Recent advances in multi-dimensional Persistent Homology

Patrizio Frosini[1,2]

[1]Department of Mathematics, University of Bologna, Italy
[2]ARCES - Vision Mathematics Group, University of Bologna, Italy
`frosini@dm.unibo.it`

ATMCS 4
Münster, 21-26 June 2010

Outline

1. **Multi-dimensional Persistent Homology**
2. **Comparing multi-dimensional Persistent Homology**
3. **Stability with respect to noisy functions and noisy sets**
4. **Computing our distance**

The results I am going to present refer to a collective work of the Vision Mathematics Group (Niccolò Cavazza, Andrea Cerri, Barbara Di Fabio, Massimo Ferri, Patrizio Frosini, Claudia Landi).

The experimental results I shall show at the end of this talk have been obtained in a joint work with the C.N.R. IMATI Group (Silvia Biasotti, Daniela Giorgi).

Content of this talk

In brief, the main message of this talk:

- A distance to compare multi-dimensional persistent homology is available.
- Multi-dimensional persistent homology is stable with respect to this distance.
- This distance can be approximated with arbitrary precision.

1. **Multi-dimensional Persistent Homology**
2. Comparing multi-dimensional Persistent Homology
3. Stability with respect to noisy functions and noisy sets
4. Computing our distance

Main definitions:

Given a space X and a continuous function $\vec{\varphi} : X \to \mathbb{R}^k$,

Lower level sets

For every $\vec{u} \in \mathbb{R}^k$, $X\langle\vec{\varphi} \preceq \vec{u}\rangle = \{x \in X : \vec{\varphi}(x) \preceq \vec{u}\}$.
($(u_1, \ldots, u_k) \preceq (v_1, \ldots, v_k)$ means $u_j \leq v_j$ for every index j.)

Definition (Carlsson&Zomorodian 2007)

The Multi-dimensional persistent homology groups of $(X, \vec{\varphi})$ are the groups

$$H_q^{\vec{u},\vec{v}}(X, \vec{\varphi}) = ImH_q\left(X\langle\vec{\varphi} \preceq \vec{u}\rangle \hookrightarrow X\langle\vec{\varphi} \preceq \vec{v}\rangle\right)$$

for $\vec{u} \prec \vec{v}$.

Homology coefficients taken in a field $\mathbb{K}$

Definition (Carlsson&Zomorodian 2007)

The rank invariants of $(X, \vec{\varphi})$ are functions

$$\rho_{(X,\vec{\varphi}),q} : \{(\vec{u}, \vec{v}) \in \mathbb{R}^k \times \mathbb{R}^k : \vec{u} \prec \vec{v}\} \to \mathbb{N} \cup \{\infty\},\ q \in \mathbb{Z},$$

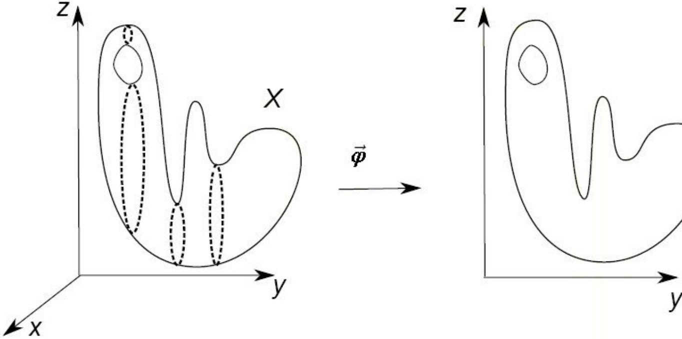
such that $\rho_{(X,\vec{\varphi}),q}(\vec{u}, \vec{v})$ equals the rank of the persistent homology group $H_q^{\vec{u},\vec{v}}(X, \vec{\varphi})$.

Case $q = 0$:

Rank invariants are also called size functions [Frosini et al. 1991,....]

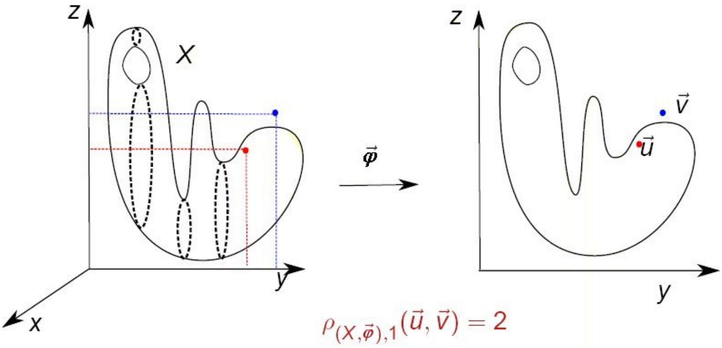
Example of Rank Invariant

$\vec{\varphi} : X \to \mathbb{R}^2$, $\vec{\varphi} = (y, z)$, $\rho_{(X,\vec{\varphi}),1} : \{(\vec{u}, \vec{v}) \in \mathbb{R}^2 \times \mathbb{R}^2 : \vec{u} \prec \vec{v}\} \to \mathbb{N}$
$\rho_{(X,\vec{\varphi}),1}(\vec{u}, \vec{v})$ = number of 1-homology classes born before $\vec{u}$ and still alive at $\vec{v}$



Example of Rank Invariant

$\vec{\varphi} : X \to \mathbb{R}^2$, $\vec{\varphi} = (y, z)$,
$\rho_{(X,\vec{\varphi}),1} : \{(\vec{u}, \vec{v}) \in \mathbb{R}^2 \times \mathbb{R}^2 : \vec{u} \prec \vec{v} \in \mathbb{R}^2 \times \mathbb{R}^2\} \to \mathbb{N}$
$\rho_{(X,\vec{\varphi}),1}(\vec{u}, \vec{v}) =$ number of 1-homology classes born before $\vec{u}$ and still alive at $\vec{v}$



$\rho_{(X,\vec{\varphi}),1}(\vec{u}, \vec{v}) = 2$

Issues

- Representation of rank invariants via persistence diagrams and *the foliation method*;
- Comparison of rank invariants via a suitable distance d_{match};
- Stability of rank invariants with respect to perturbations of $\vec{\varphi}$;
- Stability of rank invariants with respect to perturbations of X;
- Computability of d_{match}.

Representation of rank invariants via persistence diagrams

HOW CAN WE REPRESENT PERSISTENT HOMOLOGY?

In the 1-dimensional case (i.e. filtering functions taking values in $\mathbb{R}$) persistent Betti numbers can be represented via persistence diagrams (H. Edelsbrunner, D. Letscher and A. Zomorodian, *Topological persistence and simplification*, Discrete Comput. Geom. 28 (2002), 511-533.)

Persistence diagrams are collections of points in the real plane, endowed with multiplicities.

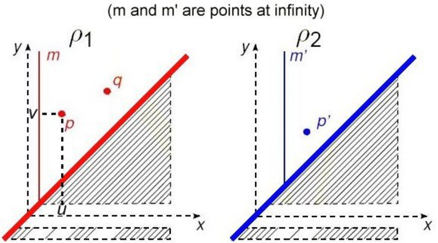
An equivalent concept (formal series of cornerpoints) has been introduced to represent size functions, i.e. 0-th Betti numbers (P. Frosini and C. Landi, *Size Theory as a Topological Tool for Computer Vision*, Pattern Recognition And Image Analysis, vol. 9 (4) (1999), pp. 596-603.)

Persistence Diagrams

Persistence diagrams have been developed to represent persistent homology in the 1-dimensional case.

The persistence diagram consists of a collection of points (with multiplicities): each point (u, v) in the collection represents the birth and the death of a homology class.

The values u and v are the “times” of birth and death, respectively.



Two examples of persistence diagrams

Foliation Method

How can we represent multi-dimensional Persistent Homology?

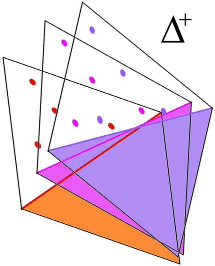
An idea to do that is the Foliation Method

References:

- S. Biasotti, A. Cerri, P. Frosini, D. Giorgi and C. Landi, *Multidimensional size functions for shape comparison*, Journal of Mathematical Imaging and Vision, vol. 32 (2008), n. 2, 161-179.
- F. Cagliari, B. Di Fabio and M. Ferri, *One-dimensional reduction of multi-dimensional persistent homology*, Proc. Amer. Math. Soc. 138 (2010), 3003-3017.

Foliation Method

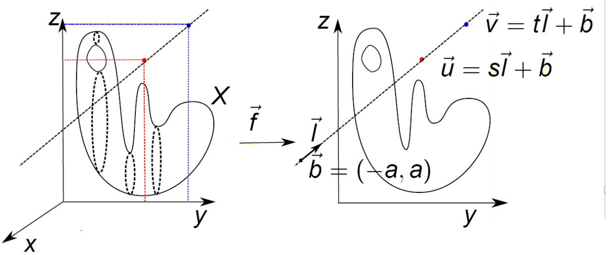
THE MAIN IDEA:
We can cut the domain Δ^+ of the multi-dimensional rank invariant (i.e. $\{(\vec{u}, \vec{v}) \in \mathbb{R}^k \times \mathbb{R}^k : \vec{u} \prec \vec{v}\}$) by a family of half-planes (indeed, a foliation of Δ^+). If this family is chosen in a suitable way, the restriction of the rank invariant to each half-plane can be seen as a 1-dimensional rank invariant (i.e. a rank invariant computed with respect to a suitable filtering function, taking values in $\mathbb{R}$).



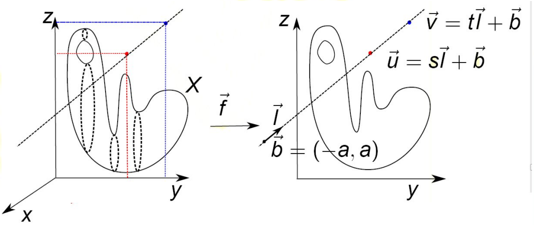
Foliation Method

HOW DO WE DO THAT, IN PRACTICE?

If we are interested in computing the rank invariant at the point $(\vec{u}, \vec{v})$, we consider the line r through $\vec{u}$ and $\vec{v}$. We set $\vec{l}$ equal to the unit vector $\frac{\vec{v}-\vec{u}}{\|\vec{v}-\vec{u}\|}$, and $\vec{b}$ equal to the unique point on the line r such that the summation of its coordinates vanishes.



Foliation Method



The keypoint is that the 1-dimensional rank invariant *"computed on the half-plane"*

$$\pi_{(\vec{l},\vec{b})} : \begin{cases} \vec{u} = s\vec{l} + \vec{b} \\ \vec{v} = t\vec{l} + \vec{b} \end{cases} \qquad s, t \in \mathbb{R}, s < t$$

with respect to a suitable filtering function on X equals the multidimensional rank invariant computed on the same plane.

Reduction to the case k = 1

The 1-dimensional filtering function:

$$F^{\vec{\varphi}}_{(\vec{l},\vec{b})}(x) = \max_{i=1,\dots,k}\left\{\frac{\varphi_i(x) - b_i}{l_i}\right\}.$$

The pairs $(\vec{l},\vec{b})$ will be called admissible pairs.

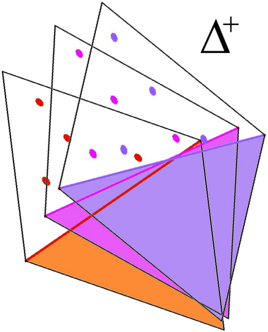
Reduction Theorem

For every $(\vec{u},\vec{v}) = (s\vec{l}+\vec{b}, t\vec{l}+\vec{b}) \in \pi_{(\vec{l},\vec{b})}$ it holds that

$$\check{\rho}_{(X,\vec{\varphi}),q}(\vec{u},\vec{v}) = \check{\rho}_{\left(X,F^{\vec{\varphi}}_{(\vec{l},\vec{b})}\right),q}(s,t).$$

Reduction to the case k = 1

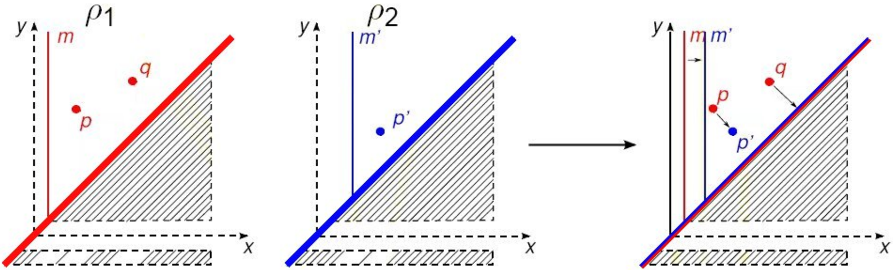
Rank invariants are represented by persistence diagrams leaf-by-leaf



No contradiction with Carlsson-Zomorodian's result....

1. Multi-dimensional Persistent Homology
2. **Comparing multi-dimensional Persistent Homology**
3. Stability with respect to noisy functions and noisy sets
4. Computing our distance

Matching distance in the 1-dimensional case



$$d_{match}(\rho_1, \rho_2) = \min_{\gamma:D\to D'} \max_{q\in D} \|q - \gamma(q)\|_\infty$$

$D = \{m, p, q, \ldots\}$, $D' = \{m', p', \ldots\}$, γ bijective.

Matching distance in the multi-dimensional case

In the multi-dimensional case we can introduce the following

Multi-dimensional Matching Distance

$$D_{match}\left(\check{\rho}_{(X,\vec{\varphi}),q},\check{\rho}_{(X,\vec{\psi}),q}\right)=\sup_{(\vec{l},\vec{b})}\min_{i=1,\dots,k} l_i\cdot d_{match}(\check{\rho}_{(X,F^{\vec{\varphi}}_{(\vec{l},\vec{b})}),q},\check{\rho}_{(X,G^{\vec{\psi}}_{(\vec{l},\vec{b})}),q})$$

where

$$F^{\vec{\varphi}}_{(\vec{l},\vec{b})}(x)=\max_{i=1,\dots,k}\left\{\frac{\varphi_i(x)-b_i}{l_i}\right\},\quad G^{\vec{\psi}}_{(\vec{l},\vec{b})}(x)=\max_{i=1,\dots,k}\left\{\frac{\psi_i(x)-b_i}{l_i}\right\}.$$

1. Multi-dimensional Persistent Homology
2. Comparing multi-dimensional Persistent Homology
3. **Stability with respect to noisy functions and noisy sets**
4. Computing our distance

Foliation Method

THE IMPORTANCE OF USING THE MULTI-DIMENSIONAL MATCHING DISTANCE

As we see in the next slides, by using the multi-dimensional Matching Distance we get stability with respect to both perturbations of the filtering functions and perturbations of the topological spaces on which these functions are defined.

This is a key point in order to apply multi-dimensional Persistent Homology to real problems.

Stability with respect to perturbation of the filtering function

Let X be a triangulable space

One-Dimensional Stability Theorem:

$\varphi, \psi : X \to \mathbb{R}$ continuous functions. Then

$$d_{match}(\check{\rho}_{(X,\varphi),q}, \check{\rho}_{(X,\psi),q}) \leq \max_{x \in X} |\varphi(x) - \psi(x)|.$$

Reference: D. Cohen-Steiner, H. Edelsbrunner, J. Harer, *Stability of persistence diagrams*, Discrete Comput. Geom. 37 (2007), no. 1, 103–120.

We have proven this new result:

Multi-dimensional Stability Theorem

$\vec{\varphi}, \vec{\psi} : X \to \mathbb{R}^k$ continuous functions. Then

$$D_{match}\left(\check{\rho}_{(X,\vec{\varphi}),q}, \check{\rho}_{(X,\vec{\psi}),q}\right) \leq \max_{x \in X} \|\vec{\varphi}(x) - \vec{\psi}(x)\|_\infty.$$

An example (persistent homology in degree 0)

- Experimental results on a set of 8 human models represented by triangular meshes.
- For each model, we take the 2-dimensional measuring function $\vec{\varphi} = (\varphi_1, \varphi_2)$ with

$$\varphi_1(P_i) = 1 - \frac{||P_i - (B + \vec{w})||}{\max_j ||P_j - (B + \vec{w})||}.$$

where B is the center of mass of the model and

$$\vec{w} = \frac{\sum_{i=1}^{n} (P_i - B)||P_i - B||^2}{\sum_{i=1}^{n} ||P_i - B||^2};$$

similarly, $\varphi_2(P_i) = 1 - \frac{||P_i-(B-\vec{w})||}{\max_j ||P_j-(B-\vec{w})||}$.

An example (continued)

The 2D matching distance w.r.t. $\vec{\varphi} = (\varphi_1, \varphi_2)$ and the max of the 1D matching distances w.r.t. φ_1 and φ_2:

								
	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

Stability with respect to perturbation of the sets

IDEA: We can translate the stability w.r.t. noisy domains into the stability w.r.t. noisy functions:

$$K \subseteq D \subseteq \mathbb{R}^n \quad \rightsquigarrow \quad d_K : D \to \mathbb{R},\ d_K(x) = \inf_{y \in K} \|y - x\|$$

$$\vec{\varphi} : D \to \mathbb{R}^k \quad \rightsquigarrow \quad \vec{\Phi} : D \to \mathbb{R}^{k+1}, \vec{\Phi} = (d_K, \vec{\varphi})$$

Stability w.r.t. noisy domains

Let K_1, K_2 be non-empty closed subsets of a triangulable subspace D of $\mathbb{R}^n$. Let $d_{K_1}, d_{K_2} : D \to \mathbb{R}$ their respective distance functions. Moreover, let $\vec{\varphi}_1, \vec{\varphi}_2 : D \to \mathbb{R}^k$ be vector-valued continuous functions. Then, defining $\vec{\Phi}_1, \vec{\Phi}_2 : D \to \mathbb{R}^{k+1}$ by $\vec{\Phi}_1 = (d_{K_1}, \vec{\varphi}_1)$ and $\vec{\Phi}_2 = (d_{K_2}, \vec{\varphi}_2)$, the following inequality holds: $D_{match}\left(\check{\rho}_{(D,\vec{\Phi}_1),q}, \check{\rho}_{(D,\vec{\Phi}_2),q}\right) \leq \max\left\{\delta_H(K_1, K_2), \|\vec{\varphi}_1 - \vec{\varphi}_2\|_\infty\right\}$. ($\delta_H =$ Hausdorff distance)

An example (persistent homology in degree 0)

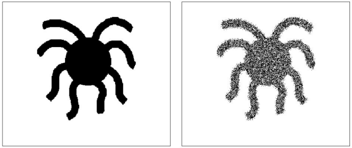


Figure: Two binary images of an octopus. The image on the right is noisy.

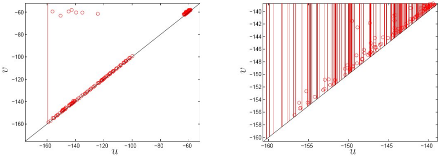


Figure: The persistence diagrams of persistent homology in degree 0, corresponding to the original octopus image and to the noisy octopus image, respectively. We can see that ordinary 1-dimensional persistent homology is not stable under perturbation of the set.

An example (continued)

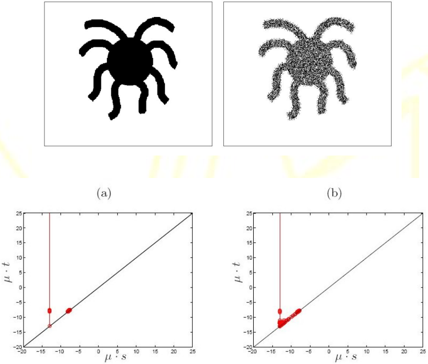


Figure: The persistent diagrams describing the rank invariants in degree 0 for the two images, restricted to a half-plane in the foliation and rescaled by $\mu = \min\{l_1, l_2\}$. Our approach allows us to get stability under perturbation of the set.

Retrieving information about $(K, \vec{\varphi}_{|K})$ from $(D, \vec{\Phi} = (d_k, \vec{\varphi}))$

BY OUR APPROACH, WE DO NOT LOSE INFORMATION ABOUT THE ORIGINAL SET AND THE ORIGINAL FILTERING FUNCTION:

Let K be a non-empty triangulable subset of a triangulable subspace D of $\mathbb{R}^n$.

Let $\vec{\varphi} : D \to \mathbb{R}^k$ be a continuous function.

Let us define $\vec{\Phi} : D \to \mathbb{R}^{k+1}$ by setting $\vec{\Phi} = (d_K, \vec{\varphi})$, where d_K is the distance from K.

Theorem

In each degree

$$\check{\rho}_{(K,\vec{\varphi}_{|K})}(\vec{u}, \vec{v}) = \lim_{\epsilon \to 0^+} \check{\rho}_{(D,\vec{\Phi})}\left((0, \vec{u}), (\epsilon, \vec{v})\right),$$

for every $\vec{u}, \vec{v} \in \mathbb{R}^k$ with $\vec{u} \prec \vec{v}$.

1. Multi-dimensional Persistent Homology
2. Comparing multi-dimensional Persistent Homology
3. Stability with respect to noisy functions and noisy sets
4. **Computing our distance**

Approximating the matching distance between rank invariants in the 2-dimensional case

We recall the definition of the matching distance in the case $k = 2$:

$$D_{match}\left(\check{\rho}_{(X,\vec{\varphi}),q},\check{\rho}_{(Y,\vec{\psi}),q}\right) = sup_{(\vec{l},\vec{b})\in L\times B}\mu(\vec{l})\cdot d_{match}\left(\check{\rho}_{\left(X,F^{\vec{\varphi}}_{(\vec{l},\vec{b})}\right)},\check{\rho}_{\left(Y,F^{\vec{\psi}}_{(\vec{l},\vec{b})}\right)}\right)$$

$$= sup_{(\vec{l},\vec{b})\in L\times B} d_{match}\left(\check{\rho}_{\left(X,\mu(\vec{l})\cdot F^{\vec{\varphi}}_{(\vec{l},\vec{b})}\right)},\check{\rho}_{\left(Y,\mu(\vec{l})\cdot F^{\vec{\psi}}_{(\vec{l},\vec{b})}\right)}\right) = sup_{(\vec{l},\vec{b})\in L\times B} d(\vec{l},\vec{b})$$

where $\mu(\vec{l}) = \min\{l_1, l_2\}$, $F^{\vec{\varphi}}_{(\vec{l},\vec{b})}(x) = \max\left\{\frac{\varphi_1(x)-b_1}{l_1}, \frac{\varphi_2(x)-b_2}{l_2}\right\}$,
$F^{\vec{\psi}}_{(\vec{l},\vec{b})}(x) = \max\left\{\frac{\psi_1(x)-b_1}{l_1}, \frac{\psi_2(x)-b_2}{l_2}\right\}$.

Approximating the matching distance between rank invariants in the 2-dimensional case

Our main result about the perturbation of the leaf in the foliation

The following statement holds:

Change of leaves and matching distance

Let us set

$$C = \max\{\|\vec{\varphi}\|_\infty, \|\vec{\psi}\|_\infty\}$$

and

$$d(\vec{l}, \vec{b}) = d_{match}\left(\check{\rho}_{\left(X, \mu \cdot F^{\vec{\varphi}}_{(\vec{l}, \vec{b})}\right)}, \check{\rho}_{\left(Y, \mu \cdot F^{\vec{\psi}}_{(\vec{l}, \vec{b})}\right)}\right).$$

Let us assume that $\|(\vec{l}, \vec{b}) - (\vec{l}', \vec{b}')\|_\infty \leq \epsilon$, with $\epsilon \leq \frac{1}{4}$. Then

$$\left| d(\vec{l}, \vec{b}) - d(\vec{l}', \vec{b}') \right| \leq \epsilon \cdot (32C + 2)$$

Approximating the matching distance between rank invariants in the 2-dimensional case

Let us simplify our notations

The strip $(0, 1) \times \mathbb{R}$

In order to simplify the study of the function $d(\vec{l}, \vec{b})$ (whose supremum is the multi-dimensional matching distance d_{match}), we observe that $(\vec{l}, \vec{b})$ can be identified by the pair (l_1, b_1) (since $l_2 = 1 - l_1$ and $b_2 = -b_1$).

In the following we shall speak of the value of $d(\vec{l}, \vec{b})$ at the point $(l_1, b_1) \in (0, 1) \times \mathbb{R}$: we shall mean the value of $d(\vec{l}, \vec{b})$ at the point $((l_1, l_2), (b_1, b_2))$.

Computing the 2-dimensional Matching Distance

Previous results open the way to the approximation of the matching distance between 2-dimensional rank invariants.
Indeed, if we take a finite grid of points G in $(0,1)\times\mathbb{R}$ in such the way that each point of $(0,1)\times\mathbb{R}$ has distance from G less than ϵ then the matching distance

$$D_{match}\left(\check{\rho}_{(X,\vec{\varphi}),q},\check{\rho}_{(Y,\vec{\psi}),q}\right)=\sup_{(l_1,b_1)\in(0,1)\times\mathbb{R}} d(\vec{l},\vec{b})$$

is approximated with an error less than $\epsilon\cdot(32C+2)$ by the pseudodistance

$$\widetilde{D}_{match}\left(\check{\rho}_{(X,\vec{\varphi}),q},\check{\rho}_{(Y,\vec{\psi}),q}\right)=\max_{(l_1,b_1)\in G} d(\vec{l},\vec{b})$$

where $C=\max\{\|\vec{\varphi}\|_\infty,\|\vec{\psi}\|_\infty\}$.

Figure A: The 2-dimensional Matching Distance in action (multi-dimensional persistent homology in degree 0)

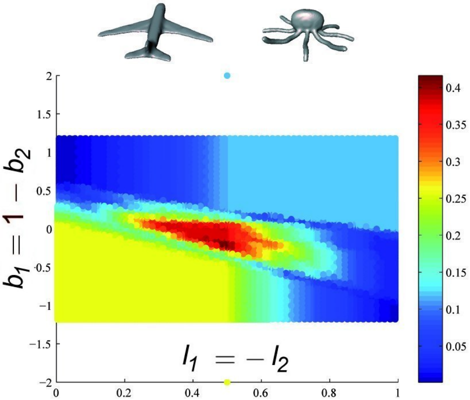


Figure: The function $d(\vec{l}, \vec{b})$ for an airplane and an octopus model, shown on top of the plot. The error tolerance is 5% of the constant C.

Figure B

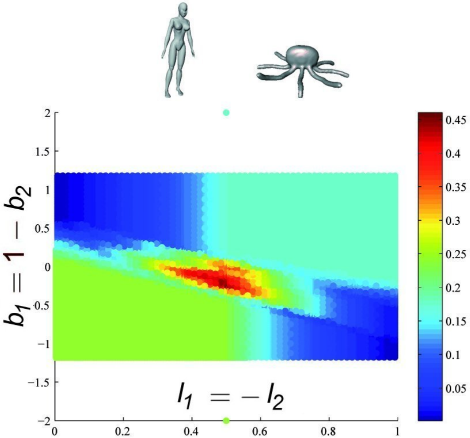


Figure: The function $d(\vec{l}, \vec{b})$ for a human and an octopus model, shown on top of the plot. The error tolerance is 5% of the constant C.

Figure C

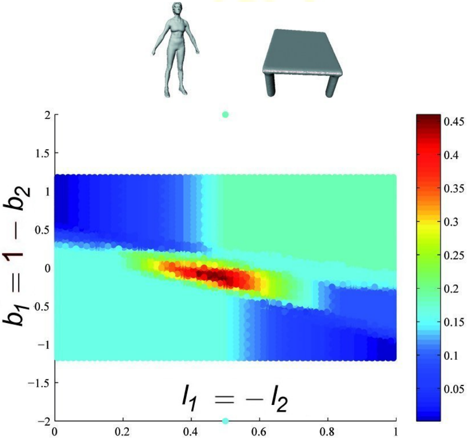


Figure: The function $d(\vec{l}, \vec{b})$ for a human and a table model, shown on top of the plot. The error tolerance is 5% of the constant C.

Figure D

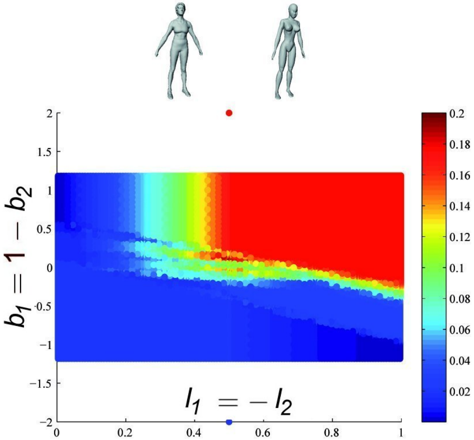


Figure: The function $d(\vec{l}, \vec{b})$ for two human models, shown on top of the plot. The error tolerance is 5% of the constant C.

Figure E

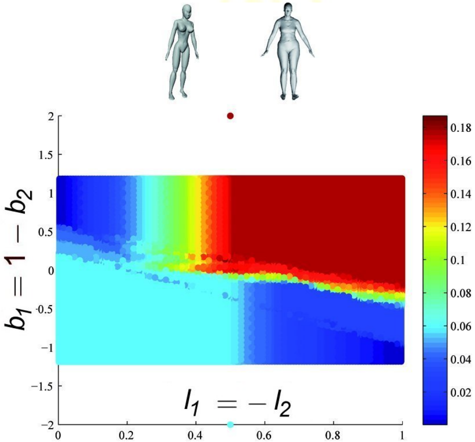


Figure: The function $d(\vec{l}, \vec{b})$ for two human models, shown on top of the plot. The error tolerance is 5% of the constant C.

Conclusions

THE FOLIATION METHOD ALLOWS US TO OBTAIN THE FOLLOWING RESULTS:

- A distance to compare multi-dimensional persistent homology is available.
- Multi-dimensional persistent homology is stable in every degree with respect to this distance, with reference both to perturbation of the filtering function $\vec{\varphi} : X \to \mathbb{R}^k$ (with respect to L_∞-norm) and to perturbation of the set X (with respect to Hausdorff distance).
- This distance can be approximated with arbitrary precision, by using an upper bound for the error associated with the change of the leaf.