An introduction to topological data analysis through the concepts of natural pseudo-distance, persistence diagram and group equivariant non-expansive operator

Patrizio Frosini

Department of Mathematics and ARCES, University of Bologna
patrizio.frosini@unibo.it

STEP 1 Webinar Series: Topology, September 25, 2020

Outline

What is Topological Data Analysis?

How can we compare the shape of data?

How can we get information about the natural pseudo-distance d_G?

How can we manage the case $G \neq \mathrm{Homeo}(X)$?

An open problem

What is Topological Data Analysis?

How can we compare the shape of data?

How can we get information about the natural pseudo-distance d_G?

How can we manage the case $G \neq \mathrm{Homeo}(X)$?

An open problem

What is Topological Data Analysis?

Topological data analysis (TDA) is an approach to the analysis of datasets using techniques from topology.

WIKIPEDIA

Main idea: adapting topological techniques to get dimensionality reduction of data and robustness to noise.

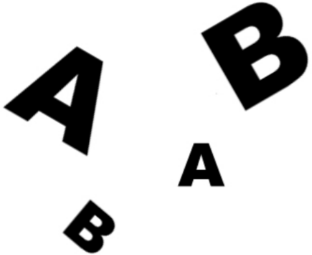
In plain words, $\boxed{\text{we look for } \textit{"shape"} \text{ in data}}$.

In some sense, "shape" is the part of information that is stable in the presence of noise and perturbation, with respect to a given observer.

What is Topological Data Analysis?

A "trivial" question: What is shape?

SIMPLE ANSWER: Shape is what is left after removing scale and rotation.



What is Topological Data Analysis?

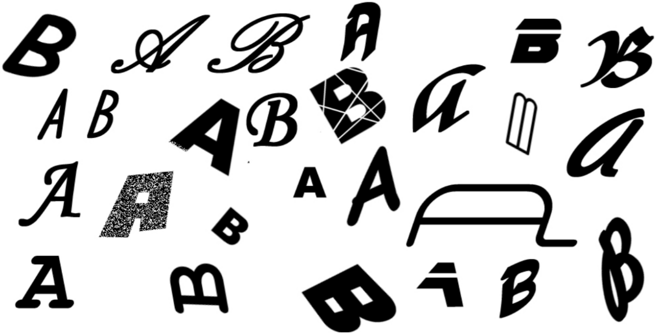
Unfortunately, this is not a good answer.

"There is always an easy solution to every human problem–
neat, plausible, and wrong."

Henry Louis Mencken (1880-1956)

What is Topological Data Analysis?

Shape is a complex concept:



Letters

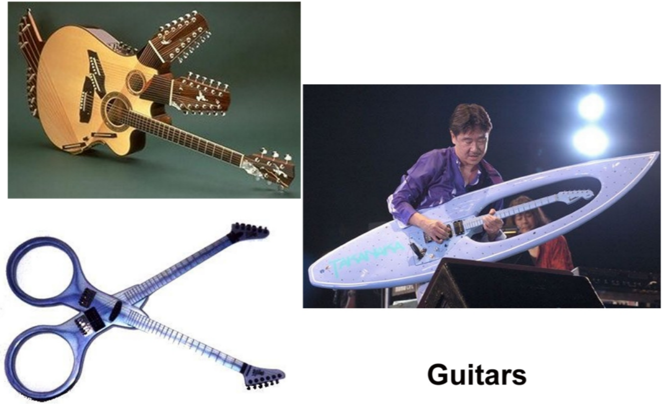
What is Topological Data Analysis?

Shape is a complex concept:

Buttons

What is Topological Data Analysis?

Shape is a complex concept:



Guitars

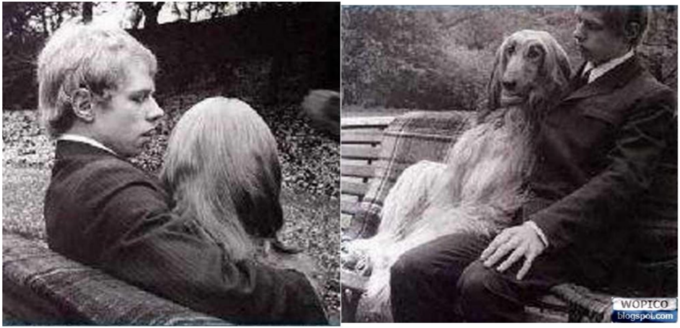
What is Topological Data Analysis?

Shape is a complex concept:

Cups

What is Topological Data Analysis?

Truth and shape often depend on the observer:



What is Topological Data Analysis?

The concept of shape is subjective and relative. It is based on the act of perceiving, depending on the chosen observer. Persistent perceptions are fundamental in order to approach this concept.

- “Science is nothing but perception.” *Plato*
- “Reality is merely an illusion, albeit a very persistent one.” *Albert Einstein*

What is Topological Data Analysis?

How can we compare the shape of data?

How can we get information about the natural pseudo-distance d_G?

How can we manage the case $G \neq \mathrm{Homeo}(X)$?

An open problem

Data as measurements

Data are usually given by measurements. Before proceeding, we have to determine what measurements are in our mathematical model.

Measurement is the assignment of a number to a characteristic of an object or event, which can be compared with other objects or events.

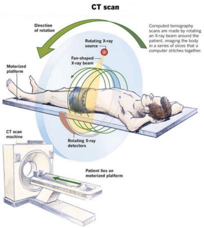
WIKIPEDIA

According to this definition, measurements (and hence data) can be seen as functions φ associating a real number $\varphi(x)$ with each point x of a set X of characteristics. (This definition admits a natural extension to vector-valued functions but, for the sake of simplicity, we will treat here the case of scalar-valued functions).

Data as measurements

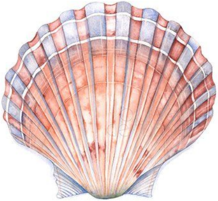
Some examples of data that can be seen as measurements (i.e., functions):

- An electrocardiogram (a function from $\mathbb{R}$ to $\mathbb{R}$);
- A gray-level image (a function from $\mathbb{R}^2$ to $\mathbb{R}$);
- A computerized tomography (CT) scan (a function from a helix to $\mathbb{R}$).



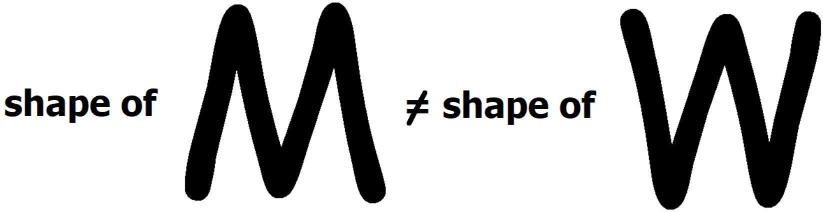
What is Topological Data Analysis?

The choice of an observer implies the choice of an invariance group G:

shape of  **= shape of** 

What is Topological Data Analysis?

The group G is not established once and forever: when the observer changes, G changes too:



How can we compare the shape of data?

Our data are represented by real-valued functions.

The elements of the transformation group G are homeomorphisms.

Two functions $\varphi_1, \varphi_2 : X \to \mathbb{R}$ are **equivalent** *with respect to a transformation group G if a transformation $g \in G$ exists, such that $\varphi_1 = \varphi_2 \circ g$.*

Two functions $\varphi_1, \varphi_2 : X \to \mathbb{R}$ are **similar** *with respect to a transformation group G if a transformation $g \in G$ exists, such that $\|\varphi_1 - \varphi_2 \circ g\|_\infty$ is small.*

These observations lead to define the concept of natural pseudo-distance with respect to the transformation group G.

The definition of d_G

Let X and G be a topological space and a subgroup of the group $\mathrm{Homeo}(X)$ of all homeomorphisms from X to X, respectively. If φ_1, φ_2 are two continuous and bounded functions from X to $\mathbb{R}$ we can consider the value $\inf_{g \in G} \|\varphi_1 - \varphi_2 \circ g\|_\infty$. This value is called the *natural pseudo-distance* $d_G(\varphi_1, \varphi_2)$ between φ_1 and φ_2 with respect to the group G.

We endow both $C^0(X,\mathbb{R})$ and G with the topology of uniform convergence, so that G becomes a topological group acting continuously on $C^0(X,\mathbb{R})$ by composition on the right. We observe that the action of G on $C^0(X,\mathbb{R})$ is continuous.

The definition of d_G

If G is the trivial group Id, then d_G is the max-norm distance $\|\varphi_1 - \varphi_2\|_\infty$. Moreover, if G_1 and G_2 are subgroups of $\mathrm{Homeo}(X)$ and $G_1 \subseteq G_2$, then

$$d_{\mathrm{Homeo}(X)}(\varphi_1, \varphi_2) \leq d_{G_2}(\varphi_1, \varphi_2) \leq d_{G_1}(\varphi_1, \varphi_2) \leq \|\varphi_1 - \varphi_2\|_\infty$$

for every $\varphi_1, \varphi_2 \in C^0(X, \mathbb{R})$.

We usually restrict d_G to $\Phi \times \Phi$, where Φ is a bounded subset of $C^0(X, \mathbb{R})$.

Our ground truth: the natural pseudo-distance d_G

The natural pseudo-distance d_G is our ground truth: it describes the differences that the observer can perceive between the measurements in Φ with respect to the equivalence expressed by the group G.

A possible objection: *"The use of the concept of homeomorphism makes the natural pseudo-distance d_G difficult to apply. For example, in shape comparison two similar objects can be non-homeomorphic, hence this pseudo-metric cannot be applied to real problems."*

A possible objection

Answer: **the homeomorphisms do not concern the "objects" but the space X where the measurements are made**.

- For example, if we are interested in gray-level images, the domain of our measurements can be modelled as the real plane and each image can be represented as a function from $\mathbb{R}^2$ to $\mathbb{R}$. Therefore, the space X is not given by the (possibly non-homeomorphic) objects displayed in the pictures, but by the topological space $\mathbb{R}^2$.
- If we make two CT scans, the topological space X is always given by a helix turning many times around a body, and no requirement is made about the topology of such a body.

In other words, it is usually legitimate to assume that the topological space X is determined only by the measuring instrument we are using to get our measurements.

Some properties of the natural pseudo-distance

Let us now have a look at some relevant properties of the natural pseudo-distance d_G.

We will examine two cases:

1. We do not assume that a homeomorphism $g \in G$ exists, such that $\|\varphi_1 - \varphi_2 \circ g\|_\infty = d_G(\varphi_1, \varphi_2)$ (general case).
2. We assume that a homeomorphism $g \in G$ exists, such that $\|\varphi_1 - \varphi_2 \circ g\|_\infty = d_G(\varphi_1, \varphi_2)$.

Let us start from case 1.

d_G and critical values: manifolds

When the filtering functions are defined on a regular closed manifold $\mathscr{M}$ and $G = \mathrm{Homeo}(\mathscr{M})$, some results restrict the range of values that can be taken by the natural pseudo-distance d_G.

Theorem

Assume that $\mathscr{M}$ is a closed manifold of class C^1 and that $\varphi_1, \varphi_2 : \mathscr{M} \to \mathbb{R}$ are two functions of class C^1. Set $d := d_{\mathrm{Homeo}(\mathscr{M})}(\varphi_1, \varphi_2)$. Then a positive integer k exists for which one of the following properties holds:

(i) *k is odd and kd is the distance between a critical value of φ_1 and a critical value of φ_2;*

(ii) *k is even and kd is either the distance between two critical values of φ_1 or the distance between two critical values of φ_2.*

d_G and critical values: surfaces

Theorem

Assume that $\mathscr{S}$ is a closed surface of class C^1 and that $\varphi_1, \varphi_2 : \mathscr{S} \to \mathbb{R}$ are two functions of class C^1. Set $d := d_{\mathrm{Homeo}(\mathscr{S})}(\varphi_1, \varphi_2)$. Then a positive integer k exists for which at least one of the following properties holds:

(i) *d is the distance between a critical value of φ_1 and a critical value of φ_2;*

(ii) *d is half the distance between two critical values of φ_1.*

(iii) *d is half the distance between two critical values of φ_2.*

(iv) *d is one third of the distance between a critical value of φ_1 and a critical value of φ_2.*

d_G and critical values: curves

Theorem

Assume that $\mathscr{C}$ is a closed curve of class C^1 and that $\varphi_1, \varphi_2 : \mathscr{C} \to \mathbb{R}$ are two functions of class C^1. Set $d := d_{\mathrm{Homeo}(\mathscr{C})}(\varphi_1, \varphi_2)$. Then a positive integer k exists for which at least one of the following properties holds:

a) *d is the distance between a critical value of φ_1 and a critical value of φ_2;*

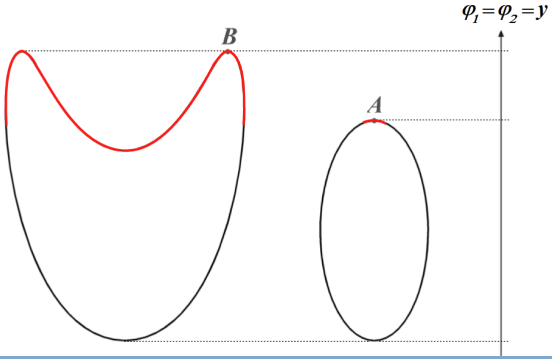
b) *d is half the distance between two critical values of φ_1.*

c) *d is half the distance between two critical values of φ_2.*

The last theorem is sharp, as shown by the following examples.

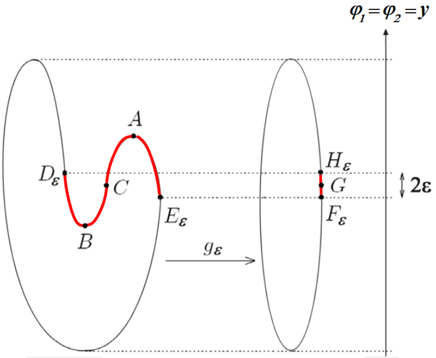
d_G and critical values: curves

Let us consider the two embeddings of S^1 in $\mathbb{R}^2$ represented in the following figure. The ordinate y defines two filtering functions φ_1, φ_2 on S^1. In this case $d_{\mathrm{Homeo}(S^1)}(\varphi_1, \varphi_2) = |\varphi_1(A) - \varphi(B)|$ is the distance between a critical value of φ_1 and a critical value of φ_2.



d_G and critical values: curves

Let us consider the two embeddings of S^1 in $\mathbb{R}^2$ represented in the following figure. The ordinate y defines two filtering functions φ_1, φ_2 on S^1. In this case $d_{\mathrm{Homeo}(S^1)}(\varphi_1, \varphi_2) = \frac{1}{2}|\varphi_1(A) - \varphi(B)|$ is half the distance between two critical values of φ_1.



Optimal homeomorphisms

Let us now consider case 2.

Assume that X is a compact topological space and $\varphi_1, \varphi_2 : X \to \mathbb{R}$ are continuous functions. Let G be a subgroup of $\mathrm{Homeo}(X)$. We say that a homeomorphism $g \in G$ is *optimal* in G for (φ_1, φ_2) if $\|\varphi_1 - \varphi_2 \circ g\|_\infty = d_G(\varphi_1, \varphi_2)$. The following results hold for optimal homeomorphisms.

Theorem

Assume that $\mathscr{M}$ is a C^1 closed manifold and that $\varphi_1, \varphi_2 : \mathscr{M} \to \mathbb{R}$ are of class C^1. If an optimal homeomorphism $g \in \mathrm{Homeo}(\mathscr{M})$ for (φ_1, φ_2) exists, then $d_{\mathrm{Homeo}(\mathscr{M})}(\varphi_1, \varphi_2)$ is the distance between a critical value of φ_1 and a critical value of φ_2.

Optimal homeomorphisms

Theorem

If $\varphi_1, \varphi_2 : S^1 \to \mathbb{R}$ are Morse functions and $d_{\mathrm{Homeo}(S^1)}(\varphi_1, \varphi_2)$ vanishes, then an optimal C^2-diffeomorphism exists in $\mathrm{Homeo}(S^1)$ for (φ_1, φ_2).

Theorem (A. De Gregorio)

The number of optimal homeomorphisms in the Lie group S^1 for a pair (φ_1, φ_2) of Morse functions from S^1 to $\mathbb{R}$ is finite.

What is Topological Data Analysis?

How can we compare the shape of data?

How can we get information about the natural pseudo-distance d_G?

How can we manage the case $G \neq \mathrm{Homeo}(X)$?

An open problem

How can we get information about the natural pseudo-distance?

In practice, we cannot directly compute the natural pseudo-distance, since its computation involves considering every transformation belonging to the group G, and the group G is usually large.

Fortunately, persistent homology and the concept of group equivariant non-expansive operator allow us to get information about the natural pseudo-distance.

What is homology?

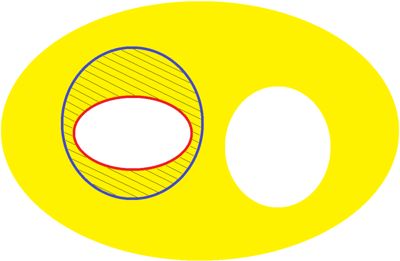
Two shapes can be often distinguished by examining their holes. Homology theory allows us to formally describe the presence of holes in a topological space by means of suitable groups, called homology groups. The definition of these groups is based on the concepts of cycle and boundary.

Speaking roughly, a k-cycle is a closed k-submanifold, a k-boundary is a k-cycle which is also the boundary of a $k+1$-submanifold, and a homology class (which represents a hole) is an equivalence class of k-cycles modulo k-boundaries.

The set of k-holes is a group (with respect to a suitable operation), called the k-th homology group.

What is homology?

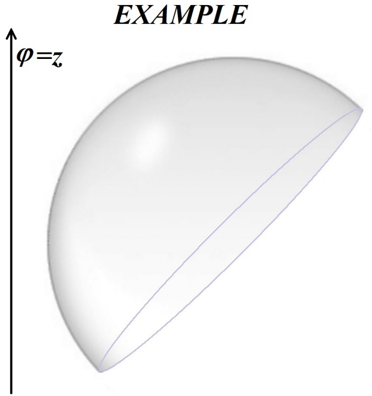
$\mathbb{R}^2$



The red cycle and the blue cycle are equivalent to each other, since their “sum” is the boundary of the marked region. Both these cycles describe the same 1-dimensional hole in the yellow subset of the real plane.

What is persistent homology?

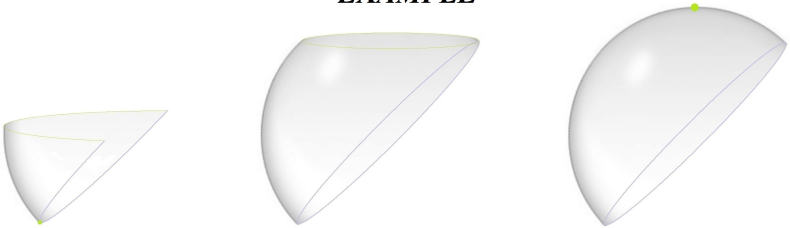
If $\varphi : X \to \mathbb{R}$ is a continuous function, we can consider the sublevel sets $X_t := \{x \in X : \varphi(x) \leq t\}$. When t varies we see the birth and death of k-dimensional holes.



What is persistent homology?

If $\varphi : X \to \mathbb{R}$ is a continuous function, we can consider the sublevel sets $X_t := \{x \in X : \varphi(x) \leq t\}$. When t varies we see the birth and death of k-dimensional holes.

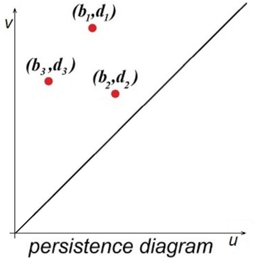
EXAMPLE



No 1-dimensional hole *Birth of a 1-dimensional hole* *Death of the 1-dimensional hole*

What is a persistence diagram?

In plain words, the persistence diagram in degree k of φ is the collection of the pairs (b_i, d_i) where b_i and d_i are the times of birth and death of the i-th hole of dimension k.



The points of the persistence diagram are endowed with multiplicity. Each point of the diagonal $u = v$ is assumed to be a point of the persistence diagram, endowed with infinite multiplicity.

What are persistent Betti numbers functions?

Persistence diagrams are not quite suitable for statistical purposes, because no <u>good</u> definition of average of persistence diagrams exists.

Persistent Betti numbers functions are more suitable for statistics.

Definition

The k-th persistent Betti numbers function $\beta_k(u, v)$ is the number of holes of dimension k whose time of birth is smaller than u and whose time of death is greater than v.

What are persistent Betti numbers functions?

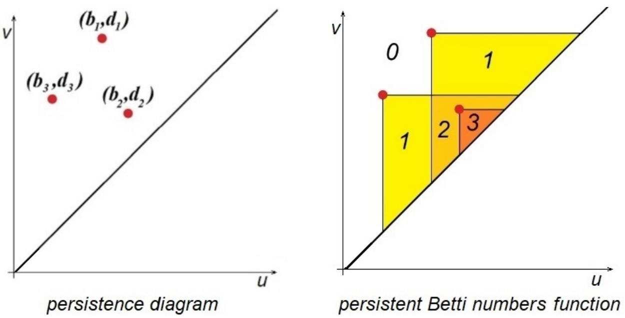
Let us give the formal definition of persistent homology group:

Definition

Let $\varphi : X \to \mathbb{R}$ be a continuous function. If $u, v \in \mathbb{R}$ and $u < v$, we can consider the inclusion i of X_u into X_v. Such an inclusion induces a homomorphism $i^* : H_k(X_u) \to H_k(X_v)$ between the homology groups of X_u and X_v in degree k. The group $PH_k^{\varphi}(u, v) := i^*(H_k(X_u))$ is called the *k-th persistent homology group with respect to the function* $\varphi : X \to \mathbb{R}$*, computed at the point* (u, v). The rank $r_k(\varphi)(u, v)$ of this group is said *the k-th persistent Betti numbers function with respect to the function* $\varphi : X \to \mathbb{R}$*, computed at the point* (u, v).

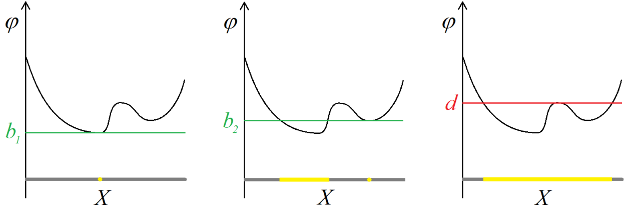
The average of persistent Betti numbers functions can be trivially defined as the usual average of real-valued functions.

What are persistent Betti numbers functions?



If we use Čech homology, persistence diagrams are equivalent to persistent Betti numbers functions.

An example for $\varphi : X = [0,1] \to \mathbb{R}$ $(k = 0)$

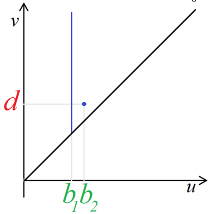


At time b_1 a connect component is born.

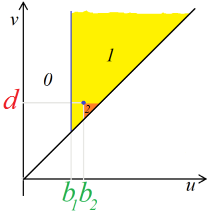
At time b_2 another connected component is born.

At time d the younger component dies by joining the older one.

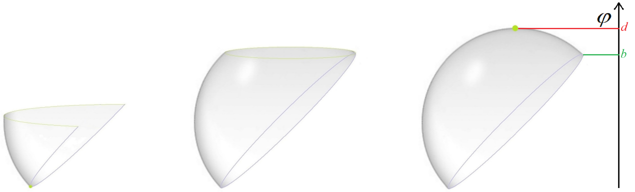
persistence diagram $\mathrm{Dgm}_0(\varphi)$



persistent Betti numbers function $\beta_0(u,v)$

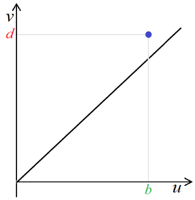


Another example of persistence diagram ($k = 1$)

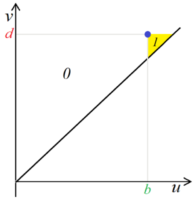


No 1-dimensional hole ***Birth of a 1-dimensional hole*** ***Death of the 1-dimensional hole***

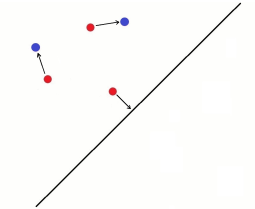
persistence diagram $\mathrm{Dgm}_1(\varphi)$



persistent Betti numbers function $\beta_1(u,v)$



Comparison of persistent Betti numbers functions



A way to change D_1 into D_2.

Persistence diagrams (and hence persistent Betti numbers functions) can be compared by means of the bottleneck distance. The bottleneck distance between two persistence diagrams D_1, D_2 is the minimum cost of changing the points of D_1 into the points of D_2, where the cost of moving each point is given by the max-norm distance in $\mathbb{R}^2$. Moving a point to the diagonal is equivalent to delete it.

Comparison of persistent Betti numbers functions

An important property of the metric d_{match} is its stability, as stated in the following result.

Theorem

If k is a natural number and $\varphi_1, \varphi_2 \in C^0(X, \mathbb{R})$, then

$$d_{\mathrm{match}}(r_k(\varphi_1), r_k(\varphi_2)) \le d_{\mathrm{Homeo}(X)}(\varphi_1, \varphi_2) \le \|\varphi_1 - \varphi_2\|_\infty .$$

The same inequality can be also stated in terms of persistence diagrams. **The matching distance between persistence diagrams is not difficult to compute. Therefore the previous inequality is a powerful tool to get a lower bound for the natural pseudo-distance d_G, when $G = \mathrm{Homeo}(X)$.**

What is Topological Data Analysis?

How can we compare the shape of data?

How can we get information about the natural pseudo-distance d_G?

How can we manage the case $G \neq \mathrm{Homeo}(X)$?

An open problem

How can we manage the case $G \neq \mathrm{Homeo}(X)$?

Since $d_{\mathrm{Homeo}(X)} \leq d_G$ for every $G \subseteq \mathrm{Homeo}(X)$, the inequality shown in the previous slide immediately implies that

$$d_{\mathrm{match}}(r_k(\varphi_1), r_k(\varphi_2)) \leq d_G(\varphi_1, \varphi_2).$$

However, the gap between $d_{\mathrm{Homeo}(X)}(\varphi_1, \varphi_2)$ and $d_G(\varphi_1, \varphi_2)$ can be pretty large.

In order to get a better result we have to use the concept of group equivariant non-expansive operator.

Group Equivariant Non-Expansive Operators

Let X and G be a topological space and a subgroup of the group $\mathrm{Homeo}(X)$ of all homeomorphisms from X to X, respectively. Let $\Phi \subseteq C^0(X, \mathbb{R})$. We now consider the set $\mathscr{F}(\Phi, G)$ of all maps from Φ to Φ that verify the following two properties:

1. $F(\varphi \circ g) = F(\varphi) \circ g$ for every $\varphi \in \Phi$ and every $g \in G$ (i.e., F is equivariant with respect to G);
2. $\|F(\varphi_1) - F(\varphi_2)\|_\infty \leq \|\varphi_1 - \varphi_2\|_\infty$ for every $\varphi_1, \varphi_2 \in \Phi$ (i.e., F is non-expansive).

Obviously, $\mathscr{F}(\Phi, G)$ is not empty, since it contains at least the identity map. The maps in $\mathscr{F}(\Phi, G)$ will be called *Group Equivariant Non-Expansive Operators* (GENEOs).

An example of GENEO

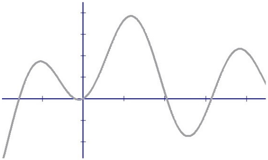
GRAYSCALE IMAGE

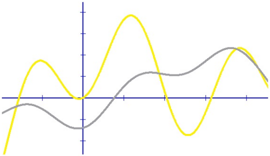
$\varphi : [0, a] \times [0, b] \longrightarrow \mathbb{R}$ $\quad$ $T(\varphi)$ = CONVOLUTION OF φ

$$F(\varphi)(x) := \frac{1}{2\pi\sigma^2} \int_{\mathbb{R}^2} \varphi(y) e^{-\frac{\|x-y\|^2}{2\sigma^2}} \, dy.$$

F is G-equivariant for G equal to the group of isometries of $\mathbb{R}^2$.
F is also non-expansive with respect to the sup-norm.

Another example of GENEO

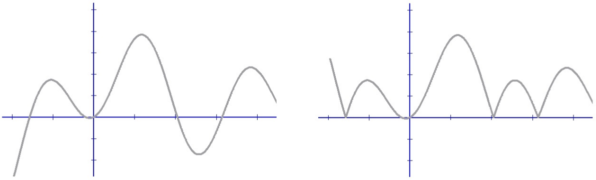




$\varphi : \mathbb{R} \to [0,1] \qquad F(\varphi)(x) := \frac{1}{2}(\varphi(x-1) + \varphi(x+1))$

F is G-equivariant for G equal to the group of translations of $\mathbb{R}$.
F is also non-expansive with respect to the sup-norm.

Another example of GENEO



$\varphi : \mathbb{R} \to [0,1]$ $\qquad$ $F(\varphi) = |\varphi|$

F is G-equivariant for G equal to the group of homeomorphisms of $\mathbb{R}$.
F is also non-expansive with respect to the sup-norm.

Choice of GENEOs

- The observer cannot usually choose the functions representing the measurement data, but he/she can often choose the operators that will be applied to those functions.
- The choice of the operators reflects the invariances that are relevant for the observer.
- In some sense we could state that the observer can be represented as a collection of (suitable) operators, endowed with the invariance he/she has chosen.

Lower bounds for d_G via persistent homology

For every fixed k and every subset $\mathscr{F} \subseteq \mathscr{F}(\Phi, G)$, we can consider the following pseudo-metric $D^{\mathscr{F},k}_{\mathrm{match}}$ on Φ:

$$D^{\mathscr{F},k}_{\mathrm{match}}(\varphi_1, \varphi_2) := \sup_{F \in \mathscr{F}} d_{\mathrm{match}}(r_k(F(\varphi_1)), r_k(F(\varphi_2)))$$

for every $\varphi_1, \varphi_2 \in \Phi$, where $r_k(\varphi)$ denotes the k-th persistent Betti numbers function with respect to the function $\varphi : X \to \mathbb{R}$. We will usually omit the index k, when its value is clear from the context or not influential.

We observe that $D^{\mathscr{F}}_{\mathrm{match}}$ is strongly invariant with respect to G, i.e., $D^{\mathscr{F}}_{\mathrm{match}}(\varphi_1, \varphi_2 \circ g) = D^{\mathscr{F}}_{\mathrm{match}}(\varphi_1 \circ g, \varphi_2) = D^{\mathscr{F}}_{\mathrm{match}}(\varphi_1, \varphi_2)$ for every $\varphi_1, \varphi_2 \in \Phi$ and every $g \in \mathrm{Homeo}(X)$.

Lower bounds for d_G via persistent homology

The importance of $D^{\mathscr{F}}_{\mathrm{match}}$ lies in the following two results, showing that it can be used to get information about the natural pseudo-distance d_G.

Theorem

If $\emptyset \neq \bar{\mathscr{F}} \subseteq \mathscr{F}(\Phi, G)$, then $D^{\bar{\mathscr{F}}}_{\mathrm{match}} \leq d_G$.

Theorem

$D^{\mathscr{F}(\Phi,G)}_{\mathrm{match}} = d_G$.

As a consequence, the topological and geometrical study of $\mathscr{F}(\Phi, G)$ is important in the research concerning the natural pseudo-distance.

Two relevant properties of $\mathscr{F}(\Phi, G)$

Two relevant properties of $\mathscr{F}(\Phi, G)$ are expressed by the following result.

Theorem

If Φ is compact, then $\mathscr{F}(\Phi, G)$ is compact.
If Φ is convex, then $\mathscr{F}(\Phi, G)$ is convex.

The compactness and convexity of $\mathscr{F}(\Phi, G)$ are important from the computational point of view.

Two further reasons to study GENEOs

There are two further reasons to study GENEOs:

1. The operator that allows to reduce multidimensional persistent Betti numbers functions to families of 1-dimensional persistent Betti numbers functions is a GENEO.
2. GENEOs appear to be of use in deep learning, since they can be seen as multi-level components that can be joined and connected in order to form neural networks by applying the operations of chaining, convex combination and direct product.

What is Topological Data Analysis?

How can we compare the shape of data?

How can we get information about the natural pseudo-distance d_G?

How can we manage the case $G \neq \mathrm{Homeo}(X)$?

An open problem

An open problem

Let us consider a closed C^1 surface $\mathscr{M}$ and two C^1 filtering functions $\varphi_1, \varphi_2 : \mathscr{M} \to \mathbb{R}$. Let $\mathrm{Homeo}(\mathscr{M})$ be the group of all self-homeomorphisms of $\mathscr{M}$. It has been proved that at least one of the following statements holds:

1. $d_{\mathrm{Homeo}(\mathscr{M})}(\varphi_1, \varphi_2)$ is the distance between a critical value of φ_1 and a critical value of φ_2;
2. $d_{\mathrm{Homeo}(\mathscr{M})}(\varphi_1, \varphi_2)$ is half the distance between two critical values of φ_1;
3. $d_{\mathrm{Homeo}(\mathscr{M})}(\varphi_1, \varphi_2)$ is half the distance between two critical values of φ_2;
4. $d_{\mathrm{Homeo}(\mathscr{M})}(\varphi_1, \varphi_2)$ is one third of the distance between a critical value of φ_1 and a critical value of φ_2.

An open problem

Interestingly, no example of two functions $\varphi_1, \varphi_2 : \mathscr{M} \to \mathbb{R}$ is known, such that (4) holds but $(1), (2), (3)$ do not hold.

A natural question arises: Can we find an example of two such functions or prove that such an example cannot exist (so improving our result)?

An open problem

We observe that the usual technique to compute the natural pseudo-distance consists in

- finding a lower bound for $d_{\mathrm{Homeo}(\mathscr{M})}(\varphi_1, \varphi_2)$ by computing the matching distance $d_{\mathrm{match}}(\mathrm{Dgm}_k(\varphi_1), \mathrm{Dgm}_k(\varphi_2))$ between the persistence diagrams in degree k of the functions φ_1 and φ_2;
- looking for a sequence (g_i) in $\mathrm{Homeo}(\mathscr{M})$, such that $\lim_{i\to\infty} \|\varphi_1 - \varphi_2 \circ g_i\|_\infty = d_{\mathrm{match}}(\mathrm{Dgm}_k(\varphi_1), \mathrm{Dgm}_k(\varphi_2))$.

If such a sequence (g_i) exists, then the value $d_{\mathrm{Homeo}(\mathscr{M})}(\varphi_1, \varphi_2)$ is equal to $d_{\mathrm{match}}(\mathrm{Dgm}_k(\varphi_1), \mathrm{Dgm}_k(\varphi_2))$.

An open problem

Unfortunately, at least one of the following statements holds:

a) $d_{\mathrm{match}}(\mathrm{Dgm}_k(\varphi_1), \mathrm{Dgm}_k(\varphi_2))$ is the distance between a critical value of φ_1 and a critical value of φ_2;

b) $d_{\mathrm{match}}(\mathrm{Dgm}_k(\varphi_1), \mathrm{Dgm}_k(\varphi_2))$ is half the distance between two critical values of φ_1;

c) $d_{\mathrm{match}}(\mathrm{Dgm}_k(\varphi_1), \mathrm{Dgm}_k(\varphi_2))$ is half the distance between two critical values of φ_2.

Therefore, if (1), (2), (3) do not hold for $\varphi_1, \varphi_2 : \mathscr{M} \to \mathbb{R}$, then $d_{\mathrm{Homeo}(\mathscr{M})}(\varphi_1, \varphi_2)$ cannot be equal to $d_{\mathrm{match}}(\mathrm{Dgm}_k(\varphi_1), \mathrm{Dgm}_k(\varphi_2))$. This means that if there exist two C^1 functions $\varphi_1, \varphi_2 : \mathscr{M} \to \mathbb{R}$ verifying (4) but not (1), (2), (3), then we need new methods to compute $d_{\mathrm{Homeo}(\mathscr{M})}(\varphi_1, \varphi_2)$ and to recognize the pair (φ_1, φ_2) as the right example. As a consequence, the answer to the question asked in this section is still unknown.

SOME REFERENCES

- P. Donatini, P. Frosini, *Natural pseudodistances between closed surfaces*, **Journal of the European Mathematical Society**, vol. 9 (2007), n. 2, 331–353.
- S. Biasotti, L. De Floriani, B. Falcidieno, P. Frosini, D. Giorgi, C. Landi, L. Papaleo, M. Spagnuolo, *Describing shapes by geometrical-topological properties of real functions*, **ACM Computing Surveys**, vol. 40 (2008), n. 4, 12:1-12:87.
- P. Frosini, G. Jabłoński, *Combining persistent homology and invariance groups for shape comparison*, **Discrete & Computational Geometry**, vol. 55 (2016), n. 2, 373-409.
- M. G. Bergomi, P. Frosini, D. Giorgi, N. Quercioli, *Towards a topological-geometrical theory of group equivariant non-expansive operators for data analysis and machine learning*, **Nature Machine Intelligence**, vol. 1, n. 9, 423–433 (2 September 2019).


Thanks for your attention!