigma \in K_p \\ f(\sigma) \in K'_p}} c_\sigma f(\sigma). \quad c_\sigma \in \mathbb{F}_2$$

To illustrate that these abstract concepts are computable. we do one example [Hatcher][Other]



$K = \{\{a\}, \{b\}, \{c\}, \{d\}, \{e\}, \{a,b\}, \{a,c\}, \{a,d\}, \{b,c\}, \{c,d\}, \{a,b,c\}\}$

$p = 2$.

$K' = \{\{g\}, \{h\}, \{l\}, \{g,h\}, \{g,l\}, \{h,l\}, \{g,h,l\}\}$

$f: K \longrightarrow K', \quad x \longmapsto \begin{cases} h & \text{if } x = a \\ g & \text{otherwise.} \end{cases}$

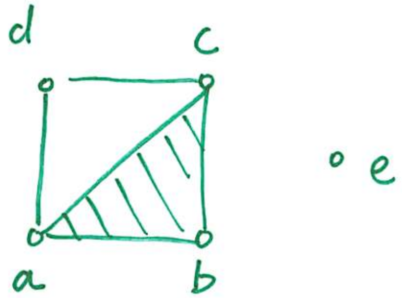
Verify this is a well-defined map between K, K' (Check $f(\sigma) \in K', \forall \sigma \in K$.)

let's compute their simplicial homologies as follows

$$0 \xrightarrow{d_{n+1}} C_n(K) \xrightarrow{d_n} \cdots \xrightarrow{d_2} C_1(K) \xrightarrow{d_1} C_0(K) \xrightarrow{d_0} 0$$

$$K: \quad 0 \xrightarrow{d_3} \mathbb{F}_2 \xrightarrow{d_2} \mathbb{F}_2^5 \xrightarrow{d_1} \mathbb{F}_2^5 \xrightarrow{d_0} 0$$

$$K' \quad 0 \xrightarrow{d_3} \mathbb{F}_2 \xrightarrow{d_2} \mathbb{F}_2^3 \xrightarrow{d_1} \mathbb{F}_2^3 \xrightarrow{d_0} 0$$



if we relabel the basis $\{a, b, c, d, e\}$ for $C_1(K)$ as $\{e_1, e_2, e_3, e_4, e_5\}$

$$d_1 : C_1(K) \to C_0(K)$$

can be represented as

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Similarly, with

$$d_2 : C_2(K) \to C_1(K)$$

its matrix representation should be.

$[1 \quad 1 \quad 1]$ (Think again!).

Table 1 in [Otter. et al]

Complex K	(worst-case) Size of K	Theoretical guarantee
Čech	$2^{\mathcal{O}(N)}$	Nerve theorem
Vietoris–Rips (VR)	$2^{\mathcal{O}(N)}$	Approximates Čech complex
Alpha	$N^{\mathcal{O}(\lceil d/2 \rceil)}$ (N points in $\mathbb{R}^d$)	Nerve theorem
Witness	$2^{\mathcal{O}(\lvert L \rvert)}$	For curves and surfaces in Euclidean space
Graph-induced complex	$2^{\mathcal{O}(\lvert Q \rvert)}$	Approximates VR complex
Sparsified Čech	$\mathcal{O}(N)$	Approximates Čech complex
Sparsified VR	$\mathcal{O}(N)$	Approximates VR complex

where N is the size of vertices set (size of observations.)

L is the set of landmark points of "Witness" complex.

Q is the set of subsamples induced by the graph.

These complexes can be categorized into:

① Čech complexes and its approximations whose structure is determined by <u>Nerve Thm</u>

② Alpha complexes.

③ Specially designed complexes for certain kinds of data. ("Witness" complex.