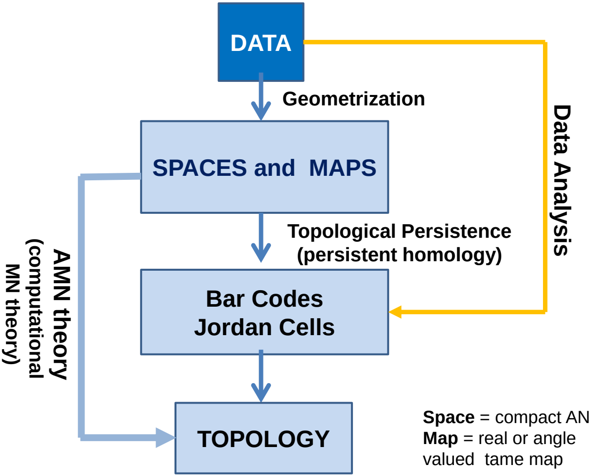
Data Analysis, Persistent homology and Computational Morse-Novikov theory

Dan Burghelea

Department of mathematics
Ohio State University, Columbuus, OH

Bowling Green , November 2013



Space = compact ANR
Map = real or angle valued tame map

CONTENTS

- **Data and Geometrization**
- **Topological Persistence (barcode and Jordan cells)**
- **A computer friendly alternative to Morse–Novikov theory (AMN)**
- **More mathematics**

1. What is **Data**
2. How the Data is obtained
3. What do we want from Data
4. What do we do

1. Mathematically data is given as:

a finite metric space (X, d)

and possibly a map

a map $f : X \to \mathbb{R}$ **or** $f : X \to \mathbb{S}^1$

2. Data are obtained :

a. By sampling (a shape in three or hiherdimensional eucliden space or a probability distribution)

b. By scanning a 2 dimensional picture

c. As a collection of two dimensional pictures (black-white) of a three dimensional environment taken by camera from different angles; each 2D picture regarded as a vector in the pixel space with a gray scale coordinate for each pixel.

d. As a list of measurements of parameters of a collection of objects/individuals; for example observations on the patients (in a hospital)

3. One wants:

i. In case of sampled geometrical objects :

to **derive geometric** and **topological features** without reconstructing the object entirely or **reconstruct a continuos shape** from a sampling.

ii. In case of an a priory unstructured observations:

to discover **patterns** and **unexpected features**, detect **missing blocks** of data, **clusterings**

4. One geometrizes the data:

one converts data into topological spaces / spaces and (real or angle valued)maps .

TOPOLOGY provides :

1. Methods to convert a finite metric space into "nice topological space " = simplicial complex or simplicial complexs and simplicial real or angle valued maps or simplicial complex with a filtration.

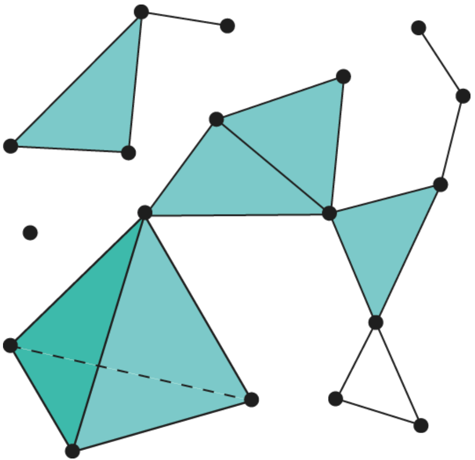
and uses

2. Homology, Betti numbers, EP characteristic (which describes all sorts of connectivity) to make mathematically precise qualitative features of the shape and then to calculate them.

SIMPLICIAL COMPLEXES

- A **solid** $k-$ **simplex** is the convex hull of $(k+1)$ linearly independent points .
- A **geometric simplicial complex** K is a "nice subspace of an Euclidean space " precisely a union of **solid simplicies** which intersect each other in faces (subsimplexes) .
- An **abstract simlicial complex** is a pair (V, Σ) with: V a finite set, Σ a family of nonempty subsets of V, so that $\boxed{\sigma \subseteq \tau \in \Sigma \Rightarrow \sigma \in \Sigma}$.

- An abstract simplicial complex determines a geometric simplicial complex and vice versa.
- A simplicial complex is determined by its incidence matrix which can be fed in as input of an algorithm.



To a finite metric space (X, d) and $\epsilon > 0$ one asociates:

1. The abstract **CECH COMPLEX,** $\mathcal{C}_\epsilon(X, d) := (\mathcal{X}, \Sigma_\epsilon)$
   - $\mathcal{X} = X$
   - $S_k := \{(x_1, x_2, \cdots x_{k+1})|$ iff $B(x_1; \epsilon) \cap \cdots B(x_{k+1}; \epsilon) \neq \emptyset\}$
2. The abstract **VIETORIS- RIPS COMPLEX**, $\mathcal{R}_\epsilon(X, d) := (\mathcal{X}, \Sigma_\epsilon)$.
   - $\mathcal{X} = X$,
   - $S_k := \{(x_1, x_2, \cdots x_{k+1})|$ iff $d(x_i, x_j) < \epsilon\}$.

- If $\epsilon < \epsilon'$ $\boxed{\mathcal{C}_\epsilon(X,d) \subseteq \mathcal{C}_{\epsilon'}(X,d)}$, $\boxed{\mathcal{R}_\epsilon(X,d) \subseteq \mathcal{R}_{\epsilon'}(X,d).}$
- The topology of $\mathcal{C}_\epsilon(X,d)$ can be very different from $\mathcal{R}_\epsilon(X,d)$, however one has:

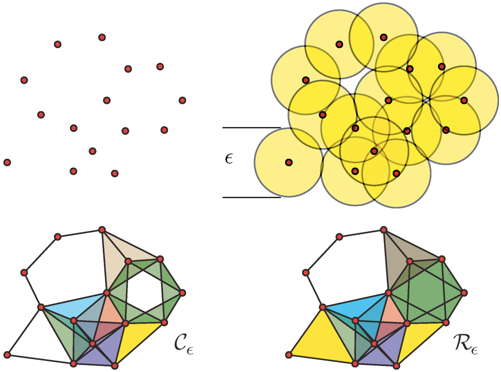
$$\boxed{\mathcal{R}_\epsilon(X,d) \subseteq \mathcal{C}_\epsilon(X,d) \subseteq \mathcal{R}_{2\epsilon}(X,d) \subseteq \mathcal{C}_{2\epsilon}(X,d)}$$

- A map $f : X \to \mathbb{R}$ provides the simplicial maps $f : \mathcal{C}_\epsilon(X,d) \to \mathbb{R}$ and $f : \mathcal{R}_\epsilon(X,d) \to \mathbb{R}$
- If $\epsilon < \pi$ a map $f : X \to \mathbb{R}$ provides the simplicial maps $f : \mathcal{C}_\epsilon(X,d) \to \mathbb{S}^1$.

If the data is a sample of a compact manifold embedded in the Euclidean space then:

Theorem

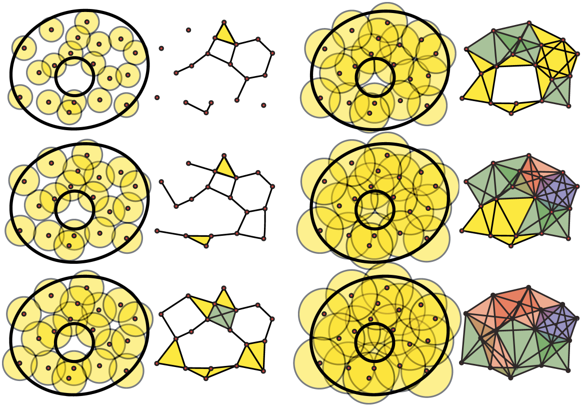
There exists $\alpha > 0$ so that for any $\epsilon-$ dense sample (X, d), $\epsilon < \alpha$, the Cech complex $\mathcal{C}_\epsilon(X, d)$ is homotopy equivalent to the manifold.



A fixed set of points can be completed to a Cech complex C_ε or to a Rips complex R_ε based on a proximity parameter ε. This Cech complex has the homotopy type of the $\varepsilon/2$ cover ($S1 \vee S1 \vee S1$), while the Rips complex has a different homotopy type ($S1 \vee S2$).

- The topology of the ϵ complexes differ, for different ϵ's .

  It is therefore desirable to consider all these complexes.



One obtains:

1. A simplicial complex,

2. A simplicial complex and a simplicial map $f : X \to \mathbb{R}$ whose sub levels $f^{-1}(-\infty, t]$ change the homology for finitely many real values $t_0 < t_1, t_2, \cdots t_N$,

3. A simplicial complex and a simplicial map $f : X \to \mathbb{R}$ or $f : X \to \mathbb{S}^1$ whose levels $f^{-1}(t)$ change the homology for finitely many (real or angle values) $t_0 < t_2 < \cdots t_N \in \mathbb{R}$.

4. A simplicial complex X with a filtration $X_0 \subset X_1 \subset \cdots X_{N-1} \subset X_N = X$; it can be interpreted as item 2 via the telescope construction.

Inspired from Morse theory/ Morse–Novikov theory:

- to $f : X \to \mathbb{R}$ based on changes in homology of sub levels $f^{-1}(-\infty, a]$ one associates a collection of *sub level bar codes* = intervals $[a, b], [a, \infty)$
- to $f : X \to \mathbb{R}$ or to $f : X \to \mathbb{S}^1$ based on changes in the homology of the levels $f^{-1}(t)$ one associates a collection for types of *bar codes* = intervals $[a, b], (a, b), [a, b), (a, b]$ and *Jordan cells* $\{(\lambda, k) \mid \lambda \in \mathbb{C} \setminus 0, k \in \mathbb{Z}_{\geq 1}\}$

Topological persistence

For: $f : X \to \mathbb{R}$,

- X compact ANR,
- f a continuous tame map,
- $a < b$, and $X_a := f^{-1}(a)$; $X_{[a,b]} := f^{-1}([a,b])$

consider

$$H_r(X_a) \xrightarrow{i_a} H_r(X_{[a.b]}) \xleftarrow{i_b} H_r(X_b)$$

The collection of these linear relations is reffered to as

(extended) **persistent homology**.

One says that:

- $x \in H_r(X_a)$ will be dead at b, $b > a$, if $i_a(x) = 0$
- $y \in H_r(X_b)$ was born after a, $a < b$, if $i_b(y) = 0$
- $x \in H_r(X_a)$ is right-observable at b, $b \geq a$ if there exists $y \in H_r(X_b)$ so that if $i_a(x) = i_b(y)$
- $y \in H_r(X_b)$ is left-observable at a, $a \leq b$ if there exists $x \in H_r(X_a)$ so that if $i_b(y) = i_b(y)$

BarCodes and Jordan cells

These concepts lead for any r to four types of intervals called *bar codes*.

- $r-$ closed bar code $[a, b]$,
- $r-$ open bar code (a, b),
- $r-$ closed-open $[a, b)$,
- $r-$ open-closed $(a, b]$.

The numbers a, b are critical values of f, i.e. values t where the homology of the fibers $X_t = f^{-1}(t)$ changes.

In case of $f : X \to \mathbb{S}^1$ to an isomorphism (the regular part of the linear relation

$$\boxed{H_r(X_t) \xrightarrow{i_t} H_r(X_{[t,t+2\pi]}) \xleftarrow{i_{t+2\pi}} H_r(X_{t+2\pi})}$$

leads for any r to

- $r-$Jordan cells $\{(\lambda, k) \mid \lambda \in \mathbb{C} \setminus 0, k \in \mathbb{Z}_{\geq 1}\}$

Existence of a closed / open, $r-$bar code with ends a and b means: for t between a and b there exists $x \in H_r(X_t)$ which is:

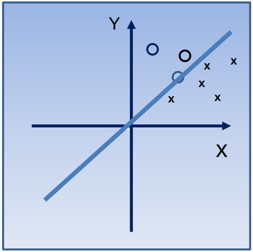
- observable at b but not at b', $b' > b$ and at a but not at a'', $a'' < a$,
- dead at b but not at b', $t < b' < b$ and born after a but not after a'', $a < a'' < t$,

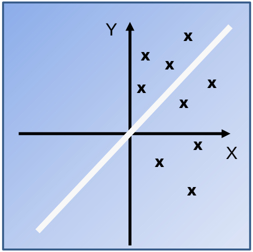
Existence of closed-open / open-closed $r-$bar code with ends a and b means that for t between a and b there exists $x \in H_r(X_t)$ which is:

- observable at a but not at a'', $a'' < a$, and dead at b but not at b', $t < b' < b$,
- observable at b but not at b', $b' > b$ and born after a but not after a', $a < a' < t$.

The multiplicity of such bar code is the number of linearly independent elements x which satisfy the properties above.

- One denotes by $\mathcal{B}^c_r(f)$, $\mathcal{B}^o_r(f)$, $\mathcal{B}^{c,o}_r(f)$ and $\mathcal{B}^{o,c}_r(f)$ the set of closed, open, closed-open, and open-closed $r-$bar codes of f.
- One collects the sets $\mathcal{B}^c_r(f)$ and $\mathcal{B}^o_{r-1}(f)$ as the finite configuration of points $C_r(f)$ in $\mathbb{C}$.
- One collects the sets $\mathcal{B}^{c,o}_r(f)$ and $\mathcal{B}^{o,c}_r(f)$ as the finite configuration of points $c_r(f)$ in $\mathbb{C} \setminus \Delta\}$
  $\Delta := \{z \in \mathbb{C} \mid \Re z = \Im z\}$





The bar code with ends *a, b, a≤b* and closed at *a* is represented as a point *a+ ib* while the bar code with ends *a, b, a < b* open at *a* is represented as a point *b + ia.*

Descripion of the configurationn $C_r(f)$

Consider the function $H^f : \mathbb{R}^2 \to \mathbb{Z}_{\geq 0}$ defined by

$$H^f(a,b) := \dim \operatorname{img} \begin{cases} H_r(f^{-1}(-\infty, a]) \to H_r(X)) \cap \\ \operatorname{img}(\mathrm{H_r(f^{-1}([b,\infty)) \to H_r(X)}) \end{cases}$$

For any square $B = [a_1, a_2] \times [b_1, b_2],\ a_1 < a_2, b_1 < b_2$, define

$$I(B) = H(a_1, b_2) + H(a_2, b_1) - H(a_1, b_1) - H(a_2, b_2)$$

and for any $z = a + ib$ the integer valued function

$$\mu^f(a,b) = \lim_{(a,b)\in \mathrm{intB}} I(B).$$

Theorem

$$C_r(f) = \mu^f.$$

Replacing the function H^f above by the function

$$h^f(a,b) := \begin{cases} \dim \mathrm{img}(H_r(f^{-1}(-\infty,a]) \to H_r(f^{-1}(-\infty,b]) & \text{if a < b} \\ \dim \mathrm{img}(H_r(f^{-1}[a,\infty)]) \to H_r(f^{-1}[b,\infty)) & \text{if a > b} \end{cases}$$

one derives for f tame, $c_r(f)$.

Relates the bar codes (bar codes and Jordan cells) of f to the topology of X ($X, \xi_f \in H^1(X;\mathbb{Z})$).
For $f : X \to \mathbb{R}$ a tame map one has.

Theorem

If $f: X \to \mathbb{R}$ is tame map then $\sharp\mathcal{B}^c_r(f) + \sharp\mathcal{B}^o_{r-1}(f)$ is a homotopy invariant of X, more precisely is equal to the Betti number $\beta_r(X)$

Theorem (Stability)

The assignment $C(X, \mathbb{S}^1) \ni f \rightsquigarrow C_r(f) \in S^n(\mathbb{C})$, $n = \beta_r(X)$, *is continuous.*

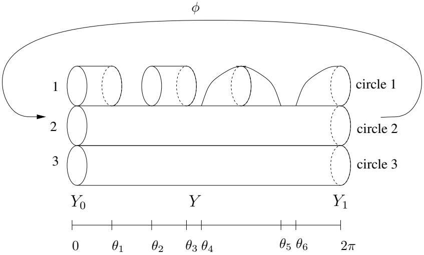
Theorem (Poincaré duality)

If M^n *is a closed* κ*-orientable*[a] *topological manifold with* $f: M \to \mathbb{R}$ *a tame map then* $C_r(f)(z) = C_{n-r}(-f)(i\bar{z})$

[a] If κ has characteristic 2 any manifold is κ-orientable if not the manifold should be orientable.

Similar but more subtle results hold for angle valued maps

EXAMPLE



map ϕ	*r*-invariants		
	dimension	bar codes	Jordan cells
circle 1: 1 time around circle 1 -3 times around 2, - 2 times around 3 circle 2: 1 time around circle 1, 4 times around 2, 1 time around 3 circle 3: 2 time around 1, 2 times around 2, 2 times around 3	0	(θ_2, θ_3)	$(1, 1)$
	1	$(\theta_6, \theta_1 + 2\pi]$ $[\theta_2, \theta_3]$ (θ_4, θ_5)	$(3, 2)$

Figure: Example of *r*-invariants for a circle valued map

Note : - If one add a cord from the $\theta_2 =$level to $\theta_3 -$ level one introduces a $0-$open bar code (θ_2, θ_3).

CREDITS.

1 H.Edelsbrunner, D. Letscher, A. Zamorodian, introduced sublevel persoistence

2. G. Carlsson V. de Silva, D. Morozov, introduced ZigZag persistence

3. D. Burghelea, T.Dey, introduced persistence for angle valued maps

4. D.Cohen-Steiner, H.Edelsbrunner, J.Harer First result on stability for sub level barcodes.

5. D. Burghelea, S.Haller . Stability fot the configurations $C_r(f)$