1. Research Project — Progetto di ricerca

The main goals of the project are the formalization and the development of a novel approach to the analysis, comparison and classification of multidimensional data, i.e. information characterized by features embedded in some multidimensional topological space. Our proposed project will be based on a topological exploration and understanding of data, finding its roots in the so-called (multidimensional) persistence, a theoretical and computational methodology recently emerged as a suitable tool for such purposes. The framework developed within the project does not aim to replace existing data analysis tools. It will provide a new tool enabling the use of additional information, hence improving the existing frameworks, making them more robust and stable. One further contribution will be the finding of important data properties with regard to the multidimensional persistence framework.

1.1. Motivation of the Research Project Proposal. The study of the topology of data is gaining more and more appeal in the scientific community. This is due to the large number of scientific contexts where it is required to deal with qualitative geometric information. Nowadays, a relevant feature in fields such as computer science, physics and engineering is the big amount of data that is produced, carrying with it a lot of information of possibly different nature. Additionally, it usually happens that such data is unavoidably affected by noise. In this context, the topological approach enables a deep reduction in the complexity of the data by confining the analysis just on their relevant part, as well as obtaining invariance under affine transformations (e.g., rotation, translation or scaling) and continuous geometrical deformations (similar to rubber sheet deformations). This research area is discussed in detail in [4]. Cognitive studies have also shown that part of the information used by the human visual system is based on topological analysis [10]; hence improving existing topological tools is definitely a desirable target.

Topological Persistence. Topological persistence, hereafter simply persistence, has revealed to be an interesting theoretical approach in studying the topology of data, with applications in a growing numbers of field, ranging from shape analysis and comparison to data simplification and hole detection in sensor networks [4, 9]. Persistence performs a topological exploration along the filtration of a given space \( X \), i.e., a nested sequence of subsets \( X_1 \subseteq X_2 \subseteq \cdots \subseteq X_n = X \). Such a sequence is usually induced by considering the sublevel sets of a function \( f : X \to \mathbb{R} \) called filtering function. The main assumption here is that the most important part of information in geometrical data is the one that “persists” with respect to the defining parameters. Focusing on the occurrence of meaningful topological events along the filtration of the sublevel sets of \( f \) (such as the birth or the death of connected components, tunnels or cavities) it is possible to provide a global description of data. Such information can be encoded in a parameterized version of the Betti numbers, known in the literature as persistent Betti numbers [9], a rank invariant [6], and, for the 0th homology (e.g., the case of connected components), a size function [4]. The key point is that these descriptors can be represented in a very simple and compact way, by means of the so-called persistence diagrams. Moreover, they are stable with respect to a suitable distance, i.e., the bottleneck distance or matching distance, thus implying resistance to noise.

Multidimensional Persistence. Multidimensional persistence studies the case when filtering functions are vector-valued. Such a generalization is highly desirable, since it would enable the study of more complex data structures, i.e., characterized by multidimensional information. This is actually the usual scenario in applications.

In data analysis, we often have to deal with a finite set of samples from some underlying topological space. Each sample is usually associated with multiple labels, representing several measurements. A typical example is the study of the simulation of a hurricane, in which independent features like temperature, pressure and 2D geographical position are studied to better understand the development of such phenomena. Another example is the analysis of 4D time-varying CT scans of human hearts. Studying features like functional value, gradient magnitude and curvature is crucial to understand anomalies related to the cycles of the heart process.
In the context of shape comparison, a shape (e.g., a 2D- or 3D-image, a triangular mesh surface or a sound) can be modeled as a topological space $X$ endowed with a filtering function $f$. The function $f$ plays the role of a descriptor for the features considered relevant within the shape comparison/retrieval problem at hand. In this perspective, the role of vector-valued filtering functions is clear: They would allow to face with applications in which shapes are characterized by properties that are intrinsically multidimensional, such as the coordinate of a point in the 3-dimensional space or the representation of color in the RGB model.

In all these settings, noise is a very relevant factor that seriously affects the analysis process. Persistence, being robust to noise, is a viable option for assessing and interpreting such information from the topological perspective.

### 2. Activity Plan — Piano di Attività

The main goal of the project is the development of a prototype framework for the analysis and classification of data based on topology descriptors. We will focus on specific situations related to the analysis of time-varying meteorological data, medical data, population statistics and the comparison of 2D- and 3D-images. This implies the introduction of suitable filtering functions, as well as the merging of possibly unrelated information from such analysis outcomes, with statistical methods. For each considered setting, public benchmark databases will be used to test the capability of the proposed framework.

This main activity within the project will be the improvement of the multidimensional framework for persistence. We will build on previous works [2, 5] to obtain a solid theoretical framework. In particular, starting from [1, 3] we will extend algorithms specifically designed for filtering function taking values in $\mathbb{R}^2$ – and applied only in the 0-th-homology situation – to obtain a robust computational platform for arbitrary multidimensional filtering functions and homology degrees.

We observe that the proposed framework is modular. Simply by changing the considered filtering functions, and without any other modification in the underlying framework, it is possible to analyze data under different perspectives. In this context, part of the project will be devoted to build/select meaningful filtering functions for applications. For example, we know by experience [3, 7, 8] that particular families of functions are better suited than others to guarantee high discriminatory power in specific shape analysis and comparison application. However, an automatic procedure for selecting suitable measuring functions according to the problem at hand has not been provided yet.

Moreover, we also want to develop some basis methodology for defining an “orthogonality” criterion for filtering functions. Indeed, this would enable us to select multidimensional filtering functions which minimize, in some sense, the redundancy in the measurements of data.

Consolidated statistical frameworks, e.g. *a contrario* methods, will be used to interpret and merge the results coming from the analysis of different, independent, possibly multidimensional filtering functions.

### References


