An Introduction to the Notion of Natural Pseudo-distance in Topological Data Analysis

Published Version:
An Introduction to the Notion of Natural Pseudo-distance in Topological Data Analysis / Frosini P.. - STAMPA. - 350:(2021), pp. 203-213. (Intervento presentato al convegno 1st International Workshop and Conference on Topological Dynamics and Topological Data Analysis, IWCTA 2018 tenutosi a Kochi, India nel December 9–11, 2018) [10.1007/978-981-16-0174-3_17].

Availability:
This version is available at: https://hdl.handle.net/11585/865276 since: 2022-02-23

Published:
DOI: http://doi.org/10.1007/978-981-16-0174-3_17

(Article begins on next page)

AN INTRODUCTION TO THE NOTION OF NATURAL PSEUDO-DISTANCE IN TOPOLOGICAL DATA ANALYSIS

PATRIZIO FROSINI

Abstract. The natural pseudo-distance d_G associated with a group G of self-homeomorphisms of a topological space X is a pseudo-metric developed to compare real valued-functions defined on X, when the equivalence between functions is expressed by the group G. In this paper we illustrate d_G, its role in topological data analysis, its main properties and its link with persistent homology.

Introduction

In topological data analysis data are frequently expressed by continuous real-valued (or vector-valued) functions defined on a topological space X, and two such functions are considered equivalent if they can be obtained from each other by composition with a suitable self-homeomorphism of X. This happens, e.g., when we are interested in comparing images with respect to the group of plane isometries, or ECG traces with respect to the group of translations in time, or temperature distributions on the earth with respect to rotations around the north pole-south pole axis. Such functions are called *filtering functions*. In order two compare this kind of data a pseudo-distance is available, quantifying the infimum of the cost of matching two functions φ_1, φ_2 by composition with a homeomorphism in the considered group G, where the cost is defined by the L^∞ norm. According to this pseudo-metric the measurements $\varphi, \varphi \circ g \in C^0(X, \mathbb{R})$ are considered equivalent to each other for every $g \in G$. In many application this property is important and useful, since it allows to choose the data equivalence the user is interested in. For the sake of simplicity, in this survey we will only consider the case of data represented by real-valued functions. This paper is devoted to illustrate this pseudo-metric, called the *natural pseudo-distance d_G associated with the group G*. After recalling the definition of d_G (Section 1), we present some theoretical results concerning the values that $d_G(\varphi_1, \varphi_2)$ can take, showing that they are strictly related with the critical values of φ_1 and φ_2, provided that these functions are regular enough (Section 2). Secondly, we observe that while d_G represents a clear ground truth in our setting, it is usually quite difficult to compute, due to the size of the group G to be examined. Therefore, efficient methods to get information about d_G are needed. The most relevant method to study the natural pseudo-distance is based on its link with persistent homology and the theory of group equivariant non-expansive operator. Section 3 is devoted to describe this link and its main consequences. In Section 4 we conclude the paper by illustrating an open problem concerning d_G.

2010 *Mathematics Subject Classification.* Primary 55N35; Secondary 57R19 65D18 68U05.

Related literature and historical notes. This survey presents the main results obtained about the natural pseudo-distance in the last three decades. These results appeared in several papers and are reported here without proof. For every statement, the paper where the interested reader can find a precise proof is referred. The concept of natural pseudo-distance appeared for the first time in the paper [14], where the distance $\|A - B\|$ between pairs (A, B) of points in a submanifold $\mathcal{M}$ of a Euclidean space was considered as a filtering function and the group G was chosen to be the group of isometries of $\mathcal{M}$. A different but strictly related distance between real-valued functions defined on a manifold had already been presented in [13], referring to the group of similarities of $\mathbb{E}^n$.

The description given in this survey is mainly based on the paper [18]. The reader can find there definitions and proofs concerning the natural pseudo-distance d_G associated with a group G, together with its link with persistent homology and the theory of group equivariant non-expansive operators. The problem of obtaining lower bounds for $d_{\mathrm{Homeo}(X)}$ by means of persistent homology in degree 0 (size functions) has been investigated in [19, 8, 6]. Lower bounds for d_G obtained by means of persistent homotopy in the case $G = \mathrm{Homeo}(X)$ and via G-invariant persistent homology in the general case have been presented in [20] and [16], respectively. A study of d_G as a quotient pseudo-metric has been done in the paper [2]. The proofs of the results concerning the link between the values that d_G can take and the critical values of the filtering functions can be found in [9, 10, 11]. The proof of the result concerning the possible values of the natural pseudo-distance in the case $X = G = S^1$ can be found in [7]. The results concerning optimal homeomorphisms are illustrated in the papers [19, 9, 4, 7]. A survey about the natural pseudo-distance in the case $G = \mathrm{Homeo}(X)$ has appeared in [15].

1. The definition of d_G

Let (X, d) and G be a finitely triangulable metric space and a subgroup of the group $\mathrm{Homeo}(X)$ of all homeomorphisms from X to X, respectively. If φ_1, φ_2 are two continuous and bounded functions from X to $\mathbb{R}$ we can consider the value $\inf_{g \in G} \|\varphi_1 - \varphi_2 \circ g\|_\infty$. This value is called *the natural pseudo-distance* $d_G(\varphi_1, \varphi_2)$ between φ_1 and φ_2 with respect to the group G. We recall that a pseudo-metric is just a metric without the property assuring that if two points have a null distance then they must coincide. We endow $C^0(X, \mathbb{R})$ with the L^∞ norm and G with the distance $D_G(g_1, g_2) := \max_{x \in X} d(g_1(x), g_2(x))$, so that G becomes a topological group acting continuously on $C^0(X, \mathbb{R})$ by composition on the right. We observe that the action of G on $C^0(X, \mathbb{R})$ is continuous [18].

If G is the trivial group Id, then d_G is the max-norm distance $\|\varphi_1 - \varphi_2\|_\infty$. Moreover, if G_1 and G_2 are subgroups of $\mathrm{Homeo}(X)$ and $G_1 \subseteq G_2$, then

$$d_{\mathrm{Homeo}(X)}(\varphi_1, \varphi_2) \le d_{G_2}(\varphi_1, \varphi_2) \le d_{G_1}(\varphi_1, \varphi_2) \le \|\varphi_1 - \varphi_2\|_\infty$$

for every $\varphi_1, \varphi_2 \in C^0(X, \mathbb{R})$.

The direct computation of d_G is usually difficult, due to the size of G. As an example, if $X = \mathbb{R}^3$ and G is the group of all isometries of $\mathbb{R}^3$, a direct computation of d_G would require to evaluate $\|\varphi_1 - \varphi_2 \circ g\|_\infty$ for every isometry $g : \mathbb{R}^3 \to \mathbb{R}^3$. The reader could think of approximating $d_G(\varphi_1, \varphi_2)$ by the value $\mu_S(\varphi_1, \varphi_2) := \inf_{g \in S} \|\varphi_1 - \varphi_2 \circ g\|_\infty$, where S is a sufficiently dense subset S of G. Unfortunately, the use of μ_S would be impractical for data retrieval for two reasons. First of all, in many cases S should be a very large set in order to obtain a good approximation of d_G, so implying a large computational cost. Secondly, S could not be assumed to be a subgroup of G, even if G is compact (cf. Section 3.1 in [18]). For example, this happens when G is the group $SO(3)$ of all orientation-preserving isometries of $\mathbb{R}^3$ that take the point $(0,0,0)$ to itself. As a consequence, the function $\mu_S(\varphi_1,\varphi_2)$ would not be a pseudo-metric. This would make the use of μ_S unsuitable for several applications. In Section 3 we will see that this difficulty can be worked around by means of persistent homology and the concept of group equivariant non-expansive operator (Theorem 3.4).

We conclude this section by observing that in many cases we are not interested in every function in $C^0(X,\mathbb{R})$, but in a bounded topological subspace Φ of $C^0(X,\mathbb{R})$. This is due to the fact that the choice of each measuring device restricts the set of functions that can be obtained as data produced by the measurement. From now on, we will assume that a bounded topological subspace Φ of $C^0(X,\mathbb{R})$ has been chosen.

1.1. The role of d_G in Topological Data Analysis.

The comparison of data is usually a process depending on an observer. We could indeed say that data comparison consists in the study of the relationship between an observer and the reality he/she can measure. In this framework, data coincide with measurements. Observers receive and transform data and are, in some sense, defined by the way they perform this transformation. It follows that observers can be defined as collections of suitable operators acting on measurements [17].

According to the dictionary, a "measurement is the assignment of a number to a characteristic of an object or event, which can be compared with other objects or events" [23]. This definition implies that measurements (and hence data) can be seen as functions φ associating a real number $\varphi(x)$ with each point x of a set X of characteristics. (This definition admits a natural extension to vector-valued functions but, for the sake of simplicity, we will treat here only the case of scalar-valued functions). If we wish to develop a theory that can be applied in real situations, we need stability with respect to noise. This justifies the use of topologies on X and on the set Φ of possible measurements on X, as illustrated in the previous section. Furthermore, observers are often endowed with some kind of equivariance, represented by a suitable group G of homeomorphisms. Therefore we are interested in models where this equivariance can be represented. For example, we usually look for pseudo-metrics that do not distinguish between the shapes of the same object in different spatial positions. The natural pseudo-distance d_G has this property, since it vanishes when the measurements φ, $\varphi \circ g$ are considered, with $\varphi \in \Phi$ and $g \in G$. For this reason, the pseudo-metric d_G can be considered as a ground truth for data comparison in our theoretical setting. This justifies our interest in its study.

2. Theoretical results about d_G

When the filtering functions are defined on a regular closed manifold, some results restrict the range of values that can be taken by the natural pseudo-distance d_G.

Theorem 2.1 ([9])**.** *Assume that $\mathcal{M}$ is a closed manifold of class C^1 and that $\varphi_1, \varphi_2 : \mathcal{M} \to \mathbb{R}$ are two functions of class C^1. Set $d := d_{\mathrm{Homeo}(\mathcal{M})}(\varphi_1, \varphi_2)$. Then a positive integer k exists for which one of the following properties holds:*

(i) k is odd and kd is the distance between a critical value of φ_1 and a critical value of φ_2;
(ii) k is even and kd is either the distance between two critical values of φ_1 or the distance between two critical values of φ_2.

Theorem 2.2 ([10])**.** *Assume that $\mathcal{S}$ is a closed surface of class C^1 and that $\varphi_1, \varphi_2 : \mathcal{S} \to \mathbb{R}$ are two functions of class C^1. Set $d := d_{\mathrm{Homeo}(\mathcal{S})}(\varphi_1, \varphi_2)$. Then a positive integer k exists for which at least one of the following properties holds:*

(i) d is the distance between a critical value of φ_1 and a critical value of φ_2;
(ii) d is half the distance between two critical values of φ_1.
(iii) d is half the distance between two critical values of φ_2.
(iv) d is one third of the distance between a critical value of φ_1 and a critical value of φ_2.

Theorem 2.3 ([11])**.** *Assume that $\mathcal{C}$ is a closed curve of class C^1 and that $\varphi_1, \varphi_2 : \mathcal{C} \to \mathbb{R}$ are two functions of class C^1. Set $d := d_{\mathrm{Homeo}(\mathcal{C})}(\varphi_1, \varphi_2)$. Then a positive integer k exists for which at least one of the following properties holds:*

a) d is the distance between a critical value of φ_1 and a critical value of φ_2;
b) d is half the distance between two critical values of φ_1.
c) d is half the distance between two critical values of φ_2.

The statement in the last theorem is sharp, as shown by the following examples.

Example 2.4. Let us consider the two embeddings of S^1 into $\mathbb{R}^2$ represented in Figure 1. The ordinate y defines two filtering functions φ_1, φ_2 on S^1. In this case $d_{\mathrm{Homeo}(S^1)}(\varphi_1, \varphi_2) = |\varphi_1(A) - \varphi(B)|$, i.e. it is the distance between a critical value of φ_1 and a critical value of φ_2.

Example 2.5. Let us consider the two embeddings of S^1 into $\mathbb{R}^2$ represented in Figure 2. The ordinate y defines two filtering functions φ_1, φ_2 on S^1. In this case $d_{\mathrm{Homeo}(S^1)}(\varphi_1, \varphi_2) = \frac{1}{2}|\varphi_1(A) - \varphi_1(B)|$, i.e. it is half the distance between two critical values of φ_1. In Figure 2 a homeomorphism $g_\varepsilon : S^1 \to S^1$ is displayed, such that $\|\varphi_1 - \varphi_2 \circ g_\varepsilon\|_\infty \le \varepsilon$ (we set $g_\varepsilon(D_\varepsilon) = H_\varepsilon$, $g_\varepsilon(C) = G$ and $g_\varepsilon(E_\varepsilon) = F_\varepsilon$; the first red arc is taken to the second red arc).

The research concerning the case that G is a proper subgroup of Homeo($\mathcal{M}$) is still at its very beginning. As an example of the results concerning this line of research we cite the following theorem.

Theorem 2.6 ([7])**.** *Let φ_1, φ_2 be Morse functions from the Lie group S^1 to $\mathbb{R}$ and set $d = d_{S^1}(\varphi_1, \varphi_2)$. At least one of the following statements holds:*

(1) There exist a critical point θ_1 for φ_1 and a critical point θ_2 for φ_2 such that $d = |\varphi_1(\theta_1) - \varphi_2(\theta_2)|$;
(2) There exist $\theta_1, \theta_2, \tilde\theta_1, \tilde\theta_2 \in S^1$ such that
- $d = |\varphi_1(\theta_1) - \varphi_2(\theta_2)| = |\varphi_1(\tilde\theta_1) - \varphi_2(\tilde\theta_2)|$;
- $\frac{d\varphi_1}{d\theta}(\theta_1) = \frac{d\varphi_2}{d\theta}(\theta_2)$ *and* $\frac{d\varphi_1}{d\theta}(\tilde\theta_1) = \frac{d\varphi_2}{d\theta}(\tilde\theta_2)$;
- $\theta_1 - \theta_2 = \tilde\theta_1 - \tilde\theta_2$;
- $\frac{d\varphi_1}{d\theta}(\theta_1) \cdot \frac{d\varphi_1}{d\theta}(\tilde\theta_1) \cdot (\varphi_1(\theta_1) - \varphi_2(\theta_2)) \cdot (\varphi_1(\tilde\theta_1) - \varphi_2(\tilde\theta_2)) < 0.$

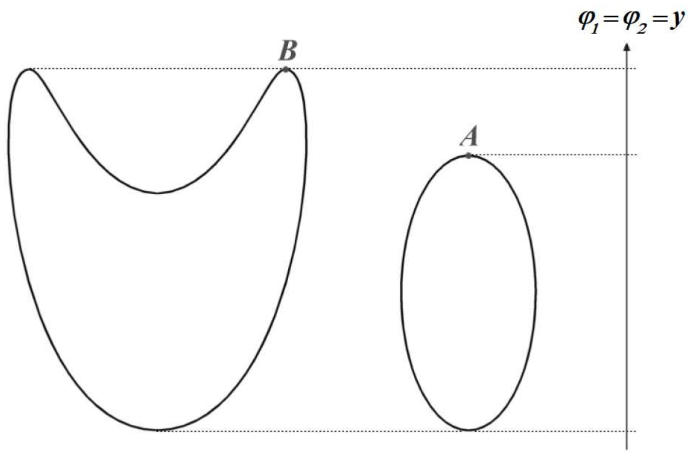

FIGURE 1. In this case the natural pseudo-distance is equal to the distance between two critical values of the filtering functions.

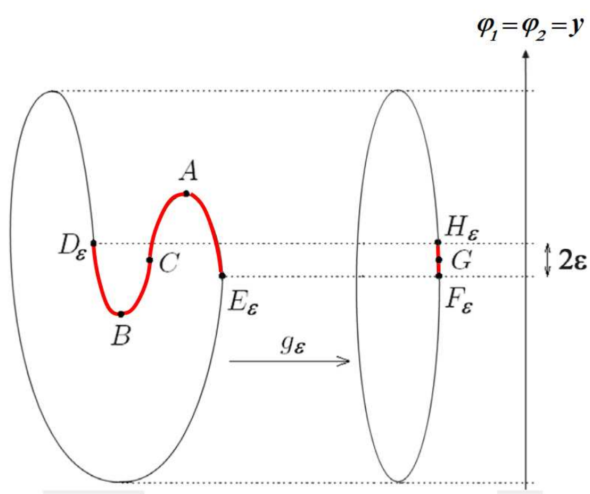

FIGURE 2. In this case the natural pseudo-distance is equal to half the distance between two critical values of the filtering function φ_1.

2.1. **Optimal homeomorphisms.** Assume that $\varphi_1, \varphi_2 : X \to \mathbb{R}$ are continuous functions. Let G be a subgroup of $\mathrm{Homeo}(X)$. We say that a homeomorphism $g \in G$ is *optimal* in G for (φ_1, φ_2) if $\|\varphi_1 - \varphi_2 \circ g\|_\infty = d_G(\varphi_1, \varphi_2)$. The following results hold for optimal homeomorphisms.

Theorem 2.7 ([9]). *Assume that $\mathcal{M}$ is a C^1 closed manifold and that $\varphi_1, \varphi_2 : \mathcal{M} \to \mathbb{R}$ are of class C^1. If an optimal homeomorphism $g \in \mathrm{Homeo}(\mathcal{M})$ for (φ_1, φ_2) exists, then $d_{\mathrm{Homeo}(\mathcal{M})}(\varphi_1, \varphi_2)$ is the distance between a critical value of φ_1 and a critical value of φ_2.*

Theorem 2.8 ([4]). *If $\varphi_1, \varphi_2 : S^1 \to \mathbb{R}$ are Morse functions and $d_{\mathrm{Homeo}(S^1)}(\varphi_1, \varphi_2)$ vanishes, then an optimal C^2-diffeomorphism exists in $\mathrm{Homeo}(S^1)$ for (φ_1, φ_2).*

Theorem 2.9 ([7]). *The number of optimal homeomorphisms in the Lie group S^1 for a pair (φ_1, φ_2) of Morse functions from S^1 to $\mathbb{R}$ is finite.*

3. A link between d_G and persistent homology

In this section we will show that the natural pseudo-distance d_G can be studied by combining persistent homology with the concept of group equivariant non-expansive operator.

Persistent homology. Persistent homology can be seen as an efficient method to compute lower bounds and good approximations for the natural pseudo-distance. We recall here some basic definitions and facts concerning persistent homology. The interested reader can find a more detailed and formal treatment in [12, 1, 3, 22]. In plain words, persistent homology is a mathematical theory describing the changes of the homology groups of the sub-level sets $X_t = \varphi^{-1}((-\infty, t])$ varying t in $\mathbb{R}$, where φ is a real-valued continuous function defined on a topological space X. We can look at the parameter t as an increasing time, whose change produces the birth and death of k-dimensional holes in the sub-level set X_t. For $k = 0, 1, 2$, the expression "k-dimensional holes" refers to gaps between connected components, tunnels and voids, respectively. The distance between the birthdate and deathdate of a hole is called its *persistence*. The more persistent is a hole, the more important it is for data comparison, since holes with small persistence are usually produced by noise.

As happens for homology, persistent homology can be introduced in several different settings. In this paper we will use the definition based on Čech homology (cf. [5]).

We start from the following definition.

Definition 3.1. Let $\varphi : X \to \mathbb{R}$ be a continuous function. If $u, v \in \mathbb{R}$ and $u < v$, we can consider the inclusion i of X_u into X_v. Such an inclusion induces a homomorphism $i^* : H_k(X_u) \to H_k(X_v)$ between the homology groups of X_u and X_v in degree k. The group $PH_k^{\varphi}(u, v) := i^*(H_k(X_u))$ is called the *k-th persistent homology group with respect to the function $\varphi : X \to \mathbb{R}$, computed at the point* (u, v). The rank $r_k(\varphi)(u, v)$ of this group is said *the k-th persistent Betti numbers function with respect to the function $\varphi : X \to \mathbb{R}$, computed at the point (u, v).*

It can be easily proved that if $g \in \mathrm{Homeo}(X)$, the groups $PH_k^{\varphi}(u, v)$, $PH_k^{\varphi \circ g}(u, v)$ are isomorphic to each other for every $(u, v) \in \mathbb{R}$ with $u < v$ and every $k \in \mathbb{Z}$.

A classical way to describe persistent Betti numbers functions is given by *persistence diagrams*. The k-th persistence diagram $\mathrm{Dgm}_k(\varphi)$ of the function φ is the set of all pairs (b_j, d_j), where b_j and d_j are the birthdate and the deathdate of the j-th k-dimensional hole, respectively, with reference to the filtration $X_t = \varphi^{-1}((-\infty, t])$ varying t in $\mathbb{R}$. When a hole never dies, we set its deathdate

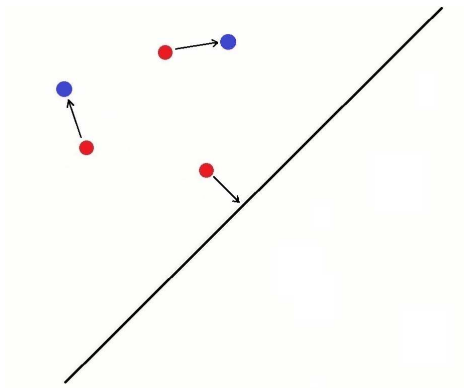

FIGURE 3. An example of matching between two persistence diagrams.

equal to ∞. For technical reasons, the points (t,t) are added to each persistence diagram. Two persistence diagrams $\mathrm{Dgm}_k(\varphi_1), \mathrm{Dgm}_k(\varphi_2)$ can be compared by means of the *bottleneck distance* $d_{BN}\left(\mathrm{Dgm}_k(\varphi_1), \mathrm{Dgm}_k(\varphi_2)\right)$. It is defined as the maximum movement of the points of $\mathrm{Dgm}_k(\varphi_1)$ that is necessary to change $\mathrm{Dgm}_k(\varphi_1)$ into $\mathrm{Dgm}_k(\varphi_2)$, measured with respect to the maximum norm (see Figure 3). If Čech homology is used, each persistent Betti numbers function $r_k(\varphi)$ is equivalent to the corresponding persistence diagram $\mathrm{Dgm}_k(\varphi)$. Therefore the bottleneck distance induces a metric d_{match} on the set of the persistent Betti numbers functions, so that $d_{\mathrm{match}}\left(r_k(\varphi_1), r_k(\varphi_2)\right) = d_{BN}\left(\mathrm{Dgm}_k(\varphi_1), \mathrm{Dgm}_k(\varphi_2)\right)$. The interested reader can find the formal definitions of persistence diagram and bottleneck distance in [12].

An important property of the metric d_{match} is its stability, as stated in the following result.

Theorem 3.2. *If k is a natural number and $\varphi_1, \varphi_2 \in C^0(X, \mathbb{R})$, then*

$$d_{\mathrm{match}}(r_k(\varphi_1), r_k(\varphi_2)) \leq d_{\mathrm{Homeo}(X)}(\varphi_1, \varphi_2) \leq \|\varphi_1 - \varphi_2\|_\infty .$$

Group equivariant non-expansive operators. Let us consider the set $\boldsymbol{\mathcal{F}}(\Phi, G)$ of all maps F from Φ to Φ that verify the following two properties:

(1) $F(\varphi \circ g) = F(\varphi) \circ g$ for every $\varphi \in \Phi$ and every $g \in G$ (i.e. F is equivariant with respect to G);
(2) $\|F(\varphi_1) - F(\varphi_2)\|_\infty \leq \|\varphi_1 - \varphi_2\|_\infty$ for every $\varphi_1, \varphi_2 \in \Phi$ (i.e. F is non-expansive).

Obviously, $\boldsymbol{\mathcal{F}}(\Phi, G)$ is not empty, since it contains at least the identity map.

The maps in $\boldsymbol{\mathcal{F}}(\Phi, G)$ are called Group Equivariant Non-Expansive Operators (GENEOs). In $\boldsymbol{\mathcal{F}}(\Phi, G)$ we define the metric $D_{GENEO}(F_1, F_2) := \sup_{\varphi \in \Phi} \|F_1(\varphi) - F_2(\varphi)\|_\infty$.

Persistent homology as a tool to get lower bounds for d_G. If $\mathcal{F}$ is a nonempty subset of $\boldsymbol{\mathcal{F}}(\Phi, G)$, then for every fixed k we can define the following pseudo-metric $D^{\mathcal{F},k}_{\mathrm{match}}$ on Φ:

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

for every $\varphi_1, \varphi_2 \in \Phi$, where $r_k(\varphi)$ denotes the k-th persistent Betti numbers function with respect to the function $\varphi : X \to \mathbb{R}$. We will usually omit the index k, when its value is clear from the context or not influential.

We observe that $D^{\mathcal{F}}_{\text{match}}(\varphi_1, \varphi_2 \circ g) = D^{\mathcal{F}}_{\text{match}}(\varphi_1 \circ g, \varphi_2) = D^{\mathcal{F}}_{\text{match}}(\varphi_1, \varphi_2)$ for every $\varphi_1, \varphi_2 \in \Phi$ and every $g \in \text{Homeo}(X)$.

The importance of $D^{\mathcal{F}}_{\text{match}}$ lies in the following two results, showing that it can be used to get information about the natural pseudo-distance d_G.

Theorem 3.3 ([18])**.** *If $\emptyset \neq \mathcal{F} \subseteq \boldsymbol{\mathcal{F}}(\Phi, G)$, then $D^{\mathcal{F}}_{\text{match}} \leq d_G$.*

Theorem 3.4 ([18])**.** *Let us assume that every function in Φ is non-negative, the k-th Betti number of X does not vanish, and Φ contains each constant function c for which a function $\varphi \in \Phi$ exists such that $0 \leq c \leq \|\varphi\|_\infty$. Then $D^{\boldsymbol{\mathcal{F}}(\Phi,G)}_{\text{match}} = d_G$.*

As a consequence, the topological and geometrical study of $\boldsymbol{\mathcal{F}}(\Phi, G)$ is important in the research concerning the natural pseudo-distance. Theorem 3.4 allows us to approximate d_G by approximating $D^{\boldsymbol{\mathcal{F}}(\Phi,G)}_{\text{match}}$.

Two relevant properties of $\boldsymbol{\mathcal{F}}(\Phi, G)$ are expressed by the following results.

Theorem 3.5 ([18])**.** *If Φ is compact, then $\boldsymbol{\mathcal{F}}(\Phi, G)$ is compact.*

Theorem 3.6 ([21])**.** *If Φ is convex, then $\boldsymbol{\mathcal{F}}(\Phi, G)$ is convex.*

4. An open problem

Let us consider a closed C^1 surface $\mathcal{S}$ and two C^1 filtering functions $\varphi_1, \varphi_2 : \mathcal{S} \to \mathbb{R}$. Let $\text{Homeo}(\mathcal{S})$ be the group of all self-homeomorphisms of $\mathcal{S}$. We know that $d_{\text{Homeo}(\mathcal{S})}(\varphi_1, \varphi_2) := \inf_{g \in \text{Homeo}(\mathcal{S})} \|\varphi_1 - \varphi_2 \circ g\|_\infty$ is the natural pseudo-distance between φ_1 and φ_2, with respect to the group $\text{Homeo}(\mathcal{S})$. As we have previously seen, it has been proved in [10] that at least one of the following statements holds:

(1) $d_{\text{Homeo}(\mathcal{S})}(\varphi_1, \varphi_2)$ is the distance between a critical value of φ_1 and a critical value of φ_2;
(2) $d_{\text{Homeo}(\mathcal{S})}(\varphi_1, \varphi_2)$ is half the distance between two critical values of φ_1;
(3) $d_{\text{Homeo}(\mathcal{S})}(\varphi_1, \varphi_2)$ is half the distance between two critical values of φ_2;
(4) $d_{\text{Homeo}(\mathcal{S})}(\varphi_1, \varphi_2)$ is one third of the distance between a critical value of φ_1 and a critical value of φ_2.

Interestingly, no example of two functions $\varphi_1, \varphi_2 : \mathcal{S} \to \mathbb{R}$ is known, such that (4) holds but (1), (2), (3) do not hold. A natural question arises: Can we find an example of two such functions or prove that such an example cannot exist (so improving Theorem 5.7 in [10])?

We recall that the usual technique to compute the natural pseudo-distance $d_{\text{Homeo}(\mathcal{S})}$ consists in

- finding a lower bound for $d_{\text{Homeo}(\mathcal{S})}(\varphi_1, \varphi_2)$ by computing the bottleneck distance $d_{BN}(\text{Dgm}_k(\varphi_1), \text{Dgm}_k(\varphi_2))$ between the persistence diagrams in degree k of the functions φ_1 and φ_2 (cf. Theorem 3.2);
- looking for a sequence (g_i) in $\text{Homeo}(\mathcal{S})$, such that $\lim_{i\to\infty} \|\varphi_1 - \varphi_2 \circ g_i\|_\infty = d_{BN}(\text{Dgm}_k(\varphi_1), \text{Dgm}_k(\varphi_2))$.

If such a sequence (g_i) exists, then the definition of natural pseudo-distance implies that $d_{\text{Homeo}(\mathcal{S})}(\varphi_1, \varphi_2)$ is equal to $d_{BN}(\text{Dgm}_k(\varphi_1), \text{Dgm}_k(\varphi_2))$.

Unfortunately, at least one of the following statements holds (cf. [8]):

a) $d_{BN}(\mathrm{Dgm}_k(\varphi_1), \mathrm{Dgm}_k(\varphi_2))$ is the distance between a critical value of φ_1 and a critical value of φ_2;

b) $d_{BN}(\mathrm{Dgm}_k(\varphi_1), \mathrm{Dgm}_k(\varphi_2))$ is half the distance between two critical values of φ_1;

c) $d_{BN}(\mathrm{Dgm}_k(\varphi_1), \mathrm{Dgm}_k(\varphi_2))$ is half the distance between two critical values of φ_2.

Therefore, if $(1), (2), (3)$ do not hold for $\varphi_1, \varphi_2 : \mathcal{S} \to \mathbb{R}$, then $d_{\mathrm{Homeo}(\mathcal{S})}(\varphi_1, \varphi_2)$ cannot be equal to $d_{BN}(\mathrm{Dgm}_k(\varphi_1), \mathrm{Dgm}_k(\varphi_2))$. This means that if there exist two C^1 functions $\varphi_1, \varphi_2 : \mathcal{S} \to \mathbb{R}$ verifying (4) but not $(1), (2), (3)$, then we need new methods to compute $d_{\mathrm{Homeo}(\mathcal{S})}(\varphi_1, \varphi_2)$ and to recognize the pair (φ_1, φ_2) as the right example. As a consequence, the answer to the question asked in this section is still unknown.

Acknowledgement

Work carried out under the auspices of INdAM-GNSAGA.

References

1. Silvia Biasotti, Leila De Floriani, Bianca Falcidieno, Patrizio Frosini, Daniela Giorgi, Claudia Landi, Laura Papaleo, and Michela Spagnuolo, *Describing shapes by geometrical-topological properties of real functions*, ACM Comput. Surv. **40** (2008), no. 4, 12:1–12:87.
2. Francesca Cagliari, Barbara Di Fabio, and Claudia Landi, *The natural pseudo-distance as a quotient pseudo-metric, and applications*, Forum Math. **27** (2015), no. 3, 1729–1742.
3. Gunnar Carlsson and Afra Zomorodian, *The theory of multidimensional persistence*, Discrete Comput. Geom. **42** (2009), no. 1, 71–93.
4. Andrea Cerri and Barbara Di Fabio, *On certain optimal diffeomorphisms between closed curves*, Forum Math. **26** (2014), no. 6, 1611–1628.
5. Andrea Cerri, Barbara Di Fabio, Massimo Ferri, Patrizio Frosini, and Claudia Landi, *Betti numbers in multidimensional persistent homology are stable functions*, Math. Methods Appl. Sci. **36** (2013), no. 12, 1543–1557.
6. Michele d'Amico, Patrizio Frosini, and Claudia Landi, *Natural pseudo-distance and optimal matching between reduced size functions*, Acta Appl. Math. **109** (2010), no. 2, 527–554.
7. Alessandro De Gregorio, *On the set of optimal homeomorphisms for the natural pseudo-distance associated with the Lie group S^1*, Topology Appl. **229** (2017), 187–195.
8. Pietro Donatini and Patrizio Frosini, *Lower bounds for natural pseudodistances via size functions*, Archives of Inequalities and Applications **2** (2004), no. 1, 1–12.
9. ______, *Natural pseudodistances between closed manifolds*, Forum Math. **16** (2004), no. 5, 695–715.
10. ______, *Natural pseudodistances between closed surfaces*, J. Eur. Math. Soc. (JEMS) **9** (2007), no. 2, 331–353.
11. ______, *Natural pseudo-distances between closed curves*, Forum Math. **21** (2009), no. 6, 981–999.
12. Herbert Edelsbrunner and John Harer, *Persistent homology—a survey*, Surveys on discrete and computational geometry, Contemp. Math., vol. 453, Amer. Math. Soc., Providence, RI, 2008, pp. 257–282.
13. Patrizio Frosini, *A distance for similarity classes of submanifolds of a Euclidean space*, Bull. Austral. Math. Soc. **42** (1990), no. 3, 407–416.
14. ______, *Measuring shapes by size functions*, Proc. SPIE **1607** (1992), 122–133.
15. ______, *Una rapida escursione fra le distanze naturali di taglia (Italian)*, Atti Accademia Peloritana dei Pericolanti Cl. Sci. Fis. Mat. Natur. **LXXIII** (1995), 25–40.
16. ______, *G-invariant persistent homology*, Math. Methods Appl. Sci. **38** (2015), no. 6, 1190–1199.
17. ______, *Towards an observer-oriented theory of shape comparison*, Eurographics Workshop on 3D Object Retrieval (Alfredo Ferreira, Andrea Giachetti, and Daniela Giorgi, eds.), The Eurographics Association, 2016, pp. 5–8.

18. Patrizio Frosini and Grzegorz Jabłoński, *Combining persistent homology and invariance groups for shape comparison*, Discrete & Computational Geometry **55** (2016), no. 2, 373–409.
19. Patrizio Frosini and Claudia Landi, *Size theory as a topological tool for computer vision*, Pattern Recognition and Image Analysis **9** (1999), no. 4, 596–603.
20. Patrizio Frosini and Michele Mulazzani, *Size homotopy groups for computation of natural size distances*, B. Belg. Math. Soc.-Sim. **6** (1999), no. 3, 455–464.
21. Patrizio Frosini and Nicola Quercioli, *Some remarks on the algebraic properties of group invariant operators in persistent homology*, Machine Learning and Knowledge Extraction (Cham) (Andreas Holzinger, Peter Kieseberg, A Min Tjoa, and Edgar Weippl, eds.), Springer International Publishing, 2017, pp. 14–24.
22. Steve Y. Oudot, *Persistence theory: from quiver representations to data analysis*, Mathematical Surveys and Monographs, vol. 209, American Mathematical Society, Providence, RI, 2015.
23. Wikipedia contributors, *Measurement — Wikipedia, the free encyclopedia*, `https://en.wikipedia.org/w/index.php?title=Measurement&oldid=870827913`, 2018, [Online; accessed 27-November-2018].

Dipartimento di Matematica, Università di Bologna, Italia
E-mail address: `patrizio.frosini@unibo.it`