"Roadmap" Paper

persistent homology advantages
- ① based on algebraic topology
- ② computable via linear algebra
- ③ robust wrt small perturbations in input data.

Given a topological space S_ε → approximate by a simplicial complex

↓ homology

~~To~~ ~~meant~~

features : # components, holes, voids.

↓ ~~barcode~~

lifetime of features represented by barcode
- left endpoint rep birth of feature
- right endpt rep death of feature

Homology associates one vector space $H_i(X)$ to a space X for each $i \in \mathbb{N}$.

- $H_0(X)$ = # path components in X
- $H_1(X)$ = # holes
- $H_2(X)$ = # voids.

~~Theorem~~

Key steps summary :

① define simplicial complex

② find boundary maps $\partial_n : \tilde{C}_n(\Delta) \to \tilde{C}_{n-1}(\Delta)$.

matrices. ∂_n with the ~~comp~~ n-simplices as basis.

e.g. $\partial_1 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$

NOTE : make sure always that $\partial_n \circ \partial_{n+1} = 0 \quad \forall n$

③. compute homology group. H_n : nth homology group.

$H_n(\Delta) = \ker(\partial_n) / \operatorname{Im}(\partial_{n+1})$ for each n.

④. compute Betti numbers from $\dim(H_n(\Delta))$.

Persistent Homology :

Overview of the steps: ① Data.

↓

② Filtered complex

↓

③ Barcodes

↓

④ Interpretation.

The key insight of persistence hom:
Consider several possible ε. As ε increases, add simplices to the complexes.
Then detect which features "persist" as ε increases.

(with a slight change of notation.
$H_n \to H_k$.)

Recall that the kth homology group is $H_k = \underset{\text{cycle group}}{Z_k} / \underset{\text{boundary group}}{B_k}$.
The elements in H_k are classes of homologous cycles.

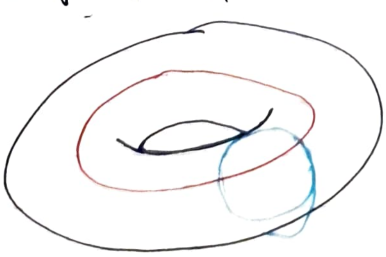
- **Reduction Algorithm.**

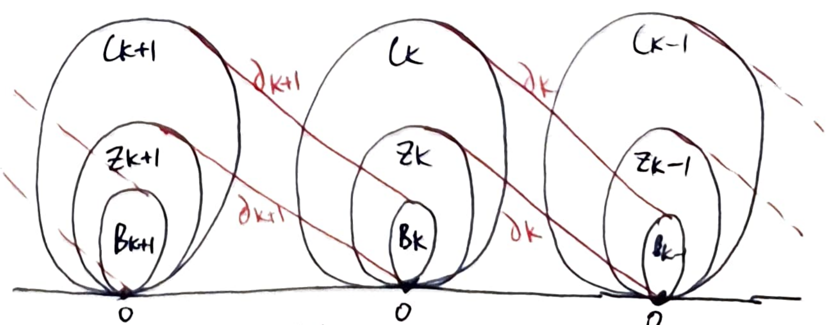
Since C_k is free, the oriented ~~k-comple~~ k-simplices form the standard basis for C_k.

For the boundary map $\partial_k : C_k \to C_{k-1}$, ~~we can represent it with the standard~~ we define M_k, the standard matrix representation of ∂_k.

- M_k has entries from $\{-1, 0, 1\}$.
- M_k has m_k columns (# of k-simplices) and m_{k-1} rows (# of $(k-1)$-simplices)
- $\left.\begin{array}{l} \text{null}(M_k) = Z_k. \\ \text{range}(M_k) = B_{k-1}. \end{array}\right\} (*)$

e.g. The basis for H_1 of the torus ~~X form~~ consists of the two 1-cycles:





(*) Visualizing $\begin{cases} \text{null}(M_k) = Z_k \\ \text{range}(M_k) = B_{k-1}. \end{cases}$

STEPS:

① $\partial_k : C_k \to C_{k-1}$ boundary map

$\downarrow$

② M_k: standard matrix representation of ∂_k (boundary matrix.)

$\downarrow$ elementary row & column operations.

① exchange row (resp. col.) i and j

② multiply row (resp. col.) by -1.

③ replace row (resp. col.) by $(\text{row/col } i) + q(\text{row/col. } j)$, $q \in \mathbb{Z}$, $j \neq i$.

③ $\tilde{M}_k$: Smith normal form.

$$\tilde{M}_k = \left[\begin{array}{ccc|c} b_1 & & 0 & \\ & b_2 & & 0 \\ & & \ddots & \\ 0 & & b_{l_k} & \\ \hline & 0 & & 0 \end{array}\right]$$

- $l_k = \operatorname{rank} M_k = \operatorname{rank} \tilde{M}_k$.
- $b_i \geq 1$ and $b_i \mid b_{i+1}$ $\forall\, 1 \leq i < l_k$.

Repeat ① – ③ for all dimensions. Then we get:

$\forall k$, (1) the torsion coefficients of H_{k-1} are the diagonal entries $b_i > 1$

(2) $\{e_i \mid l_k + 1 \leq i \leq m_k\}$ is a basis for Z_k.
$\operatorname{rank} Z_k = m_k - l_k$.

(3) $\{b_i \hat{e}_i \mid 1 \leq i \leq l_k\}$ is a basis for B_k.
$\operatorname{rank} B_k = \operatorname{rank} M_{k+1} = l_{k+1}$.

$\Downarrow$

β_k (Betti number) $= \operatorname{rank} Z_k - \operatorname{rank} B_k = m_k - l_k - l_{k+1}$.

Complexity of the reduction algorithm: $O(m^3)$ where m is the number of simplices in Δ.

(4)

• **Filtration**: motivation: ~~if we build the simplicial complex~~ on a set of points from ~~a space, the~~ For infinite space, use simplicial complex to approx the homology of the space.

A subcomplex of Δ is a subset $\Delta^i \subseteq \Delta$ that is also a simplicial complex.

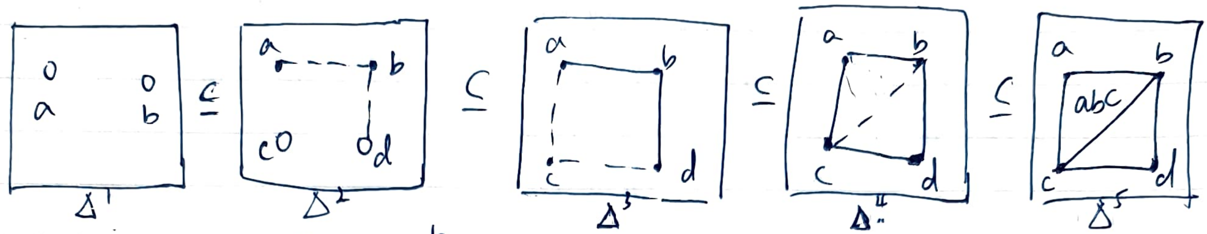
Let Δ be a finite simplicial complex and let $\Delta^1 \subset \Delta^2 \cdots \subset \Delta^m = \Delta$ be a finite sequence of nested subcomplexes of Δ. The simplicial complex Δ with such a sequence of subcomplexes is called filtered simplicial complex.

$\emptyset \subseteq \Delta^1 \subseteq \Delta^2 \subseteq \cdots \subseteq \Delta^m = \Delta$

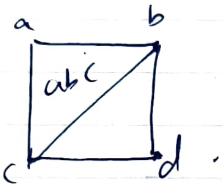
For generality, let ~~$K^i = K^m$~~ $\Delta^i = \Delta^m \quad \forall i \geq m$.

Δ is a filtered complex, i is called filtration index.

e.g. $\emptyset \subseteq$



back to eg. 1 · $\Delta = E1 =$



• **Persistence**:

Given a filtered complex, for the i^{th} complex Δ^i we can compute the associated

- boundary maps ∂_k^i for all dimensions k
- boundary matrices M_k^i (and hence $\tilde{M}_k^i$) for all dimensions k.
- groups C_k^i, Z_k^i (cycle group), B_k^i (boundary group), and H_k^i (homology group)

Then the p-persistent kth homology group of Δ^i is

- $H_k^{i,p} = Z_k^i / (B_k^{i+p} \cap Z_k^i)$ (persistence equation)

- (equivalently) define injection $\eta_k^{i,p} : H_k^i \to H_k^{i+p}$ that maps a homology class into another homology class that contains it.

$H_k^{i,p} \cong Im(\eta_k^{i,p})$.

Further, the associated p-persistent kth Betti number of Δ^i is $\beta_k^{i,p}$, and

$\beta_k^{i,p}$ = rank of the free subgp of $H_k^{i,p}$.

Persistence Module

Key idea: The persistence homology of a filtered complex is simply the standard homology of a ~~filtered complex~~ particular graded module over a polynomial ring.

Def: A persistence complex $\mathcal{C}$ is a family of complexes $\{C_*^i\}_{i\geq 0}$ over R, together with chain maps $f^i : C_*^i \to C_*^{i+1}$ s.t. ~~we have~~:

$$C_*^0 \xrightarrow{f^0} C_*^1 \xrightarrow{f^1} C_*^2 \xrightarrow{f^2} \cdots$$

$$\begin{tikzcd}
{} \arrow[d, "\partial_3"'] & {} \arrow[d, "\partial_3"'] & {} \arrow[d, "\partial_3"'] & \\
C_2^0 \arrow[r, "f^0"] \arrow[d, "\partial_2"'] & C_2^1 \arrow[r, "f^1"] \arrow[d, "\partial_2"'] & C_2^2 \arrow[r, "f^2"] \arrow[d, "\partial_2"'] & \cdots \\
C_1^0 \arrow[r, "f^0"] \arrow[d, "\partial_1"'] & C_1^1 \arrow[r, "f^1"] \arrow[d, "\partial_1"'] & C_1^2 \arrow[r, "f^2"] \arrow[d, "\partial_2"'] & \cdots \\
C_0^0 \arrow[r, "f^0"] & C_0^1 \arrow[r, "f^1"] & C_0^2 \arrow[r, "f^2"] & \cdots
\end{tikzcd}$$

Def: A persistence module $\mathcal{M}$ is a family of R-modules M^i, together with homomorphism $\varphi^i : M^i \to M^{i+1}$.

Algorithms:
Filtered simplicial complex $\to$ barcodes.
① Reduction (Prev. section move here).
②. Reading off the intervals.

(a) If $low(j) = i$: the simplex σ_j is paired with σ_i.
the entrance of σ_i causes the birth of a feature
σ_j causes the death of a feature.

(b) If $low(j)$ = undefined:
entrance of σ_j causes the birth of a feature. 2 further cases:

Case 1: If $\exists k$ s.t. $\text{low}(k) = j$, then σ_j is paired with σ_k
entrance of σ_k causes the death of the feature

Case 2: If no such k exists, then σ_j is unpaired.

- <u>Barcode</u>: A pair (σ_i, σ_j) gives the half-open interval $[dg(\sigma_i), dg(\sigma_j))$,
where $dg(\sigma)$ for a simplex $\sigma \in \Delta$ is the smallest number
ℓ s.t. $\sigma \in \Delta_\ell$.

~~Given a~~

Def: Let $\Delta_1 \subset \Delta_2 \subset \cdots \subset \Delta_m = \Delta$ be a filtered simplicial complex.
The <u>pth persistent homology of Δ</u> is the pair

$$\left(\{ H_p(\Delta_i) \}_{1 \le i \le m},\ \{ f_{i,j} \}_{1 \le i \le j \le m} \right),$$

where $\forall\, i,j \in \{1, \dots, m\}$ with $i \le j$, the linear maps $f_{i,j} : H_p(\Delta_i) \to H_p(\Delta_j)$ are the maps induced by the inclusion maps $\Delta_i \to \Delta_j$.

<u>Def</u>: For $i < j$, the <u>(i,j)-persistent homology</u> of a ~~a~~ <u>persistence complex $\mathcal{C}$</u>, denoted $H_*^{i \to j}(\mathcal{C})$, is defined to be the image of the induced homomorphism $f_* : H_*(C_*^i) \to H_*(C_*^j)$.

<u>Structure Theorem</u>

If D is a PID, then every finitely generated D-module is isomorphic to a direct sum of cyclic D-modules, i.e., it decomposes uniquely into the form:

$$D^\beta \oplus \left(\bigoplus_{i=1}^{m} D/d_i D \right), \quad d_i \in D,\ \beta \in \mathbb{Z} \text{ s.t. } d_i \mid d_{i+1}.$$

Similarly, every graded module M over a graded PID D decomposes uniquely into the form:

$$\left(\bigoplus_{i=1}^{n} \Sigma^{\alpha_i} D \right) \oplus \left(\bigoplus_{j=1}^{m} \Sigma^{\gamma_j} D/d_j D \right).$$

<u>Classification Theorem</u>

For a finite persistence module $\mathcal{M}$ with field F coefficients,

$$H_*(\mathcal{M}; F) \cong \underbrace{\bigoplus_i x^{t_i} \cdot F[x]}_{\text{free}} \oplus \underbrace{\left(\bigoplus_j x^{r_j} \cdot \left(F[x] / (x^{s_j} \cdot F[x]) \right) \right)}_{\text{torsional}}.$$

The free elements $\xleftrightarrow{\text{bij.}}$ homology generators with birth at t_i and persist for all future parameter values.

The torsional elements $\xleftrightarrow{\text{bij.}}$ homology generators with birth at r_j and death at $r_j + s_j$.

☆ The Classification Thm gives the fundamental characterization of barcodes.

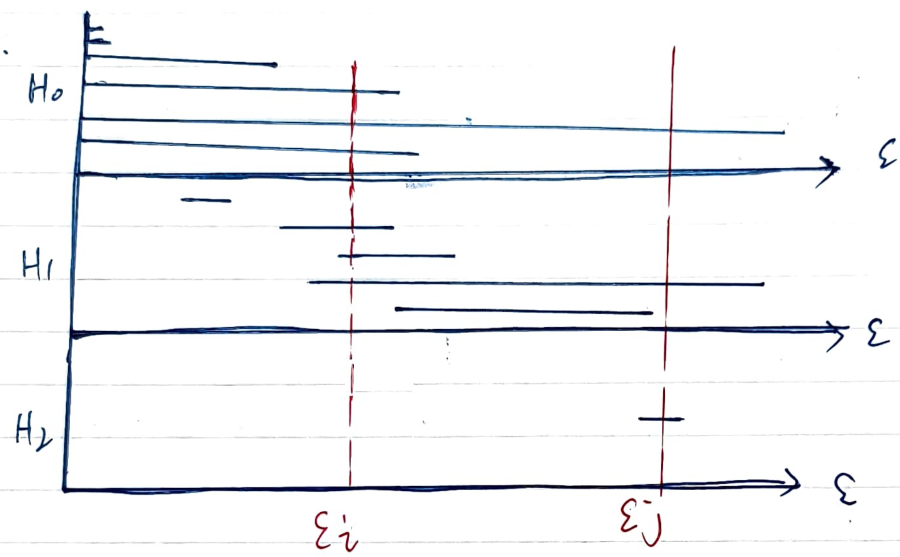
~~The classification Theorem gives the fundamental characterization~~

Thm: (Barcode as the persistence analogue of Betti number).

$\text{Rank}(H_k^{i \to j}(\mathcal{C};F)) = \#$ intervals in the barcode of $H_k(\mathcal{C};F)$ spanning the parameter interval $[i,j]$

In particular, $H_*(C_*^i;F) = \#$ intervals that contain i.

e.g.



Rank: k^{th} Betti number of a complex: $\beta_k := \text{rank}(H_k)$.
p-persistent k^{th} Betti number of complex Δ^i: $\beta_k^{i,p} := \text{rank}$ (free subgroup of $H_k^{i,p}$).

Note that as with Betti number, the barcode for H_k does not give the actual structure of the ~~...~~ homology group, but just a continuously parameterized rank. The barcode is useful in that it can qualitatively filter out topological noise (shortlived features) and capture significant features (features that persist over changing values of ε)