Science today is faced with:

1. larger and larger the amount of (recordable) data
less and less possible to analyze the data by traditional methods

2. considerably computable power (parallel computing) and (hopefully soon to be) quantum computing available

3. A computer can output only rational numbers and **does not input infinite objects**

From Data (Analysis) to (Algebraic)Topology and back

Dan Burghelea

Department of Mathematics
Ohio State University, Columbuus, OH

Bucharest, November 2024

Algebraic Topology and DATA

a. ALGEBRAIC TOPOLOGY is about

SHAPES= spatial forms (multidimensional)

$$\begin{cases} \text{smooth manifolds} \subset \\ \text{semi-algebraic sets} \subset \\ \text{SIMPLICIAL COMPLEXES, K} \end{cases}$$

and possibly real- valued or multi-valued maps $f : K \rightsquigarrow \mathbb{R}$

b. DATA or POINT CLOUD DATA (PCD) =
finite metric space (X, d).
Possibly X is equipped with a multi-valued, real-valued function on X.

c. both objects can be inputted in a computer as square matrices, with the data matrix having possibly additional columns

Data - Topology - Data

DATA = (point cloud data)

$\Downarrow$ geometrization

SHAPES (simplicial complexes)

$\Downarrow$ linear algebra

INVARIANTS (computer friendly) $\Rightarrow$ **MATHEMATICS** (new results)

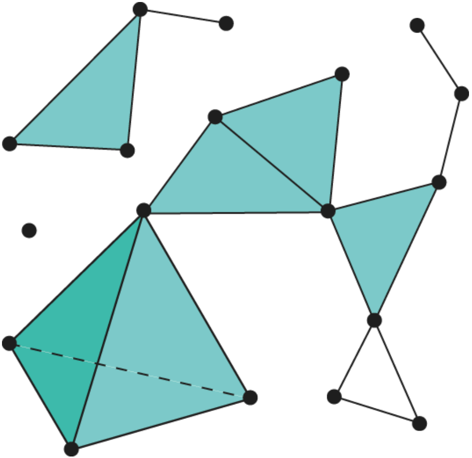
$\Downarrow$ (interpretation)

DATA (detect previously undetected) patterns (like clusters, twistings, missing blocks. relevant number of parameter the observed DATA depends on,etc)

SIMPLICIAL COMPLEXES

- A **solid** $k-$ **simplex** is the convex hull of $(k+1)$ linearly independent points in some Euclidean space .
- A **geometric simplicial complex** K is a subset of an Euclidean space which is a union of **solid simplicies** which intersect each other in faces (solid simplices) .
- For a geometric simplicial complex K denote by $\mathcal{S}_i$ the set of simpicies of dimension i.

Illustration



$\sharp\mathcal{S}_0 = 18,\ \sharp\mathcal{S}_1 = 23,\ \sharp\mathcal{S}_2 = 8,$

- An **abstract simplicial 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 with $\mathcal{S}_k = \{\sigma \in \Sigma \mid \sharp\sigma = k + 1\}$ and vice versa.
- A simplicial complex can be **inputted as a matrix with entries** ± 1 **or** 0,

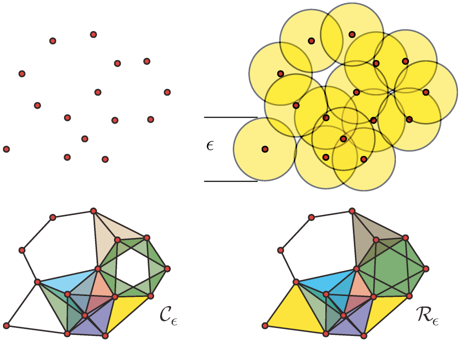
Geometrization of Data (X, d) into $(V, \Sigma, f : V \to ?)$

To a **FINITE** metric space (X, d) and $\epsilon > 0$ one can asociate:

1. The abstract **CECH COMPLEX,** $\mathcal{C}_\epsilon(X, d) := (V(X, d), \Sigma_\epsilon)$
   - $V(X, d) = X$
   - $\mathcal{S}_k := \{(x_1, x_2, \cdots x_{k+1})|$ iff $B(x_1; \epsilon) \cap \cdots B(x_{k+1}; \epsilon) \neq \emptyset\}$

   or
2. The abstract **RIPS COMPLEX**, $\mathcal{R}_\epsilon(X, d) := (V(X, d), \Sigma_\epsilon)$.
   - $V(X, d) = X$,
   - $\mathcal{S}_k := \{(x_1, x_2, \cdots x_{k+1})|$ iff $d(x_i, x_j) < \epsilon\}$.

Illustration (from R.Griest)



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 ($S1 \vee S2$).

$\mathcal{R}_\epsilon(X, d)$ or $\mathcal{C}_\epsilon(X, d)$ lead via telescope construction ?? to a simplicial complex and a simplicial map $f : X \to \mathbb{R}$ whose levels $f^{-1}(t)$ change the topology for finitely many real values $t_0 < t_1 < t_2, \cdots < t_N$.

The proposed invariants record the changes in the homology, (a concept to be described below), of the levels of a real-valued or angle-valued map

Elementary Linear Algebra / homology / torsion

- **Vector spaces**, V, $V' \subset V$, V/V',
- **inner product** on V $\langle , \rangle$; if $\dim V < \infty$ identifies V with V^*,
- **linear map** $f : V_1 \to V_2$; when V_1, V_2 are equipped with inner products $\langle , \rangle_1$, $\langle , \rangle_2$ f has volume $Vol(f)$ when $f \neq 0$,

  with the convention $Vol(f = 0) = 1$
- **Linear transformation** $\alpha : V \to V$ has $tr(\alpha), \det(\alpha), \det'(\alpha)$[1]

[1] product of nonzero eigenvalues

- **Chain complex**

$$\mathcal{V} := \ 0 \longrightarrow V_n \cdots \xrightarrow{d_{i+2}} V_{i+1} \xrightarrow{d_{i+1}} V_i \xrightarrow{d_i} V_{i-1} \xrightarrow{d_{i-1}} \cdots V_0 \longrightarrow 0$$

  for any i, $d_i \cdot d_{i+1} = 0$

- **Homology** (and) **Betti numbers**

$$H_i(\mathcal{V}) := \ker d_i / img\ d_{i+1}, \ \ \beta_i(\mathcal{V}) := \dim H_i(\mathcal{V}) \in \mathbb{Z}_{\geq 0}$$

- **E-P characteristic**
  $\chi(\mathcal{V}) := \sum (-1)^i \dim V_i = \sum (-1)^i \beta_i(\mathcal{V}) \in \mathbb{Z}$

- For $\langle , \rangle_i$ inner products in V_i, **Torsion of the chain complex with inner products**
  $T(\mathcal{V}, \langle , \rangle) := \prod_{i, d_i \neq 0} (Vol\ d_i)^{(-1)^i} \in \mathbb{R}.$

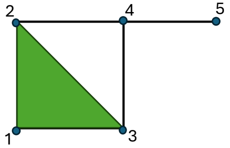
The chain complex with inner products

For a simplicial complex with $\mathcal{S}_k$ the sets of $k-$simplices define:

1. $(V_k, \langle , \rangle_k) := Maps(\mathcal{S}_k, \kappa)$, $\kappa = \mathbb{R}$ or $\mathbb{C}$,
2. $d_k(x) := \sum_{y, \in \mathcal{S}_{k-1}} I(x,y)y, \quad x \in \mathcal{S}_k, y \in \mathcal{S}_{k-1}$

I(x,y) is the incidency between x and y; a total order on the set V is supposed to be chosen

EXAMPLE :



$\mathcal{S}_0 = \{1, 2, 3, 4, 5\}$;
$\mathcal{S}_1 = \{(1,2),(1,3),(2,3),(2,4),(3,4),(4,5)\}$; $\mathcal{S}_2 = \{(1,2,3)\}$

The associated chain complex

$V_0 = \kappa^5$, $V_1 = \kappa^5$, $V_2 = \kappa$, $\kappa = \mathbb{R}/\mathbb{C}$ with the obvious inner product

$$d_2 = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 1 & 0 \\ 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1, & 1 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix} d_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$\beta_2 = 0, \beta_1 = 1, \beta_0 = 1, \chi = 0,$

$Vol\ d_2 = Vol\ d_1 = 1$ hence $T = 1$.

Classical topological invariants

For a finite simplicial complex (and then for any compact space X homeomorphic to it) using the **chain complex with inner products** associated to the simplicial complex one has

- Betti numbers, $\beta_i(X) \in \mathbb{Z}_{\geq 0}$
- E-U characteristic, $\chi(X) \in \mathbb{Z}$
- Torsion, $T(X) \in \mathbb{Q}$

For a pair $(X, \xi), \xi \in [X, S^1]$

- Monodromy, $R(X, \xi) \in [GL[Q]]$?

Interaction Topology - Data suggests new topological invariants: **barcodes with multiplicity** and **Jordan cells with multiplicity** which refine the previous one and are accessible to the computer. (i.e. computer friendly)

BarCodes and Jordan cells

These classical numerical invariants can be refined to four types of intervals with multiplicity, and Jordan blocks with multiplicity, for each $r = 0, 1, 2, \cdots$.

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

- $r-$ closed bar code $[a, b]$, $\Rightarrow z = a + ib \in \mathbb{C}$
- $r-$ open bar code (a, b), $\Rightarrow z = b + ia \in \mathbb{C}$
- $r-$ closed-open $[a, b)$, $\Rightarrow z = a + ib \in \mathbb{C}$
- $r-$ open-closed $(a, b]$ $\Rightarrow z = b + ia \in \mathbb{C}$

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

For $f : X \to \mathbb{S}^1$ similar barcodes ($z = e^{ia+(b-a)}$ or $z = e^{ib+(a-b)}$) and in addition

- $r-$ Jordan cells with multiplicities, $J(\lambda, k)$. $= k \times k$ Jordan matrix with λ on diagonal

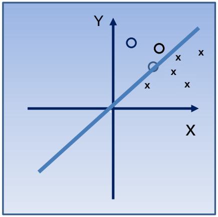
can be also be defined without reference to homology, derived directly from the matrix of the simplicial complex and the simplicial map

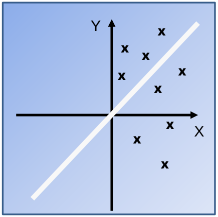
more geometrization

For both real-valued or angle-valued map denote 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, open-closed $r-$bar codes of f and by $\mathcal{J}_r$ the set of $r-$ Jordan cells

- Collect 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 $\mathbb{C} \setminus \Delta\}$
  $\Delta := \{z \in \mathbb{C} \mid \Re z = \Im z\}$ for real valued and $\Delta := \{z \in \mathbb{C} \mid |z| = 1\}$ for angle-valued map.
- Collect the sets $\mathcal{B}^c_r(f)$ and $\mathcal{B}^o_{r-1}(f)$ as the finite configuration of points $C_r(f)$ in $\mathbb{C}$.
- For $1-$cocycle analogues of configurations $C_r(f)$ and $c_r(f)$ exist but as configurations of points on $\mathbb{R}$ and $\mathbb{R} \setminus 0$ resp.
- The set of configurations of points in $\mathbb{C}$ and in $\mathbb{C} \setminus \Delta$ can be topologized appropriately.

For a real-valued map the configuration $C_r(f)$ and $c_r(f)$ are illustrated below

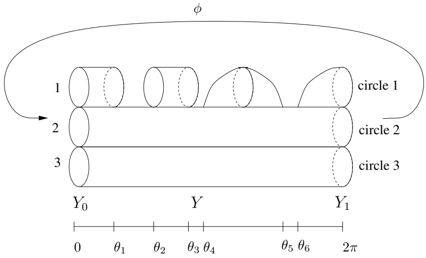




The bar code with ends *a, b, a≤b* and closed at *a* is represented as a point $a + ib$ while the bar code with ends *a, b, a* < *b* open at *a* is represented as a point $b + ia$.

the configurations for an angle valued map and 1-cocycle on blackboard

EXAMPLE barcodes and Jordan cells for an angle-valued map



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		$(1, 1)$
	1	$(\theta_6, \theta_1 + 2\pi]$ $[\theta_2, \theta_3]$ (θ_4, θ_5)	$(3, 2)$

Figure: Example of r-invariants for an angle-valued map

Why refinements?

Theorem

1. *For $f : X \to \mathbb{R}$ a tame map and X a space homeomorphic to a compact simplicial complex*

$$\beta_r(X) = \sharp\mathcal{B}^c_r(f) + \sharp\mathcal{B}^o_{r-1}(f).$$

2. *For $f : X \to S^1$ a tame map and X a space homeomorphic to a compact simplicial complex $M(X, \xi) = \oplus_{J \in \mathcal{J}_r(f)} J$ where ξ is the homotopy class of f.*
3. *Poincaré duality for orientable manifolds extend to bar codes in consistency with 1 above.*

Why computer friendly?

a.

Theorem

The assignment map $\to$ configuration, from the space of (tame)?? real-valued or angle-valued maps to the space of configurations on $\mathcal{C}$ resp. $\mathcal{C} \setminus \Delta$ with the appropriate topologies are continuous.

b. Approximation with arbitrary specified accuracy of any of the four type of bar codes or Jordan cells can be outputted by effective algorithms from the matrix describing the simplicial complex and the real or angle-valued map f.

Persitence theory - the four types of barcodes and the Jordan cells

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

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

consider

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

The collection of ALL these linear relations is referred to as

(zig-zag) **persistent homology**.

One possible definition of the four type of bar codes are based on the concepts

Death from the left to the right or otherwise

Observability from the left to the right or otherwise

and the definition of Jordan cells are based on the linear algebra of linear relations in the case when $H_r(X_a) = H_r(X_b)$

Other definitions can be derived via :

Elliptic theory (differential geometry) or

Graph representations or

Measure theory.

5. Example 1.
Lung cancer imaging.

- 3D radiological images of cancerous lungs shows both tumors and blood vessels as areas of increased density.
- Blood vessels show up as long tunnels in the image
- Tumors show up as balls.
- **Question**: How to distinguish automatically between tumors and blood vessels ?

Example 2. **Diabetes Patients**
(after Miller-Reaven Study) from G Carlsson's paper

- Study carried out in 1976 on 145 patients at Stanford Hospital; Most of patients had symptoms of diabetes although some were normal
- For each patient 6 metabolic variables (involving insulin response glucose tolerence, relative weight .....) were measured and recorded in a 6 dimensional space. Hence a point cloud of 145 points in $\mathbb{R}^6$
- **Questions**: Find the relevant number of metabolic variables needed to detect the diabetes. Find qualitative features (type of diabetes, etc).

References

1. D.Burghelea, New Topological invariants for real-and Angle-valued maps (an alternative to Morse Novikov theory, Book World Scientific , 2018

2. Dan Burghelea, Stefan Haller, Topology of angle valued maps, bar codes and Jordan blocks , J Appl. and Comput. Topology (2017) Vol 1, issue 1.

3.G.Carlson, Topology and Data , Bull . Amer. Math. Soc. 46 pp 255-308

4. Gunnar Carlsson, Mikael Vejdemo-Johansson, Topological Data Analysis with Applications (book)