Topological data analysis
Lecture 5

Anton Ayzenberg

ATA Lab, FCS NRU HSE
Noeon Research

Spring 2024
Faculty of Computer Science / Yandex Data School

Persistent homology: practical algorithm

- Previously we had $D_1, D_2, \ldots$ matrices of simplicial differentials $\partial_i \colon C_i(K) \to C_{i-1}(K)$
- D_i is the incidence matrix between i-dim simplices and $(i-1)$-dim simplices at least over $\mathbb{Z}_2$.
- Now we combine them all in a huge matrix D of size $N \times N$,
- where N is the total number of simplices in a filtration.
- The order of rows and columns as in the array Simplices.

Gauss elimination on columns

- Reduce D by elementary operations on columns.
- It is allowed to any column, to add another column multiplied by a scalar.
- We move from left to right checking all $1 \leqslant j \leqslant N$.
- Let $\mathrm{low}(j)$ denote the index of the lowest nonzero element in j-th column.
- For each j we loop through $r < j$, and look for r such that $\mathrm{low}(r) = \mathrm{low}(j)$.
- When we meet such r, we reduce j-th column by r-th column.
- We loop until $\mathrm{low}\colon [N] \to [N]$ becomes injective. This is called the reduced form of a matrix.
- Permutations are not allowed.

You have the Matrix...

	1	2	3	4	23	34	12	24	234	13	14	123	124	134	1234
1					0	0	1	0		1	1				
2					1	0	1	1		0	0				
3					1	1		0		1	0				
4						1		1	0		1				
23									1			1			
34									1			0	0	1	
12									0			1	1	0	
24									1			0	1	0	
234												0	0	0	1
13												1	0	1	0
14		0								0			1	1	0
123		0			0					0					1
124		0		0	0		0			0	0				1
134		0		0	0		0			0	0				1
1234		0		0	0		0			0	0				

After Gauss elimination in columns

	1	2	3	4	23	34	12	24	234	13	14	123	124	134	1234
1							1								
2					1		1								
3					1	1									
4						1									
23									1						
34									1						
12												1	1		
24									1				1		
234															1
13												1			
14													1		
123															1
124															1
134															1
1234															

After Gauss

Look at the reduced matrix M.

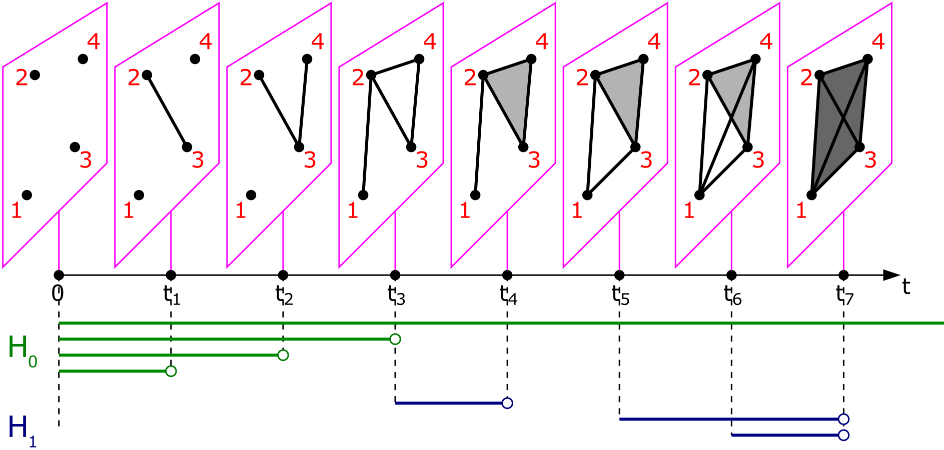
Theorem

- Let $j = \mathrm{low}(s) \neq 0$ in M. Then **Simplices[j]** gives birth to homology of rank dim Simplices[j] and **Simplices[s]** kills this homology. Each such pair (j, s) gives rise to the interval module $I_{[BirthTimes[j], BirthTimes[s])}$.
- If the whole r-th column vanish and moreover $r \notin \mathrm{low}([N])$, then Simplices[r] gives birth to homology of rank dim Simplices[r] which never dies. It gives the interval module $I_{[BirthTimes[r], +\infty)}$.

Persistent homology of the filtration is the direct sum of all the listed interval modules.

Example

Simplices = [1, 2, 3, 4, 23, 34, 12, 24, 234, 13, 14, 123, 124, 134, 1234]
BirthTimes = [0, 0, 0, 0, t1, t2, t3, t3, t4, t5, t6, t7, t7, t7, t7]



Our reduced matrix

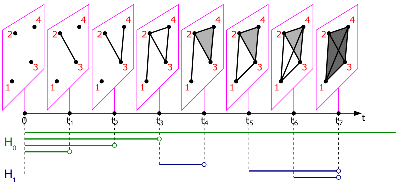
	1	2	3	4	23	34	12	24	234	13	14	123	124	134	1234
1							1								
2					1		1								
3					1	1									
4						1									
23									1						
34									1						
12												1	1		
24									1				1		
234															1
13												1			
14													1		
123															1
124															1
134															1
1234															

Circled elements have indices: $[3, 23]$, $[4, 34]$, $[2, 12]$, $[24, 234]$, $[13, 123]$, $[14, 124]$, $[134, 1234]$.

Reading births and deaths

Norns determine the fate of homology

Simplices = [1, 2, 3, 4, 23, 34, 12, 24, 234, 13, 14, 123, 124, 134, 1234]
BirthTimes = [0, 0, 0, 0, t1, t2, t3, t3, t4, t5, t6, t7, t7, t7, t7]



- We see pairs: $[3, 23]$, $[4, 34]$, $[2, 12]$, $[24, 234]$, $[13, 123]$, $[14, 124]$, $[134, 1234]$.
- For example, vertex 3 gives rise to homology which dies when 23 appears.
- 3 appears at time 0, and 23 — at time t_1.
- Hence we have interval module $I_{[0;t_1)}$ in 0-th homology.
- etc...

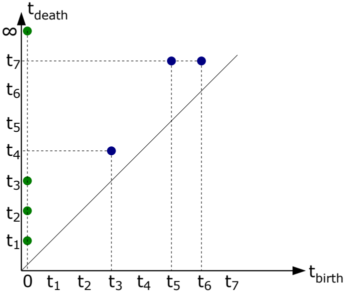
We have decomposition: $I_{[0;t_1)} \oplus I_{[0;t_2)} \oplus I_{[0;t_3)} \oplus I_{[t_3;t_4)} \oplus I_{[t_5;t_7)} \oplus I_{[t_6;t_7)}$.

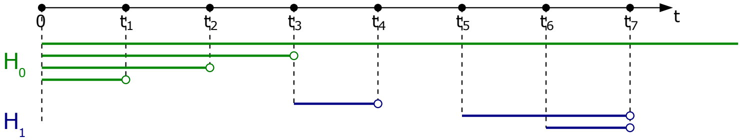
Real times

- Let K_t, $t \in \mathbb{R}$ be a collection of simplicial complexes on vertex set $[m]$,
- such that $t_1 < t_2$ implies $K_{t_1} \subseteq K_{t_2}$.
- This is called a filtration with real time.
- Changes may occur only at discrete time moments.
- Therefore, the only difference is that BirthTimes stores real values.
- The algorithm above outputs the interval decomposition.

How to encode persistent homology?

Persistence diagram instead of barcode.





Philosophy

- If some homology lives long, then it is **a meaningful homology**.
- The lifetime = death time - birth time.
- Lifetime is long, if the point of persistence diagram is far from the diagonal $y = x$.
- Points close to the diagonal are considered noise.
- Long-live-homology are important topological features of a filtration.

Filtrations: Čech

Demonstration: press to play in browser

Filtrations: Čech

Demonstration: press to play in browser

- Let $X = \{x_1, \ldots, x_m\} \subset \mathbb{R}^d$ be a point-cloud.
- Let $X_t = \bigcup B_{t/2}(x_i)$.
- X_t is a filtration, but not of simplicial complexes.

Filtrations: Čech

Demonstration: press to play in browser

- Let $X = \{x_1, \dots, x_m\} \subset \mathbb{R}^d$ be a point-cloud.
- Let $X_t = \bigcup B_{t/2}(x_i)$.
- X_t is a filtration, but not of simplicial complexes.
- Replace X_t by its **nerve**!

Nerve complex

Let $K_t^{\check{C}}$ be a simplicial complex on $[m]$, such that $\{i_1, \dots, i_s\} \in K_t^{\check{C}}$ iff

$$B_{t/2}(x_{i_1}) \cap \dots \cap B_{t/2}(x_{i_s}) \neq \varnothing$$

Let $K_t^{\check{C}}$ be a simplicial complex on $[m]$, such that $\{i_1, \ldots, i_s\} \in K_t^{\check{C}}$ iff $B_{t/2}(x_{i_1}) \cap \cdots \cap B_{t/2}(x_{i_s}) \neq \varnothing$.

Nerve theorem

$$X_t \simeq K_t^{\check{C}}$$

Nerve complex and Nerve theorem

Let $K_t^{\check{C}}$ be a simplicial complex on $[m]$, such that $\{i_1, \ldots, i_s\} \in K_t^{\check{C}}$ iff $B_{t/2}(x_{i_1}) \cap \cdots \cap B_{t/2}(x_{i_s}) \neq \varnothing$.

Nerve theorem

$$X_t \simeq K_t^{\check{C}}$$

- Therefore X_t and $K_t^{\check{C}}$ have the same homology for each t.
- Hence they have the same persistent homology.
- We can work with simplicial filtration $\{K_t^{\check{C}}\}$ called **Čech filtration**.

Nerve complex and Nerve theorem

Let $K_t^{\check{C}}$ be a simplicial complex on $[m]$, such that $\{i_1, \ldots, i_s\} \in K_t^{\check{C}}$ iff $B_{t/2}(x_{i_1}) \cap \cdots \cap B_{t/2}(x_{i_s}) \neq \varnothing$.

Nerve theorem

$$X_t \simeq K_t^{\check{C}}$$

- Therefore X_t and $K_t^{\check{C}}$ have the same homology for each t.
- Hence they have the same persistent homology.
- We can work with simplicial filtration $\{K_t^{\check{C}}\}$ called **Čech filtration**.
- Simplex I is born at the time = minimal t for which balls of radii $t/2$ around x_i, $i \in I$, intersect.

Nerve complex and Nerve theorem

Let $K_t^{\check{C}}$ be a simplicial complex on $[m]$, such that $\{i_1, \ldots, i_s\} \in K_t^{\check{C}}$ iff $B_{t/2}(x_{i_1}) \cap \cdots \cap B_{t/2}(x_{i_s}) \neq \varnothing$.

Nerve theorem

$$X_t \simeq K_t^{\check{C}}$$

- Therefore X_t and $K_t^{\check{C}}$ have the same homology for each t.
- Hence they have the same persistent homology.
- We can work with simplicial filtration $\{K_t^{\check{C}}\}$ called **Čech filtration**.
- Simplex I is born at the time = minimal t for which balls of radii $t/2$ around x_i, $i \in I$, intersect.
- It may be difficult to find this number in practice.

Vietoris–Rips filtration

- Let $X = \{x_1, \ldots, x_m\} \subset \mathbb{R}^d$ be a point-cloud.

Vietoris–Rips filtration

- Let $X = \{x_1, \dots, x_m\} \subset \mathbb{R}^d$ be a point-cloud.
- K_t^{VR} be a simplicial complex on $[m]$ such that
- $I \in K_t^{VR}$ iff all pairwise distances between x_i and x_j, for $i \neq j$, $i, j \in I$, are less than t.
- Simplex I is born at time moment $\max_{i \neq j \in I} \operatorname{dist}(x_i, x_j)$.

Vietoris–Rips filtration

- Let $X = \{x_1, \dots, x_m\} \subset \mathbb{R}^d$ be a point-cloud.
- K_t^{VR} be a simplicial complex on $[m]$ such that
- $I \in K_t^{VR}$ iff all pairwise distances between x_i and x_j, for $i \neq j$, $i, j \in I$, are less than t.
- Simplex I is born at time moment $\max_{i \neq j \in I} \operatorname{dist}(x_i, x_j)$.
- $\{K_t^{VR}\}$ is called **Vietoris–Rips filtration**.

Vietoris–Rips filtration

- Let $X = \{x_1, \ldots, x_m\} \subset \mathbb{R}^d$ be a point-cloud.
- K_t^{VR} be a simplicial complex on $[m]$ such that
- $I \in K_t^{VR}$ iff all pairwise distances between x_i and x_j, for $i \neq j$, $i, j \in I$, are less than t.
- Simplex I is born at time moment $\max_{i \neq j \in I} \mathrm{dist}(x_i, x_j)$.
- $\{K_t^{VR}\}$ is called **Vietoris–Rips filtration**.
- Very easy to compute birth times!

Vietoris–Rips filtration

- Let $X = \{x_1, \dots, x_m\} \subset \mathbb{R}^d$ be a point-cloud.
- K_t^{VR} be a simplicial complex on $[m]$ such that
- $I \in K_t^{VR}$ iff all pairwise distances between x_i and x_j, for $i \neq j$, $i, j \in I$, are less than t.
- Simplex I is born at time moment $\max_{i \neq j \in I} \mathrm{dist}(x_i, x_j)$.
- $\{K_t^{VR}\}$ is called **Vietoris–Rips filtration**.
- Very easy to compute birth times!
- Can be adapted to any finite metric space, e.g. metric graph.

How to compare filtrations and diagrams?

- We have two filtrations $\{K_t\}$ and $\{L_t\}$ on the same set $[m]$.
- Set $\mathrm{dist}(\{K_t\},\{L_t\}) = \max_{I\subset[m]} |\mathrm{BirthTime}_K[I] - \mathrm{BirthTime}_L[I]|$.

How to compare filtrations and diagrams?

- We have two filtrations $\{K_t\}$ and $\{L_t\}$ on the same set $[m]$.
- Set $\mathrm{dist}(\{K_t\},\{L_t\}) = \max_{I\subset[m]} |\mathrm{BirthTime}_K[I] - \mathrm{BirthTime}_L[I]|$.
- If $X = \{x_1,\dots,x_m\}$ and $Y = \{y_1,\dots,y_m\}$ are two data clouds and
- $\mathrm{dist}(X,Y) = \max_i \mathrm{dist}(x_i,y_i)$, and
- K_t^{VR} and L_t^{VR} are their Vietoris–Rips filtrations

How to compare filtrations and diagrams?

- We have two filtrations $\{K_t\}$ and $\{L_t\}$ on the same set $[m]$.
- Set $\mathrm{dist}(\{K_t\}, \{L_t\}) = \max_{I \subset [m]} |\mathrm{BirthTime}_K[I] - \mathrm{BirthTime}_L[I]|$.
- If $X = \{x_1, \ldots, x_m\}$ and $Y = \{y_1, \ldots, y_m\}$ are two data clouds and
- $\mathrm{dist}(X, Y) = \max_i \mathrm{dist}(x_i, y_i)$, and
- K_t^{VR} and L_t^{VR} are their Vietoris–Rips filtrations, then
- **Exercise:** $\mathrm{dist}(\{K_t^{VR}\}, \{L_t^{VR}\}) \leqslant 2\,\mathrm{dist}(X, Y)$.

How to compare filtrations and diagrams?

- We have two filtrations $\{K_t\}$ and $\{L_t\}$ on the same set $[m]$.
- Set $\mathrm{dist}(\{K_t\},\{L_t\}) = \max_{I\subset[m]} |\mathrm{BirthTime}_K[I] - \mathrm{BirthTime}_L[I]|$.
- If $X = \{x_1, \ldots, x_m\}$ and $Y = \{y_1, \ldots, y_m\}$ are two data clouds and
- $\mathrm{dist}(X, Y) = \max_i \mathrm{dist}(x_i, y_i)$, and
- K_t^{VR} and L_t^{VR} are their Vietoris–Rips filtrations, then
- **Exercise:** $\mathrm{dist}(\{K_t^{VR}\},\{L_t^{VR}\}) \leqslant 2\,\mathrm{dist}(X, Y)$.
- What about their persistent diagrams?

Metric on diagrams

Bottleneck distance (or Wasserstein, or Kantorovich metric)

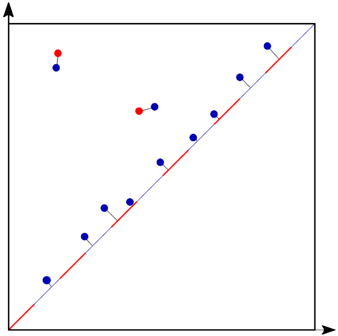


Fig.from GUDHI Library

Metric on diagrams

Bottleneck distance (or Wasserstein, or Kantorovich metric)

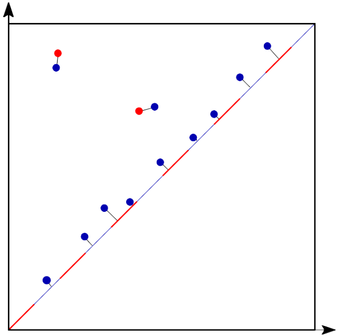


Fig.from GUDHI Library

Stability theorem

$$\mathrm{dist}(\mathrm{PD}(\{K_t\}), \mathrm{PD}(\{L_t\})) \leqslant \mathrm{dist}(\{K_t\}, \{L_t\}).$$

Generalizations of persistent modules

We can hardly see this slide. However if we do, then it is time to switch to whiteboard

Sources

S. Barannikov, *Framed Morse complex and its invariants*, Advances in Soviet Mathematics. Vol.21 (1994), pp. 93–115.

H. Y. Cheung, T. C. Kwok, L. C. Lau, *Fast matrix rank algorithms and applications*, J. ACM 60:5 (2013), Article 31.

H. Edelsbrunner, J. L. Harer, Computational Topology: An Introduction, 2010.

Morozov, Dionysus2 library https://mrzv.org/software/dionysus2/

A. J. Zomorodian, Topology for computing, 2005.