Università di Pisa

Corso di Dottorato in Matematica
XXXVI ciclo

Tesi di Dottorato

A bridge between persistent homology and group equivariant non-expansive operators: theory and applications

Candidato: Francesco Conti

Coordinatore Dottorato:
Roberto Frigerio

Supervisore:
Davide Moroni

Co-Supervisore:
Patrizio Frosini

Co-Supervisore:
Maria Antonietta Pascali

Esame finale anno 2024

Abstract

Topological Data Analysis (TDA) is proving to be an excellent tool for shape analysis of digital data. The recently found synergy with artificial intelligence gave rise to Topological Machine Learning (TML), which aims to combine the expressive power of computational topology with the accuracy of machine learning to provide a comprehensive and automatic framework for data classification. The aim of this thesis is twofold: to develop current applications of TML in practical scenarios, with emphasis on the most overlooked aspects of its pipeline, and to connect the theory of TDA with a broader class of maps, the Group Equivariant Non-Expansive Operators (GENEOs). In the first part of this dissertation, we develop a pipeline to study digital data by means of TML in order to validate the practical aspects of our theory. We apply this pipeline to benchmark and experimental datasets, achieving state-of-the-art accuracies in biomedical scenarios. Moreover, we perform an empirical but extensive study of the stability of features arising from the various homological dimensions with respect to noise and points distribution in the persistence diagram. Such a comparison is novel in the TML literature and our findings show that results coming from the concatenation of each homological dimension available are the best approach in the vectorisation step. We later expand on the main concept of TDA, proving that the functor that computes persistence diagrams can be seen as a particular instance of GENEOs (Theorem 4.1.4). The GENEO framework allows us to inject arbitrary equivariances in a machine learning setting and represents a new possible approach to neural network architecture. Next, we fully present the theory of GENEOs and their properties, such as convexity and concavity, under suitable assumptions. This thesis expand the GENEO theory with two new tools to define such operators, namely using symmetric functions (Theorem 5.3.24) and a characterization theorem of linear GENEOs between arbitrary functional spaces (Theorem 6.2.2). Finally, we develop a new neural network architecture with GENEOs instead of neurons and show its potential in a couple of applications.

Contents

Introduction

The goal of this thesis is to enhance the connection between Topological Data Analysis (TDA) and the theory of Group Equivariant Non-Expansive Operators (GENEOs), as well as develop the theory and applications of both concepts. Topological data analysis computes descriptors of data to encode its shape in a more manageable and useful object called persistence diagram. The first part of the thesis focuses on TDA applications and theory. In particular, we provide a novel attempt to standardise a fragmented part of the TDA pipeline, with empirical evidence to support our claim. Moreover, we perform three new case studies of real-world datasets using a TDA pipeline, achieving state-of-the-art accuracy results. In the second part of the thesis we generalize the theory of TDA, proving that the operator that produces persistence diagrams belongs to a family of operators that we call GENEOs. Furthermore, we develop the theory of GENEOs with two new methodologies to build them and we provide a few applications of this new framework. The two parts of the dissertation rely on different but related mathematical concepts. In this dissertation we further link them with a proof that the mathematical core of TDA can be seen as an element of the GENEO space.

In the last decades, the need to analyze and extract meaningful information from a large quantity of data is becoming fundamental in many aspects of scientific research and beyond. Deep learning methods achieve state-of-the-art performances in a huge variety of real-world tasks. Part of their success is due to the fact that raw data are sufficient, if not better, than hand-crafted features for learning a specific task. Despite their extreme effectiveness, however, little effort has been made to standardise the theory behind neural networks. Furthermore, as deep network applications grow in complexity, so do their architectures and we have reached a point where their architectures are often as task-specific as the hand-crafted features that they intended to replace. In data analysis, Topological Data Analysis (TDA) [1] is establishing as one of the most prominent lines of research since it allows to exploit symmetries and invariance of data, overcoming the infamous course of dimensionality [2], of which many deep learning methods suffer. Moreover, topological data analysis produces low dimensional models required in the so-called explainable artificial intelligence [3]. Incidentally, a low dimensional model is also less prone to overfitting. TDA allows to extract powerful representations of the data shape that are both stable with respect to noise and allow for easy low dimensional interpretations. These features can play an important role in deep learning [4, 5]. The mathematical core of TDA is Persistent Homology (PH), which has been deeply investigated both from a theoretical and an applicative point of view

[6, 7, 8, 9, 10, 11]. On a broader perspective, persistent homology can be seen as a map that transform data into multisets. Moreover, persistent homology is invariant with respect to homeomorphisms of data by construction. Defining a new mathematical model that stems from functional analysis and interpreting data as points in a function space, operators that act on data are in fact maps between functional spaces. In this optic, the computation of persistence diagrams can be seen as a group equivariant non-expansive operator [12]. In general, neural networks (or any intelligent observer) can be seen as an agent acting on data and transforming them in order to better study them. The key idea of this new mathematical framework is to shift the focus from the data to the space of transformations of the data, and the symmetries they are associated with. The goal is therefore to study the geometrical and topological properties of the space of GENEOs, of which PDs are simply an element. Moreover, the group of equivariance can be changed depending on the task and the resulting framework is more flexible than persistent homology, which is equivariant with respect to a fixed group. This model is what we call the theory of GENEOs.

Broadly speaking, persistent homology aims to extract features from data that encode its shape. More importantly, the features extracted represent the data shape not only from a static point of view, that is, from a fixed perspective like many competitors, rather it is interested in the evolution of a class of features arising from algebraic topology. The common assumption is that features that are detectable at multiple scales are the most important in describing the shape of the data. More precisely, persistent homology studies the evolution of the Betti numbers associated with a finite family of nested subcomplexes of the data, and keeps track of their appearances and vanishing. The process that computes such a family of subcomplexes is called filtration, which represent one of the main geometric component of TDA. The collection of pairs (birth, death) of topological invariants, counted with multiplicity, is usually referred to as Persistence Diagram (PD), and it is the feature extracted by persistence homology that we are going to exploit in this dissertation. In particular, each point of the PD is associated to a specific homological dimension. In lower dimensions, such points offer an easy interpretation. In particular, the collection of points in homological dimension zero represents the evolution of the connected components of the data (i.e. the number of 0-dimensional holes plus one). Points in homological dimension one represent the evolution of holes in the data (i.e. 1-dimensional holes) and analogously in higher dimensional counterpart. When endowed with the bottleneck or Wasserstein distance, the most common in literature, the space of PDs lacks fundamental properties to be directly employed in machine learning. Nevertheless, TDA solves this issue with suitable embedding in vector or Hilbert spaces ([13, 14, 15] just to name a few). Such embeddings are called vectorisations. Currently, the features extracted from TDA achieve state-of-the-art performance in many tasks [16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27]. In this field, some limitations still remain. First of all, applications to real-world datasets are still scarce. Moreover, the literature is not consistent on how to exploit the potentiality of the features extracted. We recall that different homological dimensions encode information about different degrees of connection of the shape of the data. In many cases, the various homological dimensions do not carry equal importance in a classification task. In the vectorisation step of the TDA pipeline the literature is fragmented on how to exploit the information provided by such dimensions. In particular, we find examples in bibliography where the homological dimensions deemed less important are entirely discarded in such a step [28, 29, 30, 31, 32, 33]. Other times, each dimension is vectorised and subsequently concatenated altogether [14, 34, 35, 36] (just to name a few). Again, both approaches can be used interchangeably, depending on the task [27, 37, 38]. The implicit assumption is that only the best-performing approach is reported. The first contribution of this dissertation to the theory of TDA lies in a novel attempt to standardise the literature concerning the handling of homological dimensions in the vectorisation step. We provide empirical evidence that the concatenation approach is consistently among the best performing and never experiments a drop of performance, in contrast with all other approaches. In doing so, we develop a topological machine learning pipeline alongside this dissertation that gathers all the major bibliography on TDA and we apply it to benchmark datasets. To further validate our claims, we perform a study on the noise incidence of such approaches. Again, such a study is new in the TDA literature and can provide insights on how to effectively exploit the information provided by the homological dimensions in the vectorisation step. Incidentally, in this study we also show that omitting the homological dimension of points in the persistence diagram represents a viable alternative when the data has well-distributed points in different homological dimensions of comparable cardinality. Finally, the last contribution of this dissertation to the theory of TDA is the application of the aforementioned pipeline to three new real-world datasets, thus expanding the literature of TDA applications. In particular, we performed a case study on upwelling regimes of the Ibaria / Canary current system and two cases studies of biomedical signals for chondrogenic bone cancer grading and Alzheimer disease detection. In both biomedical signals datasets we achieved state-of-the-art accuracies.

Widening the perspective and aiming to provide a more general and flexible mathematical framework to the operators involved in persistent homology, we flow in the theory of group equivariant operators. With equivariance with respect to a group we mean that the operator commutes with the action of the group, and invariance is a special case of equivariance. The invariance of persistent homology with respect to homeomorphisms is well known in literature. Moreover, the stability theorem [39] implies that the operator that maps data to their persistence diagrams is non-expansive. Originating from functional analysis is the theory of group equivariant non-expansive operators [12]. Despite originating from different mathematical branches, the connection between the theory of GENEOs and TDA is actually quite strong, but often overlooked. In particular, there are three major connections between these two concepts, and this dissertation contributes with a fourth. The invariance of persistence homology w.r.t. the group of all homeomorphisms is not requested in numerous scenarios. Moreover, the metric defined on the space of GENEOs is in many cases unfeasible. Broadly speaking, in [12] has been proven a dual synergy between GENEOs and TDA, since GENEOs can provide a metric for TDA that is not homeomorphic-invariant, and TDA can provide an approximation of the metric defined on GENEOs. The third connection already known in literature is an interaction of GENEOs with multiparameter persistent homology [40, 41]. Another contribution of this dissertation is the enhancement of the connection between TDA and GENEOs, with a formal proof that the computation of persistence diagrams is an element of the space of GENEOs from a functorial point of view (Theorem 4.1.4). As already discussed, TDA is revolutionising data analysis. However, this theory comes with certain limitations. Historically, the theory underpinning persistent homology offers a model that is invariant with respect to the group of all homeomorphisms. Despite being equivariant with respect to the largest group of transformations, the fixity of the group of equivariance may represent a constrain too severe in many situations. Furthermore, topology may not always be the best perspective to study data. It is to address all these limitations that the theory of group equivariant non-expansive operators has generated. The idea to employ group equivariant operators in deep learning is not new and its importance is well-known in literature [42, 43, 44, 45]. Yet, it lacks a formalisation. Arguably the leading example of equivariant operators in deep learning is the convolutional neural network [46]. In fact, the convolution operator is equivariant with respect to translations. The equivariance of convolutional neural networks with respect to translations allows both a reduction of the number of parameters of the network, which otherwise could be cumbersome, and also to learn features that can detect patterns in every part of the image by default. This translates both into a speed-up of the learning process and a robustness of the learned features. Convolutional neural networks represent the state-of-the-art in computer vision and part of their success is due to the equivariance with respect to a group that is relevant for the task. Of course, in other tasks equivariance with respect to different transformations is requested to mimic the same concept. Currently, data augmentation or heavy preprocessing are the most common strategies to produce networks resistant to even simple data transformations. It is our belief that formalising the framework for injecting group equivariance in the architecture of neural networks is fundamental in the success of deep learning. The benefits of using equivariant operators in deep learning are twofold. Firstly, this allows to inject pre-existing knowledge in the model. This results in gaining control of the learned features, which are forced to commute with respect to the chosen transformations [47]. Secondly, the equivariance of the model corresponds to exploiting symmetries in the data space. In mathematics, symmetry is almost exclusively expressed by means of group theory. This yields a drastically reduced space to be explored during optimisation, which results in a speed-up learning process and a model more robust to overfitting and noise. Moreover, such a mathematical framework offers an easily interpretable and transparent model. Our mathematical model focuses on group equivariant non-expansive operators. The additional request of non-expansivity, prerogative of our model, is motivated by two aspects. One is epistemological: the goal of an intelligent operator is to find representations of the data that compress the original information. Of course, there are scenarios in which this request is not locally satisfied, but the long-term goal is to reduce the information to extract only the relevant features. The second aspect is purely mathematical. The non-expansivity of such operators is fundamental in providing important mathematical properties such as the compactness of the GENEOs space, under suitable hypothesis. Such a property is extremely helpful in applications, since it guarantees to approximate the space with just a finite set. For more information on such a theory, we refer the reader to [12, 26, 48, 49, 50, 51, 52, 53, 54, 55, 56]. Our mathematical model is based on some epistemological assumptions. First of all, data are represented as functions defined on topological spaces. This is due to the fact that only data that are stable with respect to a certain criterion (i.e. some kind of measurements) can be considered for applications. Hence, stability requires a topological structure. In many real-world applications, data can be represented as $\mathbb{R}^m$-valued continuous functions defined on a topological space X. Simple examples are the coloring of 3D objects, the coordinates of the points in a planar curve, the grey-levels in X-ray images, and many more. For the sake of simplicity, in this dissertation we will restrict to real-valued functions. The second epistemological assumption is that data can not be directly studied. They are only knowable through acts (i.e. measurements or transformations) made by an agent. Hence, there is no absolute way to study data. Rather, the pair (data, agent) is what truly matters. Furthermore, agents are described by means of how they transform data while preserving some kind of invariance. That is, an agent is a group equivariant operator that acts on function spaces. Finally, data similarity depends on the output of the considered agent. In our framework, we switch the analysis from the data to the pair (data, agent). As already stated, an agent can be modeled as a group equivariant operator and we aim to present a good topological theory of the space of such operators. For more details on these assumptions, please refer to [48]. The applicability of this theory is severely restricted by our ability to generate new GENEOs, which is currently very limited. In this line of research, this dissertation contributes with two new methods to define GENEOs by means of symmetric functions (Theorem 5.3.24) and with a characterization theorem for linear GENEOs (Theorem 6.2.2). Moreover, the ability to generate a large number of GENEOs is beneficial both for the increased flexibility of the model, and for the approximation of the GENEO space with just a finite set, which is achievable due to the compactness of such a space. Finally, the last contribution of this dissertation is the design of a GENEO network and its application in two experiments, both of which with excellent results.

We stress the fact that several parts of the dissertation have already been published separately in specific papers. In those cases, we cite the respective works. This dissertation is structured as follows. In Chapter 1 we present the necessary bibliography of current topological machine learning and we formally describe the pipeline devised alongside this dissertation. Such a pipeline is a slightly expanded version of the one presented in [57]. In Chapter 2 we apply the pipeline to a series of benchmark datasets, in order to both validate its effectiveness and to provide some insights on how to handle the homological dimensions in the vectorisation step. In this chapter we define our proposed approach to the vectorisation step, and we validate our findings with a novel noise resistance study of the extracted features. In Chapter 3 we present the applications of the topological machine learning pipeline to real-world datasets coming from temperature maps of the Atlantic Ocean and Raman Spectra derived from biological samples. This chapter is derived from [16, 17, 18]. Chapter 4 represents the pivotal point of the dissertation, providing a strong link between its two main components: TDA and GENEOs. In particular, we prove from a functorial point of view that the operator that produces persistent diagrams belongs to the space of GENEOs. In Chapter 5 we describe the theory of group equivariant non-expansive operators and new ways to build them through the concept of symmetric function and permutant. This chapter is mainly derived from [12, 53, 54]. In Chapter 6 we present a characterization theorem for linear GENEOs between arbitrary finite spaces which revolves around the concept of generalized permutant measure. In Chapter 7 we develop a GENEO network and apply it to benchmark datasets.

The dissertation concludes with a recap of the work here presented and possible future developments in our line of research.

Chapter 1

Topological data analysis meets ML for data classification

In this chapter, we define and explore a promising approach to artificial intelligence and data classification, namely Topological Machine Learning (TML). TML combines Topological Data Analysis (TDA, Section 1.2) with techniques from Machine Learning (ML) in order to study digital data. The resulting model benefits both from the dimensionality reduction coming from Persistent Homology (PH, Section 1.2.1) and its stability with respect to noise and also from the discriminative power of ML algorithms. This chapter is organized as follows. In Sections 1.1 and 1.2 we are going to present the bibliographical background of our mathematical setting. For more information about such topics, we refer the reader to [1, 6, 37, 39, 58, 59, 60]. In Section 1.3 we are going to describe the topological machine learning pipeline devised alongside this dissertation, which has already been presented in [57].

1.1 Algebraic topology

Algebraic topology is a wide branch of mathematic that aims to study the shape of topological spaces. For an in-depth coverage of the subject, we refer to standard literature [6, 58, 61]. The main goal of algebraic topology is to define computable algebraic invariants associated with topological spaces (e.g. manifolds) that persist under homeomorphisms. Arguably the most significant invariant devised by algebraic topology is the **homology** of a topological space. Homology describes in a quantitative and unambiguous fashion how a topological space is connected. There are alternative formalisms to homology to study the general shape of a space. Probably the best known are the curvature of a space [62] or homotopy theory [63]. The main advantage of homology is its (relative) computational efficiency. Loosely speaking, homology captures the presence of holes in a topological space by focusing on what surrounds a hole. In this section, we are going to focus on the concept of simplicial homology, which is the homology of a simplicial complex.

Definition 1.1.1. Given $k \in \mathbb{N}$, a **k-simplex** σ is the convex hull of $k + 1$ affinely independent points $u_0, \dots, u_k \in \mathbb{R}^n$, and we write $\sigma = (u_0, \dots, u_k)$. That is,

$$\sigma = \left\{ \sum_{i=0}^{k} \lambda_i u_i \text{ s.t. } \lambda_i \in \mathbb{R}, \lambda_i \geq 0 \text{ for every } i = 0, \dots, k \text{ and } \sum_{i=0}^{k} \lambda_i = 1 \right\}.$$

Given a k-simplex σ, we call **vertices** of σ each point u_i, for $i = 0, \dots, k$. Given a k-simplex σ, we call τ a **face** of σ if it is a convex hull of a non-empty subset of vertices of σ, and we denote it with $\tau \leq \sigma$. We call a face **proper** if the subset is not the entire set.

Definition 1.1.2. A **simplicial complex** $\mathcal{K}$ is a finite collection of simplices such that if $\sigma \in \mathcal{K}$ and $\tau \leq \sigma$, then $\tau \in \mathcal{K}$ and if $\sigma_1, \sigma_2 \in \mathcal{K}$, then $\sigma_1 \cap \sigma_2$ is either empty or a face of both.

The **mesh** of a simplicial complex is the maximum diameter of its simplices. A subcomplex of a simplicial complex $\mathcal{K}$ is a subset of $\mathcal{K}$ that is still a simplicial complex.

Remark 1.1.3. By definition, simplicial complexes are sets of geometric entities. Given a simplicial complex $\mathcal{K}$, with a slight abuse of notation, we will refer to its geometric realization (i.e. the topological space given by the union of simplices in $\mathcal{K}$) with the same notation.

Definition 1.1.4. A topological space is **triangulable** if it is homeomorphic to a simplicial complex. If a topological space is triangulable, we call a triangulation any homeomorphism to a simplicial complex.

Theorem 1.1.5 [64, 65] *Every smooth manifold admits an (essentially unique) triangulation.*

According to Theorem 1.1.5, we can focus on simplicial complexes without losing generality, as they cover most of the topological spaces that appear in data analysis and machine learning applications. A common assumption in this field is the "manifold hypothesis" [66], which states that data come from an underlying manifold. Although some studies have challenged this hypothesis (see [67]), it remains a crucial premise in our research. See Figure 1.1 for an example of a triangulation of the Möbius strip and the Klein bottle. The first step in order to define the homology groups of a simplicial complex is to introduce the concept of chain complexes.

Definition 1.1.6. An **orientation** of the k-simplex σ is a choice of the ordering of its vertices $u_0, \dots, u_k$. Two orderings define the same orientation if they differ by an even permutation.

It follows directly from the definition of orientation that every simplex can have exactly two orientations. The following theory holds for any field $\mathbb{F}$. To keep the exposition simple and following standard literature on this concept, we will limit to the case $\mathbb{F} = \mathbb{Z}/2\mathbb{Z}$. Similarly, such a theory allows for more general versions with integer dimensions, but in this dissertation we consider only natural numbers. Let $p \in \mathbb{N}$.

Definition 1.1.7. Given a simplicial complex $\mathcal{K}$, a **p-chain** is a formal sum of oriented p-simplices in $\mathcal{K}$. The standard notation is $c = \sum a_i \sigma_i$, where $a_i \in \mathbb{Z}/2\mathbb{Z}$ are the coefficients.

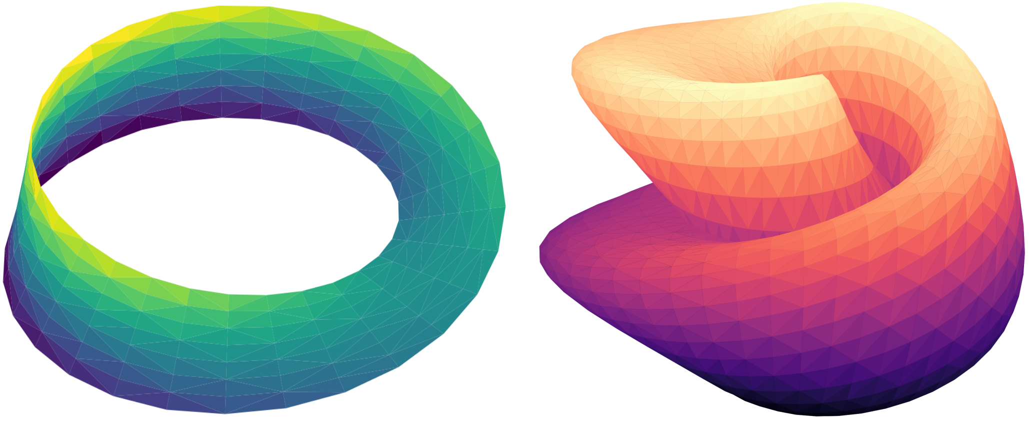


Figure 1.1: Triangulation of the Möbius strip (left) and of the Klein bottle (right).

Since we are working in $\mathbb{Z}/2\mathbb{Z}$, it can be useful to think as a p-chain as the set of p-simplices with $a_i = 1$. We denote with $C_p = C_p\left(\mathcal{K}\right)$ the set of p-chains of the simplicial complex $\mathcal{K}$. On C_p we define the sum operation:

$$+ \colon C_p \times C_p \to C_p, \quad \left(\sum a_i\sigma_i, \sum b_i\sigma_i\right) \mapsto \sum \left(a_i + b_i\right)\sigma_i.$$

The neutral element is $0 = \sum 0\sigma_i$ and the inverse of every element is itself. Since associativity follows from associativity of addition, we have that $(C_p, +)$ is a group, called the group of p-chains. With a slight abuse of notation, we will refer to this group simply with C_p, since the group operation is understood. To connect such groups in different dimensions we define the boundary of p-simplices and p-chains.

Definition 1.1.8. Given a p-simplex $\sigma = (u_0, \dots, u_p)$, the **boundary** of σ, denoted with $\partial_p\sigma$, is the $(p-1)$-chain given by

$$\partial_p\sigma = \sum_{j=0}^{p} \left(u_0, \dots, \hat{u_j}, \dots, u_p\right),$$

where $(u_0, \dots, \hat{u_j}, \dots, u_p)$ is the $(p-1)$-simplex generated by all vertices with the exception of u_j.

Definition 1.1.9. The **boundary** of a p-chain c is the sum of the boundaries of its simplices, that is $\partial_p c = \sum a_i \partial_p \sigma_i$.

The boundary of a p-chain is a $(p-1)$-chain and, since the p-simplices form a basis of C_p, we can write the linear map $\partial_p \colon C_p \to C_{p-1}$. We call this map the boundary map and it is easy to check that it commutes with addition ($\partial_p(c + c') = \partial_p c + \partial_p c'$).

Proposition 1.1.10 *The boundary map $\partial_p \colon C_p \to C_{p-1}$ is a homomorphism, since it commutes with the sum operation.*

The **chain complex** of the simplicial complex $\mathcal{K}$ is the sequence of chain groups connected by the boundary homomorphisms,

$$\dots \xrightarrow{\partial_{p+2}} C_{p+1} \xrightarrow{\partial_{p+1}} C_p \xrightarrow{\partial_p} C_{p-1} \xrightarrow{\partial_{p-1}} \dots$$

A **p-cycle** is a p-chain c with empty boundary, that is $\partial c = 0$. We denote with Z_p the group of p-cycles, which is a subgroup of C_p. We have that $Z_p = \ker \partial_p$. A **p-boundary** is a p-chain c that is the boundary of a $(p+1)$-chain, that is $c = \partial d$ with $d \in C_{p+1}$. We denote with B_p the group of p-boundaries. We have that $B_p = \mathrm{Im}\partial_{p+1}$.

Lemma 1.1.11 (Fundamental lemma of homology [39]) $\partial_p \partial_{p+1} c = 0$ *for every natural* p *and every* $(p+1)$*-chain* c*. That is, every* p*-boundary is a* p*-cycle. Equivalently,* B_p *is a subgroup of* Z_p*.*

Definition 1.1.12. The **p-th homology group** of the simplicial complex $\mathcal{K}$ is the p-th cycle group modulo the p-th boundary group, $H_p := Z_p/B_p$. The **p-th Betti number** is the rank of the group, $\beta_p := \mathrm{rank}\,(H_p)$.

The homology groups and Betti numbers are topological invariants, that is, they are invariant under homeomorphisms. Moreover, Betti numbers provide an intuitive and powerful interpretation: with the exception of β_0, the p-th Betti number counts the number of p-dimensional holes in the topological space. In the case of β_0, it counts the number of connected components. Other types of homology can be defined: see reduced homology, cohomology and others [39]. Their definition and use is beyond the scope of this dissertation, so they are not addressed. For any compact smooth manifold, homology offers a quantitative and computable way to identify it, up to homeomorphisms, with a finite sequence of integers in the form of Betti numbers. In the next section, we are going to expand the concept of homology in the presence of finite metric spaces, such as data, in the form of persistent homology.

1.2 Topological data analysis

Topological data analysis aims to extract qualitative and quantitative descriptors of a finite metric space (data) that are stable in the presence of noise. Such descriptors are not statistical, rather they rely on the underlying manifold structure of data in an algebraic fashion. The primary concept of TDA is persistent homology, which can be thought of as the extension of homology for finite and noisy metric spaces. In fact, the homology of a finite metric space, such as digital data, is essentially trivial, since the only non-null homology group is H_0. Moreover, persistent homology allows to express data as sequences of simplicial complexes, which allows for a study of the evolution of the homology. The choice on how to construct such a sequence is fundamental and shapes the point of view of such a study.

1.2.1 Persistent homology

Persistent homology measures the scale of topological features computed by homology. In real-world data, the scale at which important topological features occur is not a priori obvious. As an example, one can think of a football ball and a sponge ball of the same size. If looked at closely, the latter is full of small holes and the two balls do not look anything alike. If they are looked at from a distance, however, they do share the same shape. PH is able to track topological changes at different scales of resolution and store such information. It is composed of two main ingredients, one geometric in the form of a function on a topological space, and one algebraic, which turns the function into measurements. We stress the fact that the choice of the function is fundamental in extracting meaningful measurements. Let $\mathcal{K}$ be a simplicial complex and $f\colon \mathcal{K} \to \mathbb{R}$ a monotonic function, that is, $f(\tau) \leq f(\sigma)$ whenever τ is a face of σ. We are going to refer to f as the filtration function. The monotonicity of f guarantees that setting $\mathcal{K}(a) = f^{-1}(-\infty, a]$, it is a subcomplex of $\mathcal{K}$ for every $a \in \mathbb{R}$. Moreover, since $\mathcal{K}$ is finite, there is a finite sequence of filtration values $-\infty = a_0 < a_1 < a_2 < \cdots < a_n$ of the simplices in $\mathcal{K}$. This means that we can arrange all subcomplexes of $\mathcal{K}$ in a sequence of complexes called **filtered simplicial complex**

$$\emptyset = \mathcal{K}_0 \subseteq \mathcal{K}_1 \subseteq \cdots \subseteq \mathcal{K}_n = \mathcal{K},$$

where $\mathcal{K}_i = \mathcal{K}(a_i)$ for each i. For ease of notation, we will refer to such a filtration as $(\mathcal{K}_i)_{i=0}^n$, omitting from the notation the dependency from the filtration function f, since it is clear. Moreover, it is clear as of now that the choice of a filtration f defines different filtered simplicial complexes, hence different structures in persistent homology. In particular, we are interested in the topological evolution of this sequence of complexes. For every $i \leq j$ the inclusion map $\iota\colon \mathcal{K}_i \hookrightarrow \mathcal{K}_j$ induces an homomorphism between the respective homology groups $f_p^{i,j}\colon H_p(\mathcal{K}_i) \to H_p(\mathcal{K}_j)$, for each dimension p. That is, for each dimension p, we have the sequence of homology groups connected by homomorphisms

$$H_p(\mathcal{K}_0) \to H_p(\mathcal{K}_1) \to \cdots \to H_p(\mathcal{K}_n) = H_p(\mathcal{K}).$$

Actually, most of these homomorphisms are actually isomorphisms. In such cases, no topological events occur. Rather, we are interested in the changes of homology groups during the filtration, that is, in the emergence and disappearance of classes. By convention, when two classes merge the elder one is the one that persists. This convention is usually known to as *elder rule* and is functional in our setting.

Definition 1.2.1. The **p-th persistent homology groups** of the filtration $(\mathcal{K}_i)_{i=0}^n$ are the images of homomorphisms induced by inclusion, $H_p^{i,j} := \mathrm{Im} f_p^{i,j}$, for $0 \leq i \leq j \leq n$. The **p-th persistent Betti numbers** are the ranks of these groups, $\beta_p^{i,j} := \mathrm{rank} H_p^{i,j}$.

In what follows, it may be helpful to say that a homology class $\gamma \in H_p^{i,i} = H_p(\mathcal{K}_i)$ is **born** at $\mathcal{K}_i$ if $\gamma \notin H_p^{i-1,i}$. Similarly, we say that a class γ born at $\mathcal{K}_i$ **dies** at $\mathcal{K}_j$, with $j > i$, if $f_p^{i,j-1}(\gamma) \notin H_p^{i-1,j-1}$ but $f_p^{i,j}(\gamma) \in H_p^{i-1,j}$. That is, the class γ merges with an older class when passing from $\mathcal{K}_{j-1}$ to $\mathcal{K}_j$. The **persistence** of a class γ born at $\mathcal{K}_i$ and that dies at $\mathcal{K}_j$ is the difference $\mathrm{pers}(\gamma) := a_j - a_i$, where a_j (resp. a_i) is the filtration value of $\mathcal{K}_j$ (resp. $\mathcal{K}_i$). If a class never dies, we set its persistence to infinity. We stress the fact that multiple classes can born and die at the same time. We denote with $\mu_p^{i,j}$ the multiplicity of indipendent p-dimensional classes that are born at $\mathcal{K}_i$ and die at $\mathcal{K}_j$. It holds that $\mu_p^{i,j} := \left(\beta_p^{i,j-1} - \beta_p^{i,j}\right) - \left(\beta_p^{i-1,j-1} - \beta_p^{i-1,j}\right)$ for all $i < j$ and all p.

Definition 1.2.2. The **p-persistence diagram** (PD) of the filtration $(\mathcal{K}_i)$, denoted as $\mathrm{Dgm}_p(f)$, is the multiset of points (a_i, a_j) with multiplicity $\mu_p^{i,j}$, together with all the points (a, a) with infinite multiplicity.

Remark 1.2.3. It follows directly from the monotonicity of f that every point (a_i, a_j) of the PD verifies $a_j > a_i$. Therefore, every point of the PD lies in the closed half-plane above the diagonal.

Remark 1.2.4. Persistence diagrams are graded object, that is, there is a persistence diagram for every natural p. When the context is clear, we are going to omit the dependence from p in the notation of persistence diagrams. Moreover, given a PD D, it may be useful sometimes to refer to $\mu\colon D \to \mathbb{Z}_{>0}$ as its multiplicity function.

Lemma 1.2.5 (Fundamental lemma of persistent homology [39]) *Let $\emptyset = \mathcal{K}_0 \subseteq \mathcal{K}_1 \subseteq \cdots \subseteq \mathcal{K}_n = \mathcal{K}$ be a filtered simplicial complex. For every pair of indices $0 \leq k \leq l \leq n$ and every dimension p, the p-th persistent Betti number is $\beta_p^{k,l} = \sum_{i \leq k} \sum_{j > l} \mu_p^{i,j}$.*

The fundamental lemma of persistent homology guarantees that the persistence diagram encodes all information about persistent homology groups and it is therefore the invariant that we are interested in. Moreover, the PD offers two visualizations easily interpretable. The first one is to plot each element in the PD as a point in the upper half-plane, the second one, usually referred to as persistence barcode [68], is to plot bars of length equal to their respective persistence. Figure 1.2 shows four stages of a filtration for points sampled from a circle and corrupted with noise. Figure 1.3 shows the persistence barcode associated with the filtration in Figure 1.2. Vertical black lines correspond to the four stages depicted in Figure 1.2. Red bars indicate classes in H_0 and blue bars indicate classes in H_1.

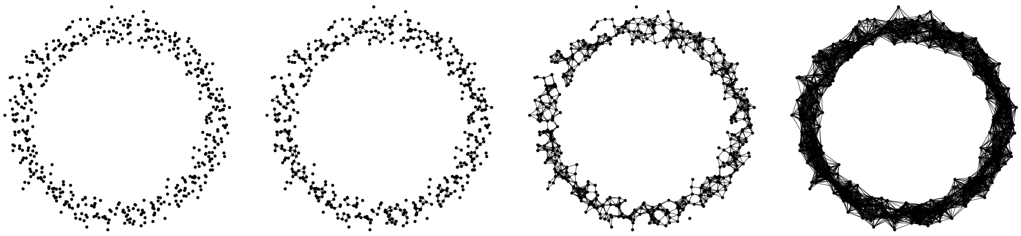


Figure 1.2: Different stages of a filtration for noisy points sampled from a circle.

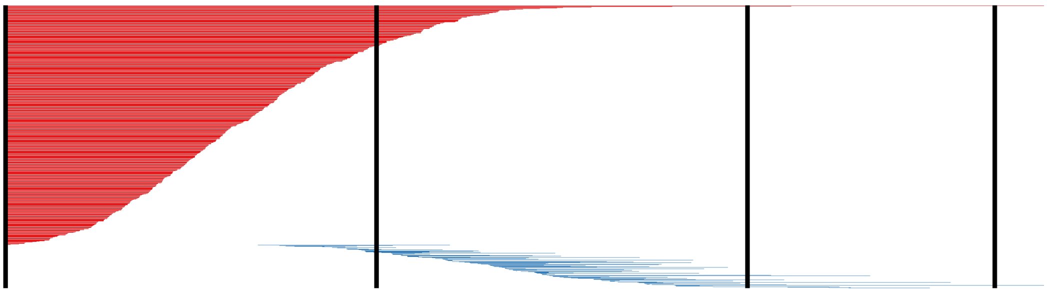


Figure 1.3: Persistence barcode associated with noisy points sampled from a circle. Vertical black lines correspond to the stages of Figure 1.2. Red bars indicate classes in H_0 and blue bars indicate classes in H_1.

Finally, we are able to address the stability of PDs, a property that makes such descriptors suitable for real-world applications. Let D, D' be two persistence diagrams and $\|\cdot\|_\infty$ be the L_∞-norm.

Definition 1.2.6. The **bottleneck distance** d_{match} between D and D' is defined as

$$d_{\text{match}}\left(D, D'\right) := \inf_{\eta\colon D \to D'} \sup_{\alpha \in D} \|\alpha - \eta(\alpha)\|_\infty ,$$

where η ranges over all multi-bijections between D and D'.

We stress the fact that the presence of every point of the diagonal with infinite multiplicity on the persistence diagram guarantees that there is a multi-bijection between each pair of PDs. Given a triangulable topological space $\mathbb{X}$ and a continuous function $f\colon \mathbb{X} \to \mathbb{R}$, $a \in \mathbb{R}$ is a **homological critical value** for f if there is no $\varepsilon > 0$ for which $f_p^{a-\varepsilon, a+\varepsilon}$ is an isomorphism for each dimension p.

Definition 1.2.7. We call a continuous function f **tame** if it has only finitely many homological critical values and all homology groups have finite rank.

Theorem 1.2.8 (Stability theorem for bottleneck distance [39]) *Let $\mathbb{X}$ be a triangulable topological space and let $f, g\colon \mathbb{X} \to \mathbb{R}$ be two tame functions. For each dimension p, the bottleneck distance between the diagrams $D = \mathrm{Dgm}_p(f)$ and $D' = \mathrm{Dgm}_p(g)$ is bounded from above by the L_∞-distance between the functions. That is, $d_{\text{match}}\left(D, D'\right) \leq \|f - g\|_\infty$.*

The main drawback of the bottleneck distance is its insensitivity to details. For this reason, it may be useful to introduce an alternative metric between persistence diagrams.

Definition 1.2.9. Let q be a positive real number and let D, D' be two persistence diagrams. The **q-Wasserstein distance** between D and D' is defined as

$$d_q\left(D, D'\right) := \left[\inf_{\eta\colon D \to D'} \sum_{\alpha \in D} \|\alpha - \eta(\alpha)\|_\infty^q\right]^{1/q} ,$$

where η ranges over all multi-bijections between D and D'.

Perhaps unsurprisingly, a similar stability result for any q-Wasserstein distance is out of reach. However, we can state something very similar for a large class of functions. Given a triangulable topological space $\mathbb{X}$, we say that the triangulation of $\mathbb{X}$ **grows polynomially** if there are constants c and j such that $N(r) \leq c/r^j$, where $N(r)$ is the minimum number of simplices in a triangulation with mesh at most r. Finally, the **k-th total persistence** of a persistence diagram D is the sum of k-th powers of the persistence of its points.

Theorem 1.2.10 (Stability theorem for Lipschitz functions [39]) *Let $f, g\colon \mathbb{X} \to \mathbb{R}$ be tame 1-Lipschitz functions on a metric space whose triangulations grow polynomially with constant exponent j. Then there are constants C and $k > j$ no smaller than 1 such that the q-Wasserstein distance between $D = \mathrm{Dgm}_p(f)$ and $D' = \mathrm{Dgm}_p(g)$ is $d_q\left(D, D'\right) \leq C \|f - g\|_\infty^{1-k/q}$ for every $q \geq k$.*

1.2.2 Vectorisation methods

The main drawback of the space of PDs, endowed with the bottleneck or Wasserstein distance, is that it is infinite-dimensional (cf. [69]) and it lacks a Hilbert space structure. Moreover, even basic statistical quantities such as the average of two persistence diagrams are not well defined. As such, PDs can not be directly employed in a ML algorithm. Hence, in literature considerable effort has been devoted to embed PDs into a more manageable space, resulting in a large number of solutions that have been devised. Essentially, all of them embed the space of PDs in a suitable Hilbert or vector space. In any case, since an infinite-dimensional space does not admit a faithful embedding in finite-dimensional vector space, and even mapping barcodes to infinite-dimensional vector space may result in a lack of discriminative power, there is no canonic way to perform this embedding. Loosely speaking, similarly to the notorious No Free Lunch theorem (cf. [70]), there is no optimal embedding suitable for every application. To be more precise, usually, the methods that embed the space of PDs in a Hilbert space are referred to as kernel methods, and the methods that embed the space of PDs in a vector space are referred to as vectorisation methods. The former of which are usually computationally prohibitive, and in literature the latter are preferred. In this section, we are going to briefly describe the vectorisation methods that have been employed in the topological machine learning pipeline devised alongside this dissertation. For a brief summary of the different vectorisation methods, their stability and parameters, refer to Table 1.1. We refer to Figure 1.4 for a graphical example of a persistence diagram and its various vectorisations by means of the techniques that we are going to introduce. For consistency with the original works, we describe the different methods with the original notation. For this reason, in this section exclusively, the same symbol could be used several times to indicate different concepts. Let D be a finite persistence diagram and $\mu\colon D \to \mathbb{Z}_{>0}$ its multiplicity function.

1.2.2.1 Persistence statistics [71]

The **persistence statistics** is the simplest descriptor defined in literature and it consists in a collection of 38 statistical quantities of the persistence diagram. More in detail, it consists of the mean, standard deviation, median, interquartile range, full range, $10^{\text{th}}, 25^{\text{th}}, 75^{\text{th}}$ and 90^{th} percentiles of the births, deaths, midpoints and lifespan of each point of the PD counted with multiplicity, the total number of bars and the entropy E_μ of the PD, which is defined as

$$E_\mu := -\sum_{[p,q]\in D} \mu_{p,q}\left(\frac{q-p}{L_\mu}\right)\log\left(\frac{q-p}{L_\mu}\right),$$

where L_μ is the weighted sum $L_\mu := \sum_{[p,q]\in D} \mu_{p,q}\,(q-p)$.

Vectorisation method	**Stability**	**Parameters**
Persistence statistics	✗	✗
Entropy summary	✓	resolution $\in \{50, 100\}$
Algebraic functions	✗	✗
Tropical coordinate function	✓	resolution $\in \{50, 100\}$
Complex polynomial	✓	number of coefficients $\in \{5, 20\}$ polynomial type $\in \{R, T\}$
Betti curve	✗	resolution $\in \{50, 100\}$
Lifespan curve	✗	resolution $\in \{50, 100\}$
Persistence landscapes	✓	number of landscapes $\in \{5, 10\}$ resolution $\in \{50, 100\}$
Persistence silhouette	✓	weight $\in \{1, 10\}$ resolution $\in \{50, 100\}$
Persistence image	✓	bandwidth $\in \{0.05, 1\}$ resolution $\in \{50, 100\}$
Template function	✓	$\delta \in \{5, 25\}$ $\pi \in \{1, 20\}$
Adaptive template system	✓	number of clusters $\in \{10, 25\}$
ATOL	✓	number of functions $\in \{2, 4\}$

Table 1.1: Summary of the vectorisation methods used in the topological machine learning pipeline, their stability and parameters.

1.2.2.2 Entropy summary [28]

The **entropy summary** is the extension of the entropy defined in the persistence statistics to a piecewise constant map $S_\mu \colon \mathbb{R} \to \mathbb{R}$ defined by

$$S_\mu(t) := -\sum_{[p,q]\in D} \mu_{p,q} \left(\frac{q-p}{L_\mu}\right) \log\left(\frac{q-p}{L_\mu}\right) \mathbb{1}_{p\le t\le q},$$

where L_μ is defined as in the persistence statistics. It arises from the idea to summarize the information about the number of intervals of the persistence barcode and their homogeneity with a simple descriptor such as a piecewise constant map.

1.2.2.3 Algebraic functions [72]

A more evolved vectorisation comes in the form of algebraic vectorisations. The next three methods define polynomial maps and evaluate them on the persistence diagram. The ring of **algebraic functions** arises from the understanding that a persistence barcode can be identified by the collection $\{x_1, y_1, \dots, x_n, y_n\} \in \mathbb{R}^{2n}$, where x_i and y_i represents the birth and death of the i-th bar. Of course, aiming to characterize the barcodes with polynomials in $2n$ variables, such polynomials should be independent on the order of the barcodes. It turns out that such a subring can be characterized algebraically by the subring of polynomials f in variables $\{x_1, y_1, \dots, x_n, y_n\} \in \mathbb{R}^{2n}$ such that there exist polynomials $\{g_i, 1 \le i \le n\}$ satisfying $\frac{\partial f}{\partial x_i} + \frac{\partial f}{\partial y_i} = (x_i - y_i)g_i$. The vectorisation by means of algebraic functions then consists on selecting a finite set of such polynomials and evaluating them at each point $(x_i, y_i) = (p_i, q_i) \in D$.

1.2.2.4 Tropical coordinate function [73]

The **tropical coordinate functions** vectorisation arises from the same concept as algebraic functions, but such polynomials are required to be both symmetric and tropical. We refer to [74] for a formal definition of both concepts. In particular, the max and min functions are tropical functions and in [73] it is shown that they are more suitable than other polynomials, given the underlying structure of the barcode. After selecting a fixed number of such polynomials, to produce the vectorisation they are evaluated at $(x_i, y_i) = (q_i - p_i, \max(r(q_i - p_i), p_i))$ or $(x_i, y_i) = (q_i - p_i, \min(r(q_i - p_i), p_i))$, where r is a positive integer parameter.

1.2.2.5 Complex polynomial [75, 76]

The **complex polynomial** vectosization of D is yet again another vectosization of persistence diagrams by means of polynomials. The motivation behind this approach was to speed up the process of computing the bottleneck distance, which can be quite burdensome, with a suitable metric defined on vectors. Such vectors are composed of the coefficients of the representation of PDs by means of complex polynomials. More in detail, we define the complex polynomial $C_X(z) := \prod_{[p,q]\in D} [z - X(p,q)]^{\mu_{p,q}}$, where $X\colon \mathbb{R}^2 \to \mathbb{C}$ is either $R(x,y) := x + iy, T(x,y) := \frac{y-x}{2}[(\cos\alpha - \sin\alpha) + i(\cos\alpha + \sin\alpha)]$ or

$$S(x,y) := \begin{cases} \frac{y-x}{\alpha\sqrt{2}}(x+iy) & \text{if } (x,y) \neq (0,0) \\ 0 & \text{otherwise} \end{cases}$$

and $\alpha := \sqrt{x^2 + y^2}$. The vectorisation is therefore obtained by considering the coefficients of such a polynomial.

1.2.2.6 Betti curve [77]

The next four methods turn a persistence diagram into one or more curves. The vectors are subsequentially obtained by sampling the given curve on a finite grid on $\mathbb{R}$. The **Betti curve** was originally introduced to study the evolution of time series by means of TDA and is based on the assumption that the main point of persistent homology consists on following the change in the number of holes corresponding to the change in the radius parameter. More formally, the Betti curve is defined as the curve $\beta_\mu\colon \mathbb{R} \to \mathbb{R}, \beta_\mu(t) := \sum_{[p,q]\in D} \mu_{p,q} \mathbb{1}_{p\leq t\leq q}$. By discretizing such a curve in a grid of $\mathbb{R}$ we obtain the Betti curve vectorisation.

1.2.2.7 Lifespan curve [78]

The **lifespan curve** tracks lifespan information over the filtration. One can think of it as a topological intensity function that accounts for the size or intensity of topological features. More specifically, the lifespan curve is the map $L_\mu\colon \mathbb{R} \to \mathbb{R}, L_\mu(t) := \sum_{[p,q]\in D} \mu_{p,q}\, (q-p)\, \mathbb{1}_{p\leq t\leq q}$.

1.2.2.8 Persistence landscapes [15]

The **persistence landscape** originated from the idea of converting the barcode into a function in an additive manner. Since the resulting descriptor belongs to a separable Banach space, it is easy to apply statistical tools to it. Informally, the persistence landscape counts the number of points in the persistence diagram in the upper left quadrant of (b, d), and we obtain a vectorisation of the PD by "stacking isosceles triangles" whose bases are the intervals in the barcode. It is important to note that the mapping from persistence diagrams to persistence landscapes is stable and invertible. Formally, the persistence landscape is the collection of curves $\Lambda_i^\mu \colon \mathbb{R} \to [-\infty, \infty]$, $\Lambda_i^\mu(t) = \sup \left\{ s \geq 0 \text{ such that } \left(\sum_{[p,q] \in D} \mathbb{1}_{[t-s,t+s] \subset [p,q]} \mu_{p,q} \right) \geq i \right\}$, with the convention that the supremum over an empty set is zero. Since D is finite, it exists $\bar{i}$ such that $\Lambda_i^\mu(t) \equiv 0$ for every $i > \bar{i}$ and this yields a finite vectorisation of the PD.

1.2.2.9 Persistence silhouette [79]

The w-weighted **persistence silhouette** originates as a variation of persistence landscapes which allows for a trade-off parameters between uniformly treating all pairs in the persistence diagram and considering only the most persistence pairs. Specifically, when w is small, the persistence silhouette is dominated by the effect of low persistence pairs. Conversely, when w is large, the persistence silhouette is dominated by the most persistent pairs. Given a function $w \colon D \to \mathbb{R}_{>0}$, the w-weighted persistence silhouette is the map $\phi_\mu^w \colon \mathbb{R} \to \mathbb{R}$ defined by

$$\phi_\mu^w := \frac{\sum_{[p,q] \in D} w(p,q) \mu_{p,q} \Delta([p,q], t)}{\sum_{[p,q] \in D} w(p,q) \mu_{p,q}},$$

where $\Delta([p,q], t) := \max\left(\min\left(t - p, q - t\right), 0\right)$. Since it is a variation of the persistence landscape, it benefits from all its statistical properties.

1.2.2.10 Persistence image [13]

An evolution of curve vectorisation comes in the form of functional vectorisation, which arises from the same idea but with a codomain different than $\mathbb{R}$. The first functional vectorisation that we are going to introduce is the **persistence image**, which aims to compute a vector in a stable, efficient way and to maintain an interpretable connection with the original PD. Moreover, the persistence image allows to adjust the relative importance of points in different regions of the PD. Given a continuous, piecewise-differentiable function $f \colon \mathbb{R}^2 \to \mathbb{R}_{\geq 0}$ such that $f(x, 0) = 0$ and a collection of smooth probability distributions $\Psi = \{\psi_{p,q}\}$ with mean $(p, q-p)$, the persistence image $\mathbf{I}_{f,\Psi}^\mu$ with respect to (f, Ψ) is the discretization on a finite grid Z of $\mathbb{R}^2$ defined by

$$\mathbf{I}_{f,\Psi}^\mu(Z) := \int \int_Z \rho_{f,\Psi}^\mu(x, y) \; dx \; dy,$$

where $\rho_{f,\Psi}^\mu(x, y) := \sum_{[p,q] \in D} \mu_{p,q} f(p, q-p) \psi_{p,q}(x, y)$.

1.2.2.11 Template function [80]

The **template function** addresses the problem of approximating continuous functions on compact subsets of the space of persistence diagrams. Given $\mathcal{S}$ a compact subset of the space of persistence diagrams, the goal is to devise provably-correct and computationally feasible approaches to approximating a function $F\colon \mathcal{S} \to \mathbb{R}$, given a finite sample $D_1, \dots, D_n \in \mathcal{S}$ and their value $F(D_1), \dots, F(D_n) \in \mathbb{R}$. More in detail, a template system T is a subset of $C_C(\Delta)$ (the space of functions from the upper half-plane to $\mathbb{R}$ with compact support) such that for every pair of barcodes D_1, D_2, there is at least one $f \in T$ such that $V_{D_1}(f) \neq V_{D_2}(f)$, where V_D is the function induced by the barcode and it is defined as $V_D\colon C_C(\Delta) \to \mathbb{R}, V_D(f) := \sum_{[p,q]\in D} \mu_{p,q} f(p, q-p)$. The template function vectorisation with respect to the template system T is therefore the vector $\tau^{\mu} := (V_D(f_1), \dots, V_D(f_n))$.

1.2.2.12 Adaptive template system [14]

The last two examples of vectorisation methods are called ensemble vectorisation methods and require a large quantity of persistence diagrams to train a suitable vectorisation. The main utility of the **adaptive template system** vectorisation is that they can be used to construct dense subsets of the space of continuous real-valued functions with domain a PD, with respect to the compact-open topology. Although this topology is not metrizable, two functions are deemed to be nearby if their values on compact sets are similar. Since the space of persistence diagrams is rather large and complicated, such comparisons are desirable. The adaptive template system defines finitely many ellipses E_j that strictly contain the support of the collection of PDs. Each of these ellipses is expressed in quadratic form by the 2×2 matrix A_j. At this point, it defines

$$g_j(z) := \begin{cases} 1 - h_j(z) & \text{if } h_j(z) < 1 \\ 0 & \text{otherwise,} \end{cases}$$

where $h_j(z) = (z - x_j)^T A_j (z - x_j)$ and $x_j \in \mathbb{R}^2$ is the center of the ellipse E_j. That is, inside the ellipse g_j measures how far a point is from the center, and it measures 0 outside of the ellipse. After these steps, it applies the same procedure as the template function.

1.2.2.13 ATOL [81]

The **ATOL** vectorisation is the last vectorisation method that we are going to introduce. It relies on a quantization of the space of diagrams that is statistically optimal and is fast and practical even for large-scale and high dimensional scenarios. This method is proven able to separate clusters of persistence diagrams. Given $z = (z_1, \dots, z_b)$ points in $\mathbb{R}^2$ sampled indipendently and identically distrubeted, the contrast functions $\{\Omega_i, 1 \leq i \leq b\}$ are given by $\Omega_i\colon \mathbb{R}^2 \to \mathbb{R}, \Omega_i(x) := \exp\left(-\frac{\|x-z_i\|}{\frac{1}{2}\max_{j\neq i}\|z_j - z_i\|_2}\right)$. The ATOL contrast function vectorisation is therefore given by the vector $\left(\Omega_1^{\mu}, \dots, \Omega_b^{\mu}\right)$, where $\Omega_i^{\mu} := \sum_{[p,q]\in D} \mu_{p,q} \Omega_i(p, q)$.

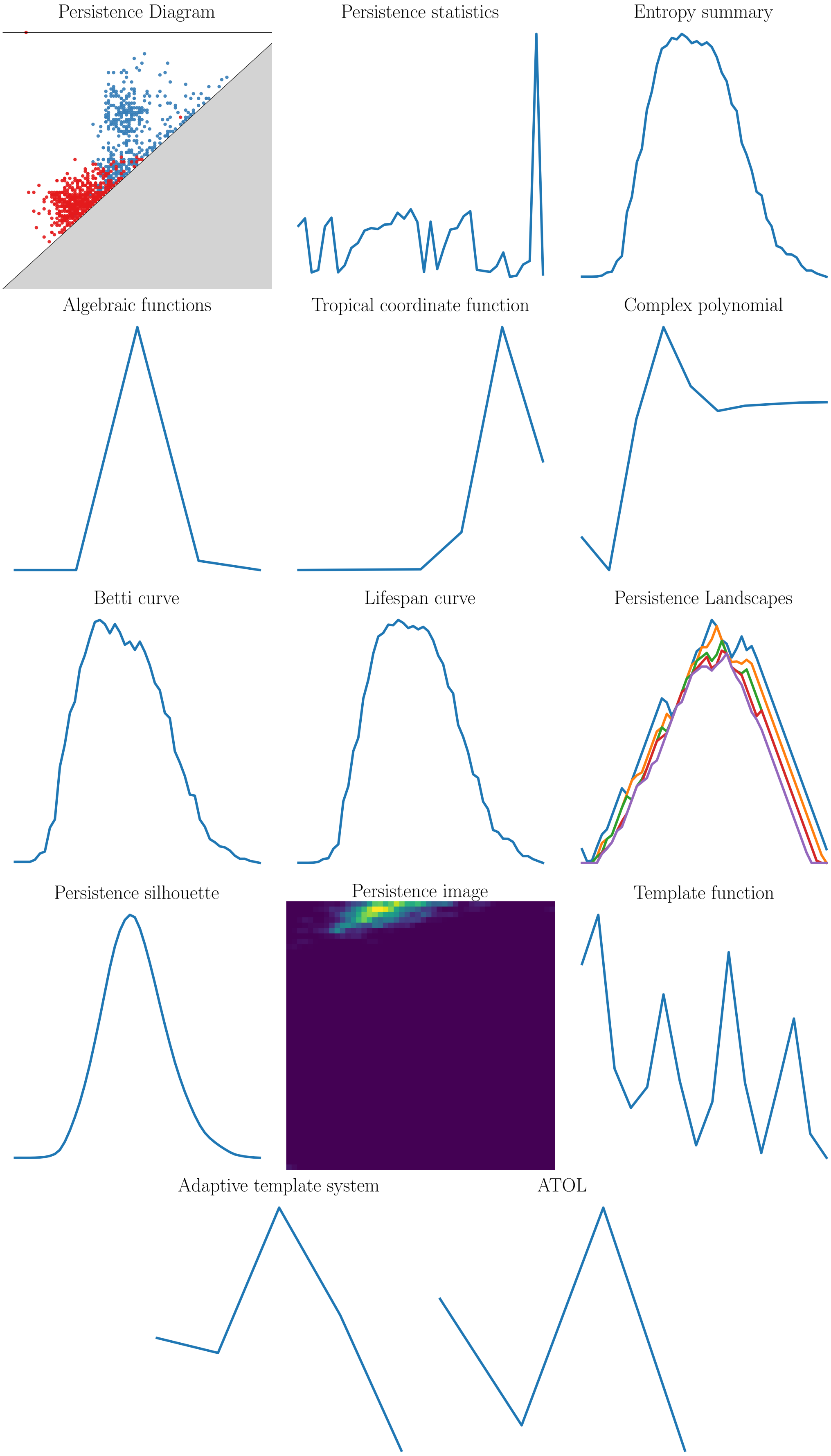


Figure 1.4: Examples of a PD and its vectorisations by means of the various techniques used in the topological machine learning pipeline.

1.3 The topological machine learning pipeline

Now that we have introduced the necessary bibliography, we can finally formally describe the topological machine learning pipeline that was devised alongside this dissertation. We highlight the fact that the pipeline here described represents a slight extension over the one already presented in [57]. We refer to Figure 1.5 for a schematic of the pipeline. This is motivated by the effectiveness of methods from computational topology to the analysis of digital data and the excellent results achieved in recent years by its descriptors. In a topological machine learning pipeline, there are three non-canonical choices that are fundamental in the nature of the study. Firstly, a suitable filtration must be chosen to compute persistence diagrams from data. This choice is fundamental since it represents the geometric component of the pipeline. More in detail, different filtrations highlight different geometric features of data and result in different persistence diagrams. However, the choice of the filtration is usually bounded by the data type. For point cloud data, the most common filtration is the Vietoris-Rips filtration [82], for images the cubical complex filtration [83] and for signals the lower star filtration [59, 84]. More filtrations are available in the literature and are used in the pipeline. Since their use is strictly related to the data type, we postpone their description to the next chapter, where different benchmark datasets are presented. After the filtration has been chosen, the pipeline computes the persistence diagrams associated to the dataset. The second non-canonical choice is how to handle the homological dimensions in the vectorisation step. This choice in particular is often overlooked in literature, in the sense that a fair comparison between the different approaches is rarely provided, with the implicit assumption that only the best results are reported. To be more precise, when entering the vectorisation process of the pipeline, the homological dimension must be omitted. This technicism admits three possible solutions. The first is to focus only on one dimension, the one deemed more important. Examples of this approach can be found in [14, 15, 28, 29, 30, 31, 32, 33, 37, 85]. The second approach is to vectorise each dimension separately and then concatenate the results, as in [14, 34, 35, 36] (among many others). Finally, perhaps unexpectedly, forgetting the homological dimension and vectorising everything altogether still represents a good alternative [16, 57]. Our results consistently favor a certain approach, as we will show in the following chapters, where an extensive comparison between the different approaches will be performed. In any case, our pipeline performs all three approaches and returns the accuracy for each of them. Finally, the last non-canonical choice in a machine learning pipeline is the choice of the vectorisation method and its possibile parameters. We refer to Table 1.1 for the vectorisation methods and parameters combination used in this dissertation. Again, via a grid-search approach, the pipeline vectorises the PD using the methods previously described. In the end, the resulting vectors enter a machine learning algorithm (support vector classifier or random forest classifier, we refer to [86] for more information on machine learning) and returns the accuracy for each method. We stress the fact that, after the persistence diagrams are computed with the suitable choice of the filtration, the pipeline is entirely automatic. For the TDA part of the pipeline, we used the methods coming from the Gudhi [87] and Giotto [88] Python library. For the ML part, we used the `scikit-learn` [89] Python library.

Now that we have introduced the necessary literature on the topic of topological machine learning, discussed its potentials and described the pipeline devised alongside this dissertation, we are ready to test its results on benchmark datasets. More importantly, the next step is to dive deep in some of the most overlooked aspects of this theory. All this is the content of the next chapter.

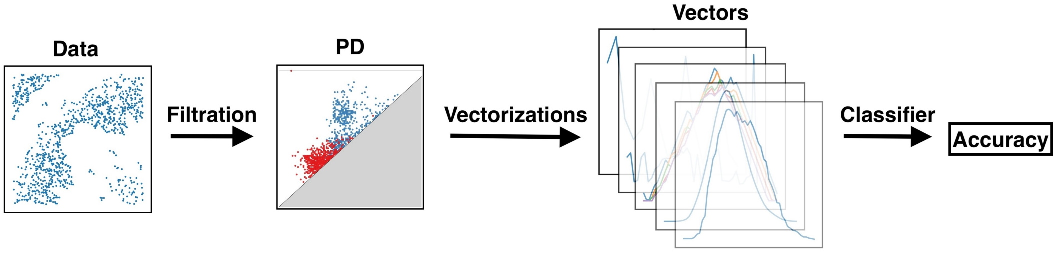


Figure 1.5: Scheme of the topological machine learning pipeline. Starting from the data we produce a persistence diagram by means of a filtration, which is then vectorised through various methods and such vectors enter a machine learning algorithm which returns a classification with a certain accuracy.

Chapter 2

The TML pipeline: experiments and results

The aim of this chapter is mainly twofold: provide some applications of the topological machine learning pipeline described in Chapter 1 on a variety of benchmark datasets and provide a fair comparison of the accuracy results of the different ways to handle homological dimensions previously described. This second goal in particular is the first contribution of this dissertation and it constitutes a novel attempt to standardise a non canonical part of the TML pipeline. To be more precise, as reported in the literature there are three ways to manage the homological dimensions in the vectorisation step:

- **Singular dimension vectorisation** [14, 15, 28, 29, 30, 31, 32, 33, 37, 85]: only a specific homological dimension is vectorised and used as train and test dataset by the machine learning classifier. We refer to this approach as $\mathbf{H_i}$, for i ranging in the homological dimensions;

- **Concatentating dimension vectorisation** [14, 34, 35, 36, 90]: every homological dimension is vectorised singularly, and then the resulting vectors are concatenated. We refer to this approach as the **concat** approach;

- **Fused dimension vectorisation** [16, 57]: perhaps surprisingly, completely neglecting the homological dimension of each point and vectorising the PD altogether is in many cases a good approach. We refer to this approach as the **fused** approach.

In many works (cf. [14, 15, 37, 85, 91]) the methods are used interchangeably, but no comparison is reported. In fact, little at all emphasis on a comparison among the three approaches is provided in the literature. Our findings, on the other hand, lean totally in favor of the concat method. To be more precise, the concat approach may not always be the absolute best, but its results are always among the top performing and it never suffers drops in performance, as is the case for all other methods. In this chapter, we are going to perform experiments in a heterogeneous setting of different datasets, filtrations, vectorisation techniques and classifiers. We want to highlight the fact that if a priori information is known for specific tasks, some of these approaches may not be very insightful. For instance, FMNIST [92] is a dataset of images of clothings. If we try to classify such a dataset considering the persistence arising from a filtration based on the gray value of the image using exclusively H_0 or H_1, it would yield obviously poor results. Nevertheless, we want to perform a blind analysis, as if no a priori information were available. This is to mimic many real-world datasets, where information about the topology of the data may not be so obvious, or even available, a priori. Moreover, the second contribution of this dissertation is a noise robustness test between of the three approaches in a synthetic dataset, in order to better assert our findings. This test is, to the best of our knowledge, novel in the TDA literature. Finally, a preliminary study on how the distribution of points in the PD influences the meaningfulness of the fused approach.

2.1 Benchmark evalutations

We first apply the topological machine learning pipeline to a heterogeneous list of benchmark datasets, with data ranging from images, signals, point cloud data and graphs. Different filtrations will be used and described in the appropriate section. In each table, we report in bold the best method in each line (i.e. each vectorisation method) and in red the overall best accuracy. The first line of each cell reports the best accuracy for a single vectorisation across all parameters and classifiers, and the second line is the average over the different parameters and classifiers combination. In every application, we performed a $70\% - 30\%$ train-test split.

2.1.1 Orbit dataset

The Orbit dataset is a collection of orbits in the plane generated by a linked twisted map, which is the equation:

$$\begin{cases} x_{n+1} = x_n + r y_n(1 - y_n) & \mod 1, \\ y_{n+1} = y_n + r x_{n+1}(1 - x_{n+1}) & \mod 1. \end{cases}$$

With mod 1 we mean that only the fractional part is considered. For different values of the parameter r, different orbits are generated. The task is to recognize the parameter based on the orbit generated. Figure 2.1a shows four different orbits generated for different choices of the parameter r. This dataset is inspired by [93] and it is composed of five different values of the parameter $r \in [2, 3.5, 4, 4.1, 4.3]$, each class with 200 orbits. Each orbit consists of $1,000$ points. For this dataset, we used the Vietoris-Rips filtration [82]. Such a filtration grows balls of radius ε centered at each point. The simplices are the collections of points with diameter less than the current filtration value ε. Following the elder rule, when two or more simplices connect, the younger one dies. We report the accuracy results for the Orbit dataset in Table 2.1. For this first application, both H_1 and concat achieve very similar results and are consistently the best-performing approaches. H_0 is not very informative and fused is informative only with very specific parameter - vectorisation combinations.

2.1.2 SHREC14

SHREC14 [94] is a dataset for shape retrieval of real and synthetic non-rigid 3D human shapes and poses. Following TDA literature on this dataset (cf. [14, 37, 95]), we focused only on the synthetic part. It consists of a classification task with 15 classes of human bodies (e.g. male neutral, male bodybuilder, female neutral) each one with 20 poses. Figure 2.1b shows a sample of the shapes and poses of SHREC14. For this dataset, we perform the same filtration as [14], which exploits the heat kernel signature function to compute persistence diagrams. We report the accuracy results for the SHREC14 dataset in Table 2.2. In this case, H_1 achieves the absolute best result, with 0.97% accuracy. Both fused and concat approaches are right behind with 0.94% and across the board are pretty much equivalent.

2.1.3 Outex dataset

Outex dataset [96] is composed of $1,360$ images divided into 68 classes. Each image is a 128×128 RGB of a sample texture. Figure 2.1c shows sample images for the Outex dataset. Since each image has three channels, we compute the cubical complex filtration [83] for each channel and concatenate the resulting vectors. In detail, the cubical complex exploit the intrisic structure of pixels. Each pixel intensity is given by its graylevel value. A pixel enter the filtration as a 0-simplex when their intensity is greater than the current treshold. In order to define a triangulation of the grid, pixels are connected with their vertical and horizontal neighbours, as well as their neighbours along the first diagonal. 1-simplices and 2-simplices enter the filtration when all pixels composing them enter the filtration. We report the accuracy results for the Outex dataset in Table 2.3. In this case, concat is clearly the only best approach and for the first time H_1 performs poorly. Conversely, H_0 and fused are able to capture key features of the data for classification, despite some notably drop in performances.

2.1.4 FMNIST

The Fashion-MNIST dataset [92] is composed of $60,000$ training images and $10,000$ test images divided in ten classes. Each picture is a 28×28 grayscale image, representing an individual article of clothing collected from Zalando's inventory. See Figure 2.1d for some sample images. We highlight that, due to convergence issues, SVC was not used for this dataset. For this dataset, we performed two different filtrations, in order to highlight the fact that even for suboptimal filtrations the concat approach still consistently performs among the best. The first filtration is the cubical filtration on the grayscale image. The second filtration is inspired by [90]. It defines eight directions and nine centers, binarizes the original image, and computes the distance from such centers and directions, in addition to the density filtration. Finally, it computes the cubical complex and concatenates the results. We refer to this filtration as the multi filtration. Figure 2.2 shows two preprocessed images, one for each filtration. We refer to Table 2.4 for the accuracy results of both filtrations. In particular, we highlight two important aspects. First, despite the suitability of the chosen filtration, the concat approach yields the best results. Second, in order to achieve satisfactory results a more complex and *ad-hoc* filtration has been devised. As already stated, such a filtration is particularly suited for the task at hand, but it is more sensitive to the parameter choice. This is seen by the fact that the gap between the best accuracy and the average is very pronounced. Moreover, the gap between the various vectorisation methods is wider than usual. Despite these details, the accuracy results obtained from this vectorisation are very satisfactory. In this case, H_1 performs considerably worse, while H_0 and fused are able in some combinations to perform reasonably well.

2.1.5 COLLAB dataset

The COLLAB dataset is a network graphs dataset of scientific collaborations presented in [97]. It consists of $5,000$ graphs derived from three public collaboration datasets which also serve as labels: high energy physics, condensed matter physics and astrophysics. Each node of the graphs is an author, and there is a link between two authors if they coauthor a scientific article. COLLAB is a dataset of weighted, undirected graphs. Every collaboration between n authors contributes to the edge weight between those authors of a factor $1/(n-1)$. The vertices are not weighted; this means that all vertices immediately enter the filtration as 0-simplexes. The filtration value of the 1-simplexes is the weight of the edge connecting the two vertices, and for 2-simplexes, we chose as the filtration value the maximum weight of the edges forming it. For computational reasons, we limit to the computation of homology up to dimension 2. We refer to Table 2.5 for the accuracy results of this evaluation. In this case, both H_2, fused and concat offer the best results, while H_0 and H_1 are not very informative.

2.1.6 Gravitational waves dataset

The gravitational waves dataset is inspired by [98, 99]. Without diving into too much detail, which would somehow require an exposition on modern physics, the gravitational wave dataset is composed of signals generated by a surrogate model of a non-spinning binary black hole generating gravitational waveforms. When detected, such a signal is heavily corrupted by noise and it is difficult to discern it from pure noise. The dataset is composed of $1,000$ signals, half of which are a corrupted version of the gravitational waveform, and half of which are simply noise. We refer to Figure 2.3 for a graphical example of the dataset. On the left, we have a purely noise signal (blue), on the right the gravitational waveform (green) and its corrupted version (red). Following [99], we have employed a Takens embedding [100] in order to pick the recurrent structure of the gravitational waveform. More in detail, for this step we used the `giotto` built-in function with both embedding dimension and time delay of 30. For a visual example of the projection of the resulting point cloud in 3D, we refer the reader to Figure 2.4. Moreover, we also performed a lower star filtration for this dataset. We refer to Table 2.6 for the accuracy results of this evaluation. Two things in particular stand out from these results. First, in agreement with what we have already seen, the concat method is among the best performing. Also, perhaps surprisingly, is not necessary a complicated filtration to achieve high accuracy results in this case, since a simple lower star filtration performs considerably better. In any case, when the Takens embedding filtration is employed, both H_1 and H_2 perform worse than the other approaches.

From these first experiments, the features extracted form the concat approach are consistently among the best performing, if not solely. However, it is the only method that never experienced a drop of performances and it is across the board more consistent with respect to vectorisations and parameters combination. Moreover, the fused approach seems to be a viable alternative, however with sporadic drop of performances. We stress the fact that the vectorisation of the fused approach is as computational expensive as a single dimensions, while the concat approach is $i\times$ as expensive, where i is the number of dimensions. The true bottleneck of the pipeline is the computation of the persistence diagrams, however for heavy datasets such as COLLAB also the vectorisation part is not negligible.

2.2 Noise robustness in homological dimensions

To further validate our findings in Section 2.1, in this section we provide a preliminary study on noise robustness of the features extracted from the various homological dimensions. In particular, this study aims to discover relationships between the dimension of the feature and its stability with respect to noise. To this aim, we developed a synthetic dataset inspired by [13]. It consists of 600 point clouds, each with 500 points sampled from six geometrical shapes: the unit cube, a sloped circle of diameter one, the sphere of diameter one, three clusters with centers randomly chosen from the unit cube, three clusters within three clusters with centers randomly chosen from the unit cube and a torus. Figure 2.5 shows samples of the six classes of the Synthetic dataset. From this dataset, we created three variants with increasing degrees of noise. Every point is perturbed by a Gaussian distribution of standard deviation $\sigma \in \{0.05, 0.1, 0.15\}$. Figure 2.6 shows the same class corrupted by the various degrees of noise. For this experiment, we also computed H_2 since it is not trivial for points in 3D. Moreover, we expanded the list of possible approaches. In particular, alongside the usual approaches that we refer to as H_0, H_1, H_2, f_{012} and c_{012} we also restricted ourselves to the case where H_2 was not computed. The resulting fused and concat approach containing only features from H_0 and H_1 are referred to as f_{01} and c_{01}. The rational behind these choices is to further validate our findings even when not all non-trivial homological dimensions are computed (e.g. for computational restrictions). We report in Table 2.7 and Table 2.8 the accuracy results of the pipeline for the various degrees of noise. From these results, we want to highlight three aspects, which are intuitively apparent but which it is nonetheless encouraging to observe in this experiment, albeit a preliminary one. Firstly, in accordance with our previous findings, concatenating all homological dimensions is yet again the best approach, regardless of the degree of noise corrupting the data. Secondly, features extracted from singular dimensions and with zero or low degrees of noise are unquestionably better in higher dimensions than in lower ones, for this dataset. However, when the incidence of noise grows, higher dimensions lose their advantage with respect to features extracted from lower dimensions. This could be explained by the fact that features in higher dimensions are generated by the interaction of a larger number of points, whose mutual position is more perturbed by noise with respect to the mutual position of fewer points. Thirdly, it is best to

	Vietoris Rips			
vectorisation	**H_0**	**H_1**	**fused**	**concat**
Pers. Statistics	0.62	0.90	0.77	**0.91**
	0.49	**0.63**	0.51	0.60
Entropy Summary	0.59	**0.94**	0.64	0.92
	0.58	**0.90**	0.58	**0.90**
Algebraic Functions	0.61	0.87	0.87	**0.88**
	0.47	**0.75**	0.58	0.61
Tropical Coordinates	0.58	0.93	0.92	**0.94**
	0.44	**0.80**	0.59	0.73
Complex Polynomials	0.55	**0.79**	0.75	0.76
	0.44	**0.56**	0.51	0.52
Betti Curve	0.20	**0.92**	0.20	**0.92**
	0.20	**0.85**	0.20	0.83
Lifespan Curve	0.57	0.94	0.57	**0.95**
	0.50	**0.91**	0.50	0.82
Pers. Landscapes	0.20	**0.96**	0.20	**0.96**
	0.20	**0.92**	0.20	0.78
Pers. Silhouette	0.53	**0.92**	0.48	0.91
	0.30	**0.85**	0.29	0.65
Pers. Images	0.56	**0.74**	0.55	0.73
	0.39	**0.55**	0.41	0.47
Template Functions	0.61	0.63	0.41	**0.78**
	0.45	0.46	0.32	**0.53**
ATS	0.57	**0.83**	0.54	**0.83**
	0.41	0.57	0.42	**0.58**
ATOL	0.59	0.74	0.60	**0.76**
	0.44	0.53	0.44	**0.54**

Table 2.1: **Orbit dataset results.** H_1 and concat achieve consistently the best results. First line of each row: best result, second line: average.

	Heat kernel signature			
vectorisation	$\mathbf{H_0}$	$\mathbf{H_1}$	**fused**	**concat**
Pers. Statistics	0.87 0.63	**0.96** **0.87**	0.91 0.75	0.94 0.70
Entropy Summary	0.24 0.24	0.70 0.62	0.73 **0.64**	**0.74** 0.63
Algebraic Functions	0.89 0.74	0.87 0.80	0.86 0.83	**0.90** **0.89**
Tropical Coordinates	0.79 0.52	**0.89** **0.78**	0.87 0.66	0.87 0.70
Complex Polynomials	0.83 0.62	0.88 **0.75**	0.86 0.70	**0.91** **0.75**
Betti Curve	0.08 0.08	**0.71** 0.61	0.66 **0.62**	0.64 0.59
Lifespan Curve	0.69 0.66	0.87 0.87	**0.91** **0.89**	0.90 0.88
Pers. Landscapes	0.69 0.64	**0.90** **0.89**	**0.90** **0.89**	**0.90** **0.89**
Pers. Silhouette	0.69 0.64	0.83 0.80	0.86 0.77	**0.90** **0.85**
Pers. Images	0.51 0.33	0.90 0.75	**0.91** **0.76**	0.87 0.62
Template Functions	0.89 0.70	**0.97** **0.91**	0.94 0.89	0.94 0.89
ATS	0.78 0.55	0.90 **0.84**	0.90 0.82	**0.91** 0.81
ATOL	0.81 0.52	0.88 0.81	0.86 0.82	**0.93** **0.83**

Table 2.2: **SHREC14 results.** H_1 achieves the overall best result, fused and concat are right behind. First line of each row: best result, second line: average.

	Cubical complex			
vectorisation	$\mathbf{H_0}$	$\mathbf{H_1}$	**fused**	**concat**
Pers. Statistics	0.88	0.60	**0.90**	**0.90**
	0.60	0.44	0.63	**0.70**
Entropy Summary	0.44	0.63	0.41	**0.72**
	0.34	0.31	0.32	**0.57**
Algebraic Functions	0.78	0.73	0.76	**0.88**
	0.50	**0.55**	0.49	**0.55**
Tropical Coordinates	0.84	0.71	0.84	**0.87**
	0.58	0.58	**0.60**	**0.60**
Complex Polynomials	0.74	0.72	0.73	**0.83**
	0.43	0.46	0.43	**0.51**
Betti Curve	0.03	0.62	0.03	**0.63**
	0.03	**0.58**	0.03	**0.58**
Lifespan Curve	0.32	0.59	0.33	**0.70**
	0.18	**0.52**	0.18	0.38
Pers. Landscapes	0.33	0.57	0.33	**0.66**
	0.26	0.49	0.26	**0.56**
Pers. Silhouette	0.43	0.52	0.41	**0.70**
	0.29	0.38	0.28	**0.43**
Pers. Images	0.74	0.65	0.73	**0.77**
	0.55	0.45	0.59	**0.56**
Template Functions	0.71	0.77	0.72	**0.83**
	0.37	0.65	0.44	**0.69**
ATS	0.77	0.77	0.81	**0.85**
	0.62	0.67	0.65	**0.73**
ATOL	0.79	0.75	0.78	**0.81**
	0.49	0.59	0.51	**0.69**

Table 2.3: **Outex dataset results.** Concat achieves consistently the best results. First line of each row: best result, second line: average.

	Cubical complex				**Multi filtration**			
vectorisation	**H_0**	**H_1**	**fused**	**concat**	**H_0**	**H_1**	**fused**	**concat**
Pers. Statistics	0.53	0.48	0.56	**0.61**	0.78	0.52	0.75	**0.80**
	0.53	0.48	0.56	**0.61**	0.69	0.47	0.67	**0.71**
Entropy Summary	0.15	0.32	0.16	**0.34**	0.10	**0.14**	0.13	0.12
	0.15	0.32	0.16	**0.34**	0.10	**0.13**	**0.13**	**0.13**
Algebraic Functions	0.27	0.35	0.29	**0.44**	0.80	0.50	0.79	**0.83**
	0.27	0.35	0.29	**0.44**	**0.77**	0.39	0.76	**0.77**
Tropical Coordinates	0.38	0.38	0.41	**0.51**	0.11	0.13	**0.14**	0.13
	0.37	0.37	0.41	**0.51**	0.10	**0.13**	0.12	0.12
Complex Polynomials	0.32	0.33	0.32	**0.42**	0.21	0.20	0.26	**0.31**
	0.30	0.32	0.30	**0.41**	0.16	0.17	0.20	**0.25**
Betti Curve	0.13	**0.35**	0.13	**0.35**	0.11	0.13	**0.14**	0.13
	0.13	0.34	0.13	**0.35**	0.10	0.10	**0.11**	0.10
Lifespan Curve	0.12	**0.32**	0.12	**0.32**	0.13	0.12	**0.14**	**0.14**
	0.12	0.31	0.12	**0.32**	0.11	0.11	**0.12**	**0.12**
Pers. Landscapes	0.11	0.40	0.11	**0.41**	0.11	0.14	**0.27**	0.26
	0.11	**0.39**	0.11	0.38	0.10	0.12	**0.22**	0.20
Pers. Silhouette	0.16	0.27	0.17	**0.31**	0.23	0.25	**0.30**	0.28
	0.13	0.24	0.14	**0.26**	0.18	0.18	**0.26**	0.25
Pers. Images	0.49	0.43	0.50	**0.55**	0.35	0.38	0.41	**0.44**
	0.43	0.38	0.44	**0.50**	0.24	0.26	0.33	**0.36**
Template Functions	0.36	**0.49**	0.35	0.48	0.36	0.33	0.40	**0.43**
	0.23	0.41	0.24	**0.45**	0.31	0.30	0.34	**0.35**
ATS	0.28	0.15	0.11	**0.30**	0.27	0.28	**0.34**	**0.34**
	0.21	0.15	0.11	**0.22**	0.25	0.24	0.25	**0.27**
ATOL	0.34	0.25	**0.40**	0.27	0.21	0.21	0.28	**0.29**
	0.24	0.21	**0.37**	0.23	0.14	0.14	**0.18**	0.18

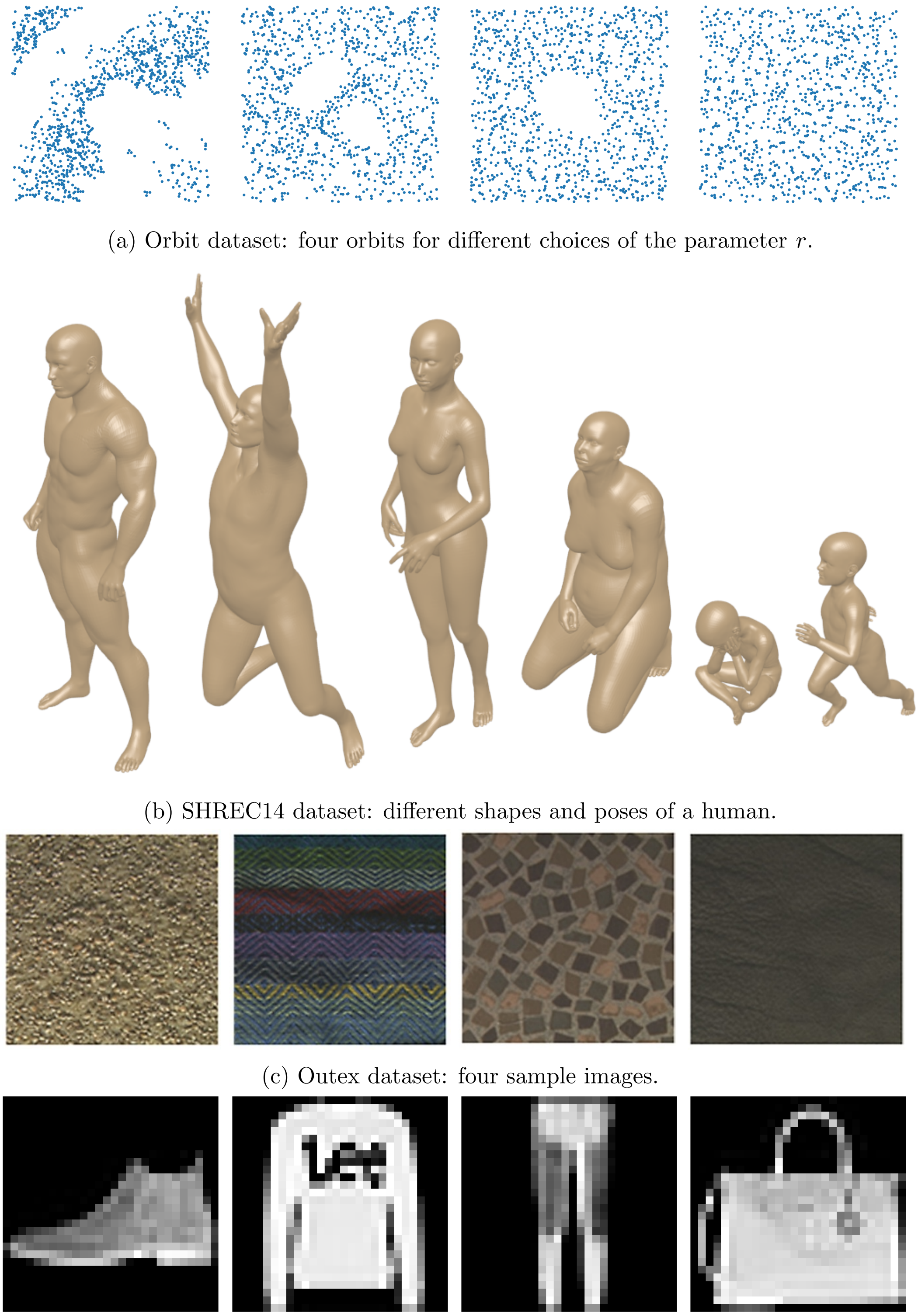
Table 2.4: **FMNIST dataset results.** Two different filtrations with very different results. In both cases, concat performs best. First line of each row: best result, second line: average.

	Edge weight filtration				
vectorisation	$\mathbf{H_0}$	$\mathbf{H_1}$	$\mathbf{H_2}$	**fused**	**concat**
Pers. Statistics	0.59	0.54	0.75	**0.76**	0.75
	0.57	0.53	**0.72**	**0.72**	0.71
Entropy Summary	0.59	0.53	0.73	0.72	**0.74**
	0.57	0.53	**0.70**	0.66	0.68
Algebraic Functions	0.59	0.54	**0.66**	**0.66**	**0.66**
	0.58	0.54	**0.60**	0.57	0.59
Tropical Coordinates	0.59	0.54	**0.68**	0.66	**0.68**
	0.58	0.52	0.58	**0.60**	**0.60**
Complex Polynomials	0.59	0.57	**0.64**	0.62	0.62
	0.58	0.55	**0.60**	0.59	**0.60**
Betti Curve	0.60	0.54	**0.71**	**0.71**	0.70
	0.60	0.53	0.62	**0.64**	0.60
Lifespan Curve	0.59	0.54	0.73	**0.74**	0.73
	0.56	0.52	0.59	**0.62**	0.61
Pers. Landscapes	0.59	0.54	0.73	0.73	**0.74**
	0.58	0.53	**0.63**	0.61	**0.63**
Pers. Silhouette	0.59	0.55	0.70	**0.71**	**0.71**
	0.58	0.54	0.66	0.65	**0.67**
Pers. Images	0.60	0.53	0.79	0.77	0.79
	0.57	0.53	**0.71**	0.70	**0.71**
Template Functions	0.52	0.49	0.68	**0.69**	0.65
	0.50	0.49	0.62	**0.64**	0.59
ATS	0.55	0.53	**0.71**	**0.71**	**0.71**
	0.52	0.51	0.66	**0.68**	**0.68**
ATOL	0.57	0.53	**0.72**	0.70	0.71
	0.55	0.52	**0.69**	0.66	**0.69**

Table 2.5: **COLLAB dataset results.** Except for H_0 and H_1, the accuracy results are high and consistent for all approaches. First line of each row: best result, second line: average.

	Takens embedding					**Lower star**
vectorisation	$\mathbf{H_0}$	$\mathbf{H_1}$	$\mathbf{H_2}$	**fused**	**concat**	$\mathbf{H_0}$
Pers. Statistics	0.56	0.56	0.55	0.64	**0.66**	0.88
	0.53	0.53	0.53	0.57	**0.58**	0.83
Entropy Summary	0.70	**0.72**	0.65	0.67	0.70	0.92
	0.67	**0.70**	0.62	0.65	0.68	0.79
Algebraic Functions	0.66	0.64	0.61	0.64	**0.68**	0.61
	0.58	0.57	0.56	0.57	**0.59**	0.56
Tropical Coordinates	0.69	**0.71**	0.66	0.70	0.70	0.54
	0.60	**0.61**	0.58	0.60	0.60	0.52
Complex Polynomials	0.67	0.70	0.65	0.69	**0.72**	0.69
	0.60	0.61	0.58	0.59	**0.64**	0.67
Betti Curve	0.69	**0.70**	0.63	0.68	**0.70**	0.86
	0.67	**0.69**	0.62	0.68	**0.69**	0.79
Lifespan Curve	0.73	0.67	0.63	**0.74**	0.73	0.68
	0.61	0.58	0.57	**0.62**	**0.62**	0.59
Pers. Landscapes	0.70	0.68	0.64	**0.73**	0.72	0.73
	0.60	0.59	0.57	**0.61**	0.60	0.61
Pers. Silhouette	0.60	0.65	0.64	**0.68**	**0.68**	0.71
	0.58	0.60	0.58	0.61	**0.62**	0.68
Pers. Images	0.50	0.56	0.50	0.57	**0.59**	0.50
	0.50	0.55	0.52	**0.57**	0.55	0.48
Template Functions	0.50	0.56	0.55	**0.58**	0.50	0.51
	0.50	0.52	0.52	**0.54**	0.50	0.49
ATS	**0.71**	0.67	0.64	0.70	**0.71**	0.85
	0.69	0.60	0.60	0.68	**0.69**	0.83
ATOL	**0.75**	0.73	0.72	**0.75**	**0.75**	**0.94**
	0.73	0.71	0.67	0.72	**0.73**	0.94

Table 2.6: **Gravitational wave dataset results.** Two different filtrations and their accuracies. When multiple approaches are available, H_0, fused and concat perform best. First line of each row: best result, second line: average.



(a) Orbit dataset: four orbits for different choices of the parameter r.

(b) SHREC14 dataset: different shapes and poses of a human.

(c) Outex dataset: four sample images.

(d) FMNIST dataset: four sample images.

Figure 2.1: Samples of the datasets used in this chapter.

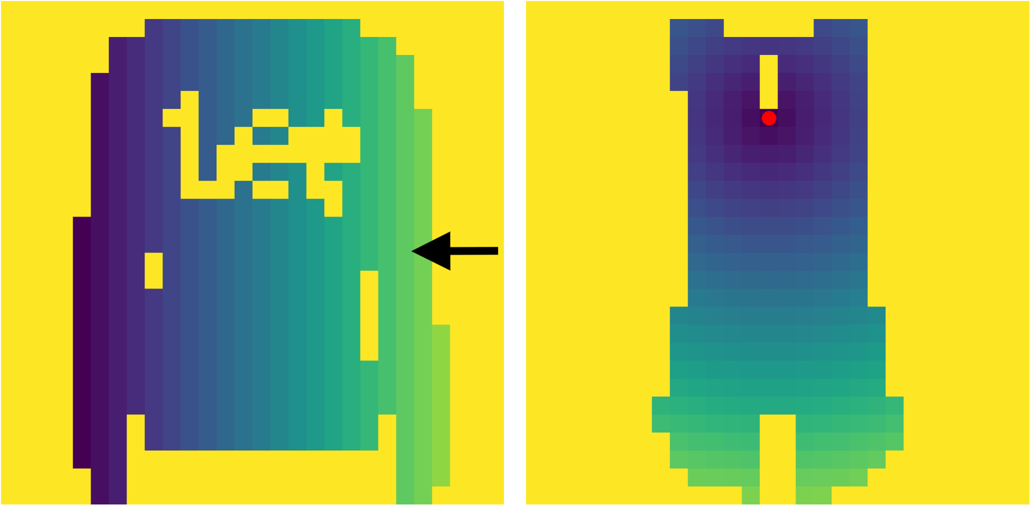


Figure 2.2: Two processed images of the FMNIST dataset. Height filtration with the arrow along which it was computed the distance (left). Center filtration with the center from which it was computed the distance to measure function (right).

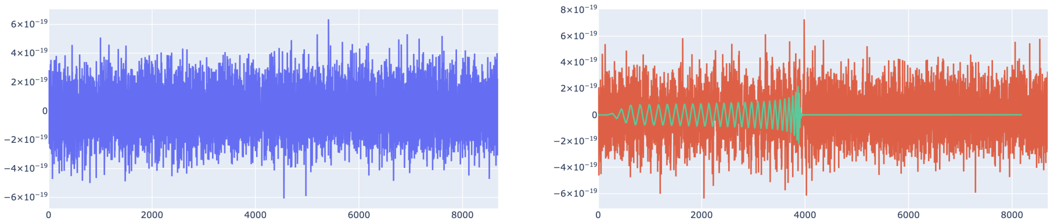


Figure 2.3: Two signals coming from the gravitational wave dataset. Pure noise signal (left), gravitational waveform (right, green) and its corrupted version (right, red).

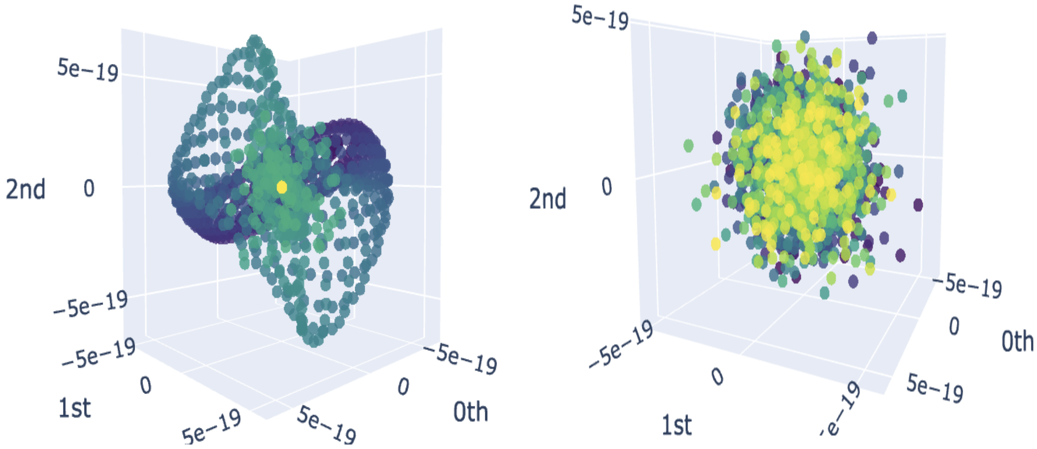


Figure 2.4: 3D projection of the embedded gravitational waveform (left) and 3D projection of the embedded noise (right).

use all homological dimensions available since f_{012} and c_{012} are superior to f_{01} and c_{01}, which are themselves superior to H_0 and H_1.

To summarize, concatenating all non-trivial homological dimensions seems to be the best approach, regardless of the noise level. If for some reason (i.e. computational cost) only fewer homological dimensions are computed, the best approach is yet to concatenate all homological dimensions available. Finally, it appears that higher dimensions are less reliable in the presence of noise. We stress the fact that these are just preliminary and empirical findings, further statistical studies should be conducted.

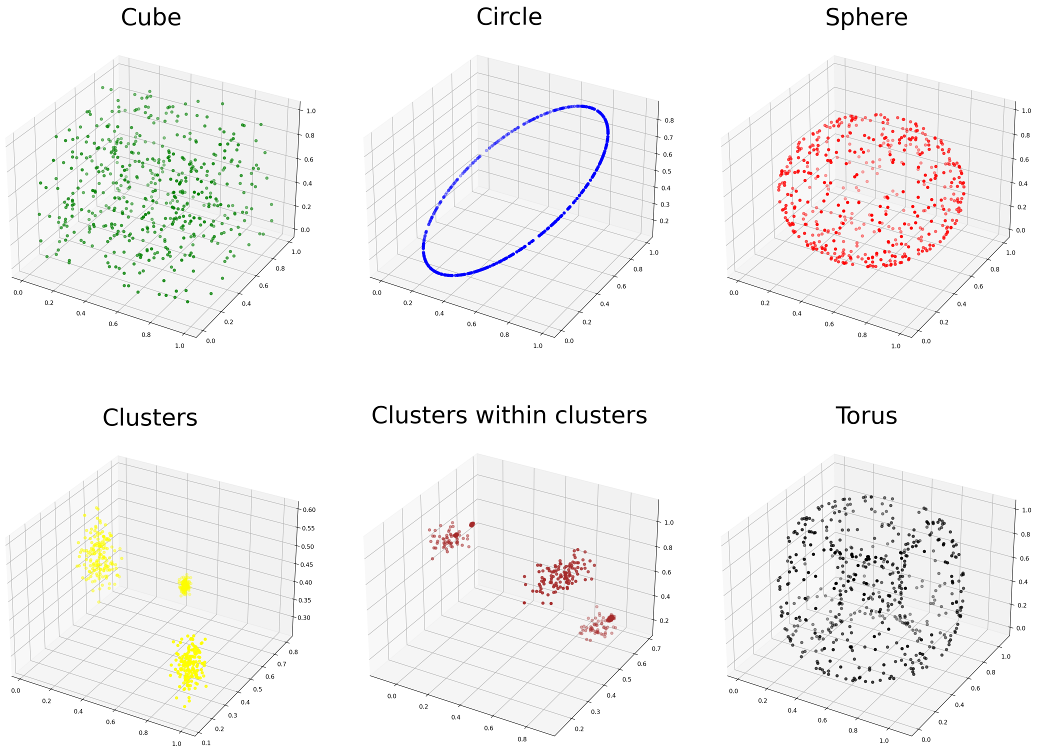


Figure 2.5: Sample of shapes for the Synthetic dataset.

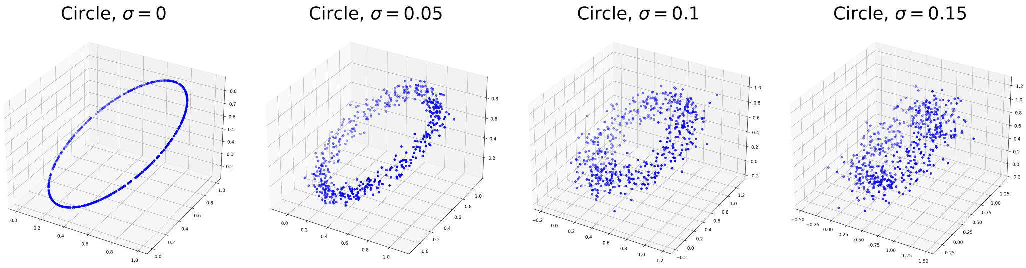


Figure 2.6: The different levels of noise corrupting the Synthetic dataset.

vectorisation	Vietoris Rips													
	$\sigma = 0$							$\sigma = 0.05$						
	H_0	H_1	H_2	f_{01}	c_{01}	f_{012}	c_{012}	H_0	H_1	H_2	f_{01}	c_{01}	f_{012}	c_{012}
Pers. Statistics	0.96	**0.99**	0.96	**0.99**	**0.99**	**0.99**	**0.99**	0.96	0.97	0.97	0.96	0.97	0.97	0.98
	0.93	**0.98**	0.92	0.96	0.97	0.94	0.96	**0.88**	0.78	**0.88**	0.78	0.79	0.81	0.79
Entropy Summary	0.91	0.90	0.91	0.91	**0.94**	0.91	**0.94**	0.86	0.84	0.86	**0.92**	**0.92**	0.88	**0.92**
	0.90	0.85	0.90	0.90	0.87	**0.91**	0.86	0.82	0.84	0.82	**0.87**	**0.87**	0.85	**0.87**
Algebraic Functions	0.96	0.92	0.95	0.95	0.96	0.94	**0.97**	0.91	0.87	0.92	**0.93**	0.92	**0.93**	0.92
	0.89	0.88	0.88	**0.93**	0.88	**0.93**	0.89	0.85	0.83	0.85	**0.91**	0.85	0.90	0.85
Tropical Coordinates	0.96	0.97	0.96	**0.99**	0.98	**0.99**	0.97	0.93	**0.97**	0.93	0.95	0.96	0.94	0.96
	0.90	0.79	0.90	**0.96**	0.74	0.94	0.80	**0.86**	0.77	**0.86**	0.77	0.66	0.80	0.71
Complex Polynomials	0.95	**0.97**	0.95	0.96	**0.97**	0.96	0.96	0.91	**0.95**	0.91	0.94	**0.95**	0.94	0.94
	0.86	**0.89**	0.86	0.86	0.84	0.85	0.84	0.84	**0.87**	0.84	0.85	0.85	0.84	0.84
Betti Curve	0.17	**0.92**	0.17	0.17	0.90	0.33	0.90	0.17	0.86	0.17	0.17	**0.88**	0.17	0.87
	0.17	**0.83**	0.17	0.17	0.80	0.33	0.80	0.17	**0.81**	0.17	0.17	0.80	0.17	0.79
Lifespan Curve	0.87	0.94	0.87	0.87	**0.95**	0.84	0.94	0.83	0.91	0.84	0.83	**0.93**	0.83	0.92
	0.82	**0.90**	0.82	0.82	**0.90**	0.79	0.89	0.81	**0.89**	0.81	0.81	0.86	0.81	0.87
Pers. Landscapes	0.17	**0.97**	0.17	0.17	**0.97**	0.17	0.96	0.17	**0.91**	0.17	0.17	**0.91**	0.17	**0.91**
	0.17	**0.91**	0.17	0.17	0.89	0.17	0.85	0.17	**0.83**	0.17	0.17	0.82	0.17	0.82
Pers. Silhouette	0.86	**0.94**	0.84	0.89	**0.94**	0.90	0.93	0.84	0.89	0.84	0.89	**0.92**	0.87	**0.92**
	0.48	0.83	0.48	0.50	**0.86**	0.50	**0.86**	0.49	0.75	0.49	0.51	**0.82**	0.51	0.81
Pers. Images	0.94	**0.99**	0.93	**0.99**	**0.99**	**0.99**	**0.99**	0.92	0.80	0.92	0.79	**0.93**	0.80	0.92
	0.77	**0.97**	0.77	0.93	0.96	0.94	0.96	0.73	0.69	0.73	0.66	**0.77**	0.67	0.76
Template Functions	0.86	0.99	1.00	0.99	0.99	0.99	1.00	0.86	**0.92**	0.91	0.68	0.90	0.74	0.91
	0.81	**0.98**	**0.98**	0.96	**0.98**	0.93	**0.98**	0.80	0.79	0.82	0.61	0.79	0.64	**0.83**
ATS	0.94	0.97	0.92	**0.99**	0.96	**0.99**	0.98	0.94	0.95	0.95	0.96	0.94	**0.97**	0.95
	0.83	0.94	0.89	0.98	0.94	**0.99**	0.95	0.80	0.92	**0.93**	0.92	0.89	**0.93**	0.90
ATOL	0.96	0.99	0.99	0.99	0.99	0.99	1.00	0.94	0.96	0.95	**0.97**	0.96	0.96	0.95
	0.86	**0.98**	**0.98**	**0.98**	**0.98**	0.96	**0.98**	0.86	**0.92**	0.91	0.77	**0.92**	0.78	**0.92**

Table 2.7: **Synthetic dataset results for noise level $\sigma = 0$ and $\sigma = 0.05$.** For low levels of noise, every approach seem equivalent, but c_{012} still achieves the best accuracy, together with H_2. First row of each cell: best result, second row: average.

vectorisation	Vietoris Rips													
	$\sigma = 0.1$							$\sigma = 0.15$						
	H_0	H_1	H_2	f_{01}	c_{01}	f_{012}	c_{012}	H_0	H_1	H_2	f_{01}	c_{01}	f_{012}	c_{012}
Pers. Statistics	0.79	0.84	0.79	0.83	0.85	0.87	0.83	0.64	0.69	0.63	0.67	0.70	0.65	0.71
	0.77	0.64	**0.77**	0.63	0.64	0.66	0.63	**0.64**	0.49	0.63	0.48	0.49	0.48	0.50
Entropy Summary	0.74	0.70	0.74	0.76	**0.85**	0.76	0.84	0.66	0.54	0.66	0.63	**0.67**	0.62	0.66
	0.72	0.66	0.72	**0.75**	**0.75**	0.74	0.74	0.64	0.50	**0.65**	0.62	0.59	0.62	0.56
Algebraic Functions	0.73	0.75	0.73	0.80	0.78	**0.84**	0.80	**0.67**	0.63	**0.67**	**0.67**	0.64	**0.67**	0.66
	0.72	0.69	0.72	0.78	0.76	**0.80**	0.77	0.65	0.60	**0.66**	**0.66**	0.64	0.65	**0.66**
Tropical Coordinates	0.75	0.81	0.75	0.76	**0.82**	0.79	**0.82**	0.65	0.67	0.66	0.62	0.67	0.67	**0.68**
	0.74	0.64	**0.75**	0.59	0.57	0.61	0.55	0.64	0.48	**0.65**	0.45	0.46	0.50	0.45
Complex Polynomials	0.75	0.78	0.75	**0.83**	0.81	0.82	0.81	0.64	0.65	0.65	0.64	0.69	0.63	**0.70**
	0.73	0.69	0.73	0.74	**0.76**	0.73	0.75	0.64	0.62	0.64	0.63	0.63	0.62	**0.65**
Betti Curve	0.17	0.67	0.17	0.17	**0.68**	0.17	0.67	0.17	0.49	0.17	0.17	**0.50**	0.17	0.46
	0.17	**0.64**	0.17	0.17	0.61	0.17	0.56	0.17	**0.45**	0.17	0.17	0.44	0.17	0.42
Lifespan Curve	0.73	0.78	0.73	0.73	**0.86**	0.73	0.83	**0.67**	0.62	**0.67**	**0.67**	0.66	**0.67**	**0.67**
	0.71	**0.76**	0.71	0.71	0.73	0.71	**0.76**	**0.64**	0.60	**0.64**	**0.64**	0.60	**0.64**	0.60
Pers. Landscapes	0.17	0.78	0.17	0.17	0.76	0.17	**0.79**	0.17	0.64	0.17	0.17	**0.67**	0.17	0.66
	0.17	**0.70**	0.17	0.17	0.64	0.17	0.66	0.17	**0.62**	0.17	0.17	0.51	0.17	0.50
Pers. Silhouette	0.72	0.70	0.72	0.74	**0.83**	0.73	**0.83**	0.64	**0.66**	0.64	0.63	**0.66**	0.62	0.64
	0.42	0.61	0.45	**0.67**	0.66	0.45	0.66	0.40	**0.55**	0.40	0.38	0.42	0.38	0.51
Pers. Images	0.73	0.68	0.73	0.57	0.76	0.58	**0.78**	**0.63**	0.48	**0.63**	0.38	0.61	0.47	0.59
	0.63	0.52	**0.65**	0.45	0.59	0.47	0.58	0.53	0.41	**0.54**	0.33	0.46	0.33	0.44
Template Functions	0.75	0.75	**0.79**	0.56	**0.79**	0.57	**0.79**	0.65	0.61	0.61	0.43	0.64	0.44	**0.67**
	0.72	0.62	0.62	0.45	0.65	0.44	0.66	**0.64**	0.45	0.43	0.38	0.42	0.32	0.48
ATS	0.73	**0.86**	0.84	0.77	0.84	0.83	0.84	0.62	0.68	0.66	0.63	0.66	0.66	0.71
	0.55	**0.85**	0.77	0.69	0.82	0.80	0.83	0.45	0.67	0.64	0.56	0.65	0.61	**0.68**
ATOL	0.83	0.83	**0.85**	0.84	0.83	0.84	**0.85**	0.65	0.67	0.63	0.61	0.66	0.64	**0.70**
	0.75	0.77	0.78	0.64	0.77	0.63	**0.81**	0.63	0.61	0.59	0.47	0.61	0.48	**0.68**

Table 2.8: **Synthetic dataset results for noise level $\sigma = 0.1$ and $\sigma = 0.15$.** For higher level of noise, singular lower dimensions gain relevance in classification. The concat approach is still the best performing. First row of each cell: best result, second row: average.

2.3 Understanding pros and cons of features from fusion and concatenation approaches

Finally, in our last experiment we want to investigate the relationship between the fused and concat approaches in more detail. In particular, we want to study from an empirical point of view how the distribution and quantity of points in the various homological dimensions of the persistence diagrams influence the usefulness of either approach. To this end, we developed four synthetic datasets that vary in the number and distribution of points. We denote them as Dataset $1-4$. In order to better control the distribution of points in the PD, in all datasets we directly synthetically generated the PDs. More in detail, each dataset has three classes C_1, C_2 and C_3 with points in homological groups H_0, H_1 and H_2. Each class C_i has a number of points in each homological dimension which is a random integer in the intervals $[0, n_0], [0, n_1], [0, n_2]$ for H_0, H_1 and H_2 respectively. Each homological dimension of each class samples points in the box $b_1, b_2, b_3 \in \mathbb{R}^2$. For ease of notation, with $b_i = [x, y]$ we mean that points are sampled from the triangle $\{(b, d)\}$, with $b \in [x, y]$ and $d \in [b, y]$. Finally, each class is corrupted with random Gaussian noise with standard deviation σ_1, σ_2 and σ_3. For a list of parameters used to generate the four datasets, we refer the reader to Table 2.9. We refer to Figure 2.7 for a graphical example of the different datasets. Each dataset is composed of 900 samples. The idea behind these combinations is that Dataset 1 is composed of classes with well-separated homological dimensions and with a comparable number of points within the homological dimensions. Dataset 2 has classes with very mixed homological dimensions but a comparable number of points in the homological dimensions. Dataset 3 has points in H_0 that are orders of magnitude higher than H_1 and H_2, but well-separated points. Finally, Dataset 4 has again points in H_0 that are orders of magnitude higher than H_1 and H_2 and additionally very mixed homological dimensions. We refer to Table 2.10 for the accuracy results of this experiment. In Dataset 1 we observe that concat and fused results are very close, although concat being slightly better and more consistent. In particular, it is clear that for this dataset fused is able to synergise information coming from all dimensions. For Dataset 2, however, the fused approach is rarely able to synergise the various homological dimensions to increase their singular performances. When it happens, it is only marginally. This is supposedly due to the fact that points in different homological dimensions are more mixed with respect to that in Dataset 1. The results for Dataset 3 are in a sense similar to those of Dataset 1, with one major difference. Again, concat and fused approaches are the best-performing methods. The concat approach is nonetheless the best performing. It is interesting to note however that the gap between the first row of each cell and the second row is wider than in Dataset 1 and 2, both for the fused and the concat approaches. That is, in the presence of homological dimensions with a huge disparity of points, using all dimensions became very sensitive to optimal parameters choice. Finally, the results with Dataset 4 are consistent with our previous findings. In particular, the gap between fused and approach is more pronounced, and again with different orders of magnitude of points in the various homological dimensions, using features from all of them became way more sensible to the parameters than in the setting with a similar amount of numbers.

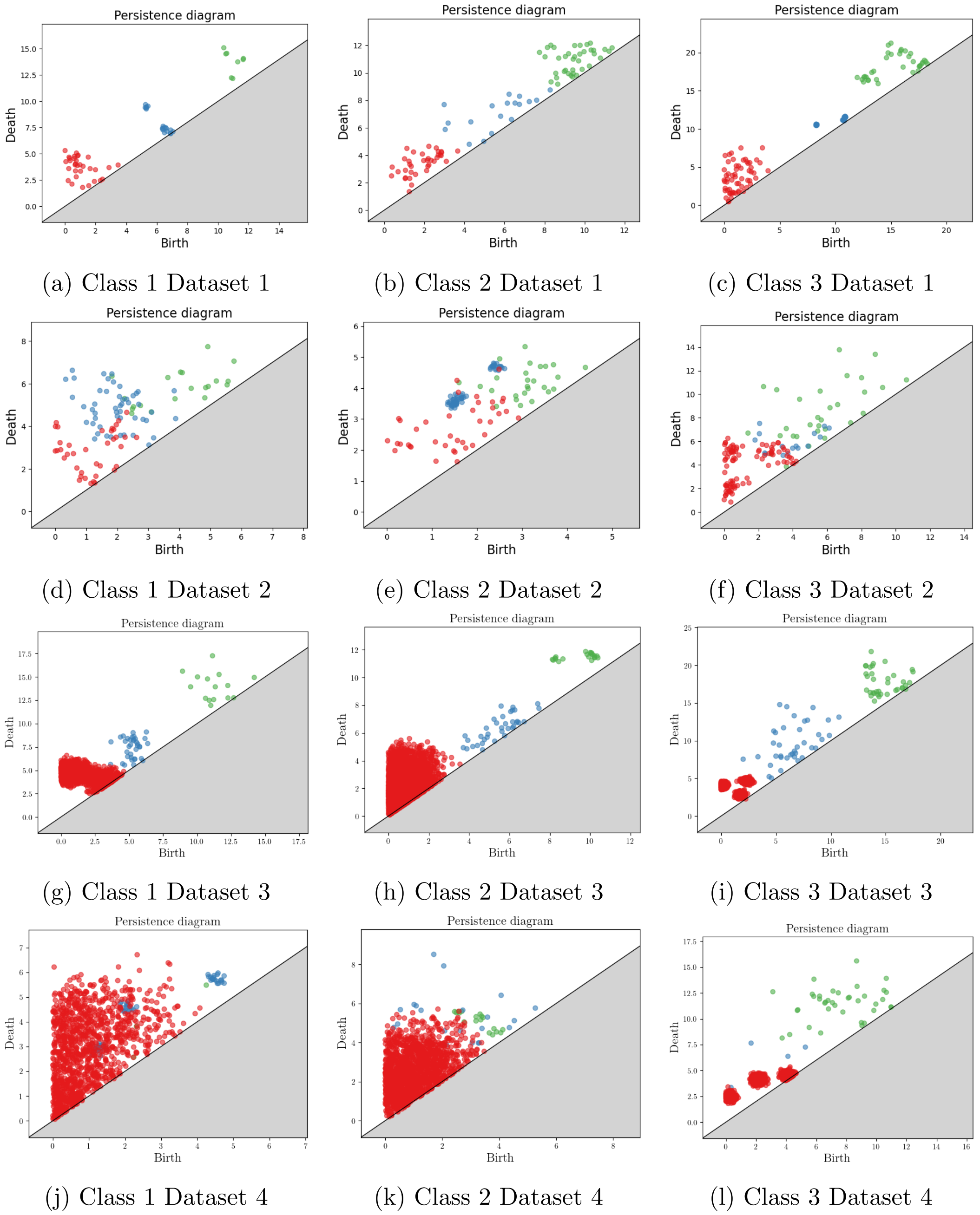
	Dataset 1	**Dataset 2**	**Dataset 3**	**Dataset 4**
$\mathbf{C_1 : (n_0, n_1, n_2)}$	$(100, 60, 20)$	$(100, 60, 20)$	$(3120, 60, 20)$	$(1600, 60, 20)$
$\mathbf{C_2 : (n_0, n_1, n_2)}$	$(120, 60, 40)$	$(120, 60, 40)$	$(2120, 60, 40)$	$(1620, 60, 40)$
$\mathbf{C_3 : (n_0, n_1, n_2)}$	$(120, 60, 40)$	$(120, 60, 40)$	$(3120, 60, 40)$	$(1620, 60, 40)$
$\mathbf{C_1 : (b_0, b_1, b_2)}$	$([0, 5], [5, 10], [10, 15])$	$([0, 5], [1, 6], [2, 7])$	$([0, 5], [5, 10], [10, 15])$	$([0, 5], [1, 6], [2, 7])$
$\mathbf{C_2 : (b_0, b_1, b_2)}$	$([0, 4], [4, 8], [8, 12])$	$([0, 4], [1, 5], [2, 6])$	$([0, 4], [4, 8], [8, 12])$	$([0, 4], [1, 5], [2, 6])$
$\mathbf{C_3 : (b_0, b_1, b_2)}$	$([0, 6], [6, 12], [12, 20])$	$([0, 6], [2, 8], [4, 12])$	$([0, 6], [6, 12], [12, 20])$	$([0, 6], [2, 8], [4, 12])$
$(\sigma_1, \sigma_2, \sigma_3)$	$(1, 1.5, 2)$	$(1, 1.5, 2)$	$(1, 1.5, 2)$	$(1, 1.5, 2)$

Table 2.9: Parameters combination for the four datasets of pros and cons of fusion and concat approaches.

To conclude the chapter, our experiments show that the concat approach is undoubtedly more reliable and consistently achieves the best accuracy results. More importantly, it never suffers from a drop of performance related to specific datasets - filtrations - vectorisations, in contrast with all other approaches. The presence of noise seems to impact it in a more limited way than the other approaches. Especially, higher dimensions seem to be more affected by noise, while concat is still very consistent. However, the concat approach can be more computational expensive and this factor may be limiting in certain scenarios. The fused approach can be an exellent alternative, easy to vectorise as a singular dimensions, but it can suffer from sharp drop of performances when the homological dimenisons are not well separated or with number of points that differs by orders of magnitude. It is much more sensible to the parameters combination but is usually able to improve the quality of features from individual homological dimensions. All these results must be considered preliminary. In any case, the premises of the topological machine learning pipeline devised alongside this dissertation are excellent and its application to real-world data is the focus of the next chapter.

	Dataset 1					Dataset 2					Dataset 3					Dataset 4				
	H_0	**H_1**	**H_2**	**fused**	**concat**	**H_0**	**H_1**	**H_2**	**fused**	**concat**	**H_0**	**H_1**	**H_2**	**fused**	**concat**	**H_0**	**H_1**	**H_2**	**fused**	**concat**
PS	0.69	0.91	0.99	0.89	1.00	0.70	0.75	0.84	0.83	0.90	0.78	0.92	0.98	0.93	**0.99**	0.71	0.78	0.86	0.82	0.94
	0.65	0.91	**0.99**	0.83	**0.99**	0.65	0.74	0.82	0.79	**0.86**	0.64	0.91	**0.98**	0.51	0.72	0.51	0.75	**0.85**	0.58	0.72
ES	0.37	0.37	0.43	**0.56**	0.43	0.34	0.34	0.46	**0.54**	0.44	0.43	0.40	0.46	**0.62**	0.46	0.42	0.43	0.49	**0.60**	0.49
	0.37	0.35	0.41	**0.56**	0.41	0.33	0.32	0.43	**0.51**	0.43	0.42	0.39	0.42	**0.49**	0.43	0.40	0.40	0.47	**0.58**	0.46
AF	0.60	0.62	0.71	**0.79**	0.77	0.61	0.63	0.71	0.74	**0.76**	0.63	0.62	0.75	**0.81**	0.79	0.63	0.64	0.77	0.76	**0.82**
	0.56	0.58	0.64	**0.74**	0.71	0.59	0.60	0.67	0.69	**0.72**	0.59	0.60	0.66	**0.76**	0.71	0.56	0.62	**0.68**	0.66	**0.68**
TC	0.56	0.54	0.67	0.67	**0.73**	0.52	0.53	0.70	0.67	**0.72**	0.62	0.51	0.69	0.60	**0.70**	0.60	0.56	0.70	0.74	**0.79**
	0.53	0.51	0.62	0.61	**0.70**	0.49	0.49	0.68	0.65	**0.71**	0.59	0.49	0.64	0.55	**0.65**	0.59	0.51	0.67	0.66	**0.69**
CP	0.53	0.66	0.71	0.72	**0.82**	0.59	0.58	0.66	0.58	**0.74**	0.66	0.65	0.68	0.70	**0.76**	0.67	0.67	0.72	0.79	**0.81**
	0.41	0.43	0.64	0.53	**0.68**	0.42	0.44	0.45	0.50	**0.64**	0.58	0.59	0.59	0.67	**0.71**	0.60	0.58	0.67	0.70	**0.72**
BC	0.35	0.35	0.43	**0.56**	0.44	0.39	0.37	0.44	**0.52**	0.47	0.43	0.39	0.47	**0.60**	0.44	0.43	0.40	0.48	**0.62**	0.47
	0.34	0.34	0.41	**0.55**	0.40	0.35	0.35	0.43	**0.50**	0.44	0.39	0.38	0.45	**0.47**	0.41	0.41	0.36	0.46	**0.60**	0.44
LC	0.45	0.41	0.56	**0.66**	0.59	0.46	0.39	**0.59**	0.52	0.57	0.51	0.46	0.56	**0.69**	0.63	0.51	0.44	0.53	**0.63**	0.60
	0.40	0.39	0.54	**0.65**	0.57	0.42	0.38	**0.58**	0.51	0.57	0.50	0.42	0.53	**0.60**	0.56	0.48	0.41	0.52	**0.61**	0.52
PL	0.61	0.51	0.64	**0.75**	0.73	0.60	0.52	0.65	**0.74**	0.73	0.65	0.52	0.67	**0.75**	**0.75**	0.63	0.58	0.64	0.69	**0.74**
	0.57	0.49	0.63	0.70	**0.71**	0.56	0.50	0.63	**0.70**	0.69	0.62	0.51	0.61	**0.72**	0.70	0.59	0.56	0.61	0.67	**0.68**
PSi	0.57	0.49	0.58	**0.66**	0.63	0.53	0.48	0.54	**0.65**	0.64	0.62	0.50	0.54	**0.73**	0.63	0.54	0.49	0.57	0.65	**0.66**
	0.52	0.45	0.55	**0.65**	0.60	0.50	0.44	0.53	**0.62**	0.60	0.56	0.45	0.52	**0.64**	0.62	0.51	0.45	0.56	0.61	**0.63**
PI	0.51	0.51	0.59	**0.76**	0.68	0.52	0.54	0.66	**0.68**	**0.68**	0.63	0.49	0.62	**0.79**	0.70	0.60	0.56	0.62	0.71	**0.77**
	0.46	0.44	0.54	**0.70**	0.57	0.49	0.50	0.57	**0.65**	0.62	0.58	0.48	0.55	**0.68**	0.62	0.59	0.52	0.58	0.60	**0.64**
TF	0.67	0.89	0.99	0.98	1.00	0.68	0.76	0.80	0.81	**0.87**	0.77	0.91	0.97	0.96	**0.99**	0.73	0.75	0.83	0.82	**0.92**
	0.64	0.81	0.94	0.86	**0.96**	0.65	0.69	0.79	0.75	**0.82**	0.73	0.86	**0.95**	0.77	0.79	0.66	0.65	**0.81**	0.74	**0.81**
ATS	0.66	0.86	0.97	0.98	**0.99**	0.69	0.72	0.78	0.81	**0.87**	0.72	0.82	0.97	0.93	**0.99**	0.71	0.71	0.82	0.79	**0.88**
	0.65	0.84	0.96	0.91	**0.96**	0.64	0.67	0.75	0.76	**0.83**	0.70	0.80	**0.95**	0.73	0.87	0.66	0.68	**0.80**	0.73	0.73
ATOL	0.57	0.80	0.95	0.78	**0.98**	0.58	0.67	0.81	0.76	**0.86**	0.66	0.85	0.98	0.62	1.00	0.62	0.67	0.80	0.62	**0.90**
	0.51	0.73	0.94	0.57	**0.95**	0.51	0.60	0.79	0.70	**0.84**	0.60	0.77	**0.95**	0.57	0.81	0.56	0.63	**0.80**	0.58	0.75

Table 2.10: **Four synthetic datasets of pros and cons of fusion and concat approaches results.** For well separated classes with a comparable number of points, the fused approach is able to synergise the various homological dimensions. When this is not the case, howerver, it suffers a drop in performance. First row of each cell: best result, second row: average.



(a) Class 1 Dataset 1

(b) Class 2 Dataset 1

(c) Class 3 Dataset 1

(d) Class 1 Dataset 2

(e) Class 2 Dataset 2

(f) Class 3 Dataset 2

(g) Class 1 Dataset 3

(h) Class 2 Dataset 3

(i) Class 3 Dataset 3

(j) Class 1 Dataset 4

(k) Class 2 Dataset 4

(l) Class 3 Dataset 4

Figure 2.7: Examples of classes $C1-3$ for Datasets $1-4$.

Chapter 3

Real-world dataset applications

The aim of this chapter is to present three applications of the topological machine learning pipeline described in Chapter 1 and tested on benchmark datasets in Chapter 2 to real-world datasets that are under analysis ISTI-CNR laboratory. This chapter represents the third contribution of this dissertation and it is derived from [16, 17, 18]. This is motivated by the good results of the pipeline on benchmark datasets, and also expand the literature of TDA applications to real-world data, which is still under-represented. The applications presented in this chapter are three, two of which are biomedical. Following the findings of the previous chapter, we are only going to present the results obtained by the pipeline for the concat approach (when applicable). We want to highlight the fact that all three applications deal with datasets heavily affected by noise, making the findings of noise robustness of the concat approach of the previous chapter even more relevant. Of course, the results are nonetheless impacted by the presence of such noise. However, results are still satisfactory, expecially in biomedical data.

3.1 Sea surface temperature analysis

The first application that we are going to present comes from the analysis of Sea Surface Temperatures (SST) and has already been presented [18]. This application is justified by the recent advances in remote sensing that provide experts with a huge amount of marine observation acquired by satellite sensors. Consequently, the need for automatic methods is increased and this application aims to provide precisely this automatic pipeline. The case of study consists in the classification of the upwelling regimes of the Iberia - Canary Current System (ICCS), one of the least studied among the mesoscale ecosystems [101]. Upwelling is a process of particular interest because it causes the transportation of deeper, colder and nutrient-rich waters to the surface. Hence, it affects the biological parameters of the habitat and enhances local biodiversity [102]. Sea surface temperature is the measure of the water's temperature performed by satellite instruments that record the energy emanating from the ocean surface, which is emitted at different wavelengths. Studying the sea surface data is important because it allows to understand changes in the environment and consequently changes in the access to food, migration patterns and mating access of the species. To the best of our knowledge, very few solutions have been developed to tackle the automation of the upwelling event classification [103, 104, 105].

3.1.1 Dataset description

In this experiment we collect satellite imagery from two satellite sources: the EUMETSAT's METOP-A and METOP-B [106] and NASA's Aqua [107]. As a preliminary step, a visual inspection of SST maps of the southwestern region of the Iberian Peninsula has been performed by experts. This leads to the identification of four typologies of mesoscale events as the most representative. The first mesoscale pattern $E1$ is associated with the meander of the southward upwelling jet to the west, near Cape St. Vincent, alongside the development of upwelling filaments. The second mesoscale pattern $E2$ is depicted by the southwards flow of the upwelling jet overpassing the Cape St. Vincent forming an extended meridional filament. Pattern $E3$ is characterized by a clear line of cool water throughout the whole southern Iberian coast. To be more precise, experts distinguish two sub-types in $E3$ [103], but we do not consider this split in our application. Finally, pattern $E4$ occurs when a warm countercurrent develops near the southern Iberian coast, surrounding Cape S. Vincent, and flowing north near the coast. A selection of 503 images (381 METOP, 122 Aqua) from years 2009 to 2016 has been collected and manually classified by experts in the mesoscale patterns $E1, E2, E3, E4$. The resulting dataset is balanced. The spatial resolution of the satellite sources is of one nadir of 1km, and temperature accuracy of 0.01°C (METOP) or 0.005°C (Aqua), with a range from -2°C to 36°C. The files were provided in either NetCDF-4 or HDF format (the latter only for pre-2014 Aqua files) and converted into 8-bit grayscale PNG images. In particular, the following steps were performed:

- information about the latitude, longitude and temperature value was extracted from the NetCDF/HDF file and stored in three `NumPy` arrays;
- a Cartopy GeoAxis was prepared with a Plate Carrée projection and an extent of $[36°\text{N}, 39.5°\text{N}] \times [10.5°\text{W}, 7°\text{W}]$;
- a grayscale colormap was defined such that a temperature of 5°C corresponds to gray 95%, a temperature of 25°C corresponds to gray 0% and the in-between values are linearly interpolated. The white color has been assigned to missing or low-quality data;
- the temperature was plotted in the GeoAxis using MatPlotLib's `pcolormesh` Python method (normalized between 5°C and 25°C) and saved using MatPlotLib's `savefig` Python method, with a 0.2-inch white padding, resulting in a 409 × 409 PNG image (370 × 370 without the white frame).

We highlight the fact that the thermal resolution of the raw data is 0.01°C at least, but when the temperature map is converted into a PNG file, such a resolution might be lower. The dataset is composed of images that are heavily affected by noise or vast areas of missing data (clouds). In the vast majority of cases, only around half of the sea surface is visible. We refer to Figure 3.1 for an example of a neat image and one corrupted by noise. In order to enhance the signal of each image, and limit the incidence of noise, we performed the following preprocessing steps:

- the Iberian peninsula is filtered out using a black mask;
- multi-threshold Otsu [108] with a 5-class segmentation (the Scikit-Image implementation is applied);
- median filtering with kernel size 7 followed by Gaussian filtering with kernel size 3 (OpenCV implementation);
- fat edge extraction using the inbuilt `ImageFilter.find_edge` function of the PIL Python package, kernel size 3.

We refer to Figure 3.2 for a graphical example of the performed preprocessing.

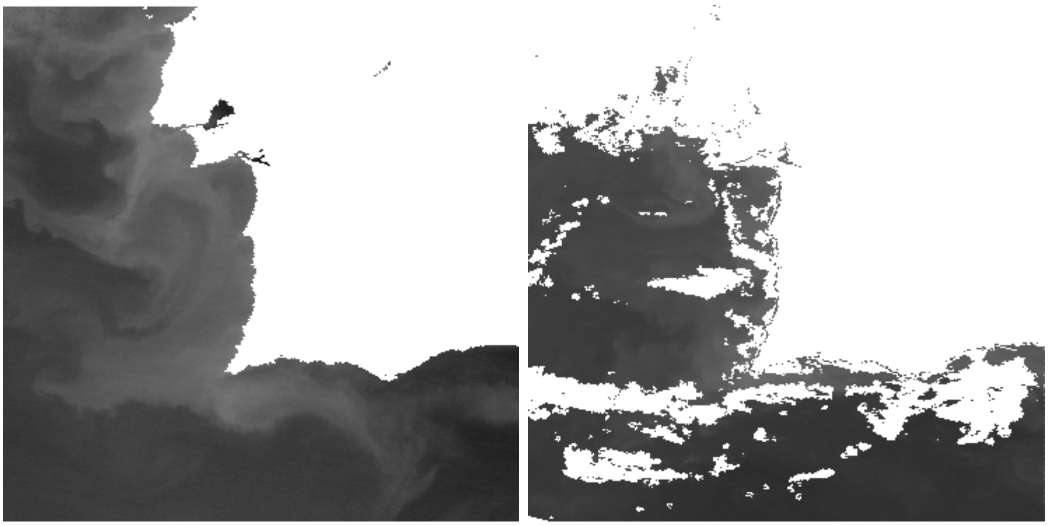


Figure 3.1: Sample images from the dataset of SST. A neat image (left) and a corrupted image (right).

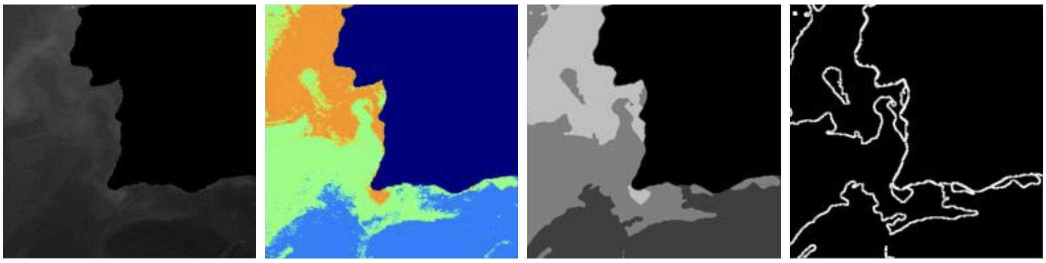


Figure 3.2: The preprocessing applied to the SST dataset. In order: The Iberian peninsula is filtered out, multi-threshold Otsu, medial filter and Gaussian filter and fat edge extraction.

		Predicted			
		$E1$	$E2$	$E3$	$E4$
True	$E1$	7	3	8	7
	$E2$	2	2	13	8
	$E3$	0	1	24	0
	$E4$	0	0	2	23

Table 3.1: Confusion table for SST dataset.

3.1.2 Results

The topological machine learning pipeline has been applied to the processed dataset described in the previous section. The best-performing method achieves a 56% overall accuracy using Betti curves. We refer to Table 3.1 for the confusion matrix of the best-performing method. The accuracy is clearly not particularly impressive and the classification of $E1$ and $E2$ in particular can be improved. For comparison, we train two competitors in the same scenario. As a fist competitor we trained a convolutional neural network, state-of-the-art in computer vision tasks [109]. We employed two architectures, both with 6 convolutional layers. The first configuration consists of $[32, 64, 128, 128, 256, 256]$ kernels of dimension 3×3. The second configuration consists of $[4, 8, 16, 16, 32, 32]$ kernels of dimension 3×3. Both CNNs are not able to improve the accuracy of our model, since the accuracy was of 34% and of 45%. Likely, such results are impacted by the scarcity of data, but still our method outperformed CNNs model. Secondly, we performed standard machine learning to the dataset. More in detail, we used the same classifiers as in the topological machine learning pipeline, without the topological part. In this case, the accuracy result was of 51%. The result of our topological machine learning pipeline, despite not being particularly satisfactory in terms of accuracy, is at least encouraging. Topological descriptors extracted from SST maps can provide support in the detection of $E3$ and $E4$ patterns, showing robustness against noise and missing signal.

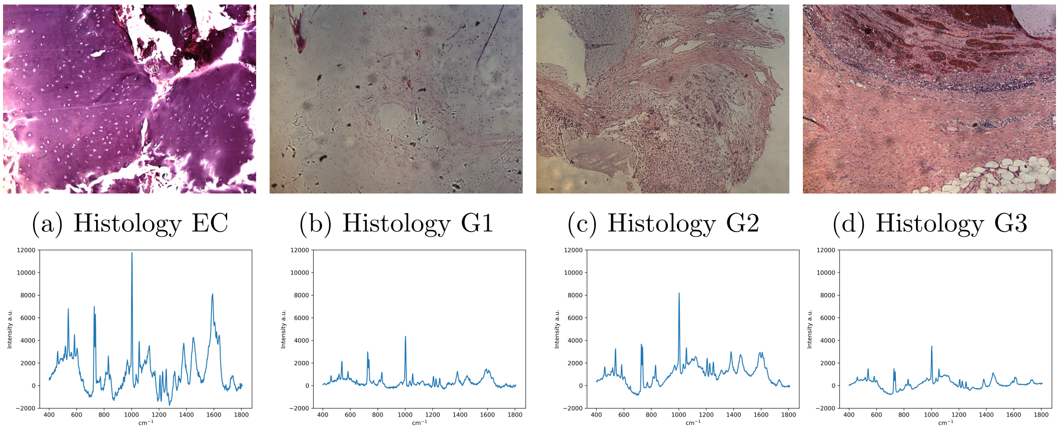
3.2 Raman spectroscopy for cancer grading

The second application comes from [16] and aims to develop an automated pipeline for cancer grading using Raman Spectroscopy (RS). Raman spectroscopy is a non-invasive optical technique sensitive to the molecular composition of biological tissues. In particular, RS can be used to optically probe the molecular changes associated with diseased tissues. The Raman spectrum is a plot of scattered intensity as a function of the energy difference between the incident and scattered photons and is obtained by pointing a monochromatic laser beam at the tissue under investigation. Hence, the loss or gain in the photon energies corresponds to the difference in the final and initial vibrational energy levels of the molecules belonging to the specific spots of the tissue investigated. The difference between final and initial vibrational energy levels denote shifts in wave-numbers, which are unique for individual molecules resulting in specific peaks that are spectrally narrow and potentially associated with the vibration of a specific chemical bond in the molecules [110]. The grading of cancer tissues is currently one of the main challenges for pathologists and RS can provide the support needed for making diagnoses more accurate and less invasive [111, 112, 113, 114, 115]. In this study, we aim to provide insights for the grading of chondrogenic tumors. Chondrogenic tumors are the second worldwide largest group of bone tumors and its malignant cells produce a cartilaginous matrix. When they occur in previously normal bones, they are generally classified as primary chondrosarcomas. At the same time, secondary chondrosarcomas result from the malignant transformation of a benign cartilaginous lesion. They are classified into three malignant degrees, the first degree (CS G1), the second one (CS G2) and the third one (CS G3). In addition to such three degrees, Enchondroma (EC) is a noncancerous version. Distinguishing between EC and CS G1 is a rather critical issue for pathologists, generating many false positive and false negative diagnoses [116, 117]. A first attempt to exploit RS for chondrogenic cancer grading has been performed [118] and was later expanded [119]. In this work, we are going to apply the topological machine learning pipeline to a dataset of Raman spectra under analysis at ISTI-CNR laboratory. The work was approved by the local Ethical Committee Comitato Etico Regionale per la Sperimentazione Clinica della Regione Toscana sezione AREA VASTA NORD OVEST (protocol number 14249).

3.2.1 Dataset description

The data acquisition was carried out with a Thermo Fisher Scientific DXR2xi Raman microscope. A total of ten patients, who were being treated at the Institution, Azienda Ospedaliera Universitaria Pisana, Pisa, were enrolled in the study under the Ethical Committee agreement. More details can be found in [118]. Formalin-fixed paraffin-embedded tumor tissue sections were collected on glass slides and subsequently submitted to RS analysis after the dewaxing step. The protocol to remove paraffin and formalin has provided the immersion of the histopathological sections in a series of two baths of xylene for 10 minutes, respectively, and then washing the sections in PolyButylene Succinate (PBS) to remove residual formalin. The Raman spectroscopy measurements were configured based on the following experimental parameters: laser wavelength 532nm; power laser of 5–10mW; 400–3400cm^{-1} full range grating; $10\times, 50\times$ and $100\times$ objectives; 25μm pinhole; 5 (FWHM) cm^{-1} spectral resolution. Integration time for recording a Raman spectrum was 1s and 10 scans for any spectrum. As a first step, the tissue morphology overview was carried out to identify the regions of interest with the collection of a number of mosaic images at low ($10\times$) and intermediate ($50\times$) magnification. Thus, the acquisition of Raman spectra was carried out with a $100\times$

objective. Optimization of signal-to-noise ratio and minimization of sample fluorescence were obtained through preliminary measurements in order to set the best experimental parameters. Multiple measurements were performed in different regions within the various samples, in order to assess intra-sample variability. In turn, no pre-treatment of the samples was necessary before Raman measurements. Minimal preprocessing, including background removal and baseline application, was performed using the tools of the DXR2xi GUI, and a 5th order polynomial correction was used to compensate for the tissue fluorescence. Peaks were identified with specific tool support by Omicron 9.0 software. Raman hyperspectral chemical maps ranging from $50 \times 50\mu\text{m}^2$ (step size 1μm) to approximately $200 \times 200\mu\text{m}^2$ (step size 4μm), recording several hundreds of spectra per map were collected. Raman maps provide the fundamental advantage of being able to localize Raman spectra to specific locations, providing local information about chemical composition. Step sizes were chosen to have a collection time for each map less than 7 hours for all the maps. Ten supplemental spectra have been acquired, making use of an Xplora Plus (Horiba) in a similar experimental setup and preprocessing procedure in order to test the classification method on never-seen data samples. This way, the results of the final test show that the classification method proposed is neither subject-dependent nor vendor-specific (DXR Thermo Fisher data for model training, Xplora Horiba data for final model testing). The dataset is balanced. Moreover, when computing the persistence diagrams we restricted to the wavenumber range 400-1800cm^{-1} and opted for a Vietoris-Rips filtration. We refer to Figure 3.3 for examples of the dataset. Figures 3.3a, 3.3b, 3.3c and 3.3d show histological images of the tumors with the different degrees of malignancy, while Figures 3.3e, 3.3f, 3.3g and 3.3h show the respective Raman spectra.



(a) Histology EC (b) Histology G1 (c) Histology G2 (d) Histology G3

(e) RS EC (f) RS G1 (g) RS G2 (h) RS G3

Figure 3.3: Representative histologic images of the tumors analyzed in this study and the respective Raman spectra.

3.2.2 Results

For this study, we performed multiple experiments but we are going to present only the two most relevant in this dissertation. For a more detailed description of all performed experiments, we refer the reader to [16]. As stated in the previous section, we had access to two datasets. One composed of 400 spectra and one composed of 10 spectra obtained from a different equipment and at a different time. For the experiments that we are going to present, the first dataset is used as training and the second one is used as test. More in detail, we report two different experiments, one with all four labels of malignancy degrees and one with just two labels: EC (benign) *vs.* CS (malignant). Table 3.2a reports the confusion matrix for the four labels experiment, with an accuracy of 80%. Table 3.2b reports the confusion matrix for the two labels experiment, with an accuracy of 90%. These results are remarkable and, together with the limitations of the dataset, show the potential of this method in large-scale applicability. Moreover, it is shown in [16] that the features extracted from TDA are more convenient (both for accuracy and significance) for a machine learning study than using directly the Raman spectra or neural networks, highlighting the power of TDA in such challenging scenarios. Also, results are very promising with respect to the state of the art, as the classification accuracy outperforms the best results in literature [119]. Due to the size of the dataset, our results should be considered preliminary but significant. Finally, the proposed pipeline provides a classification model that can be easily integrated into a workflow (as already done in the commercial workstation as for the preprocessing modules), enabling the reduction of time and cost of the grading of cancerous tissues.

Table 3.2: Confusion tables of the Raman spectroscopy for cancer grading experiment.

(a) Confusion table for the 4 degrees classification.

		Predicted			
		EC	CS G1	CS G2	CS G3
True	EC	2	0	0	0
	CS G1	1	2	0	0
	CS G2	0	0	3	0
	CS G3	1	0	0	1

(b) Confusion table for the 2 degrees classification.

		Predicted	
		EC	CS
True	EC	1	1
	CS	0	8

3.3 Raman spectroscopy for Alzheimer's disease detection

Alzheimer's disease (AD) is the most common neurodegenerative disease and, due to the population aging, its rate of affliction is likely to increase. At present, the clinical diagnosis of AD requires a series of neurological examinations (National Institute of Aging – Alzheimer's Association criteria) but the definitive diagnosis is possible only after the patient's death and brain tissue analysis. Therefore, there is a need to improve the accuracy of clinical diagnosis with innovative, cost-effective and specific approaches. As seen in the previous section, RS represents a fast, efficient, non-invasive diagnostic tool and its high-precision detection is expected to reduce or replace other AD diagnostic tests. Recently, Raman-based techniques demonstrated significant potential in identifying AD by detecting specific biomarkers in body fluids [120, 121]. The detection of Cerebrospinal Fluid (CSF) biomarkers is one of the diagnostic criteria for AD [122] because CSF is more sensitive than blood or other biofluids in the diagnosis of AD. Therefore, RS can be used as an effective tool to analyze CSF samples [123, 124]. In this work, we propose a novel method based on the collection of the vibrational Raman fingerprint of the proteomic content of cerebrospinal fluid and the topological machine learning pipeline presented in Chapter 1 in order to support the AD diagnosis. This study is an expanded version of [17].

3.3.1 Dataset description

The study population is made of 43 patients, enrolled in the framework of the Bando Salute 2018 PRAMA project ("Proteomics, RAdiomics and Machine learning-integrated strategy for precision medicine for Alzheimer's"), co-funded by the Tuscany Region, with the approval of the Institutional Ethics Committee of the Careggi University Hospital Area Vasta Centro (ref. number 17918_bio). All of them showed pathological symptoms: the majority of them have been diagnosed with AD, while the others have been considered as controls (noAD), even if diagnosed with other neurological conditions (vascular dementia, hydrocephalus and multiple sclerosis). The CSF samples were collected by lumbar puncture, then immediately centrifuged at 200g for one minute, 20°C and stored at -80°C until analysis [125, 126]. On the day of analysis, CSF samples were thawed and centrifuged again at 4000g for ten minutes at 4°C. The pellet was separated from the supernatant and further used for the analyses. A 2μl drop of the pellet was deposited onto a gold mirror support (ME1S-M01; Thorlabs, Inc., Newton, NJ), followed by air drying for 30 minutes and acquisition of Raman spectra from the outer ring of the dried drop. A set of five Raman spectra have been collected for each drop-casted sample by using a micro-Raman spectrometer (Horiba, France) in back-scattering configuration, equipped with a laser excitation source tuned at 785nm (40mW power, 20 second integration time, 10 accumulations) and a Peltier cooled CCD detector. For each patient, the average of the five acquisitions of the raw Raman spectrum is computed. This resulted in a dataset of 43 acquisitions of RS: 21 belonging to the AD class and 22 to the noAD class. The following preprocessing steps have been applied:

- baseline correction with parameters $l = 1\text{e}{+}7$ for smoothness and $p = 0.05$ for asymmetry;
- signal smoothing using the `scipy savgol_filter` Python package and parameters $w = 9$ for window and $p = 2$ for polynomial order;
- autocorrelation transform using `numpy.correlate` built-in function.

We refer to Figure 3.4 for a visualization of the preprocessing steps and to Figure 3.5 for a visual example of the final dataset. On the left side, there is the entirety of the dataset, on the right side the average of both classes with standard deviation. The filtration used in this application is the lower star, which results only in H_0 features. The validation scheme is the Leave One Out cross-validation (LOO) [127].

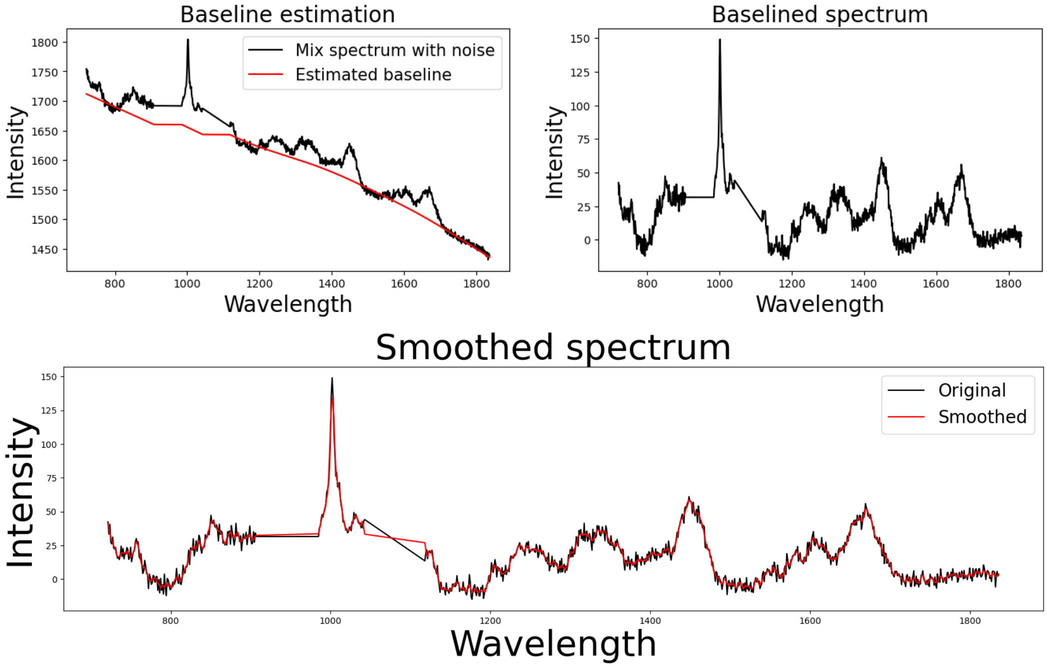


Figure 3.4: Alzheimer's disease RS dataset. The entirety of the dataset (left) and average with standard deviation (right).

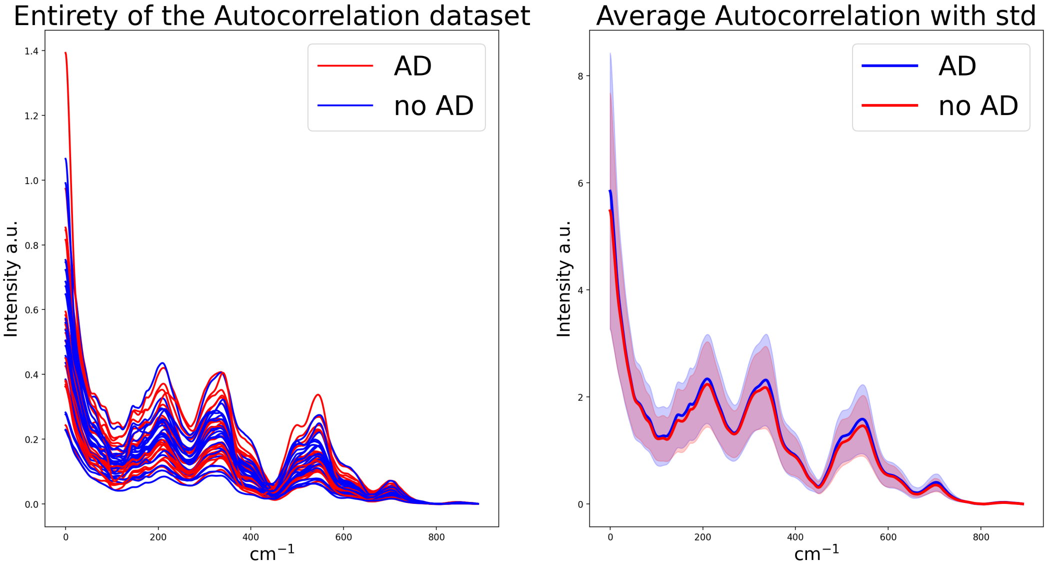


Figure 3.5: Alzheimer's disease RS dataset after the autocorrelation transform. The entirety of the dataset (left) and average with standard deviation (right).

		Predicted	
		AD	noAD
True	AD	18	3
	noAD	3	19

Table 3.3: Confusion table for Raman spectroscopy of Alzheimer's disease dataset.

3.3.2 Results

We refer to Table 3.3 for the confusion matrix of this application, which resulted in a 86% accuracy. Such results improve the current state-of-the-art [123], but must be considered preliminary due to the scarcity of data. In any case, our results strongly support that RS and topological data analysis together may provide an effective combination to the clinical diagnosis of AD. Also, our pipeline do not require the choice of any parameters, hence the proposed methodology may evolve in automatic support to AD diagnosis and could be easily embedded in a commercial platform of Raman spectroscopy. Of course, all these considerations are preliminary and require further statistical confirmation.

In conclusion, despite all these results are preliminary, they are significant and encouraging. On the two biomedical applications we have achieved state-of-the-art results with excellent accuracies. In the SST application the result were at least promising, since all competitors performed worse. However, the accuracy alone was not very satisfactory, but the dataset is extremely corrupted by noise. All these results further validate the utility of TML in real-world scenario, since it offers a powerful tool to describe and classify data where other state-of-the-art methods fail. Moreover, it is less impacted by the noise and is not as impacted as neural network by the scarcity of data. In accordance with current literature [16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27], these first chapters showed that TML is a promising new approach to data analysis and classification. Following these great premises, the remaining of the dissertation is devoted to investigate thoroughly the mathematical foundation of PH and to link it to a new category of operators that are more generale and may be more useful.

Chapter 4

A new set of equivariances for topological data analysis

This chapter represents the pivotal point of the two main concepts of the dissertation, namely topological data analysis and group equivariant non-expansive operators. The first part of the dissertation was devoted to the description, applications and possible advancements of topological data analysis, while the remaining part of the dissertation will be entirely dedicated to group equivariant non-expansive operators. In this chapter, we are going to establish a link between these two concepts in the form of Theorem 4.1.4. More in detail, we are going to prove that the operator that computes persistence diagrams can be seen as a particular instance of GENEOs from a functorial point of view.

Chapters 1, 2 and 3 presented the theory behind topological data analysis, some applications and insights on benchmark datasets, and some case studies with real-world data. In the latter, in particular, we achieved state-of-the-art accuracies. These successes are mainly due to the fact that, in most cases, features extracted by persistent homology are informative and representative of the data shape. Such features are easily exploitable in a machine learning setting to produce models that are both robust, to a certain extent interpretable, and accurate. The previous chapters have emphasized the potentials of TDA and its integration with deep learning only offers greater progress to this field [128, 129, 130, 131, 132, 133, 134]. This paradigm, however, has some weaknesses. Historically, the framework of TDA considered data as geometric objects whose shape can be studied by means of suitable filtration functions. With different filtrations available, we are equipped with different tools to study the geometry and the shape of data under various lens. A possibly more accurate approach, the one adopted by the theory of GENEOs, is that data does not carry geometric shape. Rather, data are interpreted as functions and the filtrations of TDA are simply transformations of functions (data) in other functions which are both more manageable and informative. The geometry and shape to be studied is therefore to be found not in data themselves, but in the space of such transformations. Moreover, modelling the framwork in terms of geometry of data has the severe risk to confuse the various protagonists of the TDA paradigm. That is, the border between transformations / preprocessing and the actual filtration can be very mixed. To provide an example, let us consider a setting in which we aim to apply the blur operator to images and compute the persistence diagrams by means of the cubical filtration. The confusion is whether we are defining a new filtration, which incorporates both the blur and the cubical filtration, or we are transforming the original data and then using the cubical filtration. The GENEO setting solves this issue considering only transformations and studying the shape of the resulting space. Moreover, it allows for more flexible approaches. For instance, both the blur operator and the operator that computes persistence diagrams are GENEOs. Because of the compactness and convexity of the GENEOs space (Theorems 5.2.3 and 5.2.4), we are allowed to consider every operator in the segment connecting these two, e.g. the average, and the resulting operator may be better suited for the task at hand. In the TDA setting, the average of the blur operator and the persistence diagram is not even defined. In addition, a further drawback of topological data anlysis lies in its mathematical foundations. It is a well-known fact that persistent homology is homeomorphism-invariant; it follows directly from its definition. The constrain of basic TDA to be invariant with respect to homeomoprhisms is too loose in a variety of scenarios. To solve this issue, in literature we can find examples of the design of *ad hoc* filtrations (with the same issues described above) in order to distinguish homeomorphic data. We can find examples of such behaviour also in Chapters 2 and 3. Although feasible, this process could definitely be improved upon. The theory of GENEOs solves this issue allowing to inject specific equivariances directly into the model, in the same fashion as persistent homology does relatively to homeomorphisms. From here onwards, we are going to present the theory of Group Equivariant Non-Expansive Operators (GENEOs), and when we refer to data we mean functions defined on a topological space. The theory of GENEOs consitutes the focus of the second part of the dissertation and will be described deeply in Chapter 5. In this chapter we briefly present the main concepts necessary to prove Theorem 4.1.4. This theorem is a contribution of this dissertation that aims to further establish a strong connection between the two core settings this thesis is based upon: TDA and GENEOs. Despite originating from different mathematical fields, namely representation theory and functional analysis, TDA and GENEOs are profoundly connected. Firstly, the core idea of both approaches is to study the shape of an object: data for TDA and observers for GENEOs. More importantly, their synergy is already known in literature and Theorem 4.1.4 aims to further establish a connection between these two concepts from a functorial point of view. More in detail, the use of GENEOs allows to restrict the invariance of TDA to subgroups of the group of homeomorphisms. Moreover, GENEOs interact with multiparameter persistence homology [40, 41] and we will show in the next chapter that GENEOs and TDA can provide metrics for each other that increase the descriptive power of both [12]. This chapter provides a new connection, proving that the computation of persistence diagrams is actually a GENEO. As briefly stated, the core idea behind the GENEO definition is to switch the focus from data to the observers, which can also be understood as tasks. Alongside an observer come specific transformations of data that are deemed equivariant in the model. Therefore, instead of on data, our model focuses on couples data set-transformations, referred to as **perception pairs**. Moreover, instead of focusing on the shape of data, we focus on the geometric and topological properties of the operators defined on such pairs. Generally, TDA and PH are understood as the geometric study of data; in this chapter, we are going to prove that they simply represent one of the possible studies.

Formally, let X be a non-empty topological space and Φ a subspace of $\mathbb{R}_b^X := \{\varphi\colon X \to \mathbb{R}, \varphi \text{ bounded}\}$. Let $\mathrm{Aut}_\Phi(X)$ be the group of bijections $g\colon X \to X$ such that $\varphi g \in \Phi$ for every $\varphi \in \Phi$, with the composition operation. Given $G \subseteq \mathrm{Aut}_\Phi(X)$, a perception pair is the pair (Φ, G). The group G represents the equivariances that we want to inject into the model. For example, if we are dealing with images the isometries are likely to belong to G. Given two perception pairs (Φ, G) and (Ψ, K) (with possible different domains X and Y), a Group Equivariant Non-Expansive Operator (GENEO) is a pair $(F, T)\colon (\Phi, G) \to (\Psi, K)$ such that $T\colon G \to K$ is a homomorphism and $F\colon \Phi \to \Psi$ is T-equivariant (i.e. $F(\varphi g) = F(\varphi)T(g)$) and non-expansive (i.e. $\|F(\varphi) - F(\psi)\|_\infty \leq \|\varphi - \psi\|_\infty$). For the sake of clarity, in this dissertation, we will restrict to the case where data are modeled as real-valued tame functions $\varphi\colon X \to \mathbb{R}$. We stress the fact that this request is little restrictive, since a large variety of data can be modeled in this way. An immediate example is signals, that are already modeled as functions. Also, grayscale images, where the domain is a grid of $\mathbb{R}^2$, or point clouds where $\varphi(x) = 1$ if x belong to the point cloud and $\varphi(x) = 0$ otherwise. We recall that, for functions defined on topological spaces, the homology we refer to is singular homology [58], and Definition 1.2.7 of tame function is modified accordingly.

4.1 The computation of PDs is a GENEO

We now describe the setting that will allow us to prove Theorem 4.1.4. In this chapter, we are going to treat the diagonal of persistence diagrams as a point with infinite multiplicity. For more information on such assumption, we refer the reader to [135]. Let us consider the category C_1 whose objects are tame functions $\varphi\colon X \to \mathbb{R}$, with X a topological space. Throughout the chapter we will always assume that $\varphi\colon X \to \mathbb{R}, \psi\colon Y \to \mathbb{R}$ and $\chi\colon Z \to \mathbb{R}$, with X, Y, Z topological spaces. The morphisms of C_1 from φ to ψ are pairs (h, ε), where $h\colon X \to Y$ is a homeomorphism such that $\|\varphi - \psi \circ h\|_\infty \leq \varepsilon$. We refer to the collection of morphisms of C_1 with $\mathrm{Mor}(C_1)$, and with $\mathrm{Mor}(\varphi, \psi)$ the morphisms between the objects φ and ψ. Given the morphisms $(h, \varepsilon_1) \in \mathrm{Mor}(\varphi, \psi)$ and $(k, \varepsilon_2) \in \mathrm{Mor}(\psi, \chi)$, their composition is given by $(k \circ h, \varepsilon_1 + \varepsilon_2)$. It holds that $(\mathrm{Id}_X, 0)$ is the identity morphism between (X, φ) and itself. The associativity of morphisms composition is given by the associativity of composition of homeomorphisms and associativity of the sum of real numbers, hence C_1 is truly a category. The second category C_2 we are going to define has as objects the persistence diagrams $\mathrm{Dgm}(\varphi)$ of functions in C_1. The morphisms from $\mathrm{Dgm}(\varphi)$ to $\mathrm{Dgm}(\psi)$ are pairs $(I^{\mathrm{Dgm}(\varphi),\mathrm{Dgm}(\psi)}, \eta)$, where $I^{\mathrm{Dgm}(\varphi),\mathrm{Dgm}(\psi)}$ is the collection of all matchings from $\mathrm{Dgm}(\varphi)$ to $\mathrm{Dgm}(\psi)$ with cost less or equal than $\eta \in \mathbb{R}$. For the definition of cost of a matching, we refer to [135]. We admit $I^{\mathrm{Dgm}(\varphi),\mathrm{Dgm}(\psi)}$ to be empty if η is so small that no lower cost matching is possible. In C_2, given two morphisms $(I^{\mathrm{Dgm}(\varphi),\mathrm{Dgm}(\psi)}, \eta_1) \in \mathrm{Mor}(\mathrm{Dgm}(\varphi), \mathrm{Dgm}(\psi))$ and $(I^{\mathrm{Dgm}(\psi),\mathrm{Dgm}(\chi)}, \eta_2) \in \mathrm{Mor}(\mathrm{Dgm}(\psi), \mathrm{Dgm}(\chi))$, we define the composition of morphisms to be

$$(I^{\mathrm{Dgm}(\psi),\mathrm{Dgm}(\chi)},\eta_2)\circ(I^{\mathrm{Dgm}(\varphi),\mathrm{Dgm}(\psi)},\eta_1):=(I^{\mathrm{Dgm}(\varphi),\mathrm{Dgm}(\psi)},\eta_1+\eta_2). \tag{4.1.1}$$

We note that the set $I^{\mathrm{Dgm}(\varphi),\mathrm{Dgm}(\varphi)}$ with cost less or equal than 0 consists of the singleton $\left\{\mathrm{Id}_{\mathrm{Dgm}(\varphi)}\right\}$, since we have assumed the diagonal to be a single point. Therefore, the identity morphism from $\mathrm{Dgm}(\varphi)$ to itself is given by $(\left\{\mathrm{Id}_{\mathrm{Dgm}(\varphi)}\right\},0)$. The associativity of morphisms of C_2 follows from definition. Hence, C_2 is actually a category. We now define a functor F between C_1 and C_2. Given $\varphi\in\mathrm{Ob}(C_1)$, we define $F(\varphi):=\mathrm{Dgm}(\varphi)\in\mathrm{Ob}(C_2)$. Given $(h,\varepsilon)\in\mathrm{Mor}(\varphi,\psi)$, with $h\colon X\to Y$, we define $F((h,\varepsilon)):=(I^{\mathrm{Dgm}_\varphi,\mathrm{Dgm}_\psi},\varepsilon)$. It easy to check that $I^{\mathrm{Dgm}(\varphi),\mathrm{Dgm}(\psi)}$ is not empty due to the stability theorem [136]. Let us consider in C_1 the identity morphism $(\mathrm{Id}_X,0)$ from φ to itself. It is easy to check that $F((\mathrm{Id}_X,0))=(\{\mathrm{Id}_{\mathrm{Dgm}_\varphi}\},0)$. Moreover, it holds that

$$\begin{aligned}F((k,\varepsilon_2)\circ(h,\varepsilon_1))&=F((k\circ h,\varepsilon_1+\varepsilon_2))\\&=(I^{\mathrm{Dgm}_\varphi,\mathrm{Dgm}_\chi},\varepsilon_1+\varepsilon_2)\\&=(I^{\mathrm{Dgm}_\psi,\mathrm{Dgm}_\chi},\varepsilon_2)\circ(I^{\mathrm{Dgm}_\varphi,\mathrm{Dgm}_\psi},\varepsilon_1)\\&=F((k,\varepsilon_2))\circ F((h,\varepsilon_1)).\end{aligned}$$

Hence, F is actually a functor between C_1 and C_2.

Remark 4.1.1. We highlight that, in order to formalise the functor F we are interested in, some technical care was necessary. We also note that, while each homeomorphism $h\colon X\to Y$ with cost bounded by ε induces a homomorphism between the homology groups of φ and ψ, h does not induce a matching between $\mathrm{Dgm}(\varphi)$ and $\mathrm{Dgm}(\psi)$ with cost less or equal than ε. Rather, it induces multiple matchings with bounded cost and there is no canonical choice. We refer to [137, 138] for a more detailed description of such behaviour, but the essential is that each morphism in C_1 is not associated in a natural way to a unique morphism in C_2.

Given a morphism $(h,\varepsilon)\in\mathrm{Mor}(\varphi,\psi)$, if we set $\varepsilon=0$, then necessarily the only homeomorphism we can consider (if it exists) is $h\colon X\to Y$ such that $\varphi=\psi\circ h$. In this case, there is a unique induced matching between $\mathrm{Dgm}(\varphi)$ and $\mathrm{Dgm}(\psi)$ and a unique morphism in $\mathrm{Mor}(\mathrm{Dgm}(\varphi),\mathrm{Dgm}(\psi))$, which is $(\left\{\mathrm{Id}_{\mathrm{Dgm}(\varphi)}\right\},0)$. We denote with Δ^+ the closed upper half-plane. That is, $\Delta^+:=\left\{(x,y)\in\mathbb{R}^2,x\le y\right\}$.

Remark 4.1.2. We recall that GENEOs map functions into functions. However, the support of a PD is a compact that can be identified as a function in this way: given a persistence diagram $\mathrm{Dgm}(\varphi)$, we define $f_{\mathrm{Dgm}(\varphi)}\colon\Delta^+\to\mathbb{R}$ that maps each point to the minimum distance from a point in the PD. Identifying a persitence diagram with such a function, the bottleneck distance between PDs turns into a distance between such functions. With a sligh abuse of notation, until the end of the chapter we are still going to denote with $\mathrm{Dgm}(\varphi)$ the support of the persistence diagram.

Before proceeding, we recall the following well-known theorem [139].

Theorem 4.1.3 *Let A,B be two compact sets, d_H the Hausdorff distance and d_A,d_B the distance function from A and B, respectively. Then,*

$$d_H(A,B)=\|d_A-d_B\|_\infty.$$

In our setting, the support of PDs are compact sets. Therefore it holds that $d_H\left(\mathrm{Dgm}(\varphi), \mathrm{Dgm}(\psi)\right) = \left\| d_{\mathrm{Dgm}(\varphi)} - d_{\mathrm{Dgm}(\psi)} \right\|_\infty = \left\| f_{\mathrm{Dgm}(\varphi)} - f_{\mathrm{Dgm}(\psi)} \right\|$. We are finally able to state and prove the main result of this chapter.

Theorem 4.1.4 *Let us consider $\Phi = C^0(X, \mathbb{R})$ and $D = \left\{ f_{\mathrm{Dgm}(\varphi)}, \varphi \in \Phi \right\}$. The functor $F\colon (\Phi, \mathrm{Homeo}(X)) \to (D, \mathrm{Id}_{\Delta^+}), F(\varphi) = f_{\mathrm{Dgm}(\varphi)}$ is a GENEO with respect to the trivial homomorphism $T\colon \mathrm{Homeo}(X) \to \mathrm{Id}_{\Delta^+}$.*

Proof. We have to check that F is T-equivariant and non-expansive. The T-equivariance follows directly from the invariance of persistent homology with respect to homeomorphisms. More precisely, given $\varphi \in C_1$ and $g \in \mathrm{Homeo}(X)$, we have that

$$\begin{aligned} F\left(\varphi \circ g\right) &= F\left(\varphi\right) \\ &= F\left(\varphi\right) \mathrm{Id}_{\Delta^+} \\ &= F\left(\varphi\right) T(g). \end{aligned}$$

For more details on the invariance of PH with respect to homeomorphisms, we refer the reader to Remark 10 of [52] and Theorem 2.5 of [140]. Regarding the non-expansivity, given $\varphi, \psi \in C_1$, we have that

$$\begin{aligned} \left\| F(\varphi) - F(\psi) \right\|_\infty &= \left\| f_{\mathrm{Dgm}(\varphi)} - f_{\mathrm{Dgm}(\psi)} \right\|_\infty \\ &= \left\| d_{\mathrm{Dgm}(\varphi)} - d_{\mathrm{Dgm}(\psi)} \right\|_\infty \\ &= d_H\left(\mathrm{Dgm}(\varphi), \mathrm{Dgm}(\psi)\right) \\ &\leq d_{\mathrm{match}}\left(\mathrm{Dgm}(\varphi), \mathrm{Dgm}(\psi)\right) \\ &\leq \left\| \varphi - \psi \right\|_\infty, \end{aligned}$$

where the second equality follows from Theorem 4.1.3 and the two inequalities come from Section 3.1 of [136]. □

Broadly speaking, Theorem 4.1.4 expresses the computation of persistence diagrams using the language of GENEOs and shows that such operator is actually a specific GENEO. As already stated in the Introduction and in the beginning of this chapter, GENEOs and TDA are linked in many ways, and Theorem 4.1.4 further increase this connection. It is the this strong connection that motivates the second part of the dissertation, that focuses on the topological and geometrical properties of the space of GENEOs, new ways to define them and some applications to benchmark datasets.

Chapter 5

The theory of group equivariant non-expansive operators

In the previous chapter we showed various connections between GENEOs and TDA, and most importantly we provided a new link in the form of Theorem 4.1.4. Since the computation of persistence diagrams can be thought of as a GENEO, the aim of this chapter is to fully present the theory of group equivariant non-expansive operators. The mathematical framework of GENEOs represents a novel geometric approach to the theory of deep neural networks which is proving successful due to its ability to generate models in low dimensions that are highly interpretable and transparent. Moreover, this framework is general enough to provide us new tools to explore the composition of such operators in ways that would be impossible in other frameworks. Returning to a previous example, let us suppose that we wish to apply blur to data and to compute the persistence diagram. Both the TDA setting and the GENEO setting allow us to easily perform both of these steps. The benefit of the GENEO setting is that we can actually be more malleable. Since the GENEOs space is convex and compact (Theorems 5.2.3, 5.2.3), every operator in the segment connecting the blur operator and the PD operator (e.g. the average operator) is still a GENEO and may be best suited for the study. Such an operator is immediately available and easily definable in the GENEO setting, but its meaning and definition in the TDA setting is not clear. The aim of this chapter is to present the current theory of GENEOs, some topological properties of their space and a method to build them by means of symmetric functions. Most of the contents of this chapter are not original and they are inserted for self-completion, but Section 5.3 is the first contribution of this dissertation to the theory of GENEOs. We recall that a more exhaustive bibliography of the full theory of GENEOs can be found in the Introduction. For more information about the results reported in this chapter, we refer the reader to the original papers: [12, 53, 54, 141]. For the sake of brevity, the proofs of results present in other works are omitted.

5.1 Topologies on data

One of the key aspects of this theory is the definition of suitable topologies on data. This concept follows the idea to formalize the assumption that data are stable.

The stability of data is a fundamental concept in applications, since it allows for reproducibility. Stability requires a notion of closeness and hence a topology. In our setting, a data set Φ is a set of bounded real-valued functions on a non-empty set X:

$$\Phi \subseteq \{\varphi\colon X \to \mathbb{R}, \varphi \text{ bounded}\} = \mathbb{R}_b^X,$$

where $\mathbb{R}_b^X$ is the set of all bounded real-valued functions on X. The set X is often called the **domain** of the data set Φ and we denote it with $\operatorname{dom}(\Phi)$. The idea of this setting is that the set X is the space where agents (or instruments) make measurements, and it is not accessible if not for the admissible measurements φ contained in Φ.

Example 5.1.1. A grayscale image can be formalized as a function φ from a grid on the real plane X to the real numbers. Agents have no control on the grid, e.g. the pixel density or distribution, they can only access the generated images and eventually transform them.

In our model, agents act on data by transforming it in a way that makes it easier to be studied. In doing so, they transform measurements into other measurements while preserving certain invariances and symmetries deemed important.

Definition 5.1.2. A **Φ-operation** is a function $g\colon X \to X$ such that the composition $\varphi g \in \Phi$ for every $\varphi \in \Phi$.

Remark 5.1.3. Given a Φ-operation g, the function $R_g\colon \Phi \to \Phi$ that maps φ to φg is non-expansive. Moreover, if g is a bijection then R_g is an isometry. This can be easily proven by noticing that, given $\varphi, \psi \in \Phi$, it holds that:

$$\|\varphi - \psi\|_\infty = \max_{x\in X} |\varphi(x) - \psi(x)| \geq \max_{x\in \operatorname{Im}(g)} |\varphi(x) - \psi(x)| = \|\varphi g - \psi g\|_\infty .$$

The composition of Φ-operations is yet a Φ-operation and the identity function Id_X is a Φ-operation for every $\Phi \subseteq \mathbb{R}_b^X$. We call a Φ-operation g **invertible** if there is a Φ-operation h such that $gh = hg = \mathrm{Id}_X$. From now on, we denote with g^{-1} the Φ-operation such that $gg^{-1} = g^{-1}g = \mathrm{Id}_X$. We denote the collection of all invertible Φ-operations with $\mathrm{Aut}_\Phi(X)$. More formally,

$$\mathrm{Aut}_\Phi(X) := \left\{g\colon X \to X, g \text{ is a bijection and } \varphi g, \varphi g^{-1} \in \Phi \text{ for every } \varphi \in \Phi\right\}.$$

Remark 5.1.4. $\mathrm{Aut}_\Phi(X)$ is a group with the composition operation.

The group $\mathrm{Aut}_\Phi(X)$ naturally induces an associative right action on Φ:

$$\rho\colon \Phi \times \mathrm{Aut}_\Phi(X) \to \Phi, \quad (\varphi, g) \mapsto \varphi g \tag{5.1.1}$$

where φg is the usual function composition.

Definition 5.1.5. A **perception pair** is (Φ, G) with $\Phi \subseteq \mathbb{R}_b^X$ and $G \subseteq \mathrm{Aut}_\Phi(X)$.

The choice of the group G encodes the symmetries of Φ that are deemed relevant by the agent for the task at hand. Different agents might choose different groups G even for the same data set Φ.

Example 5.1.6. Given a data set Φ, the pair $(\Phi, \mathrm{Aut}_\Phi(X))$ is called the universal perception pair.

Example 5.1.7. Let us consider the finite set $X = \{1, \ldots, n\}$ and the space $\mathbb{R}_b^X$. We note that $\mathrm{Aut}_{\mathbb{R}_b^X} = S_n$, where S_n is the set of permutation of X. Then $\left(\mathbb{R}_b^X, S_n\right)$ is a perception pair.

Example 5.1.8. As already stated, a choice of a suitable group of equivariances encodes the idea of different agents or perception pairs. This concept is often overlooked, since in many contexts the agent is subtextually clear. That is, given a dataset Φ, the perception pair (Φ, G) can be obvious and it is as if Φ is the only object of interest. We now provide an example where a dataset is given but the perception pair is not obvious. In such a case, we highlight how different choices of the group of equivariances G encode different perception pairs (or tasks). Figure 5.1 shows graylevel images sampled from the KDEF dataset [142]. In such a scenario, the task is not immediately clear, and hence neither is the group of equivariances. For instance, a possible task could be emotion recognition. Another plausible task is facial recognition. Together with the task (observer or agent), comes the different admissible transformations. For the sake of simplicity, we refer to φ_a the image represented in Figure 5.1a, and similarly for all other images, and with X the domain of the images (the grid of pixels). In an emotion recognition task, the map $g_1 \colon X \to X$ such that $\varphi_a g_1 = \varphi_b$ does not change the semantics of the image, and we would like for g_1 to belong the group of equivariances G, while $g_2 \colon X \to X$ such that $\varphi_a g_2 = \varphi_d$ would not. On the other hand, in a facial recognition task, the map g_2 would belong to G, while g_1 would not. That is, different tasks encode different symmetries, and considering only Φ (the data set) as the object of study is reductive. In our approach, we focus on pairs (Φ, G).

5.1.1 Topological structure on the data set

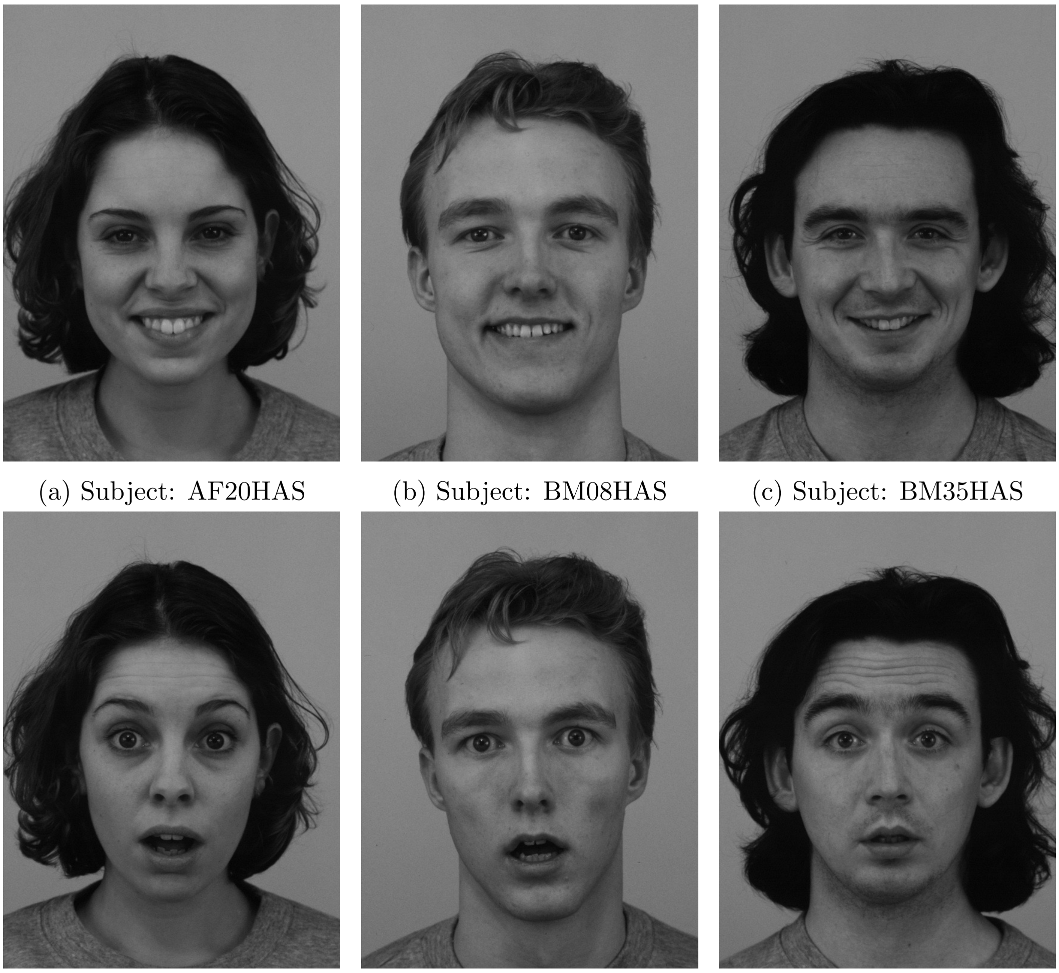
We endow the space of admissible measurements Φ with the topology of uniform convergence that is induced by the distance

$$D_\Phi\left(\varphi_1, \varphi_2\right) := \left\|\varphi_1 - \varphi_2\right\|_\infty .$$

The topological structure of X is inherited by the extended pseudo-metric D_X:

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

for $x_1, x_2 \in X$. We recall that a pseudo-metric is a distance d without the property that $d\left(x_1, x_2\right) = 0$ implies that $x_1 = x_2$ and that an extended pseudo-metric is a pseudo-metric that can take an infinite value. We stress the fact that the assumption behind the definition of D_X is that we can distinguish two points of X only if there is a measurement that maps them to different values. If Φ contains only constant functions, for instance, D_X vanishes for every couple of points of X and no discrimination can be made. This is a pathological example since, in this case, X is not even a T_0 space.



(a) Subject: AF20HAS (b) Subject: BM08HAS (c) Subject: BM35HAS

(d) Subject: AF20SUS (e) Subject: BM08SUS (f) Subject: BM35SUS

Figure 5.1: Sample of graylevel images from KDEF dataset [142].

We consider the extended pseudo-metric space (X, D_X) a topological space by choosing as base $\mathcal{B}_{D_X}$ for the topology the collection of open balls:

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

where $0 < \varepsilon < \infty$ and $x \in X$.

Remark 5.1.9. Our choice of topology allows us to deal with non-continuous functions with respect to the Euclidean topology.

In general, X is not compact with respect to the topology induced by D_X. As an example, let us consider the open interval $X =]0, 1[$ and $\Phi = \{\mathrm{Id}_X\}$. In this case, the topology induced by D_X is the Euclidean topology and X is not compact with respect to it. Nonetheless, the following results hold.

Theorem 5.1.10 *If Φ is totally bounded, then (X, D_X) is totally bounded.*

Corollary 5.1.11 *If Φ is totally bounded and (X, D_X) is complete, then (X, D_X) is compact.*

Example 5.1.12. Let $X = \{(\cos 2\pi p, \sin 2\pi p) \in \mathbb{R}^2 : p \in \mathbb{Q}\}$, Φ be the set of all non-expansive functions from X to $[0,1]$ and G be the group of all rotations $\rho_{2\pi q}$ of $2\pi q$ radiants, where $q \in \mathbb{Q}$. It holds that Φ is compact, but the topological space X is not complete, hence Φ is not compact.

We now investigate the relationship between the topology τ_{D_X} induced by the extended pseudo-metric D_X and the initial topology τ_{in} on X with respect to Φ, where the topology on $\mathbb{R}$ is simply the Euclidean topology. We recall that the initial topology is the coarsest topology on X such that each function $\varphi \in \Phi$ is continuous.

Proposition 5.1.13 *Each element $\varphi \in \Phi$ is a non-expansive map, and hence it is continuous with respect to D_X.*

Theorem 5.1.14 *If Φ is totally bounded, then the topology τ_{D_X} coincides with τ_{in}.*

Example 5.1.15. We can give a counter-example of Theorem 5.1.14 if Φ is not totally bounded. Let Φ be the set of all continuous functions from $[0,1]$ to $\mathbb{R}$, with respect to the Euclidean topologies on both $[0,1]$ and $\mathbb{R}$. In this case the initial topology τ_{in} is the Euclidean topology, while τ_{D_X} is the discrete topology.

5.1.2 Topological structure on the equivariance group

As of now, we have defined a topology on the data set Φ and a topology on the domain X. We now want to define a topology on the third and last component of a perception pair: $\mathrm{Aut}_\Phi(X)$. We define the distance between two elements of $\mathrm{Aut}_\Phi(X)$ as the difference of their actions on Φ. More specifically,

$$D_{\text{Aut}}(g_1, g_2) := \sup_{\varphi \in \Phi} D_\Phi(\varphi g_1, \varphi g_2),$$

for any $g_1, g_2 \in \mathrm{Aut}_\Phi(X)$.

Remark 5.1.16. For any $g_1, g_2 \in \mathrm{Aut}_\Phi(X)$ it holds that

$$\begin{aligned} D_{\text{Aut}}(g_1, g_2) &= \sup_{\varphi \in \Phi} D_\Phi(\varphi g_1, \varphi g_2) \\ &= \sup_{x \in X} \sup_{\varphi \in \Phi} |\varphi(g_1(x)) - \varphi(g_2(x))| \\ &= \sup_{x \in X} D_X(g_1(x), g_2(x)). \end{aligned}$$

Hence, D_{Aut} coincides with the pseudo-metric induced by the uniform convergence on $\mathrm{Aut}_\Phi(X)$.

As usual, we want to study the compactness of $\mathrm{Aut}_\Phi(X)$. We have the following results concerning the topology on $\mathrm{Aut}_\Phi(X)$ and the compactness of a subgroup of $\mathrm{Aut}_\Phi(X)$.

Theorem 5.1.17 *The following statements hold:*

- $\mathrm{Aut}_\Phi(X)$ *is a topological group with respect to the topology induced by* D_{Aut}*;*
- *the action of* $\mathrm{Aut}_\Phi(X)$ *on* Φ *is continuous.*

Theorem 5.1.18 *Let G be a subgroup of* $\mathrm{Aut}_\Phi(X)$*. If Φ is totally bounded, then* $(G, \mathrm{Aut}_\Phi(X))$ *is totally bounded.*

Corollary 5.1.19 *If Φ is totally bounded and (G, D_{Aut}) is complete, then (G, D_{Aut}) is compact.*

Example 5.1.20. We can give a counter-example of Corollary 5.1.19 if Φ is not totally bounded. Let X be the unit circle and Φ be the set of all non-expansive functions from X to $[0,1]$. Let G be the group of all rotations $\rho_{2\pi q}$ of $2\pi q$ radiants, where $q \in \mathbb{Q}$. Then, the space (G, D_{Aut}) is not complete, hence it is not compact.

Theorem 5.1.21 *Consider a compact subspace $\Phi \subseteq \mathbb{R}_b^X$. Assume that (X, D_X) is a compact metric space. Then, $\mathrm{Aut}_\Phi(X)$ is compact.*

5.1.3 The natural pseudo-distance

The last concept that we are going to introduce in this section is the natural pseudo-distance d_G [52]. This metric represents the ground truth in our model and allows for comparison between functions in Φ. Such a distance vanishes for functions that are equivalent with respect to the action of $G \subseteq \mathrm{Aut}_\Phi(X)$. Moreover, we will show that the natural pseudo-distance allows for another connection between GENEOs and TDA, different from the one presented in the previous chapter.

Definition 5.1.22. Given a group $G \subseteq \mathrm{Aut}_\Phi(X)$, the **natural pseudo-distance** d_G is defined by setting

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

where $\varphi_1, \varphi_2 \in \Phi$.

As pointed out in [143], the natural pseudo-distance between two measurements φ_1, φ_2 can be seen as the distance between the orbits $\varphi_1 G$ and $\varphi_2 G$ with respect to the action of G on Φ.

Remark 5.1.23. If $G = \{\mathrm{Id}_X\}$, then d_G and D_Φ coincides on Φ. Moreover, given two subgroups $G_1 \subseteq G_2 \subseteq \mathrm{Aut}_\Phi(X)$, it holds that

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

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

The main drawback of the natural pseudo-distance is that it is difficult to compute. Luckily, we will show in the following section that we are able to approximate the natural pseudo-distance by means of a dual approach based on group equivariant non-expansive operators and persistent homology.

5.2 The space of GENEOs

In this section, we are going to introduce the main concept of this new approach to the framework of geometric deep learning, namely Group Equivariant Non-Expansive Operators (GENEOs). In our framework, GENEOs encode the idea of agents that transform data (perception pairs) preserving symmetries and distances. Moreover, we are going to define a topological structure on the space of GENEOs and a new pseudo-distance on Φ, based on a dual approach of persistent homology and the theory of GENEOs, which allows us for an approximation of the natural pseudo-distance.

For the whole section (Φ, G) and (Ψ, K) will denote two perception pairs. We stress the fact that such perception pairs are allowed to have different groups of equivariances and different domains.

Definition 5.2.1. A map $(F, T) : (\Phi, G) \to (\Psi, K)$ such that $T \colon G \to K$ is an homomorphism and $F \colon \Phi \to \Psi$ is a continuous, T-equivariant map (i.e. $F(\varphi g) = F(\varphi) T(g)$ for every $\varphi \in \Phi$ and every $g \in G$) is called a **Group Equivariant Operator (GEO)** from (Φ, G) to (Ψ, K).

Definition 5.2.2. A GEO (F, T) from (Φ, G) to (Ψ, K) such that F is non-expansive is called a **Group Equivariant Non-Expansive Operator (GENEO)**.

In scenarios where the homomorphism T is fixed, we will refer to the map $F \colon \Phi \to \Psi$ as a GENEO (resp. GEO) if the couple (F, T) satisfies Definition 5.2.2 (resp. 5.2.1). We denote with $\mathcal{F}^{\text{all}}$ the set of all GENEOs between two perception pairs $(\Phi, G), (\Psi, K)$. Given a homomorphism $T \colon G \to K$, we denote with $\mathcal{F}^{\text{all}}_T$ the set of all GENEOs between the perception pairs (Φ, G) and (Ψ, K) with respect to T. We endow $\mathcal{F}^{\text{all}}$ with the uniform convergence distance

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

for any $F_1, F_2 \in \mathcal{F}^{\text{all}}$. As already stated, the GENEOs space benefits from good mathematical properties that make such a space more manageable and potentially more useful. In particular, the following results hold.

Theorem 5.2.3 *If Φ and Ψ are compact with respect to D_Φ and D_Ψ, respectively, then $\mathcal{F}^{\text{all}}_T$ is compact with respect to D_{GENEO}.*

Theorem 5.2.4 *If Ψ is convex, then $\mathcal{F}^{\text{all}}_T$ is convex.*

Theorem 5.2.3 and Theorem 5.2.4 are fundamental in applications. The compactness of $\mathcal{F}^{\text{all}}_T$ guarantees that the space of GENEOs can be approximated by a finite set, while the convexity allows us to obtain new GENEOs by convex combinations of pre-existing ones. Finally, we have a preliminary result of $\mathcal{F}^{\text{all}}$.

Corollary 5.2.5 *If X, Y are finite and Φ, Ψ are compact with respect to D_Φ and D_Ψ, respectively, then $\mathcal{F}^{\text{all}}$ is compact.*

Proof. If X and Y are finite, then the set of all homomorphisms between G and K is finite. The finite union of compacts is compact, hence from Theorem 5.2.3 it follows that $\mathcal{F}^{\text{all}}$ is compact. □

Remark 5.2.6. We underline that if we remove the assumption that our operators are non-expansive, the property of compactness does not hold anymore. As an example, let $\Phi = \Psi$ be equal to the set of all constant functions from $\mathbb{R}$ to $[0, 1]$, and $G = H$ be the trivial group containing just the identity on $\mathbb{R}$. We observe that $\Phi, X = \mathbb{R}$ and G are compact with respect to their topologies. Let us now consider the sequence $(F_n)_{n \in \mathbb{N}}$ of GEOs from Φ to Φ with respect to the identity homomorphism $\text{Id}_G \colon G \to G$, defined by setting $F_n(\varphi) := \varphi^n$ for every function $\varphi \in \Phi$ and every positive integer n. It is easy to check that $\lim_{n\to\infty} D_{\mathrm{GENEO}}(F_m, F_n) = 1$ for every positive integer m, and hence the sequence (F_n) does not admit any converging subsequence. This implies that the space of all GEOs from Φ to Φ with respect to Id_G is not compact.

The results shown so far have demonstrated that the non-expansivity of the GENEOs, a prerogative of this model, has important implications not only from an epistemological point of view, but also from a mathematical one.

5.2.1 Persistent homology-induced pseudo-metrics

As already stated in Chapter 4, the computation of persistence diagrams can be seen as a particular GENEO with an appropriate choice of perception pairs. In this section we are going to present more in detail some of the connections between TDA and GENEOs that are already known in literature. In particular, PH allows us to define suitable metrics for the approximation of the natural pseudo-distance and the other distances defined in Section 5.1. Let us fix two perception pairs (Φ, G), (Ψ, K) and a homomorphism $T\colon G \to K$. Let us consider a set of GENEOs $\mathcal{F} \subseteq \mathcal{F}_T^{\mathrm{all}}$. We can define another metric, $\mathcal{D}_{\mathrm{match}}^{\mathcal{F},k}$ which is computationally efficient, stable and strongly invariant. The roots of $\mathcal{D}_{\mathrm{match}}^{\mathcal{F},k}$ are in persistent homology.

Definition 5.2.7. A pseudo-metric d on Φ is **strongly G-invariant** if it is invariant under the action of G with respect to each variable. That is,

$$d(\varphi_1, \varphi_2) = d(\varphi_1 g, \varphi_2) = d(\varphi_1, \varphi_2 g) = d(\varphi_1 g, \varphi_2 g),$$

for every $\varphi_1, \varphi_2 \in \Phi$ and every $g \in G$.

Remark 5.2.8. The natural pseudo-distance d_G is strongly G-invariant.

Persistent Betti numbers (PBNs, Definition 1.2.1) are not necessarily finite in our setting, even if X is compact. Therefore, for the rest of this chapter, we will assume that PBNs of every $\varphi \in \Phi$ take a finite value at each point. We recall that d_{match} comes from Definition 1.2.6.

Example 5.2.9. Let us consider the set $X = \{0\} \cup \left\{\frac{1}{n}, n \in \mathbb{N}^+\right\}$ and $\Phi = \{\iota\colon X \hookrightarrow \mathbb{R}\}$. X is compact, but every sublevel set $X_u = \{x \in X, x \leq u\}$, for $u > 0$, has infinitely many connected components.

Let us now fix a non-empty subset $\mathcal{F}$ of $\mathcal{F}_T^{\mathrm{all}}$. For every k, we can finally define the extended pseudo-metric $\mathcal{D}_{\mathrm{match}}^{\mathcal{F},k}$ on Φ:

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

where $\varphi_1, \varphi_2 \in \Phi$ and $r_k(\varphi)$ is the k-th persistent Betti number with respect to $\varphi \in \Phi$. The following results hold.

Proposition 5.2.10 *$\mathcal{D}_{\mathrm{match}}^{\mathcal{F},k}$ is a strongly G-invariant pseudo-metric on Φ.*

Theorem 5.2.11 *Let $\mathcal{F}^{\mathrm{all}}$ be the space of all GENEOs from (Φ, G) to (Ψ, K). If $\mathcal{F}$ is a non-empty subset of $\mathcal{F}^{\mathrm{all}}$, then*

$$D_{\mathrm{match}}^{\mathcal{F},k} \leq d_G \leq D_\Phi.$$

Theorem 5.2.12 *Let $\mathcal{F}^{\mathrm{all}}$ be the space of all GENEOs from (Φ, G) to itself. Let us assume that each function $\varphi \in \Phi$ is non-negative, the k-th Betti number of X does not vanish and that for each function $\varphi \in \Phi$, also each constant function c such that $0 \leq c \leq \|\varphi\|_{\infty}$ is in Φ. Then, $\mathcal{D}_{\mathrm{match}}^{\mathcal{F},k} = d_G$.*

Remark 5.2.13. If Φ is bounded, we can add a suitable constant to every function in Φ in order to make them non-negative. Hence, if Φ is bounded such hypothesis in Theorem 5.2.12 is not restrictive.

Proposition 5.2.14 *Let $\mathcal{F}$ be a non-empty subset of $\mathcal{F}_T^{\mathrm{all}}$. For every $\varepsilon > 0$, a finite subset $\mathcal{F}^*$ of $\mathcal{F}$ exists, such that*

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

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

The previous results are of great importance in our framework. First, they allow the natural pseudo-distance to be approximated in a computationally efficient way via persistent homology. Second, they further enhance the connections between the theory of GENEOs and TDA. Although the connection was already introduced in Chapter 4, Theorem 5.2.12 establishes a connection between different theoretical concepts, namely d_G and $\mathcal{D}_{\mathrm{match}}^{\mathcal{F},k}$.

5.3 Building GENEOs via symmetric functions

The previous section explained how we are able to approximate the set of all GENEOs with just a finite subset, due to the compactness of the space of such operators. This approximation requires large and dense sets of GENEOs, each one representing a data-observer interaction. In this section, we are going to introduce a new method to produce non-linear GENEOs through the concepts of symmetric function and permutant. This result is the first contribution of this dissertation to the theory of GENEOs and it is derived from [54]. This technique benefits from the approximability of continuous symmetric functions by symmetric polynomials. The mathematical setting is the same as the previous sections, but we will consider $G = H$ and $T = \mathrm{Id}_G$.

First, we denote the image of X through the admissible measurements as $\mathrm{Im}(\Phi) = \{\varphi(x), \text{ for } \varphi \in \Phi, x \in X\}$. The following result holds:

Proposition 5.3.1 *If X and Φ are compact, then $\mathrm{Im}(\Phi)$ is compact with respect to the Euclidean topology on $\mathbb{R}$.*

Definition 5.3.2. Let S_X be the set of permutations of X. For each $g \in G$, the map $c_g \colon S_X \to S_X$ that maps $s \in S_X$ to $g \circ s \circ g^{-1}$ is called the **conjugation action** of $g \in G$ on S_X. For every subset H of S_X, we denote the set $c_g(H)$ by the symbol gHg^{-1}.

Definition 5.3.3. A finite set $H \subseteq \mathrm{Aut}_{\Phi}(X)$ is called a **permutant** for G if either $H = \emptyset$ or $gHg^{-1} = H$ for every $g \in G$.

Remark 5.3.4. In general, a permutant is not a normal subgroup of G. We stress that we require neither that a permutant is a group nor that it is a subset of G.

Example 5.3.5. The sets $\emptyset$ and $\{\mathrm{Id}_X\}$ are trivial permutants for any subgroup G of $\mathrm{Aut}_\Phi(X)$.

Example 5.3.6. If G and $\mathrm{Aut}_\Phi(X)$ are finite, both G and $\mathrm{Aut}_\Phi(X)$ are permutants for G.

Example 5.3.7. Let Φ be the set of all functions from S^1 to $[0,1]$ that are non-expansive with respect to the Euclidean distances on both S^1 and $[0,1]$.

- If G is the group of isometries of $\mathbb{R}$ and h is the clockwise rotation of ℓ radiants, for $\ell \in \mathbb{R}$, then the set $H = \left\{h, h^{-1}\right\}$ is a permutant for G.
- If G is the group generated by the reflection with respect to the axis $x = 0$, then the set $H = \left\{\mathrm{Id}_{S^1}, \rho, \rho^2, \rho^3\right\}$ is a permutant for G, where ρ is the clockwise rotation of $\pi/2$ around the origin.

Remark 5.3.8. If G is Abelian, every finite subset of G is a permutant for G. This follows from the fact that the conjugation action on an Abelian group is the identity.

Remark 5.3.9. In what follows, we will commit a slight abuse of notation with the symbol $\|\cdot\|_\infty$. It will be used both for the max-norm of functions and the max-norm of points of $\mathbb{R}^m$, for $m \in \mathbb{N}$. That is, given $\varphi\colon X \to \mathbb{R}$ and $\alpha = (\alpha_1, \dots, \alpha_m) \in \mathbb{R}^m$, $\|\varphi\|_\infty = \max_{x\in X} |\varphi(x)|$ and $\|\alpha\|_\infty = \max_{1\le i\le m} |\alpha_i|$.

5.3.1 Building GEOs from symmetric functions

Definition 5.3.10. Let X be a symmetric subset of $\mathbb{R}^n$, i.e. a subset C such that $\pi(X) = X$ for every permutation π of the coordinates. A function $f\colon X \to \mathbb{R}$ is **symmetric** on X if its value is the same no matter the order of its arguments. That is,

$$f(a_1, \dots, a_n) = f\left(a_{\pi(1)}, \dots, a_{\pi(n)}\right)$$

for every $(a_1, \dots, a_n) \in X$ and every permutation π of the set $\{1, \dots, n\}$.

Proposition 5.3.11 *Let f be a continuous real-valued symmetric function defined on a compact symmetric subset of $\mathbb{R}^n$. Then f is the restriction of a continuous real-valued symmetric function f defined on $\mathbb{R}^n$.*

Proposition 5.3.11 guarantees that a continuous real-valued symmetric function defined on a compact symmetric subset of $\mathbb{R}^n$ coincides with the restriction of a continuous real-valued symmetric function defined on $\mathbb{R}^n$. Let $\mathcal{S}\colon \mathbb{R}^n \to \mathbb{R}$ be a symmetric function and $H = \{h_i\}_{i=1}^n$ be a non-empty permutant for $G \subseteq \mathrm{Aut}_\Phi(X)$. Then, we can define an operator $\mathcal{S}_H\colon \Phi \to \mathbb{R}_b^X$ by setting, for any $\varphi \in \Phi$,

$$\mathcal{S}_H(\varphi) := \mathcal{S}(\varphi \circ h_1, \dots, \varphi \circ h_n), \tag{5.3.1}$$

where $\mathcal{S}(\varphi \circ h_1, \dots, \varphi \circ h_n)(x) := \mathcal{S}((\varphi \circ h_1)(x), \dots, (\varphi \circ h_n)(x))$ for every $x \in X$.

Proposition 5.3.12 *If $\mathcal{S}\colon \mathbb{R}^n \to \mathbb{R}$ is a symmetric function and $G \subseteq \mathrm{Aut}_\Phi(X)$, then $\mathcal{S}_H$ defined as in Equation 5.3.1 is a GEO from Φ to $\mathbb{R}_b^X$ with respect to the identity homomorphism $\mathrm{Id}_G\colon G \to G$.*

Corollary 5.3.13 *If $\mathcal{S}\colon \mathbb{R}^n \to \mathbb{R}$ is a symmetric function and its restriction to $\mathrm{Im}(\Phi)^n$ is non-expansive, then $\mathcal{S}_H$ is a GENEO from Φ to $\mathbb{R}_b^X$ with respect to Id_G.*

As of now, we have defined a GEO associated with a symmetric function and we showed that such a GEO is actually a GENEO if the function is non-expansive. In the following sections, we will show how to define a GENEO even in the presence of symmetric functions that may be expansive.

Remark 5.3.14. The key point of the use of permutants in this setting is that the size of permutants is, in many cases, smaller than the size of the equivariance group G. Indeed, an alternative approach to the construction of GENEOs may be with the integration on the equivariance group G. However, despite the possibly infinite and large size of the group G, the size of a permutant is by its very nature finite. Moreover, the largest the group of equivariances G becomes, the smaller the size of permutants for G.

5.3.2 Fundamental theorem on symmetric polynomials

For extending Corollary 5.3.13 to expansive symmetric functions, we need to approximate them by means of elementary symmetric functions. In the sequel, we will denote the symmetric group over the set $\{1, \dots, n\}$ as S_n. In this section, K will denote a compact metric space. Let $\mathcal{C}(K)$ be the vector space of continuous real-valued functions on K. In what follows, with a slight abuse of notation, we will confuse each polynomial with the function it represents, in the domain we are considering.

Definition 5.3.15. Given a natural number n and a finite subset $I \subseteq \mathbb{N}^n$, a polynomial $\sum_{(k_1,\dots,k_n)\in I} c_{k_1,\dots,k_n} y_1^{k_1} \dots y_n^{k_n}$ is said to be **symmetric** if $\pi(I) = I$ and $c_{k_1,\dots,k_n} = c_{\pi(k_1),\dots,\pi(k_n)}$ for every multi-index $(k_1, \dots, k_n) \in I$ and every permutation $\pi \in S_n$.

We recall the following well-known definitions and theorems (cf. [144]).

Definition 5.3.16. A subset A of $\mathcal{C}(K)$ is an **algebra** if it is a vector subspace of $\mathcal{C}(K)$ that is closed under multiplication. That is, given $f, g \in A$, then $f \cdot g \in A$.

Definition 5.3.17. A set S of functions on K **separates points** if for each pair of points $s, t \in K$ there is a function $f \in S$ such that $f(s) \neq f(t)$.

Definition 5.3.18. A set S of functions on K **vanishes** at $s \in K$ if $f(s) = 0$ for all $f \in S$.

Theorem 5.3.19 (Stone - Weierstrass Theorem) *An algebra A of continuous real-valued functions on a compact metric space K that separates points and does not vanish at any point is dense in $\mathcal{C}(K)$ with respect to the max-norm referred to the domain K.*

Corollary 5.3.20 *Let K be a compact subset of $\mathbb{R}^n$. The algebra of all polynomials $p(y_1, \dots, y_n)$ in n variables is dense in $\mathcal{C}(K)$ with respect to the max-norm referred to the domain K.*

Corollary 5.3.20 guarantees that we can approximate a continuous symmetric function by a polynomial with arbitrary accuracy. However, we also require such a polynomial to be symmetric. We can obtain this by a symmetrization of the previously found polynomial.

Proposition 5.3.21 *Let K be a compact subset of $\mathbb{R}^n$, verifying the property $\pi(K) = K$ for every $\pi \in S_n$. If $\mathcal{S}_{|K}\colon K \to \mathbb{R}$ is the restriction to K of a continuous symmetric function $\mathcal{S}\colon \mathbb{R}^n \to \mathbb{R}$ and $\|\cdot\|_\infty$ is the max-norm referred to the domain K, then for every $\varepsilon > 0$ there exists a symmetric polynomial q in n variables such that $\left\|\mathcal{S}_{|K} - q_{|K}\right\|_\infty \leq \varepsilon$.*

Definition 5.3.22. The **elementary symmetric polynomials** in the n variables $a_1, \dots, a_n$, also called **elementary symmetric functions**, are defined as:

- $\sigma_1 := \sum_{1 \leq i \leq n} a_i$;
- $\sigma_2 := \sum_{1 \leq i < j \leq n} a_i \cdot a_j$;
- $\sigma_r := \sum_{1 \leq i_1 < i_2 < \cdots < i_r \leq n} \prod_{j=i_1}^{i_r} a_j$;
- $\sigma_n := \prod_{1 \leq i \leq n} a_i$.

Theorem 5.3.23 (Fundamental Theorem on Symmetric Polynomials, [145]) *Any symmetric polynomial in n variables $a_1, \dots, a_n$ is representable in a unique way as a polynomial in the elementary symmetric polynomials $\sigma_1, \dots, \sigma_n$.*

Since the proofs of Theorems 5.3.19 and 5.3.23 are constructive, we are effectively able to approximate each continuous symmetric function $\mathcal{S}\colon \mathbb{R}^n \to \mathbb{R}$ restricted on K a compact symmetric subset of $\mathbb{R}^n$ with an error less than ε by a polynomial in the elementary symmetric functions restred to K.

To summarize the section, let $G \subseteq \mathrm{Aut}_\Phi(X)$ be the equivariance group and let F be the GEO defined in Proposition 5.3.12 with respect to the symmetric function $\mathcal{S}\colon \mathbb{R}^n \to \mathbb{R}$. Let X and Φ be compacts. Since $\mathrm{Im}(\Phi)^n$ is guaranteed to be compact by Proposition 5.3.1, we can approximate $\mathcal{S}$ by a polynomial $p\colon \mathbb{R}^n \to \mathbb{R}$ with an arbitrarily small error ε. Now, we define the symmetric polynomial $q(a_1, \dots, a_n) := \frac{1}{n!} \sum_{\pi \in S_n} p\left(a_{\pi(1)}, \dots, a_{\pi(n)}\right)$. Given a permutant $H = \{h_1, \dots, h_n\}$, we define the GEO

$$F'(\varphi) := q(\varphi \circ h_1, \dots, \varphi \circ h_n)$$

for every $\varphi \in \Phi$. It is easy to check that

$$\left\|F(\varphi) - F'(\varphi)\right\|_\infty \leq \varepsilon$$

for any $\varphi \in \Phi$. Hence, the operator F' is arbitrarily close to F and it is associated to a polynomial in the elementary symmetric polynomials.

5.3.3 Building GENEOs from elementary symmetric functions

Let $\mathcal{S}\colon \mathbb{R}^n \to \mathbb{R}$ be a continuous symmetric function. Since $\pi(\mathrm{Im}(\Phi)^n) = \mathrm{Im}(\Phi)^n$ trivially holds for every $\pi \in S_n$, the previous section guarantees that we can approximate $\mathcal{S}_{|\mathrm{Im}(\Phi)^n}$ with the restriction on $\mathrm{Im}(\Phi)^n$ of a polynomial in the elementary symmetric functions, defined as

$$\tilde{\mathcal{S}}(a_1, \dots, a_n) = \sum_{k_1=0}^{m_1} \cdots \sum_{k_n=0}^{m_n} c_{k_1,\dots,k_n} \prod_{i=1}^{n} \sigma_i^{k_i}(a_1, \dots, a_n),$$

where $m_i \in \mathbb{N}$ for every $i \in \{1, \ldots, n\}$, $c_{k_1,\ldots,k_n} \in \mathbb{R}$ for every $k_1 \in \{0, \ldots, m_1\}, \ldots, k_n \in \{0, \ldots, m_n\}$ and σ_i is the i-th elementary symmetric polynomial for every $i \in \{1, \ldots, n\}$. Let us now define the following constants:

$$M_{\mathrm{Im}(\Phi)^n} := \max_{\alpha \in \mathrm{Im}(\Phi)^n} \|\alpha\|_\infty = \max_{\varphi \in \Phi} \|\varphi\|_\infty$$

$$M_1 := \max_{1 \le i \le n} \left\{ k_i \binom{n}{i}^{k_i} i M_{\mathrm{Im}(\Phi)^n}^{ik_i - 1} \right\}$$

$$M_2 := \max_{1 \le i \le n} \left\{ \binom{n}{i}^{k_i} M_{\mathrm{Im}(\Phi)^n}^{ik_i} \right\}^{n-1}$$

$$C = n \sum_{k_1=0}^{m_1} \cdots \sum_{k_n=0}^{m_n} |c_{k_1,\ldots,k_n}| M_1 M_2, \tag{5.3.2}$$

Let us consider a non-empty permutant $H = \{h_i\}_{i=1}^n$ for G. We can define the operator $\hat{\mathcal{S}}_H \colon \Phi \to \mathbb{R}_b^X$ by setting

$$\hat{\mathcal{S}}_H(\varphi) := \frac{1}{C} \tilde{\mathcal{S}}(\varphi \circ h_1, \ldots, \varphi \circ h_n)$$

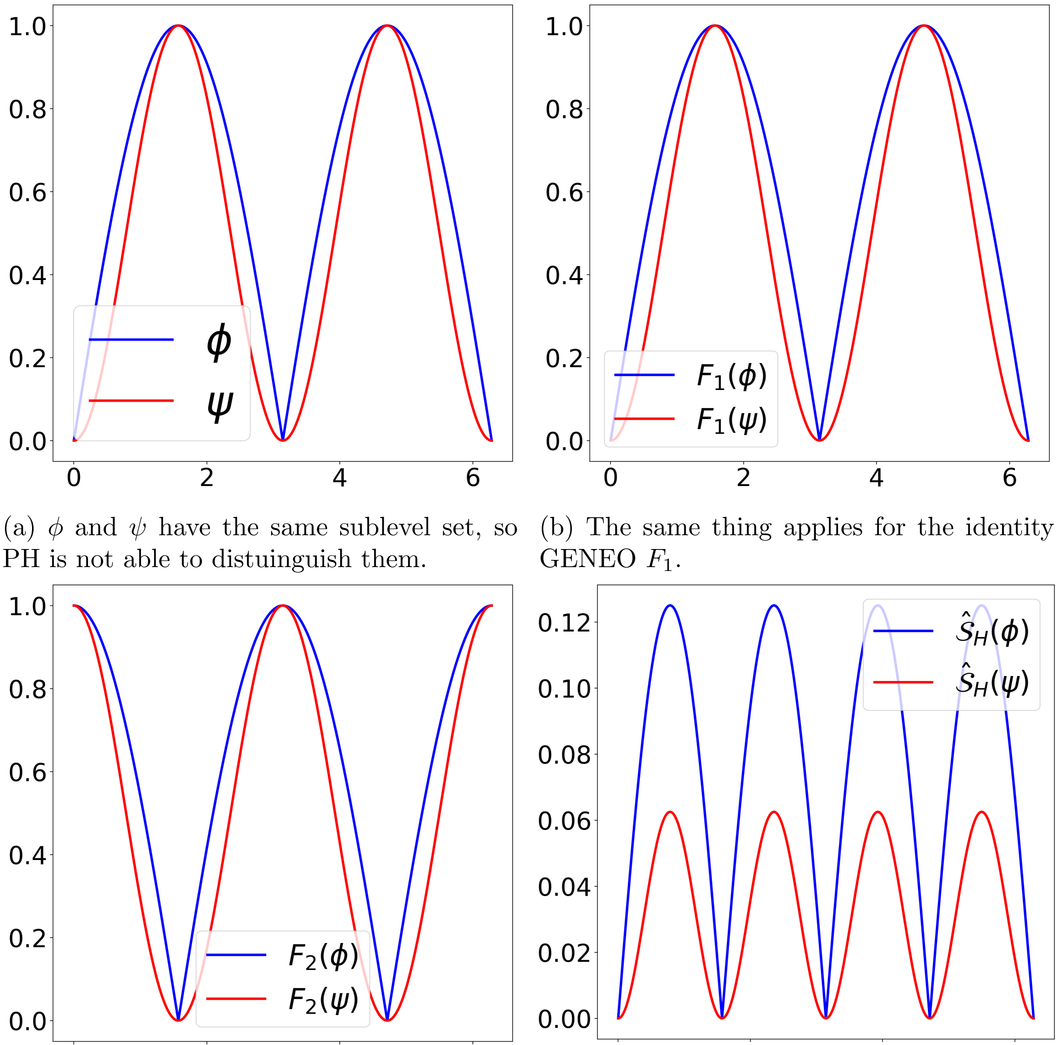
for any $\varphi \in \Phi$, where $\tilde{\mathcal{S}}(\varphi \circ h_1, \ldots, \varphi \circ h_n)(x) := \tilde{\mathcal{S}}((\varphi \circ h_1)(x), \ldots, (\varphi \circ h_n)(x))$ for every $x \in X$ and C is the constant defined in Equation 5.3.2. Finally, we can state the following.

Theorem 5.3.24 *If $\tilde{\mathcal{S}}$ is a polynomial in the n elementary symmetric functions, then $\hat{\mathcal{S}}_H$ is a GENEO from Φ to $\mathbb{R}_b^X$ with respect to* Id_G.

5.3.4 Symmetric GENEOs and data

Theorem 5.3.24 allows us to define a new set of GENEOs by means of symmetric functions and permutants. In addition to having a new tool for populating the GENEO space, main aim of this section, we are now going to explore the usefulness of operators defined in this way. More in detail, we are going to present two examples. The first example shows how symmetric GENEOs allow us to distinguish data that would otherwise be undistinguishable through persistent homology alone. In the second example, the use of non-linear GENEOs allows for more flexible pseudo-metrics $\mathcal{D}_{\mathrm{match}}^{\mathcal{F},k}$ compared to the same distance obtained through linear GENEOs, which will instead be the focus of the next chapter.

Example 5.3.25. Let us consider Φ as the space of all 1-Lipschitz functions from S^1 to $[0, 1]$, the equivariance group G of all rotations of S^1 and two function $\phi, \psi \in \Phi$ defined as $\phi(x) = |\sin(x)|, \psi(x) = \sin(x)^2$. Persistent homology is not able to distinguish ϕ from ψ, since the sublevel set of both functions is the same for every threshold value, see Figure 5.2a. Let us consider two GENEOs: $F_1 = \mathrm{Id}$ and $F_2(\varphi) = \varphi \circ \rho_{\pi/2}$, where $\rho_{\pi/2}$ is the clockwise rotation through a $\pi/2$ angle. The joint use of persistent homology, F_1 and F_2 is clearly not able to distinguish the two functions, since they both share the same sublevel set. See Figures 5.2b and 5.2c. However, let us consider the second elementary symmetric function σ_2 and the permutant $H = \left\{ \mathrm{Id}_{S^1}, \rho_{\pi/2} \right\}$. We can apply Theorem 5.3.24 and get that $\hat{\mathcal{S}}_H\left(\varphi\right) = \frac{1}{4}\left(\varphi \cdot \left(\varphi \circ \rho_{\pi/2}\right)\right)$ is a symmetric GENEO. Finally, the synergy between persistent homology and GENEOs, specifically Theorem 5.3.24, is able to distinguish ϕ and ψ, since the sublevel set of $\hat{\mathcal{S}}_H\left(\phi\right)$ and $\hat{\mathcal{S}}_H\left(\psi\right)$ is different, see Figure 5.2d.



(a) ϕ and ψ have the same sublevel set, so PH is not able to distuinguish them.

(b) The same thing applies for the identity GENEO F_1.

(c) The GENEO that translates data by $\pi/2$ radiants is not able to distinguish ϕ from ψ, when looked through PH.

(d) The symmetric GENEO $\hat{\mathcal{S}}_H$ and PH is able to distinguish the two data, since the sublevel set is different.

Figure 5.2: It is easy to check that the synergy between persistent homology and symmetric GENEOs is able to distinguish data that persistent homology alone cannot, since the sublevel set of $\hat{\mathcal{S}}_H(\phi)$ and $\hat{\mathcal{S}}_H(\psi)$ is different.

Example 5.3.26. Let Φ be the set of all functions from a set $X = \{A, B\}$ to $[0, 1]$. We can identify each function in $\varphi \in \Phi$ as the ordered pair $(\varphi(A), \varphi(B))$. Let us fix G as the group of all permutations of two elements. We recall that, from Section 5.1.1, on Φ we have the metric $D_\Phi\left(\varphi, \psi\right) = \|\varphi - \psi\|_\infty$, for $\varphi, \psi \in \Phi$. Hence, the pairwise distance between $\varphi_1 = (0,0), \varphi_2 = (0,1), \varphi_3 = (1,0)$ and $\varphi_4 = (1,1)$ is always 1. Since GENEOs encode the idea of an observer, let us suppose that we want to define a GENEO F such that the pseudo-metric $\|F(\varphi) - F(\psi)\|_\infty$ vanishes for functions with a null component, while maintaining a positive distance between φ_4 and every φ_i, for $i = 1, 2, 3$. It is easy to check that each linear transformation that maps both $(1,0)$ and $(0,1)$ to $(0,0)$ must also map $(1,1)$ to $(0,0)$. Hence, we cannot define the desired GENEO through linear maps. On the contrary, the GENEO associated with the second elementary symmetric function $\sigma_2(a_1, a_2) = a_1 \cdot a_2$ and with the permutant $H = G$ defines the pseudo-metric with the requested property.

Chapter 6

A representation theorem for linear GENEOs

The aim of this chapter is to provide a new contribution to the theory of GENEOs in the form of Theorem 6.2.2, which is a representation theorem for linear GENEOs. The ability to populate the space of GENEOs with useful operators is still one of the main limitations of the theory and this chapter operates in that direction. In particular, this chapter provides a tool to generate linear GENEOs using a weighted summation of a generalized permutant measure, which is more easily definable than GENEOs. This study expands the concepts presented in [49], but allows for arbitrary finite perception pairs and homomorphisms.

6.1 Building linear GEOs via generalized permutant measures

Throughout the chapter, we are going to restrict to finite domains for the perception pairs. Let us consider the sets $X = \{x_1, \dots, x_m\} = \{x_j\}_{j=1}^m$ and $Y = \{y_1, \dots, y_n\} = \{y_i\}_{i=1}^n$ and the function spaces $\mathbb{R}^X$ and $\mathbb{R}^Y$, equipped with the usual uniform norm $\|\cdot\|_\infty$. The finiteness of the domain allows us to simplify the key concepts of our theory. More in detail we recall that, if X is finite, $\mathbb{R}^X$ has the canonical basis $\left\{\mathbb{1}_{x_j}\right\}_j$, where $\mathbb{1}_{x_j} \colon X \to \mathbb{R}$ is the function with value 1 on x_j and 0 otherwise. Also, $\mathbb{R}^X \cong \mathbb{R}^m$ and $\mathrm{Aut}_{\mathbb{R}^X}(X) = \mathrm{Aut}(X)$. A similar result holds for $\mathbb{R}^Y$ and $\mathrm{Aut}_{\mathbb{R}^Y}(Y)$. As usual in the GENEO framework, we need to fix a group homomorphism $T \colon G \to K$, where $G \subseteq \mathrm{Aut}(X)$ and $K \subseteq \mathrm{Aut}(Y)$. The last concept that we need is the set of all functions from Y to X, which we denote with X^Y. First of all, we need to mimic the concept of the conjugation action in the presence of a homomorphism T. For each $h \in X^Y$, we consider the following action of G on X^Y:

$$\alpha_T \colon G \times X^Y \to X^Y, \quad (g, h) \mapsto ghTg^{-1}.$$

In particular, we are interested in the orbit $G(h)$ of h under the action of α_T.

Remark 6.1.1. Since X is finite, the pseudo-metric D_X induces the discrete topology on X. With this topology, $\mathbb{R}^X$ coincides with $C^0(X, \mathbb{R})$. Similarly for D_Y.

We can state the following definitions.

Definition 6.1.2. A GENEO (resp. GEO) $(F,T)\colon (\mathbb{R}^X, G) \to (\mathbb{R}^Y, K)$ is called **linear** if F is a linear map between $\mathbb{R}^X$ and $\mathbb{R}^Y$. That is, $F(\varphi+\alpha\psi) = F(\varphi)+\alpha F(\psi)$ for every $\varphi, \psi \in \mathbb{R}^X$ and every $\alpha \in \mathbb{R}$.

Definition 6.1.3. A finite signed measure μ on X^Y is called a **generalized permutant measure** with respect to T if each subset H of X^Y is measurable and μ is invariant under the action α_T of G, i.e. $\mu(H) = \mu(gHTg^{-1})$ for every $g \in G$. Equivalently, we can say that a finite signed measure μ on X^Y is a generalized permutant measure with respect to (G,T) if each singleton $\{h\} \subseteq X^Y$ is measurable and $\mu(\{h\}) = \mu(\{ghTg^{-1}\})$ for every $g \in G$.

Example 6.1.4. Let X, Y be two non-empty finite sets, with $Y \subseteq X$. We fix the equivariance group G on X as the group of all permutations of X that preserve Y, and the equivariance group K on Y is the group of all permutations of Y. Let us consider the homomorphism $T\colon G \to K$ that takes each permutation of X to its restriction to Y. For any positive integer m, we define $H_m = \{h\colon Y \to X \text{ such that } |\mathrm{Im}(h)| = m\}$, where $|\cdot|$ denotes the cardinality of the set. We define the finite signed measure μ as such: given $h \in H_m, \mu(h) := \frac{1}{m|H_m|}$. Then μ is a generalized permutant measure with respect to T.

For the rest of the chapter, we will always assume that a given homomorphism T is given, therefore we will omit the dependence from it in the definitions of GEO and GENEO. Moreover, with a slight abuse of notation, we denote with $\mu(h)$ the signed measure of the singleton $\{h\}$ for each $h \in X^Y$. Thanks to the concept of generalized permutant measure we are able to define a linear GEO associated to it. In fact, we can state the following.

Proposition 6.1.5 *Let μ be a generalized permutant measure. The operator $F\colon \mathbb{R}^X \to \mathbb{R}^Y$ defined by*

$$F(\varphi) := \sum_{h \in X^Y} \varphi h \mu(h)$$

for every $\varphi \in \mathbb{R}^X$ is a linear GEO.

Proof. First, we show that F is T-equivariant. It holds that:

$$\begin{aligned}
F(\varphi g) &= \sum_{h \in X^Y} \varphi g h \mu(h) \\
&= \sum_{h \in X^Y} \varphi g h T g^{-1} T g \mu\left(ghTg^{-1}\right) \\
&= \sum_{f \in X^Y} \varphi f T g \mu(f) \\
&= F(\varphi) T g,
\end{aligned}$$

since $\mu(h) = \mu\left(ghTg^{-1}\right)$ and the map $h \mapsto f := ghTg^{-1}$ is a bijection from X^Y to

X^Y. Now we show that F is linear. Let $\varphi, \psi \in \mathbb{R}^X$ and let $\lambda \in \mathbb{R}$. It holds that:

$$\begin{aligned} F(\varphi + \lambda\psi) &= \sum_{h \in X^Y} (\varphi + \lambda\psi)\, h\mu(h) \\ &= \sum_{h \in X^Y} \varphi h\mu(h) + \sum_{h \in X^Y} \lambda\psi h\mu(h) \\ &= F(\varphi) + \lambda F(\psi). \end{aligned}$$

Therefore, F is a linear GEO. □

Proposition 6.1.5 provides a first link between generalized permutant measures and GEOs, and its definition and proof was relatively easy. In contrast, proving the other verse of the representation theorem, i.e. associating a generalized permutant measure with a GEO, is more difficult. In any case, we are able as of now to state the result that we want to prove.

Theorem 6.1.6 *Assume that $G \subseteq \mathrm{Aut}(X), K \subseteq \mathrm{Aut}(Y)$ transitively acts on the finite set Y, $T \colon G \to K$ is a surjective homomorphism and F is a map from $\mathbb{R}^X$ to $\mathbb{R}^Y$. The map F is a linear group equivariant operator from $(\mathbb{R}^X, G)$ to $(\mathbb{R}^Y, K)$ with respect to the homomorphism T if and only if a generalized permutant measure μ exists such that $F(\varphi) = \sum_{h \in X^Y} \varphi h\mu(h)$ for every $\varphi \in \mathbb{R}^X$.*

The remaining of the section is devoted to the proof of Theorem 6.1.6. We are going to split the proof in many substeps, both for ease of reading and to separate each step in its own result.

6.1.1 Decomposition of stochastic matrices

In order to prove Theorem 6.1.6, it is necessary to introduce tools to decompose stochastic matrices. This is the aim of this section, which is mainly adapted from [146]. Let $\mathcal{M}_{n\times m}$ be the set of all $n \times m$ real-valued matrices. Moreover, with the symbol $[\ell]$ we denote the set $\{1, \ldots, \ell\}$ for every $\ell \in \mathbb{N}$.

Definition 6.1.7. A matrix $A = (a_{ij}) \in \mathcal{M}_{n\times m}$ is **(right) stochastic** if $a_{ij} \geq 0$ for all $(i, j) \in [n] \times [m]$ and $\sum_{j=1}^{m} a_{ij} = 1$ for all $i \in [n]$.

A $\{0, 1\}$-matrix is a matrix $A = (a_{ij}) \in \mathcal{M}_{n\times m}$, such that $a_{ij} \in \{0, 1\}$ for all $(i, j) \in [n] \times [m]$. We shall refer to the $\{0, 1\}$-matrices in $\mathcal{M}_{n\times m}$ with exactly one 1 in each row as **rectangular (row) permutation matrices** and $\mathcal{RP}_{n\times m}$ is the set of all rectangular permutation matrices of dimension $n \times m$. From now on, we will drop the right and row dependence from the definition of stochastic and rectangular permutation matrices, since it is understood. It is a well-known fact that the set of stochastic matrices is a convex set. We recall that an extreme point for a convex set is a point that does not lie in any open line segment joining two points in the set. With this definition, Theorem 1 of [146] can be restated as the following.

Theorem 6.1.8 *Every $n \times m$ stochastic matrix can be expressed as a convex combination of $n \times m$ rectangular permutation matrices.*

Remark 6.1.9. In general, the convex combination stated in Theorem 6.1.8 is not unique. As an example, let us consider the following stochastic matrix:

$$B = \begin{pmatrix} 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}.$$

To keep the notation simple, we refer to the rectangular permutation matrix with 1 as i-th element of the first row and j-th element of the second row as $R_{i,j}$. Using this notation, B can be expressed as the following convex combinations:

$$B = \frac{1}{12}R_{1,1} + \frac{5}{24}R_{1,2} + \frac{5}{24}R_{1,3} + \frac{1}{4}R_{3,1} + \frac{1}{8}R_{3,2} + \frac{1}{8}R_{3,3}$$

and

$$B = \frac{5}{24}R_{1,1} + \frac{1}{12}R_{1,2} + \frac{5}{24}R_{1,3} + \frac{1}{8}R_{3,1} + \frac{1}{4}R_{3,2} + \frac{1}{8}R_{3,3}.$$

6.1.2 Proof of the representation theorem for linear GEOs

In order to prove Theorem 6.1.6, we are going to split the task in several substeps that we are going to prove singularly. We stress the fact that one verse of the proof has aldready been proved by Proposition 6.1.5. Let us assume that $F\colon \mathbb{R}^X \to \mathbb{R}^Y$ is a linear GEO. Moreover, let $B = (b_{ij})$ be the matrix associated with F with respect to the bases $\{\mathbb{1}_{x_1}, \dots, \mathbb{1}_{x_m}\}$ for $\mathbb{R}^X$ and $\{\mathbb{1}_{y_1}, \dots \mathbb{1}_{y_n}\}$ for $\mathbb{R}^Y$. Given a natural number $\ell \in \mathbb{N}$ and a set Z with $|Z| = \ell$, for every permutation $p : Z \to Z$ we will denote by $\sigma_p : [\ell] \to [\ell]$ the function defined by setting $\sigma_p(j) = i$ if and only if $p(x_j) = x_i$. We observe that $\sigma_{p^{-1}} = \sigma_p^{-1}$.

Lemma 6.1.10 *For any $g \in G$, we have that $b_{ij} = b_{\sigma_{Tg}(i)\sigma_g(j)}$ for every $(i, j) \in [n] \times [m]$.*

Proof. Let us choose a function $\mathbb{1}_{x_j}$ and a permutation $g \in G$. By equivariance we have that

$$F(\mathbb{1}_{x_j} g) = F(\mathbb{1}_{x_j})Tg.$$

The left-hand side of the equation can be rewritten as:

$$F(\mathbb{1}_{x_j} g) = F(\mathbb{1}_{g^{-1}(x_j)}) = \sum_{i=1}^{n} b_{i\sigma_g^{-1}(j)}\mathbb{1}_{y_i}.$$

On the right-hand side, we get

$$F(\mathbb{1}_{x_j})Tg = \left(\sum_{i=1}^{n} b_{ij}\mathbb{1}_{y_i}\right)Tg = \sum_{i=1}^{n} b_{ij}(\mathbb{1}_{y_i}Tg) = \sum_{i=1}^{n} b_{ij}(\mathbb{1}_{(Tg)^{-1}(y_i)}) = \sum_{s=1}^{n} b_{\sigma_{Tg(s)}j}\mathbb{1}_{y_s},$$

by setting $y_s = (Tg)^{-1}(y_i)$. Therefore, we obtain the following equation:

$$\sum_{i=1}^{n} b_{i\sigma_g^{-1}(j)}\mathbb{1}_{y_i} = \sum_{s=1}^{n} b_{\sigma_{Tg(s)}j}\mathbb{1}_{y_s}.$$

This immediately implies that $b_{i\sigma_g^{-1}(j)} = b_{\sigma_{Tg(i)}j}$, for any $i \in [n]$. Since this equality holds for any $j \in [m]$ and any $g \in G$, we have that $b_{ij} = b_{\sigma_{Tg}(i)\sigma_g(j)}$ for every $(i, j) \in [n] \times [m]$ and every $g \in G$. □

For the rest of the chapter, we will assume that K is transitive on Y and T is surjective. Moreover, the surjectivity of T implies that $m > n$. These assumptions allow us to prove the following.

Lemma 6.1.11 *If K is transitive and T is surjective, an m-tuple of real numbers $\beta = (\beta_1, \dots, \beta_n)$ exists such that each row of B can be obtained by permuting β.*

Proof. Since K is transitive, for every $i \in [n]$ there exists $h_{i1} \in K$ such that $h_{i1}(y_i) = y_1$. Considering the $\bar{\imath}$-th row of B and $h_{\bar{\imath}1} \in K$, there exists $\bar{g} \in G$ such that $T(\bar{g}) = h_{\bar{\imath}1}$ because T is surjective. Hence, by Lemma 6.1.10, we have that $b_{\bar{\imath}j} = b_{\sigma_{T\bar{g}}(\bar{\imath})\sigma_{\bar{g}}(j)} = b_{\sigma_{h_{\bar{\imath}1}}(\bar{\imath})\sigma_{\bar{g}}(j)} = b_{1\sigma_{\bar{g}}(j)}$, for any $j \in [m]$. Since $\sigma_{\bar{g}}$ is a permutation, the $\bar{\imath}$-th row is a permutation of the first row. □

In order to proceed, it is helpful to split the GEO F in its positive and negative parts. That is, let us consider the linear maps $F^{\oplus}, F^{\ominus} : \mathbb{R}^X \to \mathbb{R}^Y$ defined by setting $F^{\oplus}(\mathbb{1}_{x_j}) := \sum_{i=1}^{n} \max\{b_{ij}, 0\} \mathbb{1}_{y_i}$ and $F^{\ominus}(\mathbb{1}_{x_j}) := \sum_{i=1}^{n} \max\{-b_{ij}, 0\} \mathbb{1}_{y_i}$ for every index $j \in \{1, \dots, m\}$. One can easily check that the following properties hold:

1. $F^{\oplus}, F^{\ominus}$ are T-equivariant linear maps;
2. The matrices $B^{\oplus}$ and $B^{\ominus}$ associated with $F^{\oplus}$ and $F^{\ominus}$ with respect to the bases $\{\mathbb{1}_{x_1}, \dots, \mathbb{1}_{x_m}\}$ for $\mathbb{R}^X$ and $\{\mathbb{1}_{y_1}, \dots, \mathbb{1}_{y_n}\}$ for $\mathbb{R}^Y$ are $B^{\oplus} = \left(b^{\oplus}_{ij}\right) = (\max\{b_{ij}, 0\})$ and $B^{\ominus} = \left(b^{\ominus}_{ij}\right) = (\max\{-b_{ij}, 0\})$, respectively. In particular, $B^{\oplus}$ and $B^{\ominus}$ are non-negative matrices;
3. $F = F^{\oplus} - F^{\ominus}$ and $B = B^{\oplus} - B^{\ominus}$;
4. Lemma 6.1.11 and the definitions of $B^{\oplus}, B^{\ominus}$ imply that two m-tuples of positive real numbers $\beta^{\oplus} = \left(\beta^{\oplus}_1, \dots, \beta^{\oplus}_m\right)$, $\beta^{\ominus} = \left(\beta^{\ominus}_1, \dots, \beta^{\ominus}_m\right)$ exist such that each row of $B^{\oplus}$ can be obtained by permuting $\beta^{\oplus}$, and each row of $B^{\ominus}$ can be obtained by permuting $\beta^{\ominus}$.

The next step in order to prove Theorem 6.1.6 is to express $F^{\oplus}$ and $F^{\ominus}$ as weighter sums of φh, for $h \in X^Y$. First, we need to establish a connection between the elements of X^Y and the rectangular permutation matrices. Each function $h \colon Y \to X$ can be associated with a $n \times m$ rectangular permutation matrix $R(h) = (r_{ij})$ defined by setting $r_{ij} = 1$ if $h(y_i) = x_j$, and $r_{ij} = 0$ otherwise. In the case $X = Y$ and $h \colon X \to X$ is a permutation, we denote as $P(h)$ such square matrix, which is the usual permutation matrix. Furthermore, one can prove that there is a bijection between X^Y and the set $\mathcal{RP}_{n\times m}$ of all $n \times m$ rectangular permutation matrices that sends h to $R(h)$.

Remark 6.1.12. Assume that h is a function from Y to X and $\mathcal{R}_h$ denotes the linear operator that sends each function φ in $\mathbb{R}^X$ to φh in $\mathbb{R}^Y$. One could easily check that $R(h)$ is the matrix associated with the operator $\mathcal{R}_h$ with respect to the bases $\{\mathbb{1}_{x_1}, \dots, \mathbb{1}_{x_m}\}$ for $\mathbb{R}^X$ and $\{\mathbb{1}_{y_1}, \dots, \mathbb{1}_{y_n}\}$ for $\mathbb{R}^Y$.

Proposition 6.1.13 *For every $h \in X^Y$ there exist two non-negative real numbers $c^{\oplus}(h), c^{\ominus}(h)$ such that*

$$F^{\oplus}(\varphi) = \sum_{h \in X^Y} c^{\oplus}(h)\varphi h,$$

$$F^{\ominus}(\varphi) = \sum_{h \in X^Y} c^{\ominus}(h)\varphi h,$$

for every $\varphi \in \mathbb{R}^X$*.*

Proof. Let us start by considering the statement concerning $c^\oplus$ and $F^\oplus$. If $|\beta^\oplus|_1 := \sum_{i=1}^m \beta_i^\oplus$, we note that the matrix $\frac{1}{|\beta^\oplus|_1} B^\oplus$ is stochastic. Hence, Remark 6.1.12 and Theorem 6.1.8 imply that, with a slight abuse of notation,

$$\begin{aligned} F^\oplus(\varphi) &= |\beta^\oplus|_1 \left(\frac{1}{|\beta^\oplus|_1} B^\oplus \varphi \right) \\ &= |\beta^\oplus|_1 \sum_{h \in X^Y} \gamma^\oplus(h) R(h)\varphi \\ &= \sum_{h \in X^Y} c^\oplus(h) \mathcal{R}_h(\varphi) \\ &= \sum_{h \in X^Y} c^\oplus(h) \varphi h, \end{aligned}$$

where $\sum_{h \in X^Y} \gamma^\oplus(h) = 1$, and $c^\oplus(h) = |\beta^\oplus|_1 \gamma^\oplus(h) \geq 0$ for any $h \in X^Y$. The proof of the statement concerning $c^\ominus$ and $F^\ominus$ is analogous. □

We recall that our aim is to associate a generalized permutant measure to a GEO F, and we split the task to its positive and negative parts $F^\oplus$ and $F^\ominus$. As of now, the function that maps $h \mapsto c^\oplus(h)$ is not a generalized permutant measure. In order to produce a generalized permutant measure, we need to take its average on the orbits of h under the action of G. First, we need to show that such a measure is well-defined.

Proposition 6.1.14 *If* $f_1, f_2 \in X^Y$ *and there exists* $t \in [n]$ *such that* $f_1(y_t) = f_2(y_t)$*, then either* $c^\oplus(f_1) = 0$*, or* $c^\ominus(f_2) = 0$ *or both values are null.*

Proof. Since there exists $t \in [n]$ such that $x_s = f_1(y_t) = f_2(y_t)$, we can consider $\mathbb{1}_{x_s}$ and get that

$$\begin{aligned} F^\oplus(\mathbb{1}_{x_s}) &= B^\oplus \mathbb{1}_{x_s} \\ &= \sum_{i=1}^n b_{is}^\oplus \mathbb{1}_{y_i}. \end{aligned}$$

Hence, we have that

$$\begin{aligned} b_{ts}^\oplus &= \left(\sum_{i=1}^n b_{is}^\oplus \mathbb{1}_{y_i} \right)(y_t) \\ &= F^\oplus(\mathbb{1}_{x_s})(y_t) \\ &= \sum_{h \in X^Y} c^\oplus(h) \mathbb{1}_{x_s} h(y_t) \\ &\geq c^\oplus(f_1) \mathbb{1}_{x_s} f_1(y_t) \\ &= c^\oplus(f_1). \end{aligned}$$

One could similarly check that $b_{ts}^\ominus \geq c^\ominus(f_2)$. Therefore,

$$c^\oplus(f_1) > 0 \implies b_{ts}^\oplus > 0 \implies b_{ts}^\ominus = 0 \implies c^\ominus(f_2) = 0.$$

It follows that either $c^\oplus(f_1) = 0$, or $c^\ominus(f_2) = 0$, or both. □

Corollary 6.1.15 *For every $f \in \mathrm{Aut}(X)$, either $c^{\oplus}(f) = 0$, or $c^{\ominus}(f) = 0$, or both.*

Proof. Set $f_1 = f_2$ in Proposition 6.1.14. □

We finally have the prerequisites to define the generalized permutant measures $\mu^{\oplus}$ and $\mu^{\ominus}$ associated to $F^{\oplus}$ and $F^{\ominus}$. Each of them is the average of the functions $c^{\oplus}$ and $c^{\ominus}$ along the orbit $\mathcal{O}(h)$ of h under the action of G. Formally, we define

$$\mu^{\oplus}(h) := \sum_{f \in \mathcal{O}(h)} \frac{c^{\oplus}(f)}{|\mathcal{O}(f)|} = \sum_{f \in \mathcal{O}(h)} \frac{c^{\oplus}(f)}{|\mathcal{O}(h)|},$$
$$\mu^{\ominus}(h) := \sum_{f \in \mathcal{O}(h)} \frac{c^{\ominus}(f)}{|\mathcal{O}(f)|} = \sum_{f \in \mathcal{O}(h)} \frac{c^{\ominus}(f)}{|\mathcal{O}(h)|}.$$

Proposition 6.1.16 *$\mu^{\oplus}$ and $\mu^{\ominus}$ are generalized permutant measures. As a consequence, the function $\mu = \mu^{\oplus} - \mu^{\ominus}$ is also a generalized permutant measure.*

Proof. The definition of $\mu^{\oplus}$ immediately implies that $\mu^{\oplus}(H) = \mu^{\oplus}\left(gHT(g)^{-1}\right)$ for every $g \in G$ and every subset H of X^Y. In other words, $\mu^{\oplus}$ is a non-negative generalized permutant measure. Quite analogously, we can prove that $\mu^{\ominus}$ is a non-negative generalized permutant. From Corollary 6.1.15, we can conclude that the function $\mu := \mu^{\oplus} - \mu^{\ominus}$ is a generalized permutant. □

Now that we have defined the two generalized permutant measures associated to the T-equivariant linear maps $F^{\oplus}$ and $F^{\ominus}$, we can start the steps to finally prove that the weighted sum of such measures are precisely the two maps. Let $G_h := \{g \in G \mid \alpha_T(g, h) = h\}$ be the stabilizer subgroup of G with respect to h, i.e., the subgroup of G containing the elements that fix h by the action. We recall that acting on h with respect to every element of G we obtain each element of the orbit $\mathcal{O}(h)$ exactly $|G_h|$ times, and the well known relation $|G_h|\,|\mathcal{O}(h)| = |G|$ (cf. [147]). We observe that, for $f, h \in X^Y$, if $f \in \mathcal{O}(h)$ then G_f is isomorphic to G_h. Finally we are able to prove Proposition 6.1.17, which states something very similar to Theorem 6.1.6 but regarding $F^{\oplus}$ and $F^{\ominus}$.

Proposition 6.1.17 *For any $\varphi \in \mathbb{R}^X$ we have that*

$$F^{\oplus}(\varphi) = \sum_{f \in X^Y} \varphi f \mu^{\oplus}(f),$$
$$F^{\ominus}(\varphi) = \sum_{f \in X^Y} \varphi f \mu^{\ominus}(f).$$

Proof. Recalling that $\mathcal{R}_g \colon \varphi \mapsto \varphi g$ and $\mathcal{R}_k \colon \psi \mapsto \psi k$ are linear maps for any $g \in G$ and $k \in K$, the T-equivariance condition $F^{\oplus}\mathcal{R}_{g^{-1}} = \mathcal{R}_{T(g)^{-1}}F^{\oplus}$ directly implies that $B^{\oplus}P(g) = P\left(T(g)\right)B^{\oplus}$. In particular, we have that $P\left(T(g)\right)B^{\oplus}P(g)^{-1} = B^{\oplus}$

for every $g \in G$. From Proposition 6.1.13 it follows that

$$\begin{aligned}
B^{\oplus} &= \overbrace{\frac{1}{|G|}B^{\oplus} + \cdots + \frac{1}{|G|}B^{\oplus}}^{|G| \text{ summands}} \\
&= \frac{1}{|G|}\sum_{g\in G} P(T(g))B^{\oplus}P(g)^{-1} \\
&= \frac{1}{|G|}\sum_{g\in G} P(T(g))\left(\sum_{h\in X^Y} c^{\oplus}(h)R(h)\right)P(g)^{-1} \\
&= \sum_{h\in X^Y}\sum_{g\in G}\frac{c^{\oplus}(h)}{|G|}P(T(g))R(h)P(g)^{-1} \\
&= \sum_{h\in X^Y}\frac{c^{\oplus}(h)}{|G|}\sum_{g\in G} R\left(ghT(g)^{-1}\right).
\end{aligned}$$

Therefore, with a slight abuse of notation,

$$\begin{aligned}
F^{\oplus}(\varphi) &= \sum_{h\in X^Y}\frac{c^{\oplus}(h)}{|G|}\sum_{g\in G} R\left(ghT(g)^{-1}\right)\varphi \qquad (6.1.1)\\
&= \sum_{h\in X^Y}\frac{c^{\oplus}(h)}{|G|}\sum_{g\in G} \varphi ghT(g)^{-1}. \qquad (6.1.2)
\end{aligned}$$

Let us now set $\delta(f_1, f_2) = 1$ if f_1 and f_2 belong to the same orbit under the action of G, and $\delta(f_1, f_2) = 0$ otherwise. Therefore, equality 6.1.1 implies

$$\begin{aligned}
F^{\oplus}(\varphi) &= \sum_{h\in X^Y}\frac{c^{\oplus}(h)}{|G|}\left|G_h\right|\sum_{f\in\mathcal{O}(h)}\varphi f \\
&= \sum_{h\in X^Y}\frac{c^{\oplus}(h)}{|G|}\left|G_h\right|\sum_{f\in X^Y}\delta(f,h)\varphi f \\
&= \sum_{f\in X^Y}\left(\sum_{h\in X^Y}\frac{c^{\oplus}(h)}{|G|}\left|G_h\right|\delta(f,h)\right)\varphi f \\
&= \sum_{f\in X^Y}\left(\sum_{h\in X^Y}\frac{c^{\oplus}(h)}{|\mathcal{O}(h)|}\delta(f,h)\right)\varphi f \\
&= \sum_{f\in X^Y}\left(\sum_{h\in \mathcal{O}(f)}\frac{c^{\oplus}(h)}{|\mathcal{O}(h)|}\right)\varphi f \\
&= \sum_{f\in X^Y}\varphi f \mu^{\oplus}(f).
\end{aligned}$$

The statement concerning $F^{\ominus}$ is proved analogously. □

Proposition 6.1.5, together with Proposition 6.1.16 and Proposition 6.1.17 are enough to prove Theorem 6.1.6, since they solve both verses of the representation theorem. We recall that all this section requires that K transitively acts on Y and that T is surjective, hence such hypothesis are inserted in the theorem. We have just proved a representation theorem for linear GEOs. Now, we are going to extend it to include non-expansivity. Before proceeding, we want to say something more about the relationship between $\mu^{\oplus}$ and $\mu^{\ominus}$. In what follows, we say that two matrices $A = (a_{ij}), B = (b_{ij}) \in \mathcal{M}_{n\times m}$ are mutually singular if $a_{ij} \neq 0 \implies b_{ij} = 0$ and vice versa.

Proposition 6.1.18 *$\mu^{\oplus}$ and $\mu^{\ominus}$ are mutually singular.*

Proof. If $\mu^{\oplus} \equiv 0$ there is nothing to prove. Let us assume that $\mu^{\oplus}$ is not the null measure and consider $\overline{h} \in X^Y$ such that $\mu^{\oplus}(\overline{h}) > 0$. By contradiction, we suppose that $\mu^{\ominus}(\overline{h}) > 0$. Then, by definition of $\mu^{\ominus}$ there exists $f \in \mathcal{O}(\overline{h})$ such that $c^{\ominus}(f) > 0$. Since $R(f) = (r(f)_{ij})$ is a rectangular permutation matrix, there is an index $(\bar{\imath}, \bar{\jmath})$ such that $r(f)_{\bar{\imath}\bar{\jmath}} = 1$. Hence, it holds that $b^{\ominus}_{\bar{\imath},\bar{\jmath}} = \sum_{h\in X^Y} r(h)_{\bar{\imath},\bar{\jmath}} c^{\ominus}(h) > 0$. Since $B^{\oplus}$ and $B^{\ominus}$ are mutually singular, $b^{\oplus}_{\bar{\imath},\bar{\jmath}} = 0$, and this is absurd, since from Proposition 6.1.17 we have that $b^{\oplus}_{\bar{\imath},\bar{\jmath}} = \sum_{h\in X^Y} r(h)_{\bar{\imath},\bar{\jmath}} \mu^{\oplus}(h) \geq \mu^{\oplus}(f) > 0$, because invariance property of $\mu^{\oplus}$ implies that $\mu^{\oplus}(\overline{h}) = \mu^{\oplus}(f)$. □

Proposition 6.1.18 ensures that our definitions of $\mu^{\oplus}$ and $\mu^{\ominus}$ are precisely the Hahn-Jordan decomposition of μ (cf. [148]).

6.2 Building linear GENEOs via generalized permutant measures

The last step of the chapter is to extend Theorem 6.1.6 in order to include non-expansivity. To this end, it is necessary a last step in the form of Proposition 6.2.1. Such result allows us to connect the generalized permutant measure and the norms of F and φ in the same hypothesis as Theorem 6.1.6.

Proposition 6.2.1 *Assume that $G \subseteq \mathrm{Aut}(X), K \subseteq \mathrm{Aut}(Y)$ transitively acts on the finite set Y, $T\colon G \to K$ is a surjective homomorphism and F is a map from $\mathbb{R}^X$ to $\mathbb{R}^Y$. It holds that*

$$\sum_{h\in X^Y} |\mu(h)| = \max_{\varphi\in\mathbb{R}^X\setminus\{\mathbf{0}\}} \frac{\|F(\varphi)\|_\infty}{\|\varphi\|_\infty}.$$

Proof. The statement is trivially true if F is the null operator, since in this case μ coincides with the null measure on X^Y. Hence, we can assume that F in not the null operator. Setting $c := c^{\oplus} - c^{\ominus}$, Corollary 6.1.15 implies that $|c(h)| = c^{\oplus}(h) + c^{\ominus}(h)$ for every $h \in X^Y$. By definition of $\mu^{\oplus}$ and $\mu^{\ominus}$, we have that, for any $h \in X^Y$

$$\sum_{f\in\mathcal{O}(h)} \mu^{\oplus}(f) = \sum_{f\in\mathcal{O}(h)} c^{\oplus}(f),$$
$$\sum_{f\in\mathcal{O}(h)} \mu^{\ominus}(f) = \sum_{f\in\mathcal{O}(h)} c^{\ominus}(f).$$

It follows that, for each $h \in X^Y$,

$$\begin{aligned}\sum_{f\in\mathcal{O}(h)} |\mu(f)| &\leq \sum_{f\in\mathcal{O}(h)} \mu^{\oplus}(f) + \sum_{f\in\mathcal{O}(h)} \mu^{\ominus}(f)\\ &= \sum_{f\in\mathcal{O}(h)} c^{\oplus}(f) + \sum_{f\in\mathcal{O}(h)} c^{\ominus}(f)\\ &= \sum_{f\in\mathcal{O}(h)} |c(f)|\end{aligned}$$

and hence,

$$\sum_{h\in X^Y} |c(h)| \geq \sum_{h\in X^Y} |\mu(h)|\,. \tag{6.2.1}$$

Now we set $\mathbb{1}_X := \sum_{r=1}^{m} \mathbb{1}_{x_r}$ and $\mathbb{1}_Y := \sum_{s=1}^{n} \mathbb{1}_{y_s}$. We obtain

$$F^{\oplus}\left(\mathbb{1}_X\right) = \sum_{h\in X^Y} c^{\oplus}(h)R(h)\mathbb{1}_X = \left(\sum_{h\in X^Y} c^{\oplus}(h)\right)\mathbb{1}_Y.$$

Since $F^{\oplus}$ (resp. $F^{\ominus}$) is a T-equivariant linear map, any row of $B^{\oplus}$ (resp. $B^{\ominus}$) is a permutation of the first row. Then, we get

$$\begin{aligned}F^{\oplus}(\mathbb{1}_X) = B^{\oplus}\mathbb{1}_X &= \left(\sum_{j=1}^{m} b^{\oplus}_{1j}\right)\mathbb{1}_Y,\\ F^{\ominus}(\mathbb{1}_X) = B^{\ominus}\mathbb{1}_X &= \left(\sum_{j=1}^{m} b^{\ominus}_{1j}\right)\mathbb{1}_Y.\end{aligned}$$

It follows that

$$\begin{aligned}\sum_{h\in X^Y} c^{\oplus}(h) &= \sum_{j=1}^{m} b^{\oplus}_{1j},\\ \sum_{h\in X^Y} c^{\ominus}(h) &= \sum_{j=1}^{m} b^{\ominus}_{1j}.\end{aligned}$$

Therefore

$$\begin{aligned}\sum_{h\in X^Y} |c(h)| &= \sum_{h\in X^Y} c^{\oplus}(h) + \sum_{h\in X^Y} c^{\ominus}(h)\\ &= \sum_{j=1}^{m} b^{\oplus}_{1j} + \sum_{j=1}^{m} b^{\ominus}_{1j}\\ &= \sum_{j=1}^{m} |b_{1j}|\,. \end{aligned} \tag{6.2.2}$$

Moreover, for every $i \in [n]$, we have that

$$\sum_{j=1}^{m} |b_{1j}| = \left|\sum_{j=1}^{m} b_{1j}\mathrm{sgn}(b_{1j})\right| \geq \left|\sum_{j=1}^{m} b_{ij}\mathrm{sgn}(b_{1j})\right|.$$

Considering $\overline{\varphi} = \sum_{j=1}^{m} \operatorname{sgn}(b_{1j})\mathbb{1}_{x_j} \in \mathbb{R}^X \setminus \{\mathbf{0}\}$, it follows that, with a sligh abuse of notation,

$$\begin{aligned} F(\overline{\varphi}) &= B\overline{\varphi} \\ &= B\sum_{j=1}^{m} \operatorname{sgn}(b_{1j})\mathbb{1}_{x_j} \\ &= \sum_{j=1}^{m} \operatorname{sgn}(b_{1j})B\mathbb{1}_{x_j} \\ &= \sum_{j=1}^{m} \operatorname{sgn}(b_{1j})\sum_{i=1}^{n} b_{ij}\mathbb{1}_{y_i} \\ &= \sum_{i=1}^{n}\left(\sum_{j=1}^{m} \operatorname{sgn}(b_{1j})b_{ij}\right)\mathbb{1}_{y_i}. \end{aligned}$$

Hence, from Equations 6.2.1 and 6.2.2 we have that

$$\begin{aligned} \|F(\overline{\varphi})\|_\infty &= \left\|\sum_{i=1}^{n}\left(\sum_{j=1}^{m} \operatorname{sgn}(b_{1j})b_{ij}\right)\mathbb{1}_{y_i}\right\|_\infty \\ &= \sum_{j=1}^{m} |b_{1j}| \\ &= \sum_{h\in X^Y} |c(h)| \geq \sum_{h\in X^Y} |\mu(h)|. \end{aligned}$$

Since $\|\overline{\varphi}\|_\infty = 1$, it follows that $\|F(\overline{\varphi})\|_\infty = \frac{\|F(\overline{\varphi})\|_\infty}{\|\overline{\varphi}\|_\infty} \geq \sum_{h\in X^Y} |\mu(h)|$. In particular,

$$\max_{\varphi\in\mathbb{R}^X\setminus\{\mathbf{0}\}} \frac{\|F(\varphi)\|_\infty}{\|\varphi\|_\infty} \geq \frac{\|F(\overline{\varphi})\|_\infty}{\|\overline{\varphi}\|_\infty} \geq \sum_{h\in X^Y} |\mu(h)|.$$

Now we are going to prove the other verse of the inequality. From Theorem 6.1.6 we have that, for every function $\varphi \in \mathbb{R}^X, F(\varphi) = \sum_{h\in X^Y} \varphi h\mu(h)$. Hence,

$$\begin{aligned} \|F(\varphi)\|_\infty &\leq \sum_{h\in X^Y} \|\varphi h\|_\infty\ |\mu(h)| \\ &\leq \|\varphi\|_\infty \sum_{h\in X^Y} |\mu(h)|. \end{aligned}$$

Therefore, $\frac{\|F(\varphi)\|_\infty}{\|\varphi\|_\infty} \leq \sum_{h\in X^Y} |\mu(h)|$ for every $\varphi \in \mathbb{R}^X \setminus \{\mathbf{0}\}$. In particular, the same applies for the maximum over $\varphi \in \mathbb{R}^X \setminus \{\mathbf{0}\}$. In conclusion,

$$\sum_{h\in X^Y} |\mu(h)| = \max_{\varphi\in\mathbb{R}^X\setminus\{\mathbf{0}\}} \frac{\|F(\varphi)\|_\infty}{\|\varphi\|_\infty}.$$

□

Finally, we are able to state and prove the main result of this chapter, the representation theorem for linear GENEOs.

Theorem 6.2.2 *Assume that $G \subseteq \mathrm{Aut}(X), K \subseteq \mathrm{Aut}(Y)$ transitively acts on the finite set Y, $T\colon G \to K$ is a surjective homomorphism and F is a map from $\mathbb{R}^X$ to $\mathbb{R}^Y$. The map F is a linear group equivariant non-expansive operator from $(\mathbb{R}^X, G)$ to $(\mathbb{R}^Y, K)$ with respect to the homomorphism T if and only if a generalized permutant measure μ exists such that $F(\varphi) = \sum_{h\in X^Y} \varphi h \mu(h)$ for every $\varphi \in \mathbb{R}^X$ and $\sum_{h\in X^Y} |\mu(h)| \leq 1$.*

Proof. Let us assume that F is a group equivariant non-expansive operator from $\big(\mathbb{R}^X, G\big)$ to $\big(\mathbb{R}^Y, K\big)$ with respect to T. Then, Theorem 6.1.6 guarantees that a generalized permutant measure μ exists such that $F(\varphi) = \sum_{h\in X^Y} \varphi h \mu(h)$ for every $\varphi \in \mathbb{R}^X$. Moreover, Proposition 6.2.1 guarantees that $\sum_{h\in X^Y} |\mu(h)| = \max_{\varphi\in\mathbb{R}^X\setminus\{\mathbf{0}\}} \frac{\|F(\varphi)\|_\infty}{\|\varphi\|_\infty}$. Since F is non-expansive, it follows that $\sum_{h\in X^Y} |\mu(h)| \leq 1$, which is the first implication of the statement. Let us now assume that a generalized permutant measure μ exists such that $F(\varphi) = \sum_{h\in X^Y} \varphi h \mu(h)$ for every $p \in \mathbb{R}^X$ with $\sum_{h\in X^Y} |\mu(h)| \leq 1$. Then, Proposition 6.1.5 guarantees that F is a linear group equivariant operator. Moreover, we can prove the non-expansivity of F:

$$\begin{aligned}
\|F(\varphi)\|_\infty &= \left\| \sum_{h\in X^Y} \varphi h \mu(h) \right\|_\infty \\
&\leq \sum_{h\in X^Y} \|\varphi h\|_\infty |\mu(h)| \\
&= \sum_{h\in X^Y} \|\varphi\|_\infty |\mu(h)| \\
&\leq \|\varphi\|_\infty \left(\sum_{h\in X^Y} |\mu(h)| \right) \\
&\leq \|\varphi\|_\infty .
\end{aligned}$$

□

Theorem 6.2.2 represents every linear GENEO between different perception pairs by means of generalized permutant measures. We stress the fact that generalized permutant measures are more easily definable than GENEOs, hence this theorem offers us a valuable tool in the construction of GENEOs. The concept of populating the space of GENEOs with operators that we are able to construct is fundamental in having a wide range of applications. As we will see in the next chapter, we currently have a small number of GENEOs available, which severely limits their use in applications. For this reason, Theorem 6.2.2 is not only important from a theoretical point of view, but also in terms of applicability of the theory.

Chapter 7

New neural network architectures with GENEOs

In the previous chapters we have introduced and explored the concept of GENEO, defined the topological properties of its space and proved that the operator that map data to its persistence diagram is actually a GENEO. From a theoretical point of view, the use of GENEOs in applications can provide an improvement both in accuracy, interpretability and significance of extracted features, since we are injecting specific symmetries in the model. The contribution of this chapter is to provide some first examples of applications of GENEOs. These experiments are preliminary, in the sense that they can be greatly expanded and refined. Nonetheless, these results are significant and highlight the potential that GENEOs can have in large-scale applications.

7.1 MNIST reconstruction with GENEOs

The first application that we are going to introduce deals with image reconstruction. More in detail, we corrupt the MNIST dataset [149] with a random noise that maps 80% of the pixels to the black level. The objective is to reconstruct the original image. Since we are dealing with images, we require that our model is equivariant with translations and rotations. We recall that a convolutional neural network (CNN) [150] is a GEO with respect to translations. In order to include non-expansivity and equivariance with respect to rotations, we have modified a CNN with the additional request that kernels are symmetric with respect to rotations. More in detail, each kernel of our neural network is the discretization of a rotated and normalized shifted Gaussian around the center of the kernel. Therefore, our kernel has only three parameters: the center of the Gaussian, its amplitude and its scale. An example of an admissible kernel for our network is shown in Figure 7.1. Finally, the last layer of our network is a convex combination of the outputs of the previous layer. This is done to ensure that the network is non-expansive. More in detail, our neural network architecture has four layers with $\{6, 16, 16, 16\}$ convolutions, respectively, each with kernel size 5. Since each kernel has three parameters, this results in a total of $1,842$ parameters. As a competitor, we have trained both an Autoencoder [151] and a Variational Autoencoder [152], which have been successfully applied in denoising [153]. We will not dive into the details of both of these networks, but the number of trainable parameters is $1,732,913$ and $838,337$, respectively. For the hyperparameters of the training, we used 40 epochs, a batch size of 32, a learning rate of 0.001, mean squared error loss and stochastic gradient descent optimizer. We report a graphical example of the reconstructed images for our model in Figure 7.2 and for the variational autoencoder in Figure 7.3. Results for the autoencoder are similar to the variational autoencoder, so they are not reported. The results achieved by our network are satisfactory, while the variational autoencoder is not able to learn anything. This behavious is further validate by the analysis of the loss (Figure 7.4). The training loss of the GENEO network is clearly decreasing (Figure 7.4a, especially during the first epochs), while the training loss of the variational autoencoder is overall constant during the epochs (Figure 7.4b. Moreover, the test loss of the GENEO network is greatly smaller than the respective loss for the variational autoencoder. In this first experiment, we have shown that our GENEO network is able to perform tasks that competitors are not able to perform, with satisfactory results from a human perspective and with a small amount of trainable parameters.

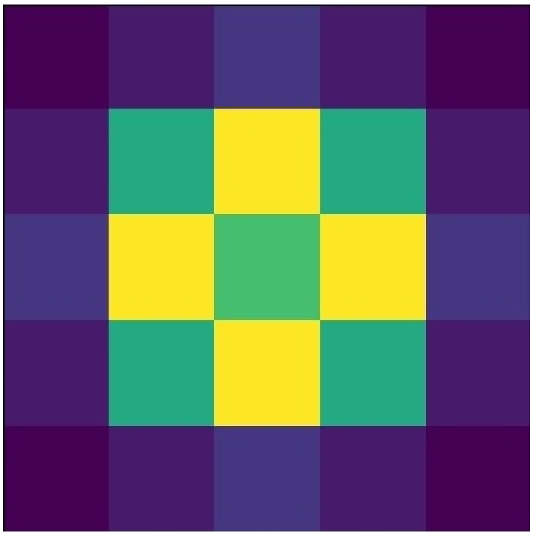


Figure 7.1: Example of a kernel of the GENEO neural network.

7.2 MNIST classification with GENEOs

The second application that we are going to present is an MNIST classification. We want to highlight that, in our current implementation, a classification task requires a final fully connected linear layer that is not guaranteed to be equivariant. This detail is very important and requires further examination. It is well known that any feedforward fully connected neural network is theoretically able to learn any decision function: it is stated in the Universal Approximation Theorem [154]. Therefore, in theory, there should be no need of the plethora of neural network architectures that are being developed. However, in real-world scenarios, the situation is quite different. Usually, fully connected neural networks learn patterns that are

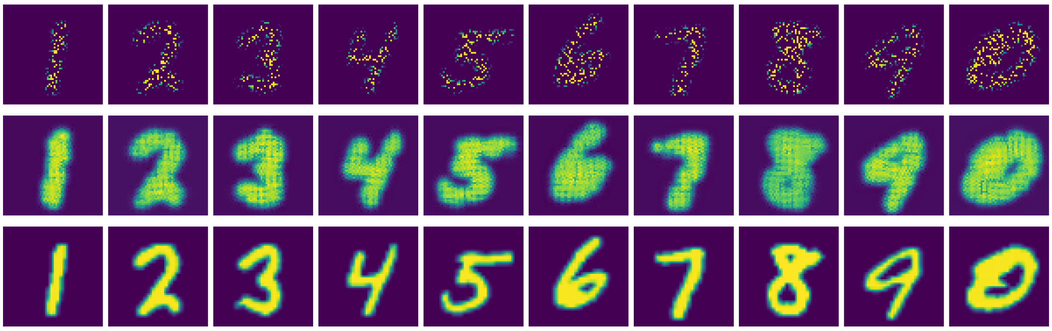


Figure 7.2: Example of corrupted MNIST digits (first row), reconstructed digits with the GENEO network (middle row) and original image (last row).

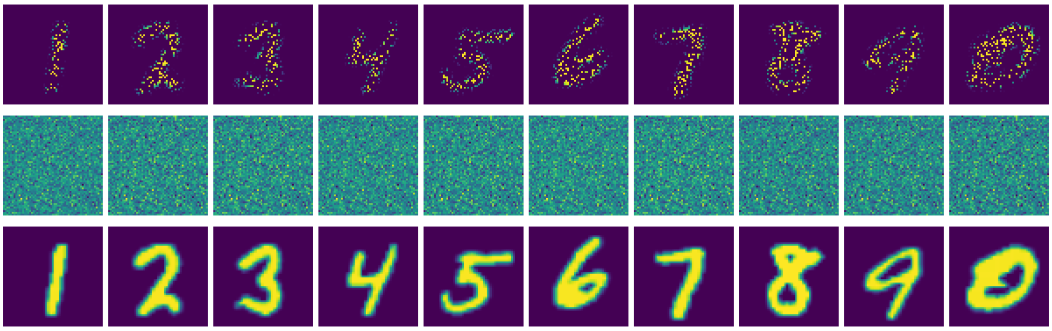
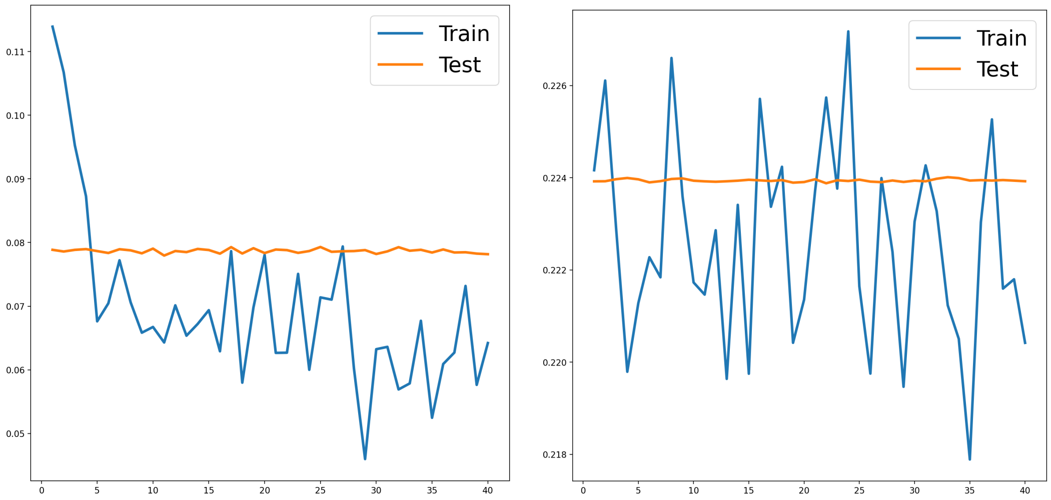


Figure 7.3: Example of corrupted MNIST digits (first row), reconstructed digits with the variational autoencoder (middle row) and original image (last row).

specific to the dataset at hand, and not features that are intrinsically important for the task. This is what is usually referred to as overfitting. This is the reason why so many neural network architectures have been devised: changing the connection of the layers (i. e. the topology of the network) we force the network to learn features that are more suitable to the task. Typically such features are more robust to noise or require fewer preprocessing steps. The most successful neural network architecture for computer vision tasks is the convolutional neural network. Although its development followed a different path, a convolutional network is translation equivariant and this property is part of its success. Indeed, it guarantees that the network learns to recognize the details of the image regardless of its position. Incidentally, this also shrinks the number of parameters (and thus the size of the space on which to do optimization). This yields a more robust model that requires a smaller training dataset. Of course, these excellent mathematical properties are lost once you insert a linear, fully connected layers at the end of the neural network to do classification. The hope, which will be confirmed empirically, is that having learned equivariant features up to a certain layer, the last linear layers will equally maintain such equivariance. A possible solution for this problem is to insert an invariant layer after the equivariant ones. This would guarantee that any following layers would still be equivariant, but in our current implementation, we do not have at our disposal such a layer. That being said, despite the networks that we introduce are not strictly GENEOs, we are still going to refer them as such. For comparison, we are



(a) Loss for the GENEO network w.r.t. epochs.

(b) Loss for the variational autoencoder w.r.t. epochs.

Figure 7.4: Loss during training and testing of both GENEO network and variational autoencoder w.r.t. epochs.

going to test three different neural network architectures: the first one is a fully connected neural network (FNN), the second one is a convolutional neural network (CNN) suitably modified in order to introduce non-expansivity (hence resulting in a GENEO network with respect to translations) and a GENEO network with respect to translations and rotations with the same kernels as the ones described in the previous section. In particular, we are going to perform three experiments. The first one is the normal classification of MNIST. The second experiment is the training on MNIST and testing of the trained networks to a rotated version of MNIST. The third experiment is the training and testing on a rotated version of MNIST. More in detail, we have the following three neural networks:

- **GENEO network**: four layers of GENEO convolutions with $\{16, 32, 64, 32\}$ kernels, respectively, each kernel of size 7. The activation function is `tanh` and there is a `maxpool` after each layer. For increased specificity, we allowed for three Gaussians in each kernel. This is followed by two fully connected layers with `leakyrelu` activation functions. The total number of trainable parameters is $207, 130$;

- **CNN**: two convolution layers with $\{6, 10\}$ kernels, respectively, each kernel of size 7. The activation function is `relu` and there is a `maxpool` after each layer. This is followed by two fully connected layers with `relu` activation functions. The total number of trainable parameters is $211, 538$;

- **FNN**: three fully connected layers with `relu` activation functions and a total number of trainable parameters of $263, 650$.

The three neural networks have been devised in order to have a similar number of parameters. Moreover, we upscaled the MNIST dataset to a 157×157 pixel image using the `torchvision.transforms.resize` built-in function.

7.2.1 Training and testing on MNIST

For the first experiment, all three neural networks achieve very similar results. In particular, the GENEO network achieves a 91% accuracy on the test set, the CNN an accuracy of 90% and FNN achieves a 88% accuracy. This was to be expected both because of the simplicity of the dataset and the heavy preprocessing that was done on MNIST. We refer to Figure 7.5 for the plot of train and test loss and accuracy during the experiment for the GENEO network. The two other neural networks share similar behavior and their plot were omitted.

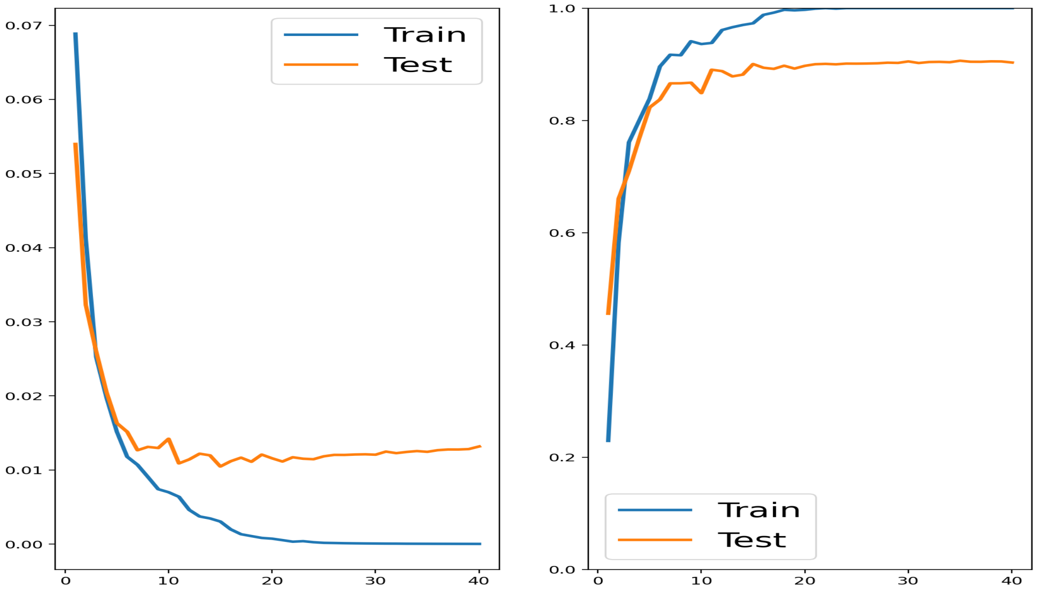


Figure 7.5: Train and test loss (left) and accuracy (right) for the GENEO network in the first experiment w.r.t. epochs.

7.2.2 Training on MNIST, testing on rotated MNIST

In the second experiment, all three neural networks are trained in the same fashion as in the previous experiment, but they are tested on a version of MNIST where images are rotated by a random angle $\in [-60°, 60°]$. In this case, the accuracy results on the test set are $66\%, 55\%$ and 53% respectively for GENEO network, CNN and FNN. This result is encouraging and shows that features extracted from the GENEO network are, in fact, more equivariant with respect to rotations than linear layers or CNNs. Therefore, despite the last linear layers do not guarantee that the equivariance is maintained throughout the neural network, this occurs at least partially. We refer to Figure 7.6 for an example of rotated MNIST digits and to Figure 7.7 for the confusion matrix of the GENEO network.

In this experiment, we can truly dive deep in how our neural network is working. In particular, in Figure 7.8 we show three trained kernels and how they act on the different digits. Blue colors correspond to negative components, and red to positive ones. In particular, the kernel shown in Figure 7.8a seems to detect the presence of circles (or nearly) in the digits, since the outputs of the digits $0, 3, 5, 6, 8, 9$ are

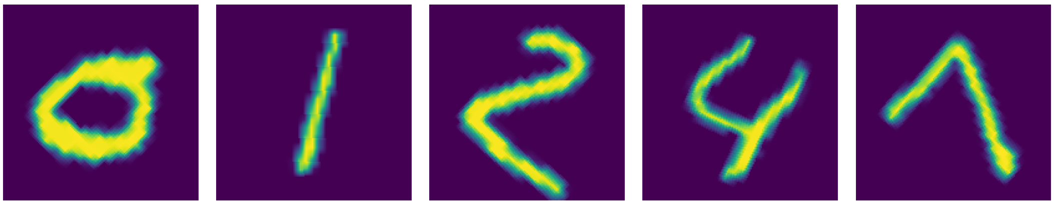


Figure 7.6: Example of rotated MNIST digits.

the only ones with positive and negative components. A similar behaviour occurs in Figure 7.8b, but such kernels seems to detect curvature, more than circles. This is supposedly to the fact that the kernel reacts to the digit 2 in the same way as the other digits with circles. We highlight the fact that the two kernels are similar in shape. Finally, Figure 7.8c shows a kernel that seems to detect segments in the digit. Another interesting concept to linger is the network activation changes when passing a digit and its rotated version. In the first row of Figure 7.9, we can see some MNIST digits and their rotated versions. In the second row there is the respective output of a GENEO kernel, which is equivariant with respect to rotations (up to finite algebra approximations). Finally, we can measure the amount of equivariance of the complete network by tracking the classification output of the GENEO network for a digit and some random rotations. It turns out that the equivariance of the GENEO network (we stress, with the addition of two final linear layers) is roughly 68%. This is quite encouraging and can explain the reason why our network outperforms the competitors in such a scenario.

7.2.3 Training and testing on rotated MNIST

In our last experiment, both the train and test datasets consist of the rotated version of MNIST. The accuracy result for the GENEO network is 87%, for the CNN is 88% and for the FNN is 85%. This result in particular shows the potentials and limitations of GENEOs. Especially in benchmark contexts, where the dataset is incredibly neat and heavily preprocessed, the usefulness of injecting equivariance in the model is mitigated but such preprocessing steps. However, it is sufficient to introduce slight alterations to what a competitor is accustomed to and the network suffers greatly. The GENEO network, on the other hand, is more robust to such alterations thanks to its equivariance. In relatively simple cases such as MNIST or rotated MNIST, a retraining of the neural network is sufficient to regain the original accuracy for all networks, but in more complex datasets this may not be the case.

In conclusion, the applications of GENEOs are still in their infancy, since not many GENEO layers have been devised yet. Nonetheless, they can offer an improvement both in accuracy and transparency and provide features that are more robust and significant to specific transformations that are known a priori. Moreover, the limited amount of parameters required helps the prevention of overfitting and in general requires a smaller dataset to be adequately trained. This can be a key factor in real-world datasets, like the one presented in Chapter 3.

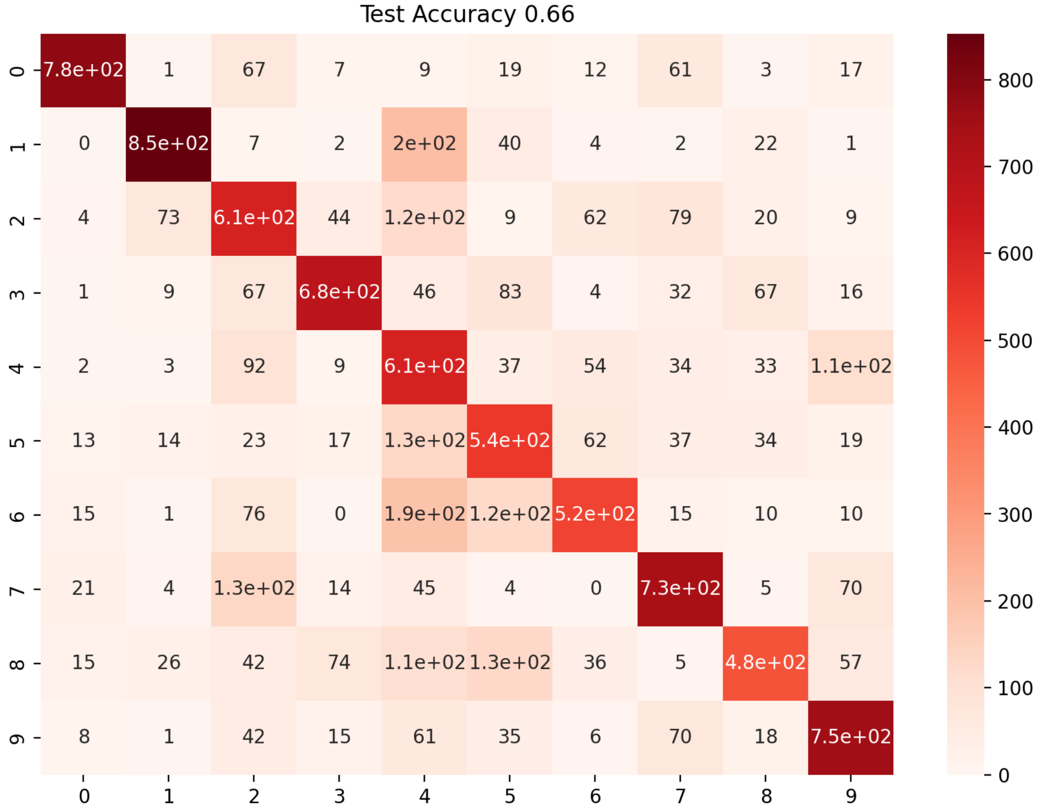
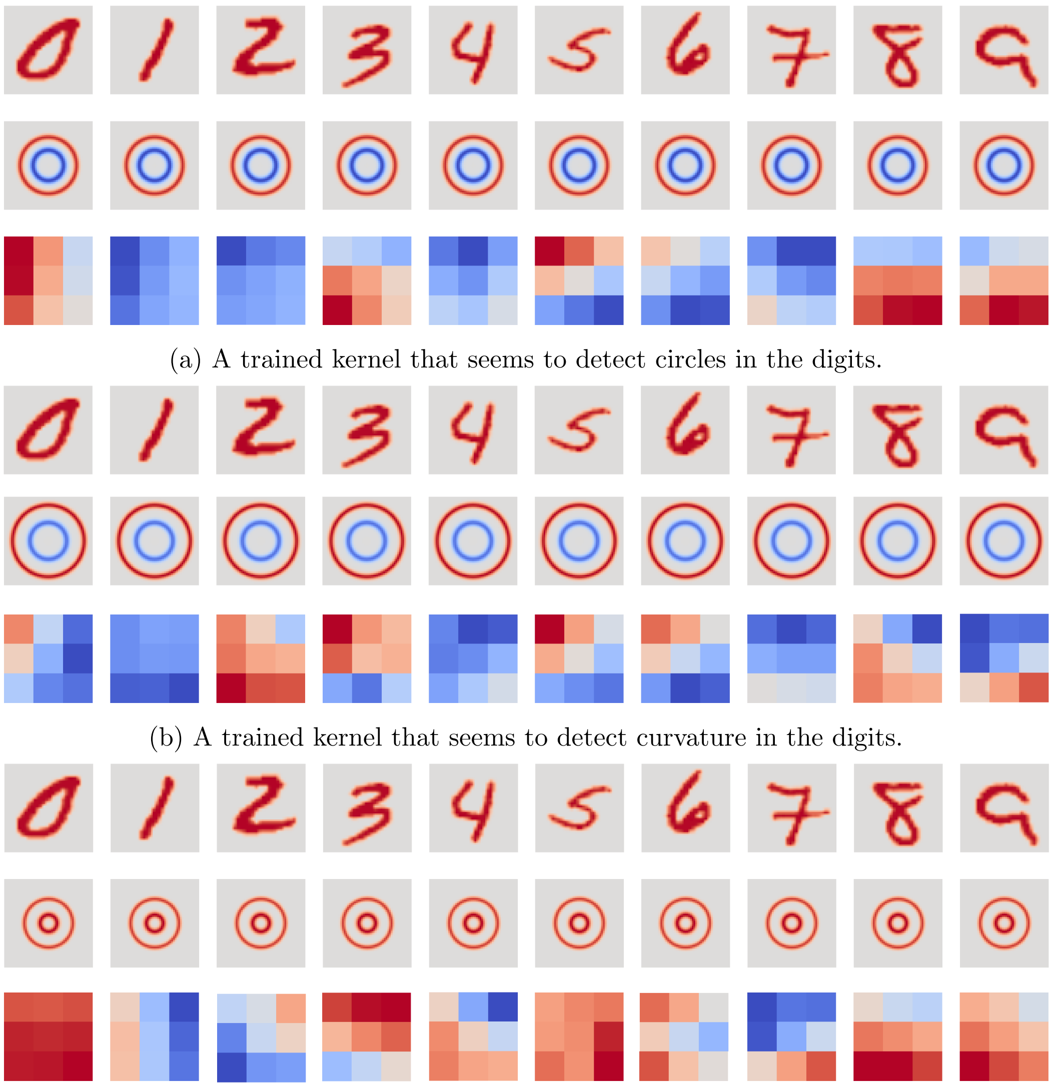


Figure 7.7: Confusion matrix of the GENEO network trained on MNIST and tested on rotated MNIST.



(a) A trained kernel that seems to detect circles in the digits.

(b) A trained kernel that seems to detect curvature in the digits.

(c) A trained kernel that seems to detect segments in the digits.

Figure 7.8: Example of trained kernels and their actions on the different digits.

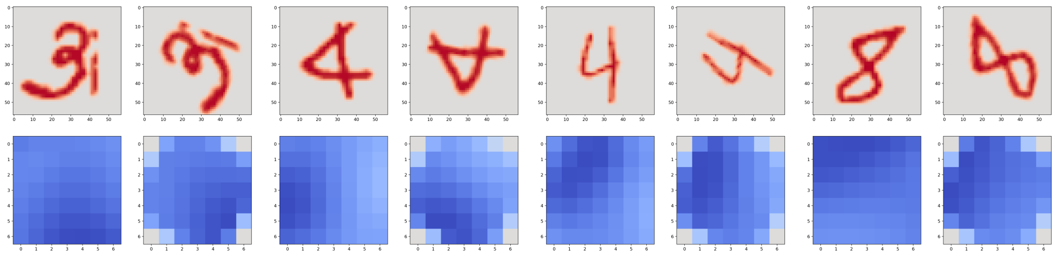


Figure 7.9: Equivariance of the GENEO network. First row: MNIST digits and their rotated version. Second row: the output of a GENEO kernel.

Chapter 8

Conclusions

The focus of this dissertation was topological data analysis. From an applicative point of view, we focused on the development of a pipeline that collects the major literature on the topic and apply it to new case studies, according to the ISTI-CNR laboratory availability. In this area, we proposed a uniform methodology in a non-canonical part of the literature, supported by our empirical findings. From a theoretical point of view, we firstly contributed to the connection of topological data analysis with a different but related concept, namely group equivariant non-expansive operators. These operators can be thought of as an expansion of the mathematical core of topological data analysis and show great potential, despite being relatively underdeveloped. In this field, we contributed with two new methodologies to generate them and with a couple of applications based on the development of a neural network based on them.

The first contribution of this dissertation consists in the development of a topological machine learning pipeline for the classification of digital data. Such a pipeline combines the features arising from computational topology for describing the data shape with the algorithms of machine learning and generates a novel, fully automated and powerful tool for data classification. More importantly, in this dissertation we proposed a novel approach to standardise a non-canonical step of the topological machine learning pipeline. In addition, our approach is supported by a noise resistance test, which is a first of this kind.

The second contribution of this dissertation consists in the application of the topological machine learning pipeline devised alongside this thesis to three novel case studies, in the form of real-world datasets coming from the biomedical and meteorological fields. In all three approaches we achieved state-of-the-art accuracies, and the biomedical scenarios were particularly satisfactory. These applications further develop the literature of TML applications, with is still somewhat limited.

The third contribution of this dissertation consists in a new connection between TDA and GENEOs, which constitutes the two main topics of this thesis. In particular, we prove that the functor that maps each function to its associated persistence diagram and maps a homeomorphism between functions to the induced matching between persistence diagrams (with a few precautions) is a group equivariant non-expansive operator. Thus, we can consider the computation of PDs as an element of the space of all group equivariant non-expansive operators. This theory is based on a dual approach that focuses on pairs data-group of equivariances. The rationale behind the use of operators that are blind to the action of the group is that we want to inject into the system preexisting knowledge, hence drastically reducing the dimensionality of the space. Moreover, the non-expansivity encodes the idea that we are interested in operators that compress the information we have in input, and grant compactness to the space of such operators. This property is fundamental in applications since it allows us to approximate such space of operators with just a finite set. In particular, this theory expand on the concept of persistent homology, allowing for more flexibility of the group of equivariance and for viewpoints that may differ from the topological one. In literature the connection between TDA and GENEOs is not new, but this thesis provides a new bridge between these concepts from a functorial point of view.

The fourth and fifth contributions of this dissertation consists of a new method to build non-linear GENEOs by means of symmetric functions and permutants (Theorem 5.3.24) and a characterization theorem of linear GENEOs between arbitrary functional spaces (Theorem 6.2.2). Both these theorems are novel in the GENEO theory and develop our ability to generate them, which is currently one of the bottleneck for applications of this framework.

The last contribution of this dissertation consists in the development of a new neural network architecture based on GENEOs. Such a network is preliminary but still expressive enough to outperform competitors in two applications, highlighting the potential of our mathematical framework.

We conclude by mentioning some new lines of research for this mathematical model. One of the most difficult aspects of this setting consists in building equivariant operators given an equivariance group. In this dissertation, we have developed a method to build linear GENEOs by means of permutants, which are in general easier to define. Nevertheless, new methods should be developed in order to get a good approximation of the space of GENEOs. Furthermore, in many applications group action may be too restrictive. As an example, a dataset consisting of digit images would not admit a group of rotations as equivariances, since the resulting operator would be blind to the digits “6” and “9”. In general, we would like less algebraic structure in order to have a more versatile mathematical model. An advancement in this research would be to generalize the framework to equivariance with respect to monoid of couplings instead on groups of bijections. Another possible development consists of a GENEO network more developed than the preliminary version presented in this dissertation. Finally, we want to highlight the potential of the GENEO network with real-world data. In Chapter 3 we highlighted the fact that current neural networks underperformed in scenarios where data are scarce. In all three real-world datasets, simple machine learning classifiers outperformed more sophisticated neural network precisely due to the scarcity of data to train on. However, this issue may be resolved with GENEO networks, since the narrowness of trainable parameters (due to the symmetries injected in the model) would also require less training data. In the future, when a more developed GENEO network is available, we aim to apply it to this kind of datasets as well.

Bibliography

[1] Gunnar Carlsson. Topology and data. *Bull. Amer. Math. Soc. (N.S.)*, 46(2):255–308, 2009.

[2] Richard Bellman and Robert Kalaba. On adaptive control processes. *IRE Transactions on Automatic Control*, 4(2):1–9, 1959.

[3] Wojciech Samek, Thomas Wiegand, and Klaus-Robert Müller. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv preprint arXiv:1708.08296*, 2017.

[4] Rickard Brüel Gabrielsson and Gunnar Carlsson. Exposition and interpretation of the topology of neural networks. In *2019 18th ieee international conference on machine learning and applications (icmla)*, pages 1069–1076. IEEE, 2019.

[5] G Carlsson and R Gabrielsson. Topological approaches to deep learning, topological data analysis. In *Abel Symposia*, volume 15, 2020.

[6] Herbert Edelsbrunner and John Harer. Persistent homology—a survey. In *Surveys on discrete and computational geometry*, volume 453 of *Contemp. Math.*, pages 257–282. Amer. Math. Soc., Providence, RI, 2008.

[7] Peter Bubenik and Jonathan A Scott. Categorification of persistent homology. *Discrete & Computational Geometry*, 51(3):600–627, 2014.

[8] Gunnar Carlsson and Afra Zomorodian. The theory of multidimensional persistence. In *Proceedings of the twenty-third annual symposium on Computational geometry*, pages 184–193, 2007.

[9] Frédéric Chazal, David Cohen-Steiner, Marc Glisse, Leonidas J Guibas, and Steve Y Oudot. Proximity of persistence modules and their diagrams. In *Proceedings of the twenty-fifth annual symposium on Computational geometry*, pages 237–246, 2009.

[10] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. In *Proceedings of the twenty-first annual symposium on Computational geometry*, pages 263–271, 2005.

[11] Herbert Edelsbrunner. Persistent homology: theory and practice. 2013.

[12] Mattia G. Bergomi, Patrizio Frosini, Daniela Giorgi, and Nicola Quercioli. Towards a topological–geometrical theory of group equivariant non-expansive operators for data analysis and machine learning. *Nature Machine Intelligence*, 1(9):423–433, Sep 2019.

[13] Henry Adams, Tegan Emerson, Michael Kirby, Rachel Neville, Chris Peterson, Patrick Shipman, Sofya Chepushtanova, Eric Hanson, Francis Motta, and Lori Ziegelmeier. Persistence images: A stable vector representation of persistent homology. *Journal of Machine Learning Research*, 18, 2017.

[14] Luis Polanco and Jose A Perea. Adaptive template systems: Data-driven feature selection for learning with persistence diagrams. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pages 1115–1121. IEEE, 2019.

[15] Peter Bubenik et al. Statistical topological data analysis using persistence landscapes. *J. Mach. Learn. Res.*, 16(1):77–102, 2015.

[16] Francesco Conti, Mario D'Acunto, Claudia Caudai, Sara Colantonio, Raffaele Gaeta, Davide Moroni, and Maria Antonietta Pascali. Raman spectroscopy and topological machine learning for cancer grading. *Scientific Reports*, 13(1):7282, 2023.

[17] Francesco Conti, Martina Banchelli, Valentina Bessi, Cristina Cecchi, Fabrizio Chiti, Sara Colantonio, Cristiano D'Andrea, Marella de Angelis, Davide Moroni, Benedetta Nacmias, Maria Antonietta Pascali, Sandro Sorbi, and Paolo Matteini. Alzheimer disease detection from raman spectroscopy of the cerebrospinal fluid via topological machine learning. *Engineering Proceedings*, 51(1), 2023.

[18] Francesco Conti, Oscar Papini, Davide Moroni, Gabriele Pieri, Marco Reggiannini, and Maria Antonietta Pascali. Analysis of sea surface temperature maps via topological machine learning. In *2023 IX International Conference on Information Technology and Nanotechnology (ITNT)*, pages 1–4, 2023.

[19] Max Z Li, Megan S Ryerson, and Hamsa Balakrishnan. Topological data analysis for aviation applications. *Transportation Research Part E: Logistics and Transportation Review*, 128:149–174, 2019.

[20] Alexander D Smith, Paweł Dłotko, and Victor M Zavala. Topological data analysis: concepts, computation, and applications in chemical engineering. *Computers & Chemical Engineering*, 146:107202, 2021.

[21] Peter Lawson, Andrew B Sholl, J Quincy Brown, Brittany Terese Fasy, and Carola Wenk. Persistent homology for the quantitative evaluation of architectural features in prostate cancer histology. *Scientific reports*, 9(1):1139, 2019.

[22] Andrew Aukerman, Mathieu Carrière, Chao Chen, Kevin Gardner, Raúl Rabadán, and Rami Vanguri. Persistent homology based characterization of

the breast cancer immune microenvironment: A feasibility study. *Journal of Computational Geometry*, 12(2):183–206, 2022.

[23] BJ Stolz, J Kaeppler, B Markelc, F Mech, F Lipsmeier, RJ Muschel, HM Byrne, and HA Harrington. Multiscale topology characterises dynamic tumour vascular networks. arxiv. *preprint*, 2020.

[24] Pablo G Cámara. Topological methods for genomics: present and future directions. *Current opinion in systems biology*, 1:95–101, 2017.

[25] Joseph Minhow Chan, Gunnar Carlsson, and Raul Rabadan. Topology of viral evolution. *Proceedings of the National Academy of Sciences*, 110(46):18566–18571, 2013.

[26] Giovanni Bocchi, Patrizio Frosini, Alessandra Micheletti, Alessandro Pedretti, Carmen Gratteri, Filippo Lunghini, Andrea Rosario Beccari, and Carmine Talarico. Geneonet: A new machine learning paradigm based on group equivariant non-expansive operators. an application to protein pocket detection. *arXiv preprint arXiv:2202.00451*, 2022.

[27] Kevin Emmett, Benjamin Schweinhart, and Raul Rabadan. Multiscale topology of chromatin folding. *arXiv preprint arXiv:1511.01426*, 2015.

[28] Nieves Atienza, Rocio Gonzalez-Díaz, and Manuel Soriano-Trigueros. On the stability of persistent entropy and new summary functions for topological data analysis. *Pattern Recognition*, 107:107509, 2020.

[29] Jan Reininghaus, Stefan Huber, Ulrich Bauer, and Roland Kwitt. A stable multi-scale kernel for topological machine learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4741–4748, 2015.

[30] Barbara Di Fabio and Massimo Ferri. Comparing persistence diagrams through complex vectors. In Vittorio Murino and Enrico Puppo, editors, *Image Analysis and Processing — ICIAP 2015*, pages 294–305, Cham, 2015. Springer International Publishing.

[31] Frédéric Chazal, Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, and Larry Wasserman. Stochastic convergence of persistence landscapes and silhouettes. In *Proceedings of the Thirtieth Annual Symposium on Computational Geometry*, SOCG'14, page 474–483, New York, NY, USA, 2014. Association for Computing Machinery.

[32] Eashwar Somasundaram, Adam Litzler, Raoul Wadhwa, Steph Owen, and Jacob Scott. Persistent homology of tumor ct scans is associated with survival in lung cancer. *Medical physics*, 48(11):7043–7051, 2021.

[33] Bastian Rieck, Tristan Yates, Christian Bock, Karsten Borgwardt, Guy Wolf, Nicholas Turk-Browne, and Smita Krishnaswamy. Uncovering the topology of time-varying fmri data using cubical persistence. *Advances in neural information processing systems*, 33:6900–6912, 2020.

[34] Audun Myers, Henry Kvinge, and Tegan Emerson. Topfusion: Using topological feature space for fusion and imputation in multi-modal data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 600–609, June 2023.

[35] Danielle Barnes, Luis Polanco, and Jose A Perea. A comparative study of machine learning methods for persistence diagrams. *Frontiers in Artificial Intelligence*, 4:681174, 2021.

[36] Yu-Min Chung and Austin Lawson. Persistence curves: A canonical framework for summarizing persistence diagrams. *Advances in Computational Mathematics*, 48(1):6, 2022.

[37] Dashti Ali, Aras Asaad, Maria-Jose Jimenez, Vidit Nanda, Eduardo Paluzo-Hidalgo, and Manuel Soriano-Trigueros. A survey of vectorization methods in topological data analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[38] Sarah Tymochko, E. Munch, and Firas A. Khasawneh. Adaptive partitioning for template functions on persistence diagrams. *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pages 1227–1234, 2019.

[39] Herbert Edelsbrunner and John L Harer. *Computational topology: an introduction*. American Mathematical Society, 2022.

[40] Silvia Biasotti, Andrea Cerri, Patrizio Frosini, Daniela Giorgi, and Claudia Landi. Multidimensional size functions for shape comparison. *Journal of Mathematical Imaging and Vision*, 32:161–179, 2008.

[41] Andrea Cerri, Barbara Di Fabio, Massimo Ferri, Patrizio Frosini, and Claudia Landi. Betti numbers in multidimensional persistent homology are stable functions. *Mathematical Methods in the Applied Sciences*, 36(12):1543–1557, 2013.

[42] Fabio Anselmi, Lorenzo Rosasco, and Tomaso Poggio. On invariance and selectivity in representation learning. *Information and Inference: A Journal of the IMA*, 5(2):134–158, 2016.

[43] Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pages 2990–2999, 2016.

[44] Diego Marcos, Michele Volpi, Nikos Komodakis, and Devis Tuia. Rotation equivariant vector field networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5048–5057, 2017.

[45] Jonathan Masci, Davide Boscaini, Michael Bronstein, and Pierre Vandergheynst. Geodesic convolutional neural networks on riemannian manifolds. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 37–45, 2015.

[46] Yann LeCun, Yoshua Bengio, et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995, 1995.

[47] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828, 2013.

[48] Patrizio Frosini. Towards an observer-oriented theory of shape comparison. In *3DOR@ Eurographics*, 2016.

[49] Giovanni Bocchi, Stefano Botteghi, Martina Brasini, Patrizio Frosini, and Nicola Quercioli. On the finite representation of linear group equivariant operators via permutant measures. *Annals of Mathematics and Artificial Intelligence*, 2023. DOI: 10.1007/s10472-022-09830-1, to appear.

[50] Patrizio Frosini and Nicola Quercioli. Some remarks on the algebraic properties of group invariant operators in persistent homology. In Andreas Holzinger, Peter Kieseberg, A Min Tjoa, and Edgar Weippl, editors, *1st International Cross-Domain Conference for Machine Learning and Knowledge Extraction (CD-MAKE)*, volume LNCS-10410 of *Machine Learning and Knowledge Extraction*, pages 14–24, Reggio, Italy, August 2017. Springer International Publishing. Part 1: MAKE Topology.

[51] Francesco Camporesi, Patrizio Frosini, and Nicola Quercioli. On a new method to build group equivariant operators by means of permutants. In Andreas Holzinger, Peter Kieseberg, A Min Tjoa, and Edgar Weippl, editors, *2nd International Cross-Domain Conference for Machine Learning and Knowledge Extraction (CD-MAKE)*, volume LNCS-11015 of *Machine Learning and Knowledge Extraction*, pages 265–272, Hamburg, Germany, August 2018. Springer International Publishing. Part 4: MAKE-Topology.

[52] Patrizio Frosini and Grzegorz Jabłoński. Combining persistent homology and invariance groups for shape comparison. *Discrete & Computational Geometry*, 55(2):373–409, 2016.

[53] Nicola Quercioli. *On the topological theory of Group Equivariant Non-Expansive Operators*. PhD thesis, alma, Maggio 2021.

[54] Francesco Conti, Patrizio Frosini, and Nicola Quercioli. On the construction of group equivariant non-expansive operators via permutants and symmetric functions. *Frontiers in Artificial Intelligence*, 5:786091, 2022.

[55] Faraz Ahmad, Massimo Ferri, and Patrizio Frosini. Generalized permutants and graph geneos. *Machine Learning and Knowledge Extraction*, 5(4):1905–1920, 2023.

[56] Giovanni Bocchi, Alessandra Micheletti, P Frosini, A Pedretti, C Gratteri, F Lunghini, AR Beccari, C Talarico, et al. A new paradigm for artificial intelligence based on group equivariant non-expansive operators (geneos) applied

to protein pocket detection. In *Proceedings of the Statistics and Data Science Conference*, pages 152–157. Pavia University Press, 2023.

[57] Francesco Conti, Davide Moroni, and Maria Antonietta Pascali. A topological machine learning pipeline for classification. *Mathematics*, 10(17):3086, 2022.

[58] Allen Hatcher. *Algebraic topology.* Cambridge University Press, Cambridge, 2002.

[59] Alessandro Verri, Claudio Uras, Patrizio Frosini, and Massimo Ferri. On the use of size functions for shape analysis. *Biological cybernetics*, 70(2):99–107, 1993.

[60] Felix Hensel, Michael Moor, and Bastian Rieck. A survey of topological machine learning methods. *Frontiers in Artificial Intelligence*, 4:681108, 2021.

[61] Glen E Bredon. *Topology and geometry*, volume 139. Springer Science & Business Media, 2013.

[62] Herbert Federer. Curvature measures. *Transactions of the American Mathematical Society*, 93(3):418–491, 1959.

[63] George W Whitehead. *Elements of homotopy theory*, volume 61. Springer Science & Business Media, 2012.

[64] S. S. Cairns. Triangulation of the manifold of class one. *Bulletin of the American Mathematical Society*, 41(8):549 – 552, 1935.

[65] J. H. C. Whitehead. On c1-complexes. 41(4):809–824, 1940.

[66] Charles Fefferman, Sanjoy Mitter, and Hariharan Narayanan. Testing the manifold hypothesis. *Journal of the American Mathematical Society*, 29(4):983–1049, 2016.

[67] Julius Von Rohrscheidt and Bastian Rieck. Topological singularity detection at multiple scales. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 35175–35197. PMLR, 23–29 Jul 2023.

[68] Gunnar Carlsson and Afra Zomorodian. The theory of multidimensional persistence. *Discrete Comput. Geom.*, 42(1):71–93, 2009.

[69] M. Zarichnyi, A. Savchenko, and V. Kiosak. Strong topology on the set of persistence diagrams. *AIP Conference Proceedings*, 2164(1):040006, 10 2019.

[70] David H Wolpert and William G Macready. No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, 1(1):67–82, 1997.

[71] Chi Seng Pun, Si Xian Lee, and Kelin Xia. Persistent-homology-based machine learning: A survey and a comparative study. *Artif. Intell. Rev.*, 55(7):5169–5213, oct 2022.

[72] Aaron B. Adcock, Erik Carlsson, and Gunnar E. Carlsson. The ring of algebraic functions on persistence bar codes. *arXiv: Rings and Algebras*, 2013.

[73] Sara Kališnik. Tropical coordinates on the space of persistence barcodes. *Foundations of Computational Mathematics*, 19(1):101–129, 2019.

[74] Gunnar Carlsson and Sara Kališnik Verovšek. Symmetric and r-symmetric tropical polynomials and rational functions. *Journal of Pure and Applied Algebra*, 220(11):3610–3627, 2016.

[75] Massimo Ferri and Claudia Landi. Representing size functions by complex polynomials. *Proc. Math. Met. in Pattern Recognition*, 9:16–19, 1999.

[76] Barbara Di Fabio and Massimo Ferri. Comparing persistence diagrams through complex vectors. In *Image Analysis and Processing—ICIAP 2015: 18th International Conference, Genoa, Italy, September 7-11, 2015, Proceedings, Part I 18*, pages 294–305. Springer, 2015.

[77] Yuhei Umeda. Time series classification via topological data analysis. *Information and Media Technologies*, 12:228–239, 2017.

[78] Yu-Min Chung and Austin Lawson. Persistence curves: A canonical framework for summarizing persistence diagrams. *Advances in Computational Mathematics*, 48(1):6, 2022.

[79] Frédéric Chazal, Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, and Larry Wasserman. Stochastic convergence of persistence landscapes and silhouettes. In *Proceedings of the thirtieth annual symposium on Computational geometry*, pages 474–483, 2014.

[80] Jose A Perea, Elizabeth Munch, and Firas A Khasawneh. Approximating continuous functions on persistence diagrams using template functions. *Foundations of Computational Mathematics*, 23(4):1215–1272, 2023.

[81] Martin Royer, Frédéric Chazal, Clément Levrard, Yuhei Umeda, and Yuichi Ike. Atol: measure vectorization for automatic topologically-oriented learning. In *International Conference on Artificial Intelligence and Statistics*, pages 1000–1008. PMLR, 2021.

[82] Jean-Claude Hausmann et al. On the vietoris-rips complexes and a cohomology theory for metric spaces. *Annals of Mathematics Studies*, 138:175–188, 1995.

[83] Tomasz Kaczynski, Konstantin Michael Mischaikow, and Marian Mrozek. *Computational homology*, volume 3. Springer, 2004.

[84] S. Biasotti, L. De Floriani, B. Falcidieno, P. Frosini, D. Giorgi, C. Landi, L. Papaleo, and M. Spagnuolo. Describing shapes by geometrical-topological properties of real functions. *ACM Comput. Surv.*, 40(4):12:1–12:87, October 2008.

[85] Mathieu Carriere, Marco Cuturi, and Steve Oudot. Sliced wasserstein kernel for persistence diagrams. In *International conference on machine learning*, pages 664–673. PMLR, 2017.

[86] Zhi-Hua Zhou. *Machine learning*. Springer Nature, 2021.

[87] The GUDHI Project. *GUDHI User and Reference Manual*. GUDHI Editorial Board, 3.9.0 edition, 2023.

[88] Guillaume Tauzin, Umberto Lupo, Lewis Tunstall, Julian Burella Pérez, Matteo Caorsi, Anibal Medina-Mardones, Alberto Dassatti, and Kathryn Hess. giotto-tda: A topological data analysis toolkit for machine learning and data exploration, 2020.

[89] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[90] Ad'elie Garin and Guillaume Tauzin. A topological "reading" lesson: Classification of mnist using tda. *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pages 1551–1556, 2019.

[91] Jan Reininghaus, Stefan Huber, Ulrich Bauer, and Roland Kwitt. A stable multi-scale kernel for topological machine learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[92] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms. *arXiv e-prints*, page arXiv:1708.07747, August 2017.

[93] Henry Adams, Tegan Emerson, Michael Kirby, Rachel Neville, Chris Peterson, Patrick Shipman, Sofya Chepushtanova, Eric Hanson, Francis Motta, and Lori Ziegelmeier. Persistence images: A stable vector representation of persistent homology. *Journal of Machine Learning Research*, 18(8):1–35, 2017.

[94] D. Pickup, X. Sun, P. L. Rosin, R. R. Martin, Z. Cheng, Z. Lian, M. Aono, A. Ben Hamza, A. Bronstein, M. Bronstein, S. Bu, U. Castellani, S. Cheng, V. Garro, A. Giachetti, A. Godil, J. Han, H. Johan, L. Lai, B. Li, C. Li, H. Li, R. Litman, X. Liu, Z. Liu, Y. Lu, A. Tatsuma, and J. Ye. SHREC'14 track: Shape retrieval of non-rigid 3d human models. In *Proceedings of the 7th Eurographics workshop on 3D Object Retrieval*, EG 3DOR'14. Eurographics Association, 2014.

[95] Danielle Barnes, Luis Polanco, and Jose A. Perea. A comparative study of machine learning methods for persistence diagrams. *Frontiers in Artificial Intelligence*, 4, 2021.

[96] T. Ojala, T. Maenpaa, M. Pietikainen, J. Viertola, J. Kyllonen, and S. Huovinen. Outex - new framework for empirical evaluation of texture analysis algorithms. In *2002 International Conference on Pattern Recognition*, volume 1, pages 701–706 vol.1, 2002.

[97] Pinar Yanardag and SVN Vishwanathan. Deep graph kernels. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1365–1374, 2015.

[98] Jonathan Blackman, Scott E Field, Chad R Galley, Béla Szilágyi, Mark A Scheel, Manuel Tiglio, and Daniel A Hemberger. Fast and accurate prediction of numerical relativity waveforms from binary black hole coalescences using surrogate models. *Physical review letters*, 115(12):121102, 2015.

[99] Christopher Bresten and Jae-Hun Jung. Detection of gravitational waves using topological data analysis and convolutional neural network: An improved approach. *arXiv preprint arXiv:1910.08245*, 2019.

[100] JP Huke. Embedding nonlinear dynamical systems: A guide to takens' theorem. 2006.

[101] Jochen Kämpf and Piers Chapman. *Upwelling systems of the world*. Springer, 2016.

[102] Rubén Varela, Fernando P Lima, Rui Seabra, Claudia Meneghesso, and Moncho Gómez-Gesteira. Coastal warming and wind-driven upwelling: a global analysis. *Science of the Total Environment*, 639:1501–1511, 2018.

[103] Marco Reggiannini, Joao Janeiro, Flávio Martins, Oscar Papini, and Gabriele Pieri. Mesoscale patterns identification through sst image processing. In *ROBOVIS*, pages 165–172, 2021.

[104] M Reggiannini, O Papini, and G Pieri. An automated analysis tool for the classification of sea surface temperature imagery. *Pattern Recognition and Image Analysis*, 32(3):631–635, 2022.

[105] Gabriele Pieri, João Janeiro, Flávio Martins, Oscar Papini, and Marco Reggiannini. Mec: A mesoscale events classifier for oceanographic imagery. *Applied Sciences*, 13(3):1565, 2023.

[106] OSI SAF. Full resolution l2p avhrr sea surface temperature metagranules (ghrsst)—metop. *EUMETSAT: Darmstadt, Germany*, 2011.

[107] NASA/JPL. Ghrsst level 2p global sea surface skin temperature from the moderate resolution imaging spectroradiometer (modis) on the nasa aqua satellite (gds2), 2020.

[108] Ping-Sung Liao, Tse-Sheng Chen, Pau-Choo Chung, et al. A fast algorithm for multilevel thresholding. *J. Inf. Sci. Eng.*, 17(5):713–727, 2001.

[109] Athanasios Voulodimos, Nikolaos Doulamis, Anastasios Doulamis, and Eftychios Protopapadakis. Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience*, 2018, 2018.

[110] Derek A Long. *The raman effect.* John Wiley & Sons Ltd, 2002.

[111] Mads Sylvest Bergholt, Wei Zheng, Kan Lin, Khek Yu Ho, Ming Teh, Khay Guan Yeoh, Jimmy Bok Yan So, and Zhiwei Huang. Raman endoscopy for in vivo differentiation between benign and malignant ulcers in the stomach. *Analyst*, 135(12):3162–3168, 2010.

[112] Mads Sylvest Bergholt. Raman endoscopy for objective diagnosis of early cancer in the gastrointestinal system. *Journal of Gastrointestinal & Digestive System*, 1(S1):008, 2013.

[113] Kenny Kong, Catherine Kendall, Nicholas Stone, and Ioan Notingher. Raman spectroscopy for medical diagnostics—from in-vitro biofluid assays to in-vivo cancer detection. *Advanced drug delivery reviews*, 89:121–134, 2015.

[114] Mustafa Çulha. Raman spectroscopy for cancer diagnosis: how far have we come? *Bioanalysis*, 7(21):2813–2824, 2015.

[115] Julietta V Rau, Valerio Graziani, Marco Fosca, Chiara Taffon, Massimiliano Rocchia, Pierfilippo Crucitti, Paolo Pozzilli, Andrea Onetti Muda, Marco Caricato, and Anna Crescenzi. Raman spectroscopy imaging improves the diagnosis of papillary thyroid carcinoma. *Scientific Reports*, 6(1):35117, 2016.

[116] C Dilara Savci-Heijink, Arjen HG Cleven, and Judith VMG Bovée. Benign and low-grade cartilaginous tumors: An update on differential diagnosis. *Diagnostic Histopathology*, 2022.

[117] David Suster, Yin Pun Hung, and G Petur Nielsen. Differential diagnosis of cartilaginous lesions of bone. *Archives of pathology & laboratory medicine*, 144(1):71–82, 2020.

[118] Mario D'Acunto, Raffaele Gaeta, Rodolfo Capanna, and Alessandro Franchi. Contribution of raman spectroscopy to diagnosis and grading of chondrogenic tumors. *Scientific Reports*, 10(1):2155, 2020.

[119] Pietro Manganelli Conforti, Mario D'Acunto, and Paolo Russo. Deep learning for chondrogenic tumor classification through wavelet transform of raman spectra. *Sensors*, 22(19):7492, 2022.

[120] Cristiano D'Andrea, Federico Angelo Cazzaniga, Edoardo Bistaffa, Andrea Barucci, Marella de Angelis, Martina Banchelli, Edoardo Farnesi, Panagis Polykretis, Chiara Marzi, Antonio Indaco, et al. Impact of seed amplification assay and surface-enhanced raman spectroscopy combined approach on the clinical diagnosis of alzheimer's disease. *Translational Neurodegeneration*, 12(1):35, 2023.

[121] Yanmei Xu, Xinyu Pan, Huan Li, Qiongfang Cao, Fan Xu, and Jianshu Zhang. Accuracy of raman spectroscopy in the diagnosis of alzheimer's disease. *Frontiers in Psychiatry*, 14:1112615, 2023.

[122] Kaj Blennow and Henrik Zetterberg. Biomarkers for alzheimer's disease: current status and prospects for the future. *Journal of internal medicine*, 284(6):643–663, 2018.

[123] Elena Ryzhikova, Nicole M Ralbovsky, Vitali Sikirzhytski, Oleksandr Kazakov, Lenka Halamkova, Joseph Quinn, Earl A Zimmerman, and Igor K Lednev. Raman spectroscopy and machine learning for biomedical applications: Alzheimer's disease diagnosis based on the analysis of cerebrospinal fluid. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 248:119188, 2021.

[124] Chia-Chi Huang and Ciro Isidoro. Raman spectrometric detection methods for early and non-invasive diagnosis of alzheimer's disease. *Journal of Alzheimer's Disease*, 57(4):1145–1156, 2017.

[125] Randy S Tashjian, Harry V Vinters, and William H Yong. Biobanking of cerebrospinal fluid. *Biobanking: Methods and Protocols*, pages 107–114, 2019.

[126] Hugo Vanderstichele, Mirko Bibl, Sebastiaan Engelborghs, Nathalie Le Bastard, Piotr Lewczuk, Jose Luis Molinuevo, Lucilla Parnetti, Armand Perret-Liaudet, Leslie M Shaw, Charlotte Teunissen, et al. Standardization of preanalytical aspects of cerebrospinal fluid biomarker testing for alzheimer's disease diagnosis: a consensus paper from the alzheimer's biomarkers standardization initiative. *Alzheimer's & Dementia*, 8(1):65–73, 2012.

[127] Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.

[128] Mathieu Carrière, Frédéric Chazal, Yuichi Ike, Théo Lacombe, Martin Royer, and Yuhei Umeda. Perslay: A neural network layer for persistence diagrams and new graph topological signatures. In *International Conference on Artificial Intelligence and Statistics*, pages 2786–2796. PMLR, 2020.

[129] Raphael Reinauer, Matteo Caorsi, and Nicolas Berkouk. Persformer: A transformer architecture for topological machine learning. *arXiv preprint arXiv:2112.15210*, 2021.

[130] Kwangho Kim, Jisu Kim, Manzil Zaheer, Joon Kim, Frédéric Chazal, and Larry Wasserman. Pllay: Efficient topological layer based on persistent landscapes. *Advances in Neural Information Processing Systems*, 33:15965–15977, 2020.

[131] Julius Von Rohrscheidt and Bastian Rieck. Topological singularity detection at multiple scales. In *International Conference on Machine Learning*, pages 35175–35197. PMLR, 2023.

[132] Archit Rathore, Nithin Chalapathi, Sourabh Palande, and Bei Wang. Topoact: Visually exploring the shape of activations in deep learning. In *Computer Graphics Forum*, volume 40, pages 382–397. Wiley Online Library, 2021.

[133] Rickard Brüel Gabrielsson, Bradley J Nelson, Anjan Dwaraknath, and Primoz Skraba. A topology layer for machine learning. In *International Conference on Artificial Intelligence and Statistics*, pages 1553–1563. PMLR, 2020.

[134] Théo Lacombe, Yuichi Ike, Mathieu Carriere, Frédéric Chazal, Marc Glisse, and Yuhei Umeda. Topological uncertainty: Monitoring trained neural networks through persistence of activation graphs. *arXiv preprint arXiv:2105.04404*, 2021.

[135] Marc Ethier, Patrizio Frosini, Nicola Quercioli, and Francesca Tombari. Geometry of the matching distance for 2d filtering functions. *Journal of Applied and Computational Topology*, 7(4):815–830, 2023.

[136] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete Comput. Geom.*, 37(1):103–120, 2007.

[137] Rocío Gonzalez-Diaz, Manuel Soriano-Trigueros, and A Torras-Casas. Partial matchings induced by morphisms between persistence modules. *Computational Geometry*, 112:101985, 2023.

[138] Ulrich Bauer and Michael Lesnick. Induced matchings of barcodes and the algebraic stability of persistence. In *Proceedings of the thirtieth annual symposium on Computational geometry*, pages 355–364, 2014.

[139] Daniel Kraft. Computing the Hausdorff distance of two sets from their distance functions. *International Journal of Computational Geometry & Applications*, 30(01):19–49, 2020.

[140] Patrizio Frosini. G-invariant persistent homology. *Mathematical Methods in the Applied Sciences*, 38(6):1190–1199, 2015.

[141] Wojciech Chachólski, Alessandro De Gregorio, Nicola Quercioli, and Francesca Tombari. Symmetries of data sets and functoriality of persistent homology. 2023.

[142] Daniel Lundqvist, Anders Flykt, and Arne Öhman. Karolinska directed emotional faces. *PsycTESTS Dataset*, 91:630, 1998.

[143] Francesca Cagliari, Barbara Di Fabio, and Claudia Landi. The natural pseudodistance as a quotient pseudo-metric, and applications. *Forum Mathematicum*, 27(3):1729–1742, 2015.

[144] Kenneth R Davidson. *Real analysis and applications.* Springer, 2010.

[145] Ben Blum-Smith and Samuel Coskey. The fundamental theorem on symmetric polynomials: history's first whiff of galois theory. *The College Mathematics Journal*, 48(1):18–29, 2017.

[146] Darian McLaren Lei Cao and Sarah Plosker. Centrosymmetric stochastic matrices. *Linear and Multilinear Algebra*, 70(3):449–464, 2022.

[147] M. Aschbacher. *Finite Group Theory*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2 edition, 2000.

[148] Tom Fischer. Existence, uniqueness, and minimality of the jordan measure decomposition. *arXiv preprint arXiv:1206.5449*, 2012.

[149] Yann LeCun. The mnist database of handwritten digits. *http://yann. lecun. com/exdb/mnist/*, 1998.

[150] Keiron O'Shea and Ryan Nash. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015.

[151] Dor Bank, Noam Koenigstein, and Raja Giryes. Autoencoders. *Machine learning for data science handbook: data mining and knowledge discovery handbook*, pages 353–374, 2023.

[152] Carl Doersch. Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908*, 2016.

[153] Zongsheng Yue, Hongwei Yong, Qian Zhao, Deyu Meng, and Lei Zhang. Variational denoising network: Toward blind noise modeling and removal. *Advances in neural information processing systems*, 32, 2019.

[154] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.