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)

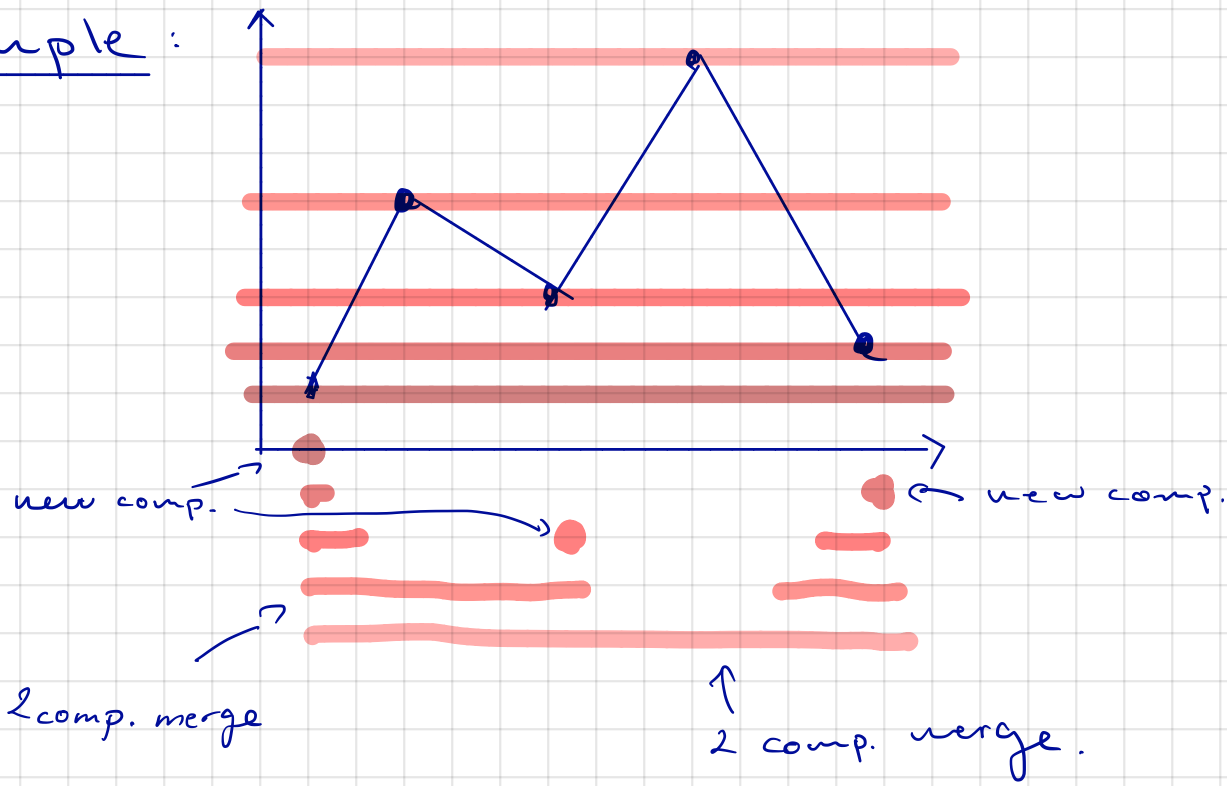
The fitration $f^{-1}(-\infty, \alpha]$ is called a lower-star fitration.

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.

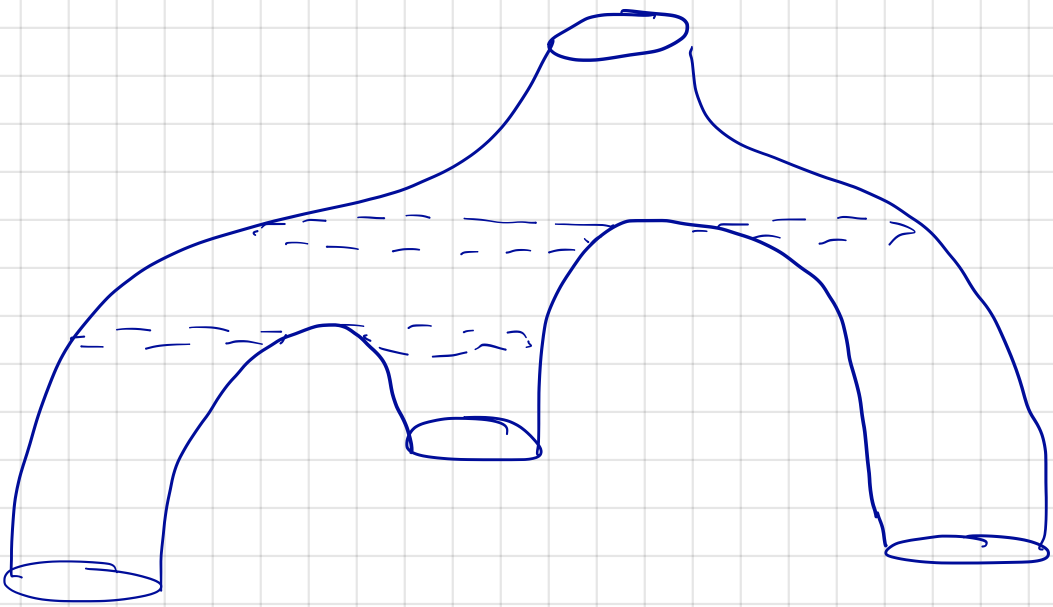
Example:



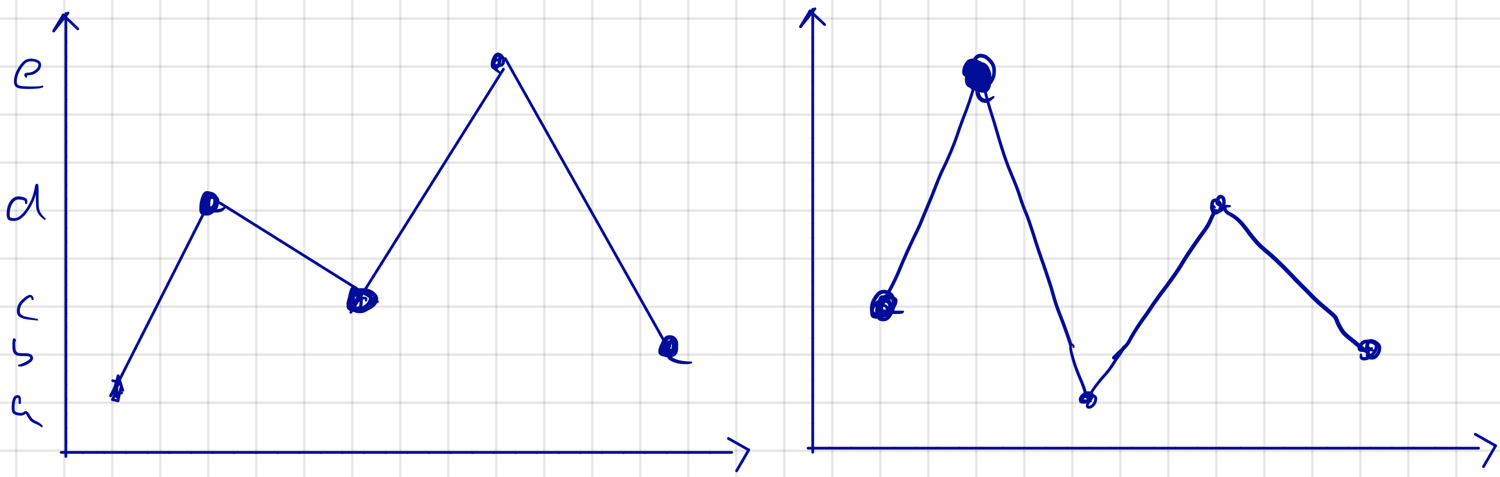
* Features can be born ($\mathrm{rk}(H_k)+1$)

  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

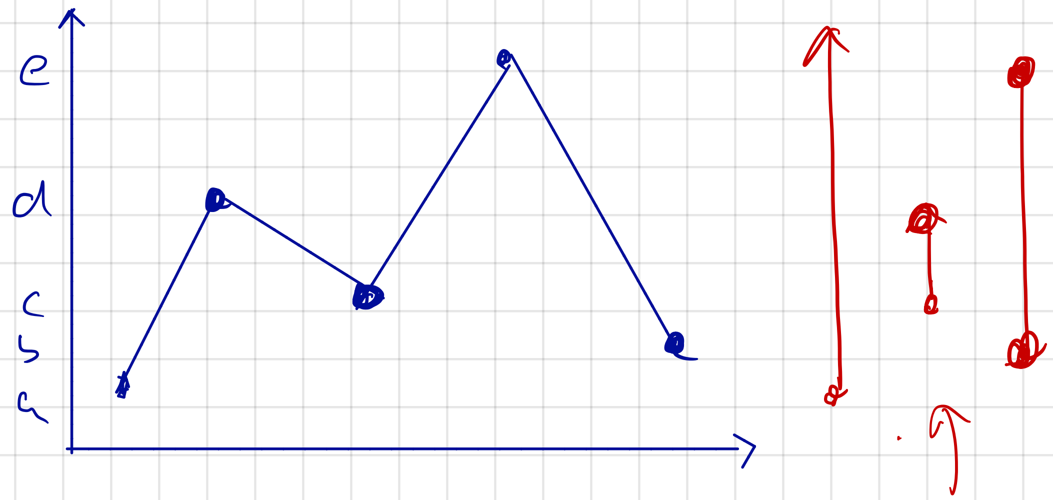


Notice +/- ranks of components is the same but what happens if we track how long live?

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

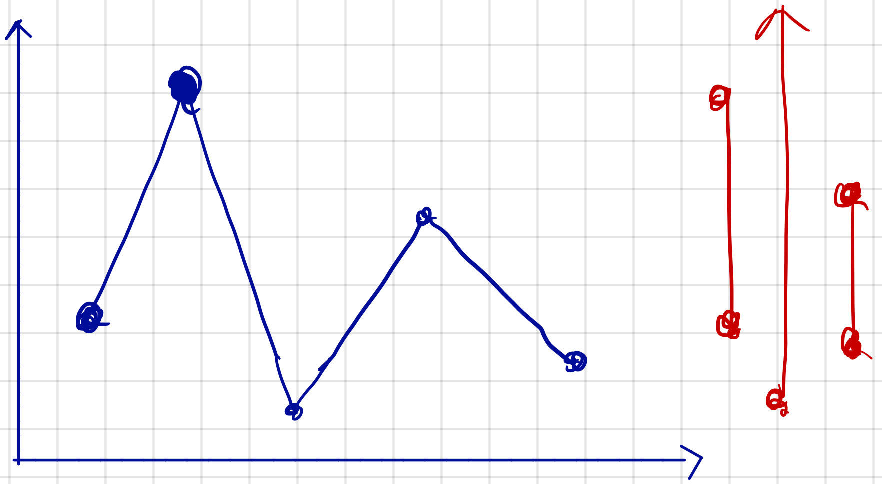
This is called the Elder rule
(There is a good algebraic reason for this)

Case 1 :



different

Case 2



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

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

Formal Definitions

Given a filtration of simplicial complexes

$$\Delta_1 \subseteq \Delta_2 \subseteq \Delta_3 \subseteq \dots \subseteq \Delta_n$$

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

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

This is called the persistent chain complex

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

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

$$\text{Module} \longleftrightarrow \text{Diagram / Barcode}$$

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

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

(These are called interval modules)

&

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

So we decompose into "intervals" such that the rank of any map is the sum of intervals which "span" the map (ie 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: $f_{12} = f_{23} \circ f_{12}$

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

Decomp: w/ slight abuse

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

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

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

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

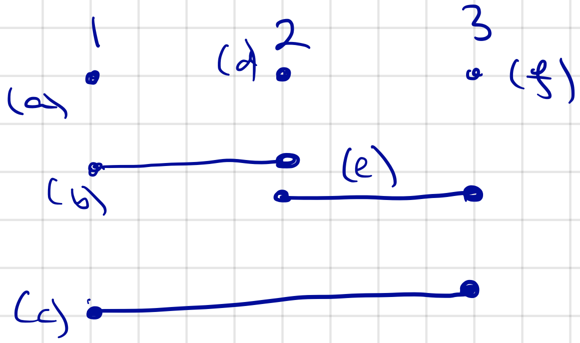
$\operatorname{coker} f_{12} \cap \operatorname{coim} f_{23} \cong [2,3]$

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

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 all maps. This makes sense since the rank determines (up to isomorphism) a vector spaces

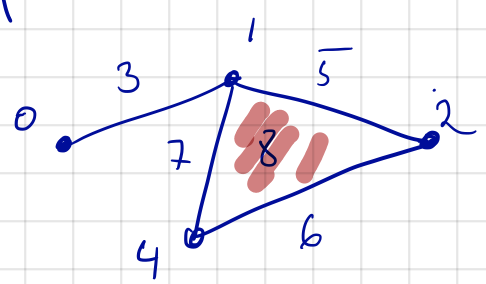
Computation

Note we did not explicitly cover this in lecture

Good news: No harder than ordinary homology

Idea: Ordered Gaussian elimination

Example:



* We will refer to simplices by function value

increasing →

$$
\partial_0 \quad
\begin{array}{c|cccc}
 & 3 & 5 & 6 & 7 \\
\hline
0 & 1 & 0 & 0 & 0 \\
1 & \boxed{1} & 1 & 0 & 1 \\
2 & 0 & \boxed{1} & 1 & 0 \\
4 & 0 & 0 & \boxed{1} & 1
\end{array}
$$

7 ← cycle

increasing ↓

$\boxed{\;}$ are the intervals

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

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

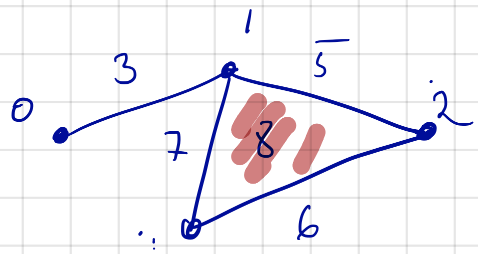
Carlsson-Zomorodian* showed that we can restrict ∂_k to just include rows which create cycles

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

* Carlsson-Zomorodian - Computing Persistent Homology (see problem set for link)

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



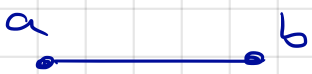
∂

	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>:



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

$f(b)=2$

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

<u>Notice</u>: these polynomials keep track of "time" (can only multiply by t not divide - ie you cannot go back in time)

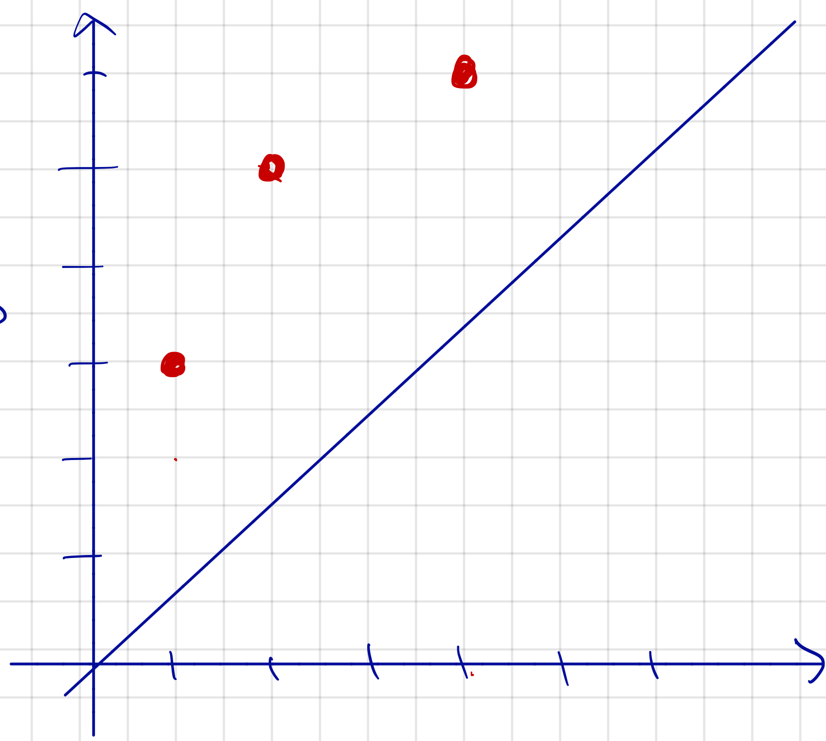
<u>Carlsson-Zomordian</u>: Persistent homology is homology over monomials w/ field coefficients $(\mathbb{Z}_p[t])$

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

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

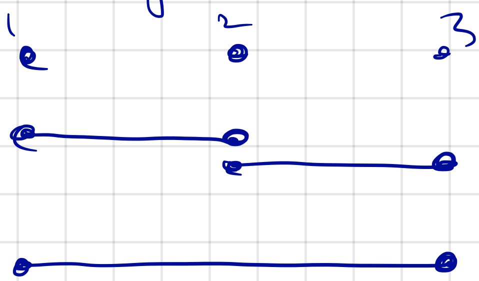


Diagrams will be useful for stability.

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.

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

$$R\{(a,b)\} = \frac{Z_a}{Z_a \cap B_b} = rk(\mathrm{im}\, H(a) \to H(b))$$

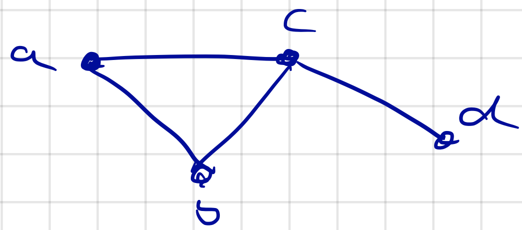
$\boxed{\text{R is an integer valued function on the space of all possible intervals}}$

Patel 2016/2018 - Generalized Persistence Diagrams

The persistence diagram is the Möbius inversion of R.

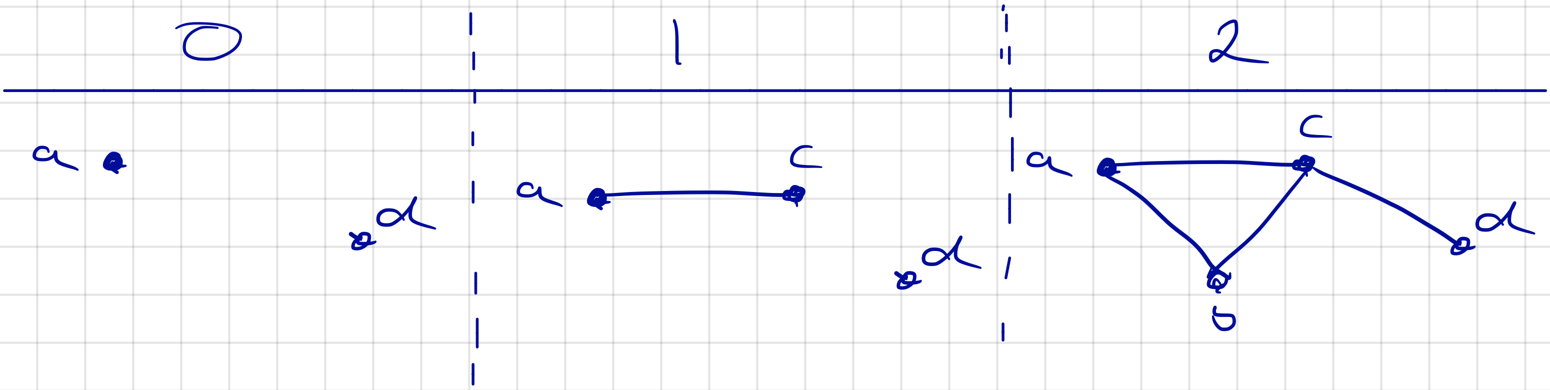
* This is just inclusion-exclusion

Example

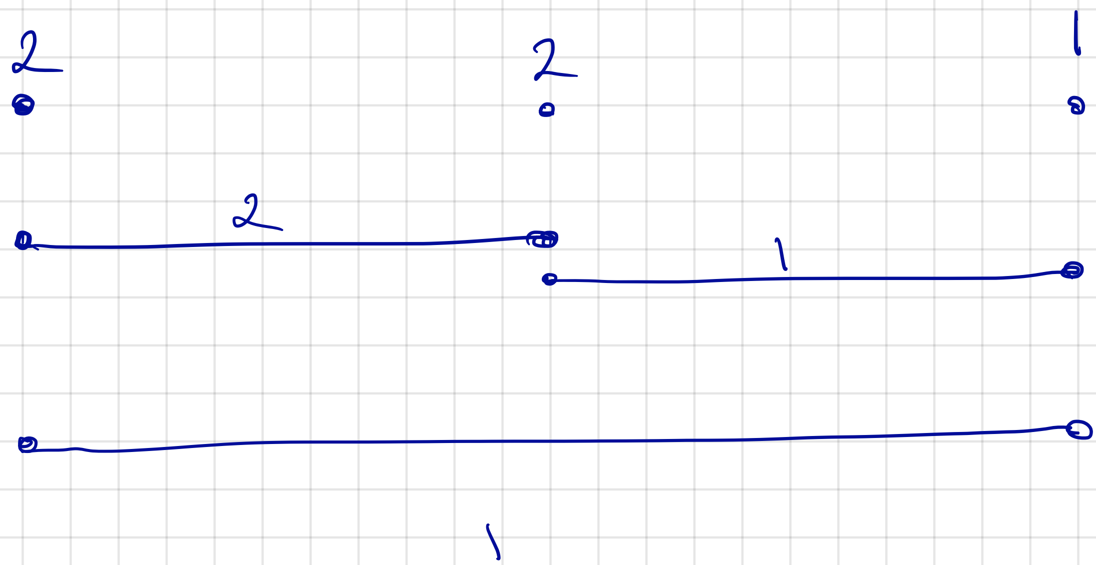


$f(a) = 0$
$f(b) = 2$
$f(c) = 1$
$f(d) = 0$

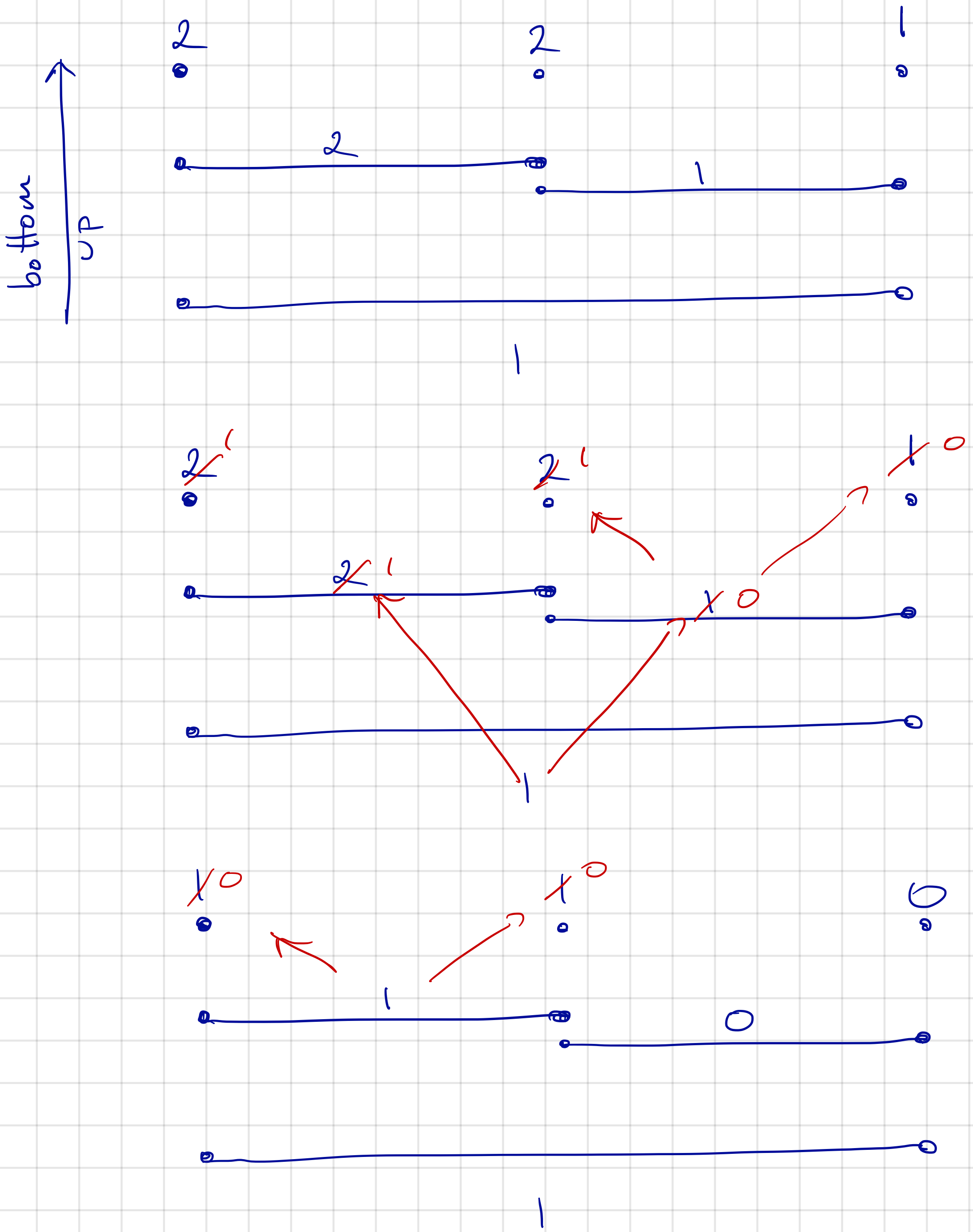
$f(ab) = 2$
$f(ac) = 1$
$f(cb) = 2$
$f(cd) = 2$



Rank



Inclusion-Exclusion



Final result: $[0,2]$

$[0,1]$

Remark: The rank of the image $H_k(X_r) \to H_k(X_s)$ is called the (r,s)-persistent Betti number, denoted $\beta_k^{r,s}$

Summing Up

Homology of a filtration

$$\Delta_1 \subseteq \Delta_2 \subseteq \Delta_3 \subseteq \dots \subseteq \Delta_n$$

$$\Downarrow$$

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

$$\Downarrow$$

$$\bigoplus_i \mathbb{I}[b_i, d_i)$$

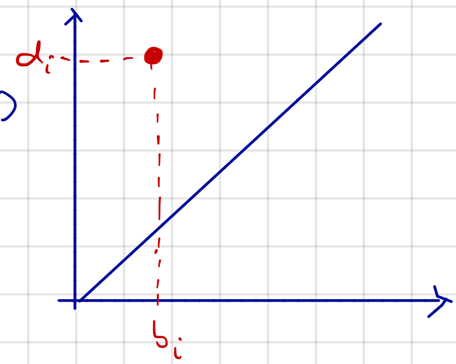
filtration

$\downarrow$

persistence module

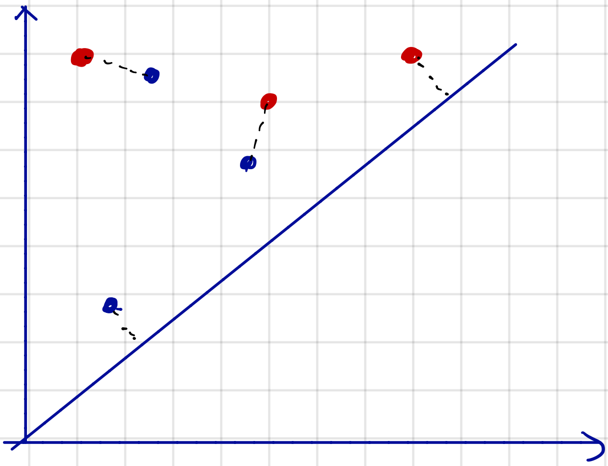
$\downarrow$

barcode/diagram



Distances between diagrams

Say we have diagrams D_1 & D_2

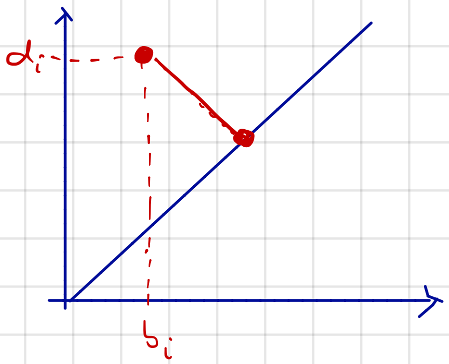


bottleneck distance: $d_B(D_1, D_2) = \inf_{\pi: D_1 \to D_2} \sup_{p \in D_1} \|p - \pi(p)\|_\infty$

Key points

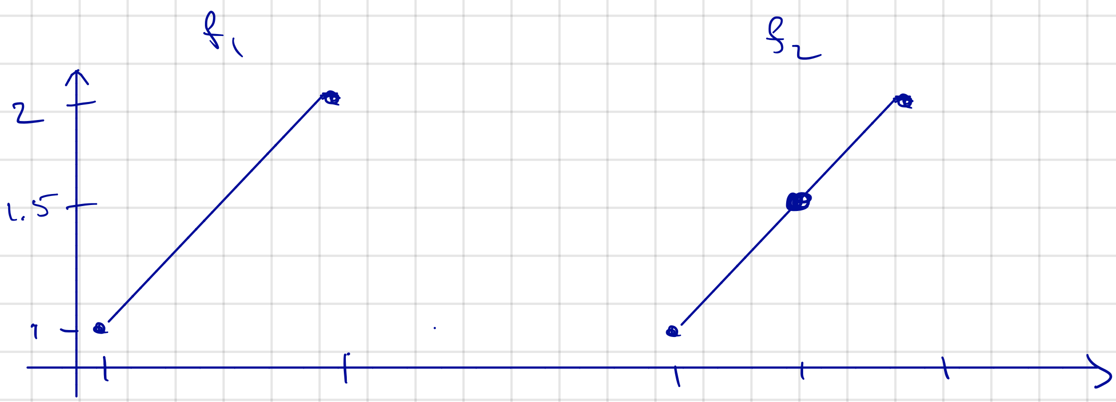
Bottleneck distance minimizes over all possible bijections between diagrams

Since different diagrams can have different numbers of points ⇒ we can project to the diagonal.

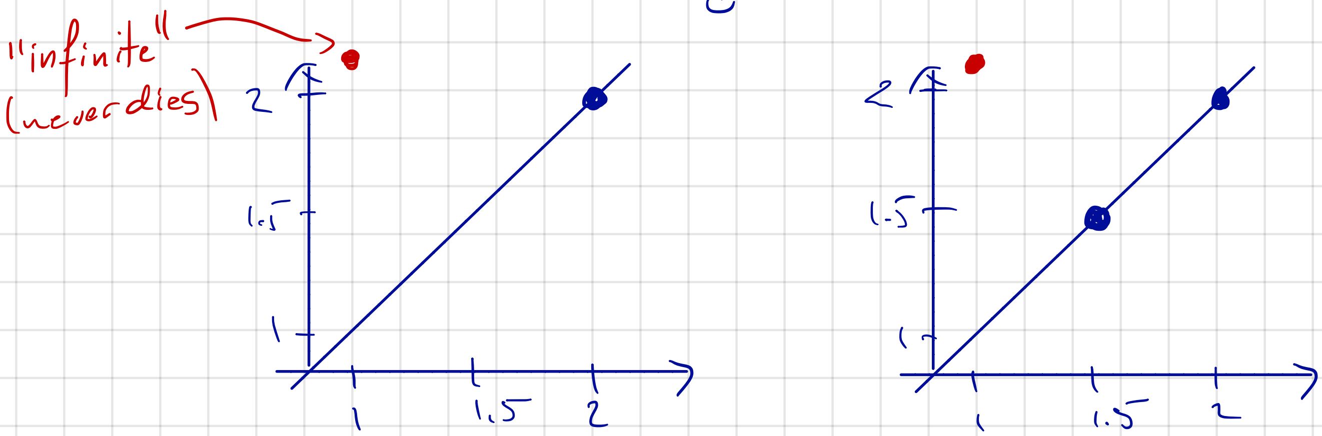


Why does this make sense?

- Diagonal represents things which are <u>not</u> persistent (i.e. are born and immediately die)



⇓



<u>Remark</u>: Points on the diagonal do not change the topology.

Stability

Let $f, g : X \to \mathbb{R}$ & $F_\alpha := f^{-1}(-\infty, \alpha]$
$G_\alpha := g^{-1}(-\infty, \alpha]$

if $\|f-g\|_\infty \leq \varepsilon$ then $F_\alpha \subseteq G_{\alpha+\varepsilon} \subseteq F_{\alpha+2\varepsilon}$

Def: If the following diagrams commute

$$\begin{tikzcd} H_k(F_\alpha) \arrow[rr] \arrow[dr] & & H_k(F_{\alpha+2\varepsilon}) \\ & H_k(G_{\alpha+\varepsilon}) \arrow[ur] & \end{tikzcd}$$

$$\begin{tikzcd} & H_k(F_{\alpha+\varepsilon}) \arrow[dr] & \\ H_k(G_\alpha) \arrow[ur] \arrow[rr, dash] & & H_k(G_{\alpha+2\varepsilon}) \end{tikzcd}$$

$$\begin{tikzcd} H_k(F_\alpha) \arrow[r] \arrow[d] & H_k(F_{\alpha+\delta}) \arrow[d] \\ H_k(G_{\alpha+\varepsilon}) \arrow[r] & H_k(G_{\alpha+\varepsilon+\delta}) \end{tikzcd}$$

$$\begin{tikzcd} H_k(F_{\alpha+\varepsilon}) \arrow[r] & H_k(G_{\alpha+\varepsilon+\delta}) \\ H_k(G_\alpha) \arrow[r] \arrow[u] & H_k(G_{\alpha+\delta}) \arrow[u] \end{tikzcd}$$

we say F & G are ε-interleaved. $(F \sim_\varepsilon G)$

Functoriality

Note that space level (chain) interleaving implies homological interleaving. This is a consequence of functoriality.

Homology is a functor:

Each space $\longmapsto$ Vector space (more generally an abelian group)

Maps between spaces $\longmapsto$ map between vector spaces

Space/chain level interleaving is a stronger condition than homological interleaving.

Remark: Category theory & Functoriality are often a useful toolbox/language for topological problems.

$$\begin{array}{ccccc} \text{Filtration} & \xrightarrow{\text{functoriality}} & \text{Persistence Module} & \longrightarrow & \text{Persistence Diagrams} \\ \|f-g\|_\varepsilon & & F \sim_\varepsilon G & ? & d_B(F,G) < \varepsilon \end{array}$$

Stability of Persistence Diagrams

Stability of Persistence Diagrams
Cohen Steiner, Edelsbrunner, Harer 2006

Theorem: Let $f, g : \Delta \to \mathbb{R}$. If $\|f-g\|_\infty \leq \varepsilon$

then $d_B(\mathrm{Dgm}(f), \mathrm{Dgm}(g)) \leq \varepsilon$
where $d_B(\cdot)$ is the bottleneck distance
& $\mathrm{Dgm}(f)$ & $\mathrm{Dgm}(g)$ are the persistence diagrams of the sublevel sets filtrations of f & g respectively.

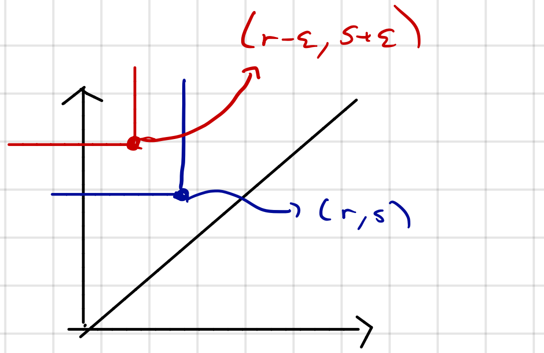
Note 1: The key point is the interleaving - this has been proven since the original result in much greater generality (we will only use the function setting at one point in the proof

Note 2: Recall that $\|f-g\|_\infty \leq \varepsilon$ implies the sublevel set persistence modules are ε-interleaved.

Next Time – Preview

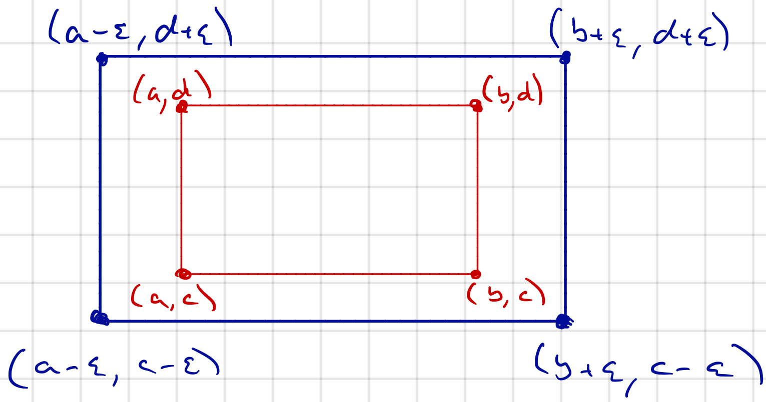
Outline:

1) Quadrant Lemma



$F \sim_\varepsilon G \implies \boxed{\beta_k^{r,s}(F) \geq \beta_k^{r-\varepsilon, s+\varepsilon}(G)}$

2) Box Lemma



$\square$: # points in blue rectangle for F

$\square$: # points in red rectangle for G

$\boxed{\square \leq \square}$

3) Easy Bijection Lemma & Intepolation

if ε is small enough

$\boxed{\square = \square}$

intepolate between f & g in small enough steps