Information Extraction from Data via Group-Equivariant Non-Expansive Operators

Patrizio Frosini

Department of Computer Science, University of Pisa
patrizio.frosini@unipi.it

Mathematics for Big Data and Artificial Intelligence
23rd ECMI Conference on Industrial and Applied Mathematics,
29 June – 3 July 2026, Kaunas, Lithuania

Outline

What is a GENEO?

Some basics on the theory of GENEOs

GENEOs and Topological Data Analysis

GENEOs and Explainable Artificial Intelligence (XAI)

What is a GENEO?

- A Group Equivariant Non-Expansive Operator (GENEO) is a mathematical tool used to approximate observers that act on data.
- The theory of GENEOs is based on the idea that the geometric characteristics of observers significantly influence the interpretation of data.
- In this talk, we will explore the core properties of GENEOs, examine their role in machine learning, and discuss their promising applications in explainable artificial intelligence.

What is a GENEO?

Some basics on the theory of GENEOs

GENEOs and Topological Data Analysis

GENEOs and Explainable Artificial Intelligence (XAI)

Perception spaces

Recall that a pseudometric is a distance function d satisfying nonnegativity, symmetry, and the triangle inequality, but not necessarily the property $d(x_1, x_2) = 0 \implies x_1 = x_2$.

Definition

Let us consider:

1. A nonempty set Φ endowed with a pseudometric D_Φ.
2. A group $(G, \circ)$ acting on Φ on the left, denoted by $*$. We assume the action is by isometries, i.e., for every $\varphi_1, \varphi_2 \in \Phi$ and every $g \in G$, $D_\Phi(g * \varphi_1, g * \varphi_2) = D_\Phi(\varphi_1, \varphi_2)$.

We call (Φ, G) a perception space.

If a perception space (Φ, G) is given, then the group G can be endowed with the pseudometric D_G, defined by $D_G(g_1, g_2) := \sup_{\varphi \in \Phi} D_\Phi(g_1 * \varphi, g_2 * \varphi)$ for all $g_1, g_2 \in G$.

Perception spaces

The set Φ represents the data we may obtain from our measuring tools (functions, graphs, point clouds, ...). The group G represents the possible transformations of the data that the observer may be interested in.

Some examples:

- $\Phi =$ the set of grey-level images viewed as functions from $\mathbb{R}^2$ to $[0, 1]$, endowed with the sup-norm distance;
  $G =$ the group of isometries of the real plane.
- $\Phi =$ the set of all subgraphs of a given graph Γ, endowed with the distance $d_{iso}(\Gamma_1, \Gamma_2) = 0$ if $\Gamma_1 \equiv \Gamma_2$ and 1 otherwise;
  $G =$ the group of graph isomorphisms of Γ.
- $\Phi =$ the set of nonempty compact subsets of the real plane, endowed with the Hausdorff distance;
  $G =$ the group of isometries of the real plane.

GEOs and GENEOs

Definition

- Let (Φ, G), (Ψ, K) be two perception spaces. If a map $F : \Phi \to \Psi$ and a group homomorphism $T : G \to K$ are given, such that $F(g * \varphi) = T(g) * F(\varphi)$ for every $\varphi \in \Phi$, $g \in G$, we say that (F, T) is an (extended) *group equivariant operator* (GEO).
- If F is non-expansive (i.e., $D_\Psi(F(\varphi_1), F(\varphi_2)) \leq D_\Phi(\varphi_1, \varphi_2)$ for every $\varphi_1, \varphi_2 \in \Phi$), we say that (F, T) is an (extended) *group equivariant non-expansive operator* (GENEO).

Remark

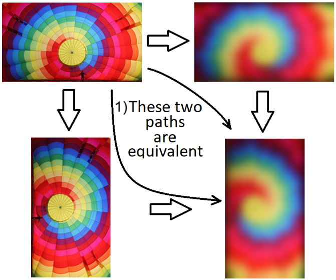
Let $(F, T) : (\Phi, G) \to (\Psi, K)$ be a GEO and assume that F is **surjective**. If $D_\Psi(F(\varphi_1), F(\varphi_2)) \leq D_\Phi(\varphi_1, \varphi_2)$ for every $\varphi_1, \varphi_2 \in \Phi$, then $D_K(T(g_1), T(g_2)) \leq D_G(g_1, g_2)$ for every $g_1, g_2 \in G$.

GENEO Example: Image Blurring

$\mathscr{I}$ = the collection of all color images, viewed as $C_c\left(\mathbb{R}^2, [0,1]^3\right)$, endowed with the max-norm distance;

- $F : \mathscr{I} \to \mathscr{I}$, and $F(\varphi)$ is the convolution of φ with a rotationally symmetric kernel k satisfying $\|k\|_1 = 1$;
- $T = \mathrm{id} : \mathrm{Isom}(\mathbb{R}^2) \to \mathrm{Isom}(\mathbb{R}^2)$.

Then (F, T) is a GENEO.



Here, the max-norm distance between functions is used.

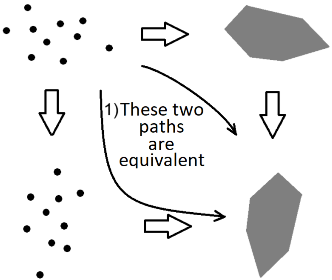
2) Blurring does not increase the distance between the images.

GENEO Example: Computing Convex Hulls

$\mathscr{S}$ = the collection of all finite nonempty subsets of the real plane;
$\mathscr{P}$ = the collection of all polyhedra in the real plane.

- $F : \mathscr{S} \to \mathscr{P}$, $F(S) = \textit{convex hull of } S$;
- $T = \mathrm{id} : \mathrm{Isom}(\mathbb{R}^2) \to \mathrm{Isom}(\mathbb{R}^2)$.

Then (F, T) is a GENEO.



Here, the Hausdorff distance between compact sets is used.

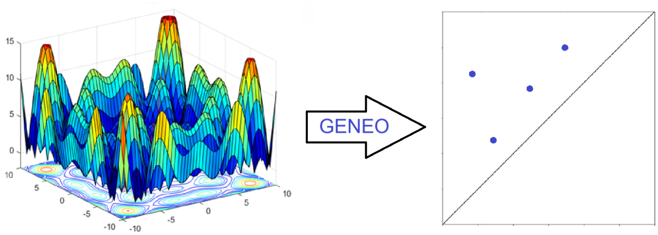
2) The operation of taking the convex hull does not increase the Hausdorff distance between sets.

GENEO Example: Computing Persistence Diagrams

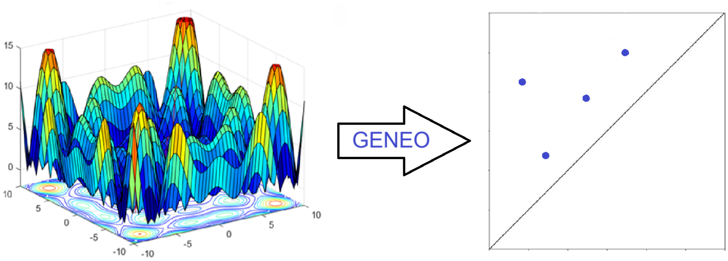
DGM = the collection of all persistence diagrams of real-valued continuous functions defined on a topological space X, where DGM is endowed with the bottleneck distance;

- $F : C_c(X, \mathbb{R}) \to \mathrm{DGM}$, $F(\varphi) = \mathrm{Dgm}_k(\varphi)$;
- $T : \mathrm{Homeo}(X) \to \{\mathrm{id}_{\mathrm{DGM}}\}$ is the trivial homomorphism.

Then (F, T) is a GENEO.



GENEO Example: Computing Persistence Diagrams



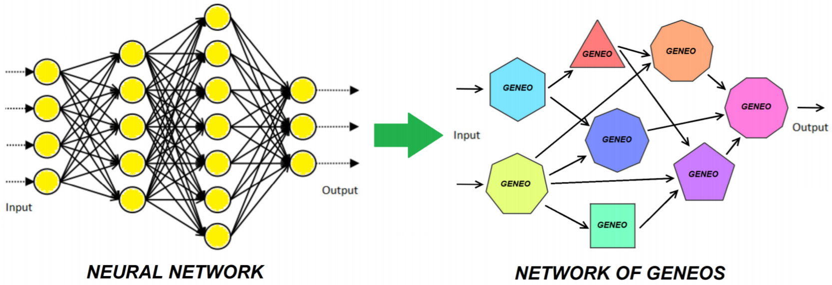
- Equivariance of (F, T) = invariance of persistence diagrams under reparameterization of the domain.
- Non-expansiveness of (F, T) = stability of persistence diagrams.

Why are GENEOs interesting?

- GENEOs rest on a **rigorous topological/geometric framework** (in what follows we outline several results).
- GENEOs **encode prior knowledge** about the chosen observer.
- The non-expansiveness property of GENEOs imposes a strong constraint that enables **meaningful data simplification**.
- GENEOs enable a **compositional approach to deep learning**.
- Analyzing the **geometry of the observer space** (as represented by GENEOs) is often more informative than analyzing the geometry of the data space.

The main point in the approach based on GENEOs

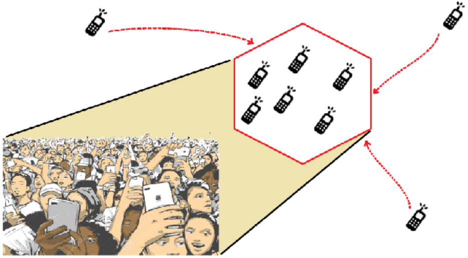
In perspective, we are looking for a good compositional theory for building efficient and transparent networks of GENEOs.
Some preliminary experiments suggest that replacing neurons with GENEOs could make some applications in deep learning more transparent and interpretable and speed up the learning process.



Some research projects concerning GENEOs (I)

CNIT / WiLab - Huawei Joint Innovation Center (JIC)

Project on GENEOs for 6G



WILAB HUAWEI

Some research projects concerning GENEOs (II)

PANDORA

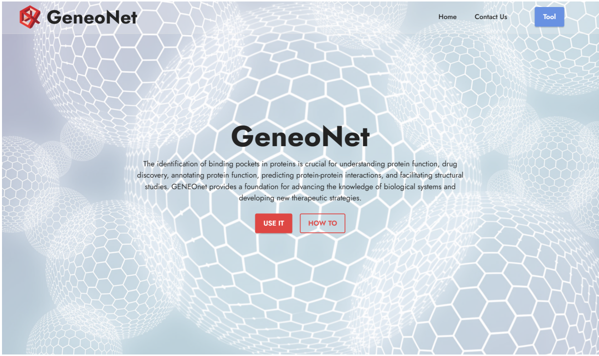
Horizon Europe (HORIZON)
Call: HORIZON-CL4-2023-HUMAN-01-CNECT
Project: 101135775-PANDORA
Funding: approximately 9 million euros.

Task 3.3 - Leveraging domain knowledge for explainable learning:
This task aims to investigate the use of domain knowledge in the development of explainable AI models. Tools like GENEOs for applications in TDA and ML and new theoretical methods of GENEOs for explainable AI are used.

The project has received funding form the European Union's Horizon Europe Framework Programme (Horizon) under grant agreement No 101135775

https://pandora-heu.eu/consortium/

Some research projects concerning GENEOs (III)



The GENEOnet webservice represents the outcome of a partnership with Italian Pharmaceutical Company Dompé Farmaceutici S.p.A.: https://geneonet.exscalate.eu/

Compactness of the space of GENEOs

From now on, let $\mathscr{F}_T^{\mathrm{all}}$ denote the set of all GENEOs (F, T) from a perception space (Φ, G) to a perception space (Ψ, K), **with fixed homomorphism** T. We equip $\mathscr{F}_T^{\mathrm{all}}$ with the following distance:

$$D_{\mathrm{GENEO}}((F, T), (F', T)) := \sup_{\varphi \in \Phi} D_\Psi\big(F(\varphi), F'(\varphi)\big),$$

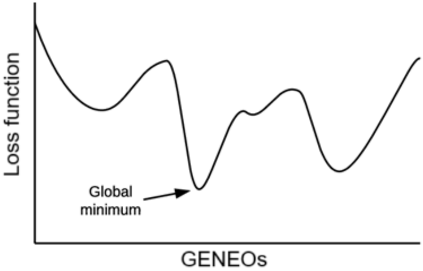
where D_Ψ is the metric chosen on Ψ.

Theorem

If (Φ, D_Φ), (Ψ, D_Ψ) *are* **compact**, *then the space* $(\mathscr{F}_T^{\mathrm{all}}, D_{\mathrm{GENEO}})$ *is* **compact**.

Compactness of the space of GENEOs

The compactness of the space of GENEOs is an important property from an applied standpoint. It implies that the space of GENEOs can be approximated arbitrarily well by a finite set of GENEOs, and that every continuous loss function defined on this space attains an absolute minimum (i.e., there exists an optimal GENEO with respect to the loss function).



Convexity of the space of GENEOs

Now assume that

- Φ, Ψ **are normed real vector spaces**, with distances D_Φ, D_Ψ induced by the norms $\|\cdot\|_\Phi$, $\|\cdot\|_\Psi$ (i.e., $D_\Phi(\varphi, \varphi') = \|\varphi - \varphi'\|_\Phi$ and $D_\Psi(\psi, \psi') = \|\psi - \psi'\|_\Phi$);
- **the group actions are linear**, i.e.,
  $g * (a\varphi + b\varphi') = a(g * \varphi) + b(g * \varphi')$,
  $k * (a\psi + b\psi') = a(k * \psi) + b(k * \psi')$,
  for all $a, b \in \mathbb{R}$, $g \in G$, $k \in K$, $\varphi, \varphi' \in \Phi$, $\psi, \psi' \in \Psi$.

Let $(F_1, T), (F_2, T), \ldots, (F_n, T)$ be GENEOs from (Φ, G) to (Ψ, K). If $(a_1, \ldots, a_n) \in \mathbb{R}^n$ and $\sum_{i=1}^n |a_i| \leq 1$, define

$$F_\Sigma(\varphi) := \sum_{i=1}^n a_i F_i(\varphi).$$

Convexity of the space of GENEOs

Proposition

(F_Σ, T) *is a GENEO from* (Φ, G) *to* (Ψ, K).

Corollary

The space $\mathscr{F}_T^{\mathrm{all}}$ *is* **convex**.

The functional case of GENEOs

A particularly important case of GENEO is the one in which the perception spaces involved consist of sets of data expressed as functions with values in $\mathbb{R}$ or $\mathbb{R}^n$.

> In the remainder of this talk we shall restrict our attention to the case of data expressed as real-valued functions.

In every perception space (Φ, G) under consideration, Φ is a set of $\mathbb{R}$-valued functions defined on a domain X, endowed with the metric $D_\Phi(\varphi, \varphi') = \|\varphi - \varphi'\|_\infty$.

The group G will be a group of permutations of X such that, if $\varphi \in \Phi$ and $g \in G$, then $\varphi \circ g^{-1} \in \Phi$ as well, and the left action of G on Φ will be given by $g * \varphi = \varphi \circ g^{-1}$. We observe that this action is **isometric**.

Methods to build GENEOs

In order to use our model profitably, we need constructive methods to produce GENEOs in the presence of pre-established data and equivariance groups.

Without going into technical details, we simply observe that, under reasonable assumptions,

- the composition of GENEOs is still a GENEO;
- the maximum and the minimum of GENEOs are still GENEOs;
- the translation of a GENEO is still a GENEO;
- the convex combination of GENEOs is still a GENEO.

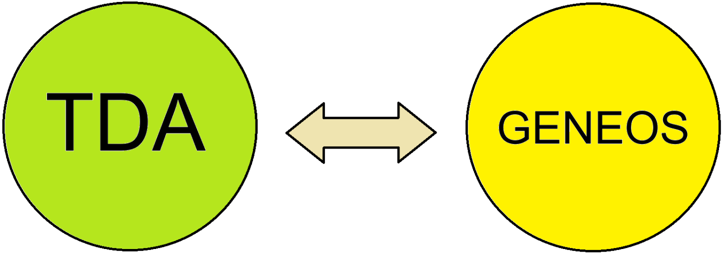
Furthermore, a representation theorem provides a construction of all linear GENEOs from (Φ, G) to itself under the assumptions that X is finite and the action of the group G is transitive.

GENEOs and TDA

We have already highlighted that the computation of persistence diagrams can be formalized as a GENEO.

However, there are also other interesting connections between the concept of GENEO and Topological Data Analysis.

In what follows, we will illustrate some of them.



GENEOs restrict the invariance of TDA

The use of GENEOs allows us **to restrict** the invariance of TDA to subgroups of the group $\mathrm{Homeo}_\Phi(X)$ of all Φ-preserving homeomorphisms of X. To show this, let us assume that a set $\mathscr{F}$ of GENEOs from a perception space (Φ, G) to a perception space (Ψ, K) is given, with respect to a fixed homomorphism $T\colon G \to K$. Let X denote the domain of the functions in Φ. Then, for every degree k, we can define a new pseudometric $\mathscr{D}_{\mathrm{match}}^{\mathscr{F},\Phi}$ on Φ by

$$\mathscr{D}_{\mathrm{match}}^{\mathscr{F},\Phi}(\varphi_1, \varphi_2) = \sup_{F \in \mathscr{F}} d_{\mathrm{match}}\big(\mathrm{Dgm}_k(F(\varphi_1)), \mathrm{Dgm}_k(F(\varphi_2))\big).$$

If $\varphi \in \Phi$, then $\mathscr{D}_{\mathrm{match}}^{\mathscr{F},\Phi}(\varphi \circ g, \varphi) = 0$ for every $g \in G$, but, in general, not for every $g \in \mathrm{Homeo}_\Phi(X)$.

Since GENEOs are non-expansive, the pseudometric $\mathscr{D}_{\mathrm{match}}^{\mathscr{F},\Phi}$ is stable:

$$\mathscr{D}_{\mathrm{match}}^{\mathscr{F},\Phi}(\varphi_1, \varphi_2) \leq \|\varphi_1 - \varphi_2\|_\infty.$$

GENEOs and biparameter persistent homology

Assume that $\boldsymbol{\varphi} = (\varphi_1, \varphi_2)$ and $\boldsymbol{\psi} = (\psi_1, \psi_2)$ belong to $C^0(X, \mathbb{R}^2)$.
For each $(a, b) \in]0, 1[\times \mathbb{R}$, let us define the map
$F_{a,b} \colon C^0(X, \mathbb{R}^2) \to C^0(X, \mathbb{R})$ by setting

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

Interestingly, $F_{a,b}$ is a GENEO.

For each degree k, we can consider the distance

$$\mathbb{D}_{\mathrm{match},k}(\boldsymbol{\varphi}, \boldsymbol{\psi}) := \sup_{(a,b) \in]0,1[\times \mathbb{R}} d_{\mathrm{match}}(\mathrm{Dgm}_k(F_{a,b}(\boldsymbol{\varphi})), \mathrm{Dgm}_k(F_{a,b}(\boldsymbol{\psi}))).$$

It is called the two-dimensional **matching distance**.

GENEOs and biparameter persistent homology

It is worth noting that, through the use of GENEOs, one can obtain many different distances for biparametric persistence. For example, if we are interested in distances based on one-parameter families of GENEOs, we can define pseudometrics in biparametric TDA using

$$\bar{F}_t(\boldsymbol{\varphi}) := (1-t)\varphi_1 + t\varphi_2, \quad t \in [0,1],$$

or

$$\hat{F}_t(\boldsymbol{\varphi}) := (1-t)\max(\varphi_1, \varphi_2) + t\min(\varphi_1, \varphi_2), \quad t \in [0,1],$$

or

$$\tilde{F}_t(\boldsymbol{\varphi}) := \left(\frac{1}{2}|\varphi_1|^t + \frac{1}{2}|\varphi_2|^t\right)^{\frac{1}{t}}, \quad t \geq 1.$$

[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 (2019).]

[N. Quercioli, *Some new methods to build group equivariant non-expansive operators in TDA*. In: Devaney, R.L., Chan, K.C., Vinod Kumar, P. (eds) Topological Dynamics and Topological Data Analysis. IWCTA 2018. **Springer Proceedings in Mathematics & Statistics**, vol 350. Springer, Singapore (2021).]

[P. Frosini, U. Fugacci, E. Mosig García, N. Quercioli, S. Scaramuccia, F. Tombari, *The convex matching distance in multiparameter persistence*, arxiv.org/abs/2512.02944 (2025).]

[N. Berkouk, F. Petit, *Projected distances for multi-parameter persistence modules*, **Annales de l'Institut Fourier**, Online first (2026), 62 p.] (a different but related approach to our research)

GENEOs and XAI

We can also apply GENEOs to explainable AI.

GENEOs make it possible to define a distance between GEOs operating on different domains and producing outputs in different codomains. This allows us to assess how similarly two neural networks behave, even when their input and output spaces differ. **The resulting metric is based on quantifying the degree of non-commutativity of suitable diagrams involving GENEOs.**

Two different operators behaving similarly

GENEOs and XAI

Informal idea: We say that the action of an agent A is **explained** by another agent B from the perspective of an agent C if:

1. C perceives A and B as being similar (this step uses GENEOs);
2. C perceives B as less complex than A.

[J. J. Colombini, F. Bonchi, F. Giannini, F. Giannotti, R. Pellungrini and P. Frosini, *Mathematical foundation of interpretable equivariant surrogate models*, 3rd World Conference on Explainable Artificial Intelligence (XAI-2025), Novel Post-hoc & Ante-hoc XAI Approaches, 09-11 July, 2025 - Istanbul, Turkey, **Communications in Computer and Information Science**, vol 2577 (2026), 294-318. Springer, Cham.]


Thanks for your attention!