Fields Institute, Toronto - November 11, 2011

Workshop on Computational Topology

Persistence based signatures for compact metric spaces

Frédéric Chazal

Joint work with Vin de Silva and Steve Oudot

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

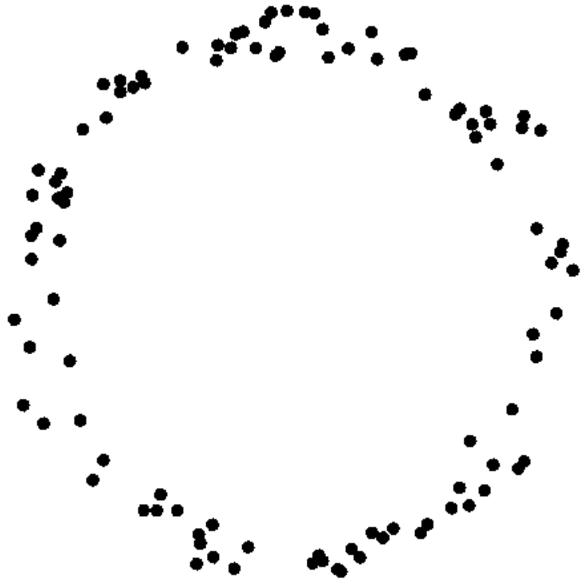
INRIA

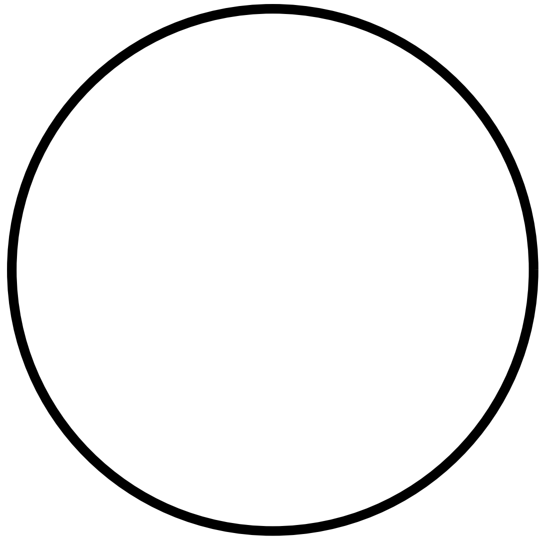
centre de recherche SACLAY - ÎLE-DE-FRANCE

GEOMETRICA

Introduction

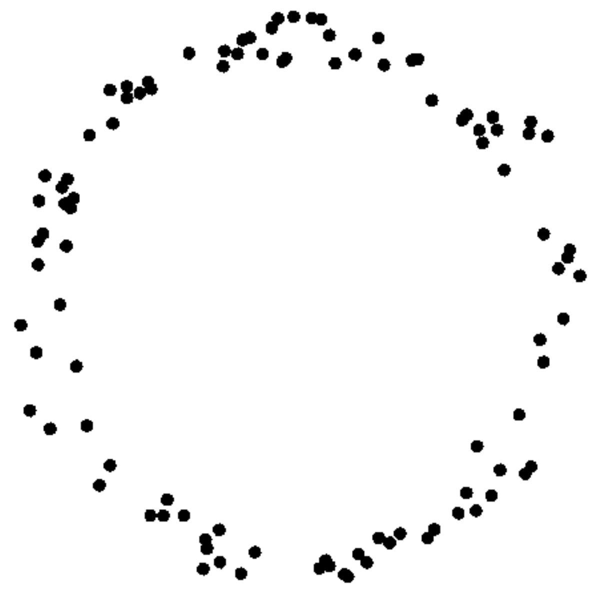
Problem: how to compare topological properties of close metric spaces (X, d_X) and (Y, d_Y)?

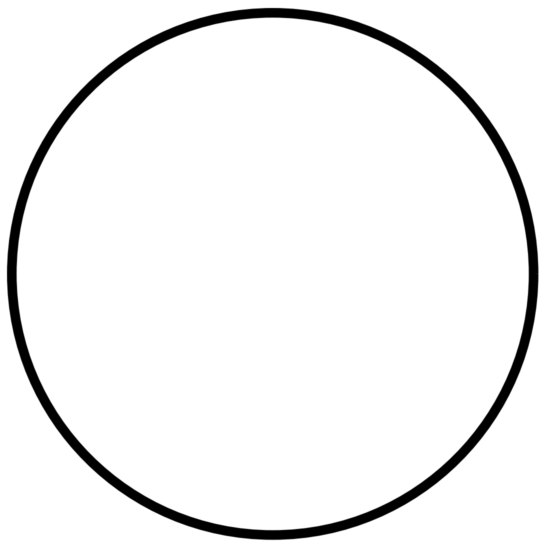




Introduction

Problem: how to compare topological properties of close metric spaces (X, d_X) and (Y, d_Y)?



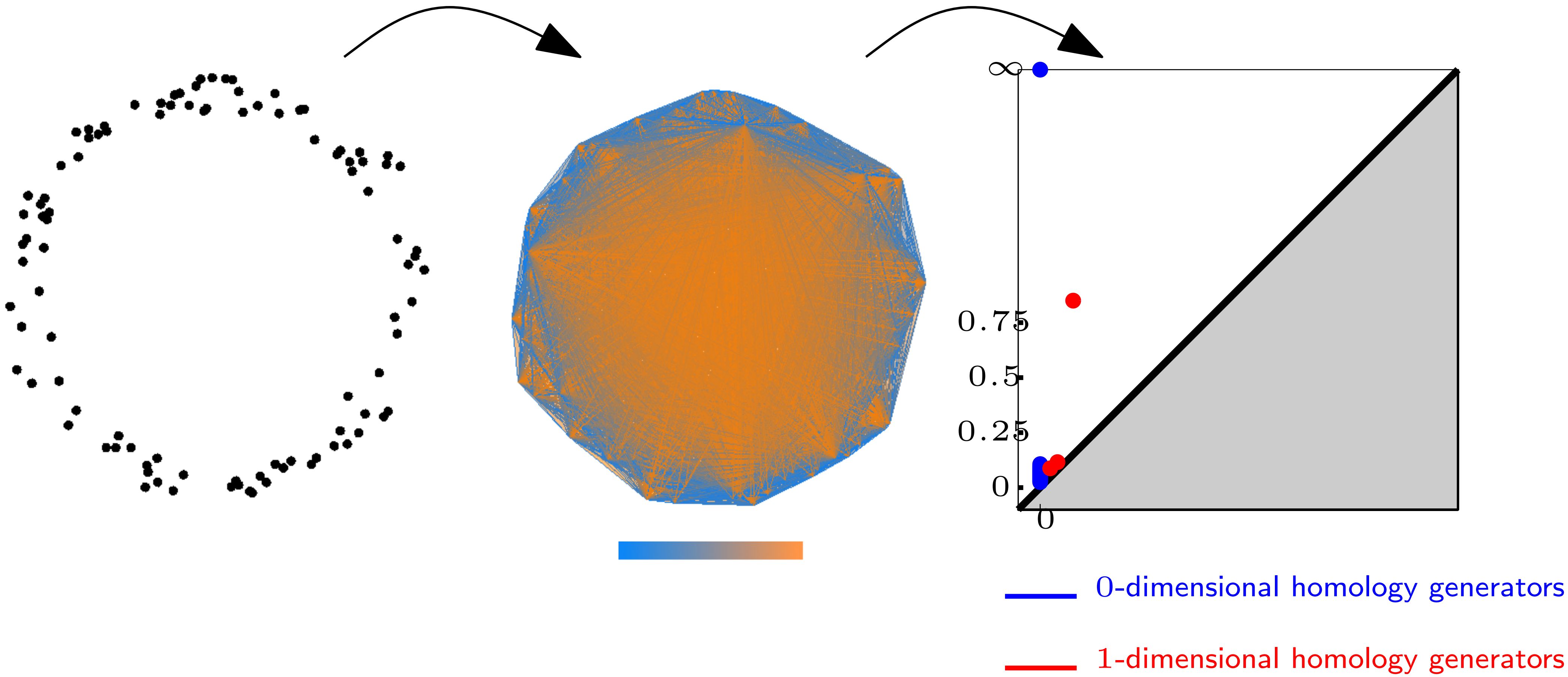


- Gromov-Hausdorff distance between (X, d_X) and (Y, d_Y)

For $\varepsilon \geq 0$, $C \subset X \times Y$ is an ε-correspondence if
i) $\forall x \in X$, $\exists y \in Y$ s. t. $(x, y) \in C$;
ii) $\forall y \in Y$, $\exists x \in X$ s. t. $(x, y) \in C$;
iii) $\forall (x, y), (x', y') \in C$, $|d_X(x, x') - d_Y(y, y')| \leq \varepsilon$.

$$d_{GH}(X, Y) = \frac{1}{2} \inf\{\varepsilon \geq 0 : \text{there exists an } \varepsilon\text{-correspondence between } X \text{and } Y\}$$

Introduction



- Scale dependent approach considering the Vietoris-Rips complexes built on X and Y.

  For $a \in \mathbb{R}$, $\mathrm{Rips}(X, a)$ is the simplicial complex with vertex set X defined by

$$[x_0, x_1, \cdots, x_k] \in \mathrm{Rips}(X, a) \Leftrightarrow d_X(x_i, x_j) \leq a, \quad \text{for all } i, j$$

- Compare the persistence of the filtrations $\mathbf{H}\mathrm{Rips}(X)$ and $\mathbf{H}\mathrm{Rips}(Y)$ where $\mathbf{H} = \mathbf{H}_p(-; \mathbb{F})$ is the simplicial homology functor in dimension p over a field $\mathbb{F}$.

Previous works

- J.-C. Hausmann 1995
- J. Latschev 2001
- D. Attali, A. Lieutier 2011

When (X, d_X) is a riemannian manifold (or a sufficiently regular subset of $\mathbb{R}^d$), Rips(X, a) is homotopy equivalent to X for small enough $a > 0$.

- F. C., D. Cohen-Steiner, L. Guibas, F. Mémoli, S. Oudot 2009

If X and Y are **finite** metric spaces, the bottleneck distance between the persistence diagrams of $\mathbf{H}\mathrm{Rips}(X)$ and $\mathbf{H}\mathrm{Rips}(Y)$ is a lower bound of $2d_{GH}(X, Y)$.

persistence diagrams used as shape signatures for classification

Tame persistent modules

Definition: A persistent module is a family $\mathcal{M} = (M^a, a \in \mathbb{R})$ of vector spaces together with linear maps $\xi_a^b : M^a \to M^b$ for all $a \leq b$ where $\xi_a^a = id_{M^a}$ and $\xi_a^c = \xi_b^c \circ \xi_a^b$ for all $a \leq b \leq c$.

Example: Given a metric space (X, d_X), the Vietoris-Rips filtration $\mathbf{H}\mathrm{Rips}(X)$ is a persistence module with linear maps the morphisms ξ_a^b induced at the homology level by the inclusion maps $\mathrm{Rips}(X, a) \hookrightarrow \mathrm{Rips}(X, b)$.

Tame persistent modules

Definition: A persistent module is a family $\mathcal{M} = (M^a, a \in \mathbb{R})$ of vector spaces together with linear maps $\xi_a^b : M^a \to M^b$ for all $a \leq b$ where $\xi_a^a = id_{M^a}$ and $\xi_a^c = \xi_b^c \circ \xi_a^b$ for all $a \leq b \leq c$.

Example: Given a metric space (X, d_X), the Vietoris-Rips filtration $\mathbf{H}\mathrm{Rips}(X)$ is a persistence module with linear maps the morphisms ξ_a^b induced at the homology level by the inclusion maps $\mathrm{Rips}(X, a) \hookrightarrow \mathrm{Rips}(X, b)$.

In the "classical" setting (finite dimentional spaces), the definition of persistence only involves the rank of the maps ξ_a^b, $a < b$.

Definition: A persistence module $\mathcal{M}$ is tame if for any $a < b$, ξ_a^b has a finite rank.

Theorem [CCGGO'09]: tame persistence modules have well-defined persistence diagrams.

Tame persistent modules

Definition: A persistent module is a family $\mathcal{M} = (M^a, a \in \mathbb{R})$ of vector spaces together with linear maps $\xi_a^b : M^a \to M^b$ for all $a \leq b$ where $\xi_a^a = id_{M^a}$ and $\xi_a^c = \xi_b^c \circ \xi_a^b$ for all $a \leq b \leq c$.

A map $\Phi : \mathcal{M} \to \mathcal{N}$ of degree t is a family of morphisms $\phi_a : M^a \to N^{a+t}$ for all a s.t. $\eta_{a+t}^{b+t} \circ \phi_a = \phi_b \circ \xi_a^b$.

$$\begin{tikzcd}
M^a \arrow[rr] \arrow[dr] & & M^b \arrow[dr] & \\
& N^{a+t} \arrow[rr] \arrow[ur, phantom, "\circlearrowright"] & & N^{b+t}
\end{tikzcd}$$

An ε-interleaving between $\mathcal{M}$ and $\mathcal{N}$ is specified by two maps $\Phi^\varepsilon : \mathcal{M} \to \mathcal{N}$ and $\Psi^\varepsilon : \mathcal{N} \to \mathcal{M}$ of degree ε s.t. $\Phi^\varepsilon \circ \Psi^\varepsilon$ and $\Psi^\varepsilon \circ \Phi^\varepsilon$ are the shifts of $\mathcal{N}$ and $\mathcal{M}$.

$$\begin{tikzcd}
\cdots \arrow[rr] & & M^a \arrow[rr, "\xi_a^{a+2\varepsilon}"] \arrow[dr, "\phi_a"'] & & M^{a+2\varepsilon} \arrow[rr] \arrow[dr] & & \cdots & \\
& \cdots \arrow[rr] \arrow[ur] & & N^{a+\varepsilon} \arrow[rr, "\eta_{a+\varepsilon}^{a+3\varepsilon}"'] \arrow[ur, "\psi_{a+\varepsilon}"'] & & N^{a+3\varepsilon} \arrow[rr] & & \cdots
\end{tikzcd}$$

Tame persistent modules

Definition: A persistent module is a family $\mathcal{M} = (M^a, a \in \mathbb{R})$ of vector spaces together with linear maps $\xi_a^b : M^a \to M^b$ for all $a \leq b$ where $\xi_a^a = id_{M^a}$ and $\xi_a^c = \xi_b^c \circ \xi_a^b$ for all $a \leq b \leq c$.

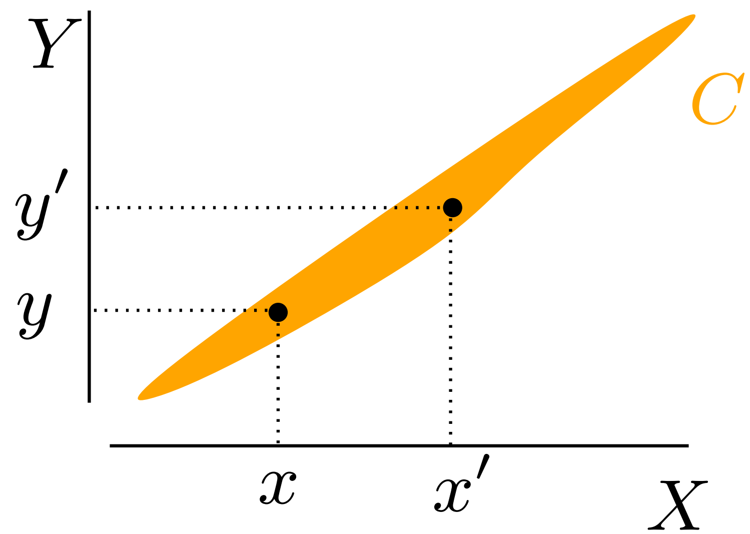
A persistence module $\mathcal{M}$ is tame if for any $a < b$, ξ_a^b has a finite rank.

Theorem [CCGGO'09]: If $\mathcal{M}$ and $\mathcal{N}$ are tame and ε-interleaved for some $\varepsilon \geq 0$ then

$$d_B(\mathrm{Dgm}(\mathcal{M}), \mathrm{Dgm}(\mathcal{N})) \leq \varepsilon$$

Correspondences and interleaving

Proposition: Let X, Y be metric spaces and let $C \subset X \times Y$ be an ε-correspondence between X and Y. Then C induces a canonical ε-interleaving between $\mathbf{H}\mathrm{Rips}(X)$ and $\mathbf{H}\mathrm{Rips}(Y)$.

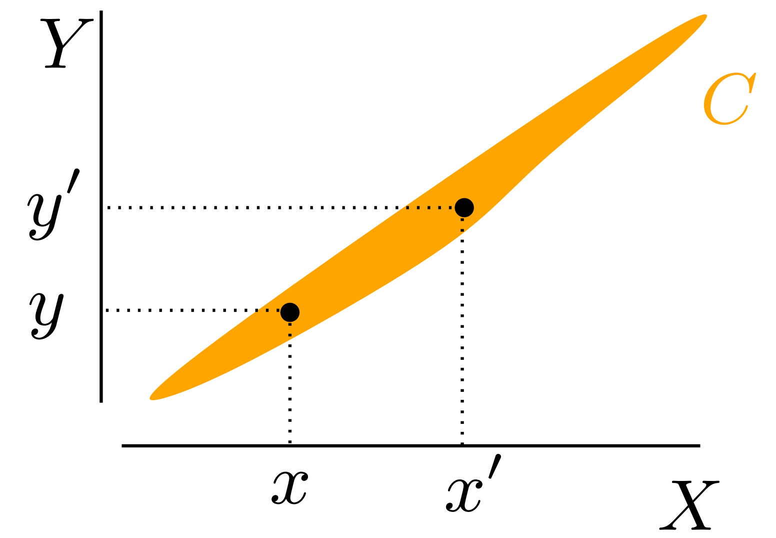


$$|d_X(x,x') - d_Y(y,y')| < \varepsilon$$

Correspondences and interleaving

Proposition: Let X, Y be metric spaces and let $C \subset X \times Y$ be an ε-correspondence between X and Y. Then C induces a canonical ε-interleaving between $\mathbf{H}\mathrm{Rips}(X)$ and $\mathbf{H}\mathrm{Rips}(Y)$.

Lemma: Let $f : X \to Y$ be compatible with C in the sense that $(x, f(x)) \in C$ for all $x \in X$. Then f induces a simplicial map $\overline{f} : \mathrm{Rips}(X, a) \to \mathrm{Rips}(Y, b)$ whenever $b \geq a + \varepsilon$. Moreover, if $g : X \to Y$ is compatible with C, $\overline{f}, \overline{g}$ are contiguous maps.



$$|d_X(x, x') - d_Y(y, y')| < \varepsilon$$

Correspondences and interleaving

Proposition: Let X, Y be metric spaces and let $C \subset X \times Y$ be an ε-correspondence between X and Y. Then C induces a canonical ε-interleaving between $\mathbf{H}\mathrm{Rips}(X)$ and $\mathbf{H}\mathrm{Rips}(Y)$.

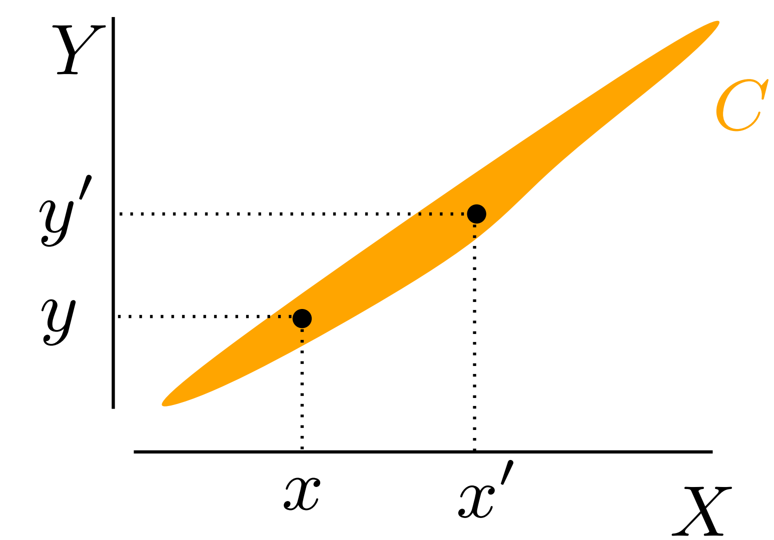
Lemma: Let $f : X \to Y$ be compatible with C in the sense that $(x, f(x)) \in C$ for all $x \in X$. Then f induces a simplicial map $\overline{f} : \mathrm{Rips}(X, a) \to \mathrm{Rips}(Y, b)$ whenever $b \geq a + \varepsilon$. Moreover, if $g : X \to Y$ is compatible with C, $\overline{f}, \overline{g}$ are contiguous maps.

Proof of lemma:

1. if $\sigma = [x_0, \cdots x_k] \in \mathrm{Rips}(X, a)$ then $\overline{f}(\sigma) = [f(x_0), \cdots f(x_k)] \in \mathrm{Rips}(Y, b)$

2. $\overline{f}(\sigma)$ and $\overline{g}(\sigma)$ are contained in a simplex of $\mathrm{Rips}(Y, b)$:

$$d_Y(f(x_i), f(x_j)) \leq d_X(x_i, x_j) + \varepsilon \leq a + \varepsilon \leq b$$
$$d_Y(g(x_i), g(x_j)) \leq d_X(x_i, x_j) + \varepsilon \leq a + \varepsilon \leq b$$
$$d_Y(f(x_i), g(x_j)) \leq d_X(x_i, x_j) + \varepsilon \leq a + \varepsilon \leq b$$



$$|d_X(x, x') - d_Y(y, y')| < \varepsilon$$

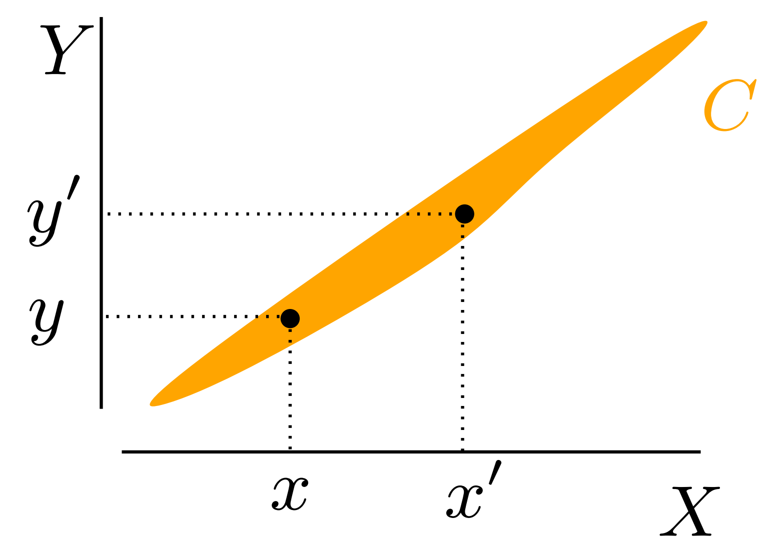
Correspondences and interleaving

Proposition: Let X, Y be metric spaces and let $C \subset X \times Y$ be an ε-correspondence between X and Y. Then C induces a canonical ε-interleaving between $\mathbf{H}\mathrm{Rips}(X)$ and $\mathbf{H}\mathrm{Rips}(Y)$.

Lemma: Let $f : X \to Y$ be compatible with C in the sense that $(x, f(x)) \in C$ for all $x \in X$. Then f induces a simplicial map $\overline{f} : \mathrm{Rips}(X, a) \to \mathrm{Rips}(Y, b)$ whenever $b \geq a + \varepsilon$. Moreover, if $g : X \to Y$ is compatible with C, $\overline{f}, \overline{g}$ are contiguous maps.

Proof of the proposition:

- Let $f : X \to Y$ and $g : Y \to X$ be compatible.
- $C^T C = \{(x, x') \in X \times X | \exists y \in Y \text{ s.t. } (x, y) \in C, (x', y) \in C\}$ is a 2ε-correspondence and $h = g \circ f$ and id_X are compatible functions $X \to X$.
- It follows from the lemma that h and id_X are contiguous, so they induce the same maps at homology level.



$|d_X(x, x') - d_Y(y, y')| < \varepsilon$

Tameness of the Rips filtration

A metric space (X, d_X) is totally bounded (or precompact) if for any $\varepsilon > 0$ there exists a finite subset $F_\varepsilon \subset X$ such that $d_H(X, F_\varepsilon) < \varepsilon$ (i.e. $\forall x \in X, \exists p \in F_\varepsilon$ s.t. $d_X(x, p) < \varepsilon$).

Theorem: Let X be a totally bounded metric space. Then $\mathbf{H}\mathrm{Rips}(X)$ is tame.
$\Rightarrow \mathrm{Dgm}(\mathbf{H}\mathrm{Rips}(X))$ is well-defined!

Tameness of the Rips filtration

A metric space (X, d_X) is totally bounded (or precompact) if for any $\varepsilon > 0$ there exists a finite subset $F_\varepsilon \subset X$ such that $d_H(X, F_\varepsilon) < \varepsilon$ (i.e. $\forall x \in X, \exists p \in F_\varepsilon$ s.t. $d_X(x,p) < \varepsilon$).

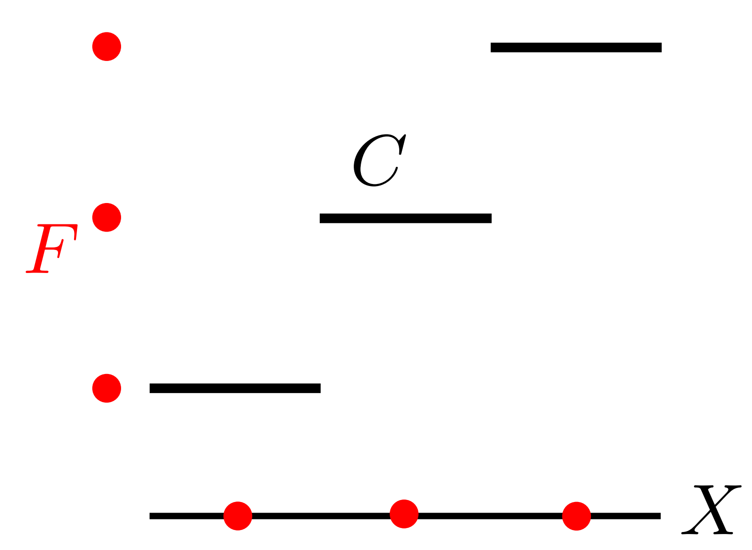
Theorem: Let X be a totally bounded metric space. Then $\mathbf{H}\mathrm{Rips}(X)$ is tame.

$\Rightarrow \mathrm{Dgm}(\mathbf{H}\mathrm{Rips}(X))$ is well-defined!

Proof: show that $I_a^b : \mathbf{H}\mathrm{Rips}(X,a) \to \mathbf{H}\mathrm{Rips}(X,b)$ has finite rank whenever $a < b$.

Let $\varepsilon = (b-a)/2$ and let $F \subset X$ be finite s. t. $d_H(X,F) \leq \varepsilon/2$.

Then $C = \{(x,f) \in X \times F | d(x,f) \leq \varepsilon\}$ is an ε-correspondence.



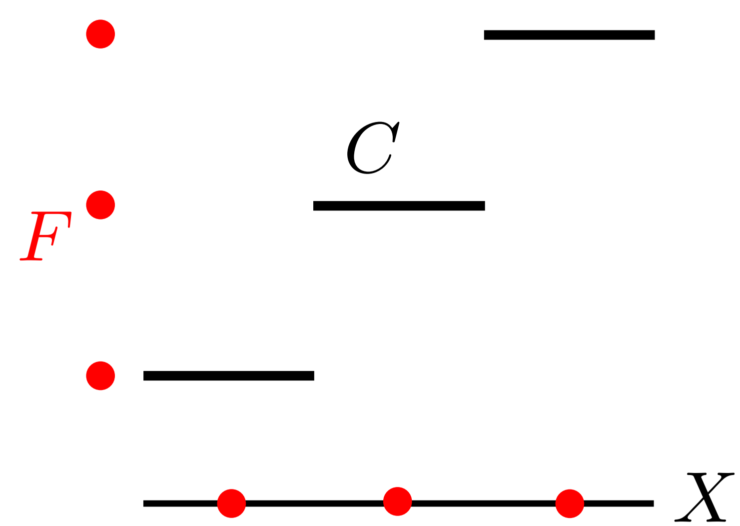
Tameness of the Rips filtration

A metric space (X, d_X) is totally bounded (or precompact) if for any $\varepsilon > 0$ there exists a finite subset $F_\varepsilon \subset X$ such that $d_H(X, F_\varepsilon) < \varepsilon$ (i.e. $\forall x \in X, \exists p \in F_\varepsilon$ s.t. $d_X(x,p) < \varepsilon$).

Theorem: Let X be a totally bounded metric space. Then $\mathbf{H}\mathrm{Rips}(X)$ is tame.

$\Rightarrow \mathrm{Dgm}(\mathbf{H}\mathrm{Rips}(X))$ is well-defined!

Proof: show that $I_a^b : \mathbf{H}\mathrm{Rips}(X,a) \to \mathbf{H}\mathrm{Rips}(X,b)$ has finite rank whenever $a < b$.



Let $\varepsilon = (b-a)/2$ and let $F \subset X$ be finite s. t. $d_H(X,F) \leq \varepsilon/2$.

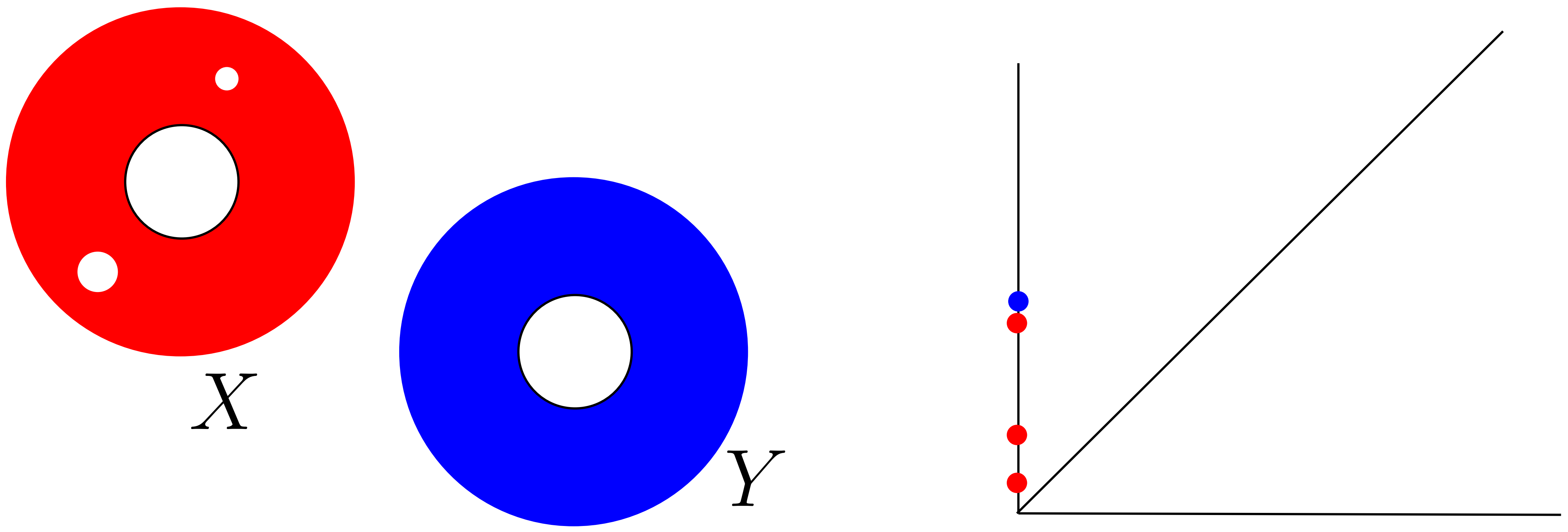
Then $C = \{(x,f) \in X \times F | d(x,f) \leq \varepsilon\}$ is an ε-correspondence.

Using the interleaving map, I_a^b factorizes as

$$\mathbf{H}\mathrm{Rips}(X,a) \to \mathbf{H}\mathrm{Rips}(F, a+\varepsilon) \to \mathbf{H}\mathrm{Rips}(X, a+2\varepsilon) = \mathbf{H}\mathrm{Rips}(X,b)$$

finite dimensional

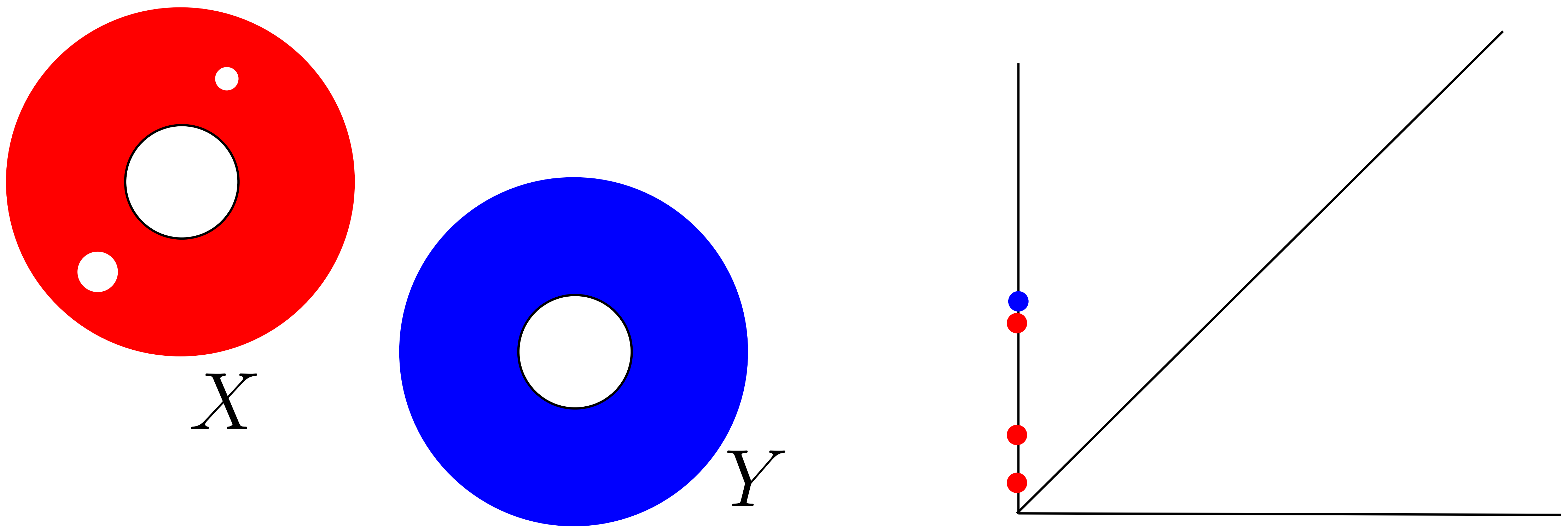
Stability of persistence diagrams



Theorem: Let X, Y be totally bounded metric spaces. Then

$$\mathbf{d}_b(\mathrm{Dgm}(\mathbf{H}\mathrm{Rips}(X)), \mathrm{Dgm}(\mathbf{H}\mathrm{Rips}(Y))) \leq 2\mathbf{d}_{GH}(X, Y)$$

Stability of persistence diagrams



Theorem: Let X, Y be totally bounded metric spaces. Then

$$\mathbf{d}_b(\mathrm{Dgm}(\mathbf{H}\mathrm{Rips}(X)), \mathrm{Dgm}(\mathbf{H}\mathrm{Rips}(Y))) \leq 2\mathbf{d}_{GH}(X, Y)$$

Proof:

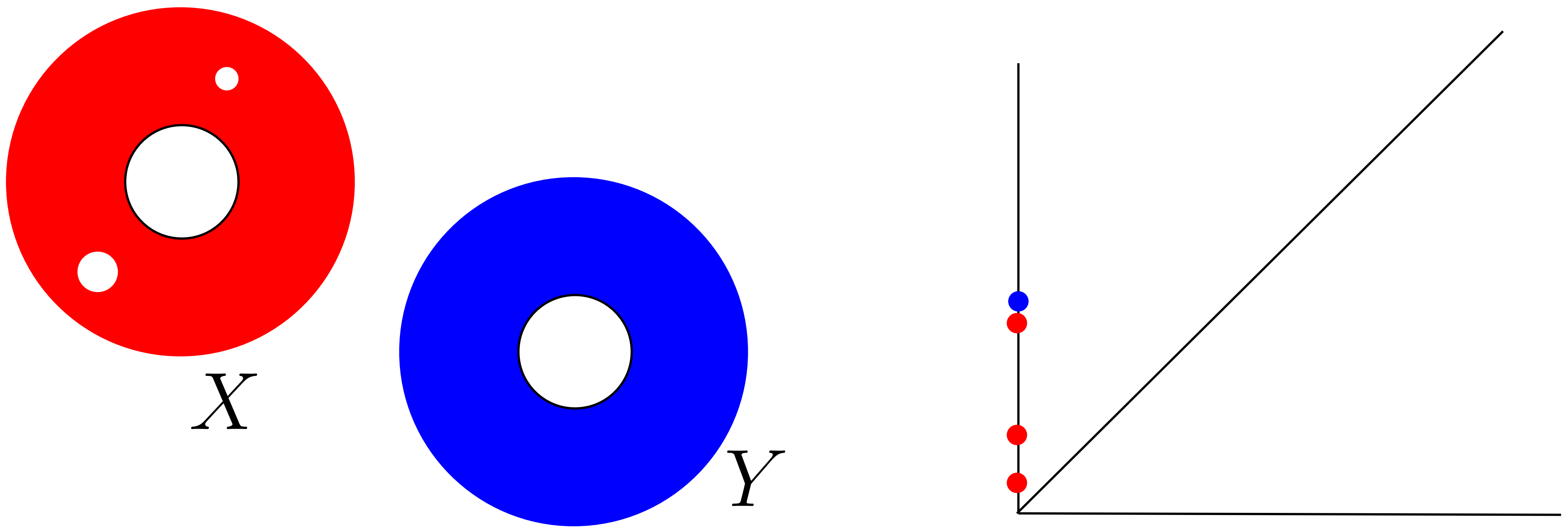
For any $\varepsilon > 2\mathbf{d}_{GH}(X, Y)$ there exists an ε-correspondence C between X and Y.

C induces a canonical ε-interleaving between $\mathbf{H}\mathrm{Rips}(X)$ and $\mathbf{H}\mathrm{Rips}(Y)$ that are tame modules.

It follows that

$$\mathbf{d}_b(\mathrm{Dgm}(\mathbf{H}\mathrm{Rips}(X)), \mathrm{Dgm}(\mathbf{H}\mathrm{Rips}(Y))) \leq \varepsilon$$

Stability of persistence diagrams



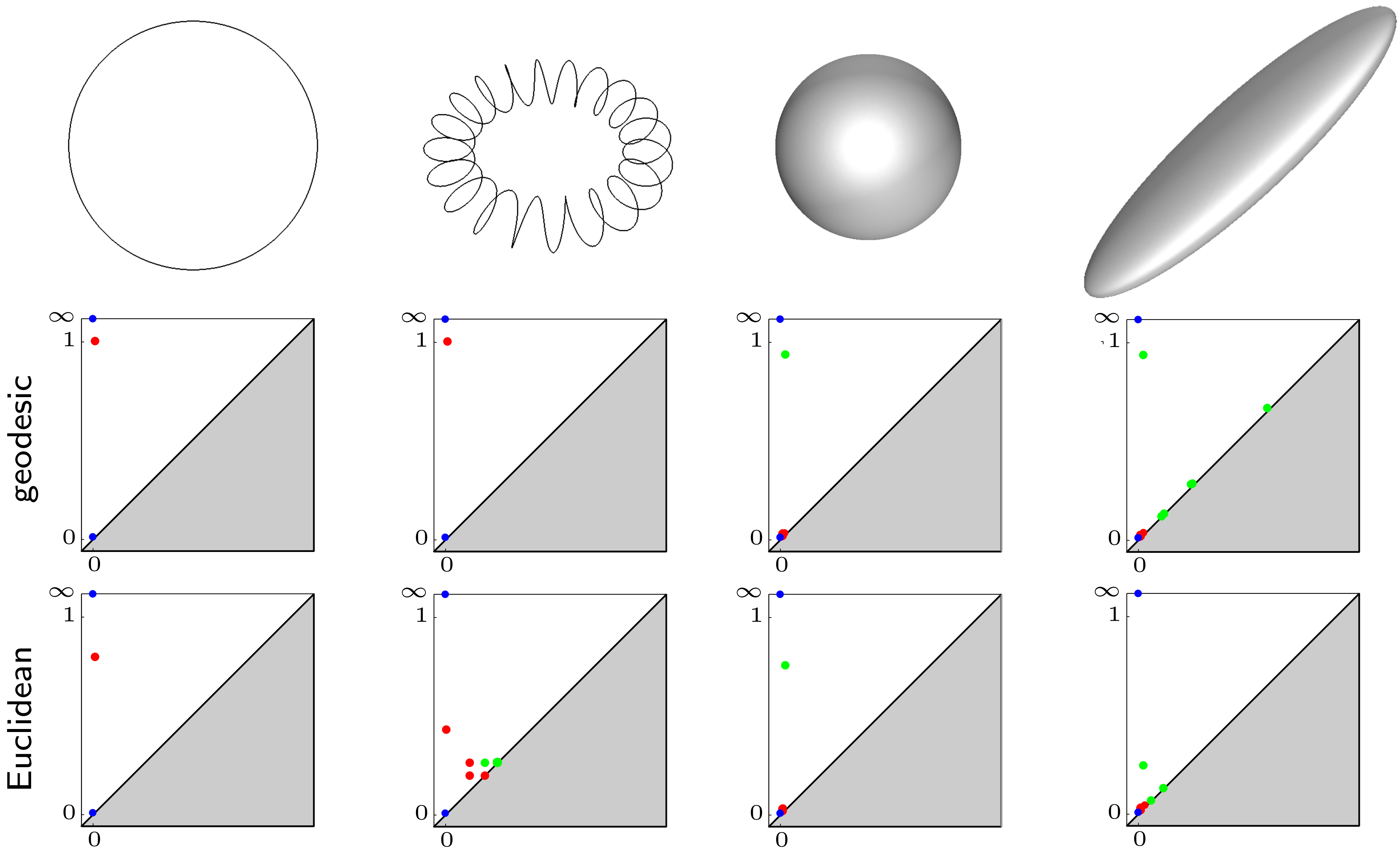
Theorem: Let X, Y be totally bounded metric spaces. Then

$$\mathbf{d}_b(\mathrm{Dgm}(\mathbf{H}\mathrm{Rips}(X)), \mathrm{Dgm}(\mathbf{H}\mathrm{Rips}(Y))) \leq 2\mathbf{d}_{GH}(X, Y)$$

- Robust multiscale signature for totally bounded spaces.
- A (topological) lower bound of the Gromov-Hausdorff distance.
- Same result with Cech filtrations.
- Similar results for pairs (X, f) where $f : X \to \mathbb{R}$ is a Lipschitz function.
- No need of the triangle inequality! $\to$ Also works for "dissimilarity measures"

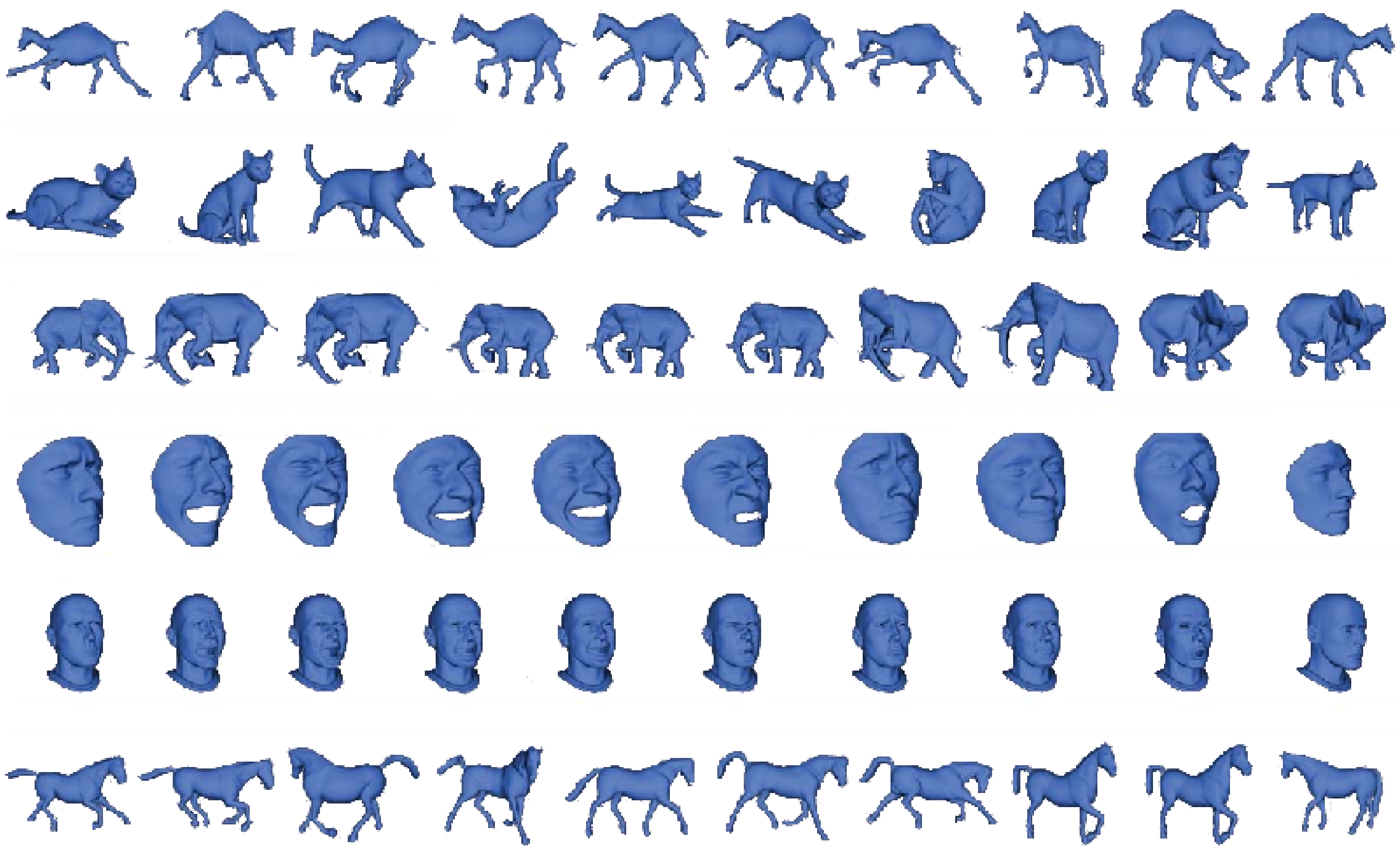
Persistence-based signatures

Signatures of some elementary shapes (approximated from finite samples):



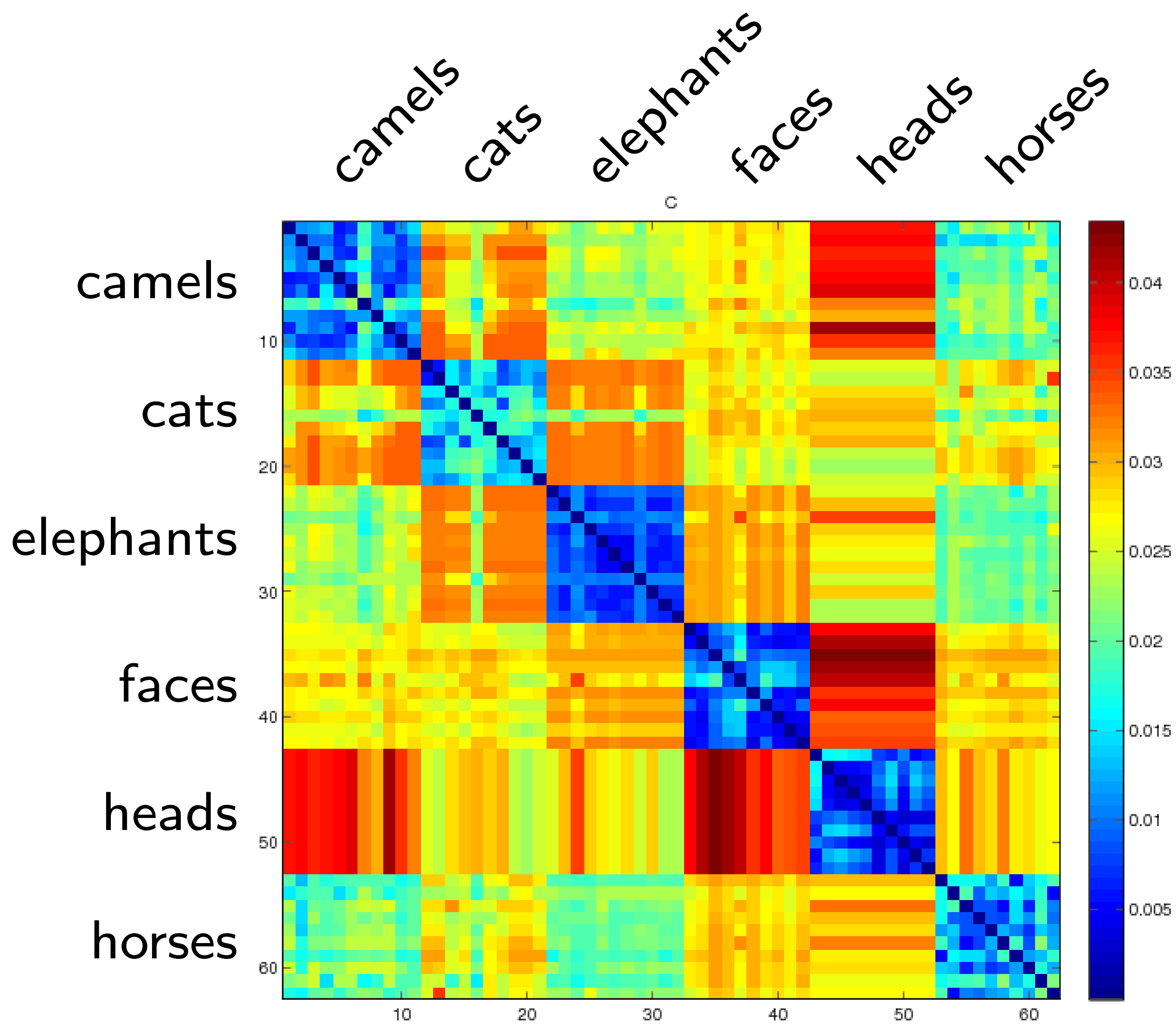
Persistence-based signatures

Experimental results:



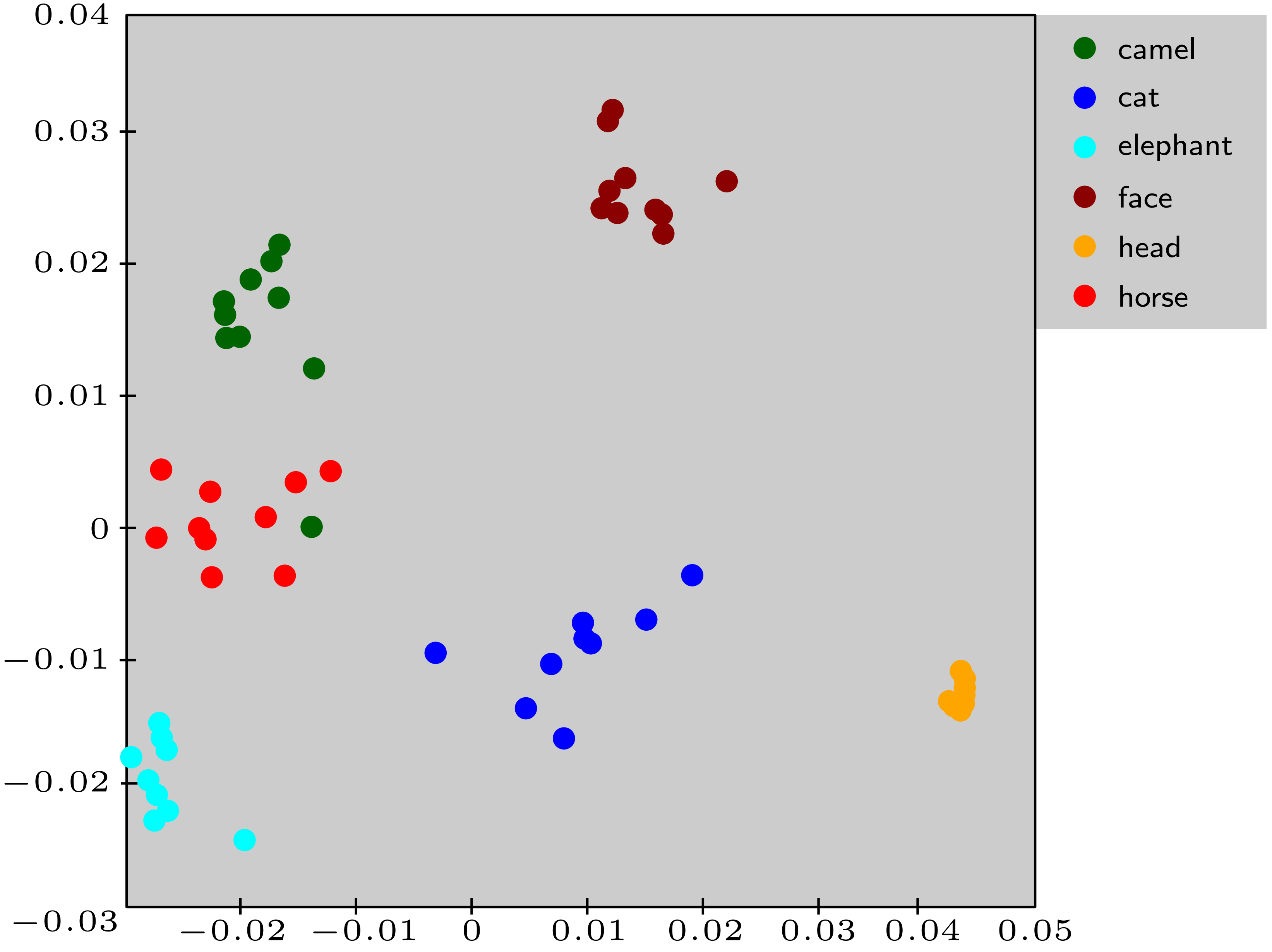
Persistence-based signatures

Experimental results:



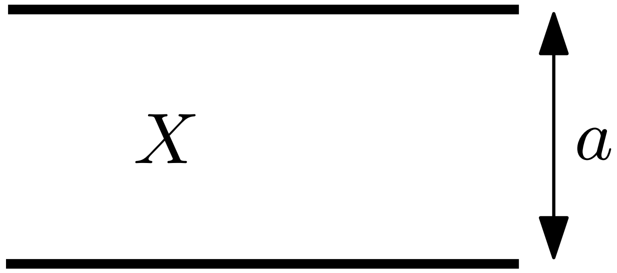
Persistence-based signatures

Experimental results:



Homework (remarks and open questions)

- Even when X is compact, $\mathbf{H}_p(\mathrm{Rips}(X,a))$, $p \geq 1$, might be infinite dimensional for some value of a:

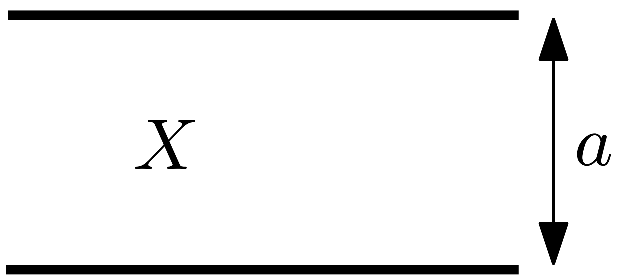


It is also possible to build such an example with the "open" Rips complex:

$$[x_0, x_1, \cdots, x_k] \in \mathrm{Rips}(X,a) \Leftrightarrow d_X(x_i, x_j) < a, \quad \text{for all } i,j$$

Homework (remarks and open questions)

- Even when X is compact, $\mathbf{H}_p(\mathrm{Rips}(X,a))$, $p \geq 1$, might be infinite dimensional for some value of a:



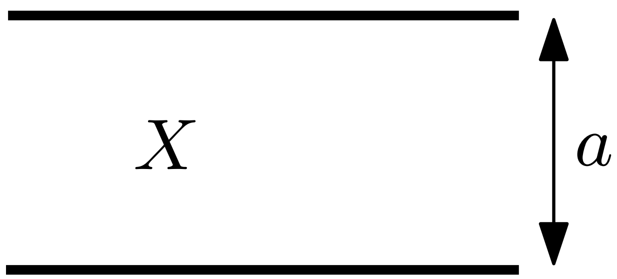
It is also possible to build such an example with the "open" Rips complex:

$$[x_0, x_1, \cdots, x_k] \in \mathrm{Rips}(X,a) \Leftrightarrow d_X(x_i, x_j) < a, \quad \text{for all } i, j$$

- If X is compact, then $\dim \mathbf{H}_1(\mathrm{Cech}(X,a)) < +\infty$ for all a ([Smale-Smale, C.-de Silva]).

Homework (remarks and open questions)

- Even when X is compact, $\mathbf{H}_p(\mathrm{Rips}(X, a))$, $p \geq 1$, might be infinite dimensional for some value of a:



It is also possible to build such an example with the "open" Rips complex:

$$[x_0, x_1, \cdots, x_k] \in \mathrm{Rips}(X, a) \Leftrightarrow d_X(x_i, x_j) < a, \quad \text{for all } i, j$$

- If X is compact, then $\dim \mathbf{H}_1(\mathrm{Cech}(X, a)) < +\infty$ for all a ([Smale-Smale, C.-de Silva]).

- If X is geodesic, then $\dim \mathbf{H}_1(\mathrm{Rips}(X, a)) < +\infty$ for all $a > 0$ and $\mathrm{Dgm}(\mathbf{H}_1 \mathrm{Rips}(X))$ is contained in the vertical line $x = 0$.

- Open problem (probably hard): for a given compact set X, how "big" can the set S of values a such that $\dim \mathbf{H}_p(\mathrm{Rips}(X, a)) = +\infty$ be?

Thank You