Towards a topological-geometrical theory of group equivariant non-expansive operators for data analysis and machine learning

Mattia G. Bergomi[a], Patrizio Frosini[b,c,*], Daniela Giorgi[d], Nicola Quercioli[b,c]

[a] *Champalimaud Research, Champalimaud Center for the Unknown - Lisbon, Portugal*
[b] *Department of Mathematics, University of Bologna*
[c] *Advanced Research Center on Electronic System "Ercole De Castro", University of Bologna*
[d] *Italian National Research Council, Institute of Information Science and Technologies "Alessandro Faedo"*

Abstract

The aim of this paper is to provide a general mathematical framework for group equivariance in the machine learning context. The framework builds on a synergy between persistent homology and the theory of group actions. We define group-equivariant non-expansive operators (GENEOs), which are maps between function spaces associated with groups of transformations. We study the topological and metric properties of the space of GENEOs to evaluate their approximating power and set the basis for general strategies to initialise and compose operators. We begin by defining suitable pseudo-metrics for the function spaces, the equivariance groups, and the set of non-expansive operators. Basing on these pseudo-metrics, we prove that the space of GENEOs is compact and convex, under the assumption that the function spaces are compact and convex. These results provide fundamental guarantees in a machine learning perspective. We show examples on the MNIST and fashion-MNIST datasets. By considering isometry-equivariant non-expansive operators, we describe a simple strategy to select and sample operators, and show how the selected and sampled operators can be used to perform both classical metric learning and an effective initialisation of the kernels of a convolutional neural network.

Keywords: Group equivariant non-expansive operator, invariance group, group action, initial topology, persistent homology, persistence diagram, bottleneck distance, natural pseudo-distance, agent, perception pair, slice category, topological data analysis
2000 MSC: Primary 55N35 Secondary 47H09, 54H15, 57S10, 68U05, 65D18

1. Introduction

Deep learning-based algorithms reached human or superhuman performance in many real-world tasks. Beyond the extreme effectiveness of deep learning, one of the main reasons for its

*Corresponding author

Email addresses: mattia.bergomi@neuro.fchampalimaud.org (Mattia G. Bergomi), patrizio.frosini@unibo.it (Patrizio Frosini), daniela.giorgi@isti.cnr.it (Daniela Giorgi), nicola.quercioli2@unibo.it (Nicola Quercioli)

success is that raw data are sufficient—if not even more suitable than hand-crafted features—for these algorithms to learn a specific task. However, only few attempts have been made to create formal theories allowing for the creation of a controllable and interpretable framework, in which deep neural networks can be formally defined and studied. Furthermore, if learning directly from raw data allows to outclass human feature engineering, the architectures of deep networks are growing more and more complex, and often are as task-specific as hand-crafted features used to be.

We aim at providing a general mathematical framework, where any agent capable of acting on a certain dataset (e.g. deep neural networks) can be formally described as a collection of transformations acting on the data. To motivate our model, we assume that data cannot be studied directly, but only through the action of operators that make them measurable. Consequently, our model stems from a functional viewpoint. By interpreting data as points of a function space, it is possible to learn and optimise operators defined on the data. In other words, we are interested in the space of transformations of the data, rather than the data themselves. Albeit unformalised, this idea is not new in deep learning. For instance, the main characteristic of convolutional neural networks [1] is the election of convolution as the operator of choice to act on the data. The convolutional kernels learned by optimising a loss function are operators that map an image to a new one that, for instance, is more easily classifiable. Moreover, convolutions are operators equivariant with respect to translations (at least in the ideal continuous case). We believe that the restriction to a specific family of operators and the equivariance with respect to interpretable transformations are key aspects of the success of this architecture. In our theory, operators are thought of as instruments allowing an agent to provide a measure of the world, as the kernels learned by a convolutional neural network allow a classifier to spot essential features to recognise objects belonging to the same category. Equivariance with respect to the action of a group, or a set of transformations corresponds to the introduction of symmetries. This allows us to both gain control on the nature of the learned operators, as well as drastically reduce the dimensionality of the space of operators to be explored during learning. Such a goal is in line with the recent interest for invariant representations in machine learning (cf., e.g., [2]).

We make use of topological data analysis to describe spaces of group-equivariant non-expansive operators (GENEOs). GENEOs are maps between function spaces associated with groups of transformations. We study the topological and metric properties of the space of GENEOs to evaluate their approximating power and set the basis for general strategies to initialise, compose operators and eventually connect them hierarchically to form operator networks. Our first contribution is to define suitable pseudo-metrics for the function spaces, the equivariance groups, and the set of non-expansive operators. Basing on these pseudo-metrics, we prove that the space of GENEOs is compact and convex, under the assumption that the function spaces are compact and convex. These results provide fundamental and provable guarantees for the goodness of this operator-based approach in a machine learning perspective. The property of compactness for instance shows how any operator belonging to a certain space can be approximated by a finite number of operators sampled in the same space. Our study of the space of GENEOs takes advantage of recent results in topological data analysis, in particular in the theory of persistent homology. This allows us to provide a general framework to previous works on group equivariance in deep learning context [3, 4]. Moreover, this approach generalises standard group-equivariance to set-equivariance, that seems much more suitable for the representation of intelligent agents.

To conclude, we validate our model with examples on the MNIST and fashion-MNIST datasets. These applications are aimed at proving the effectiveness on discrete examples, of the metrics defined and the theorems proved in the continuous case. By considering isometry-equivariant non-expansive operators (IENEO), we describe two simple algorithms allowing to select and sample IENEOs based on few labelled samples taken from the dataset. We show how selected and sampled operators can be used to perform both classical metric learning and effective initialisation of the kernels of a convolutional neural network.

We believe that the formal foundation of our model is suitable to start a new theory of *deep-learning engineering*, and that novel research lines will stem from the synergy of machine learning and topology.

The paper is structured as follows. In Section 2 the epistemological foundations of our model are discussed. The mathematical background in topological persistence is discussed in Section 3. Section 4 is the core of this work and details the mathematical models defining and studying the spaces of GENEOs, and their topology. New results in persistent homology and the extension of the theory to set-equivariance are presented in Section 5. The necessary hypotheses and theorems to prove the compactness and convexity of the space of GENEOs are described in Section 6. Finally, in Section 7, we describe the algorithms used to select and sample operators in the discrete case and apply them on both the MNIST and fashion-MNIST datasets.

2. Epistemological setting

The mathematical model described and studied in this paper is justified by an epistemological background that is briefly presented in this section. We build our model on the following assumptions:

1. Data cannot be studied in a direct and absolute way. They are only knowable through acts of measurement made by an agent. From the point of view of data analysis only the pair (data, agent) matters.
2. Any act of measurement can be represented as a function defined on a topological space, since only stable measurements can be considered for applications and stability requires a topological structure.
3. Any agent is described by the way they transform data and preserves some kind of invariance. In other words, any agent can be seen as a collection of group equivariant operators acting on function spaces.
4. Only the agent is entitled to decide about data similarity.

In other words, in our framework we assume that the analysis of data is replaced by the analysis of the pair *(data, agent)* we are considering. Since an agent can be seen as a set of group equivariant operators, from the mathematical viewpoint our purpose consists in presenting a good topological theory of suitable operators of this kind, representing agents.

In our agent-centered setting, two objects can be distinguished from each other if and only if they can be distinguished by a suitable measurement. In case every available measurement is not able to distinguish them, they must be considered as equal.

For more details about this epistemological setting we refer the interested reader to [5].

3. Mathematical background

Our mathematical model builds on functional analysis and Topological Data Analysis (TDA) [Carlsson]. TDA is an emerging field of research which studies topological approaches to explore and make sense of complex, high-dimensional data, such as artificial and biological networks [Carlsson, Scientific Reports 2012, Bulletin of the AMS 2009]. The basic idea is that topology can help to recognize patters within data, and therefore to turn data into useful knowledge. One of the main concepts in TDA is Persistent Homology (PH), a mathematical tool that captures topological information at multiple scales. Our mathematical model proposes an integration between the theory of group actions and Persistent Homology. In what follows, we briefly summarize the main concepts in PH [6, 7, 8].

3.1. Persistent Homology

In PH, data are modeled as objects in a metric space. The first step is to filter the data so to obtain a family of nested topological spaces that captures the topological information at multiple scales. A common way to obtain a filtration is by sublevel sets of a continuous function, hence the name *sublevelset persistence*. Let φ be a real-valued continuous function on a topological space X. Persistent homology represents the changes of the homology groups of the sub-level set $X_t = \varphi^{-1}((-\infty, t])$ varying t in $\mathbb{R}$. We can see the parameter t as an increasing time, whose changes produce the birth and the death of k-dimensional holes in the sub-level set X_t. For example the number of 0-dimensional holes equals the number of the connected components of X, 1-dimensional holes refer to tunnels and 2-dimensional holes to voids.

Definition 3.1. If $u, v \in \mathbb{R}$ and $u < v$, we can consider the inclusion i of X_u into X_v. If $\check{H}$ denotes the Čech homology functor, such an inclusion induces a homomorphism $i_k : \check{H}_k(X_u) \to \check{H}_k(X_v)$ between the homology groups of X_u and X_v in degree k. The group $PH_k^{\varphi}(u, v) := i_k(\check{H}_k(X_u))$ is called the kth *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 $PH_k^{\varphi}(u, v)$ is said the kth *persistent Betti number function* (PBN) with respect to the function $\varphi : X \to \mathbb{R}$, computed at the point (u, v).

Persistent Betti number functions can be completely described by multisets called *persistence diagrams*. The kth persistence diagram is the multiset of all the pairs $p_j = (b_j, d_j)$, where b_j and d_j are the times of birth and death of the jth k-dimensional hole, respectively. When a hole never dies, we set its time to death equal to ∞. The multiplicity $m(p_j)$ says how many holes share both the time of birth b_j and the time of death d_j. For technical reasons, the points (t, t) are added to each persistence diagram, each one with infinite multiplicity.

Each persistence diagram D can contain an infinite number of points. For every $q \in \Delta^* := \{(x, y) \in \mathbb{R}^2 \ : \ x < y\} \cup \{(x, \infty) : x \in \mathbb{R}\}$, the equality $m(q) = 0$ means that q does not belong to the persistence diagram D. We define on $\bar{\Delta}^* := \{(x, y) \in \mathbb{R}^2 \ : \ x \leq y\} \cup \{(x, \infty) : x \in \mathbb{R}\}$ a pseudo-metric as follows

$$d^*((x, y), (x', y')) := \min \left\{ \max\{|x - x'|, |y - y'|\}, \max \left\{ \frac{y - x}{2}, \frac{y' - x'}{2} \right\} \right\} \tag{1}$$

by agreeing that $\infty - y = \infty$, $y - \infty = -\infty$ for $y \neq \infty$, $\infty - \infty = 0$, $\frac{\infty}{2} = \infty$, $|\pm\infty| = \infty$, $\min\{\infty, c\} = c$, $\max\{\infty, c\} = \infty$.

The pseudo-metric d^* between two points p and p' takes the smaller value between the cost of moving p to p' and the cost of moving p' and p onto $\Delta := \{(x, y) \in \mathbb{R}^2 \ : \ x = y\}$. Obviously, $d^*(p, p') = 0$ for every $p, p' \in \Delta$. If $p \in \Delta^+ := \{(x, y) \in \mathbb{R}^2 \ : \ x < y\}$ and $p' \in \Delta$, then $d^*(p, p')$ equals the distance, endowed by the max-norm, between p and Δ. Points at infinity have a finite distance only to the other points at infinity, and their distance equals the Euclidean distance between abscissas.

We can compare persistence diagrams by means of the *bottleneck distance* (also called *matching distance*) δ_{match}.

Definition 3.2. Let D, D' be two persistence diagrams. We define the *bottleneck distance* δ_{match} between D and D' by setting

$$\delta_{\mathrm{match}}(D, D') := \inf_{\sigma} \sup_{p \in D} d^*(p, \sigma(p)), \tag{2}$$

where σ varies in the set of all bijections from the multiset D to the multiset D'.

For further informations about persistence diagrams and the bottleneck distance, we refer the reader to [8, 9]. Each persistent Betti number function is associated with exactly one persistence diagram, and (if we use Čech homology) every persistence diagram is associated with exactly one persistent Betti number function. Then the metric δ_{match} induces a pseudo-metric d_{match} on the sets of the persistent Betti number functions [10].

4. Mathematical model

In our mathematical model, data are represented as function spaces, that is, as sets of real-valued functions on some topological space (Subsection 4.1). Function spaces come with invariance groups representing the transformations on data which are admissible for some agent (Subsection 4.2). The groups of transformations are specific to different agents, and can be either learned or part of prior knowledge. The operators on data are then defined as group-equivariant non-expansive operators (GENEOs) (Subsection 4.3).

4.1. Data representation

Let us consider a set $X \neq \emptyset$ and a topological subspace Φ of the set of all bounded functions φ from X to $\mathbb{R}$, denoted by $\mathbb{R}^X_b$ and endowed with the topology induced by the distance

$$D_\Phi(\varphi_1, \varphi_2) := \|\varphi_1 - \varphi_2\|_\infty . \tag{3}$$

If Φ is compact, then it is also bounded, i.e., there exists a non-negative real value L, such that $\|\varphi\|_\infty \leq L$ for every $\varphi \in \Phi$. We can think of X as the space where one makes measurements, and of Φ as the *set of admissible measurements* (also called *set of admissible functions*). In other words, Φ is the set of functions from X to $\mathbb{R}$ that can be produced by measuring instruments. For example, an image can be represented as a function φ from the real plane X to the real numbers.

To quantify the distance between two points $x_1, x_2 \in X$, we compare the values taken at x_1 and x_2 by the functions in the space of possible measurements Φ. Therefore, we endow X with the extended pseudo-metric[1] D_X defined by setting

$$D_X(x_1, x_2) = \sup_{\varphi \in \Phi} |\varphi(x_1) - \varphi(x_2)| \tag{4}$$

for every $x_1, x_2 \in X$ (see Appendix A).

The assumption behind the definition of D_X is that two points can be distinguished only if they assume different values while being measured. As an example, if Φ contains only constant functions, no discrimination can be made between points in X and hence $D_X(x_1, x_2)$ vanishes for every $x_1, x_2 \in X$.

The pseudo-metric space (X, D_X) can be considered as a topological space by choosing as a base $\mathcal{B}_{D_X}$ the collection of all the sets

$$B_X(x, \varepsilon) = \{x' \in X : D_X(x, x') < \varepsilon\} \tag{5}$$

where $\varepsilon > 0$ and $x \in X$ (see, [11]).

The reason to endow the measurement space X with a topology, rather than considering just a set, follows from the need of formalizing the assumption that measurements are stable. To formalize stability we have to use a topology (or a pseudo-metric inducing a topology).

It is interesting to stress the link between the topology τ_{D_X} associated with D_X and the initial topology[2] τ_{in} on X with respect to Φ, when we take the Euclidean topology $\mathcal{T}_E$ on $\mathbb{R}$.

Theorem 4.1. *The topology τ_{D_X} on X induced by the pseudo-metric D_X is finer than the initial topology τ_{in} on X with respect to Φ. If Φ is totally bounded, then the topology τ_{D_X} coincides with τ_{in}.*

(The proof is in Appendix C.)

In general X is not compact with respect to the topology τ_{D_X}, even if Φ is compact. For example, if X is the open interval $]0, 1[$ and Φ contains only the identity from $]0, 1[$ to $]0, 1[$, the topology induced by D_X is simply the Euclidean topology and hence X is not compact. However, the next result holds.

---

[1]We recall that a pseudo-metric is just a distance d without the property that $d(a, b) = 0$ implies $a = b$. An extended pseudo-metric is a pseudo-metric that may take the value ∞. If Φ is bounded, then D_X is a pseudo-metric.

[2]We recall that τ_{in} is the coarsest topology on X such that each function $\varphi \in \Phi$ is continuous. Explicitly, the open sets in τ_{in} are the sets that can be obtained as unions of finite intersections of sets $\varphi^{-1}(U)$, where $\varphi \in \Phi$ and $U \in \mathcal{T}_E$. In other words, a base $\mathcal{B}_{\text{in}}$ of τ_{in} is given by the collection of all sets that can be represented as $\bigcap_{i \in I} \varphi_i^{-1}(U_i)$, where I is a finite set of indexes and $\varphi_i \in \Phi$, $U_i \in \mathcal{T}_E$ for every $i \in I$ [11].

Theorem 4.2. *If Φ is compact and X is complete then X is also compact.*

(The proof is in Appendix C.)

Since τ_{in} is the coarsest topology on X such that $\varphi \in \Phi$ is continuous, Theorem 4.1 guarantees that the assumption that the functions are continuous is not restrictive in practice, for example while dealing with images, which often contain discontinuities. Indeed, our functions are not required to be continuous with respect to other topologies (e.g., the Euclidean topology τ_E on $X = \mathbb{R}^2$).

4.1.1. A remark on the use of pseudo-metrics

The reader could think better to change the pseudo-metric D_X into a metric D' by quotienting out X by the equivalence relation $x_1 R x_2 \iff D_X(x_1, x_2) = 0$ and defining $D'([x_1], [x_2]) = D_X(x_1, x_2)$ for any $[x_1], [x_2] \in X/R$. The reason we do not do this is that several different sets of admissible measurements can be considered on the same set X. For two different sets Φ_1, Φ_2 of admissible functions, we obtain two different quotient spaces X/R_1, X/R_2. If we forget about the original space X, we lose the possibility of linking the equivalence classes in X/R_1 with the ones in X/R_2. On the contrary, we prefer to preserve the identity of points in X, studying how they link to each other when we change the set Φ. This observation leads us to work with pseudo-metrics instead of metrics.

Before proceeding, we observe that the map π taking each point $x \in X$ to the equivalence class $[x] \in X/R$ is continuous with respect to D_X and D', and surjective. Moreover, π takes each ball with respect to D_X to a ball with respect to D', while the inverse image under π of each ball with respect to D' is a ball with respect to D_X. It follows that if a subset $S \subseteq X$ is compact (sequentially compact) for D_X then $\pi(S)$ is compact (sequentially compact) for D', and that if a subset $\mathcal{S} \subseteq X/R$ is compact (sequentially compact) for D' then $\pi^{-1}(\mathcal{S})$ is compact (sequentially compact) for D_X. Finally, given a sequence (x_i) in X, we observe that (x_i) converges to $\bar{x}$ in X if and only if the sequence $([x_i])$ converges to $[\bar{x}]$ in X/R. These facts imply that the development of our theory in terms of pseudo-metrics is not far from the analysis in terms of metrics.

4.2. Transformations on data

In our model, we assume that data are transformed through maps from X to X which are Φ-preserving homeomorphisms with respect to the pseudo-metric D_X. Let $\text{Homeo}(X)$ denote the set of homeomorphisms from X to X with respect to D_X, and $\text{Homeo}_\Phi(X)$ denote the set of Φ-preserving homeomorphisms, namely the homeomorphisms $g \in \text{Homeo}(X)$ such that $\varphi \circ g \in \Phi$ and $\varphi \circ g^{-1} \in \Phi$ for every $\varphi \in \Phi$.

The following Proposition 4.3 implies that $\text{Homeo}_\Phi(X)$ is exactly the set of all bijections $g : X \to X$ such that $\varphi \circ g \in \Phi$ and $\varphi \circ g^{-1} \in \Phi$ for every $\varphi \in \Phi$.

Proposition 4.3. *If g is a bijection from X to X such that $\varphi \circ g \in \Phi$ and $\varphi \circ g^{-1} \in \Phi$ for every $\varphi \in \Phi$, then g is an isometry*[3] *(and hence a homeomorphism) with respect to D_X.*

[3]The definition of isometry between pseudo-metric spaces can be considered as a special case of isometry

(The proof is in Appendix C.)

Remark 4.4. In general, $\text{Homeo}(X) \neq \text{Homeo}_\Phi(X)$. As an example, take $X = [0,1]$ and $\Phi = \{id\}$. In this case $D_X(x_1, x_2) = |x_1 - x_2|$ and $\text{Homeo}_\Phi = \{id\}$, while Homeo is the set of all homeomorphisms from the interval $[0,1]$ to itself with respect to the Euclidean distance.

Remark 4.5. For each $g \in \text{Homeo}_\Phi(X)$, we consider the bijective map $R_g : \Phi \longrightarrow \Phi$ defined by setting $R_g(\varphi) = \varphi \circ g$ for every $\varphi \in \Phi$. We claim that R_g preserves the pseudo-distance D_Φ in Equation 3. Indeed, if $\varphi, \varphi' \in \Phi$ and $g \in G$ then

$$\begin{aligned} D_\Phi(\varphi \circ g, \varphi' \circ g) &= \sup_{x \in X} |(\varphi \circ g)(x) - (\varphi' \circ g)(x)| \\ &= \sup_{x \in X} |\varphi(g(x)) - \varphi'(g(x))| \\ &= \sup_{y \in X} |\varphi(y) - \varphi'(y)| = D_\Phi(\varphi, \varphi'), \end{aligned} \tag{6}$$

because g is a bijection. Since R_g is a bijection preserving D_Φ, then R_g is an isometry with respect to D_Φ.

In the rest of this paper we will assume that Φ is compact with respect to the topology induced by D_Φ, and that X is complete (and hence compact) with respect to the topology induced by D_X.

Let us now consider a subgroup G of the group $\text{Homeo}_\Phi(X)$. G represents the set of transformations on data for which we require equivariance to be respected.

We can define the pseudo-distance D_G on G:

$$D_G(g_1, g_2) := \sup_{\varphi \in \Phi} D_\Phi(\varphi \circ g_1, \varphi \circ g_2) \tag{7}$$

from $G \times G$ to $\mathbb{R}$ (see Appendix A).

Remark 4.6. D_G can be expressed as:

$$D_G(g_1, g_2) = \sup_{x \in X} D_X(g_1(x), g_2(x)) = \sup_{x \in X} \sup_{\varphi \in \Phi} |\varphi(g_1(x)) - \varphi(g_2(x))|. \tag{8}$$

We can now state the following theorems:

Theorem 4.7. *G is a topological group with respect to the pseudo-metric topology and the action of G on Φ through right composition is continuous.*

(The proof is in Appendix C.)

between metric spaces. Let (X_1, d_1) and (X_2, d_2) be two pseudo-metric spaces. It is easy to check that if $f : X_1 \longrightarrow X_2$ is a function verifying the equality $d_1(x, y) = d_2(f(x), f(y))$ for every $x, y \in X_1$, then f is continuous with respect to the topologies induced by d_1 and d_2. If f verifies the previous equality and is bijective, we say that it is an *isometry* between the considered pseudo-metric spaces. If f is an isometry, we can trivially observe that f^{-1} is also an isometry, and that f is a homeomorphism.

Theorem 4.8. *If G is complete then it is also compact with respect to D_G.*

(The proof is in Appendix C.)

From now on we will suppose that G is complete (and hence compact) with respect to the topology induced by D_G.

4.2.1. The natural pseudo-distance d_G

We can consider the natural pseudo-distance d_G on the space Φ [12]:

Definition 4.9. The pseudo-distance $d_G : \Phi \times \Phi \to \mathbb{R}$ is defined by setting

$$d_G(\varphi_1, \varphi_2) = \inf_{g \in G} D_\Phi(\varphi_1, \varphi_2 \circ g). \tag{9}$$

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

The natural pseudo-distance d_G represents the ground truth in our model. It is based on comparing measurements, and vanishes for pairs of measurements that are equivalent with respect to the action of our group of homeomorphisms G, which expresses the equivalences between data.

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 subgroups of $\mathrm{Homeo}_\Phi(X)$ and $G_1 \subseteq G_2$, then the definition of d_G implies that

$$d_{\mathrm{Homeo}_\Phi(X)}(\varphi_1, \varphi_2) \le d_{G_2}(\varphi_1, \varphi_2) \le d_{G_1}(\varphi_1, \varphi_2) \le D_\Phi(\varphi_1, \varphi_2) \tag{10}$$

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

4.2.2. A remark on the use of homeomorphisms

The reader could criticize the choice of grounding our approach on the concept of homeomorphism. After all, most of the objects that are considered for purposes of shape comparison "are not homeomorphic". Therefore, the definition of natural pseudo-distance could seem not to be sufficiently flexible, since it does not allow to compare non-homeomorphic objects. Though, it is important to note that the space X we use in our model does *not* represent the objects, but the space where one takes measurements *about* the objects. As such, X is unique. For example, two images are considered as functions from the real plane X to the real numbers, independently of the topological properties of the 3D objects represented in the images. 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, the topological space X is determined only by the measuring instrument and not by the single object instances.

4.3. Group-Equivariant Non-Expansive Operators

Under the assumptions made in the previous sections, the pair (Φ, G) is called a *perception pair*.

Let us now assume that two perception pairs (Φ, G), (Ψ, H) are given together with a fixed 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 said to be a *perception map* from (Φ, G) to (Ψ, H) associated with the homomorphism T. More briefly, we will also say that F is a *group equivariant operator*. If T is equal to the identity homomorphism $I : G \longrightarrow G$, we can say that F is a G-map. We observe that the functions in Φ and the functions in Ψ are defined on spaces that are generally different from each other.

Each perception pair (Φ, G) can be seen as a category, whose objects are the functions in Φ and the morphisms between two functions $\varphi_1, \varphi_2 \in \Phi$ are the elements $g \in G$ such that $\varphi_2 = \varphi_1 \circ g$. As usual, if $\varphi_2 = \varphi_1 \circ g$ and $\varphi_2' = \varphi_1' \circ g$ we wish to distinguish g as a morphism between φ_1 and φ_2 from g as a morphism between φ_1' and φ_2', so we make different copies $g_{(\varphi_1,\varphi_2)}$, $g_{(\varphi_1',\varphi_2')}$ of the homeomorphism g by labelling it. As natural, $g'_{(\varphi_2,\varphi_3)} \circ g_{(\varphi_1,\varphi_2)} = (g \circ g')_{(\varphi_1,\varphi_3)}$. A precise formalization of this procedure can be done in terms of slice categories. For more details we refer the reader to Appendix B.

When two perception pairs (Φ, G), (Ψ, H) are considered as categories and a homomorphism $T : G \to H$ is fixed, each perception map F from (Φ, G) to (Ψ, H) is naturally associated with a functor between the two categories, taking each function $\varphi \in \Phi$ to $F(\varphi) \in \Psi$ and each morphism $g_{(\varphi_1,\varphi_2)} \in G$ to the morphism $T(g)_{(F(\varphi_1),F(\varphi_2))} \in H$.

Definition 4.10. Assume that (Φ, G),(Ψ, H) are two perception pairs and that a homomorphism $T : G \to H$ has been fixed. Each non-expansive perception map F from (Φ, G) to (Ψ, H) with respect to T is called a *Group Equivariant Non-Expansive Operator (GENEO) associated with* $T : G \to H$.

Obviously, the non-expansivity of F means that $D_\Psi (F(\varphi_1), F(\varphi_2)) \leq D_\Phi (\varphi_1, \varphi_2)$ for every $\varphi_1, \varphi_2 \in \Phi$.

Example 4.11. As a reference for the reader, we give the following basic example of GENEO. Let Φ be the set containing all 1-Lipschitz functions from $X = S^2 = \{(x, y, z) \in \mathbb{R}^3 : x^2+y^2+z^2 = 1\}$ to $[0, 1]$, and G be the group of all rotations of S^2 around the z-axis. Let Ψ be the set containing all 1-Lipschitz functions from $Y = S^1 = \{(x, y) \in \mathbb{R}^2 : x^2+y^2 = 1\}$ to $[0, 1]$, and H be the group of all rotations of S^1. We observe that (Φ, G) and (Ψ, H) are two perception pairs. Now, let us consider the map $F : \Phi \to \Psi$ taking each function $\varphi \in \Phi$ to the function $\psi \in \Psi$ defined by setting $\psi(\theta) := \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \varphi(\theta, \alpha)\, d\alpha$ (with θ, α polar coordinates), and the homomorphism T taking the rotation of S^2 of α radians around the z-axis positively oriented to the counter-clock rotation of α radians of S^1. We can easily check that F is a perception map and a GENEO from (Φ, G) to (Ψ, H), associated with the homomorphism T. In this example F and T are surjective, but an example where F and T are not surjective can be easily found, e.g. by restricting Φ to the singleton $\bar{\Phi}$ containing only the null function and G to the trivial group $\bar{G}$ containing only the identical homomorphism.

We can study how GENEOs act on the natural pseudo-distances:

Proposition 4.12. *If F is a GENEO from (Φ, G) to (Ψ, H) associated with $T : G \to H$, then it is a contraction with respect to the natural pseudo-distances d_G, d_H.*

(The proof is in Appendix C.)

4.3.1. Pseudo-metrics on GENEO $((\Phi, G), (\Psi, H))$

Let us denote by $\mathrm{GENEO}\left((\Phi, G), (\Psi, H)\right)$ the set of all GENEOs between two perception pairs (Φ, G), (Ψ, H) associated with $T : G \to H$. We can endow this set with the following pseudo-distances D_{GENEO}, $D_{\mathrm{GENEO,H}}$.

Definition 4.13. If $F_1, F_2 \in \mathrm{GENEO}\left((\Phi, G), (\Psi, H)\right)$, we set

$$\begin{aligned} D_{\mathrm{GENEO}}\left(F_1, F_2\right) &:= \sup_{\varphi\in\Phi} D_{\Psi}\left(F_1(\varphi), F_2(\varphi)\right) \\ D_{\mathrm{GENEO,H}}\left(F_1, F_2\right) &:= \sup_{\varphi\in\Phi} d_H\left(F_1(\varphi), F_2(\varphi)\right). \end{aligned} \tag{11}$$

The next result can be easily proved by applying the inequality $d_H \le D_\Psi$ (see Theorem 5.1) and recalling that the supremum of a family of bounded pseudo-metrics is still a pseudo-metric.

Proposition 4.14. *D_{GENEO} and $D_{\mathrm{GENEO,H}}$ are pseudo-metrics on* $\mathrm{GENEO}\left((\Phi, G), (\Psi, H)\right)$*. Moreover, $D_{\mathrm{GENEO,H}} \le D_{\mathrm{GENEO}}$.*

It would be easy to check that $D_{\mathrm{GENEO,H}}$ is a metric.

For the sake of conciseness, in the following we will set $\mathcal{F}^{\mathrm{all}} := \mathrm{GENEO}\left((\Phi, G), (\Psi, H)\right)$.

This simple statement holds:

Proposition 4.15. *For every $F \in \mathcal{F}^{\mathrm{all}}$ and every $\varphi \in \Phi$: $\|F(\varphi)\|_\infty \le \|\varphi\|_\infty + \|F(\boldsymbol{0})\|_\infty$, where $\boldsymbol{0}$ denotes the function taking the value 0 everywhere.*

(The proof is in Appendix C.)

4.4. GENEOs as agents in our model

In our model the agents are represented by GENEOs. Indeed, each agent can be seen as a black box that receives and transforms data. If a nonempty subset $\mathcal{F}$ of $\mathrm{GENEO}\left((\Phi, G), (\Psi, H)\right)$ is fixed, a simple pseudo-distance $D_{\mathcal{F},\Phi}(\varphi_1, \varphi_2)$ to compare two admissible functions $\varphi_1, \varphi_2 \in \Phi$ can be defined by setting $D_{\mathcal{F},\Phi}(\varphi_1, \varphi_2) := \sup_{F\in\mathcal{F}} \|F(\varphi_1) - F(\varphi_2)\|_\infty$. This definition expresses our assumption that the comparison of data strongly depends on the choice of the agents. However, we note that the computation of $D_{\mathcal{F},\Phi}(\varphi_1, \varphi_2)$ for every pair (φ_1, φ_2) of admissible functions is computationally expensive. We will see how persistent homology allows us to replace $D_{\mathcal{F},\Phi}$ with a pseudo-metric $\mathcal{D}^{\mathcal{F},k}_{\mathrm{match}}$ that is quicker to compute.

5. A strongly group-invariant pseudo-metric induced by Persistent Homology

In this section, we show how Persistent Homology supports the definition of a strongly group-invariant pseudo-metric on Φ, for which we prove some theoretical results.

We begin by recalling the stability of the classical pseudo-distance d_{match} between Persistent Betti Numbers (BPN) (cf. Definition 3.2) with respect to the pseudo-metrics D_{Φ} and $d_{\mathrm{Homeo}(X)}$. We assume the finiteness of PBNs [4]. Then, the stability of d_{match} easily follows from the stability theorem of the interleaving distance and the isometry theorem (cf., [13]).

Theorem 5.1. *If k is a natural number, $G_1 \subseteq G_2 \subseteq \mathrm{Homeo}_{\Phi}(X)$ and $\varphi_1, \varphi_2 \in \Phi$, then*

$$d_{\mathrm{match}}(r_k(\varphi_1), r_k(\varphi_2)) \leq d_{\mathrm{Homeo}(X)}(\varphi_1, \varphi_2) \leq d_{G_2}(\varphi_1, \varphi_2) \leq d_{G_1}(\varphi_1, \varphi_2) \leq D_{\Phi}(\varphi_1, \varphi_2). \quad (12)$$

The proof of the first inequality $d_{\mathrm{match}}(r_k(\varphi_1), r_k(\varphi_2)) \leq d_{\mathrm{Homeo}(X)}(\varphi_1, \varphi_2)$ in Theorem 5.1 can be found in [10]. The other inequalities follow from the definition of the natural pseudo-distance.

5.1. Strongly group invariant comparison of filtering functions via persistent homology

The proofs reported in the rest of this section are just a straightforward generalization of the proofs given in [12] for the case $\Phi = \Psi$, $T = id$.

Let us consider a subset $\mathcal{F} \neq \emptyset$ of $\mathcal{F}^{\mathrm{all}}$. For every fixed k, we can consider the following pseudo-metric $\mathcal{D}^{\mathcal{F},k}_{\mathrm{match}}$ on Φ:

$$\mathcal{D}^{\mathcal{F},k}_{\mathrm{match}}(\varphi_1, \varphi_2) := \sup_{F \in \mathcal{F}} d_{\mathrm{match}}(r_k(F(\varphi_1)), r_k(F(\varphi_2))) \quad (13)$$

for every $\varphi_1 \varphi_2 \in \Phi$, where $r_k(\varphi)$ denotes the kth persistent Betti number function with respect to the function $\varphi : X \to \mathbb{R}$.

In this work, we will say that a pseudo-metric $\hat{d}$ on Φ is *strongly G-invariant* if it is invariant under the action of G with respect to each variable, that is, if $\hat{d}(\varphi_1, \varphi_2) = \hat{d}(\varphi_1 \circ g, \varphi_2) = \hat{d}(\varphi_1, \varphi_2 \circ g) = \hat{d}(\varphi_1 \circ g, \varphi_2 \circ g)$ for every $\varphi_1, \varphi_2 \in \Phi$ and every $g \in G$.

Remark 5.2. It is easily seen that the natural pseudo-distance d_G is strongly G-invariant.

Proposition 5.3. *$\mathcal{D}^{\mathcal{F},k}_{\mathrm{match}}$ is a strongly G-invariant pseudo-metric on Φ.*

(The proof is in Appendix C.)

[4] Though in our setting, the space X is assumed to be compact, PBN functions are not necessarily finite. For example, let us consider the set $X = \{0\} \cup \{\frac{1}{n}, \text{ with } n \in \mathbb{N}\}$ and $\Phi = \{\mathrm{Id} : X \longrightarrow X,\ \mathrm{Id}(x) = x\}$. Even if X is compact, every sublevel set $X_u = \{x \in X, \text{ such that } x \leq u\}$ has infinite connected components, and hence the 0th persiste nt Betti Number function takes infinite value everywhere.

We add the assumption on the finiteness of PBN (i.e., the assumption that the PBN function of every $\varphi \in \Phi$ takes a finite value at each point $(u, v) \in \Delta^+$) to get stability and discard pathological cases (for example the case that the set Φ of admissible functions is the set of all maps from X to $\mathbb{R}$).

Since the PBN functions of the pseudo-metric space (X, D_X) coincide with the persistent Betti number functions of its Kolmogorov quotient $\bar{X}$, the finiteness of the persistent Betti number functions can be obtained when $\bar{X}$ is finitely triangulable (cf. [10]).

5.2. *Some theoretical results on the pseudo-metric* $\mathcal{D}_{\text{match}}^{\mathcal{F},k}$

The proofs reported in this section are a generalization of the proofs given in [12] for the case $\Phi = \Psi$, $T = id$. At first we want to show that the pseudo-metric $\mathcal{D}_{\text{match}}^{\mathcal{F},k}$ is stable with respect to both the natural pseudo-distance d_G associated with the group G and the distance D_Φ.

Remark 5.4. Let X and Y be two homeomorphic spaces and let $h : Y \to X$ be a homeomorphism. Then the persistent homology group with respect to the function $\varphi : X \to \mathbb{R}$ and the persistent homology group with respect to the function $\varphi \circ h : Y \to \mathbb{R}$ are isomorphic at each point (u, v) in the domain. Therefore we can say that the persistent homology groups and the persistent Betti number functions are invariant under the action of $\mathrm{Homeo}(X)$.

Theorem 5.5. *If $\mathcal{F}$ is a non-empty subset of $\mathcal{F}^{\text{all}}$, then*

$$\mathcal{D}_{\text{match}}^{\mathcal{F},k} \leq d_G \leq D_\Phi. \tag{14}$$

(The proof is in Appendix C.)

The definitions of the natural pseudo-distance d_G and the pseudo-distance $\mathcal{D}_{\text{match}}^{\mathcal{F},k}$ come from different theoretical concepts. The former is based on a variation approach involving the set of all homeomorphisms in G, while the latter refers only to a comparison of persistent homologies depending on a family of group equivariant non-expansive operators. Given those comments, the next result may appear unexpected.

Theorem 5.6. *Let us assume that $\Phi = \Psi$, every function in Φ is non-negative, the k-th Betti number of X does not vanish, and Φ contains each constant function c for which a function $\varphi \in \Phi$ exists such that $0 \leq c \leq \|\varphi\|_\infty$. Then $\mathcal{D}_{\text{match}}^{\mathcal{F}^{\text{all}},k} = d_G$.*

(The proof is in Appendix C.)

We observe that if Φ is bounded, the assumption that every function in Φ is non-negative is not quite restrictive. Indeed, we can obtain it by adding a suitable constant value to every admissible function.

5.3. *Beyond group equivariance*

We observe that while the definition of the natural pseudo-distance d_G requires that G has the structure of a group, the definition of $\mathcal{D}_{\text{match}}^{\mathcal{F},k}$ does not need this assumption. In other words, our approach based on GENEOs can be used also when we wish to have equivariance with respect to a *set* instead of a *group* of homeomorphisms. This property is promising for extending the application of our theory to the cases in which the agent is equivariant with respect to each element of a finite set of homeomorphisms that is not closed with respect to composition and computation of the inverse.

5.4. Pseudo-metrics induced by persistent homology

Persistent homology can be seen as a topological method to build new and easily computable pseudo-metrics for the sets Φ, G and $\mathcal{F}^{\text{all}}$. These new pseudo-metrics Δ_Φ, Δ_G, Δ_{GENEO} can be used as proxies for d_G (and hence D_Φ), D_G, D_{GENEO}, respectively:

- If $\varphi_1, \varphi_2 \in \Phi$, we can set $\Delta_\Phi(\varphi_1, \varphi_2) := d_{\text{match}}(r_k(\varphi_1), r_k(\varphi_2))$. The stability theorem for persistence diagrams (Theorem 5.1) can be reformulated as the inequality $\Delta_\Phi \leq d_G$.
- If $g_1, g_2 \in G$, we can set $\Delta_G(g_1, g_2) := \sup_{\varphi \in \Phi} d_{\text{match}}(r_k(\varphi \circ g_1), r_k(\varphi \circ g_2))$. From Theorem 5.1 the inequality $\Delta_G \leq D_G$ follows.
- If $F_1, F_2 \in \mathcal{F}^{\text{all}}$, we can set $\Delta_{\text{GENEO}}(F_1, F_2) := \sup_{\varphi \in \Phi} d_{\text{match}}(r_k(F_1(\varphi)), r_k(F_2(\varphi)))$. From Theorem 5.1 the inequality $\Delta_{\text{GENEO}} \leq D_{\text{GENEO,H}}$ follows.

In particular, Δ_Φ and a discretized version of the pseudo-metric Δ_{GENEO} will be used in the experiments described in Section 7. We underline that the use of persistent homology is a key tool in our approach: it allows for a fast comparison between functions and between GENEOs. Without persistent homology, this comparison would be much more computationally expensive.

5.5. Approximating $\mathcal{D}^{\mathcal{F},k}_{\text{match}}$

The next result will be of use for the approximation of $\mathcal{D}^{\mathcal{F},k}_{\text{match}}$.

Proposition 5.7. *Let $\mathcal{F}, \mathcal{F}' \subseteq \mathcal{F}^{\text{all}}$. If the Hausdorff distance*

$$HD(\mathcal{F}, \mathcal{F}') := \max \left\{ \sup_{F \in \mathcal{F}} \inf_{F' \in \mathcal{F}'} D_{GENEO,H}(F, F'), \sup_{F' \in \mathcal{F}'} \inf_{F \in \mathcal{F}} D_{GENEO,H}(F, F') \right\}$$

is not larger than ε, then

$$\left| \mathcal{D}^{\mathcal{F},k}_{\text{match}}(\varphi_1, \varphi_2) - \mathcal{D}^{\mathcal{F}',k}_{\text{match}}(\varphi_1, \varphi_2) \right| \leq 2\varepsilon \tag{15}$$

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

(The proof is in Appendix C.)

Therefore, if we can cover $\mathcal{F}$ by a finite set of balls in $\mathcal{F}^{\text{all}}$ of radius ε, centered at points of a finite set $\mathcal{F}' \subseteq \mathcal{F}$, the approximation of $\mathcal{D}^{\mathcal{F},k}_{\text{match}}(\varphi_1, \varphi_2)$ can be reduced to the computation of $\mathcal{D}^{\mathcal{F}',k}_{\text{match}}(\varphi_1, \varphi_2)$, i.e. the maximum of a finite set of bottleneck distances between persistence diagrams, which are well-known to be computable by means of efficient algorithms.

This fact leads us to study, in the following section, the properties of the topological space $\mathcal{F}^{\text{all}}$.

6. On the compactness and convexity of the space of GENEOs

In this section we show that if the function spaces we are considering are compact and convex, then the space of GENEOs is compact and convex too. This property has important consequences from the computational point of view, since it guarantees that the space of GENEOs can be approximated by a finite set and that new GENEOs can be obtained by convex combination of preexisting GENEOs.

6.1. The space of GENEOs is compact with respect to D_{GENEO}

We start by recalling that we are assuming Φ and Ψ compact with respect to D_{Φ} and D_{Ψ}, respectively.

Theorem 6.1. *$\mathcal{F}^{\mathrm{all}}$ is compact with respect to D_{GENEO}.*

(The proof is in Appendix C.)

Corollary 6.2. *Let $\mathcal{F}$ be a non-empty subset of $\mathcal{F}^{\mathrm{all}}$. For every $\varepsilon > 0$, a finite subset $\mathcal{F}^*$ of $\mathcal{F}$ exists, such that*

$$|\mathcal{D}_{\mathrm{match}}^{\mathcal{F}^*,k}(\varphi_1, \varphi_2) - \mathcal{D}_{\mathrm{match}}^{\mathcal{F},k}(\varphi_1, \varphi_2)| \leq \varepsilon \tag{16}$$

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

(The proof is in Appendix C.)

The previous corollary shows that, under suitable hypotheses, the computation of $\mathcal{D}_{\mathrm{match}}^{\mathcal{F}^{\mathrm{all}},k}$ can be reduced to the computation of the maximum of a finite set of bottleneck distances between persistence diagrams, for every $\varphi_1, \varphi_2 \in \Phi$.

Remark 6.3. Theorem 6.1 and the inequalities $\Delta_{\mathrm{GENEO}} \leq D_{\mathrm{GENEO,H}} \leq D_{\mathrm{GENEO}}$ stated in Subsection 5.4 immediately imply that $\mathcal{F}^{\mathrm{all}}$ is compact also with respect to the topologies induced by Δ_{GENEO} and $D_{\mathrm{GENEO,H}}$.

6.2. The set of GENEOs is convex

Let $F_1, F_2, \ldots, F_n$ be GENEOs from (Φ, G) to (Ψ, H) associated with the homomorphism T. Let $(a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$ with $\sum_{i=1}^{n} |a_i| \leq 1$. Consider the function

$$F_{\Sigma}(\varphi) := \sum_{i=1}^{n} a_i F_i(\varphi) \tag{17}$$

from Φ to the set $C^0(Y, \mathbb{R})$ of the continuous functions from Y to $\mathbb{R}$, where Y is the domain of the functions in Ψ.

Proposition 6.4. *If $F_{\Sigma}(\Phi) \subseteq \Psi$, then F_{Σ} is a GENEO from (Φ, G) to (Ψ, H) with respect to T.*

(The proof is in Appendix C.)

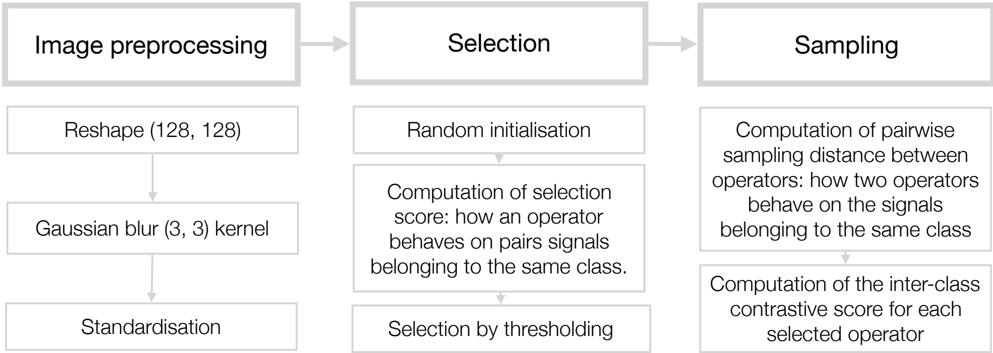

Figure 1: Experimental pipeline. Images are first reshaped and then blurred to obtain smooth border. Finally, images are then standardised. Operators are initialised randomly. Thereafter, they are selected according to their output when evaluated on object belonging to a chosen class, and sampled, so as to avoid redundant calculations.

Theorem 6.5. *If Ψ is convex, then the set of GENEOs from (Φ, G) to (Ψ, H) with respect to T is convex.*

(The proof is in Appendix C.)

7. Experimental validation

We validate our model on the MNIST and fashion-MNIST datasets. We first define isometry-equivariant non-expansive operators (IENEOs), then describe two simple algorithms allowing to select and sample IENEOs based on few labelled samples taken from the dataset. We show how selected and sampled operators can be used to perform both classical metric learning and effective initialisation of the kernels of a convolutional neural network.

7.1. Isometry-equivariant non-expansive operators (IENEOs)

We define a parametric family of non-expansive operators which are equivariant with respect to Euclidean plane isometries.

Given $\sigma > 0$ and $\tau \in \mathbb{R}$, we consider the 1-dimensional Gaussian function with width σ and center τ

$$g_\tau(t) := e^{-\frac{(t-\tau)^2}{2\sigma^2}},$$

$g_\tau : \mathbb{R} \to \mathbb{R}$. For a positive integer k, we take the set S of the $2k$-tuples $(a_1, \tau_1, \ldots, a_k, \tau_k) \in \mathbb{R}^{2k}$ for which $\sum_{i=1}^k a_i^2 = \sum_{i=1}^k \tau_i^2 = 1$. S is a submanifold of $\mathbb{R}^{2k}$.

For each $p = (a_1, \tau_1, \ldots, a_k, \tau_k) \in S$, we then consider the function $G_p : \mathbb{R}^2 \to \mathbb{R}$ defined as

$$G_p(x, y) := \sum_{i=1}^{k} a_i g_{\tau_i}\left(\sqrt{x^2 + y^2}\right).$$

If we denote by F_p the convolutional operator mapping each continuous function with compact support $\varphi : \mathbb{R}^2 \to \mathbb{R}$ to the continuous and with compact support function $\psi : \mathbb{R}^2 \to \mathbb{R}$ defined as

$$\psi(x, y) := \int_{\mathbb{R}^2} \varphi(\alpha, \beta) \cdot \frac{G_p(x - \alpha, y - \beta)}{\|G_p\|_{L^1}}\, d\alpha\, d\beta,$$

then the operator F_p is a group-equivariant non-expansive operator with respect to the group I of Euclidean plane isometries. We call F_p a IENEO (Isometry-Equivariant Non-Expansive Operator).

The IENEO F_p is parametric with respect to the $2k$-tuple $p = (a_1, \tau_1, \ldots, a_k, \tau_k) \in S$. Therefore, we define a parametric family of IENEOs

$$\mathcal{F} = \{F_p\}_{p \in S}.$$

The next section shows how to select a finite subfamily of $\mathcal{F}$ of operators which are suited to image classification in the MNIST and fashion-MNIST datasets, using the pseudo-metrics defined in Section 5.

7.2. Selection and sampling of IENEOs

We begin by randomly sampling n operators in $\mathcal{F}$. We then select those operators that consider as similar the objects belonging to the same class. To this end, let $\{\varphi_i^c\}_{i=1}^m$ be the set of functions representing the objects in a class c of cardinality $m \in \mathbb{N}$. For each of the n randomly sampled operators $F_p \in \mathcal{F}$, we define the class-dependent value

$$s_c\,(F_p) = \max_{\varphi_i^c, \varphi_j^c} d_{\text{match}}(r_k(F_p(\varphi_i^c)), r_k(F_p(\varphi_j^c))).$$

An operator F_p is *selected* if $s_c(F_p)$ is smaller than a threshold ϵ_c, for each class c.

Once we have selected a smaller number of operators according to the criterion above, we then sample operators to avoid storing operators that would focus on the same or similar characteristic across classes. To this end, given a class c, we define the distance between two operators F_p and $F_{p'}$ (cf. Section 5.4)

$$\Delta^c_{\text{IENEO}}\,(F_p, F_{p'}) := \max_{\varphi_i^c} d_{\text{match}}(r_k(F_p(\varphi_i^c)), r_k(F_{p'}(\varphi_i^c))).$$

For every class c, we sort the pairs $(F_p, F_{p'})$ in ascending order of Δ^c_{IENEO}, and assign each pair with its index in the sorted list. We then define an interclass contrastive score of the pair $(F_p, F_{p'})$ as the sum of its indices over all classes. Finally, an operator F_p is *sampled* if it belongs to a pair whose score is above a threshold t.

When we have generated a finite family $\mathcal{F}'$ of selected and sampled GENEOs, two objects φ_1 and φ_2 can be compared by computing the strongly G-invariant pseudo-metric $\mathcal{D}^{\mathcal{F}'}_{\text{match}}(\varphi_1, \varphi_2)$, defined in Section 5.

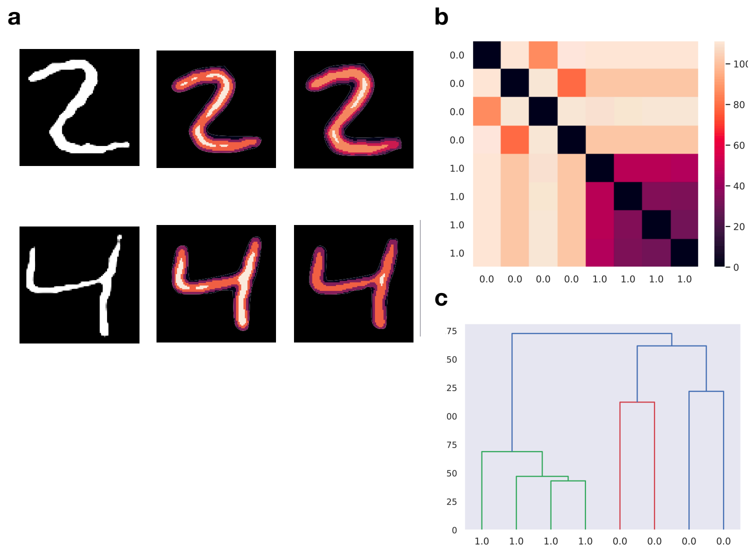

Figure 2: IENEOs on the MNIST dataset. (a) Two samples from MNIST and the action of two IENEOs. The images are computed by convolving selected and sampled operators with the original samples. (b) The distance matrix and the hierarchical clustering (c) obtained by selecting and sampling operators on 20 samples of MNIST (10 per class). Starting from 200 randomly initialised IENEOs, 24 were extracted by selection and sampling. Those operators were used to compute pairwise distances between validation samples.

7.3. Results

The aim of our applications is twofold. On one side, we want to validate the distances defined in Section 5; on the other side, we want to prove that selected and sampled IENEOs provide information that can be integrated in the current machine learning algorithms. These objectives motivate the applications in classical metric learning described in Figures 2 and 3, and the more deep learning-oriented experiment in Figure 4.

The image preprocessing and the selection and sampling protocols are represented as a flow-chart in Figure 1. More precisely, in our experiments, we initialise 200 2-dimensional IENEOs of size 11. The centres of the 5 Gaussians constituting each of the operators are sampled randomly in the interval $[-1, 1]$, whilst the standard deviation is constant for all Gaussians and set to 0.1. For each dataset we consider, in these examples, only two classes randomly selected among the available ones. After choosing the two classes we select randomly ten samples per class. These samples will be used to evaluate the operators during selection and sampling. The selection threshold ϵ_c is $\tau \cdot \sigma(s_c)$, with $\tau = 1.5$. The standard deviation σ is computed by considering the values of $s_c(F_p)$ across all operators. The threshold t for sampling is defined as the $75th$ percentile of all contrastive scores. 0-degree persistent homology is used to compute the matching distances.

Figures 2 and 3 show how selecting and sampling operators amounts to learn an invariant metric on samples from the MNIST and fashion-MNIST datasets.

Figure 4 compares the performance obtained by classifying two classes of MNIST with a dense classifier evaluated on 64 non-trainable filters, with kernels initialised by selected and

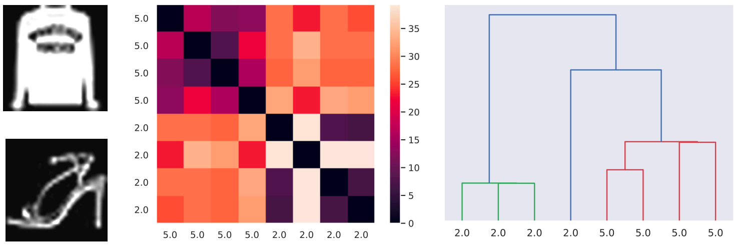

Figure 3: The metric learned on fashion-MNIST by selecting and sampling 200 IENEOs. The selection and sampling procedures outputted only 4 operators, that have been used to compute the pairwise distances between validation samples.

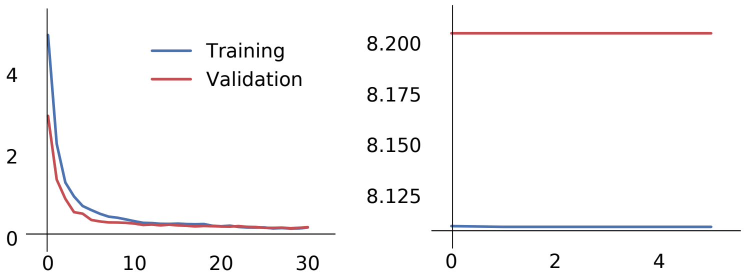

Figure 4: IENEOs versus random: Comparison between the performance obtained by classifying two classes of MNIST with a dense classifier evaluated on 64 non-trainable filters. (Left) Kernels are initialised by using selected and samples GENEOs. (Right) Random filters are initialised with Glorot initialisation [14].

samples GENEOs *vs* random filters initialised with Glorot initialisation [14].

Additional experiments are in progress and will be included in a forthcoming version of the paper.

8. Conclusion

We suggested a novel mathematical framework for machine learning based on the study of metric and topological properties of function spaces. This approach is dual to the classical one: data are thought of as points in a function space. Thus, learning consists in choosing a suitable set of operators defined on the function space that represents the data. Of all possible type of operators, we study the space of non-expansive, group equivariant operators (GENEOs). When building a machine learning system, choosing to work on a space of operators equivariant with respect to specific transformations allows us to inject in the system pre-existing knowledge. Indeed, the operators will be blind to the action of the group on the data, hence reducing the dimensionality of the space to be explored during optimisation.

Presenting our mathematical model, we first show how the space of GENEOs is suitable for machine learning. By using pseudo-metrics, we define a topology on the space of GENEOs which is induced by the one we define on the function space of data. We build the necessary machinery to define maps between GENEOs whose groups of equivariance are different. This definition is fundamental, because it allows to compose operators hierarchically, in the same fashion as computational units are linked in an artificial neural network. Thereafter, by taking advantage of known and novel results in persistent homology, we prove compactness and convexity of the space of GENEOs under suitable hypotheses. Moreover and importantly, we show how the suggested framework can be used to study operators that are equivariant with respect to set of transformations, rather than groups.

We give two algorithms that allow to select and sample from a space of operators given a dataset labelled for a classification task. These procedures allow to first select a subset of operators belonging to a certain GENEOs space, that give meaningful representation of the data with respect to their labelling, always invariant under the transformations induced by the action of G. Thenceforth, the sampling algorithm allows to eliminate redundant operators. These two strategies are used to perform metric learning and kernel on MNIST and fashion-MNIST. In addition, we show how convolutional filters initialised by selecting and sampling on few samples effectively grasp useful knowledge, that can be utilised to classify the remainder of the samples, for instance by a dense classifier.

Our forward-looking goal is the one of defining a novel artificial neural network model based on functional modules. Modules would be more complex computational units than the standard artificial neuron. The core of each module would be a GENEO, thus each module would be defined a priori to be equivariant with respect to a set of transformations. On one hand, this choice would allow us to dramatically reduce the dimensionality of the manifold to be studied during optimisation. On the other hand, choosing the transformation equivariances to be respected at each layer would allow us to inject knowledge in the networks before training, and would assure that information is not acquired by relying on unwanted noisy regularities in the training data. Module networks would learn optimal transformations of the data to achieve a task, rather than operating on data themselves.

Module networks could be built by composing modules hierarchically and knowledge could be injected in the model by engineering the proper set of equivariances. These transformations would be easily interpretable and could offer a rigorous way to compare learning dynamics of different architectures during optimisation. In particular, we are investigating the possibility to generalize capsule networks [15, 16] and modify the dynamic routing algorithm, by using the metrics on the space of GENEOs to update the connectivity strength between modules.

Appendix A Additional propositions

Proposition A.1. *The function D_X is an extended pseudo-metric on X.*

Remark A.2. We recall that a pseudo-metric is just a distance d without the property: if $d(a, b) = 0$, then $a = b$.

Proof.

- D_X is obviously symmetrical.
- The definition of D_X immediately implies that $D_X(x, x) = 0$ for any $x \in X$.

- The triangle inequality holds, since

$$\begin{aligned} D_X(x_1,x_2) &= \sup_{\varphi\in\Phi}|\varphi(x_1)-\varphi(x_2)| \\ &\le \sup_{\varphi\in\Phi}(|\varphi(x_1)-\varphi(x_3)|+|\varphi(x_3)-\varphi(x_2)|) \\ &\le \sup_{\varphi\in\Phi}|\varphi(x_1)-\varphi(x_3)|+\sup_{\varphi\in\Phi}|\varphi(x_3)-\varphi(x_2)| \\ &= D_X(x_1,x_3)+D_X(x_3,x_2) \end{aligned}$$

  for any $x_1, x_2, x_3 \in X$.

□

Proposition A.3. *If Φ is totally bounded, then for any $\delta > 0$ there exists a finite subset Φ_δ of Φ such that*

$$\left|\sup_{\varphi\in\Phi}|\varphi(x_1)-\varphi(x_2)|-\max_{\varphi\in\Phi_\delta}|\varphi(x_1)-\varphi(x_2)|\right|<2\delta$$

for every $x_1, x_2 \in X$.

Proof. Let us fix $x_1, x_2 \in X$. Since Φ is totally bounded, we can find a finite subset $\Phi_\delta = \{\varphi_1, \ldots, \varphi_n\}$ such that for each $\varphi \in \Phi$ there exists $\varphi_i \in \Phi_\delta$, for which $\|\varphi - \varphi_i\|_\infty < \delta$. It follows that for any $x \in X$, $|\varphi(x) - \varphi_i(x)| < \delta$. Because of the definition of supremum of a subset of the set $\mathbb{R}^+$ of all positive real numbers, for any $\varepsilon > 0$ we can choose a $\bar{\varphi} \in \Phi$ such that

$$\sup_{\varphi\in\Phi}|\varphi(x_1)-\varphi(x_2)|-|\bar{\varphi}(x_1)-\bar{\varphi}(x_2)|\le\varepsilon.$$

Now, if we take an index i, for which $\|\bar{\varphi} - \varphi_i\|_\infty < \delta$, we have that:

$$\begin{aligned} |\bar{\varphi}(x_1)-\bar{\varphi}(x_2)| &= |\bar{\varphi}(x_1)-\varphi_i(x_1)+\varphi_i(x_1)-\varphi_i(x_2)+\varphi_i(x_2)-\bar{\varphi}(x_2)| \\ &\le |\bar{\varphi}(x_1)-\varphi_i(x_1)|+|\varphi_i(x_1)-\varphi_i(x_2)|+|\varphi_i(x_2)-\bar{\varphi}(x_2)| \\ &< |\varphi_i(x_1)-\varphi_i(x_2)|+2\delta \\ &\le \max_{\varphi_j\in\Phi_\delta}|\varphi_j(x_1)-\varphi_j(x_2)|+2\delta. \end{aligned}$$

Hence,

$$\sup_{\varphi\in\Phi}|\varphi(x_1)-\varphi(x_2)|-\varepsilon<|\bar{\varphi}(x_1)-\bar{\varphi}(x_2)|<\max_{\varphi_j\in\Phi_\delta}|\varphi_j(x_1)-\varphi_j(x_2)|+2\delta.$$

Finally, as ε goes to zero, we have that

$$\sup_{\varphi\in\Phi}|\varphi(x_1)-\varphi(x_2)|<\max_{\varphi_j\in\Phi_\delta}|\varphi_j(x_1)-\varphi_j(x_2)|+2\delta.$$

On the other hand, since $\Phi_\delta \subseteq \Phi$:

$$\sup_{\varphi\in\Phi}|\varphi(x_1)-\varphi(x_2)|>\max_{\varphi_j\in\Phi_\delta}|\varphi_j(x_1)-\varphi_j(x_2)|-2\delta.$$

Therefore we proved the statement.

□

Proposition A.4. *The function D_G is a pseudo-metric on G.*

Proof.

- The value $D_G(g_1, g_2)$ is finite for every $g_1, g_2 \in G$, because Φ is compact and hence bounded. Indeed, a finite constant L exists such that $\|\varphi\|_\infty \le L$ for every $\varphi \in \Phi$. Hence, $\|\varphi \circ g_1 - \varphi \circ g_2\|_\infty \le \|\varphi\|_\infty + \|\varphi\|_\infty \le 2L$ for any $\varphi \in \Phi$ and any $g_1, g_2 \in G$, since $\varphi \circ g_1, \varphi \circ g_2 \in \Phi$. This implies that $D_G(g_1, g_2) \le 2L$ for every $g_1, g_2 \in G$.
- D_G is obviously symmetrical.
- The definition of D_G immediately implies that $D_G(g, g) = 0$ for any $g \in G$.
- The triangle inequality holds, since

$$\begin{aligned} D_G(g_1, g_2) &= \sup_{\varphi\in\Phi} \|\varphi \circ g_1 - \varphi \circ g_2\|_\infty \\ &\le \sup_{\varphi\in\Phi}(\|\varphi \circ g_1 - \varphi \circ g_3\|_\infty + \|\varphi \circ g_3 - \varphi \circ g_2\|_\infty) \\ &\le \sup_{\varphi\in\Phi} \|\varphi \circ g_1 - \varphi \circ g_3\|_\infty + \sup_{\varphi\in\Phi} \|\varphi \circ g_3 - \varphi \circ g_2\|_\infty \\ &= D_G(g_1, g_3) + D_G(g_3, g_2) \end{aligned} \tag{18}$$

  for any $g_1, g_2, g_3 \in G$.

□

Appendix B Our approach in terms of slice categories

In this section, we will apply the concept of slice category to our framework in order to formalize the concept of perception pairs, which are considered as subcategories of a larger category denoted by ${}^{\mathrm{PMet}}\!\diagup_{(\mathbb{R}, d_e)}$, as we explain further. Moreover we explore the link between GENEOs and functors between categories of this kind.

Let PMet be the category whose objects are pseudo-metric spaces and morphisms are the continuous functions between them. Let us fix the space $(\mathbb{R}, d_e)$, that is the real line equipped with the usual Euclidean metric, and consider the slice category over $(\mathbb{R}, d_e)$.

Now we recall the definition of slice category:

Definition B.1. The slice category ${}^{C}\!\diagup_{c}$ of a category C over an object $c \in C$ has

- objects that are all arrows $f \in C$ such that $\mathrm{cod}(f) = c$,
- morphisms that are all triples $g_{f,f'} := (f, g, f')$ where $f : X \longrightarrow c$ and $f' : X' \longrightarrow c$ are two objects of ${}^{C}\!\diagup_{c}$, $g : X \longrightarrow X'$ is a morphism of C such that $f = f' \circ g$; $\mathrm{dom}(g_{f,f'}) = f$ and $\mathrm{cod}(g_{f,f'}) = f'$.

The slice category is a special case of a comma category.

Remark B.2. There is a forgetful functor $U_c : {}^{C}\!\diagup_{c} \longrightarrow C$ which maps each object $f : X \longrightarrow c$ to its domain X and each morphism $g_{f,f'}$ between $f : X \longrightarrow c$ and $f' : X' \longrightarrow c$ to the morphism $g : X \longrightarrow X'$.

We are going to associate a perception pair $(\varPhi, G)$ with a subcategory $C(\varPhi, G)$ of $\mathrm{PMet}\diagup_{(\mathbb{R}, d_e)}$ defined as follows:

- the objects of $C(\varPhi, G)$ are the elements of $\varPhi$;
- the arrows of $C(\varPhi, G)$ are the triples $(f, g, f \circ g)$, where $f \in \varPhi$ and $g \in G$.

We observe that the action of G on $\varPhi$ ensures us that the arrow $(f, g, f \circ g)$ is well-defined for any $f \in \varPhi$ and any $g \in G$.

Now we can define a "functorial" version of the concept of GENEO.

Definition B.3. Let us consider two categories $C(\varPhi, G)$ and $C(\Psi, H)$. A functor F from $C(\varPhi, G)$ to $C(\Psi, H)$ is a C-GENEO if:

- $D_\Psi(F(\varphi), F(\varphi')) \le D_\varPhi(\varphi, \varphi')$ for any $\varphi,\ \varphi' \in \varPhi$;
- for any pair of morphisms $m,\ m' \in \hom(C(\varPhi, G))$ such that $U_\mathbb{R}(m) = U_\mathbb{R}(m')$ we have that $U_\mathbb{R}(F(m)) = U_\mathbb{R}(F(m'))$.

GENEOs and C-GENEOs share the non-expansivity condition. The proposition below shows that the second conditions respectively required in the definitions of GENEO and C-GENEO correspond to each other in a suitable sense. We omit its trivial proof.

Proposition B.4. *Let F be a functor from $C(\varPhi, G)$ and $C(\Psi, H)$. The following conditions are equivalent:*

- *there exists a group homomorphism $T : G \longrightarrow H$ such that $F(\varphi \circ g) = F(\varphi) \circ T(g)$ for any $\varphi \in \varPhi$ and any $g \in G$;*
- *for any pair of morphisms $m,\ m' \in \hom(C(\varPhi, G))$ such that $U_\mathbb{R}(m) = U_\mathbb{R}(m')$ we have that $U_\mathbb{R}(F(m)) = U_\mathbb{R}(F(m'))$.*

Appendix C Proofs

Theorem (4.1)**.** *The topology τ_{D_X} on X induced by the pseudo-metric D_X is finer than the initial topology τ_{in} on X with respect to $\varPhi$. If $\varPhi$ is totally bounded, then the topology τ_{D_X} coincides with τ_{in}.*

Proof. We know that the set $\mathcal{B}_{D_X} = \{B_X(x, \varepsilon) : x \in X, \varepsilon > 0\}$ is a base for the topology τ_{D_X} and the set $\mathcal{B}_{\text{in}} = \left\{\bigcap_{i \in I} \varphi_i^{-1}(U_i) : |I| < \infty, U_i \in \mathcal{T}_E\ \forall i \in I\right\}$ is a base for the topology τ_{in}.

First of all we have to show that the topology τ_{D_X} is finer than the initial topology τ_{in}. Let us take a set in the base $\mathcal{B}_{\text{in}}$ of τ_{in}, i.e. a set $\bigcap_{i \in I} \varphi_i^{-1}(U_i)$, where I is a finite set of indexes and $U_i \in \mathcal{T}_E$ for every index $i \in I$. It will be sufficient to show that for every $y \in \bigcap_{i \in I} \varphi_i^{-1}(U_i)$ a ball $B_X(y, \varepsilon) \in \mathcal{B}_{D_X}$ exists, such that $B_X(y, \varepsilon) \subseteq \bigcap_{i \in I} \varphi_i^{-1}(U_i)$. Since $y \in \bigcap_{i \in I} \varphi_i^{-1}(U_i)$, we have that $\varphi_i(y) \in U_i$ for every $i \in I$. Therefore, for each $i \in I$ we can find an open interval $]a_i, b_i[$ such that $\varphi_i(y) \in]a_i, b_i[\subseteq U_i$. Let us set $\varepsilon := \min_{i \in I} \min\{\varphi_i(y) - a_i, b_i - \varphi_i(y)\}$, and observe that $\varepsilon > 0$. If $z \in B_X(y, \varepsilon)$, then $|\varphi(y) - \varphi(z)| < \varepsilon$ for every $\varphi \in \varPhi$, and in particular $|\varphi_i(y) - \varphi_i(z)| < \varepsilon$ for every $i \in I$. Hence the definition of ε immediately implies that $\varphi_i(z) \in]a_i, b_i[$ for every $i \in I$, so that $z \in \bigcap_{i\in I} \varphi_i^{-1}(]a_i, b_i[)$. It follows that $B_X(y,\varepsilon) \subseteq \bigcap_{i\in I} \varphi_i^{-1}(]a_i, b_i[) \subseteq \bigcap_{i\in I} \varphi_i^{-1}(U_i)$. Therefore, $y \in B_X(y,\varepsilon) \subseteq \bigcap_{i\in I} \varphi_i^{-1}(U_i)$, and our first statement is proved.

If Φ is totally bounded, Proposition A.3 in Appendix A guarantees that for every $\delta > 0$ a finite subset Φ_δ of Φ exists such that

$$\left| \sup_{\varphi\in\Phi} |\varphi(x_1) - \varphi(x_2)| - \max_{\varphi\in\Phi_\delta} |\varphi(x_1) - \varphi(x_2)| \right| < 2\delta \tag{19}$$

for every $x_1, x_2 \in X$. Let us now set $B_\delta(x,r) := \left\{ x' \in X \middle| \max_{\varphi_i\in\Phi_\delta} |\varphi_i(x) - \varphi_i(x')| < r \right\}$ for every $x \in X$ and every $r > 0$. We have to prove that the initial topology τ_{in} is finer than the topology τ_{D_X}. In order to do this, it will be sufficient to show that for every $y \in B_X(x,\varepsilon) \in \mathcal{B}_{D_X}$ a set $\bigcap_{i\in I} \varphi_i^{-1}(U_i) \in \mathcal{B}_{\text{in}}$ exists, such that $y \in \bigcap_{i\in I} \varphi_i^{-1}(U_i) \subseteq B_X(x,\varepsilon)$.

Let us choose a positive δ such that $2\delta < \varepsilon$. Inequality (19) implies that $B_\delta(y, \varepsilon - 2\delta) \subseteq B_X(y,\varepsilon)$. We now set $U_i :=]\varphi_i(y) - \varepsilon + 2\delta, \varphi_i(y) + \varepsilon - 2\delta[$ for $i \in I$. Obviously, $y \in \bigcap_{\varphi_i\in\Phi_\delta} \varphi_i^{-1}(U_i)$. If $z \in \bigcap_{\varphi_i\in\Phi_\delta} \varphi_i^{-1}(U_i)$, then $|\varphi_i(z) - \varphi_i(y)| < \varepsilon - 2\delta$ for every $\varphi_i \in \Phi_\delta$. Hence, $z \in B_\delta(y, \varepsilon - 2\delta)$. It follows that $\bigcap_{\varphi_i\in\Phi_\delta} \varphi_i^{-1}(U_i) \subseteq B_\delta(y, \varepsilon - 2\delta)$. Therefore, $y \in \bigcap_{\varphi_i\in\Phi_\delta} \varphi_i^{-1}(U_i) \subseteq B_X(x,\varepsilon)$ because of the inclusion $B_\delta(y, \varepsilon - 2\delta) \subseteq B_X(y,\varepsilon)$. This means that τ_{in} is finer than τ_{D_X}. Since we already know that τ_{D_X} is finer than τ_{in}, it follows that τ_{D_X} coincides with τ_{in}. □

Remark. The second statement of Theorem 4.1 becomes false if Φ is not totally bounded. For example, assume Φ equal to the set of all functions from $X = [0,1]$ to $\mathbb{R}$ that are continuous with respect to the Euclidean topologies on $[0,1]$ and $\mathbb{R}$. Indeed, it is easy to check that in this case τ_{D_X} is the discrete topology, while the initial topology τ_{in} is the Euclidean topology on $[0,1]$.

Remark. The pseudo-metric space (X, D_X) may not be a T_0-space. For example, this happens if X is a space containing at least two points and Φ is the set of all the constant functions from X to $[0,1]$.

Theorem (4.2)**.** *If Φ is compact and X is complete then X is also compact.*

Proof. First of all we want to prove that every sequence (x_i) in X admits a Cauchy subsequence in X. After that, the statement follows immediately because every Cauchy sequence in a complete space is convergent, so that X is sequentially compact, and hence compact, since X is a pseudo-metric space [11].

Let us consider an arbitrary sequence (x_i) in X and an arbitrarily small $\varepsilon > 0$. Since Φ is compact, we can find a finite subset $\Phi_\varepsilon = \{\varphi_1, \ldots, \varphi_n\}$ such that $\Phi = \bigcup_{i=1}^n B_\Phi(\varphi_i, \varepsilon)$, where $B_\Phi(\varphi, \varepsilon) = \{\varphi' \in \Phi : D_\Phi(\varphi', \varphi) < \varepsilon\}$. In particular, we can say that for any $\varphi \in \Phi$ there exists $\varphi_{\bar{k}} \in \Phi_\varepsilon$ such that $\|\varphi - \varphi_{\bar{k}}\|_\infty < \varepsilon$. Now, we consider the real sequence $\varphi_1(x_i)$ that is bounded because all the functions in Φ are bounded. From Bolzano-Weierstrass Theorem it follows that we can extract a convergent subsequence $\varphi_1(x_{i_h})$. Then we consider the sequence

$\varphi_2(x_{i_h})$. Since φ_2 is bounded, we can extract a convergent subsequence $\varphi_2(x_{i_{h_t}})$. We can repeat the same argument for any $\varphi_k \in \Phi_\varepsilon$. Thus, we obtain a subsequence (x_{i_j}) of (x_i), such that $\varphi_k(x_{i_j})$ is a real convergent sequence for any $k \in \{1, \dots, n\}$, and hence a Cauchy sequence in $\mathbb{R}$. Moreover, since Φ_ε is a finite set, there exists an index $\bar{\jmath}$ such that for any $k \in \{1, \dots, n\}$ we have that

$$|\varphi_k(x_{i_r}) - \varphi_k(x_{i_s})| < \varepsilon, \quad \forall\, r, s \geq \bar{\jmath}. \tag{20}$$

We observe that $\bar{\jmath}$ does not depend on φ, but only on ε and Φ_ε.

In order to prove that (x_{i_j}) is a Cauchy sequence in X, we observe that for any $r, s \in \mathbb{N}$ and any $\varphi \in \Phi$ we have:

$$\begin{aligned} |\varphi(x_{i_r}) - \varphi(x_{i_s})| &= |\varphi(x_{i_r}) - \varphi_k(x_{i_r}) + \varphi_k(x_{i_r}) - \varphi_k(x_{i_s}) + \varphi_k(x_{i_s}) - \varphi(x_{i_s})| \\ &\leq |\varphi(x_{i_r}) - \varphi_k(x_{i_r})| + |\varphi_k(x_{i_r}) - \varphi_k(x_{i_s})| + |\varphi_k(x_{i_s}) - \varphi(x_{i_s})| \\ &\leq \|\varphi - \varphi_k\|_\infty + |\varphi_k(x_{i_r}) - \varphi_k(x_{i_s})| + \|\varphi_k - \varphi\|_\infty. \end{aligned} \tag{21}$$

It follows that $|\varphi(x_{i_r}) - \varphi(x_{i_s})| < 3\varepsilon$ for every $\varphi \in \Phi$ and every $r, s \geq \bar{\jmath}$. Thus, $\sup_{\varphi \in \Phi} |\varphi(x_{i_r}) - \varphi(x_{i_s})| = D_X(x_{i_r}, x_{i_s}) \leq 3\varepsilon$. Hence, the sequence (x_{i_j}) is a Cauchy sequence in X. The completeness of X implies that the statement of Theorem 4.2 is true. □

Example. Let Φ be the set containing all the 1-Lipschitz functions from $X = \{(x, y) \in \mathbb{R}^3 : x^2 + y^2 = 1, \arcsin(x) \in \mathbb{Q}\}$ to $[0, 1]$, and G be the group of all rotations $\rho_{2\pi q}$ of $2\pi q$ radians with $q \in \mathbb{Q}$. The topological space X is neither complete nor compact.

Proposition (4.3)**.** *If g is a bijection from X to X such that $\varphi \circ g \in \Phi$ and $\varphi \circ g^{-1} \in \Phi$ for every $\varphi \in \Phi$, then g is an isometry (and hence a homeomorphism) with respect to D_X.*

Proof. Let us fix two arbitrary points x, x' in X. Obviously, the map $R_g : \Phi \to \Phi$ taking each function φ to $\varphi \circ g$ is surjective, since $\varphi = R_g\left(R_{g^{-1}}(\varphi)\right)$. Hence $R_g(\Phi) = \Phi$. Therefore, g preserves the pseudo-distance D_X:

$$\begin{aligned} D_X(g(x), g(x')) &= \sup_{\varphi \in \Phi} |\varphi(g(x)) - \varphi(g(x'))| \\ &= \sup_{\varphi \in \Phi} |(\varphi \circ g)(x) - (\varphi \circ g)(x')| \qquad (22) \\ &= \sup_{\varphi \in R_g(\Phi)} |\varphi(x) - \varphi(x')| \qquad (23) \\ &= \sup_{\varphi \in \Phi} |\varphi(x) - \varphi(x')| = D_X(x, x'), \end{aligned}$$

Since g is bijective, it follows that g is an isometry with respect to D_X. □

Theorem (4.7)**.** *G is a topological group with respect to the pseudo-metric topology and the action of G on Φ through right composition is continuous.*

Proof. It will suffice to prove that if $f = \lim_{i\to+\infty} f_i$ and $g = \lim_{i\to+\infty} g_i$ in G with respect to the pseudo-metric D_G, then $g \circ f = \lim_{i\to+\infty} g_i \circ f_i$ and $f^{-1} = \lim_{i\to+\infty} f_i^{-1}$.

Because of the compactness of Φ and Proposition A.3, for every $\delta > 0$ we can take a finite subset Φ_δ of Φ such that

$$\left| \sup_{\varphi\in\Phi} |\varphi(x_1) - \varphi(x_2)| - \max_{\varphi\in\Phi_\delta} |\varphi(x_1) - \varphi(x_2)| \right| < 2\delta$$

for every $x_1, x_2 \in X$. We have that

$$\begin{aligned}
&D_G(g_i \circ f_i, g \circ f) \le D_G(g_i \circ f_i, g \circ f_i) + D_G(g \circ f_i, g \circ f) = \\
&= \sup_{\varphi\in\Phi} \|\varphi \circ (g_i \circ f_i) - \varphi \circ (g \circ f_i)\|_\infty + \sup_{\varphi\in\Phi} \|\varphi \circ (g \circ f_i) - \varphi \circ (g \circ f)\|_\infty \qquad (24)\\
&= \sup_{\varphi\in\Phi} \sup_{x\in X} |\varphi(g_i(f_i(x)) - \varphi(g(f_i(x))| + \sup_{\varphi\in\Phi} \sup_{x\in X} |\varphi(g(f_i(x)) - \varphi(g(f(x))| \\
&= \sup_{\varphi\in\Phi} \sup_{y\in X} |\varphi(g_i(y)) - \varphi(g(y))| + \sup_{\varphi\in\Phi} \sup_{x\in X} |\varphi(g(f_i(x)) - \varphi(g(f(x))| \\
&< D_G(g_i, g) + \max_{\varphi\in\Phi_\delta} \sup_{x\in X} |\varphi(g(f_i(x)) - \varphi(g(f(x))| + 2\delta. \qquad (25)
\end{aligned}$$

Since $g = \lim_{i\to+\infty} g_i$, $\lim_{i\to+\infty} D_G(g_i, g) = 0$. Because of Theorem 4.2, X is compact and hence $\varphi \circ g : X \to \mathbb{R}$ is a uniformly continuous function. Since $f = \lim_{i\to+\infty} f_i$, it follows that $\lim_{i\to+\infty} \sup_{x\in X} |\varphi(g(f_i(x)) - \varphi(g(f(x))| = 0$ for every $\varphi \in \Phi_\delta$, and hence $\lim_{i\to+\infty} \max_{\varphi\in\Phi_\delta} \sup_{x\in X} |\varphi(g(f_i(x)) - \varphi(g(f(x))| = 0$. Given that δ can be taken arbitrarily small, we get $g \circ f = \lim_{i\to+\infty} g_i \circ f_i$.

We also want to prove that $f^{-1} = \lim_{i\to+\infty} f_i^{-1}$. By contradiction, if we had not that $\lim_{i\to\infty} D_G(f_i^{-1}, f^{-1}) = 0$, then there would exist a constant $c > 0$ and a subsequence (f_{i_j}) of (f_i) such that $D_G(f_{i_j}^{-1}, f^{-1}) \ge c > 0$ for every index j. However, we should still have $\lim_{j\to\infty} D_G(f_{i_j}, f) = 0$ because (f_{i_j}) is a subsequence of (f_i). Since $D_G(f_{i_j}^{-1}, f^{-1}) \ge c > 0$ for every index j, a $\varphi_j \in \Phi$ should exist such that $\|\varphi_j \circ f_{i_j}^{-1} - \varphi_j \circ f^{-1}\|_\infty \ge c > 0$.

Because of the compactness of Φ, it would not be restrictive to assume (possibly by considering subsequences) the existence of the following limits: $\bar\varphi = \lim_{j\to\infty} \varphi_j$ and $\hat\varphi = \lim_{j\to\infty} \varphi_j \circ f_{i_j}^{-1}$. We would have that

$$\begin{aligned}
D_\Phi(\hat\varphi, \bar\varphi \circ f^{-1}) &= D_\Phi(\lim_{j\to\infty} \varphi_j \circ f_{i_j}^{-1}, \lim_{j\to\infty} \varphi_j \circ f^{-1}) \\
&= \lim_{j\to\infty} D_\Phi(\varphi_j \circ f_{i_j}^{-1}, \varphi_j \circ f^{-1}) \ge c > 0 \qquad (26)
\end{aligned}$$

so that $\hat\varphi \ne \bar\varphi \circ f^{-1}$.
On the other hand, we should have

$$\begin{aligned}
D_\Phi(\hat\varphi \circ f, \bar\varphi) &= D_\Phi((\lim_{j\to\infty} \varphi_j \circ f_{i_j}^{-1}) \circ f, \lim_{j\to\infty} \varphi_j) \\
&= \lim_{j\to\infty} D_\Phi((\varphi_j \circ f_{i_j}^{-1}) \circ f, (\varphi_j \circ f_{i_j}^{-1}) \circ f_{i_j}) \qquad (27)\\
&\le \lim_{j\to\infty} D_G(f_{i_j}, f) = 0
\end{aligned}$$

so that $\hat{\varphi} \circ f = \bar{\varphi}$.

It follows that R_f is not injective, against our assumptions.

This contradiction proves that $\lim_{i\to\infty} f_i^{-1} = f^{-1}$.

Therefore, G is a topological group.

Let now ε be a positive real number. If $D_\Phi(\varphi, \psi), D_G(f, g) < \delta := \varepsilon/2$ then

$$\begin{aligned} D_\Phi(\varphi \circ f, \psi \circ g) &\leq D_\Phi(\varphi \circ f, \varphi \circ g) + D_\Phi(\varphi \circ g, \psi \circ g) \\ &= D_\Phi(\varphi \circ f, \varphi \circ g) + D_\Phi(\varphi, \psi) \\ &\leq D_G(f, g) + D_\Phi(\varphi, \psi) < \varepsilon/2 + \varepsilon/2 = \varepsilon. \end{aligned} \tag{28}$$

This proves that the action of G on Φ through right composition is continuous. □

Theorem (4.8)**.** *If G is complete then it is also compact with respect to D_G.*

Proof. We want to show that G is sequentially compact, and hence compact. Let (g_i) be a sequence in G and take a real number $\varepsilon > 0$. Given that Φ is compact, we can find a finite subset $\Phi_\varepsilon = \{\varphi_1, \ldots, \varphi_n\}$ such that for every $\varphi \in \Phi$ there exists $\varphi_h \in \Phi_\varepsilon$ for which $D_\Phi(\varphi_h, \varphi) < \varepsilon$. For any fixed $k \in \{1, \ldots, n\}$, let us consider the sequence $(\varphi_k \circ g_i)$ in Φ. Applying the same argument as in the proof of Theorem 4.2, we can extract a subsequence (g_{i_j}) of (g_i) such that $(\varphi_k \circ g_{i_j})$ converges in Φ with respect to D_Φ and hence it is a Cauchy sequence for any $k \in \{1, \ldots, n\}$. For the finiteness of set Φ_ε, we can find an index $\bar{\jmath}$ such that

$$D_\Phi(\varphi_k \circ g_{i_r}, \varphi_k \circ g_{i_s}) < \varepsilon, \text{ for every } s, r \geq \bar{\jmath}. \tag{29}$$

In order to prove that (g_{i_j}) is a Cauchy sequence, we observe that for any $\varphi \in \Phi$, any $\varphi_k \in \Phi_\varepsilon$, and any $r, s \in \mathbb{N}$ we have

$$\begin{aligned} &D_\Phi(\varphi \circ g_{i_r}, \varphi \circ g_{i_s})) \\ &\leq D_\Phi(\varphi \circ g_{i_r}, \varphi_k \circ g_{i_r}) + D_\Phi(\varphi_k \circ g_{i_r}, \varphi_k \circ g_{i_s}) + D_\Phi(\varphi_k \circ g_{i_s}, \varphi \circ g_{i_s}) \\ &= D_\Phi(\varphi, \varphi_k) + D_\Phi(\varphi_k \circ g_{i_r}, \varphi_k \circ g_{i_s}) + D_\Phi(\varphi_k, \varphi). \end{aligned} \tag{30}$$

We observe that $\bar{\jmath}$ does not depend on φ, but only on ε and Φ_ε. By choosing a $\varphi_k \in \Phi_\varepsilon$ such that $D_\Phi(\varphi_k, \varphi) < \varepsilon$, we get $D_\Phi(\varphi \circ g_{i_r}, \varphi \circ g_{i_s})) < \varepsilon$ for every $\varphi \in \Phi$ and every $r, s \geq \bar{\jmath}$. Thus, $D_G(g_{i_r}, g_{i_s}) < 3\varepsilon$. Hence, the sequence (g_{i_j}) is a Cauchy sequence. Finally, given that G is complete, (g_{i_j}) is convergent. Therefore, G is sequentially compact. □

Example. Let Φ be the set containing all the 1-Lipschitz functions from $X = S^1 = \{(x, y) \in \mathbb{R}^3 : x^2 + y^2 = 1\}$ to $[0, 1]$, and G be the group of all rotations $\rho_{2\pi q}$ of X of $2\pi q$ radians with q rational number. The space (G, D_G) is neither complete nor compact.

Proposition (4.12)**.** *If F is a GENEO from (Φ, G) to (Ψ, H) associated with $T : G \to H$, then it is a contraction with respect to the natural pseudo-distances d_G, d_H.*

Proof. Since F is a GENEO, it follows that

$$\begin{aligned} d_H(F(\varphi_1), F(\varphi_2)) &= \inf_{h\in H} D_\Psi\left(F(\varphi_1), F(\varphi_2)\circ h\right) \\ &\le \inf_{g\in G} D_\Psi\left(F(\varphi_1), F(\varphi_2)\circ T(g)\right) \\ &= \inf_{g\in G} D_\Psi\left(F(\varphi_1), F(\varphi_2\circ g)\right) \\ &\le \inf_{g\in G} D_\Phi\left(\varphi_1, \varphi_2\circ g\right) = d_G(\varphi_1,\varphi_2). \end{aligned} \tag{31}$$

□

Proposition (4.15)**.** *For every $F \in \mathcal{F}^{\mathrm{all}}$ and every $\varphi \in \Phi$: $\|F(\varphi)\|_\infty \le \|\varphi\|_\infty + \|F(\boldsymbol{0})\|_\infty$, where $\boldsymbol{0}$ denotes the function taking the value 0 everywhere.*

Proof. Since F is non-expansive, we have that

$$\begin{aligned} \|F(\varphi)\|_\infty &= \|F(\varphi) - F(\mathbf{0}) + F(\mathbf{0})\|_\infty \\ &\le \|F(\varphi) - F(\mathbf{0})\|_\infty + \|F(\mathbf{0})\|_\infty \\ &\le \|\varphi - \mathbf{0}\|_\infty + \|F(\mathbf{0})\|_\infty = \|\varphi\|_\infty + \|F(\mathbf{0})\|_\infty \end{aligned}$$

□

Proposition. *$\mathcal{D}^{\mathcal{F},k}_{\mathrm{match}}$ is a strongly G-invariant pseudo-metric on Φ.*

Proof. Theorem 5.1 and the non-expansivity of every $F \in \mathcal{F}$ imply that

$$\begin{aligned} d_{\mathrm{match}}(r_k(F(\varphi_1)), r_k(F(\varphi_2))) &\le D_\Psi\left(F(\varphi_1), F(\varphi_2)\right) \\ &\le D_\Phi\left(\varphi_1, \varphi_2\right). \end{aligned}$$

Therefore $\mathcal{D}^{\mathcal{F},k}_{\mathrm{match}}$ is a pseudo-metric, since it is the supremum of a family of pseudo-metrics that are bounded at each pair (φ_1, φ_2). Moreover, for every $\varphi_1, \varphi_2 \in \Phi$ and every $g \in G$

$$\begin{aligned} \mathcal{D}^{\mathcal{F},k}_{\mathrm{match}}(\varphi_1, \varphi_2\circ g) &:= \sup_{F\in\mathcal{F}} d_{\mathrm{match}}(r_k(F(\varphi_1)), r_k(F(\varphi_2\circ g))) \\ &= \sup_{F\in\mathcal{F}} d_{\mathrm{match}}(r_k(F(\varphi_1)), r_k(F(\varphi_2)\circ T(g))) \\ &= \sup_{F\in\mathcal{F}} d_{\mathrm{match}}(r_k(F(\varphi_1)), r_k(F(\varphi_2)) \\ &= \mathcal{D}^{\mathcal{F},k}_{\mathrm{match}}(\varphi_1, \varphi_2) \end{aligned}$$

because of the equality $F(\varphi\circ g) = F(\varphi)\circ T(g)$ for every $\varphi \in \Phi$ and every $g \in G$ and the invariance of persistent homology under the action of the homeomorphisms. Since the function $\mathcal{D}^{\mathcal{F},k}_{\mathrm{match}}$ is symmetric, this is sufficient to guarantee that $\mathcal{D}^{\mathcal{F},k}_{\mathrm{match}}$ is strongly G-invariant. □

Theorem (5.5)**.** *If $\mathcal{F}$ is a non-empty subset of $\mathcal{F}^{\text{all}}$, then*

$$\mathcal{D}_{\text{match}}^{\mathcal{F},k} \leq d_G \leq D_{\Phi} \tag{32}$$

Proof. For every $F \in \mathcal{D}_{\text{match}}^{\mathcal{F},k}$, every $g \in G$ and every $\varphi_1, \varphi_2 \in \Phi$, we have that

$$\begin{aligned} d_{\text{match}}(r_k(F(\varphi_1)), r_k(F(\varphi_2))) &= d_{\text{match}}(r_k(F(\varphi_1)), r_k(F(\varphi_2) \circ T(g))) \\ &= d_{\text{match}}(r_k(F(\varphi_1)), r_k(F(\varphi_2 \circ g))) \\ &\leq D_{\Psi}\left(F(\varphi_1), F(\varphi_2 \circ g)\right) \leq D_{\Phi}\left(\varphi_1, \varphi_2 \circ g\right). \end{aligned}$$

The first equality follows from the invariance of persistent homology under action of Homeo(X) (see Remark 5.4), and the second equality follows from the fact F is a group equivariant operator. The first inequality follows from the stability of persistent homology (Theorem 5.1), while the second inequality follows from the non-expansivity of F. It follows that, if $\mathcal{F} \subseteq \mathcal{F}^{\text{all}}$, then for every $g \in G$ and every $\varphi_1, \varphi_2 \in \Phi$

$$\mathcal{D}_{\text{match}}^{\mathcal{F},k}(\varphi_1, \varphi_2) \leq D_{\Phi}\left(\varphi_1, \varphi_2 \circ g\right). \tag{33}$$

Hence, the inequality $\mathcal{D}_{\text{match}}^{\mathcal{F},k} \leq d_G$ follows, while $d_G \leq D_{\Phi}$ is stated in Theorem 5.1. □

Theorem (5.6)**.** *Let us assume that $\Phi = \Psi$, every function in Φ is non-negative, the k-th Betti number of X does not vanish, and Φ contains each constant function c for which a function $\varphi \in \Phi$ exists such that $0 \leq c \leq \|\varphi\|_\infty$. Then $\mathcal{D}_{\text{match}}^{\mathcal{F}^{\text{all}},k} = d_G$.*

Proof. For every $\varphi' \in \Phi$ let us consider the operator $F_{\varphi'} : \Phi \to \Phi$ defined by setting $F_{\varphi'}(\varphi)$ equal to the constant function taking everywhere the value $d_G(\varphi, \varphi')$ for every $\varphi \in \Phi$ (i.e., $F_{\varphi'}(\varphi)(x) = d_G(\varphi, \varphi')$ for any $x \in X$). Our assumptions guarantee that such a constant function belongs to $\Phi = \Psi$. We also set $T = id : G \to G$.

We observe that

1. $F_{\varphi'}$ is a group equivariant operator on Φ, because the strong invariance of the natural pseudo-distance d_G with respect to the group G (Remark 5.2) implies that if $\varphi \in \Phi$ and $g \in G$, then $F_{\varphi'}(\varphi \circ g)(x) = d_G(\varphi \circ g, \varphi') = F_{\varphi'}(\varphi)(g(x)) = (F_{\varphi'}(\varphi) \circ g)(x) = (F_{\varphi'}(\varphi) \circ T(g))(x)$, for every $x \in X$.
2. $F_{\varphi'}$ is non-expansive on Φ, because for every $\varphi_1, \varphi_2 \in \Phi$

$$\begin{aligned} D_{\Psi}\left(F_{\varphi'}(\varphi_1), F_{\varphi'}(\varphi_2)\right) &= |d_G(\varphi_1, \varphi') - d_G(\varphi_2, \varphi')| \\ &\leq d_G(\varphi_1, \varphi_2) \leq D_{\Phi}\left(\varphi_1, \varphi_2\right). \end{aligned}$$

Therefore, $F_{\varphi'}$ is a GENEO.

For every $\varphi_1, \varphi_2, \varphi' \in \Phi$ we have that

$$d_{\text{match}}(r_k(F_{\varphi'}(\varphi_1)), r_k(F_{\varphi'}(\varphi_2))) = |d_G(\varphi_1, \varphi') - d_G(\varphi_2, \varphi')|. \tag{34}$$

Indeed, apart from the trivial points on the line $\{(u, v) \in \mathbb{R}^2 \ : \ u = v\}$, the persistence diagram associated with $r_k(F_{\varphi'}(\varphi_1))$ contains only the point $(d_G(\varphi_1, \varphi'), \infty)$, while the persistence diagram associated with $r_k(F_{\varphi'}(\varphi_2))$ contains only the point $(d_G(\varphi_2, \varphi'), \infty)$. Both the points have the same multiplicity, which equals the (non-null) k-th Betti number of X.

Setting $\varphi' = \varphi_2$, we have that

$$d_{\text{match}}(r_k(F_{\varphi'}(\varphi_1)), r_k(F_{\varphi'}(\varphi_2))) = d_G(\varphi_1, \varphi_2). \tag{35}$$

As a consequence, we have that

$$\mathcal{D}_{\text{match}}^{\mathcal{F}^{\text{all}},k}(\varphi_1, \varphi_2) \geq d_G(\varphi_1, \varphi_2). \tag{36}$$

By applying Theorem 5.5, we get

$$\mathcal{D}_{\text{match}}^{\mathcal{F}^{\text{all}},k}(\varphi_1, \varphi_2) = d_G(\varphi_1, \varphi_2) \tag{37}$$

for every φ_1, φ_2. □

Proposition (5.7). *Let $\mathcal{F}, \mathcal{F}' \subseteq \mathcal{F}^{\text{all}}$. If the Hausdorff distance*

$$HD(\mathcal{F}, \mathcal{F}') := \max \left\{ \sup_{F \in \mathcal{F}} \inf_{F' \in \mathcal{F}'} D_{GENEO,H}(F, F'), \sup_{F' \in \mathcal{F}'} \inf_{F \in \mathcal{F}} D_{GENEO,H}(F, F') \right\}$$

is not larger than ε, then

$$\left| \mathcal{D}_{\text{match}}^{\mathcal{F},k}(\varphi_1, \varphi_2) - \mathcal{D}_{\text{match}}^{\mathcal{F}',k}(\varphi_1, \varphi_2) \right| \leq 2\varepsilon \tag{38}$$

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

Proof. Since $HD(\mathcal{F}, \mathcal{F}') \leq \varepsilon$, for every $F \in \mathcal{F}$ a $F' \in \mathcal{F}'$ and an $\eta > 0$ exist such that $D_{GENEO,H}(F, F') \leq \varepsilon + \eta$. The definition of $D_{GENEO,H}$ implies that $d_H(F(\varphi), F'(\varphi)) \leq \varepsilon + \eta$ for every $\varphi \in \Phi$. From Theorem 5.1 it follows that

$$d_{\text{match}}(r_k(F(\varphi_1)), r_k(F'(\varphi_1)) \leq \varepsilon + \eta \tag{39}$$

and

$$d_{\text{match}}(r_k(F(\varphi_2)), r_k(F'(\varphi_2)) \leq \varepsilon + \eta \tag{40}$$

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

Therefore,

$$|d_{\text{match}}(r_k(F(\varphi_1)), r_k(F(\varphi_2)) - d_{\text{match}}(r_k(F'(\varphi_1)), r_k(F'(\varphi_2))| \leq 2(\varepsilon + \eta). \tag{41}$$

As a consequence, $\mathcal{D}_{\text{match}}^{\mathcal{F},k}(\varphi_1, \varphi_2) \leq \mathcal{D}_{\text{match}}^{\mathcal{F}',k}(\varphi_1, \varphi_2) + 2(\varepsilon + \eta)$. We can show analogously that $\mathcal{D}_{\text{match}}^{\mathcal{F}',k}(\varphi_1, \varphi_2) \leq \mathcal{D}_{\text{match}}^{\mathcal{F},k}(\varphi_1, \varphi_2) + 2(\varepsilon + \eta)$. Since η can be chosen arbitrarily small, from the previous two inequalities the proof of our statement follows. □

Theorem (6.1). *$\mathcal{F}^{\text{all}}$ is compact with respect to D_{GENEO}.*

Proof. We know that $(\mathcal{F}^{\text{all}}, D_{GENEO})$ is a metric space. Therefore it will suffice to prove that $\mathcal{F}^{\text{all}}$ is sequentially compact. In order to do this, let us assume that a sequence (F_i) in $\mathcal{F}^{\text{all}}$ is given. Given that Φ is a compact (and hence separable) metric space, we can find a countable and dense subset $\Phi^* = \{\varphi_j\}_{j\in\mathbb{N}}$ of Φ. By means of a diagonalization process, we can extract a subsequence (F'_i) from (F_i), such that for every fixed index j the sequence $(F'_i(\varphi_j))$ converges to a function in Ψ with respect to D_Ψ. Now, let us consider the function $\bar{F} : \Phi \to \Psi$ defined by setting $\bar{F}(\varphi_j) := \lim_{i\to\infty} F'_i(\varphi_j)$ for each $\varphi_j \in \Phi^*$.

We extend $\bar{F}$ to Φ as follows. For every $\varphi \in \Phi$ we choose a sequence (φ_{j_r}) in Φ^*, converging to $\varphi \in \Phi$, and set $\bar{F}(\varphi) := \lim_{r\to\infty} \bar{F}(\varphi_{j_r})$. We claim that such a limit exists in Ψ and does not depend on the sequence that we have chosen, converging to $\varphi \in \Phi$. In order to prove that the previous limit exists, we observe that for every $r,\ s \in \mathbb{N}$

$$\begin{aligned} D_\Psi\left(\bar{F}(\varphi_{j_r}), \bar{F}(\varphi_{j_s})\right) &= D_\Psi\left(\lim_{i\to\infty} F'_i(\varphi_{j_r}), \lim_{i\to\infty} F'_i(\varphi_{j_s})\right) \\ &= \lim_{i\to\infty} D_\Psi\left(F'_i(\varphi_{j_r}), F'_i(\varphi_{j_s})\right) \\ &\le \lim_{i\to\infty} D_\Phi\left(\varphi_{j_r}, \varphi_{j_s}\right) = D_\Phi\left(\varphi_{j_r}, \varphi_{j_s}\right), \end{aligned}$$

because each F'_i is non-expansive.

Since the sequence (φ_{j_r}) converges to $\varphi \in \Phi$, it follows that $(\bar{F}(\varphi_{j_r}))$ is a Cauchy sequence with respect to D_Ψ. The compactness of Ψ implies that $(\bar{F}(\varphi_{j_r}))$ converges in Ψ.
If another sequence (φ_{k_r}) in given in Φ^*, converging to $\varphi \in \Phi$, then for every index $r \in \mathbb{N}$

$$\begin{aligned} D_\Psi\left(\bar{F}(\varphi_{j_r}), \bar{F}(\varphi_{k_r})\right) &= D_\Psi\left(\lim_{i\to\infty} F'_i(\varphi_{j_r}), \lim_{i\to\infty} F'_i(\varphi_{k_r})\right) \\ &= \lim_{i\to\infty} D_\Psi\left(F'_i(\varphi_{j_r}), F'_i(\varphi_{k_r})\right) \\ &\le \lim_{i\to\infty} D_\Phi\left(\varphi_{j_r}, \varphi_{k_r}\right) \\ &= D_\Phi\left(\varphi_{j_r}, \varphi_{k_r}\right). \end{aligned}$$

Since both (φ_{j_r}) and (φ_{k_r}) converge to φ it follows that $\lim_{r\to\infty} \bar{F}(\varphi_{j_r}) = \lim_{r\to\infty} \bar{F}(\varphi_{k_r})$. Therefore the definition of $\bar{F}(\varphi)$ does not depend on the sequence (φ_{j_r}) that we have chosen, converging to φ.

Now we have to prove that $\bar{F} \in \mathcal{F}^{\text{all}}$, i.e., that $\bar{F}$ verifies the properties defining this set of operators. We have already seen that $\bar{F} : \Phi \to \Psi$.

For every φ, φ' we can consider two sequences (φ_{j_r}), (φ_{k_r}) in Φ^*, converging to φ and φ', respectively. Due to the fact that the operators F'_i are non-expansive, we have that

$$\begin{aligned}
D_\Psi\left(\bar{F}(\varphi), \bar{F}(\varphi')\right) &= D_\Psi\left(\lim_{r\to\infty} \bar{F}(\varphi_{j_r}), \lim_{r\to\infty} \bar{F}(\varphi_{k_r})\right) \\
&= D_\Psi\left(\lim_{r\to\infty}\lim_{i\to\infty} F'_i(\varphi_{j_r}), \lim_{r\to\infty}\lim_{i\to\infty} F'_i(\varphi_{k_r})\right) \\
&= \lim_{r\to\infty}\lim_{i\to\infty} D_\Psi\left(F'_i(\varphi_{j_r}), F'_i(\varphi_{k_r})\right) \\
&\le \lim_{r\to\infty}\lim_{i\to\infty} D_\Phi\left(\varphi_{j_r}, \varphi_{k_r}\right) \\
&= \lim_{r\to\infty} D_\Phi\left(\varphi_{j_r}, \varphi_{k_r}\right) \\
&= D_\Phi\left(\varphi, \varphi'\right).
\end{aligned}$$

Therefore, $\bar{F} : \Phi \to \Psi$ is non-expansive. As a consequence, it is also continuous.

We can now prove that the sequence (F'_i) converges to $\bar{F}$ with respect to D_{GENEO}.

Let us consider an arbitrarily small $\varepsilon > 0$. Since Φ is compact and Φ^* is dense in Φ, we can find a finite subset $\{\varphi_{j_1}, \ldots, \varphi_{j_n}\}$ of Φ^* such that for every $\varphi \in \Phi$, there exists an index $r \in \{1, \ldots, n\}$, for which $D_\Phi\left(\varphi, \varphi_{j_r}\right) < \varepsilon$.

Since the sequence (F'_i) converges pointwise to $\bar{F}$ on the set Φ^*, an index $\bar{\imath}$ exists, such that $D_\Psi\left(\bar{F}(\varphi_{j_r}), F'_i(\varphi_{j_r})\right) < \varepsilon$ for any $i \ge \bar{\imath}$ and any $r \in \{1, \ldots, n\}$. Therefore, for every $\varphi \in \Phi$ we can find an index $r \in \{1, \ldots, n\}$ such that $D_\Phi\left(\varphi, \varphi_{j_r}\right) < \varepsilon$ and the following inequalities hold for every index $i \ge \bar{\imath}$, because of the non-expansivity of $\bar{F}$ and F'_i:

$$\begin{aligned}
D_\Psi\left(\bar{F}(\varphi), F'_i(\varphi)\right) &\le D_\Psi\left(\bar{F}(\varphi), \bar{F}(\varphi_{j_r})\right) + D_\Psi\left(\bar{F}(\varphi_{j_r}), F'_i(\varphi_{j_r})\right) + D_\Psi\left(F'_i(\varphi_{j_r}), F'_i(\varphi)\right) \\
&\le D_\Phi\left(\varphi, \varphi_{j_r}\right) + D_\Psi\left(\bar{F}(\varphi_{j_r}), F'_i(\varphi_{j_r})\right) + D_\Phi\left(\varphi_{j_r}, \varphi\right) < 3\varepsilon.
\end{aligned}$$

We observe that $\bar{\imath}$ does not depend on φ, but only on ε and on the set $\{\varphi_{j_1}, \ldots, \varphi_{j_n}\}$. It follows that $D_\Psi\left(\bar{F}(\varphi), F'_i(\varphi)\right) < 3\varepsilon$ for every $\varphi \in \Phi$ and every $i \ge \bar{\imath}$.

Hence, $\sup_{\varphi\in\Phi} D_\Psi\left(\bar{F}(\varphi), F'_i(\varphi)\right) \le 3\varepsilon$ for every $i \ge \bar{\imath}$. Therefore, the sequence (F'_i) converges to $\bar{F}$ with respect to D_{GENEO}.

The last thing that we have to show is that $\bar{F}$ is group equivariant. Let us consider a $\varphi \in \Phi$, a sequence (φ_{j_r}) in Φ^* converging to φ in Φ and a $g \in G$. Obviously, $D_\Phi(\varphi_{j_r} \circ g, \varphi \circ g) = D_\Phi(\varphi_{j_r}, \varphi)$ and hence the sequence $(\varphi_{j_r} \circ g)$ converges to $\varphi \circ g$ in Φ with respect to D_Φ. We recall that the right action of G on Φ is continuous, $\bar{F}$ is continuous and each F'_i is group equivariant. Hence, given that the sequence (F'_i) converges to $\bar{F}$ with respect to D_{GENEO}, the following equalities hold:

$$\begin{aligned}
\bar{F}(\varphi \circ g) &= \bar{F}(\lim_{r\to\infty} (\varphi_{j_r} \circ g)) \\
&= \lim_{r\to\infty} \bar{F}(\varphi_{j_r} \circ g) \\
&= \lim_{r\to\infty}\lim_{i\to\infty} F'_i(\varphi_{j_r} \circ g) \\
&= \lim_{r\to\infty}\lim_{i\to\infty} F'_i(\varphi_{j_r}) \circ T(g) \\
&= \lim_{r\to\infty} \bar{F}(\varphi_{j_r}) \circ T(g) \\
&= \bar{F}(\varphi) \circ T(g).
\end{aligned}$$

This proves that $\bar{F}$ is group equivariant, and hence a perception map. In conclusion, $\bar{F}$ is a GENEO. From the fact that the sequence F'_i converges to $\bar{F}$ with respect to D_{GENEO}, it follows that $(\mathcal{F}^{\text{all}}, D_{GENEO})$ is sequentially compact. □

Corollary (6.2)**.** *Let $\mathcal{F}$ be a non-empty subset of $\mathcal{F}^{\text{all}}$. For every $\varepsilon > 0$, a finite subset $\mathcal{F}^*$ of $\mathcal{F}$ exists, such that*

$$|\mathcal{D}_{\text{match}}^{\mathcal{F}^*,k}(\varphi_1, \varphi_2) - \mathcal{D}_{\text{match}}^{\mathcal{F},k}(\varphi_1, \varphi_2)| \leq \varepsilon \tag{42}$$

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

Proof. Let us consider the closure $\bar{\mathcal{F}}$ of $\mathcal{F}$ in $\mathcal{F}^{\text{all}}$. Let us also consider the covering $\mathcal{U}$ of $\bar{\mathcal{F}}$ obtained by taking all the open balls of radius $\frac{\varepsilon}{2}$ centered at points of $\mathcal{F}$, with respect to D_{GENEO}. Theorem 6.1 guarantees that $\mathcal{F}^{\text{all}}$ is compact, hence also $\bar{\mathcal{F}}$ is compact. Therefore we can extract a finite covering $\{B_1, \ldots, B_m\}$ of $\bar{\mathcal{F}}$ from $\mathcal{U}$. We can set $\mathcal{F}^*$ equal to the set of centers of the balls $B_1, \ldots, B_m$. The statement of our corollary immediately follows from Proposition 5.7, by recalling that $D_{GENEO,H} \leq D_{GENEO}$ and hence $HD\left(\bar{\mathcal{F}}, \mathcal{F}^*\right) \leq \varepsilon/2$. □

Proposition (6.4)**.** *If $F_\Sigma(\Phi) \subseteq \Psi$, then F_Σ is a GENEO from (Φ, G) to (Ψ, H) with respect to T.*

Proof. First we prove that F_Σ is a perception map with respect to T. Since every F_i is a perception map we have that:

$$F_\Sigma(\varphi \circ g) = \sum_{i=1}^{n} a_i F_i(\varphi \circ g) = \sum_{i=1}^{n} a_i (F_i(\varphi) \circ T(g)) = \sum_{i=1}^{n} (a_i F_i(\varphi)) \circ T(g) = F_\Sigma(\varphi) \circ T(g). \tag{43}$$

Since every F_i is non-expansive, F_Σ is non-expansive:

$$\begin{aligned} D_\Psi\left(F_\Sigma(\varphi_1), F_\Sigma(\varphi_2)\right) &= \left\| \sum_{i=1}^{n} a_i F_i(\varphi_1) - \sum_{i=1}^{n} a_i F_i(\varphi_2) \right\|_\infty & (44) \\ &= \left\| \sum_{i=1}^{n} a_i (F_i(\varphi_1) - F_i(\varphi_2)) \right\|_\infty & (45) \\ &\leq \sum_{i=1}^{n} |a_i| \, \|(F_i(\varphi_1) - F_i(\varphi_2))\|_\infty & (46) \\ &\leq \sum_{i=1}^{n} |a_i| \, \|\varphi_1 - \varphi_2\|_\infty \leq D_\Phi\left(\varphi_1, \varphi_2\right). & (47) \end{aligned}$$

Therefore F_Σ is a GENEO. □

Theorem (6.5)**.** *If Ψ is convex, then the set of GENEOs from (Φ, G) to (Ψ, H) with respect to T is convex.*

Proof. It is sufficient to apply Proposition 6.4 for $n = 2$, by setting $a_1 = t$, $a_2 = 1 - t$ for $0 \leq t \leq 1$, and observing that the convexity of Ψ implies $F_\Sigma(\Phi) \subseteq \Psi$. □

References

[1] Y. LeCun, Y. Bengio, et al., Convolutional networks for images, speech, and time series, The handbook of brain theory and neural networks 3361 (10) (1995) 1995.

[2] F. Anselmi, L. Rosasco, T. Poggio, On invariance and selectivity in representation learning, Information and Inference: A Journal of the IMA 5 (2) (2016) 134–158. arXiv:/oup/backfile/content_public/journal/imaiai/5/2/10.1093_imaiai_iaw009/2/iaw009.pdf, doi:10.1093/imaiai/iaw009.
URL http://dx.doi.org/10.1093/imaiai/iaw009

[3] T. Cohen, M. Welling, Group equivariant convolutional networks, in: International conference on machine learning, 2016, pp. 2990–2999.

[4] D. E. Worrall, S. J. Garbin, D. Turmukhambetov, G. J. Brostow, Harmonic networks: Deep translation and rotation equivariance, in: Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Vol. 2, 2017.

[5] P. Frosini, Towards an Observer-oriented Theory of Shape Comparison, in: A. Ferreira, A. Giachetti, D. Giorgi (Eds.), Eurographics Workshop on 3D Object Retrieval, The Eurographics Association, 2016. doi:10.2312/3dor.20161080.

[6] 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 Comput. Surv. 40 (4) (2008) 12:1–12:87. doi:10.1145/1391729.1391731.
URL http://doi.acm.org/10.1145/1391729.1391731

[7] G. Carlsson, A. Zomorodian, The theory of multidimensional persistence, Discrete Comput. Geom. 42 (1) (2009) 71–93. doi:10.1007/s00454-009-9176-0.
URL http://dx.doi.org/10.1007/s00454-009-9176-0

[8] H. Edelsbrunner, J. Harer, Persistent homology—a survey, in: Surveys on discrete and computational geometry, Vol. 453 of Contemp. Math., Amer. Math. Soc., Providence, RI, 2008, pp. 257–282. doi:10.1090/conm/453/08802.
URL http://dx.doi.org/10.1090/conm/453/08802

[9] D. Cohen-Steiner, H. Edelsbrunner, J. Harer, Stability of persistence diagrams, Discrete Comput. Geom. 37 (1) (2007) 103–120. doi:10.1007/s00454-006-1276-5.
URL http://dx.doi.org/10.1007/s00454-006-1276-5

[10] A. Cerri, B. Di Fabio, M. Ferri, P. Frosini, C. Landi, Betti numbers in multidimensional persistent homology are stable functions, Math. Methods Appl. Sci. 36 (12) (2013) 1543–1557. doi:10.1002/mma.2704.
URL http://dx.doi.org/10.1002/mma.2704

[11] S. A. Gaal, Point set topology, Pure and Applied Mathematics, Vol. XVI, Academic Press, New York-London, 1964.

[12] P. Frosini, G. Jabłoński, Combining persistent homology and invariance groups for shape comparison, Discrete Comput. Geom. 55 (2) (2016) 373–409. doi:10.1007/s00454-016-9761-y.
URL http://dx.doi.org/10.1007/s00454-016-9761-y

[13] S. Y. Oudot, Persistence theory: from quiver representations to data analysis, Vol. 209 of Mathematical Surveys and Monographs, American Mathematical Society, Providence, RI, 2015. doi:10.1090/surv/209.
URL https://doi.org/10.1090/surv/209

[14] X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in: Proceedings of the thirteenth international conference on artificial intelligence and statistics, 2010, pp. 249–256.

[15] G. E. Hinton, A. Krizhevsky, S. D. Wang, Transforming auto-encoders, in: International Conference on Artificial Neural Networks, Springer, 2011, pp. 44–51.

[16] S. Sabour, N. Frosst, G. E. Hinton, Dynamic routing between capsules, in: I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems 30, Curran Associates, Inc., 2017, pp. 3856–3866.
URL http://papers.nips.cc/paper/6975-dynamic-routing-between-capsules.pdf