

# Metric shape comparison via multidimensional persistent homology

Patrizio Frosini<sup>1,2</sup>

<sup>1</sup>Department of Mathematics, University of Bologna, Italy

<sup>2</sup>ARCES - Vision Mathematics Group, University of Bologna, Italy  
`patrizio.frosini@unibo.it`

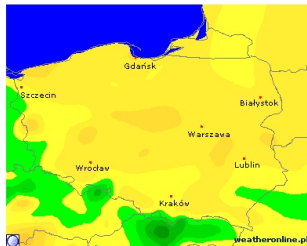
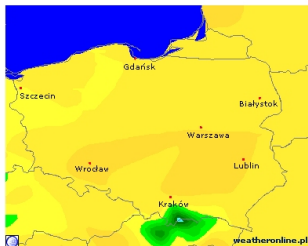
Mini-symposium on *Applied and Computational Algebraic Topology*  
6ECM, Kraków, 4 July 2012

- 1 Comparing topological spaces endowed with vector-valued functions: the natural pseudo-distance  $d_{Hom}(X)$
- 2 Persistent Homology as a tool to get LOWER bounds for the natural pseudo-distance  $d_{Hom}(X)$
- 3 Persistent Homology as a tool to get UPPER bounds for the natural pseudo-distance  $d_{Hom}(X)$
- 4 Extending Persistent Homology to the case of proper subgroups of  $Hom(X)$  (ongoing research).

- 1 Comparing topological spaces endowed with vector-valued functions: the natural pseudo-distance  $d_{Hom(X)}$**
- 2 Persistent Homology as a tool to get LOWER bounds for the natural pseudo-distance  $d_{Hom(X)}$
- 3 Persistent Homology as a tool to get UPPER bounds for the natural pseudo-distance  $d_{Hom(X)}$
- 4 Extending Persistent Homology to the case of proper subgroups of  $Hom(X)$  (ongoing research).

## Exemplifying our problem

In many applications we have to compare two vector-valued functions defined on the same topological space  $X$ , where two functions  $\varphi, \psi : X \rightarrow \mathbb{R}^k$  are considered equivalent if  $\psi = \varphi \circ g$  for a homeomorphism  $g : X \rightarrow X$  chosen in a fixed group  $G$  of homeomorphisms.



*Here the function is the temperature,  
and  $G$  is the set of rigid motions of  $S^2$ .*

## Exemplifying our problem

Another example:



*Here the function is the color, and  $G$  is the set of rotations around the  $z$ -axis.*

## Our formal setting

Assume that  $X$  is a compact topological space. For each subgroup  $G$  of the group  $Hom(X)$  of all homeomorphisms from  $X$  onto  $X$ , we can define the following pseudo-metric on the set of the continuous functions from  $X$  to  $\mathbb{R}^k$ :

$$d_G(\varphi, \psi) = \inf_{g \in G} \max_i \max_{x \in X} |\varphi_i(x) - \psi_i \circ g(x)|.$$

We shall call  $d_G(\varphi, \psi)$  the **natural pseudo-distance** between  $\varphi = (\varphi_1, \dots, \varphi_k)$  and  $\psi = (\psi_1, \dots, \psi_k)$ , associated with the group  $G$ . The  $\mathbb{R}^k$ -valued functions we are considering will be called *filtering (or measuring) functions* on  $X$ .

The functional  $\Theta(g) = \max_i \max_{x \in X} |\varphi_i(x) - \psi_i \circ g(x)|$  represents the “cost of the matching” between filtering functions induced by  $g$ . The lower this cost, the better the matching between the two filtering functions.

## Our formal setting

We observe that

- the natural pseudo-distance  $d_G$  measures the dissimilarity between the functions  $\varphi, \psi$  with respect to the group  $G$ ;
- the value  $d_G$  is small if and only if we can find a homeomorphism from  $X$  onto  $X$  in  $G$  that induces a small change of the filtering function.

## Some results

Until now, most of the results about the natural pseudo-distance  $d_G$  have been proven for the case when  $X$  is a closed manifold  $M$ ,  $G = \text{Hom}(X)$  and  $k = 1$ . Here we recall three of them:

### Theorem (for manifolds)

*Assume that  $M$  is a closed manifold of class  $C^1$  and that  $\varphi, \psi : M \rightarrow \mathbb{R}$  are two functions of class  $C^1$ . Then, if  $d$  denotes the natural pseudo-distance between  $\varphi$  and  $\psi$ , at least one of the following properties holds:*

- *a positive odd integer  $m$  exists, such that  $m \cdot d$  equals the distance between a critical value of  $\varphi$  and a critical value of  $\psi$ ;*
- *a positive even integer  $m$  exists, such that  $m \cdot d$  equals the distance between either two critical values of  $\varphi$  or two critical values of  $\psi$ .*



## Some results

### Theorem (for surfaces)

Assume that  $M$  is a closed surface of class  $C^1$  and that  $\varphi, \psi : M \rightarrow \mathbb{R}$  are two functions of class  $C^1$ . Then, if  $d$  denotes the natural pseudo-distance between  $\varphi$  and  $\psi$ , at least one of the following properties holds:

- $d$  equals the distance between a critical value of  $\varphi$  and a critical value of  $\psi$ ;
- $d$  equals half the distance between two critical values of  $\varphi$ ;
- $d$  equals half the distance between two critical values of  $\psi$ ;
- $d$  equals one third of the distance between a critical value of  $\varphi$  and a critical value of  $\psi$ .

In plain words, referring to the theorem for manifolds, we have that  $m = 1$ , or  $m = 2$ , or  $m = 3$ .

## Some results

### Theorem (for curves)

Assume that  $M$  is a closed curve of class  $C^1$  and that  $\varphi, \psi : M \rightarrow \mathbb{R}$  are two functions of class  $C^1$ . Then, if  $d$  denotes the natural pseudo-distance between  $\varphi$  and  $\psi$ , at least one of the following properties holds:

- $d$  equals the distance between a critical value of  $\varphi$  and a critical value of  $\psi$ ;
- $d$  equals half the distance between two critical values of  $\varphi$ ;
- $d$  equals half the distance between two critical values of  $\psi$ .

In plain words, referring to the theorem for manifolds, we have that  $m = 1$ , or  $m = 2$ .

## Some references about the natural pseudo-distance for topological spaces

- P. Frosini, M. Mulazzani, *Size homotopy groups for computation of natural size distances*, Bulletin of the Belgian Mathematical Society - Simon Stevin, 6 (1999), 455-464.
- P. Donatini, P. Frosini, *Natural pseudodistances between closed topological spaces*, Forum Mathematicum, 16 (2004), n. 5, 695-715.
- P. Donatini, P. Frosini, *Natural pseudodistances between closed surfaces*, Journal of the European Mathematical Society, 9 (2007), 331-353.
- P. Donatini, P. Frosini, *Natural pseudodistances between closed curves*, Forum Mathematicum, 21 (2009), n. 6, 981-999.

- 1 Comparing topological spaces endowed with vector-valued functions: the natural pseudo-distance  $d_{Hom(X)}$
- 2 Persistent Homology as a tool to get LOWER bounds for the natural pseudo-distance  $d_{Hom(X)}$**
- 3 Persistent Homology as a tool to get UPPER bounds for the natural pseudo-distance  $d_{Hom(X)}$
- 4 Extending Persistent Homology to the case of proper subgroups of  $Hom(X)$  (ongoing research).

## Some important remarks

- The natural pseudo-distance is usually difficult to compute

BUT

- in this section we will illustrate how lower bounds for the natural pseudo-distance  $d_{Hom(X)}$  can be obtained by computing Persistent Homology.

NOTE: If  $H \subseteq G$  then  $d_G \leq d_H$ . Therefore,  $d_{Hom(X)} \leq d_G$  for every subgroup  $G$  of the group of all homeomorphisms from  $X$  onto  $X$ . Hence, each lower bound for  $d_{Hom(X)}$  is also a lower bound for  $d_G$ .

## Persistent homology groups in the multidimensional case

Let us recall the definition of persistent homology group.

Let  $X$  be a compact and locally connected topological space. Let  $\varphi = (\varphi_i) : X \rightarrow \mathbb{R}^k$  be a continuous function. The function  $\varphi$  is called a  $k$ -dimensional *filtering function*. Let us consider the sub-level sets  $\{\varphi \preceq u\} = \{x \in X : \varphi_i(x) \leq u_i, i = 1, \dots, k\}$ .

### Definition (Persistent Homology Group)

The  $q$ th *Persistent Homology Group*  $H_q^\varphi(u, v)$  of  $\varphi$  (computed at  $(u, v)$ , with  $u \prec v$ ) is the image of the homomorphism  $H_q(\{\varphi \preceq u\}) \rightarrow H_q(\{\varphi \preceq v\})$  induced by the inclusion map  $\{\varphi \preceq u\} \hookrightarrow \{\varphi \preceq v\}$ .

Here,  $H_q$  denotes the  $q$ th Čech homology group. In this talk we will take coefficients in the field  $\mathbb{R}$ , so that the persistent homology groups are vector spaces.

## A metric to compare persistent homology groups

Several metrics to compare persistent homology groups exist. In this talk, we will use the **multidimensional matching distance (MMD)**  $D_{match}$ .

It is based on a foliation of the space where the parameters  $u, v$  of persistent homology groups vary. The foliation is made by half-planes. On each leaf of this foliation the multidimensional persistent homology group reduces to a **1-dimensional** persistent homology group. The MMD is defined as the sup of the ordinary matching distance, varying the half-plane in the foliation, after a suitable normalization. Here we skip its formal definition.

Biasotti, S., Cerri, A., Frosini, P., Giorgi, D., Landi, C.,  
*Multidimensional size functions for shape comparison*,  
 Journal of Mathematical Imaging and Vision, vol. 32 (2008), n. 2, 161-179.

A. Cerri, B. Di Fabio, M. Ferri, P. Frosini, C. Landi,  
*Betti numbers in multidimensional persistent homology are stable functions*,  
 AMS Acta, University of Bologna, n. 2923, (2010).

## A metric to compare persistent homology groups

Other (pseudo-)distances between MPHGs exist. This is another example:

### Definition

Let  $\varphi, \psi : X \rightarrow \mathbb{R}^k$  be two filtering functions on a compact topological space  $X$ . Let  $H_q^\varphi(u, v)$  and  $H_q^\psi(u, v)$  be the  $q$ -th multidimensional persistent homology groups associated with the functions  $\varphi$  and  $\psi$ , respectively. Let us consider the set  $E$  of all  $\epsilon \geq 0$  such that, setting  $\vec{\epsilon} = (\epsilon, \dots, \epsilon) \in \mathbb{R}^k$ , the following statements hold for each  $(u, v)$  with  $u \prec v$ :

- 1 a surjective homomorphism from a subgroup of  $H_q^\varphi(u, v)$  onto  $H_q^\psi(u - \vec{\epsilon}, v + \vec{\epsilon})$  exists;
- 2 a surjective homomorphism from a subgroup of  $H_q^\psi(u, v)$  onto  $H_q^\varphi(u - \vec{\epsilon}, v + \vec{\epsilon})$  exists.

We define  $d_T(H_q^\varphi, H_q^\psi)$  equal to  $\inf E$ .



## A metric to compare persistent homology groups

The previous distance is described in the paper

P. Frosini, *Stable comparison of multidimensional persistent homology groups with torsion*, Acta Applicandae Mathematicae, 2012 (in press - electronic version available at [www.springerlink.com](http://www.springerlink.com)).

It is related to another approach to the metric comparison of Persistent Homology, i.e. the concept of  $\varepsilon$ -interleaved modules:

F. Chazal, D. Cohen-Steiner, M. Glisse, L.J. Guibas, S.Y. Oudot, *Proximity of persistence modules and their diagrams*, SCG '09: Proceedings of the 25th annual symposium on Computational geometry, Aarhus, Denmark, 237–246 (2009).

M. Lesnick, *The optimality of the interleaving distance on multidimensional persistence modules*, arXiv:1106.5305 (2011).

## Stability of the multidimensional matching distance between persistent Betti numbers

An important property of the multidimensional matching distance is that Persistent Betti numbers (i.e. the ranks of persistent homology groups) are stable with respect to it. This stability can be expressed by means of the inequality


$$D_{match} \leq d_{Hom}(X)$$

















Therefore, the multidimensional matching distance between persistent Betti numbers gives a lower bound for the natural pseudo-distance.

In plain words, the stability of  $D_{match}$  means that a small change of the filtering function induces a small change of the function describing the Persistent Betti numbers.

## Comparing shapes

Stability allows us to use the matching distance to compare shapes:



								
	0.0000 0.0000	0.0181 0.0003	0.1411 0.0025	0.1470 0.0026	0.1325 0.0023	0.1287 0.0022	0.1171 0.0020	0.1187 0.0021
	0.0181 0.0003	0.0000 0.0000	0.1419 0.0026	0.1478 0.0026	0.1304 0.0023	0.1265 0.0022	0.1171 0.0020	0.1187 0.0021
	0.1411 0.0025	0.1419 0.0025	0.0000 0.0000	0.0137 0.0002	0.1583 0.0028	0.1370 0.0024	0.1127 0.0020	0.1017 0.0018
	0.1470 0.0026	0.1478 0.0026	0.0137 0.0002	0.0000 0.0000	0.1533 0.0027	0.1381 0.0024	0.1137 0.0020	0.1021 0.0018
	0.1325 0.0023	0.1304 0.0023	0.1583 0.0028	0.1533 0.0027	0.0000 0.0000	0.0921 0.0014	0.0588 0.0016	0.1000 0.0017
	0.1287 0.0022	0.1265 0.0022	0.1370 0.0024	0.1381 0.0024	0.0921 0.0014	0.0000 0.0000	0.1069 0.0019	0.1048 0.0018
	0.1171 0.0020	0.1171 0.0020	0.1127 0.0020	0.1137 0.0020	0.0588 0.0016	0.1069 0.0019	0.0000 0.0000	0.0350 0.0006
	0.1187 0.0021	0.1187 0.0021	0.1017 0.0018	0.1021 0.0018	0.1000 0.0017	0.1048 0.0018	0.0350 0.0006	0.0000 0.0000

*Upper lines refer to the 2-dimensional matching distance associated with a suitable function  $\varphi = (\varphi_1, \varphi_2)$  taking values in  $\mathbb{R}^2$ .*

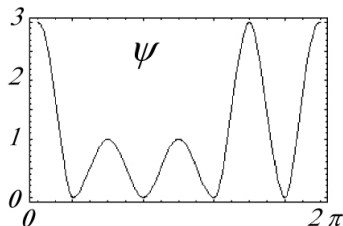
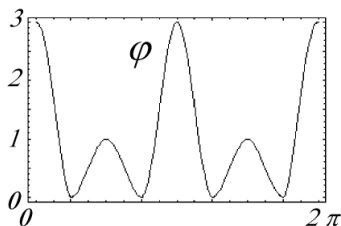
*Lower lines refer to the maximum between the 1-dimensional matching distances associated with  $\varphi_1$  and  $\varphi_2$ , respectively.*

## An important remark

We observe that the natural pseudo-distance contains more information than persistent homology groups.

Example:

Let us consider these filtering functions  $\varphi, \psi : S^1 \rightarrow \mathbb{R}$ :



The persistent homology groups  $H_q^\varphi(u, v)$  and  $H_q^\psi(u, v)$  are isomorphic for every pair  $(u, v)$  in their domain, but  $d_{Hom}(X)(\varphi, \psi) > 0$  (cf. F. Cagliari, M. Ferri and P. Pozzi, *Size functions from a categorical viewpoint*, Acta Appl. Math. 67 (2001), no. 3, 225–235).

- 1 Comparing topological spaces endowed with vector-valued functions: the natural pseudo-distance  $d_{Hom(X)}$
- 2 Persistent Homology as a tool to get LOWER bounds for the natural pseudo-distance  $d_{Hom(X)}$
- 3 Persistent Homology as a tool to get UPPER bounds for the natural pseudo-distance  $d_{Hom(X)}$**
- 4 Extending Persistent Homology to the case of proper subgroups of  $Hom(X)$  (ongoing research).

## Persistent Homology as a tool to get UPPER bounds for $d_{Hom}(X)$

Persistent Homology can be also used as a tool to get UPPER bounds for  $d_{Hom}(X)$ , but this goal is much more difficult than the one of obtaining lower bounds for the same pseudo-distance.

Here we show a recent result concerning regular curves in the plane.

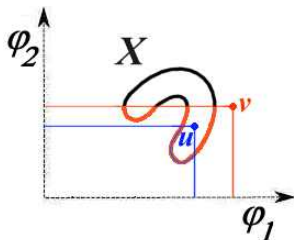
Before proceeding we recall the definition of size function.

## A look at persistent homology in degree zero, i.e. size functions

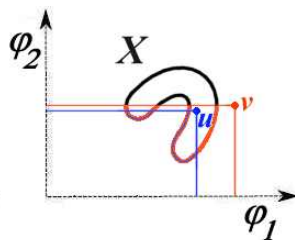
The 0-th persistent homology group at  $(u, v)$  is equivalent to the size function  $\ell(u, v)$ :

### Definition (Frosini 1991, Frosini and Mulazzani 1999)

The **Size Function** of  $\varphi$  is the function  $\ell$  that takes each pair  $(u, v)$  with  $u < v$  to the number  $\ell(u, v)$  of connected components of the set  $\{\varphi \preceq v\}$  that contain at least one point of the set  $\{\varphi \preceq u\}$ .



*In this case the size function takes the value 1.*



*In this case the size function takes the value 2.*

## The family of curves $F_r$

In order to state our theorem we need to focus on a suitable family  $F_r$  of curves.

### Definition

For every positive real number  $r$ , we define  $F_r$  to be the set of all curves represented by functions  $f$  in  $C^2(S^1, \mathbb{R}^2)$  such that

- 1  $f$  is generic;
- 2  $f(S^1)$  is contained in the disk of  $\mathbb{R}^2$  centered at  $(0, 0)$  with radius  $r$ ;
- 3  $f$  is a curve of length  $l_f$  with  $l_f \leq r$ ;
- 4 The (unsigned) curvature of the curve  $f$  is everywhere not greater than  $r$ ;
- 5 Every  $C^1$  curve  $f' : S^1 \rightarrow \mathbb{R}^2$  such that  $f'$  has a distance less than  $\frac{1}{r}$  from  $f$ , with respect to the  $C^1$ -norm, is generic.



## Persistent Homology as a tool to get UPPER bounds for $d_{Hom}(X)$

Let  $s_i : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ , with  $i = 1, 2$ , be the reflections with respect to the coordinate axes:  $s_1(x_1, x_2) = (-x_1, x_2)$ ,  $s_2(x_1, x_2) = (x_1, -x_2)$ .

Let  $\Sigma_2$  be the set of functions obtainable through finite composition of the reflections  $s_1, s_2$  (obviously,  $id \in \Sigma_2$ ).

This result holds:

### Theorem (Frosini and Landi, 2011)

*Let  $r > 0$ . For every  $\epsilon > 0$ , a  $\delta > 0$  exists such that if two curves  $f, g$  belong to the family  $F_r$  and the matching distance between the size functions of  $s \circ f$  and  $s \circ g$  is less than  $\delta$  for every  $s \in \Sigma_2$ , then*

$$d_{Hom}(S^1)(f, g) \leq \epsilon.$$

- 1 Comparing topological spaces endowed with vector-valued functions: the natural pseudo-distance  $d_{Hom(X)}$
- 2 Persistent Homology as a tool to get LOWER bounds for the natural pseudo-distance  $d_{Hom(X)}$
- 3 Persistent Homology as a tool to get UPPER bounds for the natural pseudo-distance  $d_{Hom(X)}$
- 4 **Extending Persistent Homology to the case of proper subgroups of  $Hom(X)$  (ongoing research).**

## The idea of $G$ -invariant Persistent Homology

Classical persistent homology is not tailored on the group  $G$ . In some sense, it is tailored on the group  $\text{Hom}(X)$ .

In order to obtain better lower bounds for the pseudo-distance  $d_G$  we need to adapt persistent homology, and to consider  $G$ -invariant persistent homology.

Roughly speaking, the main idea consists in using a set of chains that is invariant under the action of  $G$ , before applying the ideas of persistent homology.

## The idea of $G$ -invariant Persistent Homology

### Definition

Let us consider a topological space  $X$  and fix a subgroup  $G$  of the group  $\text{Hom}(X)$  of all homeomorphisms from  $X$  onto  $X$ . Let  $(\hat{C}, \partial)$  be a subcomplex of the singular chain complex  $(C(X), \partial)$  of  $X$ , verifying the following property in any degree  $n$ :

(\*) If  $c = \sum_i a_i \sigma_i \in \hat{C}_n$  then  $g(c) = \sum_i a_i (g \circ \sigma_i) \in \hat{C}_n$  for every  $g \in G$  (i.e.  $\hat{C}$  is invariant under the action of the group  $G$ ).

The chain complex  $\hat{C}$  will be said a  $G$ -invariant chain subcomplex of  $C(X)$ . We shall call the group  $H_n(\hat{C}) = \ker \partial_n / \text{im } \partial_{n+1}$  the  $n$ -th homology group associated with the  $G$ -invariant chain complex  $(\hat{C}, \partial)$ .

## The idea of $G$ -invariant Persistent Homology

Let  $\hat{C}$  be a  $G$ -invariant chain subcomplex of  $C(X)$ . For every topological subspace  $X'$  of  $X$ , we can consider the new chain complex  $\dots \xrightarrow{\partial_{n+1}} \hat{C}_n \cap C_n(X') \xrightarrow{\partial_n} \hat{C}_{n-1} \cap C_{n-1}(X') \xrightarrow{\partial_{n-1}} \dots$ . (This new chain complex is not requested to be  $G$ -invariant).

### Definition

This chain complex will be called *chain subcomplex of  $\hat{C}$  induced by restriction to the topological subspace  $X'$* .

Now we can apply this idea to Persistent Homology.

## The idea of $G$ -invariant Persistent Homology

### Definition

Let  $\hat{C}$  be a  $G$ -invariant chain subcomplex of  $C(X)$ . Assume that a continuous function  $\varphi : X \rightarrow \mathbb{R}^k$  is given and consider the chain subcomplex  $\hat{C}^{\varphi \prec u}$  of  $\hat{C}$  induced by restriction to the topological subspace given by the sublevel set  $\{\varphi \prec u\} = \{x \in X : \varphi(x) \prec u\}$ . If  $u, v \in \mathbb{R}^k$  and  $u \prec v$  (i.e.,  $u_j < v_j$  for every index  $j$ ), we can consider the inclusion  $i_{(u,v)}$  of the chain complex  $\hat{C}^{\varphi \prec u}$  into the chain complex  $\hat{C}^{\varphi \prec v}$ . Such an inclusion induces a homomorphism  $i_{(u,v)}^* : H_n(\hat{C}^{\varphi \prec u}) \rightarrow H_n(\hat{C}^{\varphi \prec v})$ . We shall call the group  $H_n(u, v) = i_{(u,v)}^*(H_n(\hat{C}^{\varphi \prec u}))$  the  $n$ -th persistent homology group associated with the  $G$ -invariant chain complex  $\hat{C}$ , computed at point  $(u, v)$ . The rank  $\rho_\varphi(u, v)$  of this group will be said  $n$ -th persistent Betti number associated with the  $G$ -invariant chain complex  $\hat{C}$ , computed at point  $(u, v)$ .

## The idea of $G$ -invariant Persistent Homology

The following result holds:

### Theorem

*Let us consider the persistent Betti number functions  $\rho_\varphi, \rho_\psi$  associated with a  $G$ -invariant chain complex  $\hat{C}$ , with respect to two continuous functions  $\varphi : X \rightarrow \mathbb{R}^k, \psi : X \rightarrow \mathbb{R}^k$ . Let us assume that the groups  $H_n(\hat{C})$  are finitely generated. Then*

$$D_{\text{match}}(\rho_\varphi, \rho_\psi) \leq d_G(\varphi, \psi)$$

This result can be used to obtain lower bounds for  $d_G$ , which are better than the ones obtained by computing classical persistent homology.

## Conclusions

In this short survey we have illustrated the concept of natural pseudo-distance  $d_G$  and the use of Persistent Homology to obtain lower and upper bounds for  $d_G$ .

The proofs of these results are based on a "foliation method" and on the stability of multidimensional persistent homology with respect to function perturbation.