Topological Data Analysis and Persistence Theory

NSF/CBMS Conference

Valdosta State University

August 8-12, 2022

Peter Bubenik, University of Florida

Lecture 2 : Foundational Results

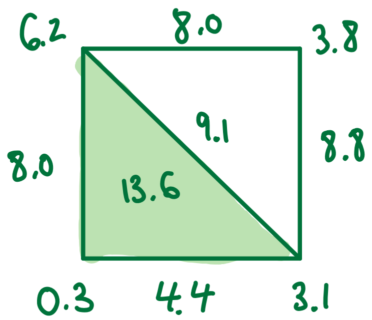
Outline:
1. The Persistence Algorithm
2. Distances
3. Stability
4. Generalized Persistence

Please interrupt me!!!

1. The Persistence Algorithm

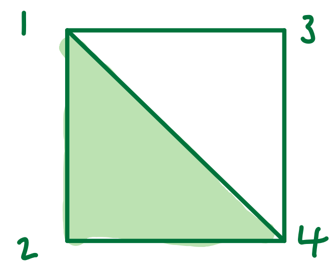
1.1 Weighted Complexes and Filtrations

Input: weighted simplicial complex

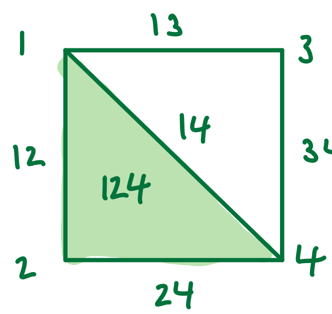


A

Assume: vertices are ordered

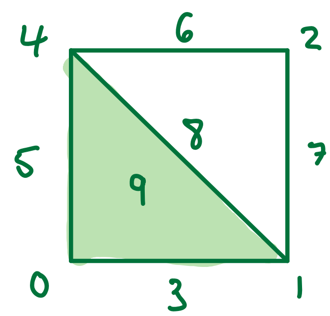


Order simplices using lexicographic order



B

Order simplices using A. Break ties with B



The Persistence Algorithm (below) only depends on the ordering of the simplices.

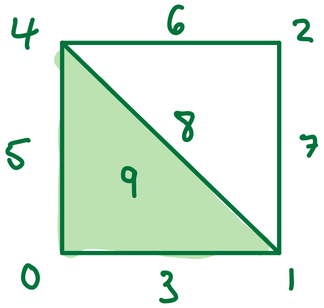
Let $d =$ # simplices.

need $w(\text{edge}) \geq w(\text{its vertices})$

The possible weights are parametrized by a subset of $\mathbb{R}^d$.

We can partition this subset by the orderings of the simplices.

1.2 The Persistence Algorithm



Boundary matrix:

$$D = \begin{array}{c|cccccccccc} & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ \hline 0 & & & & 1 & & 1 & & & & \\ 1 & & & & 1 & & & & 1 & 1 & \\ 2 & & & & & & & 1 & 1 & & \\ 3 & & & & & & & & & & 1 \\ 4 & & & & & & 1 & 1 & & 1 & \\ 5 & & & & & & & & & & 1 \\ 6 & & & & & & & & & & \\ 7 & & & & & & & & & & \\ 8 & & & & & & & & & & 1 \\ 9 & & & & & & & & & & \end{array}$$

Idea: Reduce D by iterating along columns left to right.
Zero column j from bottom to top as much as possible
by adding to column j from columns to its left.

For column j, pivot(j) = max row index for which column j has a nonzero entry
= 0 if column j is zero.

Algorithm

```
R := D
for j = 0 to n do
    while ∃ k<j with pivot(k) = pivot(j) do
        add column k to column j
    end while
end for
```

Proposition The algorithm terminates (in $O(n^3)$) and the resulting matrix R is "reduced": every row contains at most one pivot.

$$D = \begin{array}{cc} & \begin{array}{cccccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{array} & \begin{bmatrix} & & & 1 & & 1 & & & & \\ & & & 1 & & & & 1 & 1 & \\ & & & & & & 1 & 1 & & \\ & & & & & & & & & 1 \\ & & & & & 1 & 1 & & 1 & \\ & & & & & & & & & 1 \\ & & & & & & & & & \\ & & & & & & & & & \\ & & & & & & & & & 1 \\ & & & & & & & & & \end{bmatrix} \end{array} \longrightarrow \begin{array}{cc} & \begin{array}{cccccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{array} & \begin{bmatrix} & & & 1 & & 1 & 1 & & & \\ & & & 1 & & & & & & \\ & & & & & & 1 & & & \\ & & & & & & & & & 1 \\ & & & & & 1 & & & & \\ & & & & & & & & & 1 \\ & & & & & & & & & \\ & & & & & & & & & \\ & & & & & & & & & 1 \\ & & & & & & & & & \end{bmatrix} \end{array} = R$$

Reading the Persistence Diagram from R:

PD for H_p

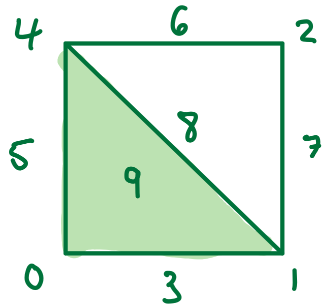
If $\text{pivot}(j) = i \neq 0$ then $(i, j) \in PD_p$, $p = \dim$ simplex i

If $\text{pivot}(i) = 0$ and $\nexists j$ s.t. $\text{pivot}(j) = i$ then $(i, \infty) \in PD_p$

$$R = \begin{array}{cc} & \begin{array}{cccccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{array} & \begin{bmatrix} & & & 1 & & 1 & 1 & & & \\ & & & 1 & & & & & & \\ & & & & & & 1 & & & \\ & & & & & & & & & 1 \\ & & & & & 1 & & & & \\ & & & & & & & & & 1 \\ & & & & & & & & & \\ & & & & & & & & & \\ & & & & & & & & & 1 \\ & & & & & & & & & \end{bmatrix} \end{array}$$

$PD_0 = \{ (1,3), (4,5), (2,6), (0,\infty) \}$

$PD_1 = \{ (8,9), (7,\infty) \}$



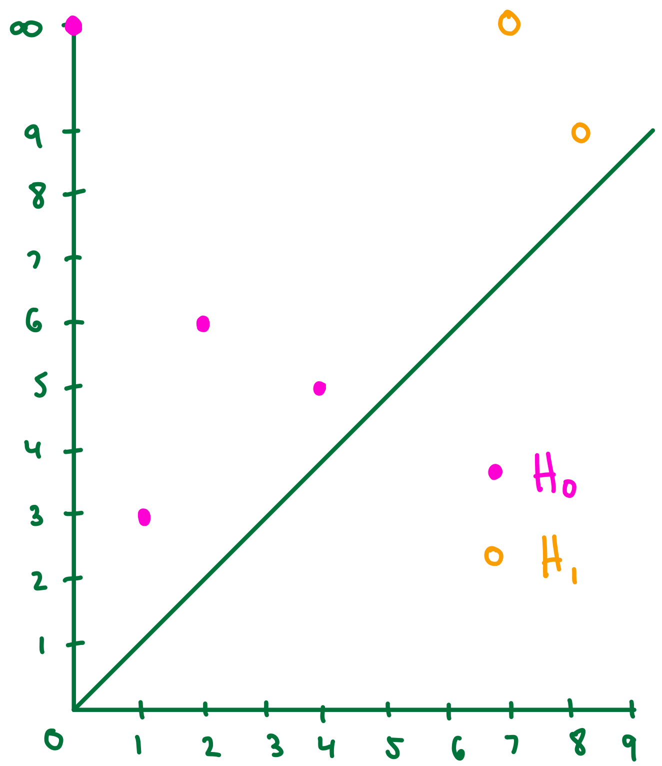
1.3 Plotting the Persistence Diagram and Barcode

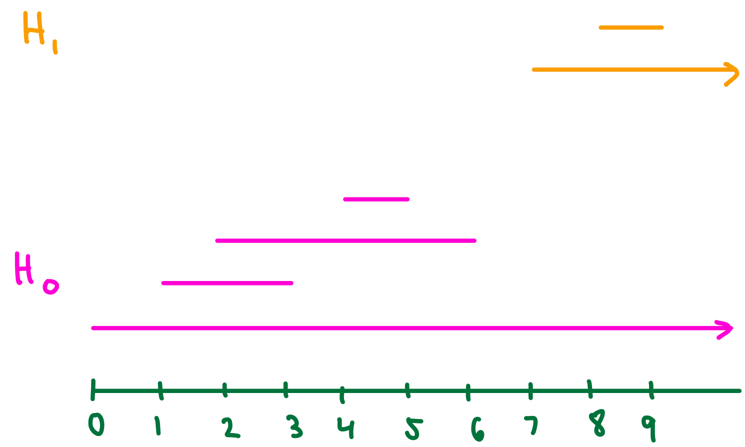
$PD_0 = \{ (1,3), (4,5), (2,6), (0,\infty) \}$

$PD_1 = \{ (8,9), (7,\infty) \}$

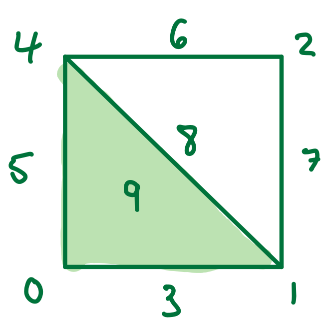
$B_0 = \{ [1,3), [4,5), [2,6), [0,\infty) \}$

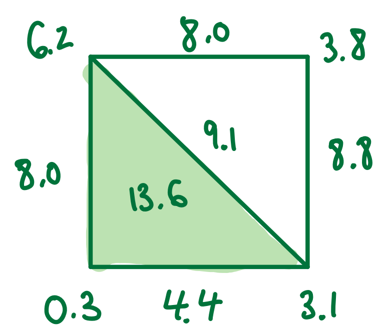
$B_1 = \{ [8,9), [7,\infty) \}$





To return to the starting real values, replace $0, 1, \ldots, n$ with the corresponding real numbers.





$PD_0 = \{(1,3), (4,5), (2,6), (0,\infty)\}$

$PD_1 = \{ (8,9), (7,\infty) \}$

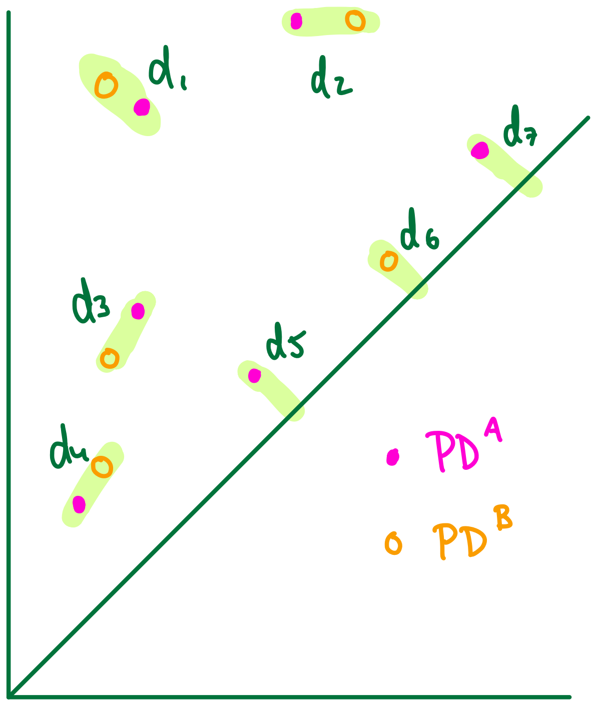
$PD_0 = \{(3.1, 4.4), (6.2, 8.0), (3.8, 8.0), (0.3, \infty)\}$

$PD_1 = \{ (9.1, 13.6), (8.8, \infty) \}$

2. Distances and Stability

2.1 Wasserstein Distance

We want to compare two persistence diagrams.



Idea: match points

where points may be matched with the diagonal.

Measure distances

Take p-norm of this vector.

Take the minimum of all matchings.

$$W_p(PD^A, PD^B) = \| (d_1, \dots, d_7) \|_p$$

Special case $p = \infty$ called the Bottleneck distance.

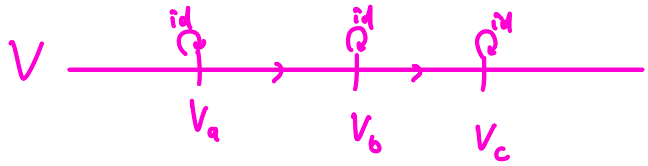
$1 \leq p \leq q \leq \infty$ $\quad W_1 \geq W_p \geq W_q \geq W_\infty$

most discriminating

2.2 Interleaving Distance

Persistence modules $\mathbb{R} \to \mathrm{Vect}$ $a \in \mathbb{R}$ $V_a \in \mathrm{Vect}$

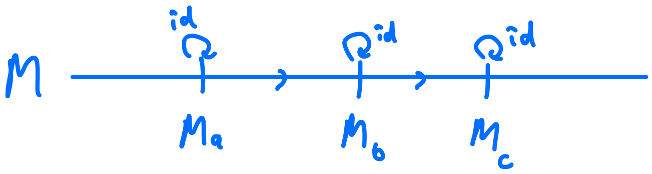
$a \le b$ $V_a \to V_b$



More generally: $\mathbb{R} \to C$, C some category

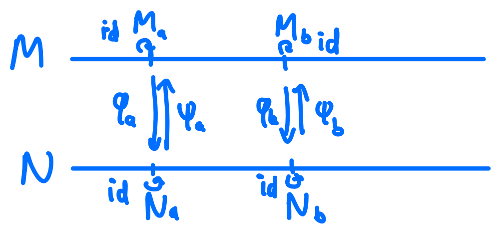
$a \in \mathbb{R}$ $M_a \in C$

$a \le b$ $M_a \to M_b \in C$



All maps comute: $M_{b \le c} \circ M_{a \le b} = M_{a \le c}$

Isomorphic:



all maps commute:

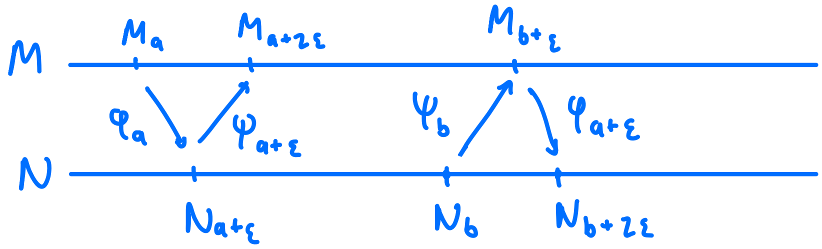
$\phi_a \varphi_a = \mathrm{id}_{M_a}$

$\varphi_b \phi_b = \mathrm{id}_{N_b}$

<u>Theorem</u> The persistence modules given by the Čech complex and the Delaunay complex are isomorphic.

Let $\varepsilon \geq 0$.

ε-interleaving:



All maps commute : triangles and parallelograms commute

$$d_I(M,N) = \inf \{ \varepsilon \geq 0 \mid \exists \ \varepsilon\text{-interleaving of } M, N \}$$

$= \infty$ ow

<u>Proposition</u> $d_I(M,M) = 0$, $d_I(M,N) = d_I(N,M)$

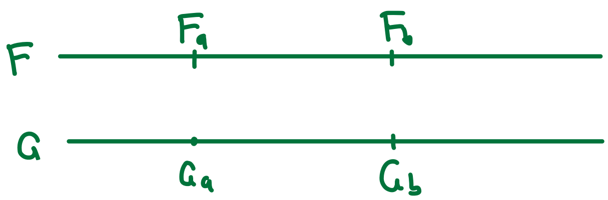
$d_I(M,P) \leq d(M,N) + d(N,P)$ extended pseudo metric

<u>Example</u> $X \in \mathrm{Top}$ $f : X \to \mathbb{R}$ $a \in \mathbb{R}$, $F_a = \{ x \in X, f(x) \leq a \}$

$\mathrm{Sub}(X,f) =: F : \mathbb{R} \to \mathrm{Top}$ $a \leq b$ $F_a \hookrightarrow F_b$

$(X,f) \in \mathbb{R}\mathrm{Top}$

Similarly $g : X \to \mathbb{R}$ gives $G : \mathbb{R} \to \mathrm{Top}$.



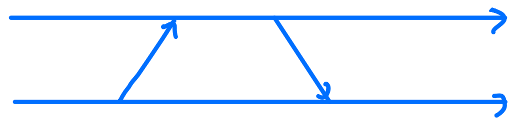
Let $d_\infty(f,g) = \sup_{x \in X} |f(x) - g(x)|$

<u>Lemma</u> $d_I(F,G) = d_\infty(f,g)$.

3. Stability

3.1 Categorical Stability

Let $I_\varepsilon = (\mathbb{R} \times \{0,1\}, \leq_\varepsilon)$



$\forall a \quad (a,0) \leq (a+\varepsilon, 1)$

$\qquad (a,1) \leq (a+\varepsilon, 0)$

$\forall a \leq b \quad (a,0) \leq (b,0)$

$\qquad (a,1) \leq (b,1)$

M, N ε-interleaved iff

$$\begin{tikzcd} \mathbb{R} \arrow[r, "0"] \arrow[dr, "M"'] & I_\varepsilon \arrow[d, dashed, "\exists"'] & \mathbb{R} \arrow[l, "1"'] \arrow[dl, "N"] \\ & C & \end{tikzcd}$$

__Theorem__ (B, de Silva, Scott) $M, N : \mathbb{R} \to C$ ε-interleaved and $H : C \to D$

$\Rightarrow HM, HN$ are ε-interleaved.

__Proof__

$$\begin{tikzcd} \mathbb{R} \arrow[r, "0"] \arrow[dr, "M"'] \arrow[ddr, bend right, "HM"'] & I_\varepsilon \arrow[d] & \mathbb{R} \arrow[l, "1"'] \arrow[dl, "N"] \arrow[ddl, bend left, "HN"] \\ & C \arrow[d, "H"] & \\ & D & \end{tikzcd}$$

D

__Corollary__ $d_I(HM, HN) \leq d_I(M, N)$.

__Corollary__ $d_I(\underset{\mathrm{Sub}(X,f)}{\underset{\shortparallel}{HF}}, \underset{\mathrm{Sub}(X,g)}{\underset{\shortparallel}{HG}}) \leq d_\infty(f, g)$

3.2 The Isometry Theorem

Isometry Theorem

Let $M, N : \mathbb{R} \to \text{Vect}$ be persistence modules arising from a finite weighted complex.

Let DM, DN denote their persistence diagrams.

Then $W_\infty(DM, DN) = d_I(M, N)$

Corollary (Sublevel set Stability Theorem)

$$W_\infty\big(\underset{\substack{\shortparallel \\ \text{Sub}(X,f)}}{DF}, \underset{\substack{\shortparallel \\ \text{Sub}(X,g)}}{DG}\big) \leq d_\infty(f, g)$$

4. Generalized Persistence

4.0 Ordinary persistence

Example $X \in F_{\mathbb{R}}\,\mathrm{Top}$ $\quad \forall a \in \mathbb{R}$ $\quad X_a \in \mathrm{Top}$ $\quad a \le b$ $\quad X_a \hookrightarrow X_b$

$$X : \mathbb{R} \to \mathrm{Top}$$

$H = H_j$ $\qquad \mathbb{R} \xrightarrow{X} \mathrm{Top} \xrightarrow{H} \mathrm{Vect} \qquad HX : \mathbb{R} \to \mathrm{Vect}$

Example $K \in F_{[n]}\,\mathrm{Simp}$

$K : [n] \to \mathrm{Simp} \qquad K_0 \hookrightarrow K_1 \hookrightarrow K_2 \hookrightarrow \cdots \hookrightarrow K_n$

$HK : [n] \to \mathrm{Vect} \qquad HK_0 \to HK_1 \to HK_2 \to \cdots \to HK_n$

4.1 Zigzag persistence

Problem: What if we have $K_0, K_1, \dots K_n \in \mathrm{Simp}$
but don't have any maps $K_i \to K_j$?

Solution : $K_0 \hookrightarrow K_0 \cup K_1 \hookleftarrow K_1 \hookrightarrow K_1 \cup K_2 \hookleftarrow K_2 \hookrightarrow \cdots \hookleftarrow K_n$

[or $\quad K_0 \hookleftarrow K_0 \cap K_1 \hookrightarrow K_1 \hookleftarrow K_1 \cap K_2 \hookrightarrow K_2 \hookleftarrow \cdots \hookrightarrow K_n$]

Let $\mathbb{Z}_{2n+1}$ be the poset

$$0 \le 1 \ge 2 \le 3 \ge 4 \le \cdots \ge 2n+1$$

We have $K : \mathbb{Z}_{2n+1} \to \mathrm{Simp}$

Apply homology to get $HK : \mathbb{Z}_{2n+1} \to \mathrm{Vect}$

4.2 Multiparameter Persistence

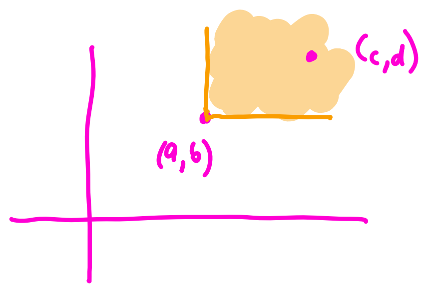
2-parameter persistence:

$\forall a, b \in \mathbb{R} \quad X_{a,b} \in \text{Top} \qquad a \le c,\ b \le d \quad X_{a,b} \hookrightarrow X_{c,d}$

$X \in F_{\mathbb{R}^2}\text{Top} \qquad X : \mathbb{R}^2 \to \text{Top} \qquad HX : \mathbb{R}^2 \to \text{Vect}$

$\mathbb{R}^2$ is the poset $(\mathbb{R}^2, \le)$

$(a,b) \le (c,d) \iff a \le c$ and $b \le d$



d-parameter persistence: $\quad X : \mathbb{R}^d \to \text{Top} \qquad HX : \mathbb{R}^d \to \text{Vect}$

$\mathbb{R}^d$ is the poset $(\mathbb{R}^d, \le)$

$(x_1, \ldots, x_d) \le (y_1, \ldots, y_d) \iff x_i \le y_i \ \forall i$

4.3 Poset-indexed Persistence

Let P be a poset. $\quad X : P \to \text{Top}$ (or $X : P \to \text{Simp}$)

$P \xrightarrow{X} \text{Top} \xrightarrow{H} \text{Vect}$

4.4 Generalized Persistence

a functor

a functor

$$P \xrightarrow{X} C \xrightarrow{H} A$$

a small category

a category

an (abelian) category

additive / exact / Grothendieck

$$P \xrightarrow{X} C$$

is a "diagram"

in C indexed by P