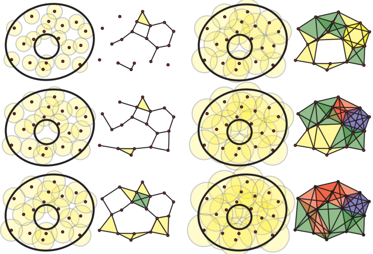
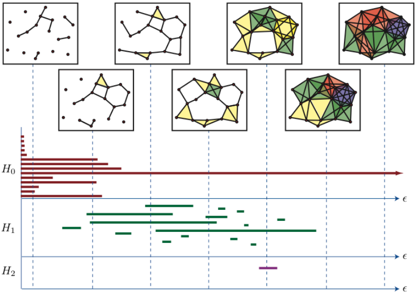
Persistent Homology and TDA

10/14/20 10:46 AM





[Ghrist 08]

Main Topic of the day: Barcodes. — parametrized ver. of Betti numbers

- Existence structure.
- Computation efficiency
- Stability. noise-robust

⑤

Input: Point cloud X

Output: "Shape" of the point cloud X

1st Idea: Build complexes from X

Čech complex $\check{C}_\varepsilon(X)$:

Form a d-simplex on $(x_0 \dots x_d)$ if $\bigcap_i B(x_i, \varepsilon/2) \neq \phi$.

By Nerve Thm, $\check{C}_\varepsilon(X) \simeq \bigcup_{x \in X} B(x, \varepsilon/2)$.

Vietoris-Rips complex $VR_\varepsilon(X)$:

Form a d-simplex on $(x_0 .. x_d)$ if the balls pairwise intersects.

VR Lemma: $\check{C}_\varepsilon(X) \subseteq VR_\varepsilon(X) \subseteq \check{C}_{2\varepsilon}(X)$.

Problem: How to choose ε?

2nd Idea: Don't choose, try it all!

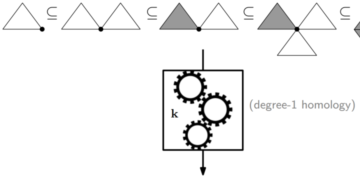
Persistent Homology.

[ELZ02, ZC05]

sequence of complex inclusions

Persistent complex $\mathcal{K}$: $C^1 \overset{i_1}{\hookrightarrow} C^2 \overset{i_2}{\hookrightarrow} C^3 \overset{i_3}{\hookrightarrow} \cdots \hookrightarrow C^N$

Goal. Homology of $i_* : H_\bullet(C^i) \xrightarrow{i_*} H_\bullet(C^j) \quad \forall i<j$



$$\mathbf{k} \xrightarrow{\binom{1}{0}} \mathbf{k}^2 \xrightarrow{(0\ 1)} \mathbf{k} \xrightarrow{\binom{0}{1}} \mathbf{k}^2 \xrightarrow{\begin{pmatrix}1&0\\0&1\end{pmatrix}} \mathbf{k}^2 \cdots$$

[Oudot20]

(i,j)-homology group of $\mathcal{K}$:

$$H_\bullet^{i\to j}(\mathcal{K}) := \text{im}\left(H_\bullet(C^i) \to H_\bullet(C^j)\right)$$

p-persistent Betti number: $\beta_p^{i\to j} := \text{rank}\, H_p^{i\to j}$.

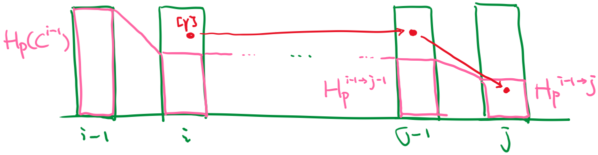
Consider the induced maps:

$$0 \to H_p(C^1) \xrightarrow{i_*} H_p(C^2) \xrightarrow{i_*} \cdots \xrightarrow{i_*} H_p(C^N)$$

$H_p^{i\to j}(\mathcal{K})$ contains "cycles" that survived from C^i to C^j.

$$H_p^{i\to j} = \text{Cycle}_p(C^i) \Big/ \left(\text{Bdry}_p(C^j) \cap \text{Cycle}_p(C^i)\right)$$

Birth and Death.



$[\gamma]$ is born at i if $[\gamma] \in H_p(C^i)$ but $[\gamma] \notin H_p(C^{i-1})$

$[\gamma]$ dies at j if it merges w/ classes born earlier:

$i_*[\gamma] \notin H_p^{i-1\to j-1}$ but $i_*[\gamma] \in H_p^{i-1\to j}$

Notice that the later-born class gets merged.

Thm. For finite persistent complex $\mathcal{C}$ over field F. (say $\mathbb{R}$ or $\mathbb{Z}_2$)
[ZC05]

$$H(\mathcal{C};F) \simeq \bigoplus_i x^{b_i}\cdot F[x] \oplus \left(\bigoplus_j x^{b_j}\cdot\left(F[x]/x^{d_j-b_j}\cdot F[x]\right)\right)$$

$x^k = i_k$ sending $C^i \to C^{i+k}$

classes born at b_i & lived since

classes born at b_j & die at d_j

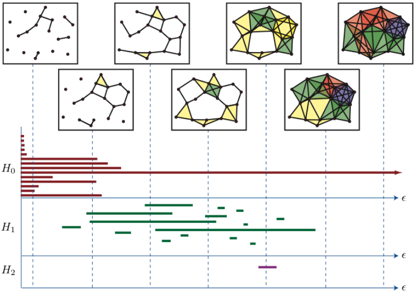
In other words, $H(\mathcal{C})$ decomposes into modules $F_{[b_i,d_i]}$:

$$0\to\cdots\to 0\to \underset{b_i}{F}\to\cdots\to \underset{d_i}{F}\to 0\to\cdots\to 0$$

$$H(\mathcal{C}) \simeq \bigoplus_i F_{[b_i,d_i]}$$

Cor. Bar code exists!

rank $H_k^{i\to j}$ = # $\overset{(k\text{-})}{\text{intervals}}$ survived from i to j



Persistence $\mathrm{Pers}[\gamma] := \varepsilon_j - \varepsilon_i$

long intervals: real cycles.

short intervals: noise.

Remark. This also explains why $VR_\varepsilon(X)$ is useful:

Any cycle $[\gamma]$ persists under $VR_\varepsilon \to VR_{2\varepsilon}$ is also a cycle in $\check{C}_{2\varepsilon}$

$$H_\bullet(VR_\varepsilon)\to H_\bullet(\check{C}_{2\varepsilon})\to H_\bullet(VR_{2\varepsilon})$$

* * *

Computation.

Incremental algorithm, adding simplices one-by-one.

Lemma. When p-simplex Δ is added, either

(1) A non-bdy p-cycle $[\gamma]$ is created, or positive Δ

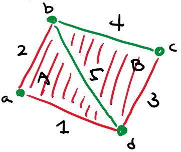
(2) An existing $(p-1)$-cycle $[\gamma]$ is destroyed. negative Δ

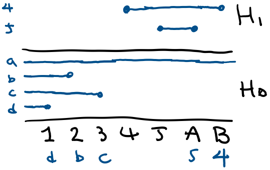
By birth & death, each negative Δ_p is paired w/ a positive σ_{p-1}

intuition: paired w/ youngest σ_{p-1} in $\partial\Delta_p$ available.

PAIR(Δ):

```
$\gamma \leftarrow \partial\Delta$    // (p-1)-cycle
$\sigma \leftarrow$ youngest (p-1)-simplex in $\gamma$.
while ($\sigma$, $\Delta'$) paired & $\gamma \neq \emptyset$:
    $\gamma' \leftarrow$ cycle killed by $\Delta'$
    $\gamma \leftarrow \gamma + \gamma'$
    $\sigma \leftarrow$ youngest (p-1)-simplex in $\gamma$
if $\gamma \neq \emptyset$:
    $\Delta$ is negative, pair ($\sigma$, $\Delta$)
else:
    $\Delta$ positive
```





So we need ways to detect the creation of simplices.

Easier w/ $VR_\varepsilon(X)$ ⟶ Huge & unmanagable still.

Witness Complex. Do subsampling $Y \subseteq X$.

[de Silva - Carlsson 04]

simplex $\sigma = (y_1, \dots, y_d)$ is weakly witnessed by x if

$$d(y_i, x) \leq d(y, x) \quad \forall y_i \ \& \ y \in Y \setminus \{y_1 \dots y_d\}$$

" strongly witnessed by x if

$$d(y_1, x) = \dots = d(y_d, x)$$

Prop. Simplex σ is strongly witnessed iff $\forall \tau \subseteq \sigma$, τ weakly witnessed.

$W_X(Y) :=$ all simplices on Y strongly witnessed by some x in X

Connection to Delaunay triangulation:

Prop. $\sigma \in Del(Y)$ iff σ strongly witnessed by $x \in \mathbb{R}^n$.

i.e. $Del(Y)$ is the witness complex $W_{\mathbb{R}^n}(Y)$.

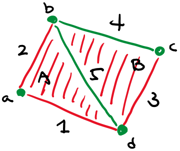
Cor. $W_X(Y) \subseteq W_{\mathbb{R}^n}(Y) = Del(Y)$.

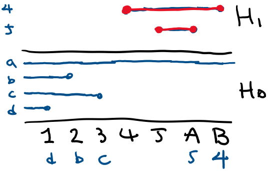
Not capturing topology again.

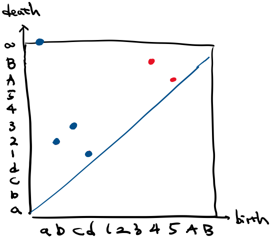
Modification: Graph induced complex [Dey-Fan-Wang13]

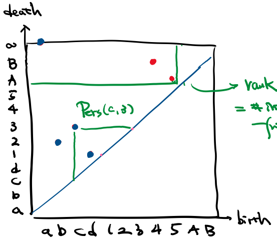
Stability.

Persistent Diagram Dg(𝒢).

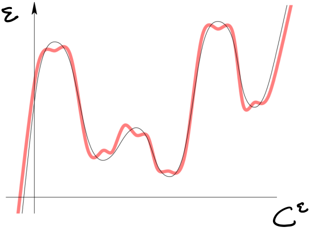


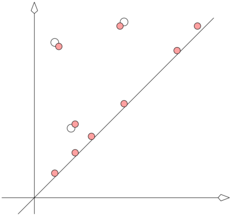






$\to \operatorname{rank} H^{5\to A}$
$= \#$ intervals persist from 5 to A.





Stability Thm. [CEH07. CSM06]

$$D_\infty(Dg_p(C^\varepsilon), Dg_p(\hat{C}^\varepsilon)) \le \|C^\varepsilon - \hat{C}^\varepsilon\|_\infty$$

$$\underset{\text{bijection } \eta}{\inf} \ \sup_{p \in Dg} \|p - \eta(p)\|_\infty \qquad \max_\sigma |\text{birth } \sigma - \text{birth } \hat{\sigma}|$$

"bottleneck dist." "moving the pts X slightly"

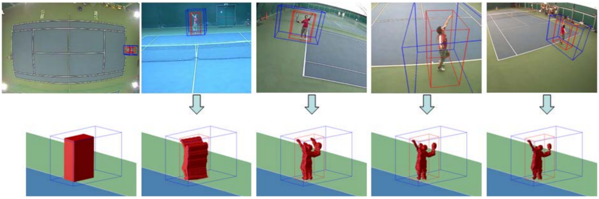
Moral: Robust to noise!

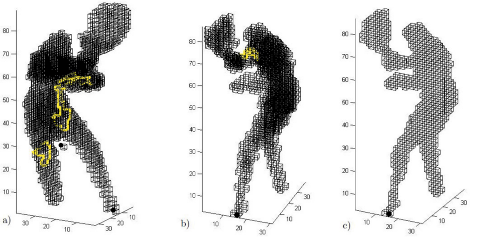
Applications. Homology inference, Shape comparison, Time series.

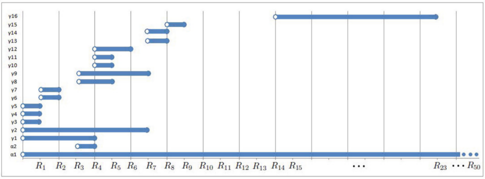
⇒ Morse Theory.

* * *

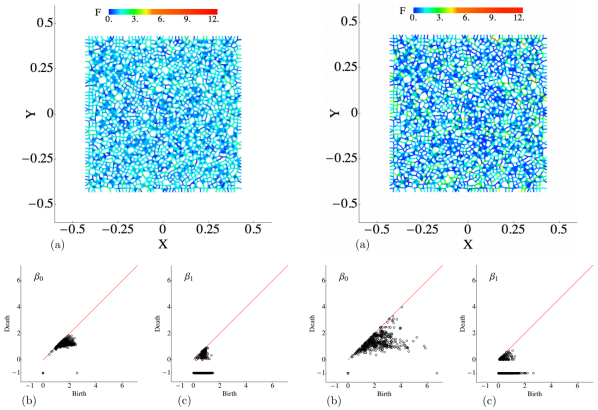
Example: 3D reconstruction by voxel carving. [GMJ012]



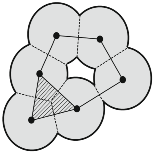


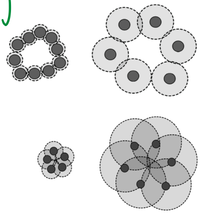


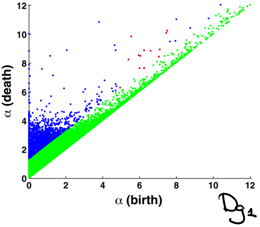
Example. Force distribution in particulate system [KGKM14]

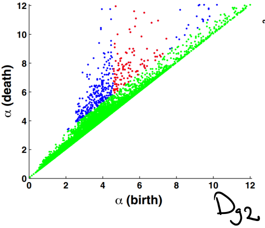


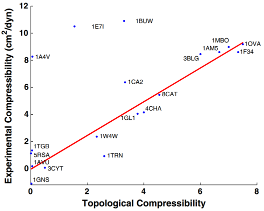
Example. Protein compressibility. [GHIKMN14]











Example. Wheeze Detection. [TGK14]

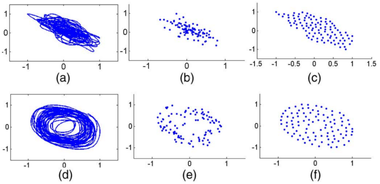


Fig. 5. Point cloud subsampling, (a) and (d): the delay embedding of a non-wheeze and a wheeze signal including 4000 points. 100 subsamples selected using random method ((b),(e)) and maxmin method((c),(f)).

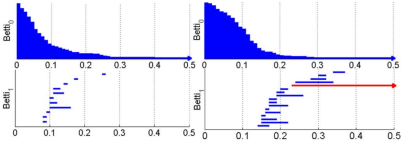


Fig. 6. The barcodes for a non-wheeze breathing sound signal (left) and a wheeze signal (right). The “significant” barcode is highlighted in red and is used to distinguish wheeze signals from non-wheeze signals.