

# Some new results in 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  
`frosini@dm.unibo.it`

Applied Algebraic Topology, ETH Zürich  
4-8 July 2011

- 1 Multi-dimensional persistent homology
- 2 The inverse problem in persistent homology
- 3 Stability of persistent homology with respect to noise affecting both the set and the filtering function

- 1 Multi-dimensional persistent homology**
- 2 The inverse problem in persistent homology
- 3 Stability of persistent homology with respect to noise affecting both the set and the filtering function

## Persistent homology groups in the multidimensional case

Let  $X$  be a compact and locally connected subspace of  $\mathbb{R}^n$ . 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  $X\langle\varphi \preceq u\rangle = \{x \in X : \varphi_i(x) \leq u_i, i = 1, \dots, k\}$ .

### Definition (Persistent Homology Group)

The  $q$ th Persistent Homology Group  $\check{H}_q^{(X, \varphi)}(u, v)$  of  $\varphi$  (computed at  $(u, v)$ ,  $u \prec v$ ) is the image of the homomorphism  $\check{H}_q(X\langle\varphi \preceq u\rangle) \rightarrow \check{H}_q(X\langle\varphi \preceq v\rangle)$  induced by the inclusion map  $X\langle\varphi \preceq u\rangle \hookrightarrow X\langle\varphi \preceq v\rangle$ .

Here,  $\check{H}_q$  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.

## Distances between multidimensional persistent homology groups

The new results described in this talk depend on the stability of multidimensional persistent homology groups.

At the present time, some distances exist, with respect to which multidimensional persistent homology groups have been proven stable.

## Distances between multidimensional persistent homology groups

The distance we use in this talk is the **multidimensional matching distance - MMD**. 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.

### References:

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).

## Distances between multidimensional persistent homology groups

Other (pseudo)distances between MPHGs exist. For example,

### Definition

Let  $\varphi : X \rightarrow \mathbb{R}^k$ ,  $\psi : Y \rightarrow \mathbb{R}^k$  be two continuous filtering functions. Let  $H_q^{(X,\varphi)}(u, v)$  and  $H_q^{(Y,\psi)}(u, v)$  the  $q$ -th multidimensional persistent homology groups associated with the pairs  $(X, \varphi)$  and  $(Y, \psi)$ , 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  $(u, v) \in \Delta^+$ :

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

We define  $d_T \left( H_q^{(X,\varphi)}, H_q^{(Y,\psi)} \right)$  equal to  $\inf E$  if  $E$  is not empty, and equal to  $\infty$  otherwise.

## Distances between multidimensional persistent homology groups

The previous distance is described in the paper

P. Frosini, *Stable comparison of multidimensional persistent homology groups with torsion*, AMS Acta, University of Bologna, n. 2922, (2010).

It is related to 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).



1 Multi-dimensional persistent homology

**2 The inverse problem in persistent homology**

3 Stability of persistent homology with respect to noise affecting both the set and the filtering function

## The inverse problem in persistent homology

Persistent homology transforms each topological space  $X$  endowed with a continuous  $\mathbb{R}^k$ -valued function  $\varphi$  into persistent homology groups, depending on the parameters  $u, v \in \mathbb{R}^k$ .

One of the most natural and important open question in persistent homology is the following one:

*How much information is lost when we change  $(X, \varphi)$  into its persistent homology groups?*

## A trivial observation

The knowledge of persistent homology does not allow us to reconstruct the pair  $(X, \varphi)$ , in general.

Indeed, the following simple statement holds:

### Proposition

*Let us assume that two filtering functions  $\varphi : X \rightarrow \mathbb{R}^k$ ,  $\psi : Y \rightarrow \mathbb{R}^k$  are given. If a homeomorphism  $h : X \rightarrow Y$  exists, such that  $\varphi = \psi \circ h$ , then the persistent homology groups of  $(X, \varphi)$  and  $(Y, \psi)$  coincide.*

Two such pairs will be said **equivalent**.

As a consequence, the best that we can hope for is to reconstruct  $(X, \varphi)$  up to homeomorphisms that preserve the filtering function.

## A natural equivalence (and the corresponding metric)

Previous proposition suggests to use the following (pseudo)metric:

$$\delta((X, \varphi), (Y, \psi)) = \inf_{h \in \text{Hom}(X, Y)} \max_i \max_{x \in X} |\varphi_i(x) - \psi_i \circ h(x)|$$

if the set  $\text{Hom}(X, Y)$  of the homeomorphisms between  $X$  and  $Y$  is not empty, while  $\delta$  takes the value  $+\infty$  otherwise.

We call  $\delta((X, \varphi), (Y, \psi))$  the **natural pseudo-distance** between  $(X, \varphi)$  and  $(Y, \psi)$ .

Obviously, if the pairs  $(X, \varphi), (Y, \psi)$  are equivalent, then their natural pseudo-distance vanishes.

## Some references about the natural pseudo-distance

- 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 manifolds*, 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.

## The case of curves

Let us examine the inverse problem in persistent homology, starting from the simplest case: the case of connected closed curves.

Let us observe that, for a connected closed curve  $X$  endowed with a filtering function  $\varphi$ ,

- The persistent homology groups are trivial in each degree  $q \geq 2$ ;
- The persistent homology groups in degree 1 contain little information: they equal  $\mathbb{R}$  if and only if  $u_i \geq \max_X \varphi_i$  for every index  $i$ , otherwise they coincide with the trivial group.

Therefore persistent homology groups in positive degree does not allow us to reconstruct the pair  $(X, \varphi)$  up to homeomorphisms.

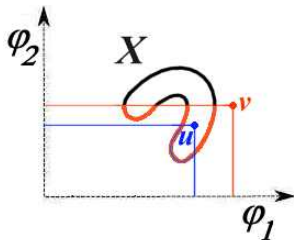
## The degree zero: size functions

Let us examine what happens for the degree 0.

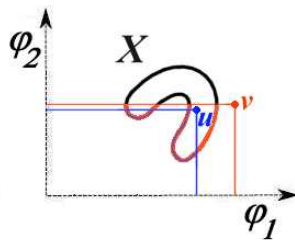
The 0-th persistent homology group at  $(u, v)$  is equivalent to the size function  $\ell(u, v)$ :

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

The **Size Function** of  $(X, \varphi)$  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  $X\langle\varphi \leq v\rangle$  that contain at least one point of the set  $X\langle\varphi \leq u\rangle$ .



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



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

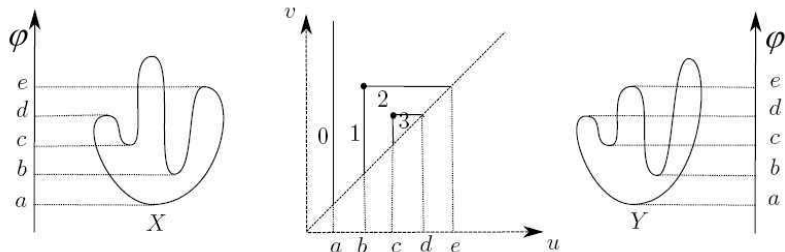
## Some important theoretical facts concerning size functions (=persistent homology groups in degree 0):

- The theory of size functions for filtering functions taking values in  $\mathbb{R}^k$  can be reduced to the case of size functions taking values in  $\mathbb{R}$ , by a suitable foliation of their domain;
- On each leaf of the foliation, size functions are described by a collection of points (corresponding to Persistent Diagrams in persistent homology);
- Size functions can be compared by measuring the difference between these collections of points, by a matching distance;
- Size functions (and persistent homology groups) are stable with respect to perturbations of the filtering functions (measured via the max-norm). More precisely, the matching distance between two size functions is a lower bound for the corresponding natural pseudo-distance.



## Some examples concerning curves

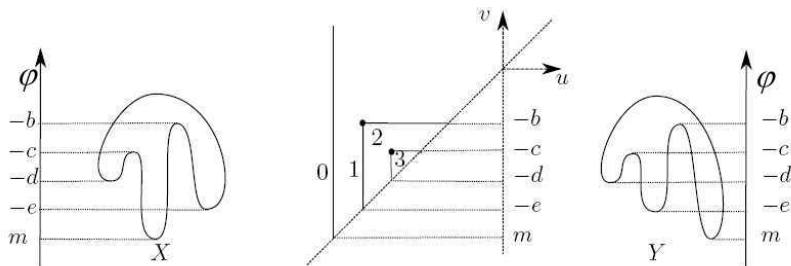
The filtering functions taking values in  $\mathbb{R}$  are not sufficient to reconstruct a curve:



These two curves cannot be distinguished by persistent homology with respect to the filtering function  $\varphi(x, y) = y$ . However, they are not equivalent.

## Some examples concerning curves

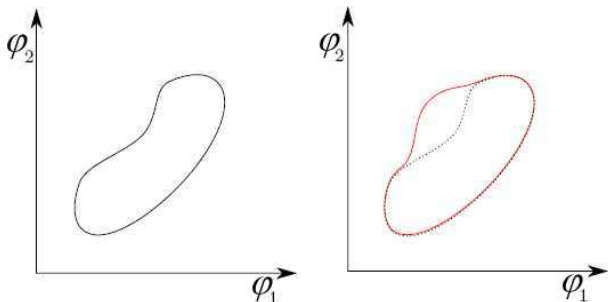
What happens if we add the information given by reversing the filtering function?



Previous two curves cannot be distinguished by persistent homology even with respect to the filtering function  $\varphi(x, y) = -y$ .

## Some examples concerning curves

What happens if we consider a filtering function taking values in  $\mathbb{R}^2$ ?



These two curves cannot be distinguished by persistent homology with respect to the filtering function  $\varphi(x, y) = (x, y)$ . However, they are not equivalent.

## Uniqueness of models of curves in persistent homology

These examples seem to suggest that, also in the case of curves, the pair  $(X, \varphi)$  cannot be reconstructed by persistent homology.

Therefore, it could be surprising that the following statement holds:

### Theorem (Frosini and Landi, 2011)

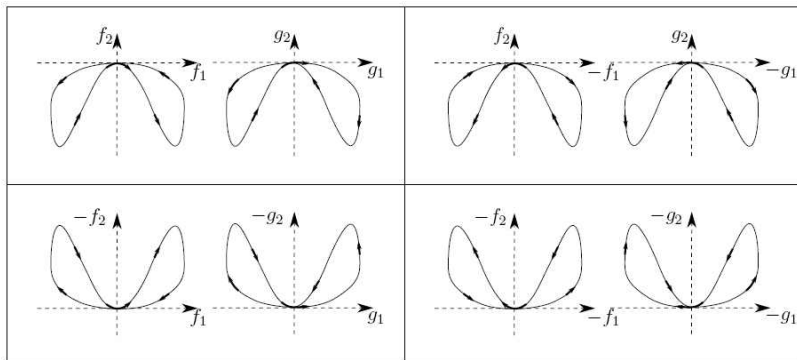
*Let  $f, g : S^1 \rightarrow \mathbb{R}^2$  be “generic” curves in  $\mathbb{R}^2$ . If the size functions of the four pairs of filtering functions  $(\pm f_1, \pm f_2), (\pm g_1, \pm g_2)$  (with corresponding signs) coincide, then there exists a  $C^1$ -diffeomorphism  $h : S^1 \rightarrow S^1$  such that  $g \circ h = f$ . Moreover, it is unique.*

In other words, **two generic closed curves  $f, g : S^1 \rightarrow \mathbb{R}^2$  are equivalent if and only if they have the same persistent homology groups in degree 0, with respect to the filtering functions  $(\pm f_1, \pm f_2), (\pm g_1, \pm g_2)$ .**

## Uniqueness of models of curves in persistent homology

The assumption that  $f$  and  $g$  are generic is important, as the following example shows.

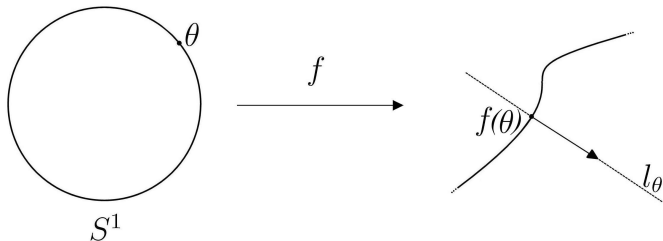
Let us consider the cases  $\varphi = (\pm f_1, \pm f_2)$  and  $\varphi = (\pm g_1, \pm g_2)$ .



## Sketch of proof

The proof can be divided into three lemmas.  
The first lemma requires a preliminary definition.

Let  $f$  be a generic function. For each point  $\theta \in S^1$ , let  $l_\theta$  be the line orthogonal to the line tangent to  $f(S^1)$  at  $f(\theta)$ . The set  $N_f$  is the set of all points  $\theta$  of  $S^1$  such that  $l_\theta$  is not parallel to the coordinate axes of  $\mathbb{R}^2$ . By applying reflections we can assume that the slope of  $l_\theta$  is positive.



## Sketch of proof

Now, we can give the first lemma used in the proof of our theorem:

### Lemma

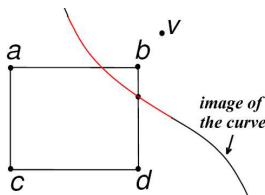
*Let  $f, g : S^1 \rightarrow \mathbb{R}^2$  be generic functions. If the size functions with respect to the four pairs of filtering functions  $(\pm f_1, \pm f_2)$ ,  $(\pm g_1, \pm g_2)$  (with corresponding signs) coincide, then  $f(\bar{N}_f) \subseteq g(S^1)$  and  $g(\bar{N}_g) \subseteq f(S^1)$ , where  $\bar{N}_f$  and  $\bar{N}_g$  denote the closure of  $N_f$  and  $N_g$ , respectively, in  $S^1$ .*

## Sketch of proof

The proof of the previous lemma is based on this formula:

$$\text{rk } H_0(b, v) - \text{rk } H_0(d, v) - \text{rk } H_0(a, v) + \text{rk } H_0(c, v) = 1 \quad (1)$$

where  $\text{rk } H_0(u, v)$  denotes the rank of the 0-th persistent homology group with respect to the filtering function  $f$ , computed at the point  $(u, v)$  in the case described in the following figure.



### Observation

The first term in equation (1) *is not* the multiplicity of a point in a persistence diagram.



## Sketch of proof

The second lemma used in the proof of our theorem:

### Lemma

*Let  $f, g : S^1 \rightarrow \mathbb{R}^2$  be generic functions. If  $f(\bar{N}_f) \subseteq g(S^1)$  and  $g(\bar{N}_g) \subseteq f(S^1)$ , then  $f(S^1) \setminus f(\bar{N}_f) \subseteq g(S^1)$  and  $g(S^1) \setminus g(\bar{N}_g) \subseteq f(S^1)$ . Therefore  $f(S^1) = g(S^1)$ .*

## Sketch of proof

The two previous lemmas imply this third statement:

### Lemma

*Let  $f, g : S^1 \rightarrow \mathbb{R}^2$  be generic functions. If the size functions with respect to the four pairs of filtering functions  $(\pm f_1, \pm f_2), (\pm g_1, \pm g_2)$  (with corresponding signs) coincide, then  $f(S^1) = g(S^1)$ .*

## Sketch of proof

The proof of our theorem ends with two parts:

**Existence of the diffeomorphism.** We define the diffeomorphism  $h$  by taking each point  $P$  of the first curve to the unique point  $Q$  of the second curve such that  $f(P) = g(Q)$  and the tangent spaces at  $f(P)$  and  $g(Q)$  coincide.

**Uniqueness of the diffeomorphism.** It follows from the fact that the filtering functions are supposed to be generic, so that the curves have a finite number of double points.

## Uniqueness of models of curves in persistent homology

Previous theorem requires the exact equality of the 0-th persistent homology groups, in order to be applied. However, it can be adapted to the case that the 0-th persistent homology groups are just close to each other with respect to the MMD. In order to state the next theorem, we need to consider a restricted space of functions.

For every positive real number  $m$ , we define  $F_m$  to be the subset of the 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 centered at  $(0, 0)$  with radius  $m$ ;
- 3  $f$  is a curve of length  $l_f$  with  $l_f \leq m$ ;
- 4 The (unsigned) curvature of the curve  $f$  is everywhere not greater than  $m$ ;
- 5 Every  $C^1$  function  $f^* : S^1 \rightarrow \mathbb{R}^2$  such that  $f^*$  has a distance less than  $\frac{1}{m}$  from  $f$ , with respect to the  $C^1$ -norm, is generic.

## Uniqueness of models of curves in persistent homology

Now we can give our main result:

### Theorem (Frosini and Landi, 2011)

*Let  $m > 0$ . For every  $\epsilon > 0$ , an  $\eta > 0$  exists such that if  $f, g \in F_m$  and the matching distance between the size functions of the four pairs of filtering functions  $(\pm f_1, \pm f_2), (\pm g_1, \pm g_2)$  (with corresponding signs) is not greater than  $\eta$ , then there exists a  $C^1$ -diffeomorphism  $h : S^1 \rightarrow S^1$  such that  $\|f - g \circ h\|_\infty \leq \epsilon$ .*

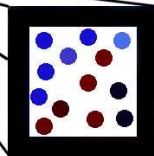
In other words, two generic closed curves  $f, g \in F_m$  are close to each other with respect to the natural pseudo-distance if and only if their persistent homology groups in degree 0, with respect to the filtering functions  $(\pm f_1, \pm f_2), (\pm g_1, \pm g_2)$ , are close to each other with respect to the multidimensional matching distance.

## Sketch of proof

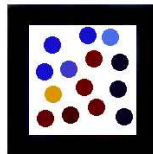
- Assume that our claim is false and show that then the definition of  $F_m$  allows us to apply the **Ascoli-Arzelà Theorem**, obtaining two new limit curves, endowed with the filtering functions  $\bar{f} = (\bar{f}_1, \bar{f}_2), \bar{g} = (\bar{g}_1, \bar{g}_2)$ ;
- Apply the **Multidimensional Stability Theorem in degree 0** and deduce that the persistent homology groups in degree 0 of these curves coincide, with respect to the filtering functions  $(\pm\bar{f}_1, \pm\bar{f}_2), (\pm\bar{g}_1, \pm\bar{g}_2)$ ;
- Apply the **previous theorem (concerning uniqueness of curves)** to our new curves  $\bar{f}, \bar{g}$ , so obtaining that their natural pseudodistance vanishes;
- Show that the fact the natural pseudodistance between  $\bar{f}$  and  $\bar{g}$  vanishes implies a contradiction.

- 1 Multi-dimensional persistent homology
- 2 The inverse problem in persistent homology
- 3 Stability of persistent homology with respect to noise affecting both the set and the filtering function**

# Stability with respect to data perturbation



DATA  
PERTURBATION



*Jacob and the Angel*  
 Marc Chagall  
 Minster of Our Lady (Fraumünster) - Zürich



## An important remark

In many applicative problems we have to manage data represented by a cloud of points, labelled by vectors (in our example, a vector RGB describing the color). They approximate a topological space  $X$  endowed with an  $\mathbb{R}^k$ -valued function  $\varphi$ .

If we are interested to study  $X$  with respect to the property described by  $\varphi$ , we cannot confine ourselves to try to reconstruct  $X$  and forget the function  $\varphi$  (in our example, this would be equivalent to study the painting without considering its colors).

On the other hand, the study of  $\varphi$  from a topological point of view is difficult, without considering the problem given by the fact that our cloud of points is topologically quite different from the original topological space  $X$ .

Is there a mathematical approach that allows us to manage both these problems at the same time?

## Our main stability result

The idea is adding to the filtering function a component that describes the “*belonging level*” of each point in the domain. Indeed, the following result holds, with respect to the Hausdorff distance  $\delta_H$ :

### Theorem (Frosini and Landi, 2010)

*Let  $K_1, K_2$  be non-empty closed subsets of a triangulable subspace  $X$  of  $\mathbb{R}^n$ . Let  $d_{K_1}, d_{K_2} : X \rightarrow \mathbb{R}$  be their respective distance functions. Moreover, let  $\varphi_1, \varphi_2 : X \rightarrow \mathbb{R}^k$  be vector-valued continuous functions. Then, defining  $\Phi_1, \Phi_2 : X \rightarrow \mathbb{R}^{k+1}$  by  $\Phi_1 = (d_{K_1}, \varphi_1)$  and  $\Phi_2 = (d_{K_2}, \varphi_2)$ , the following inequality holds:*

$$D_{\text{match}} \left( \check{H}_q^{(X, \Phi_1)}, \check{H}_q^{(X, \Phi_2)} \right) \leq \max \{ \delta_H(K_1, K_2), \|\varphi_1 - \varphi_2\|_\infty \}.$$

## Our main stability result

In other words, if the cloud of points is close to the topological space  $X$  with respect to the Hausdorff distance and the labelling function is close to the original filtering function  $\varphi$  with respect to the max-norm, then the persistent homology of the labelled cloud of points and the persistent homology of  $(X, \varphi)$  are close to each other.

## Two related papers:

- 1 F. Chazal, D. Cohen-Steiner, L. J. Guibas, F. Mémoli and S. Y. Oudot, *Gromov-Hausdorff stable signatures for shapes using persistence*, Computer Graphics Forum (proc. SGP 2009), 1393–1403.
- 2 D. Cohen-Steiner, H. Edelsbrunner, J. Harer and D. Morozov, *Persistent homology for kernels, images, and cokernels*, SODA 2009: Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms, 1011-1020.

## Comments

- 1 As regards the first paper, we underline that the problem we are interested in does not concern just the approximation of a topological space, but at the same time also the approximation of an arbitrary multidimensional filtering function.
- 2 In the other paper the problem of approximating both the topological space and the filtering function is studied and a strong result is presented (*Stability Theorem for noisy domains*). However, it has some limitations in applicability. For example, the request that the norm of the gradient of the function  $\tilde{h}$  defining the topological space has a positive lower bound prevents us to use that result when the topological space is a curve in  $\mathbb{R}^2$  or a surface in  $\mathbb{R}^3$ . Furthermore the request that  $\tilde{h}$  is smooth prevents us to study topological spaces that are not smooth sets.  
The theorem illustrated in this talk has not these limitations.

## An approximation result

A possible objection:

*“Previous theorem guarantees stability, but we have to pay a price: both the topological space and the filtering function have been changed. So we are studying a problem that differs from the original one.”*

The next result shows that our approach **does not forget** the original problem. Indeed, **the original problem can be seen as a limit case in our model.**

## An approximation result

### Theorem (Frosini and Landi, 2010)

Let  $K$  be a non-empty triangulable subset of a triangulable subspace  $X$  of  $\mathbb{R}^n$ . Moreover, let  $\varphi : X \rightarrow \mathbb{R}^k$  be a continuous function. Setting  $\Phi : X \rightarrow \mathbb{R}^{k+1}$ ,  $\Phi = (d_K, \varphi)$ , for every  $u, v \in \mathbb{R}^k$  with  $u \prec v$  and every small enough  $b > 0$  we have that

$$\check{H}_q^{(X, \varphi|_K)}(u, v) = \check{H}_q^{(X, \Phi)}((0, u), (b, v)).$$

## Conclusions

In this talk we have shown that the foliation method and the stability of multidimensional persistent homology with respect to function perturbation have at least two relevant consequences:

- The proof that 2-dimensional persistent homology allows us to solve the **inverse problem for curves**;
- The proof of the stability of persistent homology groups with respect to **both the topological space and function** perturbation.