On the role of group equivariant non-expansive operators in Topological Data Analysis

Patrizio Frosini

Department of Mathematics, ALMA-AI, AM^2 and ARCES, University of Bologna
patrizio.frosini@unibo.it

Online Workshop "Beyond TDA", NTU, August 28-30, 2021

Outline

Data analysis is not just about data

Topological and metric basics for the theory of GENEOs

Some relevant links between GENEOs and TDA

Compactness and convexity of the space of GENEOs

Methods to build GENEOs

GENEOs in the probabilistic setting

Data analysis is not just about data

Topological and metric basics for the theory of GENEOs

Some relevant links between GENEOs and TDA

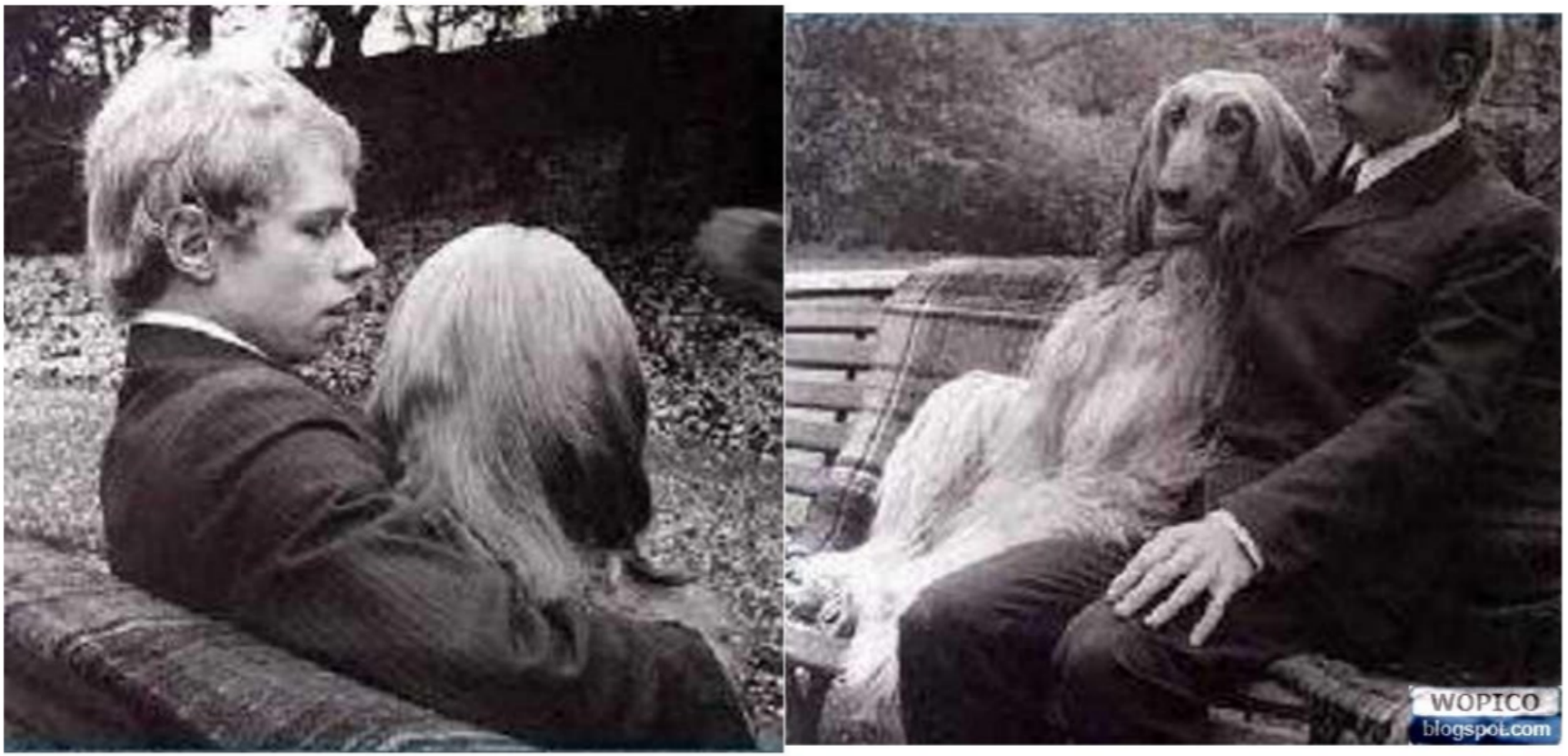
Compactness and convexity of the space of GENEOs

Methods to build GENEOs

GENEOs in the probabilistic setting

Data analysis is not just about data

Data interpretation depends on the observer:



Observers are often more important than data

We are usually not directly interested in data, but in data observers. For example, a patient is usually interested not in the data representing a computerized axial tomography of her body, but in the diagnosis that her doctor can make from these data.

Data analysis strongly depends on the chosen observer. If data analysis were not dependant on the chosen observer, then physicians' diagnoses would always be identical, scientists would always see the same causes for each phenomenon, and all people would agree in judging who the heroes and villains in a movie or a political event are.

It is indeed well known that different agents can have different reactions in the presence of the same data, and this suggests that **data analysis should study the pairs (data, observer) instead of data alone**.

What are data?

Data are usually produced by measurements (or actions) made by observers. 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.

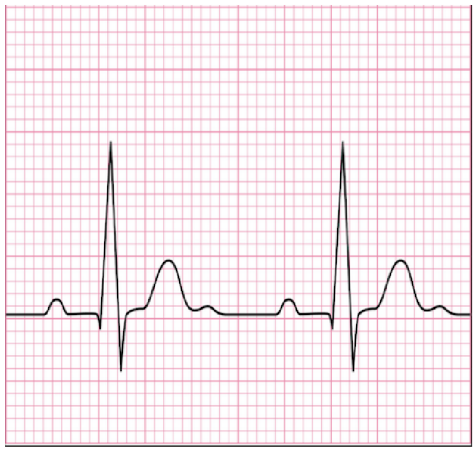
WIKIPEDIA

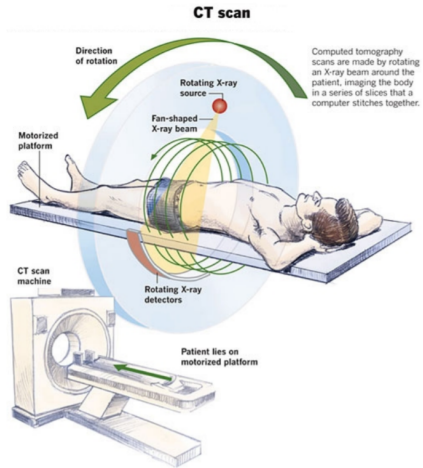
According to this definition, measurements (and hence data) can be often seen as functions φ associating a real number $\varphi(x)$ with each point x of a set X of characteristics. (This definition admits a natural extension to vector-valued functions but, for the sake of simplicity, we will treat here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

Data are measurements made by observers

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

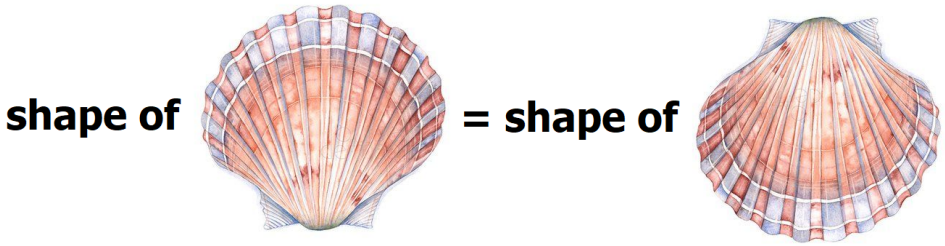
- An electrocardiogram (a function from $\mathbb{R}$ to $\mathbb{R}$);
- A gray-level image (a function from $\mathbb{R}^2$ to $\mathbb{R}$);
- A computerized tomography (CT) scan (a function from a helix to $\mathbb{R}$).



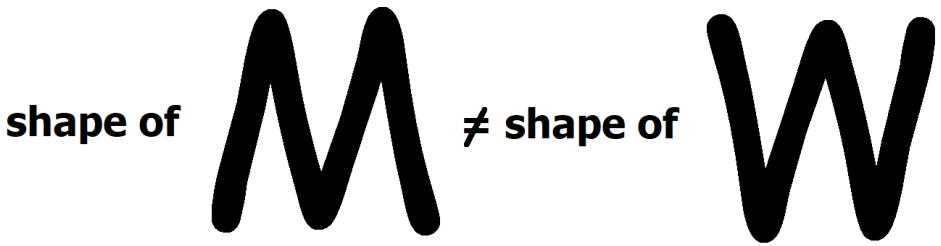


Observers are often associated with invariance groups

Observers often think that some data are equivalent to each other, according to an invariance group.



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



Data equivalence w.r.t. a group of permutations

Our data are represented by real-valued functions.

What do the expressions "data equivalence" and "data similarity" mean in our setting?

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

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

The natural pseudo-distance 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. Let us assume that φ_1, φ_2 are two continuous and bounded functions from X to $\mathbb{R}$, and 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 recall that a pseudo-distance is just a distance d without the assumption that $d(x_1, x_2) = 0$ implies $x_1 = x_2$.)

We could look at d_G as the ground truth in data comparison, when data equivalence is expressed by the group G.

The natural pseudo-distance 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 general assumptions about data and observers

Our mathematical model is based on these assumptions:

- The space of observers is often more important than the space of data;
- The study of the space of observers requires the development of a new topological-geometric model.
- This new model could be of great use in data analysis, when the role of the observers is not negligible.

These assumptions suggest us to move from **Topological Data Analysis** to the new field of **Topological Observer Analysis**.

Observers can be seen as equivariant operators

Observers are structures able to change data into other data, and usually do that by respecting some data equivalences, i.e., by commuting with some transformations.

As a first approximation, observers can be represented as group equivariant operators (GEOs).

In this talk we will give some results on the theory of **Group Equivariant Non-Expansive Operators** (**GENEOs**).
(Why "non-expansive?" Because observers are often assumed to simplify the metric structure of data in order to produce meaningful interpretations.)

We will also illustrate some interesting links between GENEOs and TDA.

Data analysis is not just about data

Topological and metric basics for the theory of GENEOs

Some relevant links between GENEOs and TDA

Compactness and convexity of the space of GENEOs

Methods to build GENEOs

GENEOs in the probabilistic setting

Measurements as admissible functions

Let X be a nonempty set. Let Φ be a topological subspace of the set $\mathbb{R}_b^X$ of all bounded functions φ from X to $\mathbb{R}$, endowed with the topology induced by the metric

$$D_\Phi(\varphi_1, \varphi_2) := \|\varphi_1 - \varphi_2\|_\infty . \tag{0.1}$$

We can see X as the space where we can make our measurements, and Φ as the space of all possible measurements. We will say that Φ is the *set of admissible functions*. In other words, Φ **is the set of all functions from X to $\mathbb{R}$ that can be produced by our measuring instruments**. For example, a gray-level image can be represented as a function from the real plane to the interval $[0,1]$ (in this case $X = \mathbb{R}^2$).

We recall that the initial topology τ_{in} on X with respect to Φ is the coarsest topology on X such that every function φ in Φ is continuous.

A pseudo-metric on X

Let us define on X the pseudo-metric

$$D_X(x_1, x_2) = \sup_{\varphi \in \Phi} |\varphi(x_1) - \varphi(x_2)|. \tag{0.2}$$

D_X induces a topology τ_{D_X} on X.

Theorem

The topology τ_{D_X} is finer than the initial topology τ_{in} on X with respect to Φ. If Φ is totally bounded, then τ_{D_X} coincides with τ_{in}.

The use of D_X implies that we can distinguish two points only if a measurement exists, taking those points to different values.

Theorem

If Φ is compact and X is complete, then X is compact.

Each bijection is an isometry

Let $\mathrm{Bij}(X)$ be the set of all bijections from X to X, and denote by $\mathrm{Bij}_\Phi(X)$ the set of all $g \in \mathrm{Bij}(X)$ such that $\varphi \circ g \in \Phi$ and $\varphi \circ g^{-1} \in \Phi$ for every $\varphi \in \Phi$. Let $\mathrm{Homeo}(X)$ be the set of all homeomorphisms from X a X with respect to D_X, and denote by $\mathrm{Homeo}_\Phi(X)$ the set of all $g \in \mathrm{Homeo}(X)$ such that $\varphi \circ g \in \Phi$ and $\varphi \circ g^{-1} \in \Phi$ for every $\varphi \in \Phi$. Let $\mathrm{Iso}(X)$ be the set of all isometries from X a X, and denote by $\mathrm{Iso}_\Phi(X)$ the set of all $g \in \mathrm{Iso}(X)$ such that $\varphi \circ g \in \Phi$ and $\varphi \circ g^{-1} \in \Phi$ for every $\varphi \in \Phi$.

Proposition

$\mathrm{Bij}_\Phi(X) = \mathrm{Homeo}_\Phi(X) = \mathrm{Iso}_\Phi(X)$.

A pseudo-metric on G

Let us now focus our attention on a subgroup G of $\mathrm{Homeo}_\Phi(X)$.
We can define a pseudo-metric D_G on G by setting

$$D_G(g_1, g_2) := \sup_{\varphi \in \Phi} D_\Phi(\varphi \circ g_1, \varphi \circ g_2). \tag{0.3}$$

Theorem

G is a topological group with respect to D_G and the action of G on Φ by right composition is continuous.

Theorem

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

GEOs and GENEOs

Each pair (Φ, G) with $G \subseteq \mathrm{Homeo}_\Phi(X)$ is called a *perception pair*.

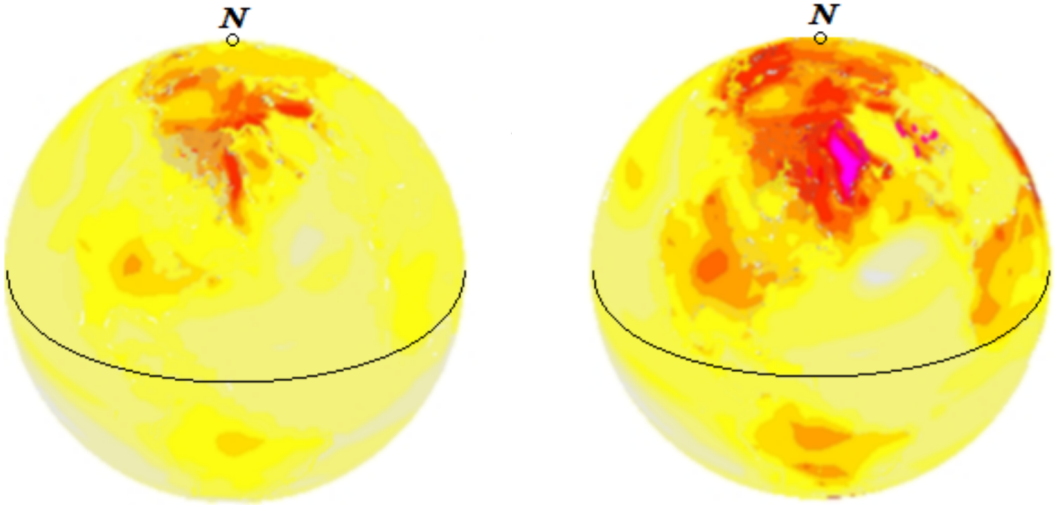
Let us assume that two perception pairs (Φ, G), (Ψ, H) are given, and fix a group homomorphism $T : G \to H$.

Each function $F : \Phi \to \Psi$ such that $F(\varphi \circ g) = F(\varphi) \circ T(g)$ for every $\varphi \in \Phi, g \in G$ is called a *Group Equivariant Operator (GEO)* associated with the homomorphism T.

If F is also non-expansive (i.e., $D_\Psi(F(\varphi_1), F(\varphi_2)) \leq D_\Phi(\varphi_1, \varphi_2)$ for every $\varphi_1, \varphi_2 \in \Phi$), then F is called a *Group Equivariant Non-Expansive Operator (GENEO)* associated with the homomorphism T.

An example of GENEO

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 assume that only two opposite points N, S can be localized on the sphere.

An example of GENEO

In this case we can set

- $X = S^2$
- Φ = set of 1-Lipschitz 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
- Ψ = set of 1-Lipschitz functions from S^1 to $[a, b]$
- H = group of rotations of S^1

An example of GENEO

This is a simple example of GENEO from (Φ, G) to (Ψ, H):

- $T(g)$ is the rotation $h \in H$ of the equator S^1 that is induced by the rotation g of S^2, for every $g \in G$.
- $F(\varphi)$ is 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 isomorphism $T : G \to H$.

Some relevant links between GENEOs and TDA

What are the main links
between GENEOs and TDA?

Some relevant links between GENEOs and TDA

- The use of GENEOs allows us to restrict the invariance of TDA;
- The operator taking each regular function to a suitable representation of its persistence diagram is a GENEO;
- GENEOs and persistent homology allow us to approximate the natural pseudo-distance d_G;
- GENEOs interact with multiparameter persistent homology;
- GENEOs can be efficiently compared by means of TDA.

In the next slides the symbol $\mathrm{GENEO}((\Phi, G), (\Psi, H))$ will denote the set of all GENEOs between the perception pairs (Φ, G), (Ψ, H) with respect to a fixed homomorphism $T : G \to H$.

1.The use of GENEOs restricts the invariance of TDA

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

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

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

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

We stress that, **differently from** $d_{\mathrm{match}}\left(\mathrm{Dgm}(\varphi_1), \mathrm{Dgm}(\varphi_2)\right)$, $\mathscr{D}^{\mathscr{F},k}_{\mathrm{match}}(\varphi_1, \varphi_2)$ **is not** invariant with respect to every $g \in \mathrm{Homeo}(X)$.

$\mathscr{D}^{\mathscr{F},k}_{\mathrm{match}}$ restricts to G the invariance of the classical bottleneck distance.

2.Persistence diagrams can be seen as a GENEO

Let us assume that:

- Φ is the set of all Morse functions from a closed regular manifold M to $[0,1]$.
- G is the group of all self-diffeomorphisms of M.
- Ψ is the set of all linear combinations $\sum_i f(\|x-p_i\|_\infty)$, where $\{p_i\}_i$ is a finite subset of the extended plane and f is a tent function.
- H is the trivial group containing only the identity of $\mathbb{R}^2$.
- $T : G \to H$ is the trivial homomorphism.

Then the operator taking each function $\varphi \in \Phi$ to the linear combination associated with its persistence diagram is a GENEO from (Φ, G) to (Ψ, H) with respect to T.

(Equivariance follows from the invariance of persistence diagrams under the action of self-homeomorphisms of M. Non-expansivity is a consequence of the stability of persistence diagrams.)

3.Computing d_G via GENEOs and persistent homology

If the perception pairs (Φ, G), (Ψ, H) coincide, the distance $\mathscr{D}^{\mathscr{F}}_{\text{match}}$ has the following two properties, showing that it can be also used to get information about the natural pseudo-distance d_G.

Theorem

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

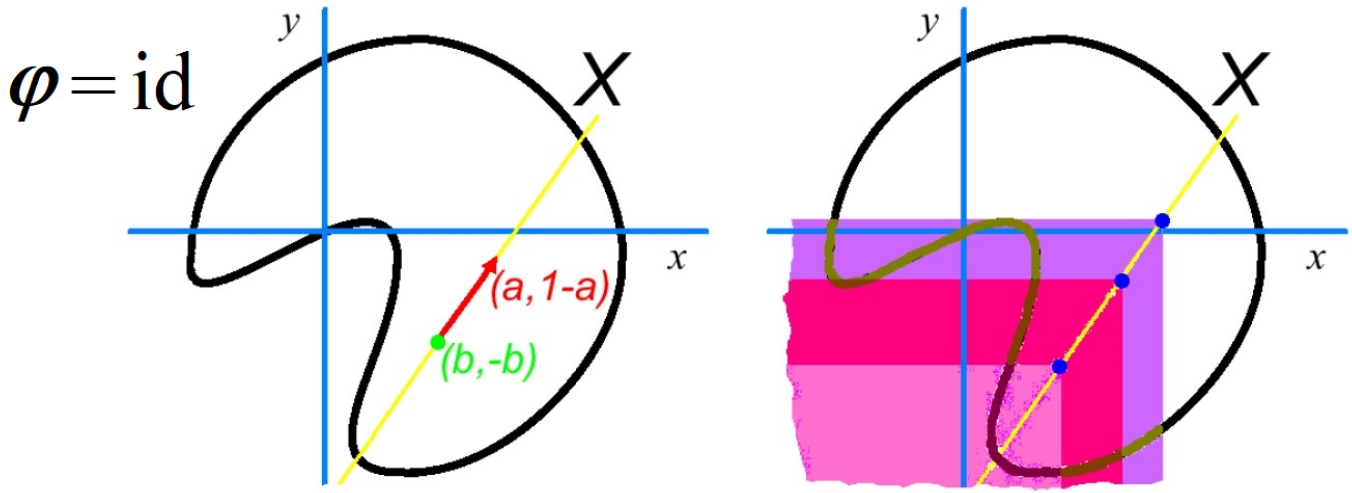
Theorem

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

GENEOs allow us to approximate the natural pseudo-distance d_G.

4.GENEOs and 2-parameter persistent homology (1/4)

If we have a bifiltration given by a function $\varphi = (\varphi_1, \varphi_2) : X \to \mathbb{R}^2$, we can consider a unit vector (w.r.t. $\|\cdot\|_1$) $w = (a, 1-a)$ with a positive slope, and a point $P = (b, -b)$. Every choice of P and w defines a filtration $\{X_t\}$ of X, where X_t is the set of points of X whose image by φ is both under and on the left of the point $P + tw$.
As a consequence, each choice of P and w defines a persistence diagram.



4.GENEOs and 2-parameter persistent homology (2/4)

If we set $(x, y) = P + tw = (at + b, (1-a)t - b)$ and define the function $\varphi_{(a,b)}(p) := \max\left\{\frac{\varphi_1(p)-b}{a}, \frac{\varphi_2(p)+b}{1-a}\right\}$, we can write $X_t = \{p \in X : \varphi_1(p) \leq x, \varphi_2(p) \leq y\}$ as the set $\{p \in X : \varphi_{(a,b)}(p) \leq t\}$.

As a consequence, the filtration $\{X_t\}$ of X leads us to consider the persistence diagram $\mathrm{Dgm}(\varphi_{(a,b)})$ of the function $\varphi_{(a,b)}$.
In order to get a stability theorem we have to normalize $\varphi_{(a,b)}$ by setting

$$\varphi^*_{(a,b)}(p) := \min\{a, 1-a\} \cdot \varphi_{(a,b)}(p).$$

4.GENEOs and 2-parameter persistent homology (3/4)

We can define a 2D matching distance $D_{\text{match}}(\varphi, \psi)$ by setting $D_{\text{match}}(\varphi, \psi) := \sup_{(a,b)\in]0,1[\times\mathbb{R}} d_{\text{match}}\left(\text{Dgm}(\varphi^*_{(a,b)}), \text{Dgm}(\psi^*_{(a,b)})\right)$.

The following theorem is well known:

Theorem (Stability Theorem)

$D_{\text{match}}(\varphi, \psi) \leq \|\varphi - \psi\|_\infty$.

This result can be easily extended to n-parameter persistent homology.

S. Biasotti, A. Cerri, P. Frosini, D. Giorgi, C. Landi, Multidimensional size functions for shape comparison, *Journal of Mathematical Imaging and Vision*, vol. 32 (2008), n. 2, 161-179.

A. Cerri, B. Di Fabio, M. Ferri, P. Frosini, C. Landi, Betti numbers in multidimensional persistent homology are stable functions, *Mathematical Methods in the Applied Sciences*, vol. 36 (2013), 1543-1557.

4.GENEOs and 2-parameter persistent homology (4/4)

In summary, the definition of the matching distance between two bifiltrations $\varphi, \psi : X \to \mathbb{R}^n$ of a topological space X can be seen as the supremum of the classical bottleneck distance between the persistence diagrams associated with the filtrations $F_{a,b}(\varphi), F_{a,b}(\psi) : X \to \mathbb{R}$, where the operator $F_{a,b}$ is defined by setting

$$F_{a,b}(\varphi) = \varphi^*_{(a,b)} = \max\left\{ \frac{\min\{a, 1-a\}}{a} \cdot (\varphi_1 - b), \frac{\min\{a, 1-a\}}{1-a} \cdot (\varphi_2 + b) \right\}.$$

Key fact: the operator $F_{a,b}$ is a GENEO for any value of a and b.

Therefore, the concept of rank invariant for a bifiltration can be introduced by the GENEOs $F_{a,b}$.

5.TDA gives a distance to compare GENEOs

We have just given four different examples showing how GENEOs can be of help for the development of TDA. We can now give an example showing how TDA can be of help in the development of a theory of GENEOs.

Persistent homology can indeed be used to define a computable and stable pseudo-metric Δ_{GENEO} between GENEOs by setting

$$\Delta_{\mathrm{GENEO}}(F_1, F_2) := \sup_{\varphi \in \Phi} d_{\mathrm{match}}\left(\mathrm{Dgm}(F_1(\varphi)), \mathrm{Dgm}(F_2(\varphi))\right)$$

for every $F_1, F_2 \in \mathrm{GENEO}((\Phi, G), (\Psi, H))$.

TDA makes available an efficient pseudo-metric to compare GENEOs.

Remark. Persistent homology also gives a shortcut to compare elements of each equivariance group G, by the pseudo-distance

$$\Delta_G(g_1, g_2) := \sup_{\varphi \in \Phi} d_{\mathrm{match}}\left(\mathrm{Dgm}(\varphi \circ g_1), \mathrm{Dgm}(\varphi \circ g_2)\right).$$

Data analysis is not just about data

Topological and metric basics for the theory of GENEOs

Some relevant links between GENEOs and TDA

Compactness and convexity of the space of GENEOs

Methods to build GENEOs

GENEOs in the probabilistic setting

Two key results

Let us assume that a homomorphism $T : G \to H$ has been fixed.
Let us define a metric D_{GENEO} on $\mathrm{GENEO}((\Phi, G), (\Psi, H))$ by setting

$$D_{\mathrm{GENEO}}(F_1, F_2) := \sup_{\varphi \in \Phi} D_{\Psi}(F_1(\varphi), F_2(\varphi)).$$

Theorem

If Φ and Ψ are compact, then $\mathrm{GENEO}((\Phi, G), (\Psi, H))$ is compact with respect to D_{GENEO}.

Theorem

If Ψ is convex, then $\mathrm{GENEO}((\Phi, G), (\Psi, H))$ is convex.

A consequence of the compactness theorem

Proposition

Let $\mathscr{F}$ *be a nonempty subset of* $\mathrm{GENEO}((\Phi, G), (\Psi, H))$. *For every* $\varepsilon > 0$, *a finite subset* $\mathscr{F}^*$ *of* $\mathscr{F}$ *exists, such that*

$$|\mathscr{D}^{\mathscr{F}^*,k}_{\mathrm{match}}(\varphi_1, \varphi_2) - \mathscr{D}^{\mathscr{F},k}_{\mathrm{match}}(\varphi_1, \varphi_2)| \leq \varepsilon$$

per ogni $\varphi_1, \varphi_2 \in \Phi$.

Elementary methods to build GENEOs

Proposition (Composition)

If $F_1 \in \mathrm{GENEO}((\Phi, G), (\Psi, H))$ *w.r.t.* $T_1 : G \to H$ *and*
$F_2 \in \mathrm{GENEO}((\Psi, H), (\chi, K))$ *w.r.t.* $T_2 : H \to K$ *then*
$F_2 \circ F_1 \in \mathrm{GENEO}((\Phi, G), (\chi, K))$ *w.r.t.* $T_2 \circ T_1 : G \to K$.

Proposition (Image by a 1-Lipschitz function)

If $F_1, \ldots, F_n \in \mathrm{GENEO}((\Phi, G), (\Psi, H))$ *w.r.t.* $T : G \to H$, L *is a 1-Lipschitz map from* $\mathbb{R}^n$ *to* $\mathbb{R}$, *and* $L^*(F_1, \ldots, F_n)(\Phi) \subseteq \Phi$ *(where* L^* *is the map induced by* L*), then*
$L^*(F_1, \ldots, F_n) \in \mathrm{GENEO}((\Phi, G), (\Psi, H))$ *w.r.t.* T.

The next three statements follow from the last proposition.

Elementary methods to build GENEOs

Proposition (Maximization)

If $F_1, \ldots, F_n \in \mathrm{GENEO}((\Phi, G), (\Psi, H))$ *w.r.t.* $T : G \to H$ *e* $\max(F_1, \ldots, F_n)(\Phi) \subseteq \Phi$, *allora* $\max(F_1, \ldots, F_n) \in \mathrm{GENEO}((\Phi, G), (\Psi, H))$ *w.r.t.* T.

Proposition (Translation)

Se $F \in \mathrm{GENEO}((\Phi, G), (\Psi, H))$ *w.r.t.* $T : G \to H$, *and* $F_b(\Phi) \subseteq \Phi$ *for* $F_b(\varphi) := F(\varphi) - b$, *then* $F_b \in \mathrm{GENEO}((\Phi, G), (\Psi, H))$ *w.r.t.* T.

Proposition (Convex combination)

If $F_1, \ldots, F_n \in \mathrm{GENEO}((\Phi, G), (\Psi, H))$ *w.r.t.* $T : G \to H$, $(a_1, \ldots, a_n) \in \mathbb{R}^n$ *con* $\sum_{i=1}^n |a_i| \leq 1$ *and* $F_\Sigma(\Phi) \subseteq \Phi$ *for* $F_\Sigma(\varphi) := \sum_{i=1}^n a_i F_i(\varphi)$, *then* $F_\Sigma \in \mathrm{GENEO}((\Phi, G), (\Psi, H))$ *w.r.t.* T.

Permutant measures

Let us consider the set $\Phi = \mathbb{R}^X \cong \mathbb{R}^n$ of all functions from a finite set $X = \{x_1, \ldots, x_n\}$ to $\mathbb{R}$, and a subgroup G of the group $\mathrm{Bij}(X)$ of all permutations of X.

Definition

A finite (signed) measure μ on $\mathrm{Bij}(X)$ is called a *permutant measure* with respect to G if every subset H of $\mathrm{Bij}(X)$ is measurable and μ is invariant under the conjugacy action of G (i.e., $\mu(H) = \mu(gHg^{-1})$ for every $g \in G$).

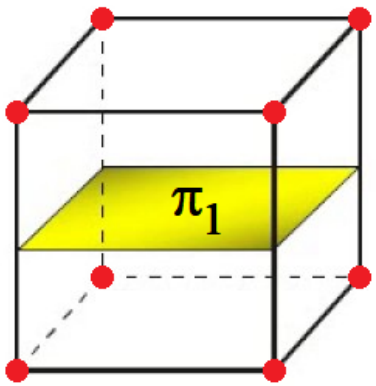
Proposition

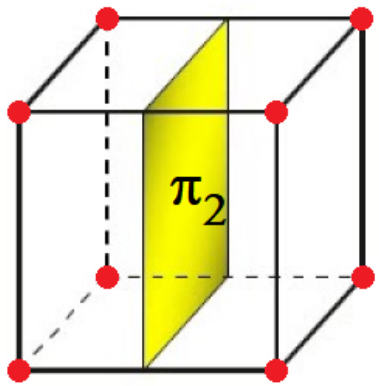
If μ is a permutant measure with respect to G, then the map $F_\mu : \mathbb{R}^X \to \mathbb{R}^X$ defined by setting $F_\mu(\varphi) := \sum_{h \in \mathrm{Bij}(X)} \varphi h^{-1}\ \mu(h)$ is a linear GEO.

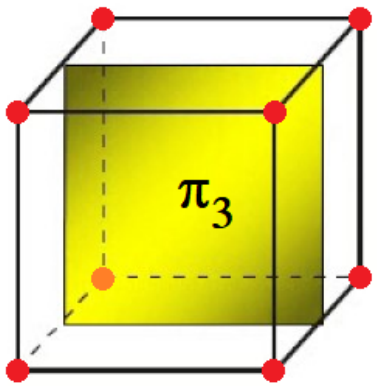
An example of permutant measure

Let us consider the set X of the vertices of a cube in $\mathbb{R}^3$, and the group G of the orientation-preserving isometries of $\mathbb{R}^3$ that take X to X. Let π_1, π_2, π_3 be the three planes that contain the center of mass of X and are parallel to a face of the cube. Let $h_i : X \to X$ be the orthogonal symmetry with respect to π_i, for $i \in \{1, 2, 3\}$.

We can now define a permutant measure μ on the group $\mathrm{Bij}(X)$ by setting $\mu(h_1) = \mu(h_2) = \mu(h_3) = c$, where c is a positive real number, and $\mu(h) = 0$ for any $h \in \mathrm{Bij}(X)$ with $h \notin \{h_1, h_2, h_3\}$.







Building GENEOs by permutant measures

The following representation theorem holds.

Theorem

Let us assume that $G \subseteq \mathrm{Bij}(X)$ *transitively acts on the finite set* X *and that* F *is a map from* $\mathbb{R}^X$ *to* $\mathbb{R}^X$*. The map* F *is a linear GENEO from* $(\mathbb{R}^X, G)$ *to* $(\mathbb{R}^X, G)$ *(with respect to the identical homomorphism* $\mathrm{id}_G : g \mapsto g$*) if and only if a permutant measure* μ *with respect to* G *exists, such that* $F(\varphi) = \sum_{h \in \mathrm{Bij}(X)} \varphi h^{-1} \ \mu(h)$ *for every* $\varphi \in \mathbb{R}^X$*, and* $\sum_{h \in \mathrm{Bij}(X)} |\mu(h)| \leq 1$*.*

Further details can be found in this preprint:

S. Botteghi, M. Brasini, P. Frosini and N. Quercioli, On the finite representation of group equivariant operators via permutant measures **https://arxiv.org/pdf/2008.06340.pdf**

Data analysis is not just about data

Topological and metric basics for the theory of GENEOs

Some relevant links between GENEOs and TDA

Compactness and convexity of the space of GENEOs

Methods to build GENEOs

GENEOs in the probabilistic setting

GENEOs in the probabilistic setting

When the space of data Φ is endowed with a probability measure, we can replace the pseudo-metrics D_X, D_G with the following two pseudo-metrics Δ_X, Δ_G, after choosing a G-invariant probability density f and a G-invariant probability measure λ on Φ:

$$\Delta_X(x_1, x_2) = \int_\Phi |\varphi(x_1) - \varphi(x_2)| f(\varphi)\, d\lambda, \quad \forall\, x_1, x_2 \in X;$$

$$\Delta_G(g_1, g_2) = \int_\Phi \|\varphi g_1 - \varphi g_2\| f(\varphi)\, d\lambda, \quad \forall\, g_1, g_2 \in G.$$

We will assume that Φ is compact.

GENEOs in the probabilistic setting

The following statements hold:

Proposition

Every function $\varphi \in \Phi$ is continuous with respect to Δ_X.

Proposition

X is totally bounded with respect to Δ_X.

Hence

Corollary

If X is complete, then X is compact.

GENEOs in the probabilistic setting

Proposition

G is a topological group and the action of G on Φ by right composition is continuous.

Proposition

G is totally bounded with respect to Δ_G.

Hence

Corollary

If G is complete, then it is also compact.

GENEOs in the probabilistic setting

The concepts of GEO and GENEO can be adapted to the probabilistic setting and the following results can be proved:

Theorem

If the spaces of data are compact, then the space of all GENEOs is compact with respect to the norm $\|F\|_{L^2} := \left(\int_{\Phi} \|F(\varphi)\|^2 f(\varphi)\ d\lambda\right)^{\frac{1}{2}}$.

Proposition

If the spaces of data are convex, then also the set of all GENEOs is convex.

A Riemannian structure for manifolds of GENEOs

Important remark. $L^2(\Phi, V)$ is endowed with the inner product $\langle F_1, F_2 \rangle := \int_\Phi \langle F_1(\varphi), F_2(\varphi) \rangle f(\varphi)\, d\lambda$. Therefore, any C^k-submanifold of GENEOs in $L^2(\Phi, V)$ naturally inherits a Riemannian structure from $L^2(\Phi, V)$ (I am skipping some technical details here).

As a consequence, we can use the gradient flow of cost functions to look for optimal GENEOs in manifolds of GENEOs.

Further details can be found in this preprint:

P. Cascarano, P. Frosini, N. Quercioli and A. Saki, *On the geometric and Riemannian structure of the spaces of group equivariant non-expansive operators*, **https://arxiv.org/pdf/2103.02543.pdf**

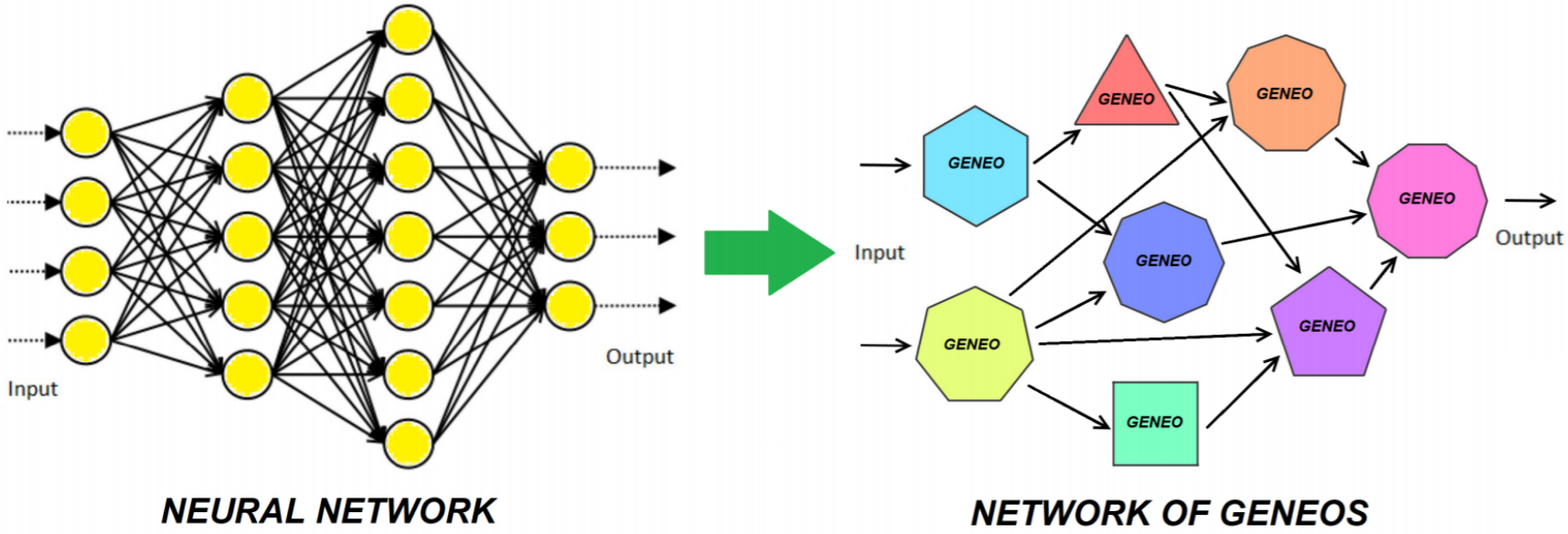
Conclusions (1/2)

There are many links between TDA and the theory of GENEOs:

- The theory of GENEOs could be a starting point for moving from TDA to Topological Observer Analysis.
- The computation of (suitable representations of) persistence diagrams can be seen as a GENEO.
- The use of GENEOs allows us to restrict the invariance of TDA;
- GENEOs and persistent homology allow us to approximate the natural pseudo-distance d_G;
- GENEOs interact with multiparameter persistent homology;
- GENEOs can be efficiently compared by means of TDA.

Conclusions (2/2)

In perspective, we would like to obtain a good compositional theory for building efficient and transparent networks of GENEOs.
Some preliminary experiments suggest that replacing neurons with GENEOs could make deep learning more transparent and interpretable and speed up the learning process.



Open questions

- How can we approximate a real observer (let us say, e.g., a physician) by GENEOs, in order to emulate her behaviour with respect to data?
- Can we devise constructive procedures, allowing us to build any possible GENEO with respect to a given equivariance group?
- What is the right way of comparing GENEOs in a topological-statistical setting?
- How should we select representative sets in a probability space of GENEOs?
- How can we compute the basic statistics for GENEOs?
- How can we predict the behaviour of networks of GENEOs and control their actions?
- How can we evaluate advantages and limits of an approach to data analysis based on the interaction of GENEOs and TDA?

Collaborators

Many results illustrated in this talk have been obtained in joint research conducted together with several collaborators and students: Faraz Ahmad, Mattia Bergomi, Silvia Biasotti, Stefano Botteghi, Martina Brasini, Francesco Camporesi, Pasquale Cascarano, Andrea Cerri, Barbara Di Fabio, Pietro Donatini, Bianca Falcidieno, Massimo Ferri, Leila De Floriani, Daniela Giorgi, Grzegorz Jabłoński, Claudia Landi, Laura Papaleo, Nicola Quercioli, Amir Saki, Michela Spagnuolo.

SOME FURTHER 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