Introduction to Applied Topology

Lecture Notes

Introduction to Applied Topology

In the next 5 weeks, we will (roughly) cover:

- invariants : euler characteristic, (co) homology, persistence diagrams
- notions of stability : classical & proofs ; recent thms
- applications : functionals of persistence diagrams
  manifold learning
  topology of random spaces
- generalisations : multiparameter persistence
  sheaf theory
  other applications ...

What we will not cover : homotopy
limit theorems
....

Note that these note will be expanded during the course

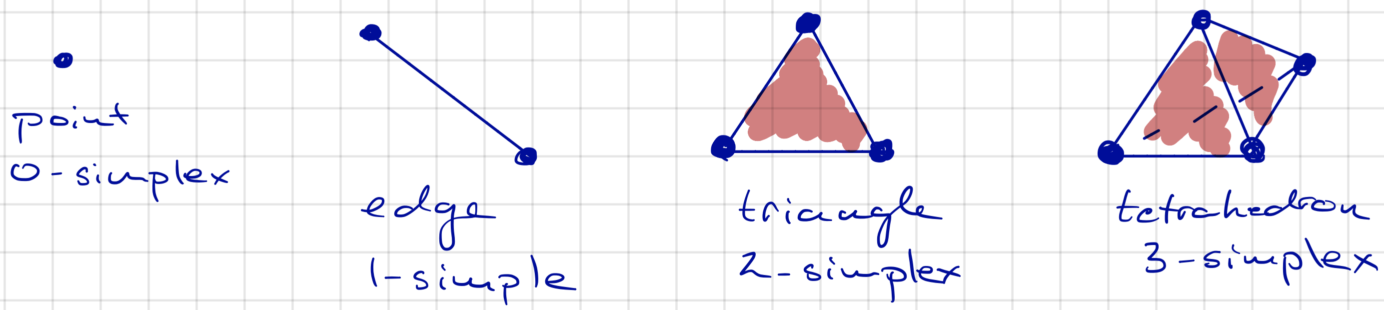
Preliminaries

Topological space - very general
- defined in terms of neighborhoods
- lead to many pathological examples

Other types of spaces: Manifolds, stratified spaces, metric spaces, ...

1^{st} model: Simplicial complex

Def: A k-simplex is the convex combination of $(k+1)$-points



If points are embedded in Euclidean space then a simplex is just the convex hull

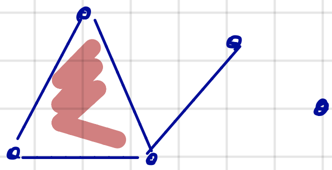
Note: Simplices can be abstract, not necessarily embedded (we will see this later)

One simplex is not terribly interesting
– consider simplicial complexes.

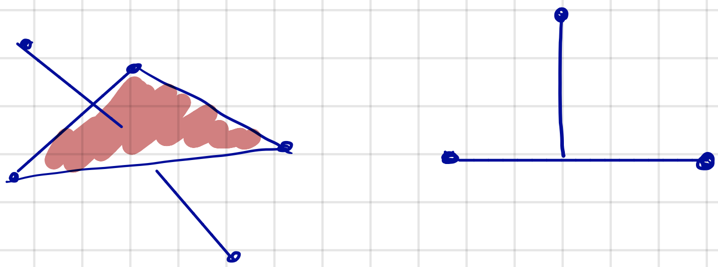
Def: A simplicial complex Δ is a set of simplicies $\{\sigma\}$ such that

1) if $\sigma \in \Delta$ & $\tau \le \sigma$ then $\tau \in \Delta$
2) if $\sigma_1 \cap \sigma_2 \neq \phi$ then $\sigma_1 \cap \sigma_2 \in \Delta$

Examples



simplicial complex



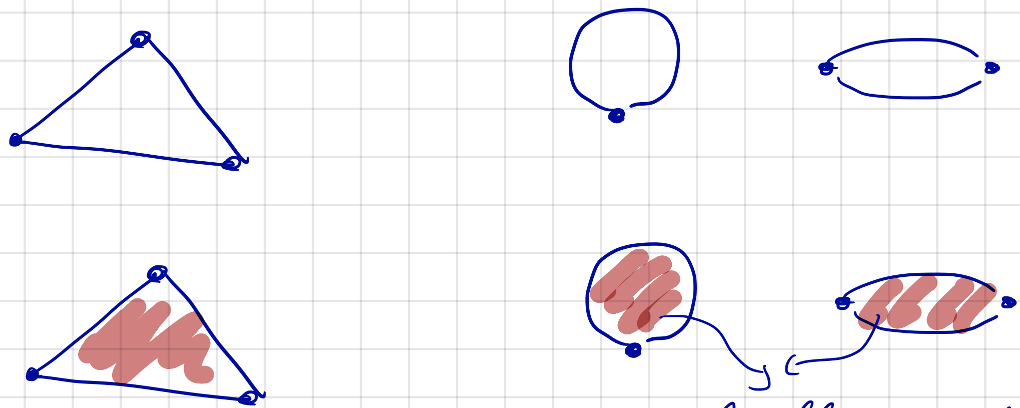
not simplicial complexes

We will introduce related concepts as we need them, e.g. carrier, closure, star, link, ...

Note: A simplicial complex is often the easiest to work with in a computer, but for hand computed examples we will often use cellular complexes

Def: A cellular complex consists of k-cells & attaching maps (how to glue the cell to lower dimensional cells

Example:



2 cells are discs.

simplicial

cellular

There are other models: Δ-complexes, simplicial sets

Here we will mainly use simplicial complexes (with cellular for some small examples)

Note: Simplicial $\Longleftrightarrow$ cellular

* Simplicial is always cellular

cellular is "equivalent" to simplicial

→ We will see exactly how later, but one can subdivide a cells into simplices.

Equivalences

We will see many type of equivalences however, topological invariants are invariant under continuous transformations
(this could be any continuous functions, homeomorphisms, diffeomorphisms, etc)

For our purposes, if f is a homeomorphism for $f: X \to Y$ then we consider $X \sim Y$ equivalent. As we introduce these concepts we will place them into context (stronger vs. weaker)

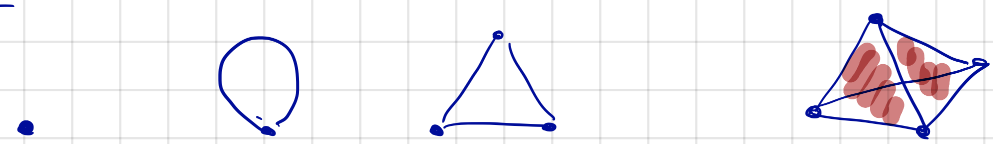
Euler Characteristic

One of the most basic & most ubiquitous topological invariants.

For surfaces (Euler)

$$\chi = |V| - |E| + |F|$$

Example:



$\chi = 1$ $\chi = 0$ $\chi = 3 - 3 = 0$ $\chi = 4 - 6 + 4 = 2$

General formula

$$\chi(K) = \sum_{i=0}^{d} (-1)^i (\# \, k\text{-simplices in } \Delta)$$

Facts:

- Topological invariant
- Easy to compute
- Ubiquitous (topology, differential geometry, category theory, random fields, ...)

Chain Complexes

We are after computable invariants. Much of what we will use will be based on linear algebra. Our starting point is the chain complex.

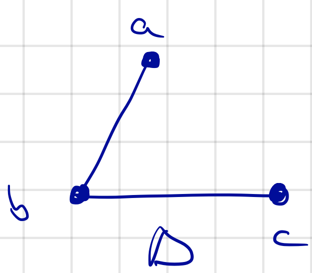
We will first describe the most relevant example before formally defining it.

Let K_d be the set of all d-simplices in K

Define $C_d(\Delta) := I_{|K_d|}$

The d-th chain group $C_d(\Delta)$ is the vector space where each simplex is its own dimension

Example



$$\begin{array}{c|ccc} & a & b & c \\ \hline a & 1 & 0 & \\ b & & 1 & \\ c & 0 & & 1 \end{array} \qquad \begin{array}{c|cc} & ab & bc \\ \hline ab & 1 & 0 \\ bc & 0 & 1 \end{array}$$

$$C_0(K) \qquad\qquad C_1(K$$

<u>Note 1</u> : In everything we consider, the chain group is really the chain vector space.

* Why is it called the chain group?

We can add simplices together (or more precisely, we can take linear combinations of simplices)

<u>Def</u> : A k-chain is the linear combination of k-simplicies

$$\boxed{c = \sum_{\sigma \in K_d} \lambda_\sigma \sigma}$$

or in other words, it is a vector indexed by the k-simplicies.

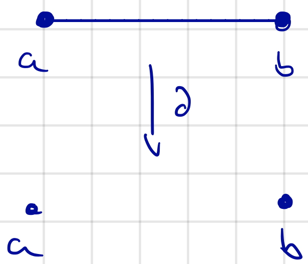
So far, we have only defined chains what is a chain complex?

Recall: Simplicies (or cells) of different dimensions are related (if τ is a face of a simplex σ in K then τ must be in K)

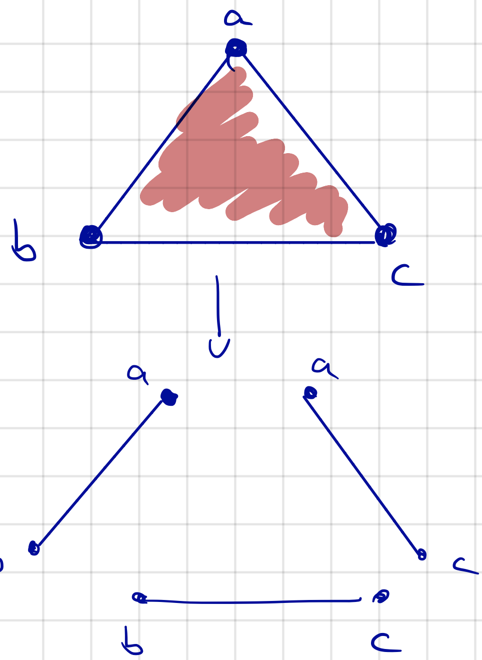
This gives rise to maps $\partial_k : C_k \to C_{k-1}$ $\forall k$

∂_k is called the k-th boundary operator. It describes how a k-simplex maps to a $(k-1)$-chain.

Example: (over $\mathbb{Z}_2$)



$\partial(ab) \longmapsto a+b$



$\partial(abc) = ab + bc + ac$

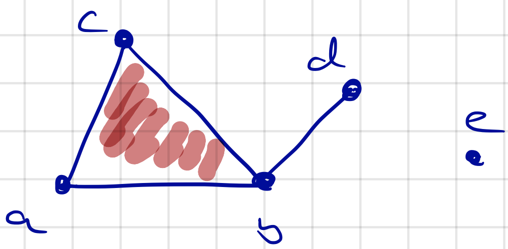
In general, we will have orientation (-1's) as well

Notice that ∂_k takes a k-simplex & returns a $(k-1)$ chain. Since it is linear, more generally it takes a k-chain & returns a $(k-1)$-chain.

$$c = \sum \lambda_\sigma \sigma \implies \partial(c) = \partial\left(\sum \lambda_\sigma \sigma\right)$$
$$= \sum \lambda_\sigma \partial(\sigma)$$

Note: vertices map to 0, $\partial(\cdot) = 0$

Example:



∂_0

$$\begin{array}{ccccc} a & b & c & d & e \\ \end{array}$$
$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

∂_1

$$\begin{array}{c|c} & \begin{array}{cccc} ab & ac & bc & bd \end{array} \\ \begin{array}{c} a \\ b \\ c \\ d \\ e \end{array} & \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ & & & \end{bmatrix} \end{array}$$

∂_2

$$\begin{array}{c|c} & abc \\ \begin{array}{c} ab \\ ac \\ bc \\ bd \end{array} & \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} \end{array}$$

Since $\partial_k : C_k \to C_{k-1}$, the chain complex is

$$\cdots\; C_{k+1} \xrightarrow{\partial_{k+1}} C_k \xrightarrow{\partial_k} C_{k-1} \xrightarrow{\partial_{k-1}} \cdots \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_1} C_0 \to 0$$

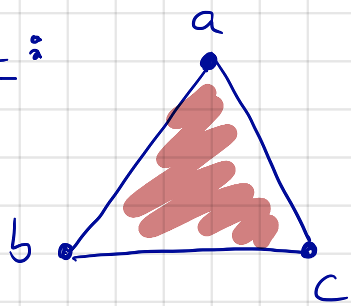
The collection $\{C_k, \partial_k\}_{k\in\mathbb{Z}}$ is the chain complex.

<u>Property</u>: A chain complex must also have the property $\boxed{\partial_k \circ \partial_{k+1} = 0}$

this is the key condition

Alternatively $\operatorname{im} \partial_{k+1} \subseteq \ker \partial_k$

<u>Example</u>:



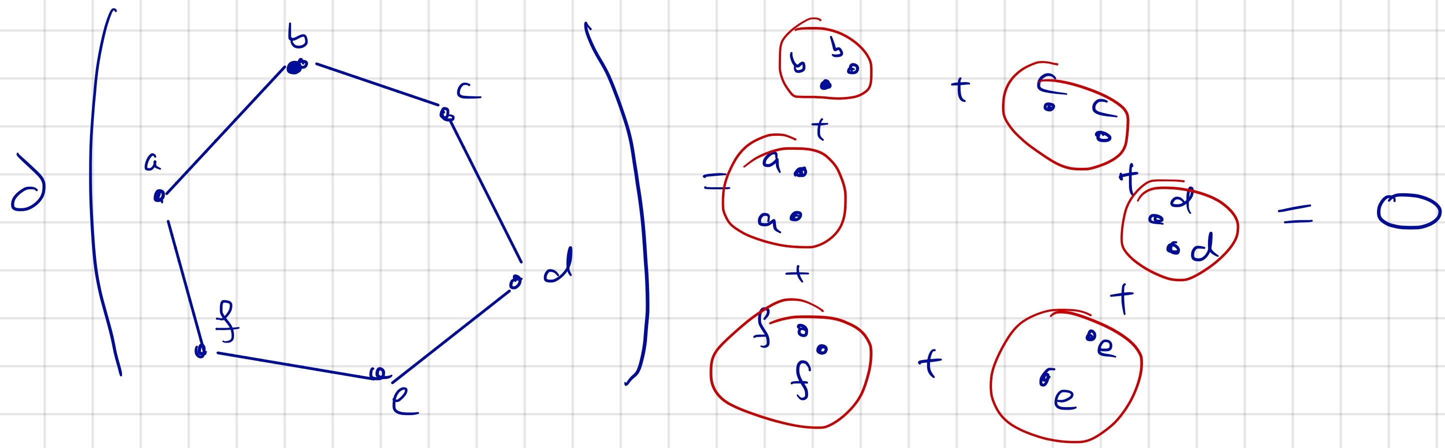
$$\partial(abc) = ab + ac + bc$$

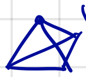
$$\begin{aligned}\partial \circ \partial(abc) &= \partial(ab + ac + bc)\\ &= \partial(ab) + \partial(ac) + \partial(bc)\\ &= a + b + a + c + b + c\\ &= 0\end{aligned}$$

($a+a=0$, $b+b=0$, $c+c=0$) since we are over $\mathbb{Z}_2$

<u>Def</u>: An element of $\ker \partial_k$ is called a k-cycle.

Why? Think cycles in graph theory



<u>Exercise</u>: Verify that $\partial \circ \partial$ () $= 0$, that the boundary of an empty tetrahedron (4 triangles) is 0.

<u>Def</u>: The space of all k-cycles is called the cycle space.

<u>Def</u>: An element of $\operatorname{im} \partial_{k+1} \in C_k$ is called a k-boundary (because it bounds a k-cycle)

<u>Obs</u>: These are all vector spaces (though this holds only in our case)

Homology

Given a chain complex $\{C_k, \partial_k\}$

let Z_k denote the space of cycles ($\ker \partial_k$)

B_{k+1} denote the space of boundaries ($\mathrm{im}\, \partial_{k+1}$)

The k-th homology group is

$$\boxed{H_k(\Delta) = \ker \partial_k / \mathrm{im}\, \partial_{k+1}}$$

"All cycles which are not bounded"

(In our case again a vector space)

<u>Def</u>: The k-th Betti number is the rank of H_k

$$\boxed{\begin{aligned} \beta_k = \mathrm{rk}(H_k) &= \mathrm{rk}\left(\frac{\ker \partial_k}{\mathrm{im}\, \partial_{k+1}}\right) = \\ &= \mathrm{rk} \ker \partial_k - \mathrm{rk}\, \mathrm{im}\, \partial_{k+1} \end{aligned}}$$

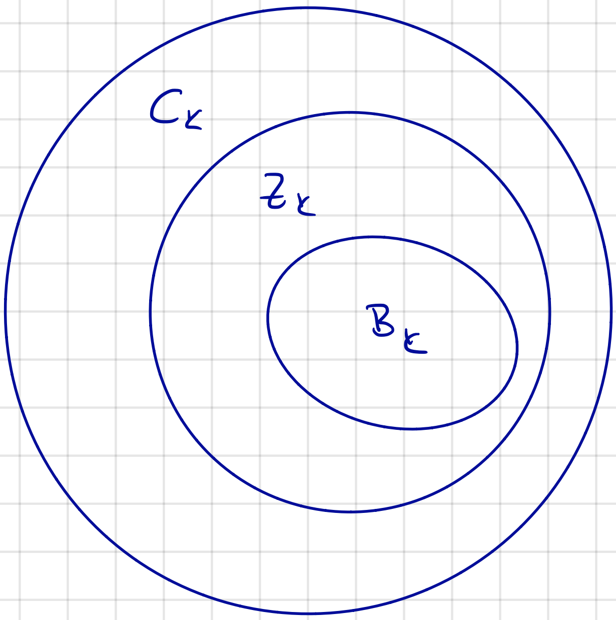
Why is this well-defined?

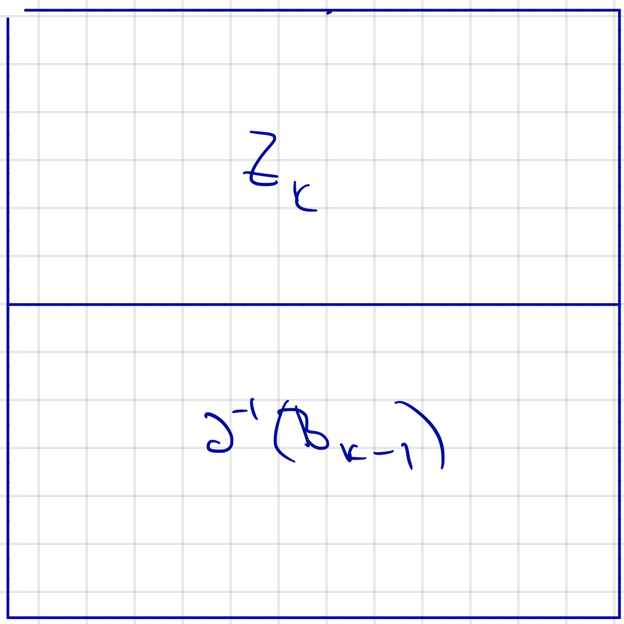
Recall: $\partial_k \circ \partial_{k+1} = 0$, so

$$\underset{\text{boundaries}}{\operatorname{im}\partial_{k+1}} \subseteq \underset{\text{cycles}}{\ker \partial_k} \subseteq \underset{\text{chains}}{C_k}$$

$$\operatorname{rk}(\operatorname{im}\partial_{k+1}) \leq \operatorname{rk}(\ker\partial_k) \leq \operatorname{rk}(C_k)$$

Picture:





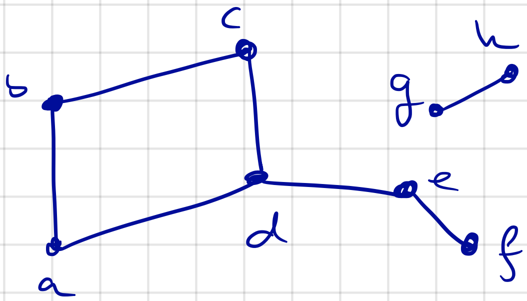
How can we compute this?

$$\boxed{\text{Gaussian elimination}}$$

Example 1: Graphs

$\forall v \in C_0 \qquad \partial(v) = 0 \qquad$ All vertices are cycles

$\Rightarrow C_0 = \ker \partial_0 \qquad$ so homology are all the equivalence classes of vertices modulo edges e.g. connected components



$\mathrm{im}\, \partial_1 = ?$

$$\begin{array}{c|ccccccc}
 & ab & bc & cd & ad & de & ef & gh \\
a & 1 & & & 1 & & & \\
b & \boxed{1} & 1 & & & & & \\
c & & \boxed{1} & 1 & & & & \\
d & & & \boxed{1} & 1 & 1 & & \\
e & & & & & \boxed{1} & 1 & \\
f & & & & & & \boxed{1} & \\
g & & & & & & & 1 \\
h & & & & & & & \boxed{1}
\end{array}$$

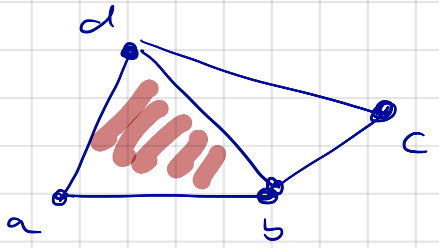
There are 8 cycles (vertices) & $\mathrm{rk}(\mathrm{im}\, \partial_1) = 6$

So $\mathrm{rk}(H_0) = 8 - 6 = 2$

The equivalence classes are the components & sum of components

Formally, $a+b+c+d+e+f$, $g+h$

Example 2: H_1



$\ker \partial_1 = ?$

	ab	bc	cd	bd	ad
a	1				1
b	$\boxed{1}$	1		1	
c		$\boxed{1}$	1		
d			$\boxed{1}$	1	1

$\ker \partial_1$ spanned by $\{bd + cd + bc,\ ad + cd + bc + ab\}$

$\operatorname{im} \partial_2 = ?$

	abd
ab	1
bc	
cd	
bd	1
ad	1

$\rightarrow$ need to make row space $= \ker \partial_1$

$\Downarrow$

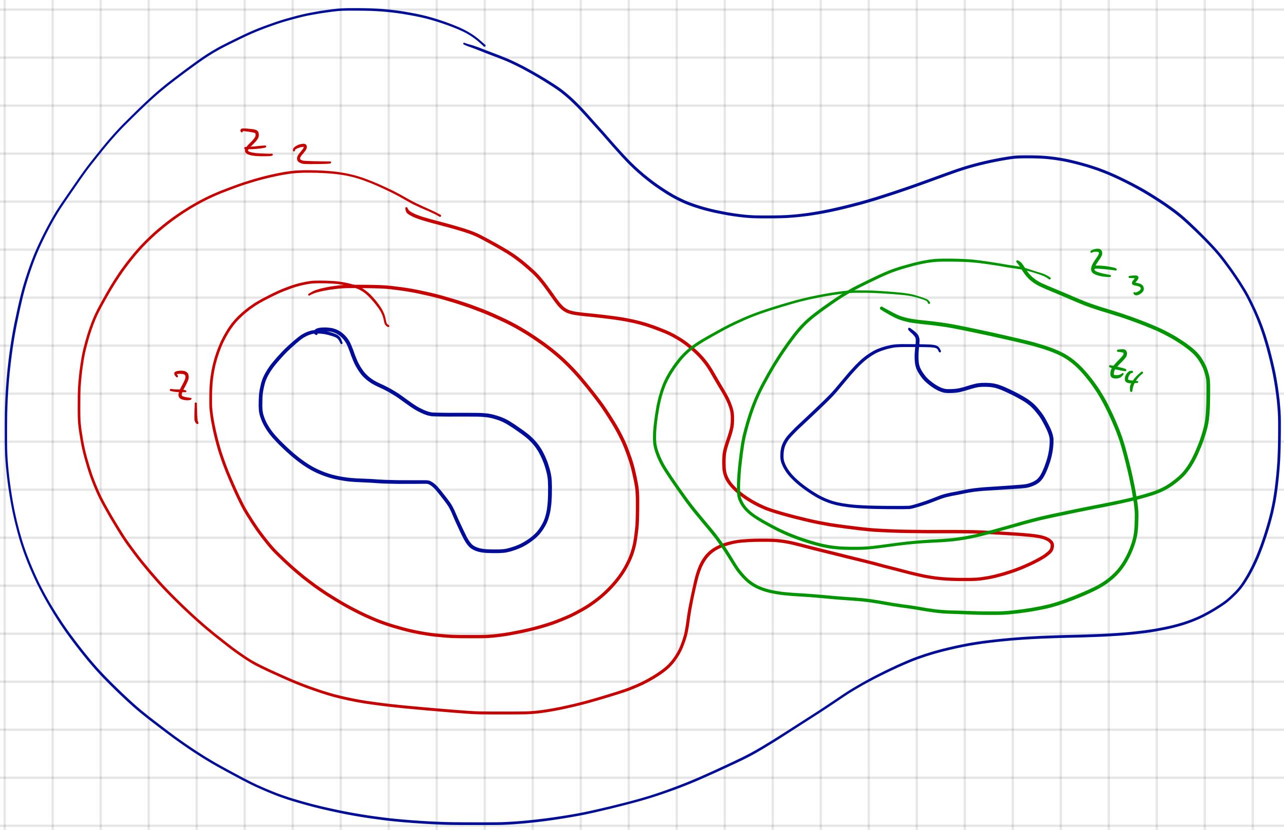
	abd
bd + cd + bc	1+0+0
ad+cd+bc+ab	1+0+0+1

$\Rightarrow$

	ab
z_1	1
z_2	0

Each simplex bounds 1 cycle
(if our chosen basis is the sum of multiple cycles then we need to make a choice)

Picture of homologous 1-cycles



* $z_1 \sim z_2 \Rightarrow$ we can write $z_1 + \sum_{c_i \in \operatorname{im} \partial_1} \lambda_{c_i} c_i$

* z_1 is not homologous to $z_2 + z_3$

  $\Rightarrow$ because $z_3 \notin \operatorname{im} \partial_1$ (by definition)

Euler Characteristic & Homology

Recall $\chi = \sum_{i=0}^{d} (-1)^i |K_d|$

Now: $|K_d| = rk(C_k)$ (since C_k is just the identity matrix indexed by k-simplices)

By the rank nullity theorem.

$$rk(C_k) = rk(\ker \partial_k) + rk(\operatorname{im} \partial_k)$$

So

$$\chi = \sum_{i=0}^{d} (-1)^i rk(C_k) = \sum_{i=0}^{d} (-1)^i (rk(\ker \partial_k) + rk(\operatorname{im} \partial_k))$$

$$= \sum_{i=0}^{d} (-1)^i (rk(\ker \partial_k) - rk(\operatorname{im} \partial_{k+1})) + rk(\operatorname{im} \partial_0)$$

$$= \sum_{i=0}^{d} (-1)^i rk(H_i) = \sum_{i=0}^{d} (-1)^i \beta_i$$

Oddly, computing alternating sums of Betti numbers is much easier than computing Betti numbers.

Persistent Homology

Until now, we have had only one space Δ

Now we consider a filtration

$$\Delta_1 \subseteq \Delta_2 \subseteq \dots \subseteq \Delta_n$$

What is the homology of a filtration?

Digression: What are some typical filtrations we will look at?

1) Functions on simplicial complexes

$$f : \Delta \to \mathbb{R}$$

(we assume for simplicity, that f is constant on each simplex)

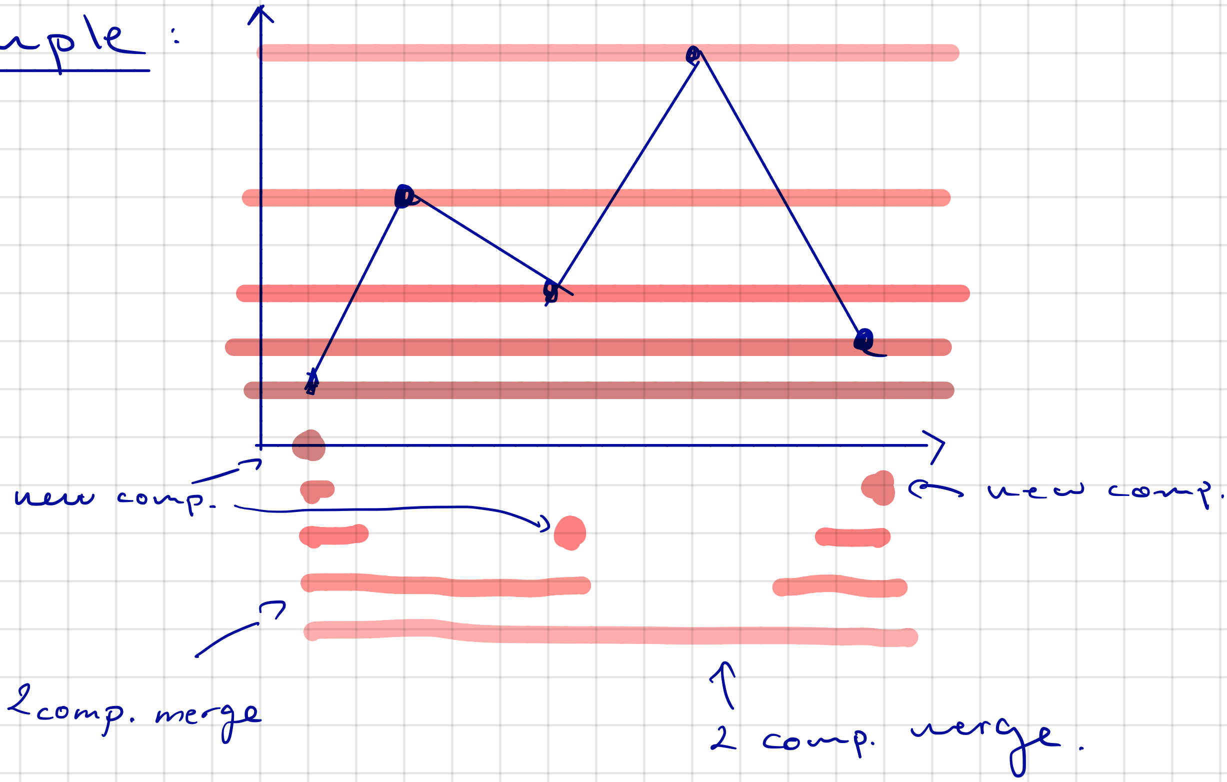
The fitration $f^{-1}(-\infty, \alpha]$ is called a lower-star fitration.

We require that $f^{-1}(-\infty, \alpha]$ is a simplicial complex (Edelsbrunner & Harer call this a monotone function)

We will not prove it here - but there is a closely related notion - the sublevel set filtration.

Intuition: Track homological features over the filtration.

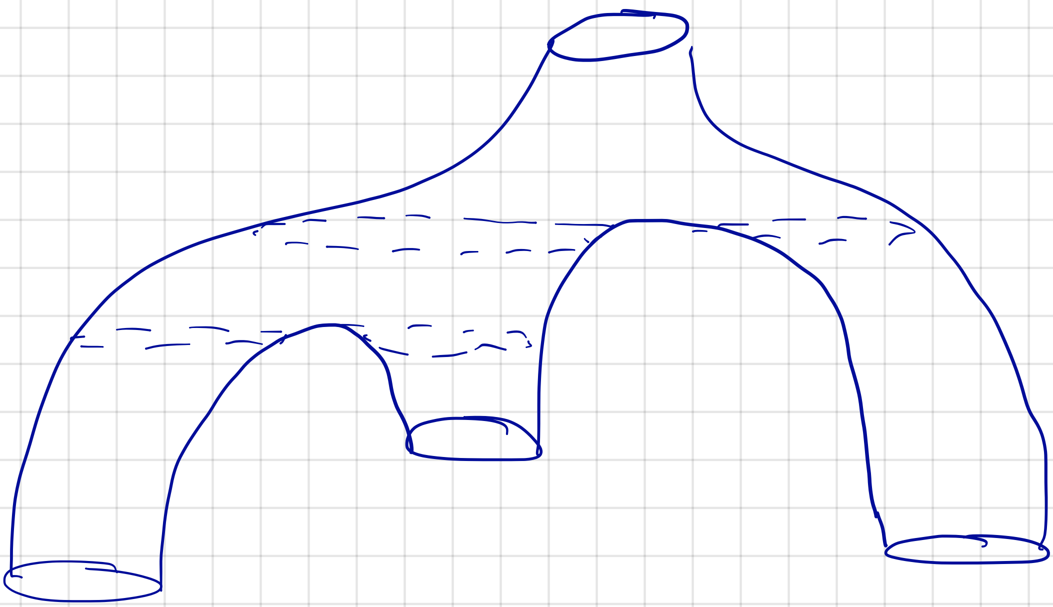
Example:



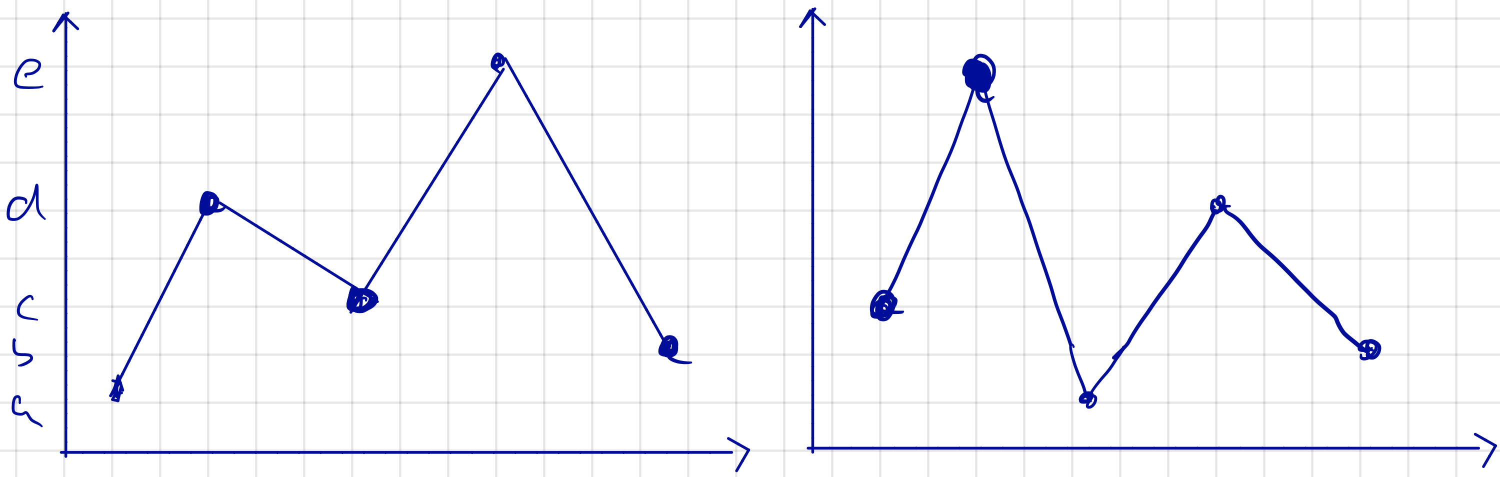
* Features can be born $(\mathrm{rk}(H_k)+1)$

  Features can die $(\mathrm{rk}(H_k)-1)$

Same example for H_1 (with apologies)



Note: There is more information than just +1/-1 of the rank

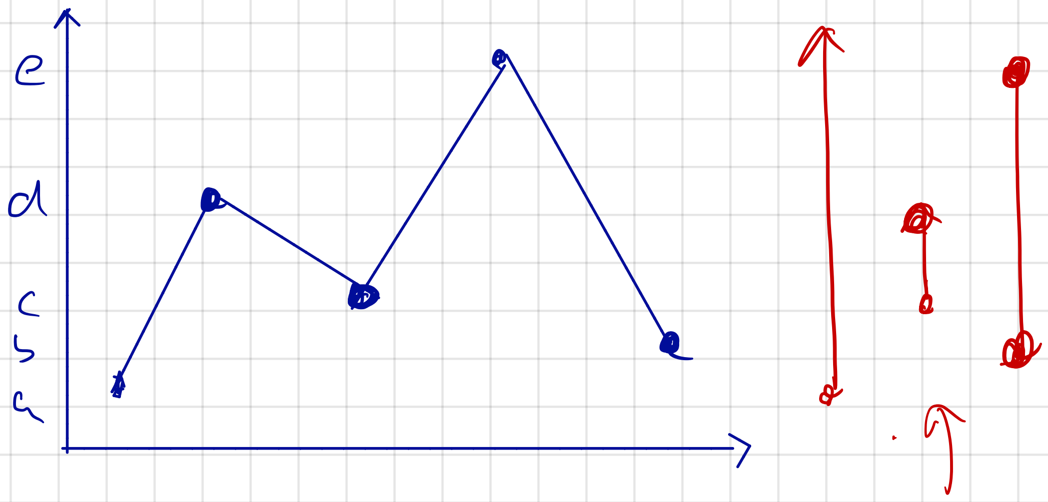


Notice +/- ranks of components is the same but what happens if we track how long live?

Key Fact: When 2 components/cycles merge, we kill the youngest one, ie the one which was born last.

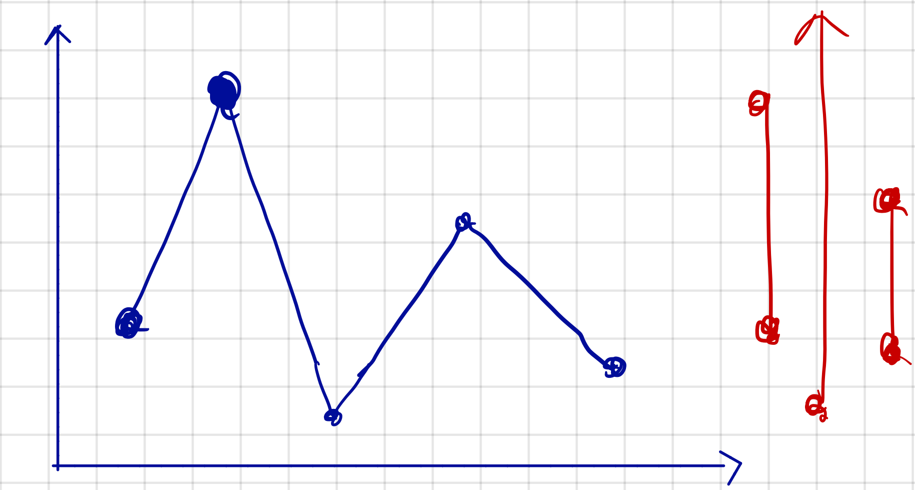
This is called the Elder rule
(There is a good algebraic reason for this)

Case 1:



different

Case 2



In red we see barcodes - notice they distinguish between the 2 functions

Remark: A completely general characterisation of what the "barcode" detects is still open.

Formal Definitions

Given a filtration of simplicial complexes

$$\Delta_1 \subseteq \Delta_2 \subseteq \Delta_3 \subseteq \cdots \subseteq \Delta_n$$

This gives rise to an increasing sequence of chain groups for all k

$$C_k(\Delta_1) \hookrightarrow C_k(\Delta_2) \hookrightarrow C_k(\Delta_3) \hookrightarrow \cdots \hookrightarrow C_k(\Delta_n)$$

This is called the persistent chain complex

$$\begin{tikzcd}
\; \arrow[d] & \; \arrow[d] & \; \arrow[d] & & \; \arrow[d] \\
C_{k+1}(\Delta_1) \arrow[r, hook] \arrow[d, "\partial_{k+1}"] & C_{k+1}(\Delta_2) \arrow[r, hook] \arrow[d, "\partial_{k+1}"] & C_{k+1}(\Delta_3) \arrow[r, hook] \arrow[d, "\partial_{k+1}"] & \cdots \arrow[r, hook] & C_{k+1}(\Delta_n) \arrow[d, "\partial_{k+1}"] \\
C_k(\Delta_1) \arrow[r, hook] \arrow[d, "\partial_k"] & C_k(\Delta_2) \arrow[r, hook] \arrow[d, "\partial_k"] & C_k(\Delta_3) \arrow[r, hook] \arrow[d, "\partial_k"] & \cdots \arrow[r, hook] & C_k(\Delta_n) \arrow[d, "\partial_k"] \\
C_{k-1}(\Delta_1) \arrow[r, hook] \arrow[d] & C_{k-1}(\Delta_2) \arrow[r, hook] \arrow[d] & C_{k-1}(\Delta_3) \arrow[r, hook] \arrow[d] & \cdots \arrow[r, hook] & C_{k-1}(\Delta_n) \arrow[d] \\
\; & \; & \; & & \;
\end{tikzcd}$$

Each column is a chain complex & $\hookrightarrow$ denotes monomorphisms

Applying homology in each column

$$C_k(\Delta_1) \hookrightarrow C_k(\Delta_2) \hookrightarrow C_k(\Delta_3) \hookrightarrow \cdots \hookrightarrow C_k(\Delta_n)$$

$$\Downarrow$$

$$H_k(\Delta_1) \longrightarrow H_k(\Delta_2) \longrightarrow H_k(\Delta_3) \longrightarrow \cdots \longrightarrow H_k(\Delta_n)$$

<u>Note</u>: Usually in topology we compute homology over $\mathbb{Z}$ (ie we treat the entries in ∂_k as integers) but we will treat them as $\mathbb{Z}_2$ (or $\mathbb{Z}_p$ for p prime).

This means $H_k(\Delta_i)$ are vector spaces & $H_k(\Delta_i) \longrightarrow H_k(\Delta_j)$ are linear maps.

<u>Remark 1</u>: The linear maps are induced from the inclusions on the chain groups

<u>Ex</u>: Check that the maps are well defined.

<u>Hint</u>: $C_k(\Delta_i) \hookrightarrow C_k(\Delta_j)$

$\Rightarrow Z_k(\Delta_i) \hookrightarrow Z_k(\Delta_j)$

$\Rightarrow B_k(\Delta_i) \hookrightarrow B_k(\Delta_j)$

Remark 2: The linear maps are just matrices again

Remark 3: Though the chain maps are monomorphisms, this is not the case for the linear maps.

Def.: A persistence module is the collection of vector spaces and maps between them:

$$\{H_k(\Delta_i)\}_{i \in \mathbb{Z}} \; ; \; H_k(\Delta_i) \to H_k(\Delta_j) \;\; \forall i,j$$

This is just:

$$H_k(\Delta_1) \longrightarrow H_k(\Delta_2) \to H_k(\Delta_3) \longrightarrow \dots \to H_k(\Delta_n)$$

One thing which makes persistence useful

Module $\longleftrightarrow$ Diagram / Barcode

Def: A barcode is a decomposition of a persistence module $P_k(\Delta)$

$$P_k(\Delta) \cong I_1^k(t) \oplus I_2^k(t) \oplus \cdots \oplus I_N^k(t)$$

such that each

$$I_j(t) = \begin{cases} \text{rank } 1 & b_j \le t < d_j \quad b_j, d_j \in \mathbb{R} \\ 0 & \text{otherwise} \end{cases}$$

(These are called interval modules)

i.e.

$$rk\left(H_k(\Delta_s) \to H_k(\Delta_t)\right) = \sum_{j=1}^{N} \min I_j(t) \cap [s,t)$$

So we decompose into "intervals" such that the rank of any map is the sum of intervals which "span" the map (i.e. the interval contains the end points of the map)

Why does this exist?

Short answer:

Gabriel's Thm (Representation theory)

A quiver of the form

$$\bullet \hookrightarrow \bullet \hookrightarrow \bullet \hookrightarrow \cdots \hookrightarrow \bullet$$

admits a decomposition.

More constructive viewpoint:

Say we build an interval decomposition incrementally:

$$H_k(\Delta_1) \xrightarrow{f_{12}} H_k(\Delta_2) \xrightarrow{f_{23}} H_k(\Delta_3)$$

Note: $f_{12} = f_{23} \circ f_{12}$

$$\begin{array}{ccccc} H_k(\Delta_1) & & H_k(\Delta_2) & & H_k(\Delta_3) \\ \wr\wr & & \wr\wr & & \\ \ker f_{12} \oplus \operatorname{coim} f_{12} & \longrightarrow & \operatorname{im} f_{12} \oplus \operatorname{coker} f_{12} & & \\ & & \wr\wr & & \\ & & \ker f_{23} \oplus \operatorname{coim} f_{23} & \longrightarrow & \operatorname{im} f_{23} \oplus \operatorname{coker} f_{23} \end{array}$$

Decomp: w/ slight abuse

a) $\ker f_{12} \simeq [1,2)$

b) $\operatorname{im} f_{12} \cap \ker f_{23} \simeq [1,3)$

c) $\operatorname{im} f_{12} \cap \operatorname{coim} f_{23} \simeq [1,3]$

$\operatorname{coker} f_{12} \cap \ker f_{23} \simeq [2,3)$

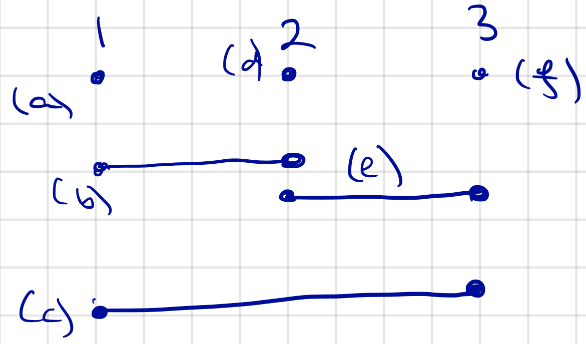
$\operatorname{coker} f_{12} \cap \operatorname{coim} f_{23} \simeq [2,3]$

$\operatorname{coker} f_{23} \simeq [3]$

Remark

$[i,j)$ is usually considered as $[i, j-1]$

Picture:



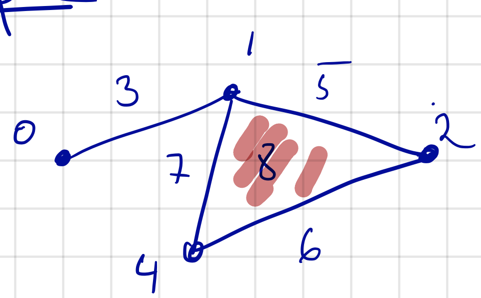
Observation: All the information is contained in ranks of all the vector spaces and <u>all</u> maps. This makes sense since the rank determines (up to isomorphism) a vector spaces

Computation

Good news: No harder than ordinary homology

Idea: Ordered Gaussian elimination

Example:



* We will refer to simplices by function value

$\longrightarrow$ increasing

∂_0 7 $\leftarrow$ cycle

increasing $\downarrow$

$$\begin{array}{c|cccc} & 3 & 5 & 6 & 7 \\ \hline 0 & 1 & 0 & 0 & 0 \\ 1 & \boxed{1} & 1 & 0 & 1 \\ 2 & 0 & \boxed{1} & 1 & 0 \\ 4 & 0 & 0 & \boxed{1} & 1 \end{array}$$

$\boxed{1}$ are the intervals

$[1,3]$, $[2,5]$, $[4,6]$

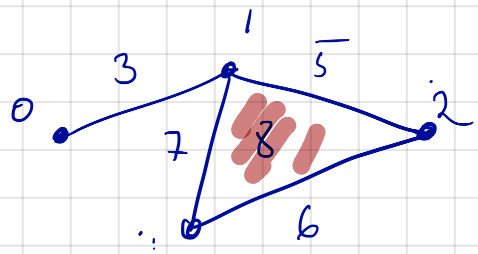
$[0,\infty)$ - unmatched rows are infinite bars

Carlsson-Zomorodian showed that we can restrict ∂_k to just include rows which create cycles

$$\begin{array}{c|c} \partial_1 & 8 \\ \hline 7 & \boxed{1} \end{array} \longrightarrow [7,8]$$

This restriction makes computation easier but it is important to keep in mind why this works.

<u>Algebraic interpretation</u>: Treat coefficients as monomials in 1 variable, w/ powers encoding time



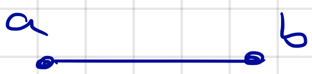
∂

	3	5	6	7
0	t^3	0	0	0
1	t^2	t^4	0	t^6
2	0	t^3	t^4	0
4	0	0	t^2	t^3

The power is the difference between edge "time" & vertex "time"

* Multiplication by t moves things forward in time

<u>Example</u>:



$f(a) = 1$ $\quad f(ab) = 3$

$f(b) = 2$

$$\partial(ab) = t^2 a + tb$$

<u>Notice</u>: these polynomials keep track of "time"

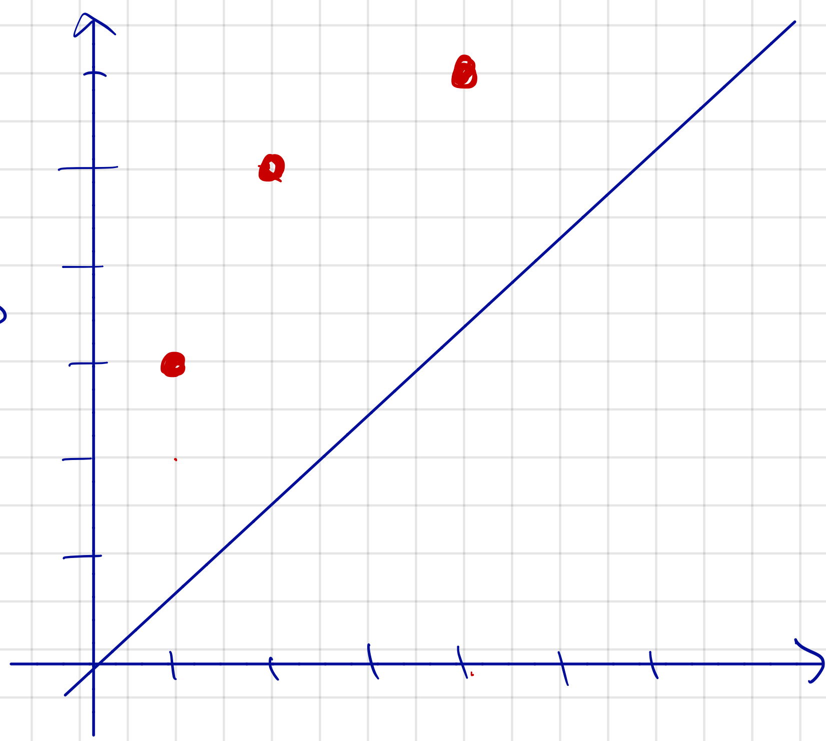
<u>Carlsson-Zomordian</u>: Persistent homology is homology over monomials w/ field coefficients ($\mathbb{Z}_p[t]$)

Upshot: $\mathbb{Z}_p[t]$ is a principle ideal domain, so a persistence module is a module over a p.i.d & admits a decomposition into a free part (infinite bars) & torsion (finite bars) * these are precisely the interval modules.

From barcodes to diagrams

It will often be useful to present barcodes as a diagram. For each interval draw a point w/ the start point on the x-axis & the end point on the y-axis

$[1,3], [2,5], [4,6]$ ⟶

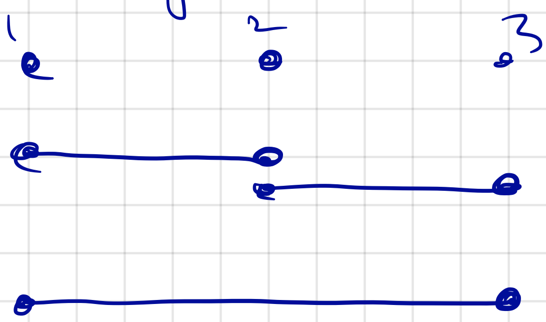


Diagrams will be useful for stability.

But first,

1 more interpretation:

Recall, all the information we want is in the rank of all the maps. For a filtration of length 3, this is represented by



Hence, we need 6 integers (ranks)
This is called the rank function.

$$R : I \to \mathbb{Z}$$

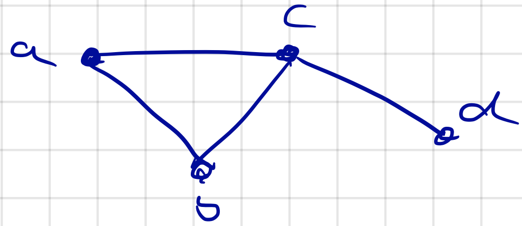
$\boxed{\text{$R$ is an integer valued function on the space of all possible intervals}}$

Patel 2016

The persistence diagram is the Möbius inversion of R.

* This is just inclusion-exclusion

Example



$f(a) = 0$

$f(b) = 2$

$f(c) = 1$

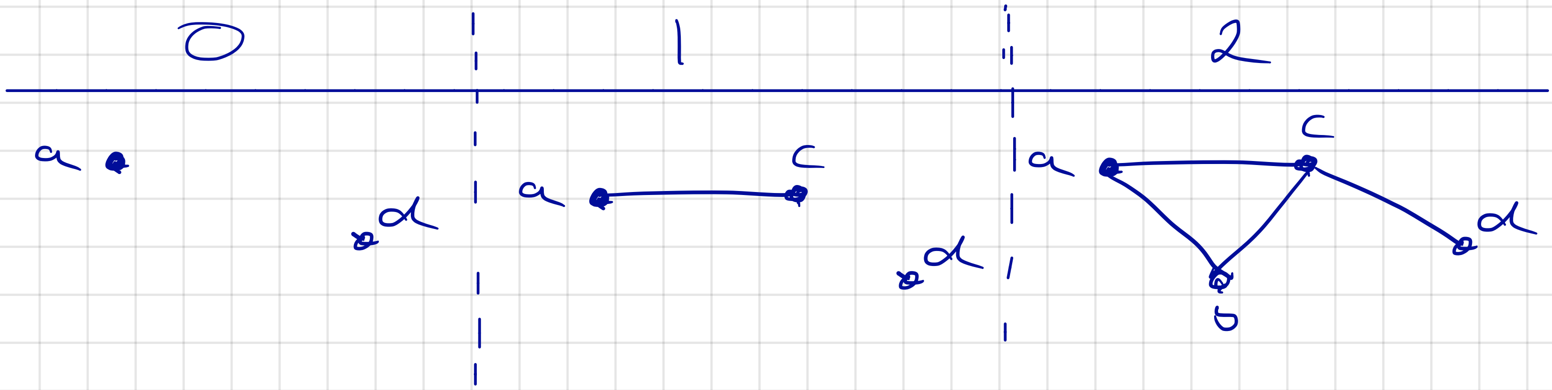
$f(d) = 0$

$f(ab) = 2$

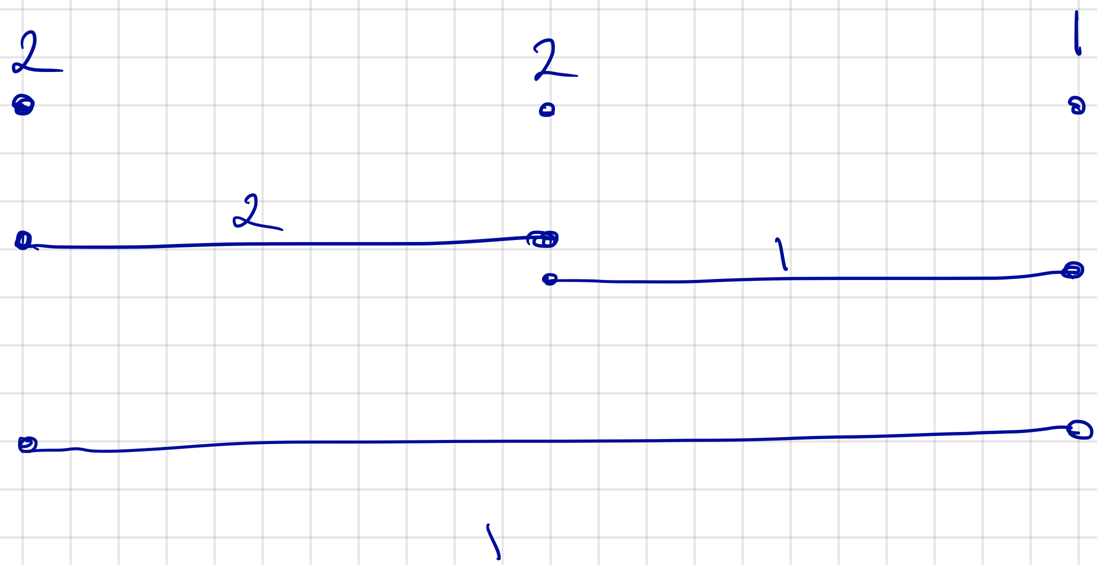
$f(ac) = 1$

$f(cb) = 2$

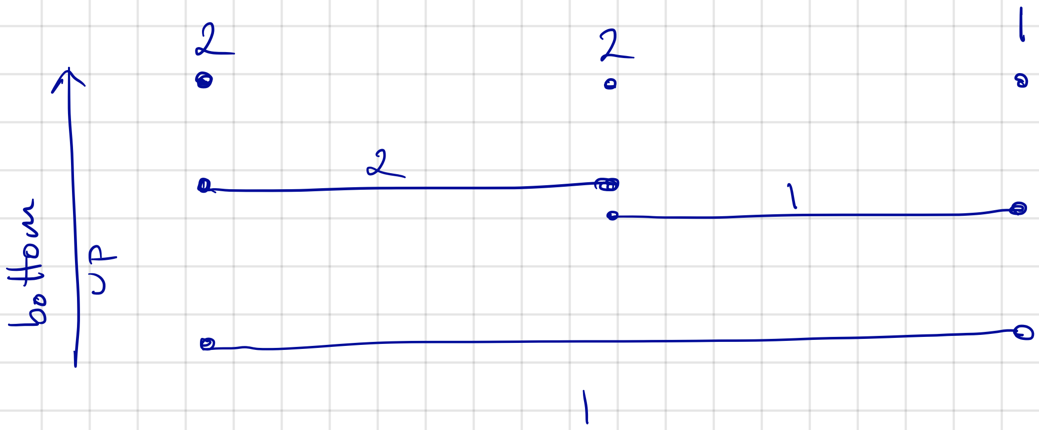
$f(cd) = 2$

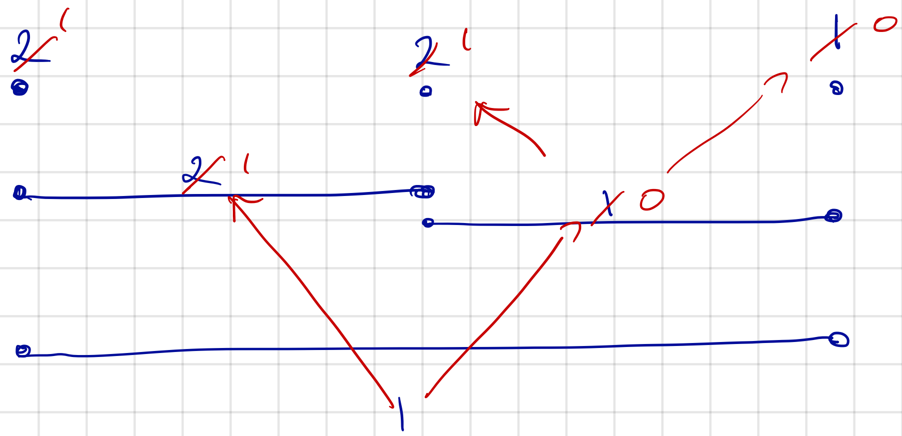


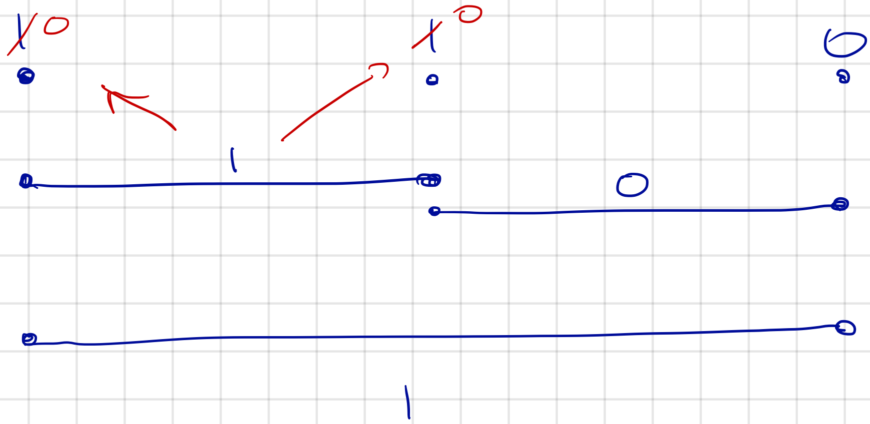
Rank



Inclusion-Exclusion







<u>Final result</u> : [0, 2]
[0, 1]