

The theory of group equivariant non-expansive operators in topological data analysis

Patrizio Frosini

(Joint work with Mattia Bergomi (Fundação Champalimaud, PT), Daniela Giorgi (ISTI-CNR, IT) and Nicola Quercioli (University of Bologna, IT))

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

GETCO, Oaxaca, September 10-14, 2018

Outline



A mathematical framework for data comparison

The natural pseudo-distance d_G

Group equivariant non-expansive operators

The link between the natural pseudo-distance and persistent homology via GENEOS

Building new GENEOS



A mathematical framework for data comparison

The natural pseudo-distance d_G

Group equivariant non-expansive operators

The link between the natural pseudo-distance and persistent homology via GNEOs

Building new GNEOs

What does DATA COMPARISON mean?



The comparison of DATA is always a process depending on an observer. We could say that **data comparison consists in the study of the relationship between an observer and the reality he/she can MEASURE**. In this setting data coincide with measurements.

Observers receive and transform data. In some sense, they are defined by the way they perform this transformation.

In our approach **observers are defined as a collection of suitable operators acting on measurements**.

What does MEASUREMENT mean?



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). If we wish to develop a theory that can be applied in real situations, we need stability with respect to noise. This naturally leads us to use a topology on the set Φ of possible measurements on X .

A topology on the space X of characteristics



A natural topology on the set Φ of possible measurements is the one induced by the L^∞ metric $D_\Phi(\varphi_1, \varphi_2) := \|\varphi_1 - \varphi_2\|_\infty$.

Since measurements are the central concept in our approach, the topology on X is derived from D_Φ .

We define this pseudometric D_X on X by setting

$$D_X(x_1, x_2) := \sup_{\varphi \in \Phi} |\varphi(x_1) - \varphi(x_2)|.$$

In plain words: Two points $x_1, x_2 \in X$ are close to each other if and only if every measurement in Φ takes similar values at those points.



Every function in Φ is continuous

In this talk we will assume that the topological space Φ is compact.

EXAMPLE 1. $X := S^1 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$, $\Phi =$ set of all 1-Lipschitzian functions from S^1 to $[0, 1]$.

EXAMPLE 2. $X := [-1, 1] \times [-1, 1]$, $\Phi =$ set of all functions from X to $[0, 1]$ that are 1-Lipschitzian both in $X_1 := [-1, 0] \times [-1, 1]$ and in $X_2 := (0, 1] \times [-1, 1]$. Please observe that the functions in Φ can be discontinuous at the points (x, y) with $x = 0$, with respect to the Euclidean topology on X . However, every function in Φ is continuous with respect to the topology induced by D_X .

Theorem

If Φ is compact, then the topology induced by D_X coincides with the *initial topology* on X , i.e. the coarsest topology on X such that each function $\varphi \in \Phi$ is continuous.



Homeomorphisms with respect to D_X

The next step consists in understanding what a Φ -preserving homeomorphism with respect to D_X is (a bijection $g : X \rightarrow X$ is called Φ -preserving if $\varphi \circ g \in \Phi$ for every $\varphi \in \Phi$).

Theorem

The Φ -preserving homeomorphisms with respect to D_X are exactly the Φ -preserving bijections from X to X .

Let us now consider a group G of homeomorphisms from X to X , whose elements preserve Φ by right composition.

We will say that (Φ, G) is a PERCEPTION PAIR.

A pseudo-metric on our Φ -preserving group G



If a perception pair (Φ, G) is given, we can define the function

$$D_G(g_1, g_2) = \sup_{\varphi \in \Phi} D_\varphi(\varphi \circ g_1, \varphi \circ g_2) \quad (0.1)$$

from $G \times G$ to \mathbb{R} .

The function D_G is a pseudo-metric on G .

Please note that also the definition of D_G is inherited from the definition of D_φ .

Theorem

G is a topological group with respect to the pseudo-metric topology and the action of G on Φ through right composition is continuous.



Compactness of X and G

We recall that we are assuming Φ compact.

Theorem

If X is complete then it is also compact with respect to D_X .

Theorem

If G is complete then it is also compact with respect to D_G .

In this talk we will assume that X and G are complete, and hence compact.



A mathematical framework for data comparison

The natural pseudo-distance d_G

Group equivariant non-expansive operators

The link between the natural pseudo-distance and persistent homology via GNEOs

Building new GNEOs

Our ground truth: the natural pseudo-distance d_G



Definition

The pseudo-distance $d_G : \Phi \times \Phi \rightarrow \mathbb{R}$ is defined by setting

$$d_G(\varphi_1, \varphi_2) = \inf_{g \in G} D_\Phi(\varphi_1, \varphi_2 \circ g).$$

It is called the **natural pseudo-distance** associated with the group G .

If $G = \{Id : x \mapsto x\}$, then d_G equals the sup-norm distance D_Φ on Φ .
If G_1 and G_2 are groups of Φ -preserving self-homeomorphisms of X and $G_1 \subseteq G_2$, then the definition of d_G implies that

$$d_{G_2}(\varphi_1, \varphi_2) \leq d_{G_1}(\varphi_1, \varphi_2) \leq D_\Phi(\varphi_1, \varphi_2)$$

for every $\varphi_1, \varphi_2 \in \Phi$.

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 grey 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 CAT scans, the topological space X is always given by an 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.



A mathematical framework for data comparison

The natural pseudo-distance d_G

Group equivariant non-expansive operators

The link between the natural pseudo-distance and persistent homology via GNEOs

Building new GNEOs



Group equivariant non-expansive operators

The natural pseudo-distance d_G represents our ground truth.

Unfortunately, d_G is difficult to compute. This is also a consequence of the fact that we can easily find subgroups G of $\text{Homeo}(X)$ that cannot be approximated with arbitrary precision by smaller **finite** subgroups of G (i.e. $G = \text{group of rigid motions of } X = \mathbb{R}^3$).

Nevertheless, in this talk we will show that d_G can be approximated with arbitrary precision by means of a **DUAL** approach based on persistent homology and group equivariant non-expansive operators (**GENEOs**).

The space of GENEOS



Definition

Assume that (Φ, G) , (Ψ, H) are two perception pairs and that a homomorphism $T : G \rightarrow H$ has been fixed. A *Group Equivariant Non-Expansive Operator (GENEO)* from (Φ, G) to (Ψ, H) is a map $F : \Phi \rightarrow \Psi$ such that the following properties hold for every $\varphi_1, \varphi_2 \in \Phi$:

1. $F(\varphi \circ g) = F(\varphi) \circ T(g)$ for every $g \in G$;
2. $D_\Psi(F(\varphi_1), F(\varphi_2)) \leq D_\Phi(\varphi_1, \varphi_2)$.

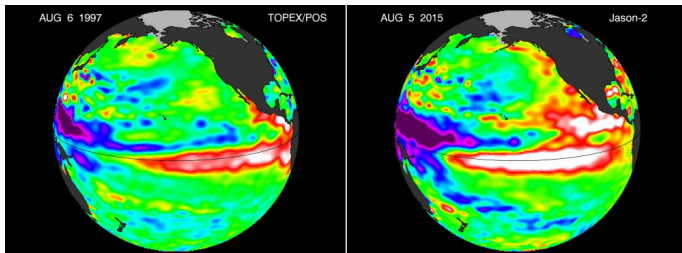
We will use the symbol \mathcal{F}^{all} to denote the set of all GENEOS from (Φ, G) to (Ψ, H) with respect to T .

An example of GENE0



We give an example of the use of the definition of GENE0 between two different perception pairs (Φ, G) , (Ψ, H) .

Let us assume to be interested in the comparison of the distributions of temperatures on a sphere, taken at two different times:



Let us also imagine that only two opposite points N, S can be localized on the sphere.



An example of GENEEO

In this case we can set

- $X = S^2$
- $\Phi =$ set of 1-Lischitzian functions from S^2 to a fixed interval $[a, b]$
- $G =$ group of rotations of S^2 around the axis $N - S$

We can also consider the “equator” of our sphere, represented as the space S^1 .

Therefore, we can also set

- $Y =$ the equator S^1 of S^2
- $\Psi =$ set of 1-Lischitzian functions from S^1 to $[a, b]$
- $H =$ group of rotations of S^1

An example of GENEIO



In this case we can build a simple example of GENEIO from (Φ, G) to (Ψ, H) by setting

- $T(g)$ equal to the rotation $h \in H$ of the equator S^1 that is induced by the rotation g of S^2 , for every $g \in G$.
- $F(\varphi)$ equal to the function ψ that takes each point y belonging to the equator S^1 to the average of the temperatures along the meridian containing y , for every $\varphi \in \Phi$;

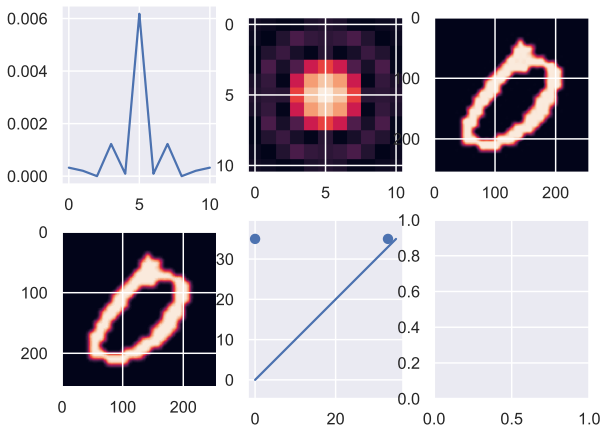
We can easily check that F verifies the properties defining the concept of group equivariant non-expansive operator with respect to the homomorphism $T : G \rightarrow H$.



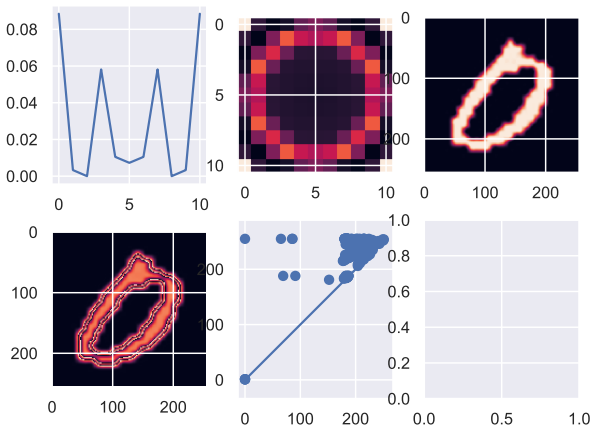
Another example of GENE0

Let us consider a set Φ of grey level images represented as 1-Lipschitzian compact supported functions from $X = \mathbb{R}^2$ to a suitable interval $[a, b]$. We set $G =$ group of isometries of \mathbb{R}^2 . We can build a simple example of GENE0 from (Φ, G) to (Φ, G) with respect to $id : G \rightarrow G$ by taking a compact supported integrable function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ that is invariant with respect to every rotation around the point $(0,0)$, and defining $F(\varphi)$ as the convolution of φ with f . If we also assume that $\int_{\mathbb{R}^2} f(x, y) dx dy = 1$, we can easily check that F verifies the properties defining the concept of group equivariant non-expansive operator with respect to the homomorphism $id : G \rightarrow G$.

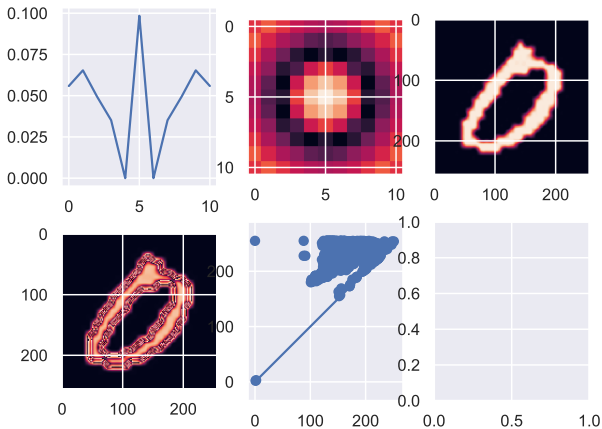
Another example of GENE0



Another example of GENE0



Another example of GENE0



Two pseudo-metrics for the space \mathcal{F}^{all}



The following two pseudo-metrics can be of use:

Definition

If $F_1, F_2 \in \mathcal{F}^{\text{all}}$, we set

$$\begin{aligned} D_{\text{GENEO}}(F_1, F_2) &:= \sup_{\varphi \in \Phi} D_{\Psi}(F_1(\varphi), F_2(\varphi)) \\ D_{\text{GENEO},H}(F_1, F_2) &:= \sup_{\varphi \in \Phi} d_H(F_1(\varphi), F_2(\varphi)). \end{aligned} \quad (0.2)$$

Proposition

D_{GENEO} and $D_{\text{GENEO},H}$ are pseudo-metrics on \mathcal{F}^{all} . Moreover, $D_{\text{GENEO},H} \leq D_{\text{GENEO}}$.



Some good news

Let \mathcal{F}^{all} be the set of all GENEOS from (Φ, G) to (Ψ, H) with respect to a fixed homomorphism $T : G \rightarrow H$.

Theorem

\mathcal{F}^{all} **is compact** with respect to both D_{GENEO} and $D_{\text{GENEO},H}$.

Corollary

\mathcal{F}^{all} can be ε -approximated by a finite subset for every $\varepsilon > 0$.

Theorem

If Ψ is convex, then \mathcal{F}^{all} **is convex** .



A mathematical framework for data comparison

The natural pseudo-distance d_G

Group equivariant non-expansive operators

The link between the natural pseudo-distance and persistent homology via GNEOs

Building new GNEOs



The link between the natural pseudo-distance and persistent homology via GNEOs

Persistent homology enters this theoretical framework by means of an equality allowing us to approximate the natural pseudo-distance:

Theorem

If $(\Phi, G) = (\Psi, H)$, then

$$d_G(\varphi_1, \varphi_2) = \sup_{F \in \mathcal{F}^{\text{all}}} d_{\text{match}}(\text{Dgm}(F(\varphi_1)), \text{Dgm}(F(\varphi_2)))$$

where $\text{Dgm}(F(\varphi))$ is the persistence diagram of the function $F(\varphi)$ and d_{match} is the usual bottleneck distance.

(More details in the paper [P. Frosini, G. Jabłoński, *Combining persistent homology and invariance groups for shape comparison*, *Discrete & Comput. Geometry*, vol. 55 (2016), n. 2, pages 373-409.])

The pseudo-metric $D_{\text{match}}^{\mathcal{F}}$



Let us take a finite ε -approximation \mathcal{F} of \mathcal{F}^{all} . We can then define the pseudo-metric

$$D_{\text{match}}^{\mathcal{F}}(\varphi_1, \varphi_2) := \sup_{F \in \mathcal{F}} d_{\text{match}}(\text{Dgm}(F(\varphi_1)), \text{Dgm}(F(\varphi_2))).$$

The following properties hold for every $\varphi_1, \varphi_2 \in \Phi$ and every $g \in G$:

- $D_{\text{match}}^{\mathcal{F}}(\varphi_1, \varphi_2 \circ g) = D_{\text{match}}^{\mathcal{F}}(\varphi_1, \varphi_2)$;
- $D_{\text{match}}^{\mathcal{F}}(\varphi_1, \varphi_2) \leq d_G(\varphi_1, \varphi_2) \leq \|\varphi_1 - \varphi_2\|_{\infty}$;
- $|d_G(\varphi_1, \varphi_2) - D_{\text{match}}^{\mathcal{F}}(\varphi_1, \varphi_2)| \leq 2\varepsilon$.



A mathematical framework for data comparison

The natural pseudo-distance d_G

Group equivariant non-expansive operators

The link between the natural pseudo-distance and persistent homology via GENEOS

Building new GENEOS



Building new GNEOs

Our approach to group equivariant Topological Data Analysis is based on the availability of GNEOs.

How could we build new GNEOs from other GNEOs?

A simple method consists in composing GNEOs:

Proposition

If F_1 is a GNEO from (Φ_1, G_1) to (Φ_2, G_2) with respect to $T_1 : G_1 \rightarrow G_2$ and F_2 is a GNEO from (Φ_2, G_2) to (Φ_3, G_3) , then $F_2 \circ F_1$ is a GNEO from (Φ_1, G_1) to (Φ_3, G_3) with respect to $T_2 \circ T_1 : G_1 \rightarrow G_3$.

Building GENEOS via 1-Lipschitzian functions



We can also produce new GENEOS by means of a 1-Lipschitzian function applied to other GENEOS:

Proposition

Assume that two perception pairs (Φ, G) , (Ψ, H) and a homomorphism $T : G \rightarrow H$ are given. Let \mathcal{L} be a 1-Lipschitzian map from \mathbb{R}^n to \mathbb{R} , where \mathbb{R}^n is endowed with the norm

$\|(x_1, \dots, x_n)\|_\infty := \max_{1 \leq i \leq n} |x_i|$. Assume also that F_1, \dots, F_n are GENEOS from (Φ, G) to (Ψ, H) with respect to T . Let us define


$\mathcal{L}^(F_1, \dots, F_n)$ by setting*

$\mathcal{L}^(F_1, \dots, F_n)(\varphi)(x) := \mathcal{L}(F_1(\varphi)(x), \dots, F_n(\varphi)(x))$. If*

$\mathcal{L}^(F_1, \dots, F_n)(\Phi) \subseteq \Psi$, then $\mathcal{L}^*(F_1, \dots, F_n)$ is a GENEOS from (Φ, G) to (Ψ, H) with respect to T .*

From this proposition the following three results follow.

Building new GENEOS via translations, the maximum operator and weighted averages



Assume that two perception pairs (Φ, G) , (Ψ, H) and a homomorphism $T : G \rightarrow H$ are given.

Proposition (Translation)

Let F be a GENEOS from (Φ, G) to (Ψ, H) with respect to T . The operator $F_b(\varphi) := \varphi - b$ is a GENEOS from (Φ, G) to (Ψ, H) with respect to T , for every $b \in \mathbb{R}$ such that $F_b(\Phi) \subseteq \Psi$.

Proposition (Maximum)

If F_1, \dots, F_n are GENEOSs from (Φ, G) to (Ψ, H) with respect to T , then the operator $F(\varphi) := \max_i F_i(\varphi)$ is a GENEOS from (Φ, G) to (Ψ, H) with respect to T , provided that $F(\Phi) \subseteq \Psi$.



Building new GENEOS via translations, weighted averages and the maximum operator

Proposition (Weighted average)

If F_1, \dots, F_n are GENEOS from (Φ, G) to (Ψ, H) with respect to T and $(a_1, \dots, a_n) \in \mathbb{R}^n$ with $\sum_{i=1}^n |a_i| \leq 1$, then the operator $F(\varphi) := \sum_{i=1}^n a_i F_i(\varphi)$ is a GENEOS from (Φ, G) to (Ψ, H) with respect to T , provided that $F(\Phi) \subseteq \Psi$.

Our results show that if we work with spaces Φ, Ψ of measurements that are compact and convex, then the topological space of all GENEOSs from (Φ, G) to (Ψ, H) with respect to T is compact and convex.

An interesting GENE0 in kD persistent homology



Previous propositions imply the following statement.

Proposition

Assume F_1, \dots, F_n are GENE0s from (Φ, G) to (Ψ, H) with respect to T , and that $(a_1, \dots, a_n), (b_1, \dots, b_n) \in \mathbb{R}^n$, with $a_1, \dots, a_n > 0$, $\sum_{i=1}^n a_i = 1$ and $\sum_{i=1}^n b_i = 0$. Then the operator

$$F(\varphi) := \max \left\{ \frac{\min_j a_j}{a_1} \cdot (F_1(\varphi) - b_1), \dots, \frac{\min_j a_j}{a_n} \cdot (F_n(\varphi) - b_n) \right\}$$

is a GENE0 from (Φ, G) to (Ψ, H) with respect to T , provided that $F(\Phi) \subseteq \Psi$.

This result can be easily generalized from the case $\Phi \subseteq C^0(X, \mathbb{R})$ to the case $\Phi \subseteq C^0(X, \mathbb{R}^m)$.

An interesting GENEIO in kD persistent homology



Let us now take $G = H$, $T = id$ and $n = m$ in the extended version of the previous proposition. By considering the projection operators $F_i(\varphi) := \varphi_i$ for every $\varphi = (\varphi_1, \dots, \varphi_n) \in \Phi \subseteq C^0(X, \mathbb{R}^n)$, we obtain the operator

$$F(\varphi) = \max \left\{ \frac{\min_j a_j}{a_1} \cdot (\varphi_1 - b_1), \dots, \frac{\min_j a_j}{a_n} \cdot (\varphi_n - b_n) \right\}.$$

This operator is important in kD persistent homology, as a key tool to reduce kD persistent Betti number functions to families of 1D persistent Betti number functions. It is interesting to observe that such an operator is a group equivariant non-expansive operator.



Open questions

After defining an observer as a collection of GENEOS, our purpose consists in looking for methods to approximate the observer by a finite (and possible small) set of simple GENEOS.

This leads us to the following open questions:

- How can we build a good library of GENEOS?
- How can we find a method to choose a finite set \mathcal{F}^* of GENEOS that allows for both a good approximation of the natural pseudo-distance d_G and a fast computation?
- How can we provide a suitable statistical theory for group equivariant non-expansive operators?

Further research is needed.



Conclusions

- In our model, data comparison is based on measurements made by an observer. Each measurement can be represented as a function defined on a topological space X .
- The observer can be seen as a collection of GENEOS, applied to the measurements. The operators are allowed to change both the space of measurements and the invariance group.
- The functions describing the measurements can be compared by means of the natural pseudo-distance associated with a group G .
- Persistent homology can be used to approximate the natural pseudo-metric d_G . This can be done by means of a method that is based on GENEOS. This method is stable with respect to noise.
- **The topological space of GENEOS deserves further research.**

