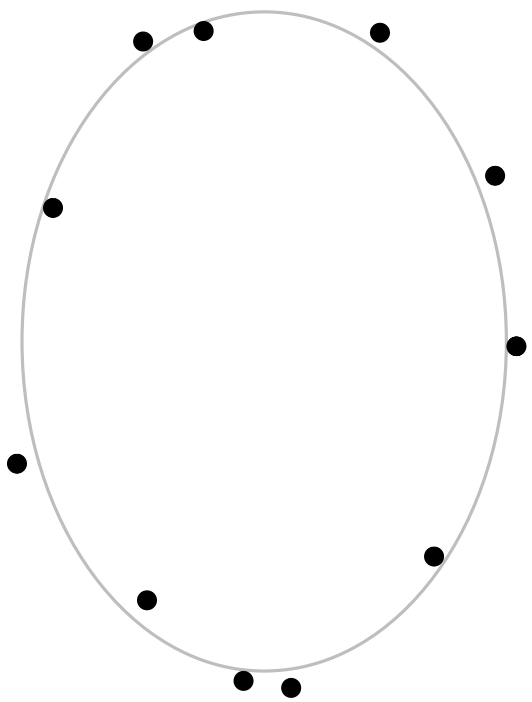
Université du Luxembourg
March 12 - 16, 2018

Topological Descriptors for Geometric Data

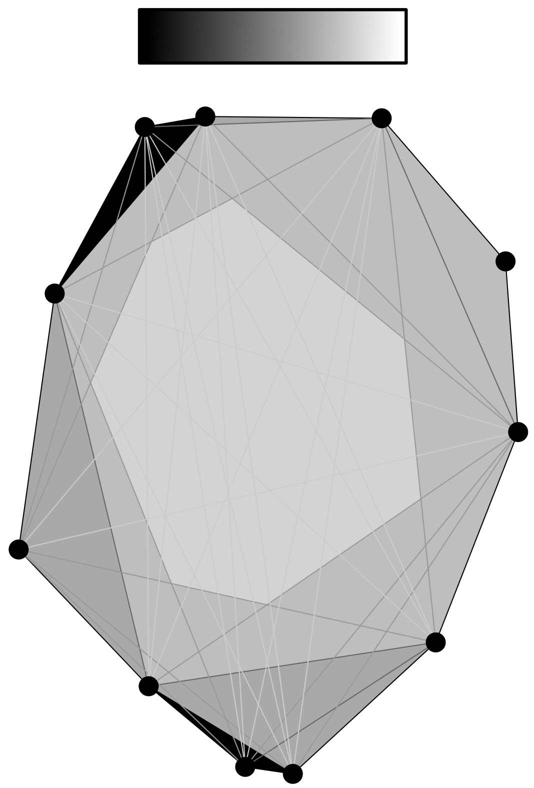
Reminder: the TDA pipeline



point cloud P

proximity rule

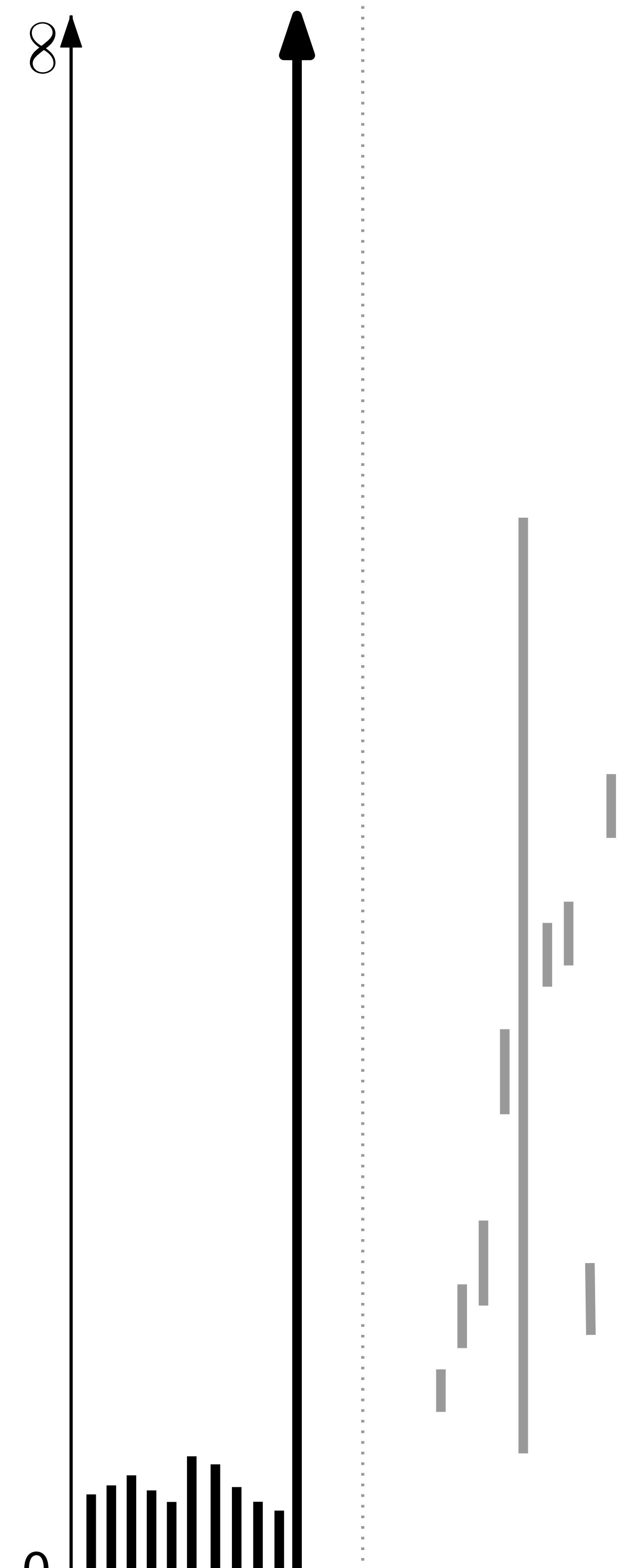
filter $P \to \mathbb{R}$



filtration $\mathcal{F}$

homology

persistence



barcode (descriptor)

The 5 pillars of the theory (persistence theory):

- decomposition theorems (existence of barcodes)
- algorithms (computation of barcodes)
- stability theorems (barcodes as stable descriptors)
- statistical frameworks for barcodes
- vectorizations and kernels on barcodes for learning

Outline

1. Descriptors and stability

2. Vectorizations and kernels

3. Statistics

4. Discrimination power

Outline

1. Descriptors and stability

2. Vectorizations and kernels

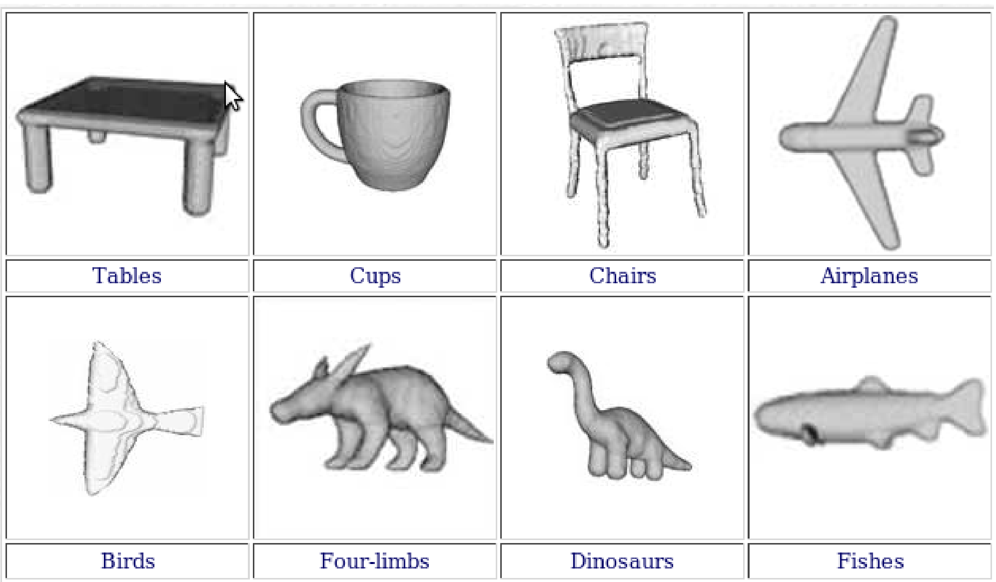
3. Statistics

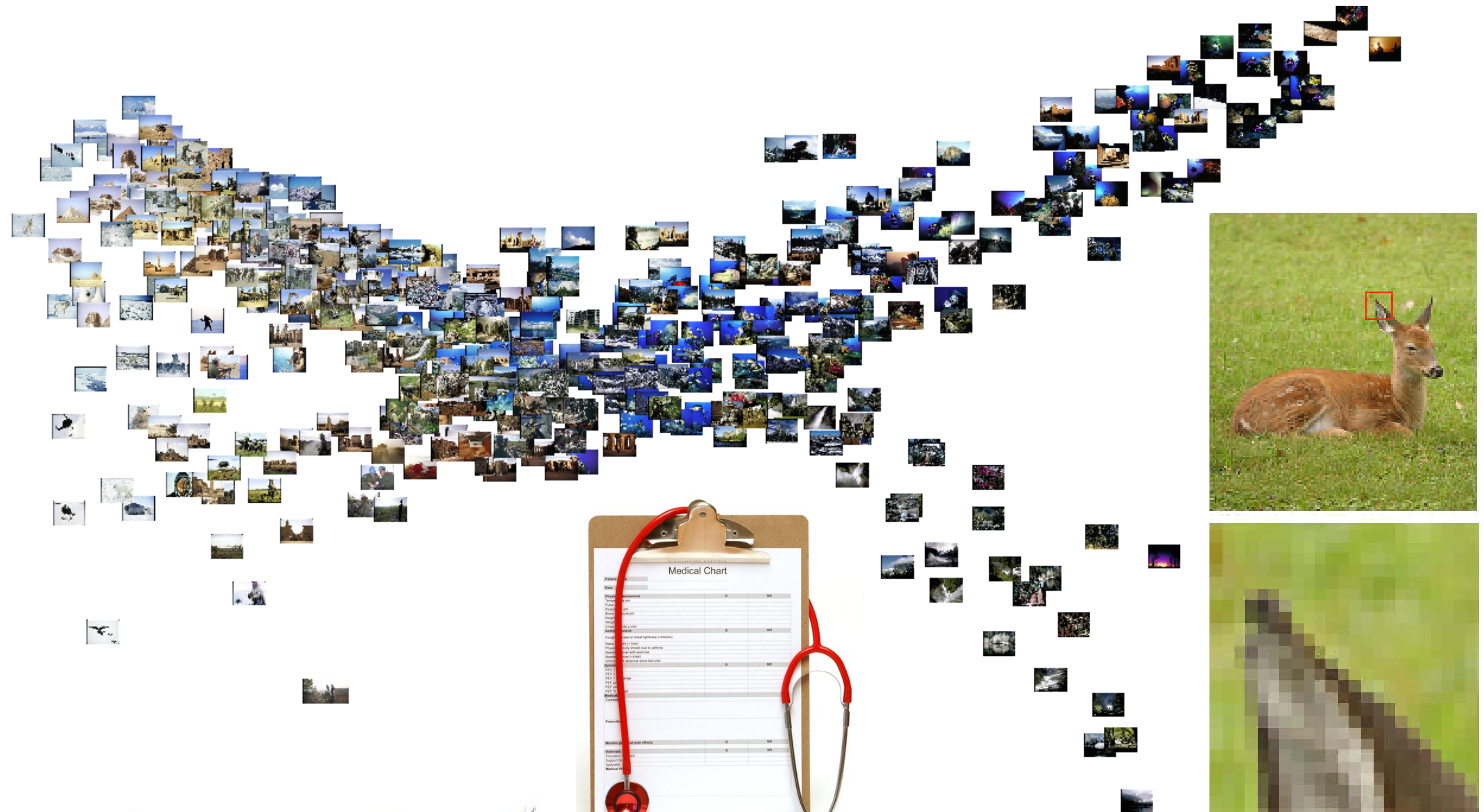
4. Discrimination power

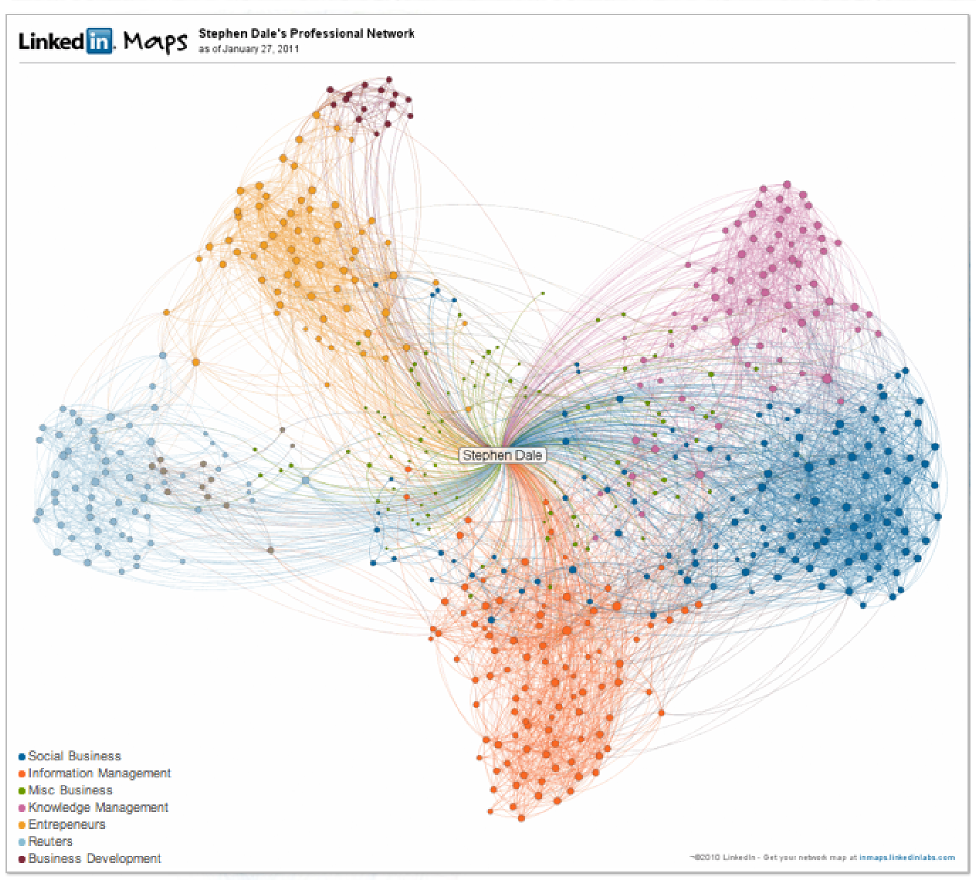
Geometric Data

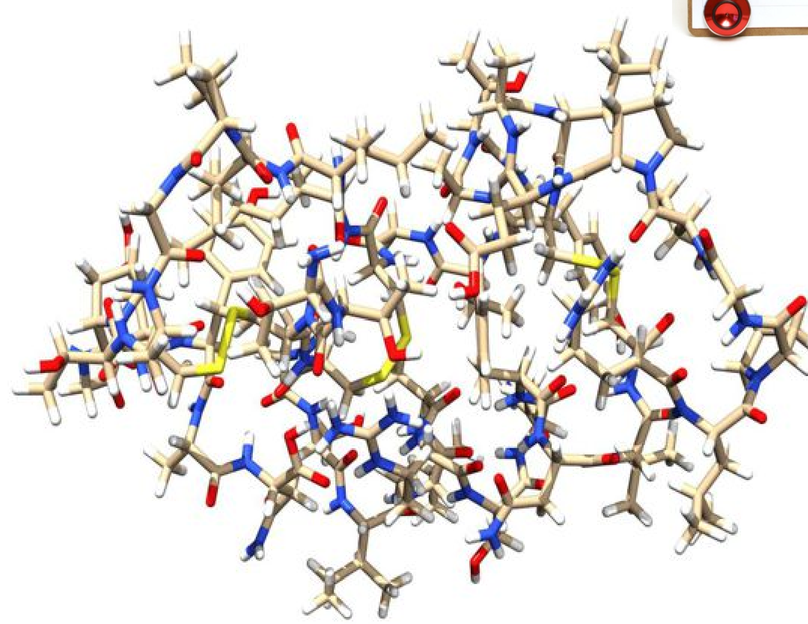
Input: point cloud equipped with a **metric** or **(dis-)similarity measure**

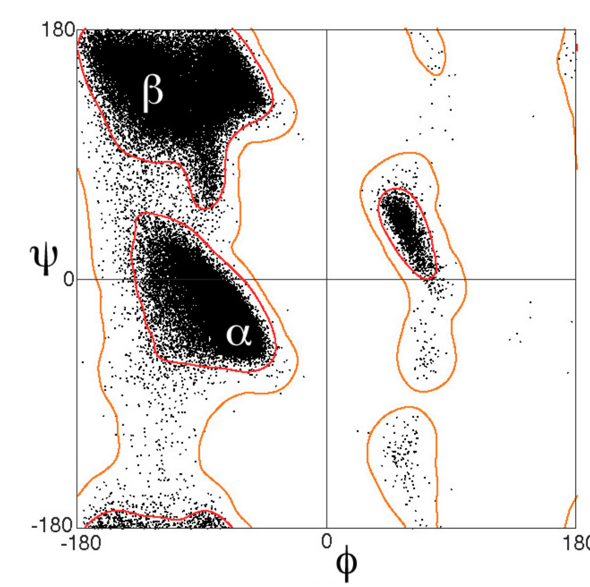
data point $\equiv$ image/patch, geometric shape, protein conformation, patient, LinkedIn user...







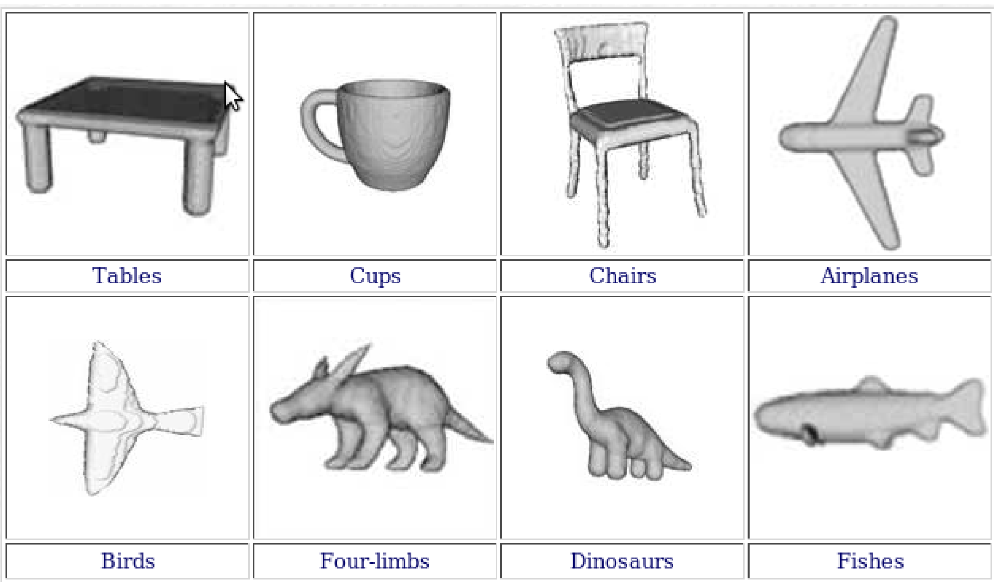


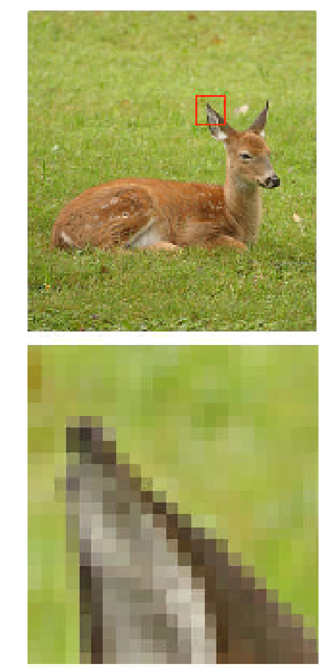


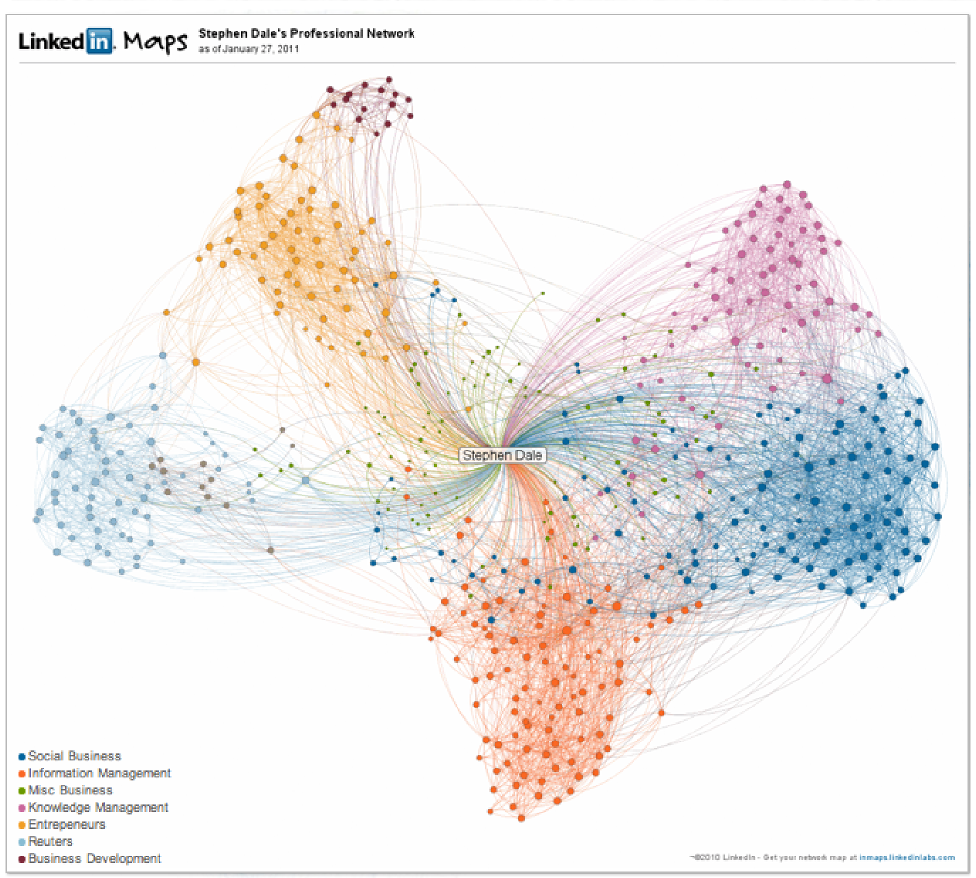
Geometric Data

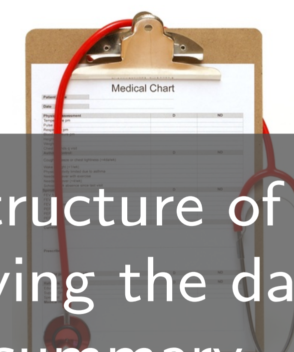
Input: point cloud equipped with a **metric** or **(dis-)similarity measure**

data point $\equiv$ image/patch, geometric shape, protein conformation, patient, LinkedIn user...

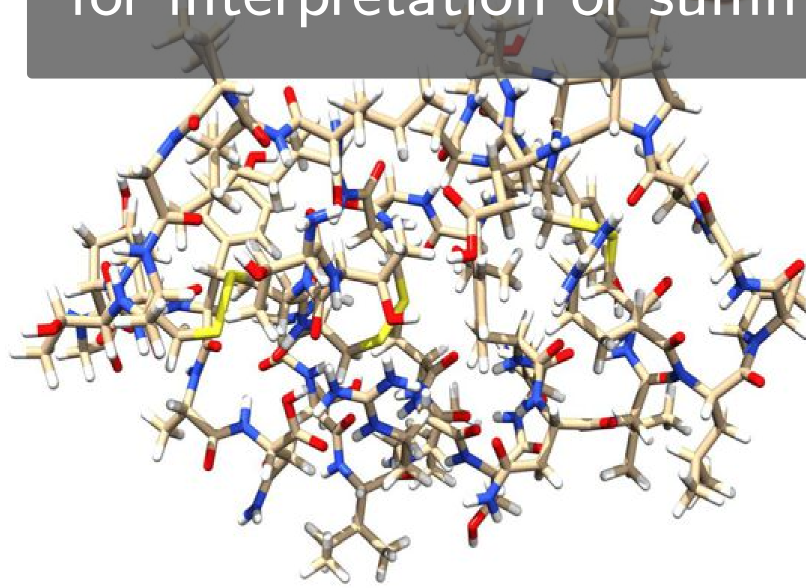


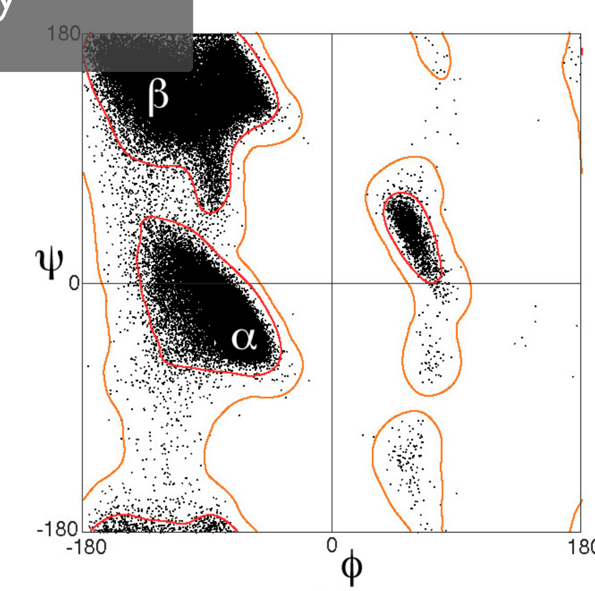






Goal: describe the structure of the geometry underlying the data, for interpretation or summary



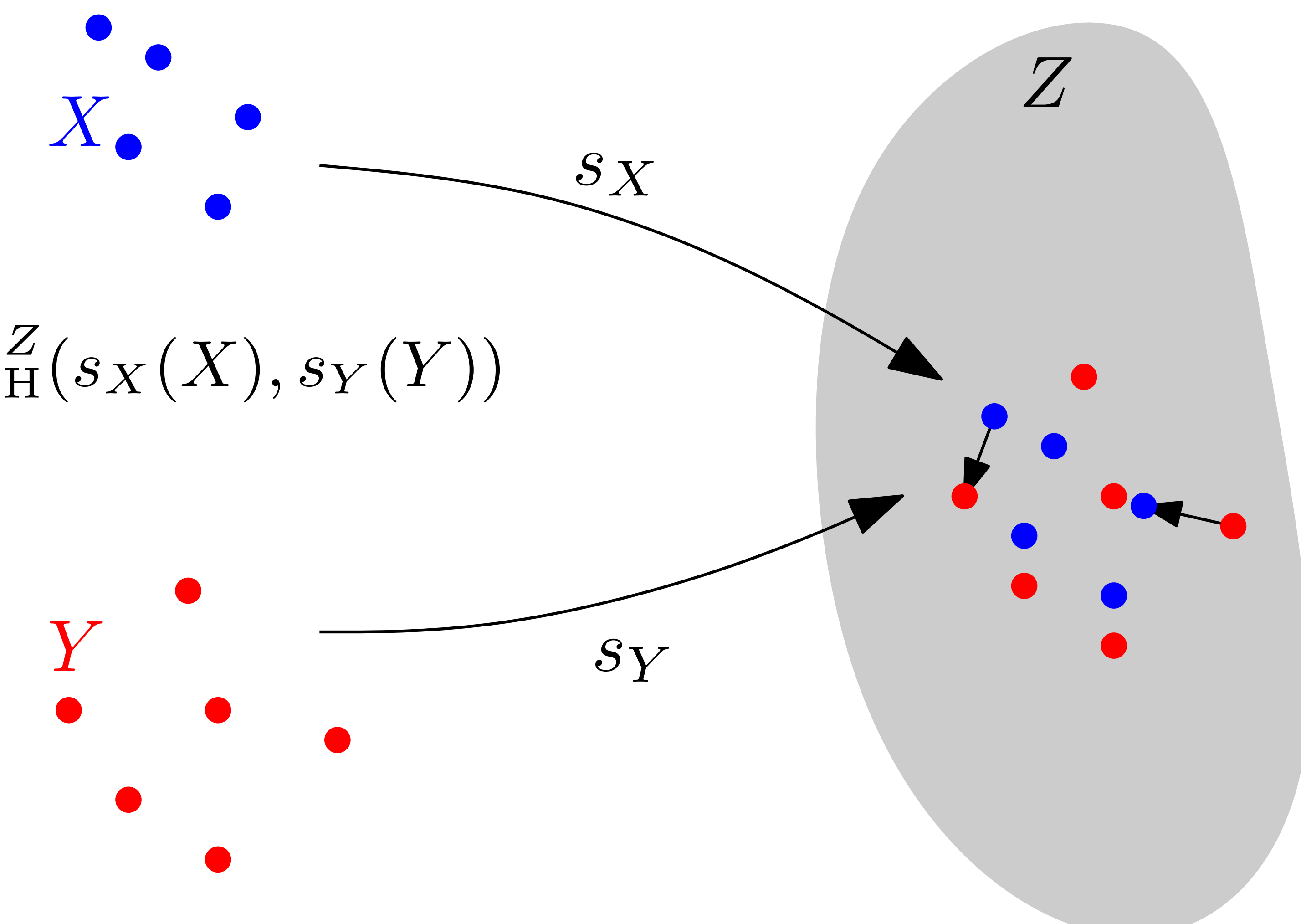


Mathematical framework

- geometric data set / underlying space $\equiv$ compact metric space

- distance between compact metric spaces $\equiv$ Gromov-Hausdorff (GH) distance

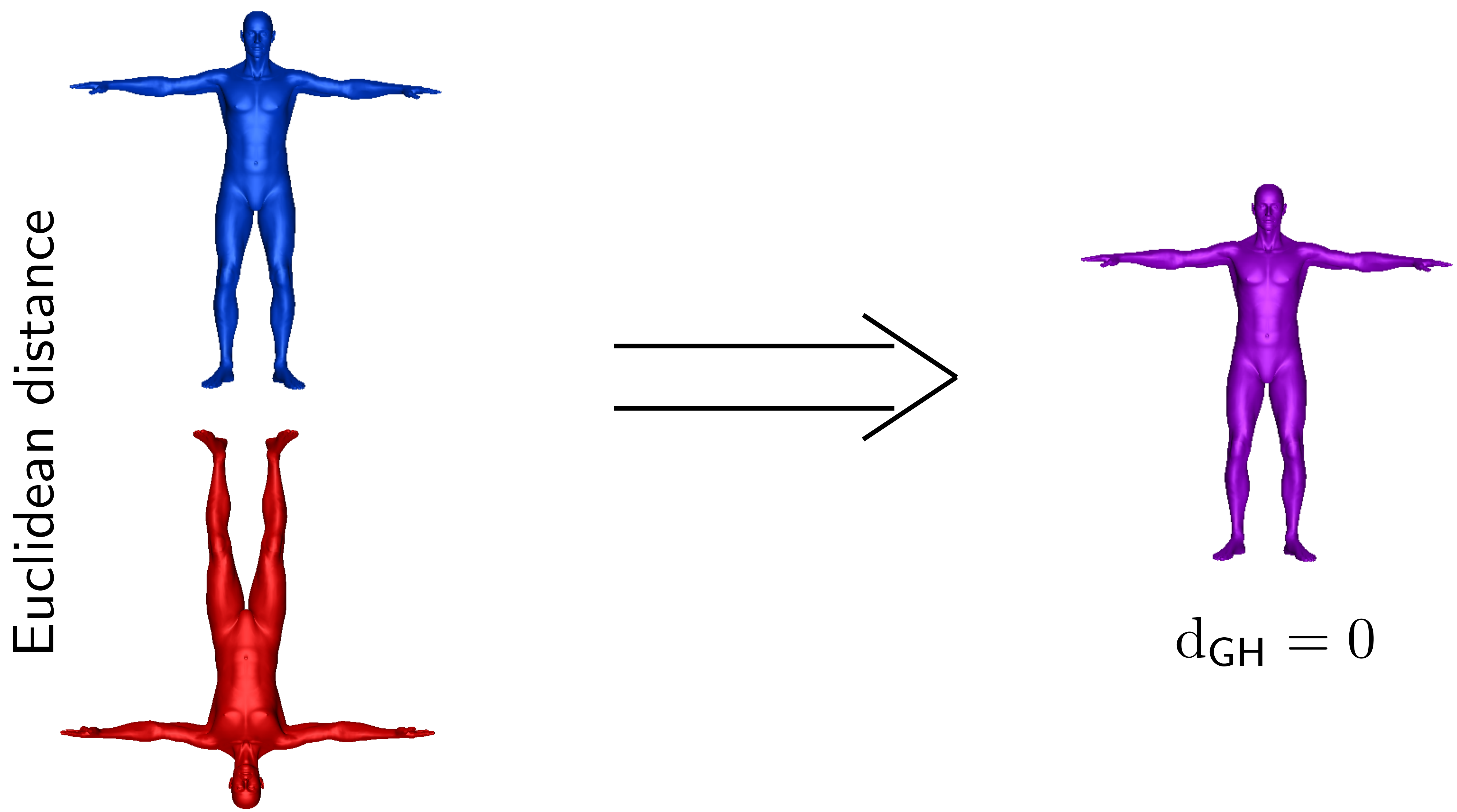
$$\mathrm{d}_{\mathrm{GH}}(X,Y) := \inf_{\substack{s_Y : X \to Z \\ s_Y : Y \to Z}} \mathrm{d}_{\mathrm{H}}^{Z}(s_X(X), s_Y(Y))$$



Mathematical framework

- geometric data set / underlying space $\equiv$ compact metric space

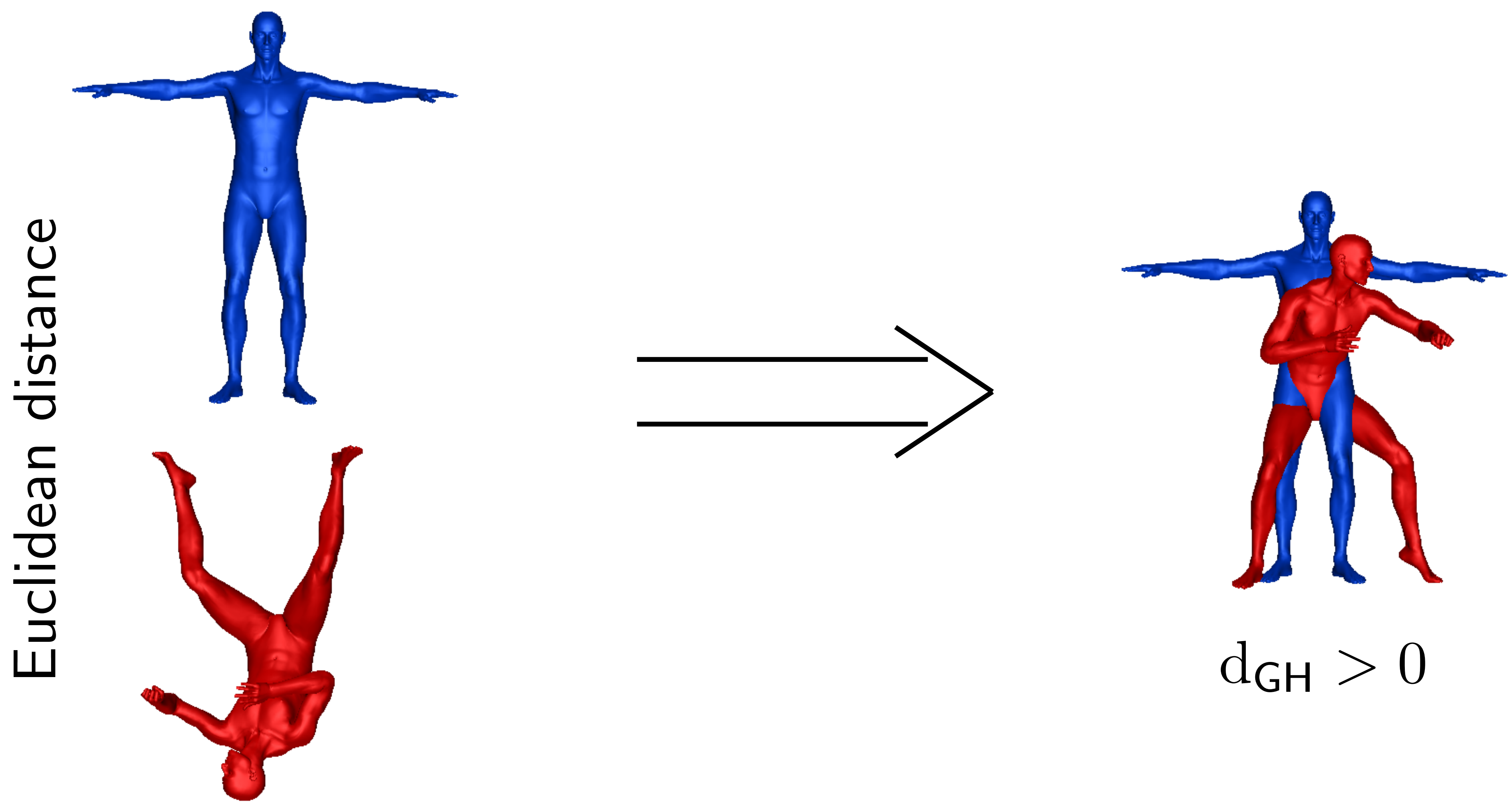
- distance between compact metric spaces $\equiv$ Gromov-Hausdorff (GH) distance



Mathematical framework

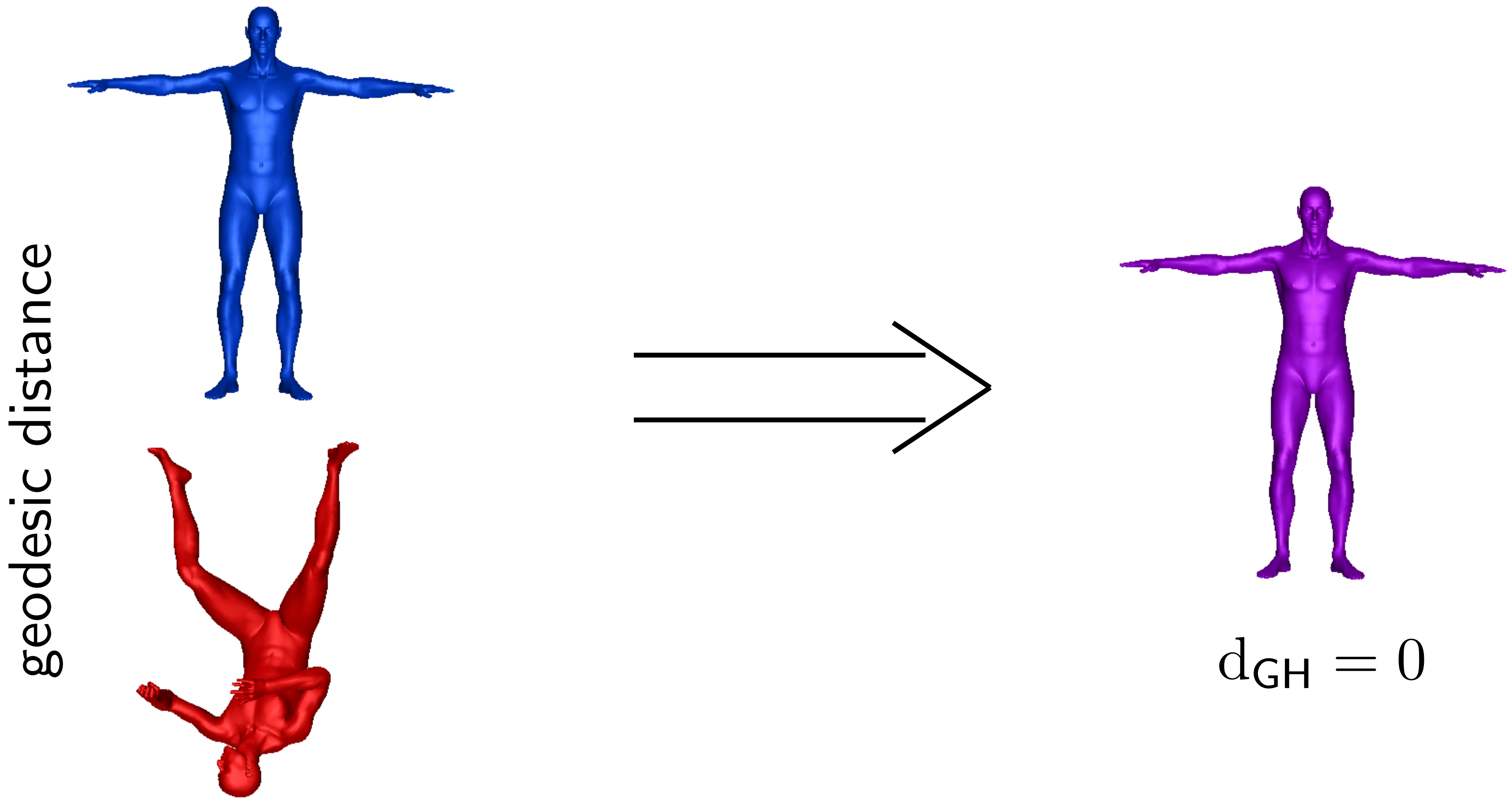
- geometric data set / underlying space $\equiv$ compact metric space

- distance between compact metric spaces $\equiv$ Gromov-Hausdorff (GH) distance



Mathematical framework

- geometric data set / underlying space $\equiv$ compact metric space
- distance between compact metric spaces $\equiv$ Gromov-Hausdorff (GH) distance



$\mathrm{d}_{\mathsf{GH}} = 0$

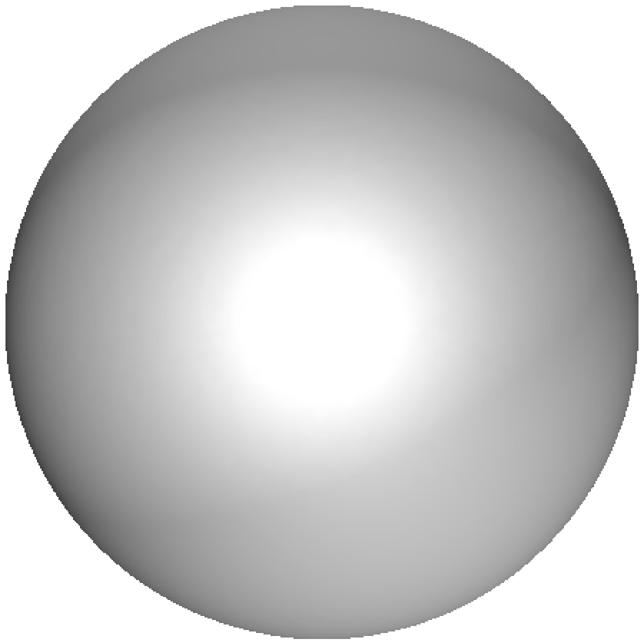
Mathematical framework

- geometric data set / underlying space $\equiv$ compact metric space

- distance between compact metric spaces $\equiv$ Gromov-Hausdorff (GH) distance

- descriptor / signature $\equiv$ persistence diagram / feature vector

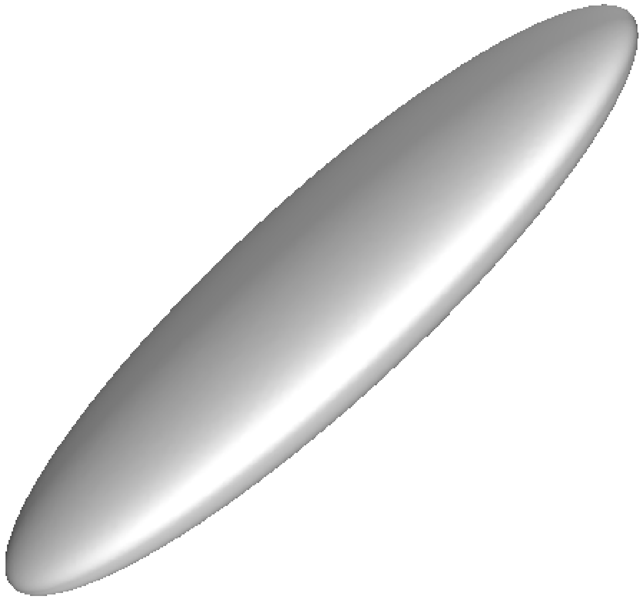
Why use descriptors



isometries
GH distance

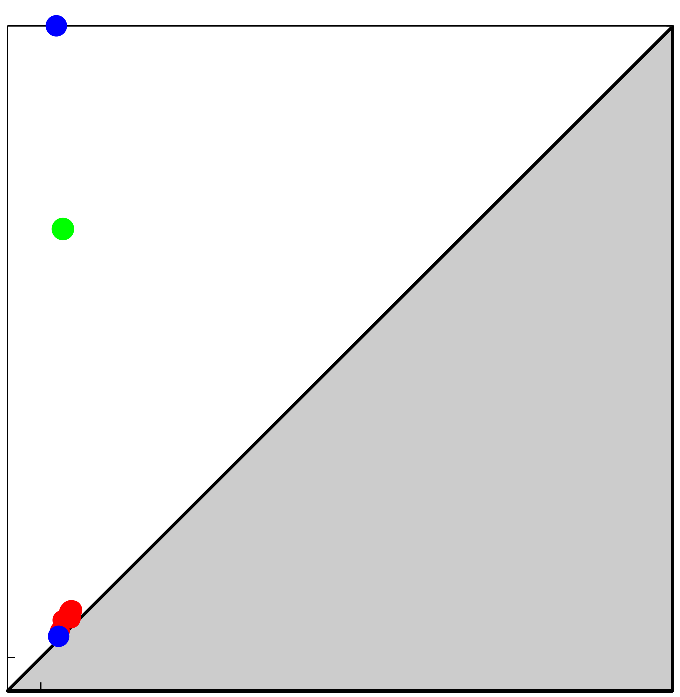
hard to compute

[Agarwal et al. 2015]



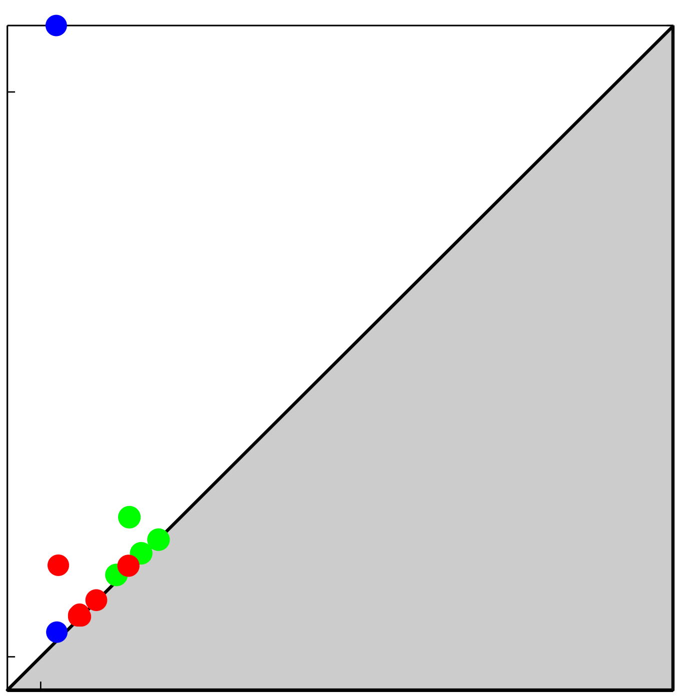
shape space

descriptors space

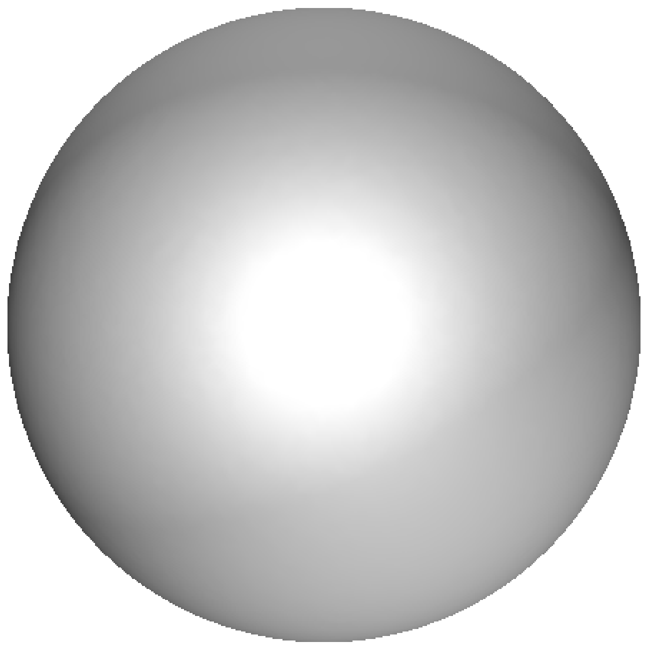


equality
distance

easy to compute



Why use descriptors

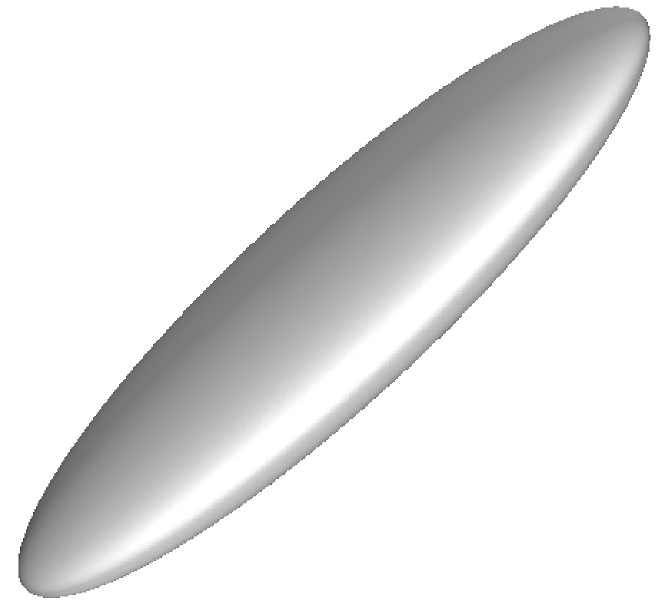


isometries

GH distance

hard to compute

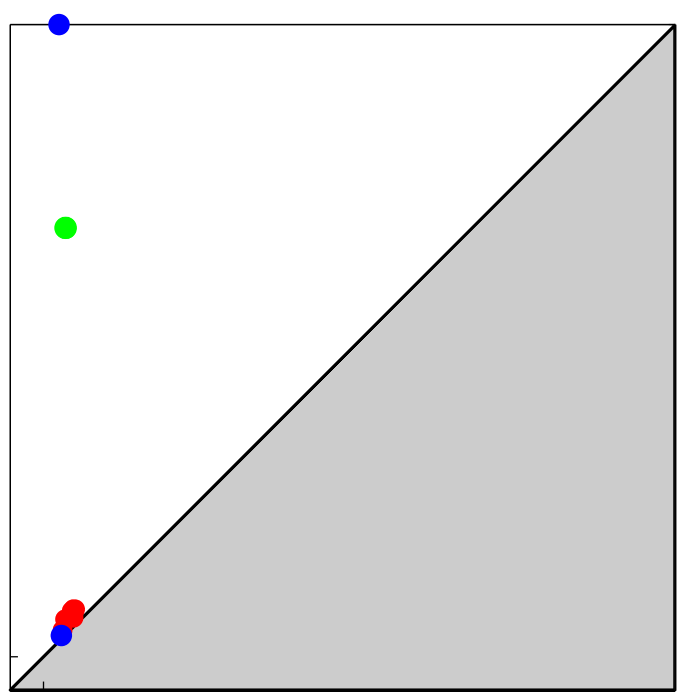
[Agarwal et al. 2015]



shape space

Ideally, descriptors distance = GH distance

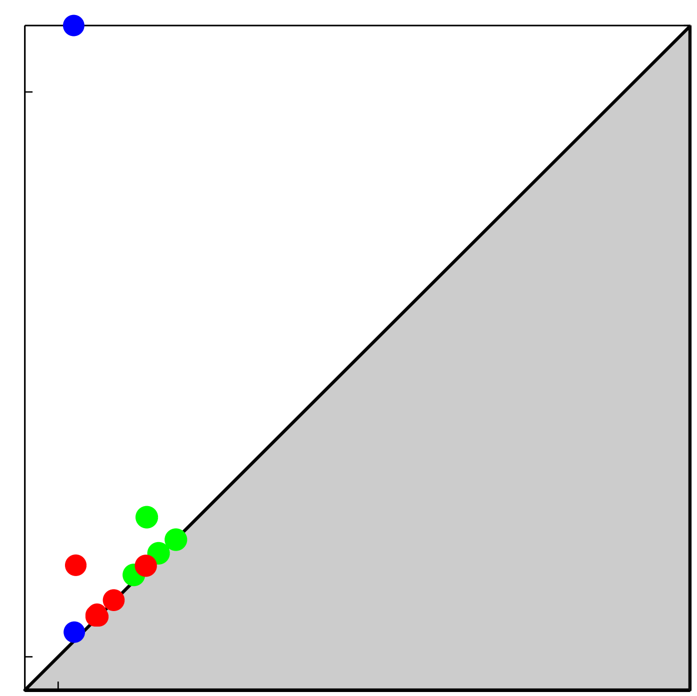
descriptors space



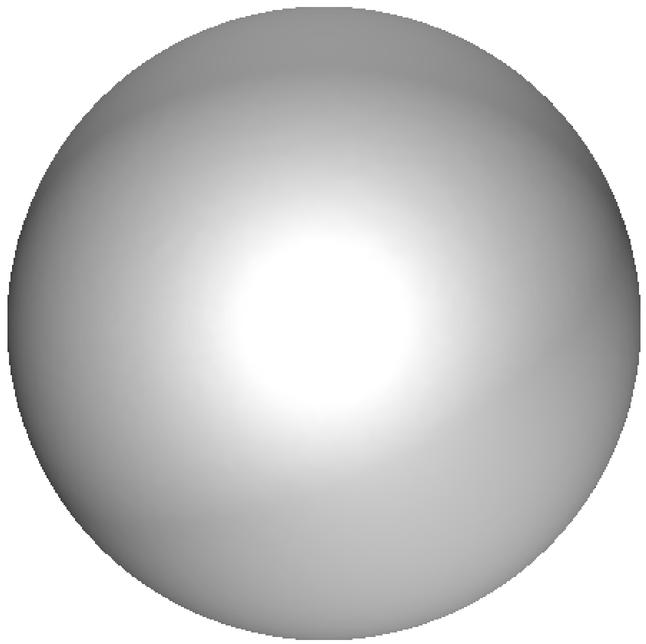
equality

distance

easy to compute



Why use descriptors

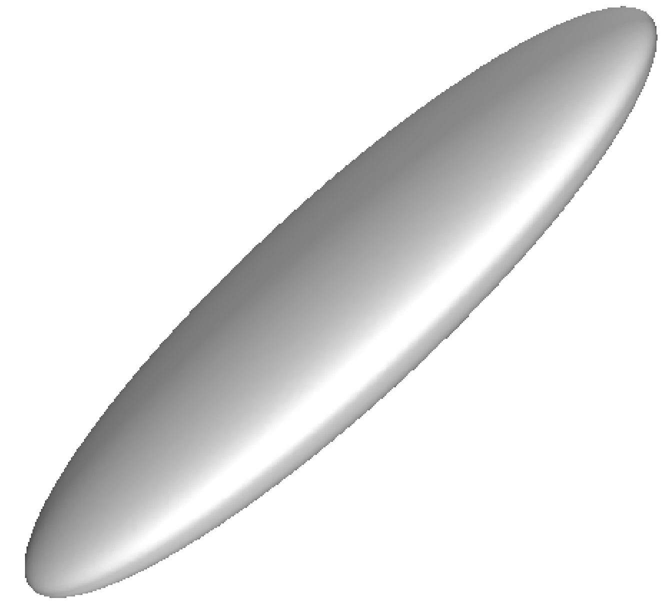


isometries

GH distance

hard to compute

[Agarwal et al. 2015]

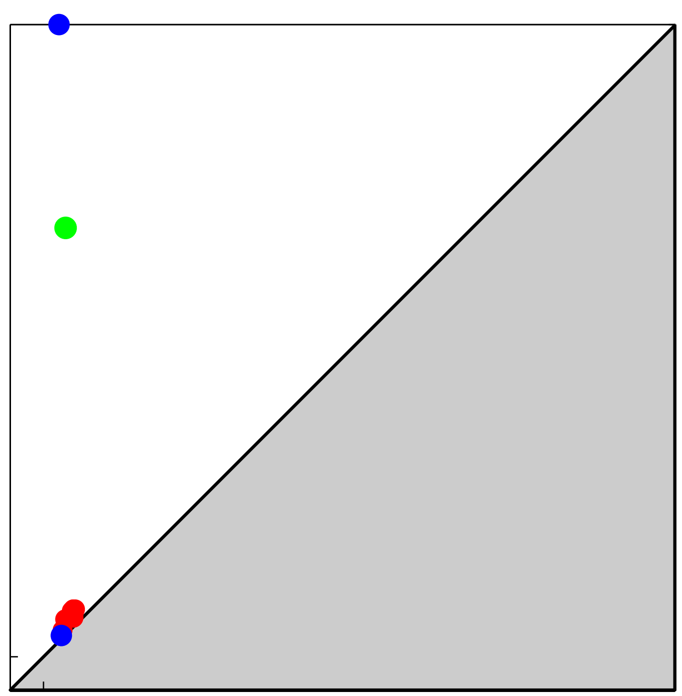


shape space

descriptors space

Ideally, descriptors distance $=$ GH distance

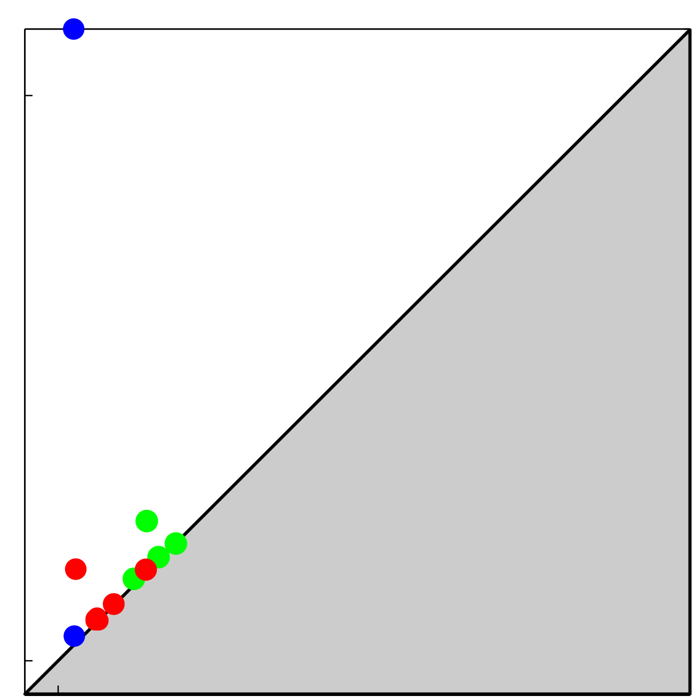
In reality, $\leq$



equality

distance

easy to compute



Why use descriptors

Some descriptors for images / 3d shapes / metric spaces:

- diameter
- curvature (mean, Gaussian, sectional)
- shape context (distribution of distances)
- heat kernel signature (heat diffusion)
- wave kernel signature (Maxwell's equations)
- spin image (local neighborhood parametrization)
- SIFT features (local distribution of gradient orientations)
- etc.

Why use descriptors

Some descriptors for images / 3d shapes / metric spaces:

- diameter
- curvature (mean, Gaussian, sectional)
- shape context (distribution of distances)
- heat kernel signature (heat diffusion)
- wave kernel signature (Maxwell's equations)
- spin image (local neighborhood parametrization)
- SIFT features (local distribution of gradient orientations)
- etc.

geometry

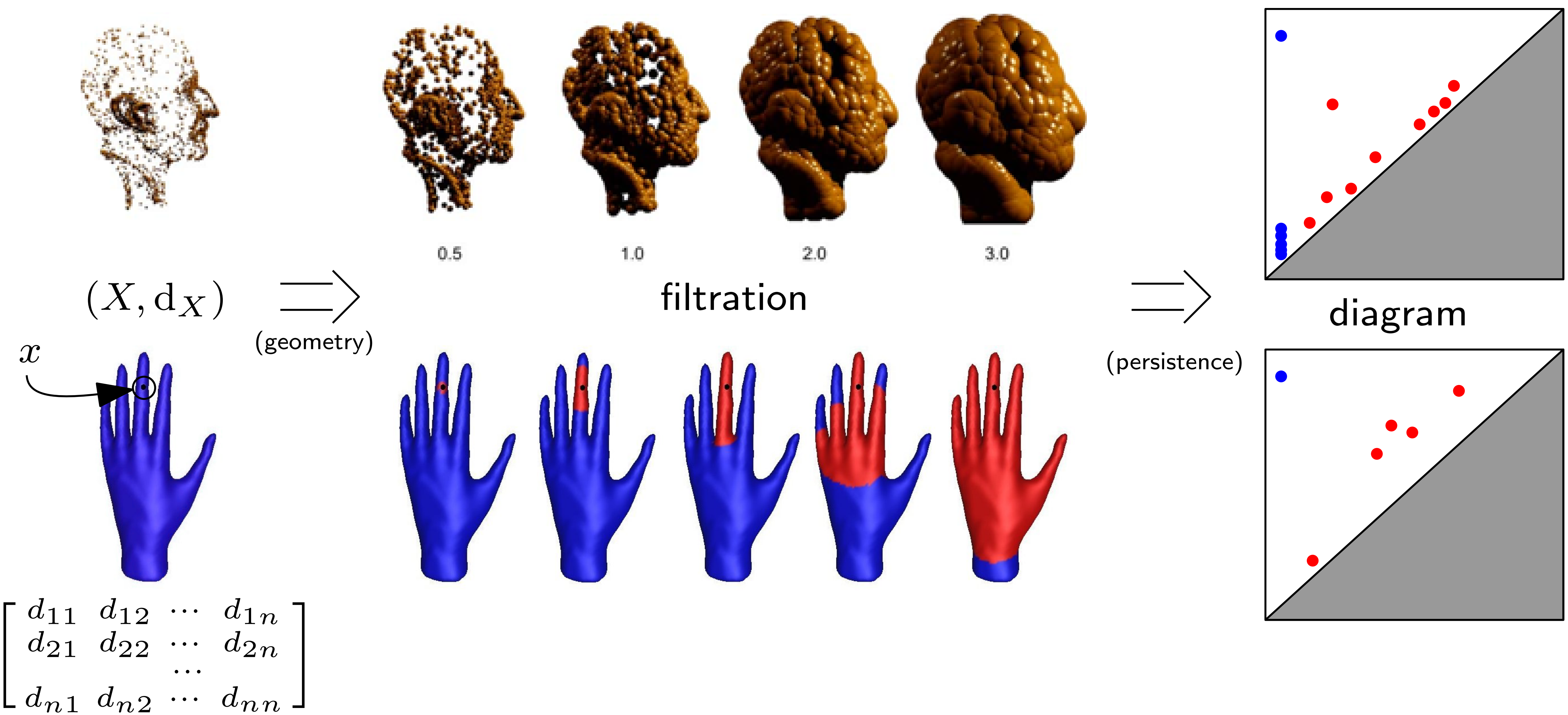
statistics

Topological descriptors

Input: a finite/compact **metric space** (X, d_X), a **basepoint** $x \in X$

Construction: a **filtration** (nested family of sublevel-sets of real-valued function)

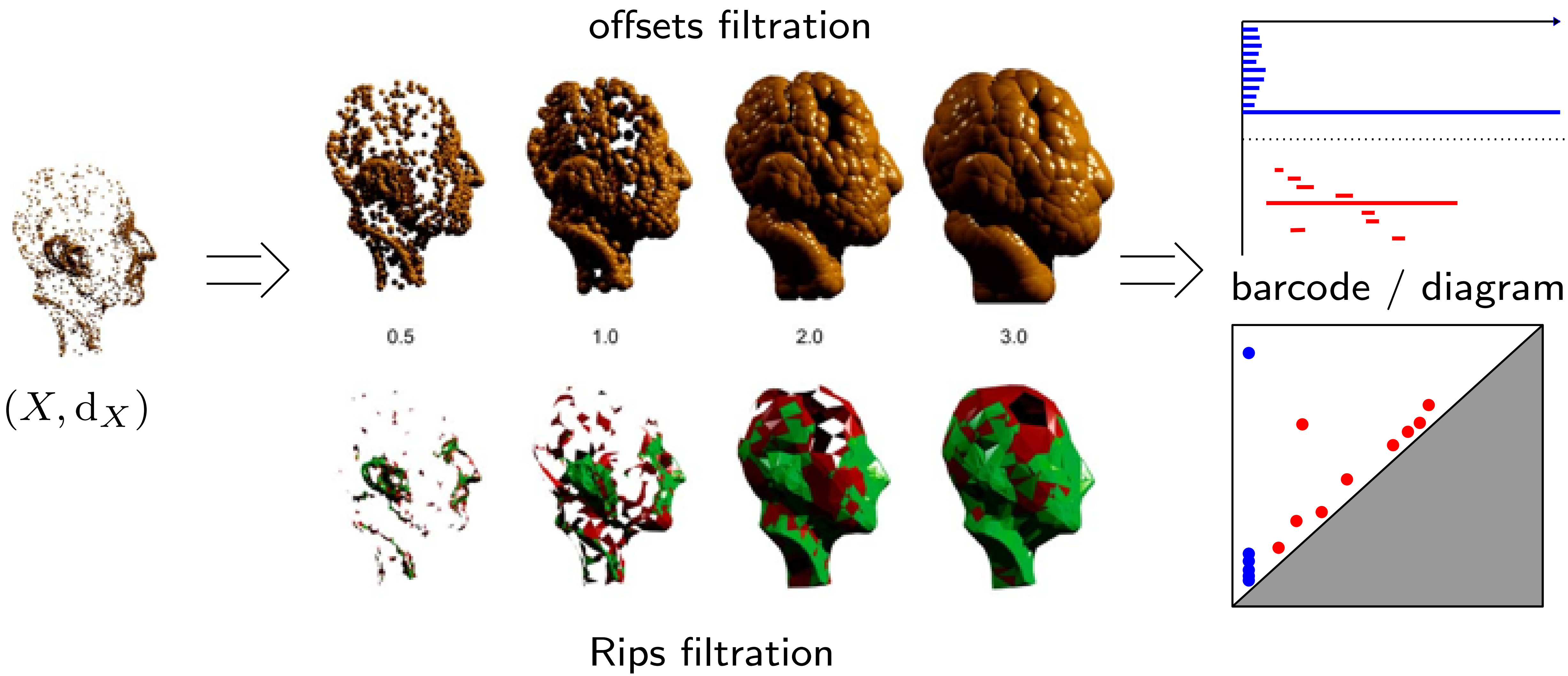
Signature: the **persistence diagram** associated with the filtration



Global topological descriptors

Input: a compact metric space (X, d_X)

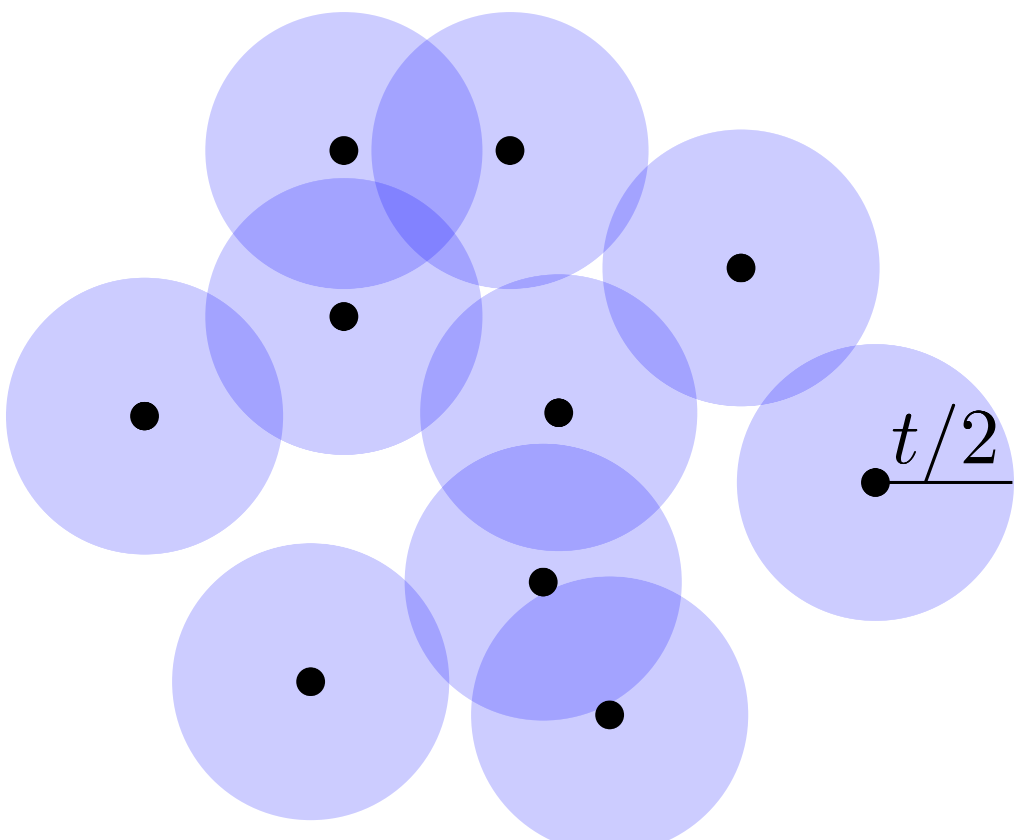
Descriptor: $\mathrm{dgm}\,\mathcal{R}(X, \mathrm{d}_X)$ where $\mathcal{R}$ stands for *Vietoris-Rips filtration*

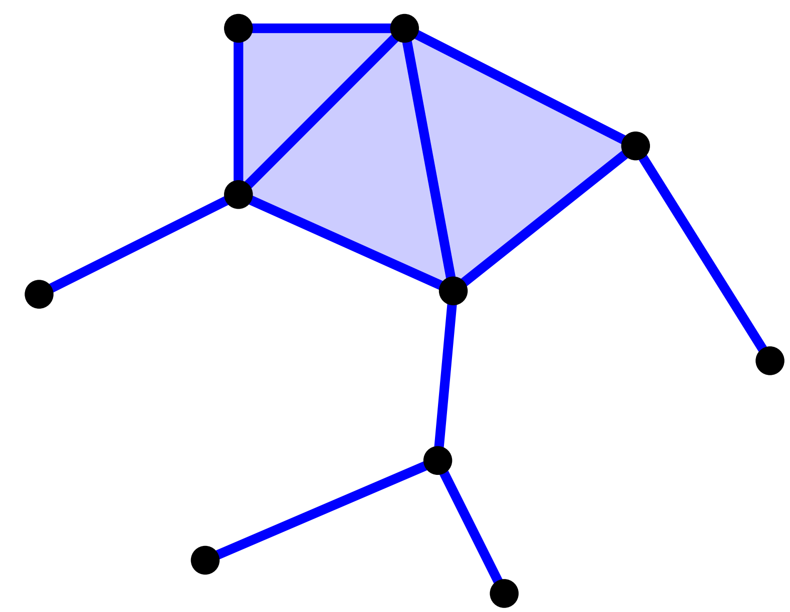


Global topological descriptors

Input: a compact metric space (X, d_X)

Descriptor: $\operatorname{dgm} \mathcal{R}(X, \mathrm{d}_X)$ where $\mathcal{R}$ stands for *Vietoris-Rips filtration*



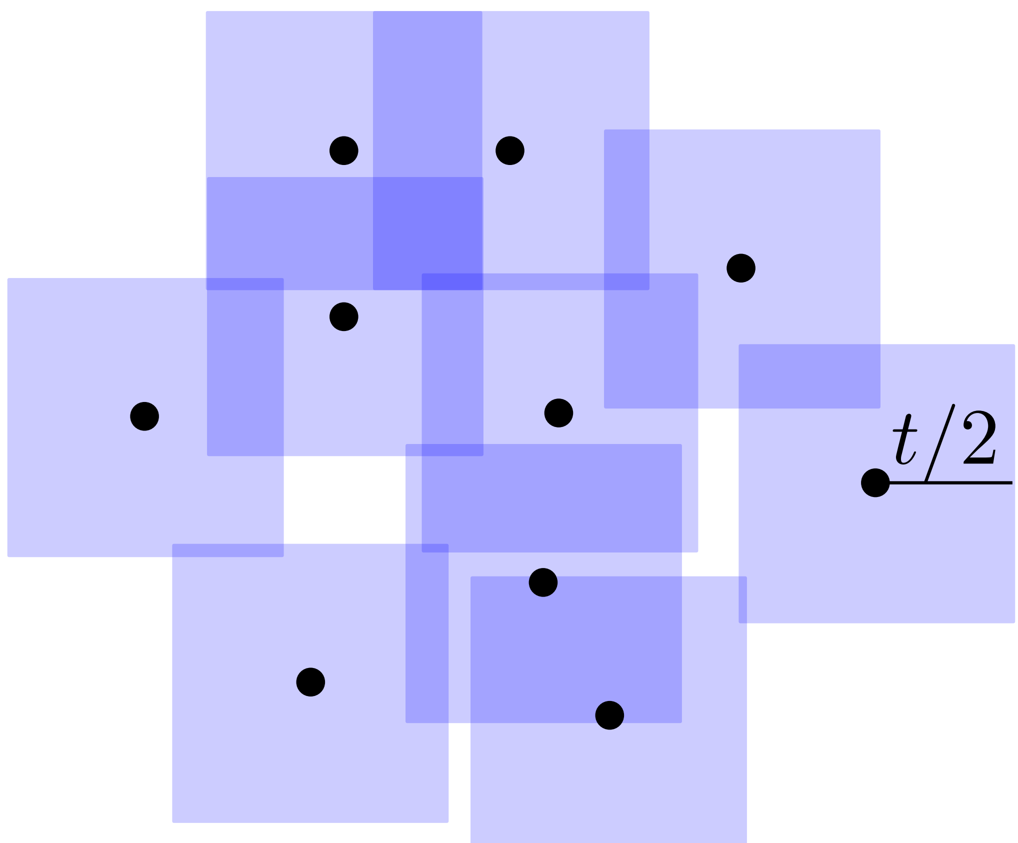


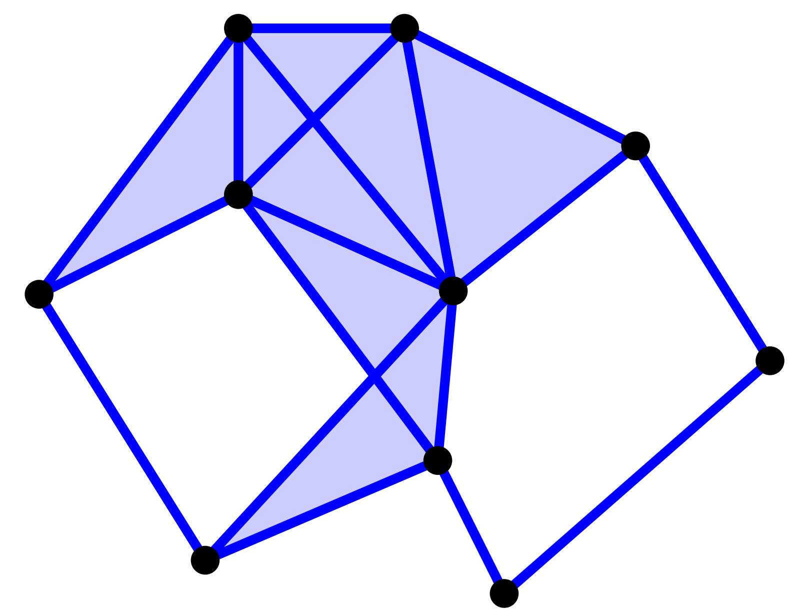
$$\{x_0, \cdots, x_r\} \in R_t(X, \mathrm{d}_X) \quad \Longleftrightarrow \quad t \geq \max_{i,j} \mathrm{d}_X(x_i, x_j)$$

Global topological descriptors

Input: a compact metric space (X, d_X)

Descriptor: $\mathrm{dgm}\,\mathcal{R}(X, \mathrm{d}_X)$ where $\mathcal{R}$ stands for *Vietoris-Rips filtration*

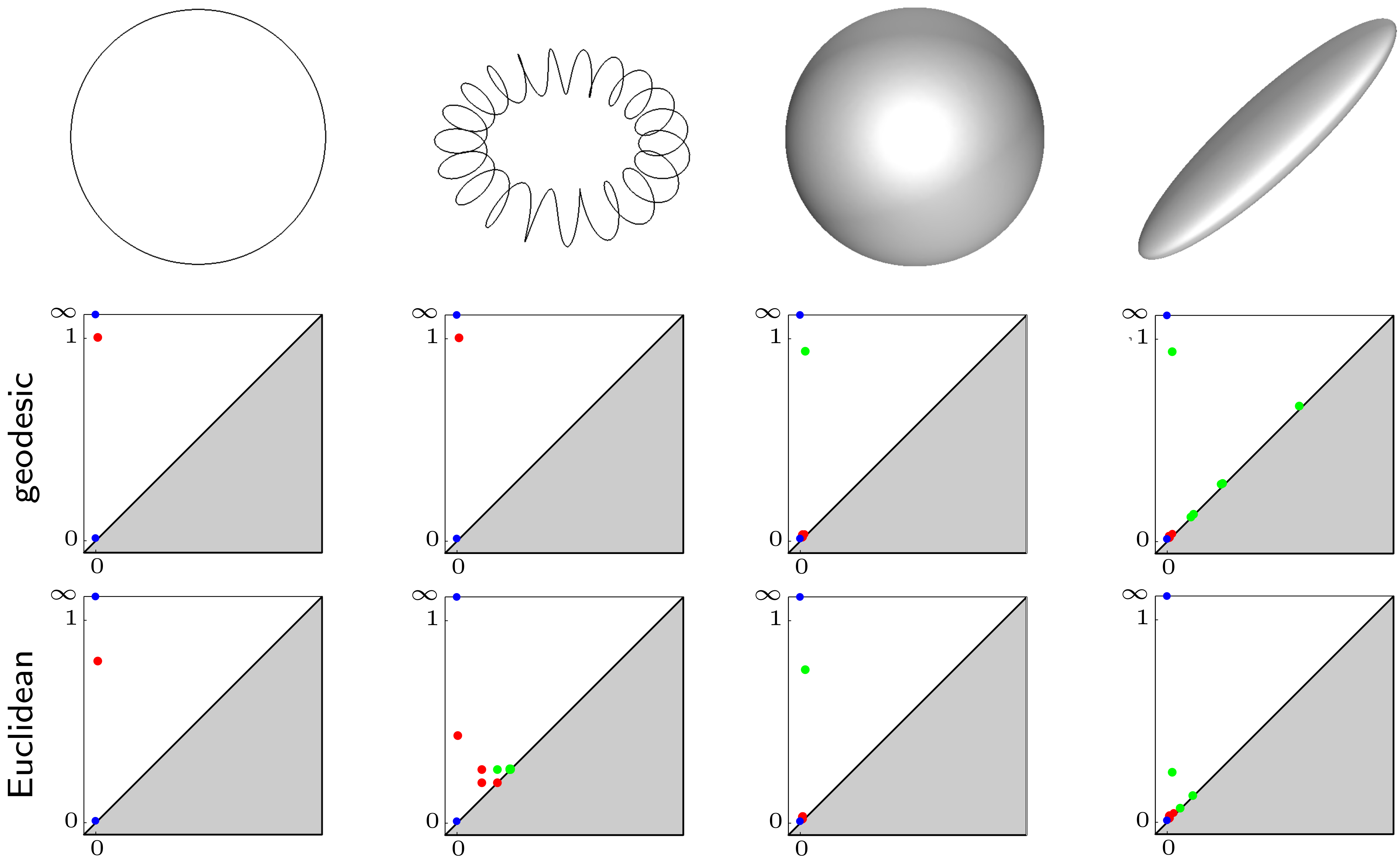




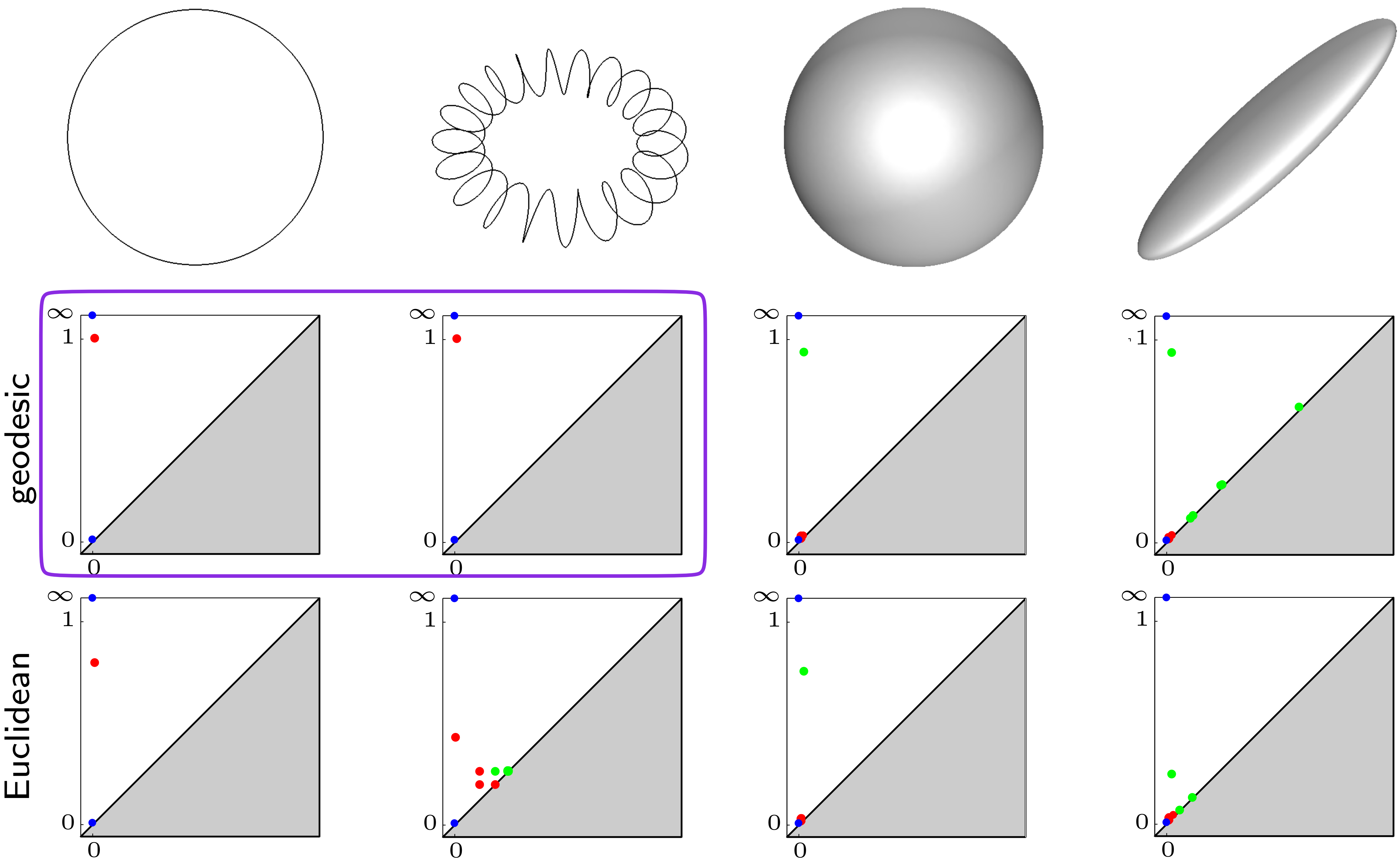
$\mathrm{R}_t(X, \mathrm{d}_X)$

$$\{x_0, \cdots, x_r\} \in R_t(X, \mathrm{d}_X) \quad \Longleftrightarrow \quad t \geq \max_{i,j} \mathrm{d}_X(x_i, x_j)$$

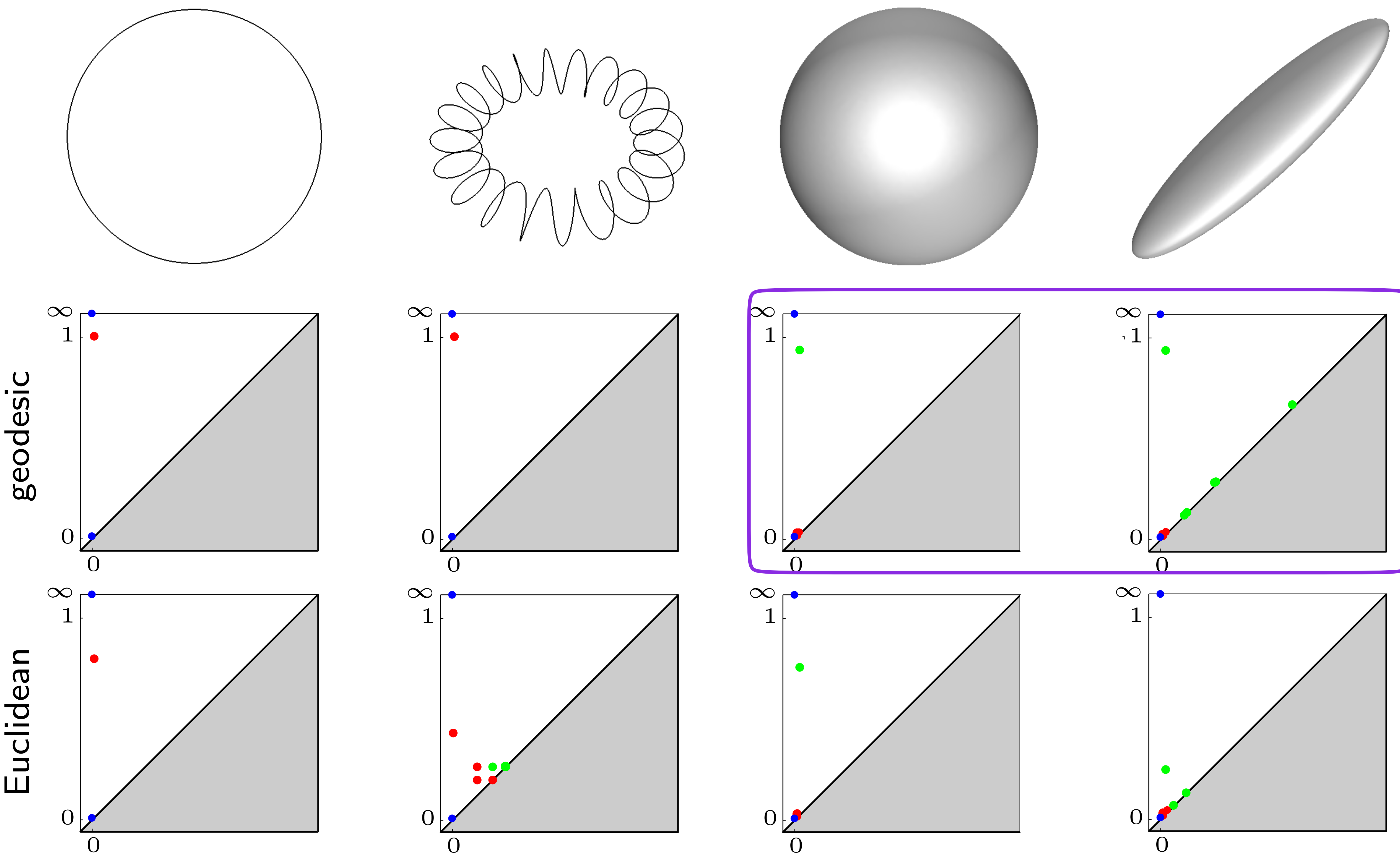
Some examples



Some examples

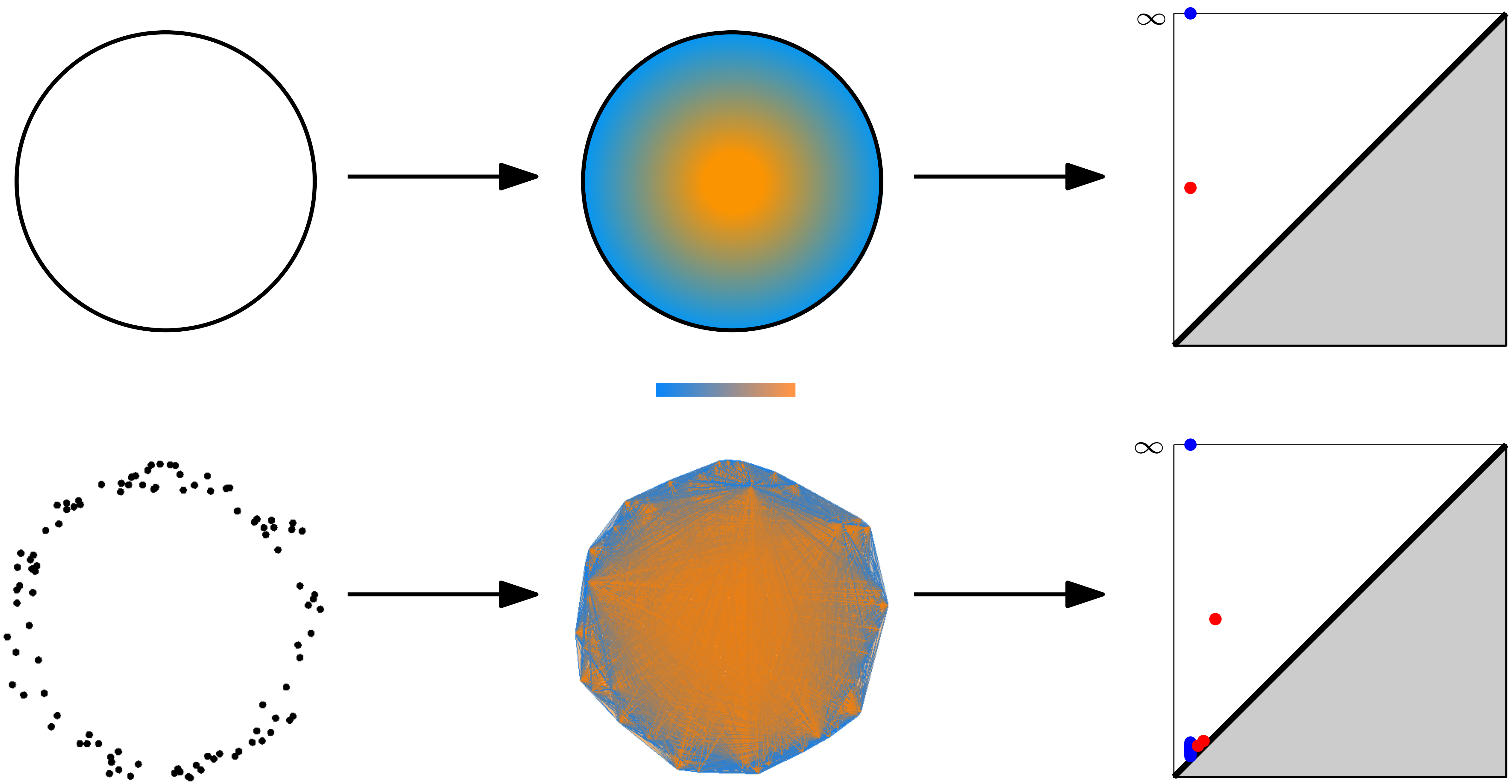


Some examples



Stability

Theorem: [Chazal, de Silva, O. 2013]
For any compact metric spaces (X, d_X) and (Y, d_Y),
$\mathrm{d}_\infty(\mathrm{dgm}\,\mathcal{R}(X, \mathrm{d}_X),\ \mathrm{dgm}\,\mathcal{R}(Y, \mathrm{d}_Y)) \leq 2\mathrm{d}_{\mathrm{GH}}(X, Y)$.

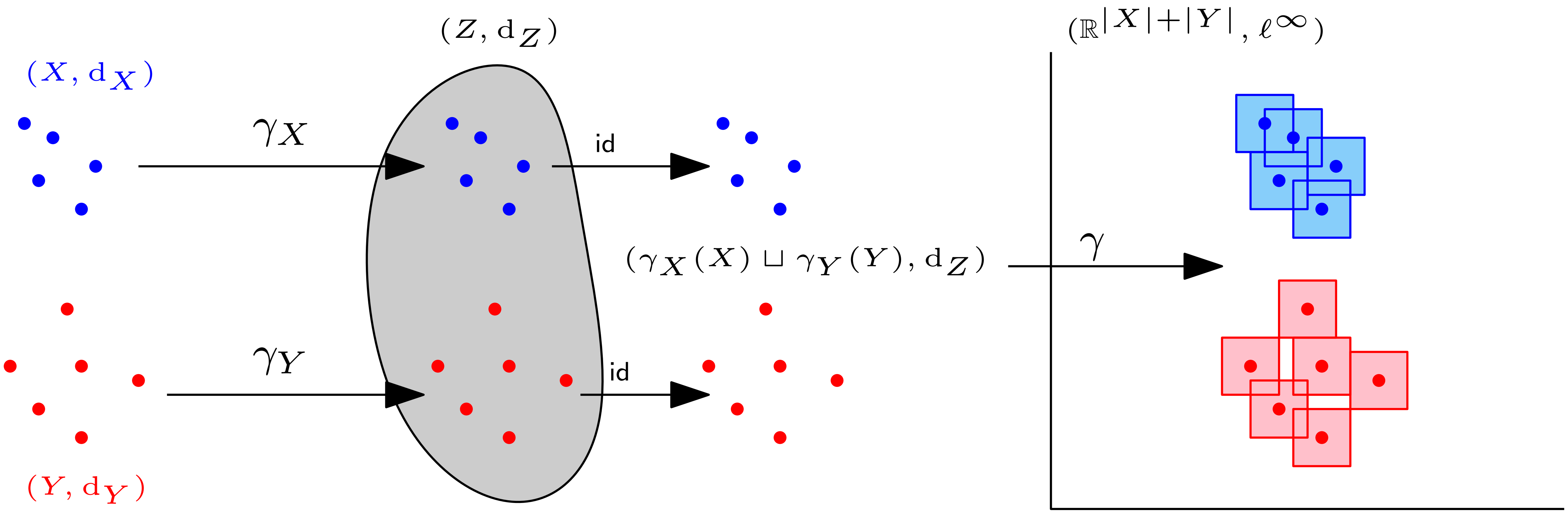


Stability

finite

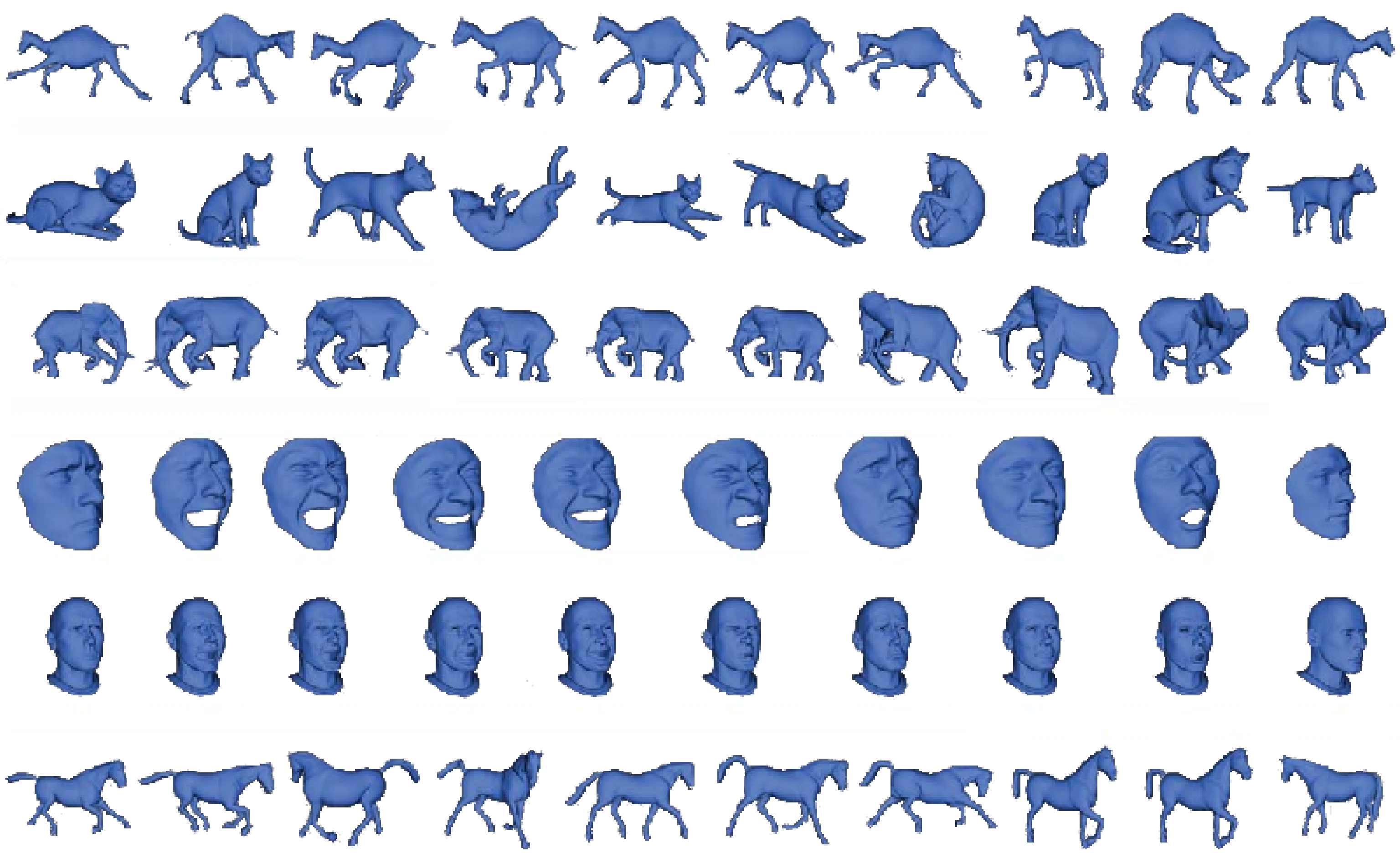
Theorem: [Chazal, de Silva, O. 2013]
For any ~~compact~~ metric spaces (X, d_X) and (Y, d_Y),
$\mathrm{d}_\infty(\mathrm{dgm}\,\mathcal{R}(X, \mathrm{d}_X),\ \mathrm{dgm}\,\mathcal{R}(Y, \mathrm{d}_Y)) \leq 2\mathrm{d}_{\mathrm{GH}}(X, Y)$.

Proof outline:

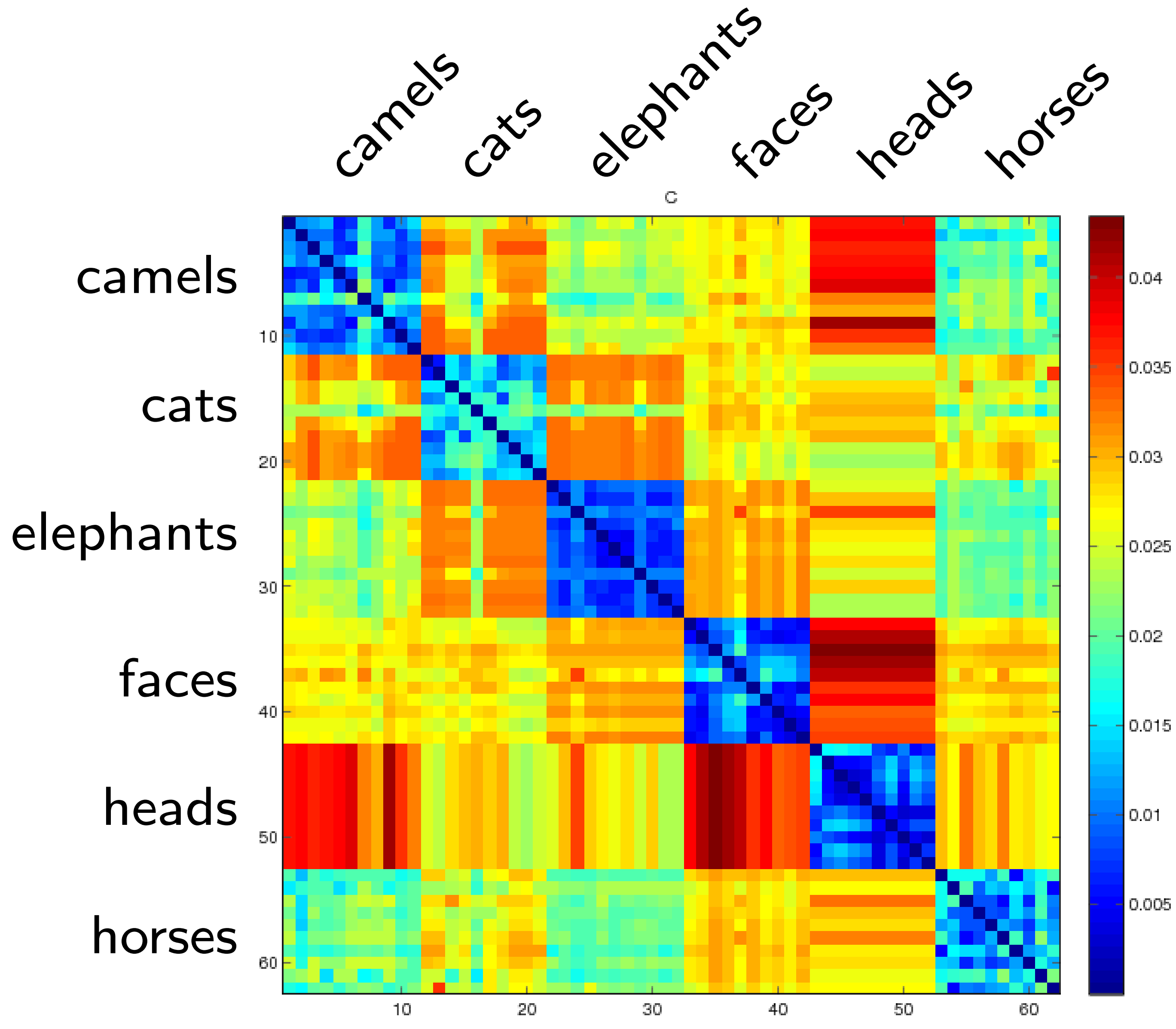


Toy application (unsupervised shape classification)

60 shapes (represented as point clouds with approximate geodesic distances)

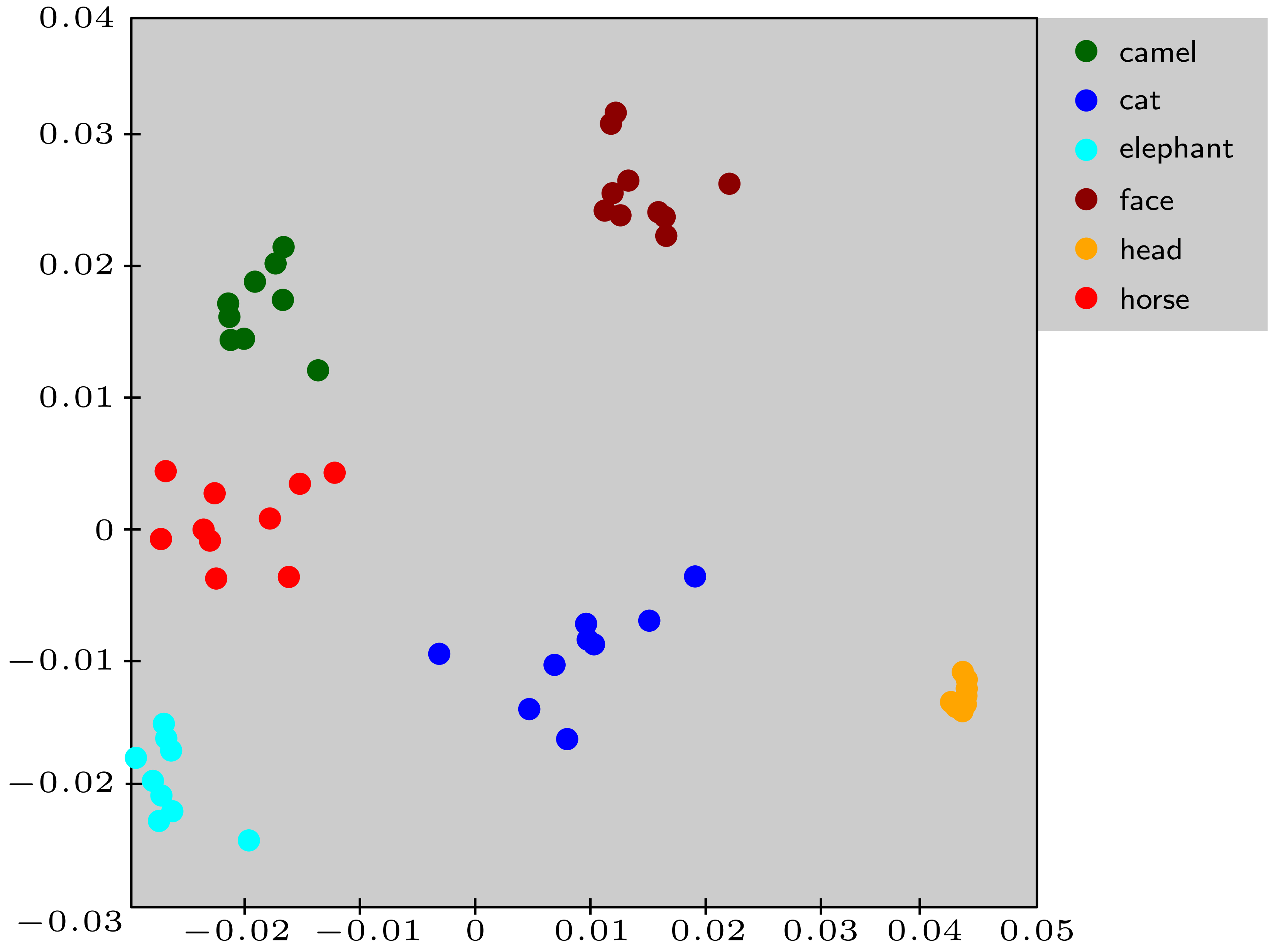


Toy application (unsupervised shape classification)



computation time $\approx$ 1 hour (pacing phase: bottleneck distances computation)

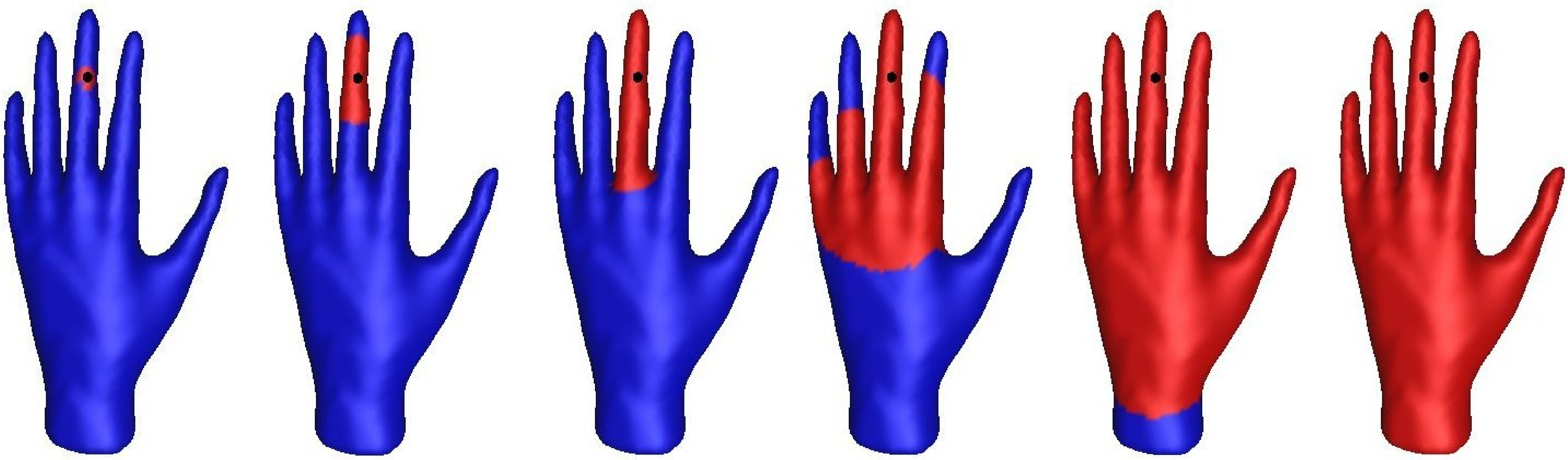
Toy application (unsupervised shape classification)



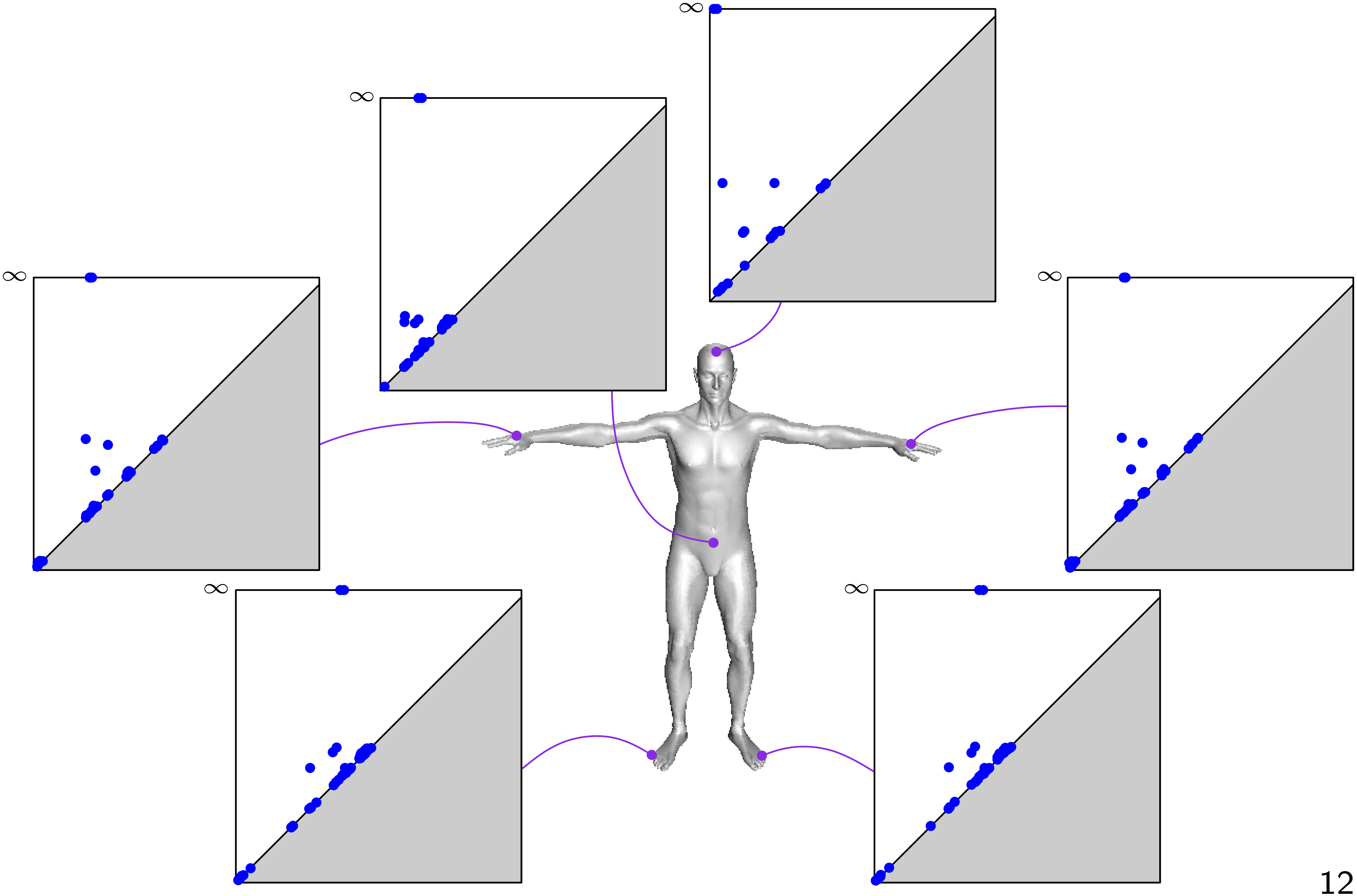
Local topological descriptors

Input: a compact metric space (X, d_X), a basepoint $x \in X$

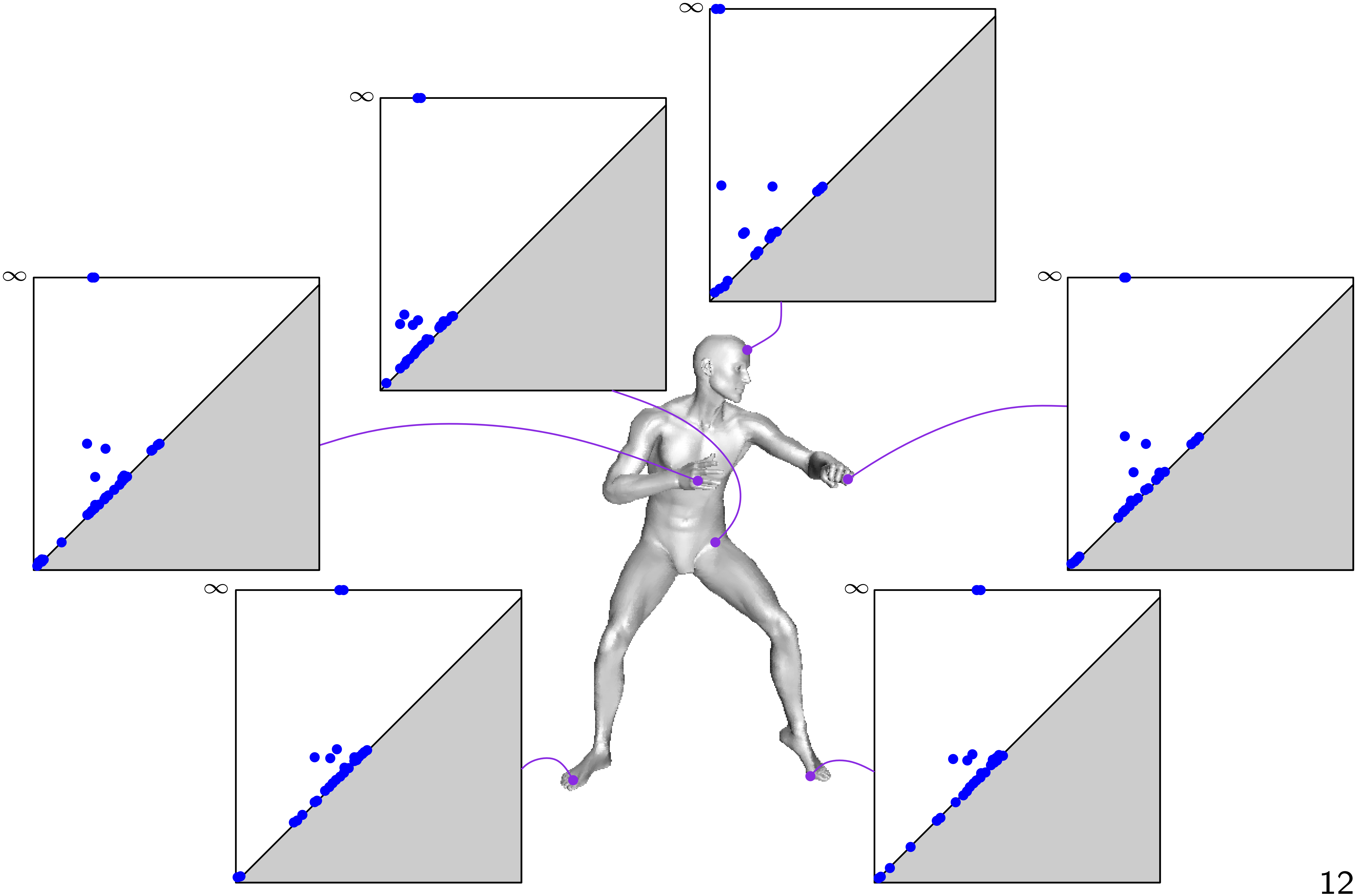
Descriptor: $\operatorname{dgm} \mathrm{d}_X(x, \cdot)$



Some examples



Some examples



Stability

Theorem: (local descriptors) [Carrière, O., Ovsjanikov 2015]
Let (X, d_X) and (Y, d_Y) be two compact length spaces with bounded curvature, and let $x \in X$ and $y \in Y$. If $\mathrm{d}_{\mathrm{GH}}((X,x),(Y,y)) \leq \frac{1}{20} \min\{\varrho(X),\ \varrho(Y)\}$, then $\mathrm{d}_\infty(\mathrm{dgm}\,\mathrm{d}_X(\cdot,x), \mathrm{dgm}\,\mathrm{d}_Y(\cdot,y)) \leq 20\ \underbrace{\mathrm{d}_{\mathrm{GH}}((X,x),(Y,y))}$.

(adaptation of d_{GH} to pointed spaces)

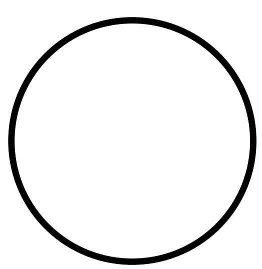
(convexity radii)

Stability

Theorem: (local descriptors) [Carrière, O., Ovsjanikov 2015]
Let (X, d_X) and (Y, d_Y) be two compact length spaces with bounded curvature, and let $x \in X$ and $y \in Y$. If $\mathrm{d_{GH}}((X,x),(Y,y)) \leq \frac{1}{20} \min\{\varrho(X),\ \varrho(Y)\}$, then $\mathrm{d}_\infty(\mathrm{dgm}\, \mathrm{d}_X(\cdot, x), \mathrm{dgm}\, \mathrm{d}_Y(\cdot, y)) \leq 20\ \underbrace{\mathrm{d_{GH}}((X,x),(Y,y))}_{\text{(adaptation of } \mathrm{d_{GH}} \text{ to pointed spaces)}}$.

(convexity radii)

Prerequisite: $\mathrm{d_{GH}}(X, Y) < \frac{1}{20} \min\{\varrho(X),\ \varrho(Y)\}$

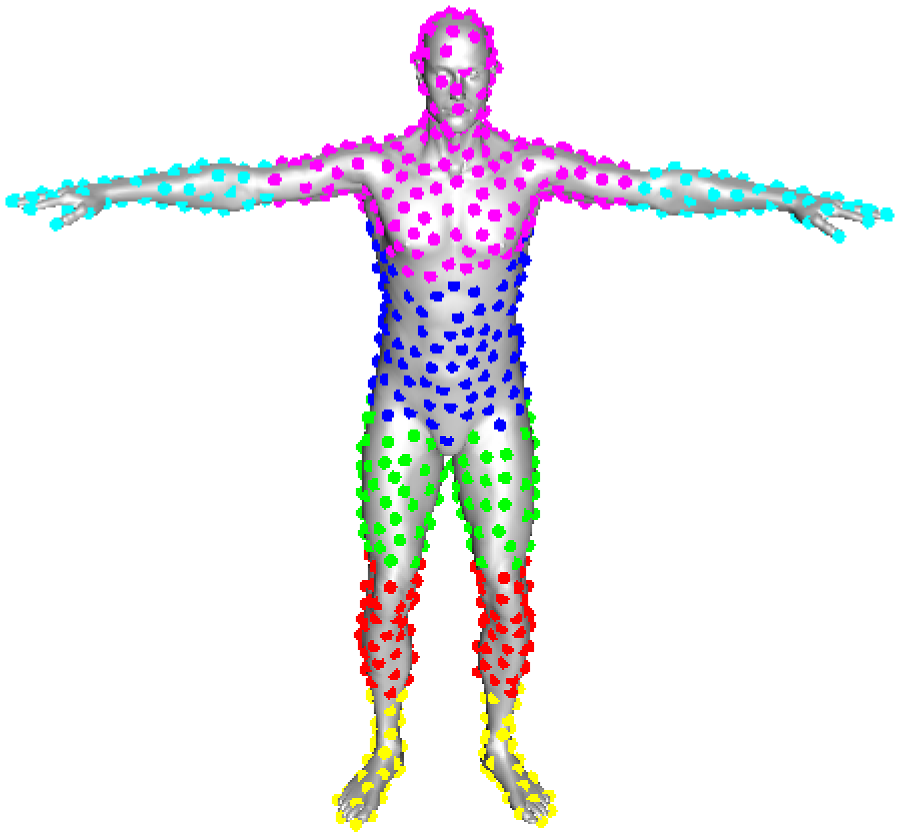
$X =$ 

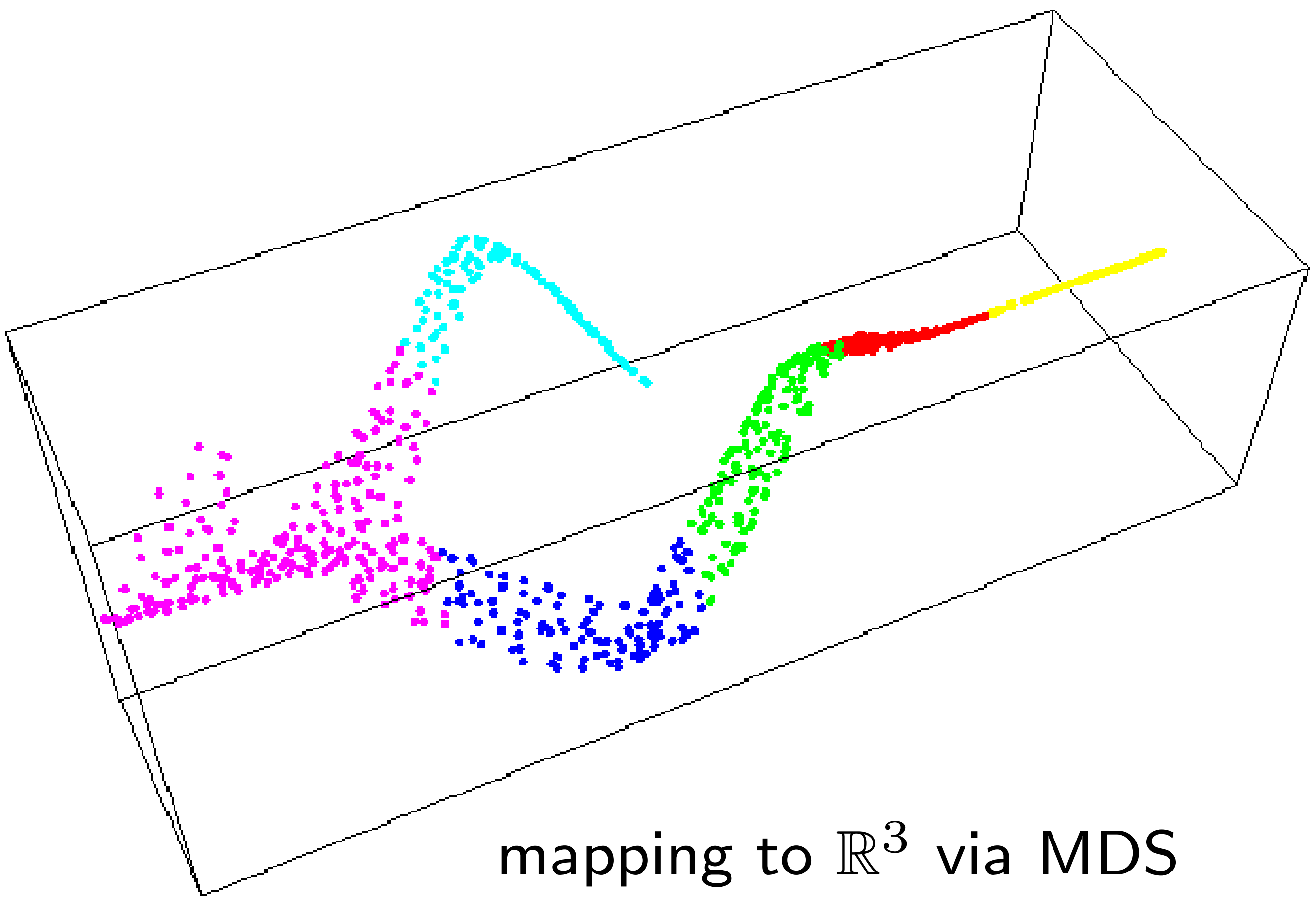
$$\mathrm{d_{GH}}(X, Y) < \infty = \varrho(Y)$$

$Y =$ 

$$\forall f, g,\ \mathrm{d}_\infty(\mathrm{dgm}\, f, \mathrm{dgm}\, g) = \infty$$

Toy application (unsupervised shape segmentation)

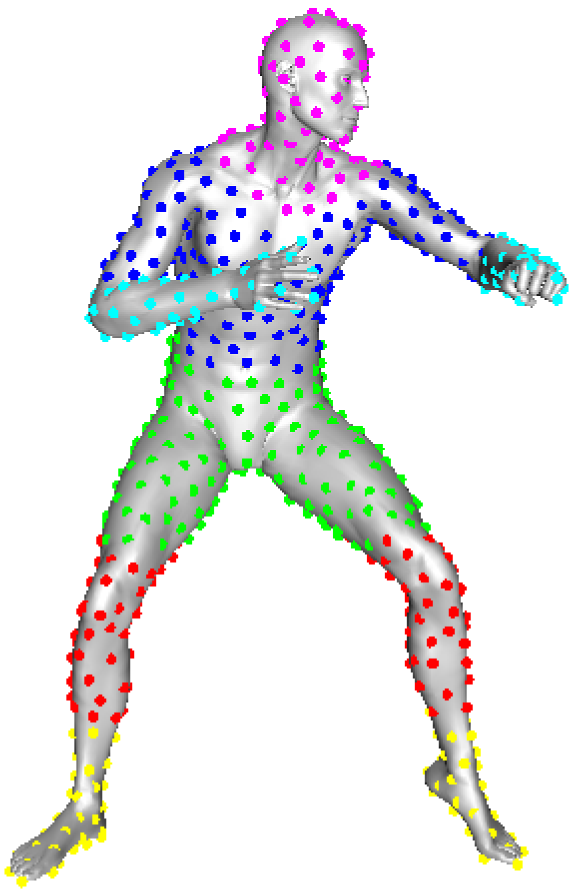


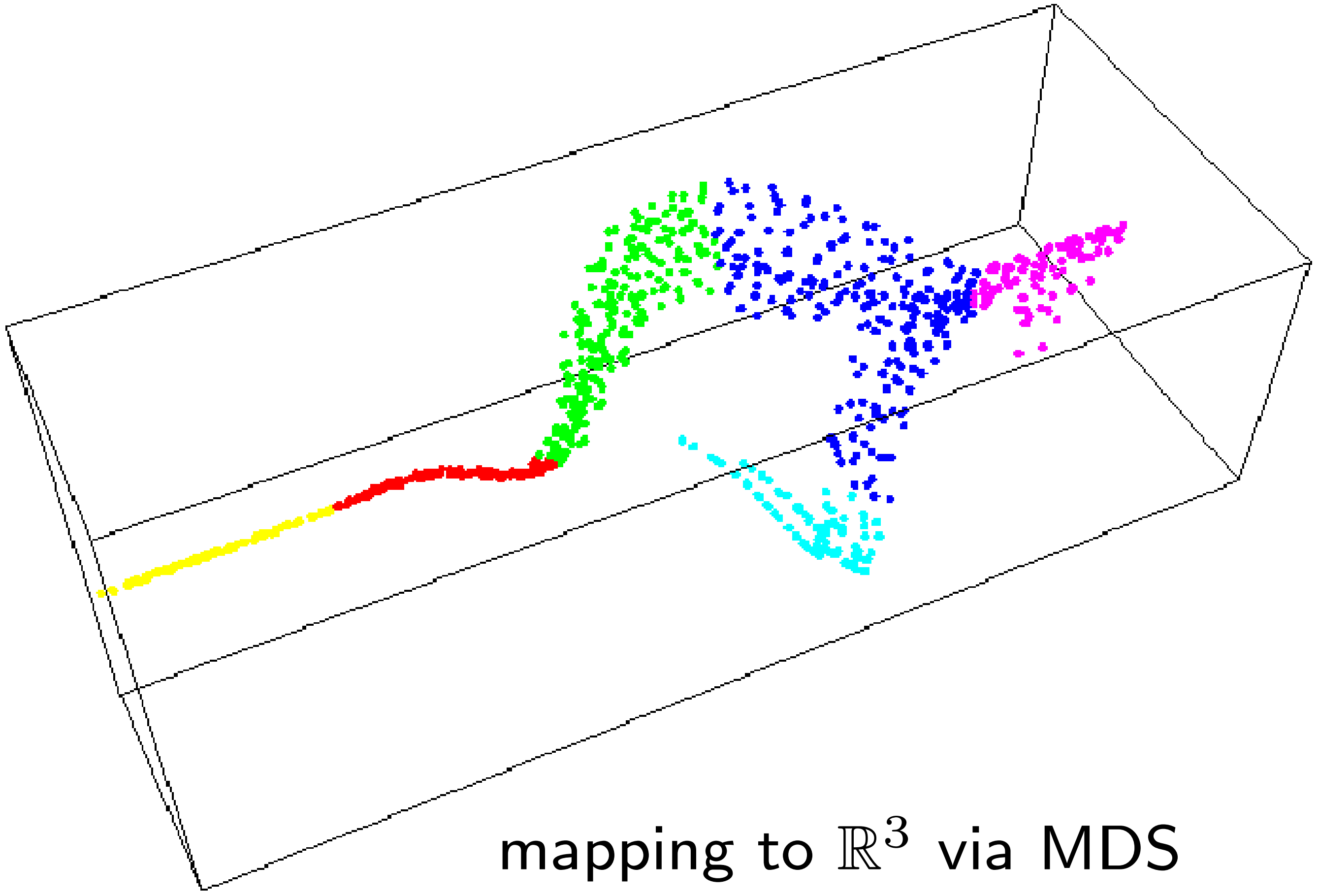


mapping to $\mathbb{R}^3$ via MDS

k-means in $\mathbb{R}^3$

Toy application (unsupervised shape segmentation)



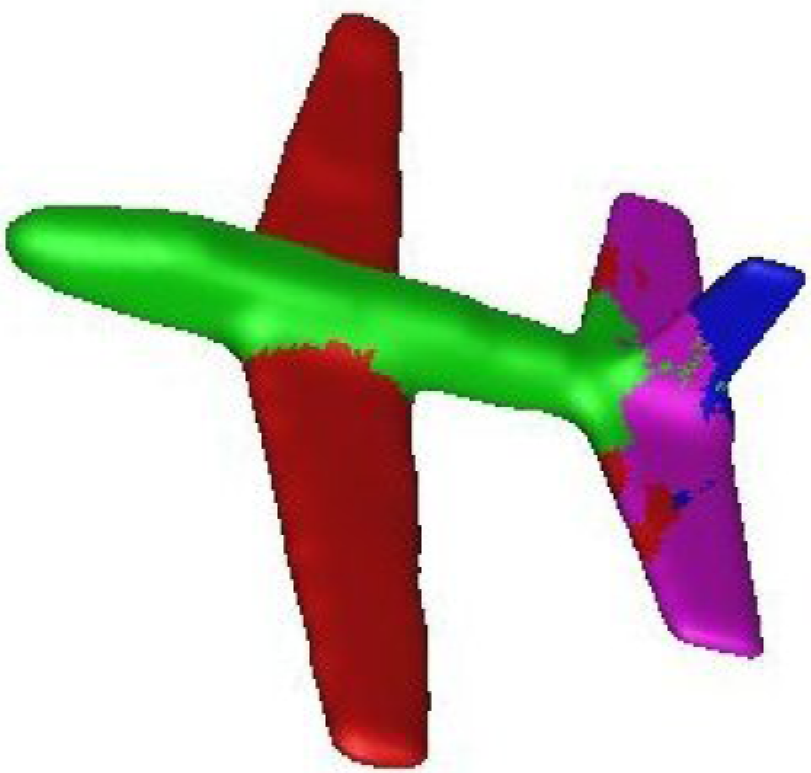


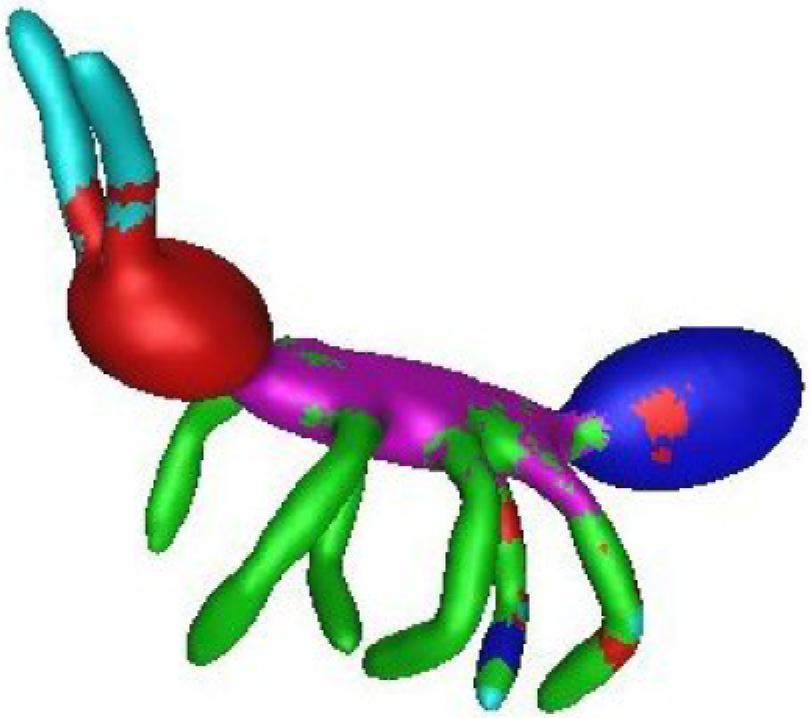
mapping to $\mathbb{R}^3$ via MDS

k-means in $\mathbb{R}^3$

Toy application (supervised shape segmentation)

Strategy: use k-NN classifier in diagram space (equipped with d_∞)





Outline

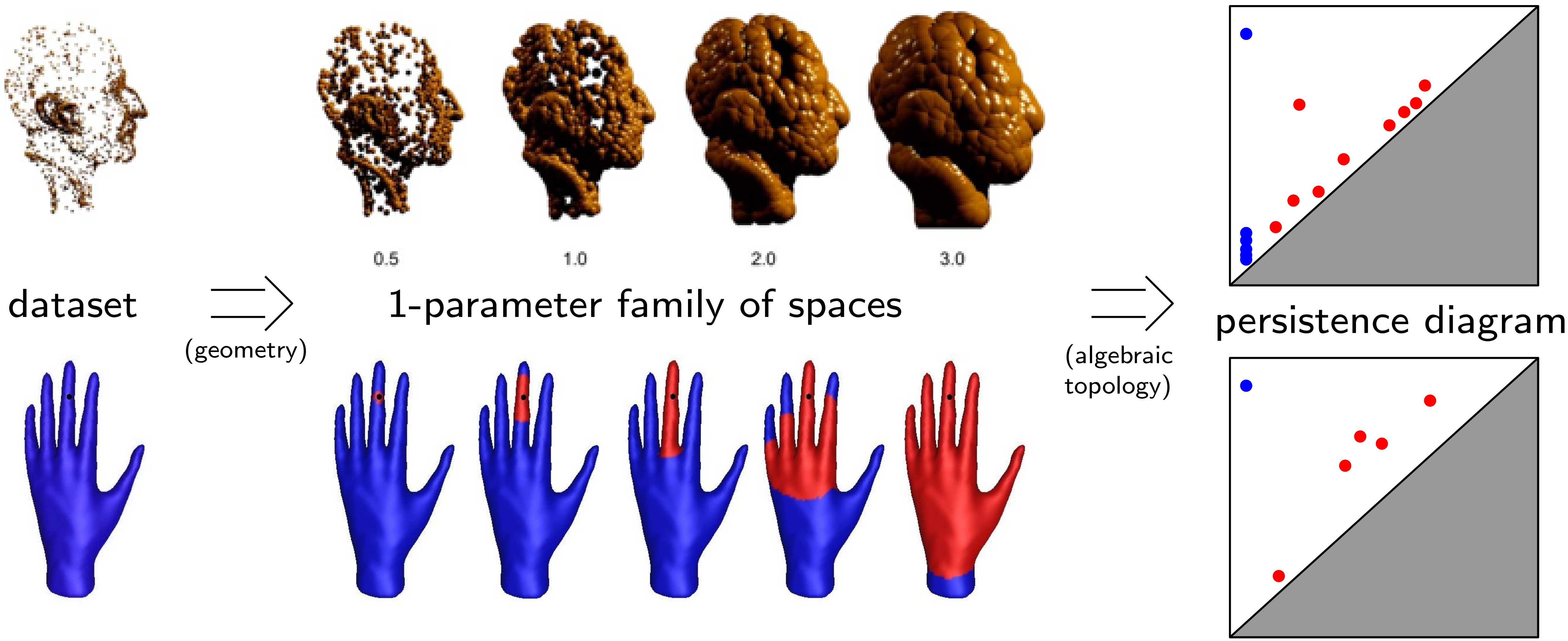
1. Descriptors and stability

2. Vectorizations and kernels

3. Statistics

4. Discrimination power

Persistence diagrams as descriptors for data



Pros:

- strong invariance and stability:

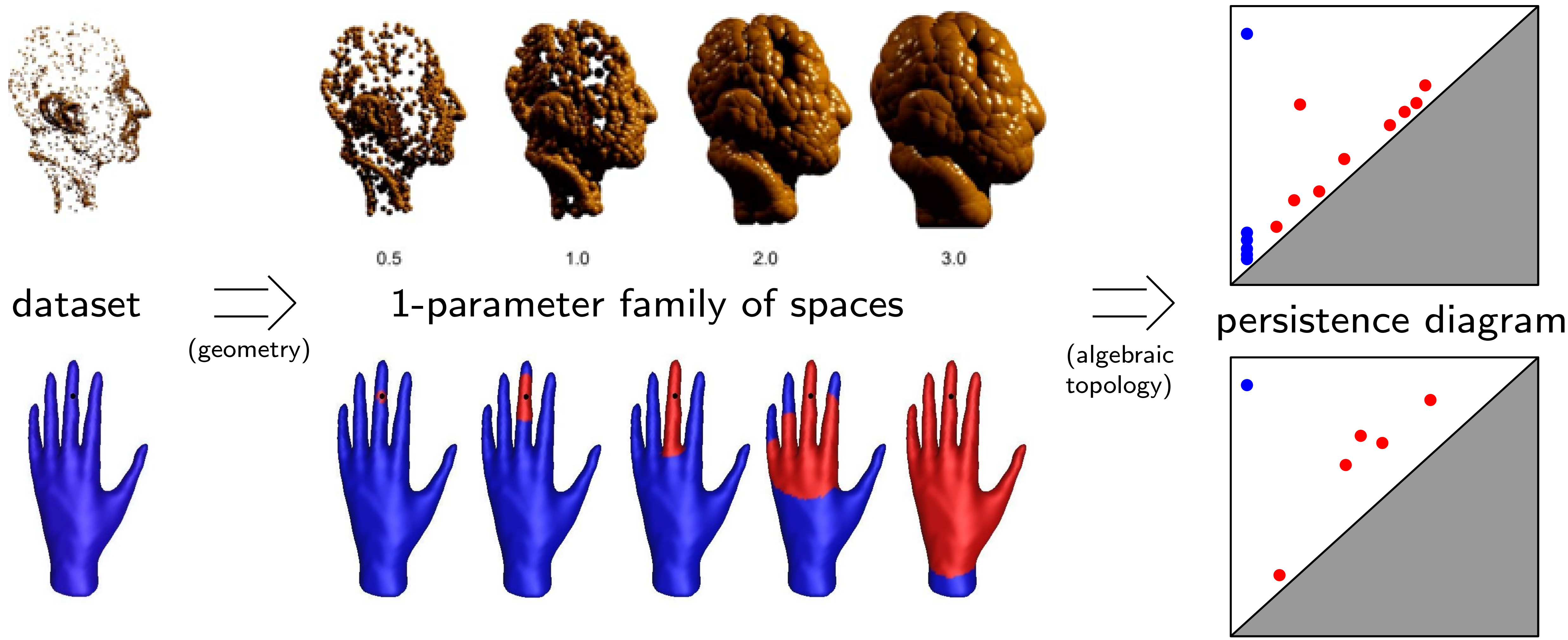
  $\mathrm{d}_\infty(\mathrm{dgm}\, X, \mathrm{dgm}\, Y) \leq \mathrm{cst}\, \mathrm{d}_{\mathrm{GH}}(X, Y)$

- information of a different nature
- flexible and versatile

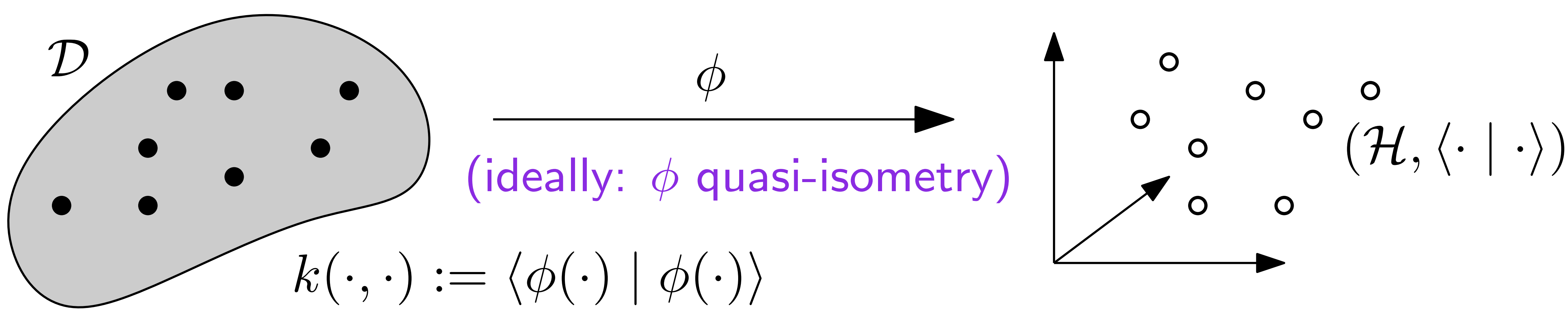
Cons:

- slow to compare
- space of diagrams is not linear
- positive intrinsic curvature

Persistence diagrams as descriptors for data



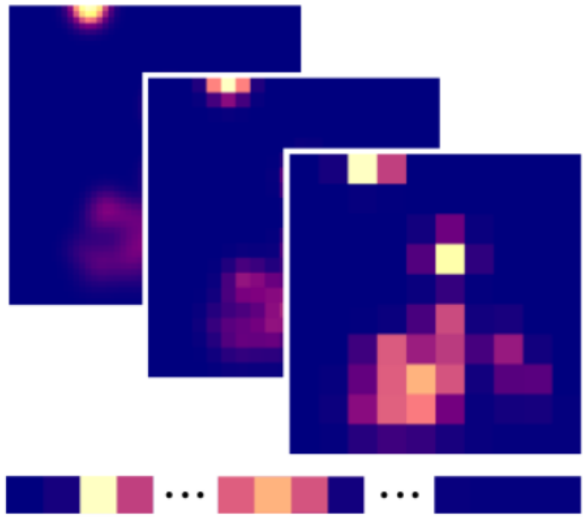
A solution: map diagrams to Hilbert space and use kernel trick



Kernels for persistence diagrams

State of the Art: define ϕ explicitly (vectorization) via:

- images [Adams et al. 2015]

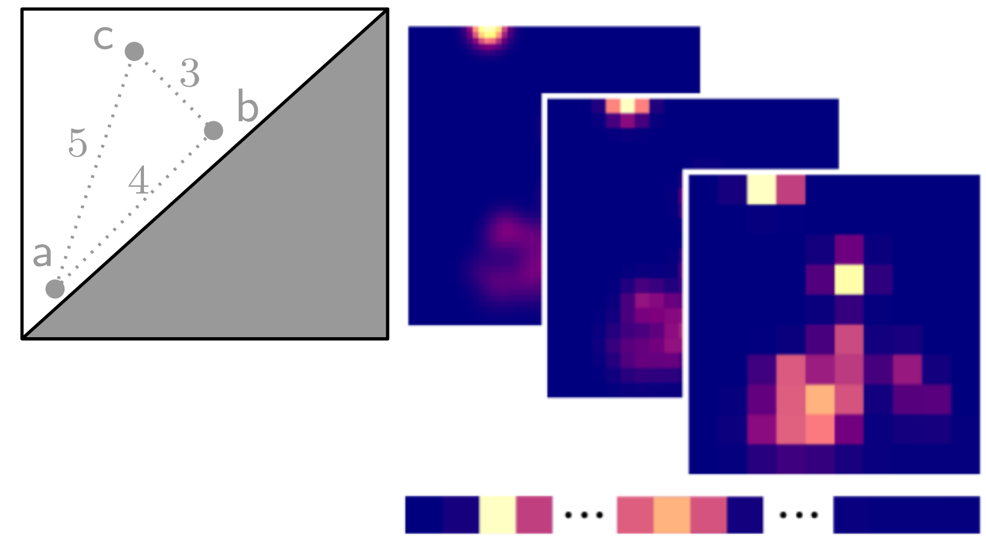


Kernels for persistence diagrams

State of the Art: define ϕ explicitly (vectorization) via:

- images [Adams et al. 2015]
- finite metric spaces [Carrière, O., Ovsjanikov 2015]

$$\begin{array}{cc} & \begin{array}{ccc} a & b & c \end{array} \\ \begin{array}{c} a \\ b \\ c \end{array} & \begin{bmatrix} 0 & 4 & 5 \\ 4 & 0 & 3 \\ 5 & 3 & 0 \end{bmatrix} \end{array}$$



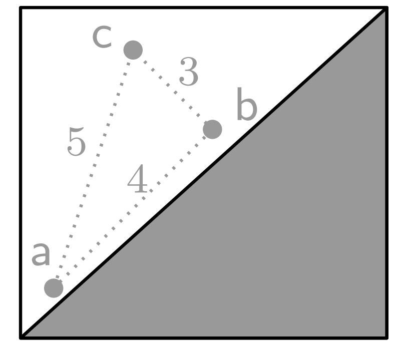
Kernels for persistence diagrams

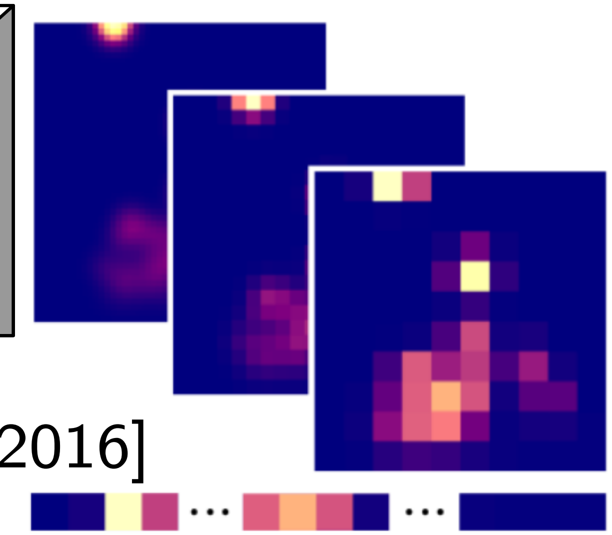
State of the Art: define ϕ explicitly (vectorization) via:

- images [Adams et al. 2015]

- finite metric spaces [Carrière, O., Ovsjanikov 2015]

$$\begin{array}{cc} & \begin{array}{ccc} a & b & c \end{array} \\ \begin{array}{c} a \\ b \\ c \end{array} & \begin{bmatrix} 0 & 4 & 5 \\ 4 & 0 & 3 \\ 5 & 3 & 0 \end{bmatrix} \end{array}$$





- polynomial roots or evaluations [Di Fabio, Ferri 2015] [Kališnik 2016]

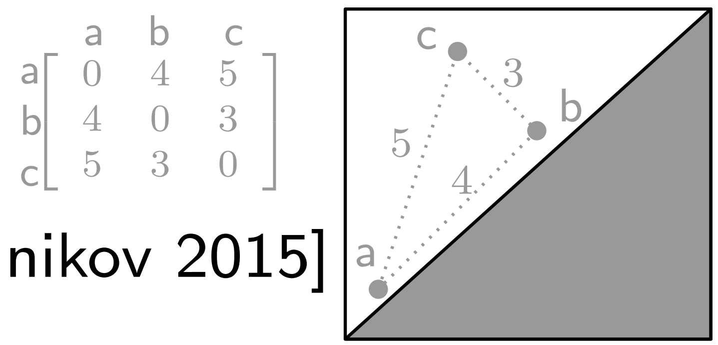
$$\{p_1, \cdots, p_n\} \mapsto (P_1(p_1, \cdots, p_n), \cdots, P_r(p_1, \cdots, p_n), \cdots)$$

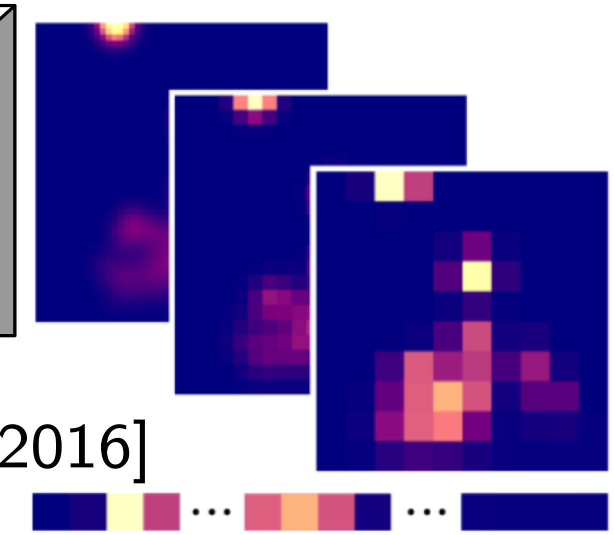
Kernels for persistence diagrams

State of the Art: define ϕ explicitly (vectorization) via:

- images [Adams et al. 2015]

- finite metric spaces [Carrière, O., Ovsjanikov 2015]

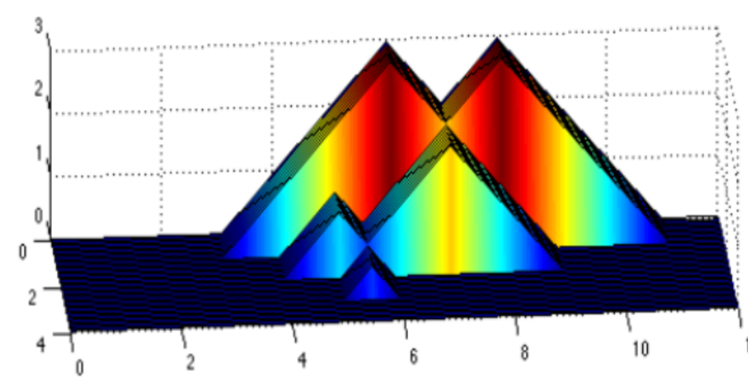




- polynomial roots or evaluations [Di Fabio, Ferri 2015] [Kališnik 2016]

$$\{p_1, \cdots, p_n\} \mapsto (P_1(p_1, \cdots, p_n), \cdots, P_r(p_1, \cdots, p_n), \cdots)$$

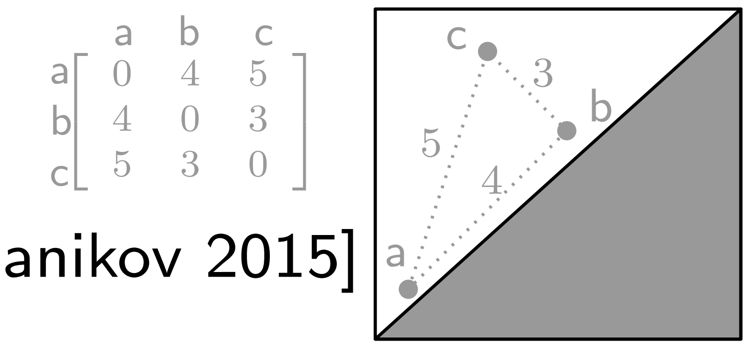
- landscapes [Bubenik 2012] [Bubenik, Dłotko 2015]



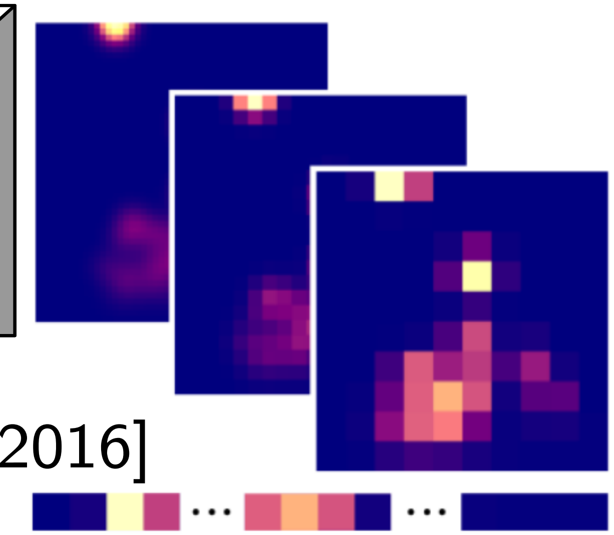
Kernels for persistence diagrams

State of the Art: define ϕ explicitly (vectorization) via:

- images [Adams et al. 2015]



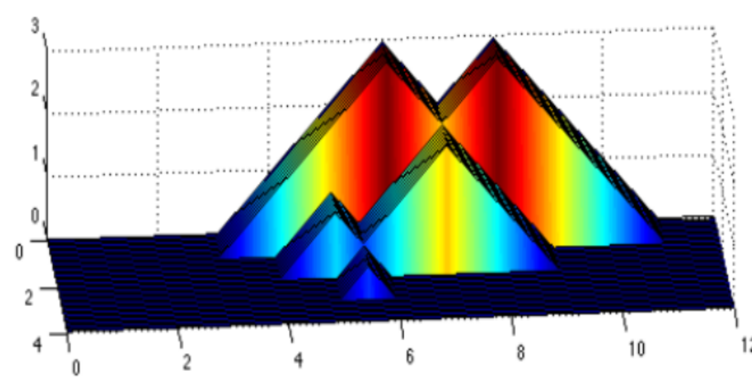
- finite metric spaces [Carrière, O., Ovsjanikov 2015]



- polynomial roots or evaluations [Di Fabio, Ferri 2015] [Kališnik 2016]

$$\{p_1, \cdots, p_n\} \mapsto (P_1(p_1, \cdots, p_n), \cdots, P_r(p_1, \cdots, p_n), \cdots)$$

- landscapes [Bubenik 2012] [Bubenik, Dłotko 2015]



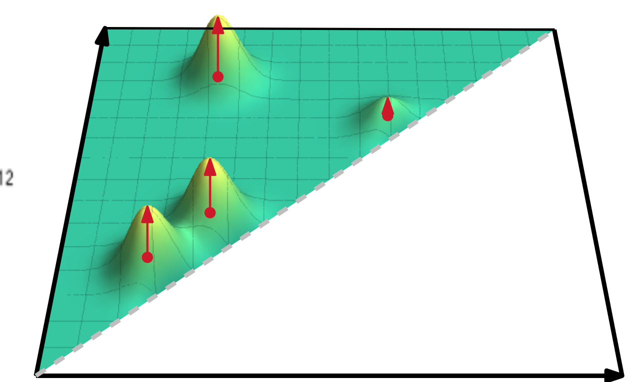
- discrete measures:

  → histogram [Bendich et al. 2014]

  → regularize optimal transport [Carrière, Cuturi, O. 2017]

  → convolution with weighted kernel [Kusano, Fukumisu, Hiraoka 2016-17]

  → heat diffusion [Reininghaus et al. 2015] + exponential [Kwit et al. 2015]



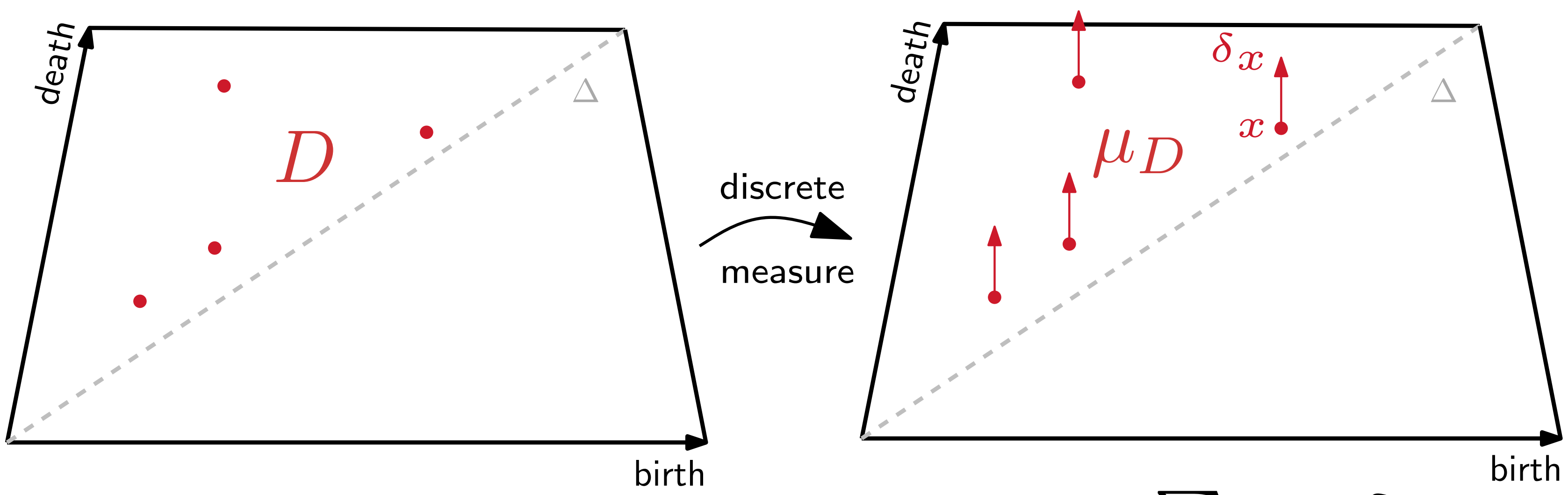
Kernels for persistence diagrams

	images	metric spaces	polynomials	landscapes	discrete measures
ambient Hilbert space	$(\mathbb{R}^d, \lVert.\rVert_2)$	$(\mathbb{R}^d, \lVert.\rVert_2)$	$\ell_2(\mathbb{R})$	$L_2(\mathbb{N} \times \mathbb{R})$	$L_2(\mathbb{R}^2)$
positive (semi-)definiteness	✓	✓	✓	✓	✓
$\lVert\phi(\cdot) - \phi(\cdot)\rVert_{\mathcal{H}} \leq g(\mathrm{d}_p)$	✓	✓	✓	✓	✓
$\lVert\phi(\cdot) - \phi(\cdot)\rVert_{\mathcal{H}} \geq f(\mathrm{d}_p)$	✗	✗	✗	✗	✓
injectivity	✗	✗	✓	✓	✓
universality	✗	✗	✗	✗	✓
algorithmic cost	f. map: $O(n^2)$ kernel: $O(d)$	f. map: $O(n^2)$ kernel: $O(d)$	f. map: $O(nd)$ kernel: $O(d)$	$O(n^2)$	$O(n^2)$

Kernels for persistence diagrams

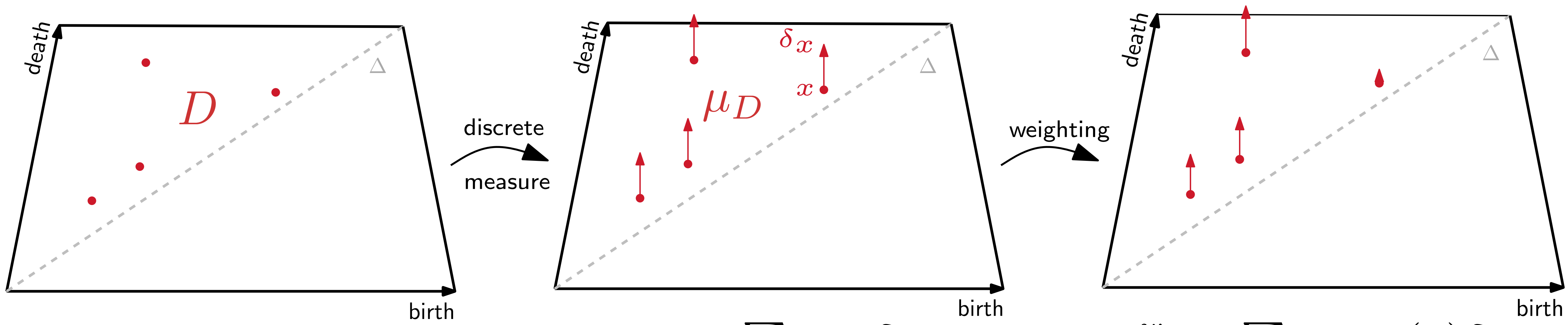
	images	metric spaces	polynomials	landscapes	discrete measures
ambient Hilbert space	$(\mathbb{R}^d, \|.\|_2)$	$(\mathbb{R}^d, \|.\|_2)$	$\ell_2(\mathbb{R})$	$L_2(\mathbb{N} \times \mathbb{R})$	$L_2(\mathbb{R}^2)$
positive (semi-)definiteness	✓	✓	✓	✓	✓
$\|\phi(\cdot) - \phi(\cdot)\|_{\mathcal{H}} \leq g(\mathrm{d}_p)$	✓	✓	✓	✓	✓
$\|\phi(\cdot) - \phi(\cdot)\|_{\mathcal{H}} \geq f(\mathrm{d}_p)$	✗	✗	✗	✗	✓
injectivity	✗	✗	✓	✓	✓
universality	✗	✗	✗	✗	✓
algorithmic cost	f. map: $O(n^2)$ kernel: $O(d)$	f. map: $O(n^2)$ kernel: $O(d)$	f. map: $O(nd)$ kernel: $O(d)$	$O(n^2)$	$O(n^2)$

Persistence diagrams as discrete measures (I)



$$\mu_D := \sum_{x \in D} \delta_x$$

Persistence diagrams as discrete measures (I)



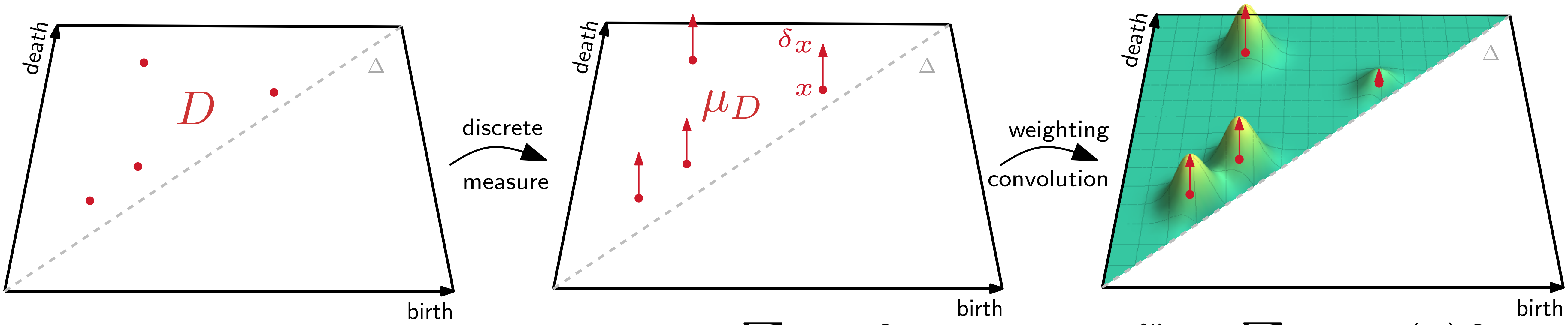
$\mu_D := \sum_{x \in D} \delta_x$

$\mu_D^w := \sum_{x \in D} w(x)\delta_{\bar{x}}$

Pb: μ_D is unstable (points on diagonal disappear)

$$w(x) := \arctan\left(c\,\mathrm{d}(x,\Delta)^r\right),\ c, r > 0$$

Persistence diagrams as discrete measures (I)



$\mu_D := \sum_{x \in D} \delta_x$

$\mu_D^w := \sum_{x \in D} w(x) \delta_{\bar{x}}$

$\tilde{\mu}_D^w := \mu_D^w * \mathcal{N}(0, \sigma)$

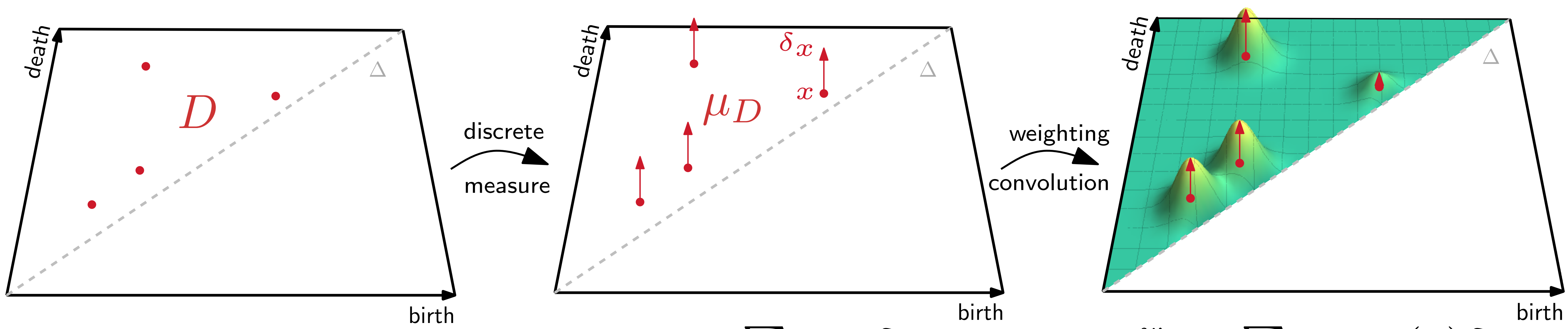
Pb: μ_D is unstable (points on diagonal disappear)

$w(x) := \arctan\left(c\, \mathrm{d}(x, \Delta)^r\right),\ c, r > 0$

Def: $\phi(D)$ is the density function of $\mu_D^w * \mathcal{N}(0, \sigma)$ w.r.t. Lebesgue measure

$$\begin{cases} \phi(D) := \dfrac{1}{\sqrt{2\pi}\sigma} \displaystyle\sum_{x \in D} \arctan(c\, \mathrm{d}(x, \Delta)^r) \exp\left(-\dfrac{\| \cdot - x\|^2}{2\sigma^2}\right) \\ k(D, D') := \langle \phi(D),\ \phi(D') \rangle_{L_2(\Delta \times \mathbb{R}_+)} \end{cases}$$

Persistence diagrams as discrete measures (I)



$\mu_D := \sum_{x \in D} \delta_x$

$\mu_D^w := \sum_{x \in D} w(x) \delta_{\bar{x}}$

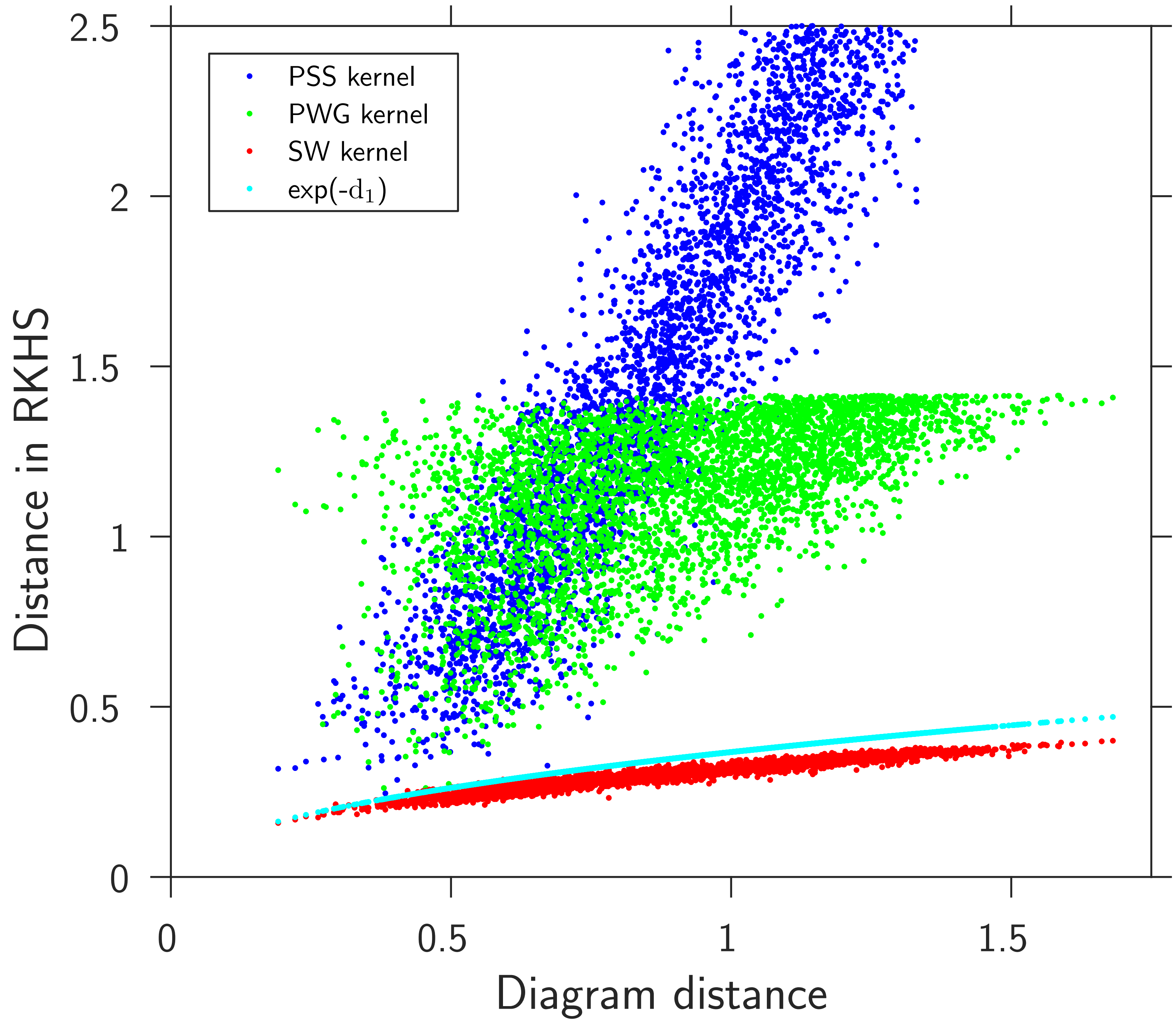
$\tilde{\mu}_D^w := \mu_D^w * \mathcal{N}(0, \sigma)$

Prop.: [Kusano, Fukumisu, Hiraoka 2016-17]

- $\|\phi(D) - \phi(D')\|_{\mathcal{H}} \le C\, d_p(D, D')$.
- ϕ is injective and $\exp(k)$ is universal

$$\left(\begin{array}{l} \phi(D) := \dfrac{1}{\sqrt{2\pi}\sigma} \displaystyle\sum_{x \in D} \arctan(c\, \mathrm{d}(x, \Delta)^r) \exp\left(-\frac{\| \cdot - x\|^2}{2\sigma^2} \right) \\ k(D, D') := \langle \phi(D),\, \phi(D') \rangle_{L_2(\Delta \times \mathbb{R}_+)} \end{array} \right.$$

Metric distortion in practice

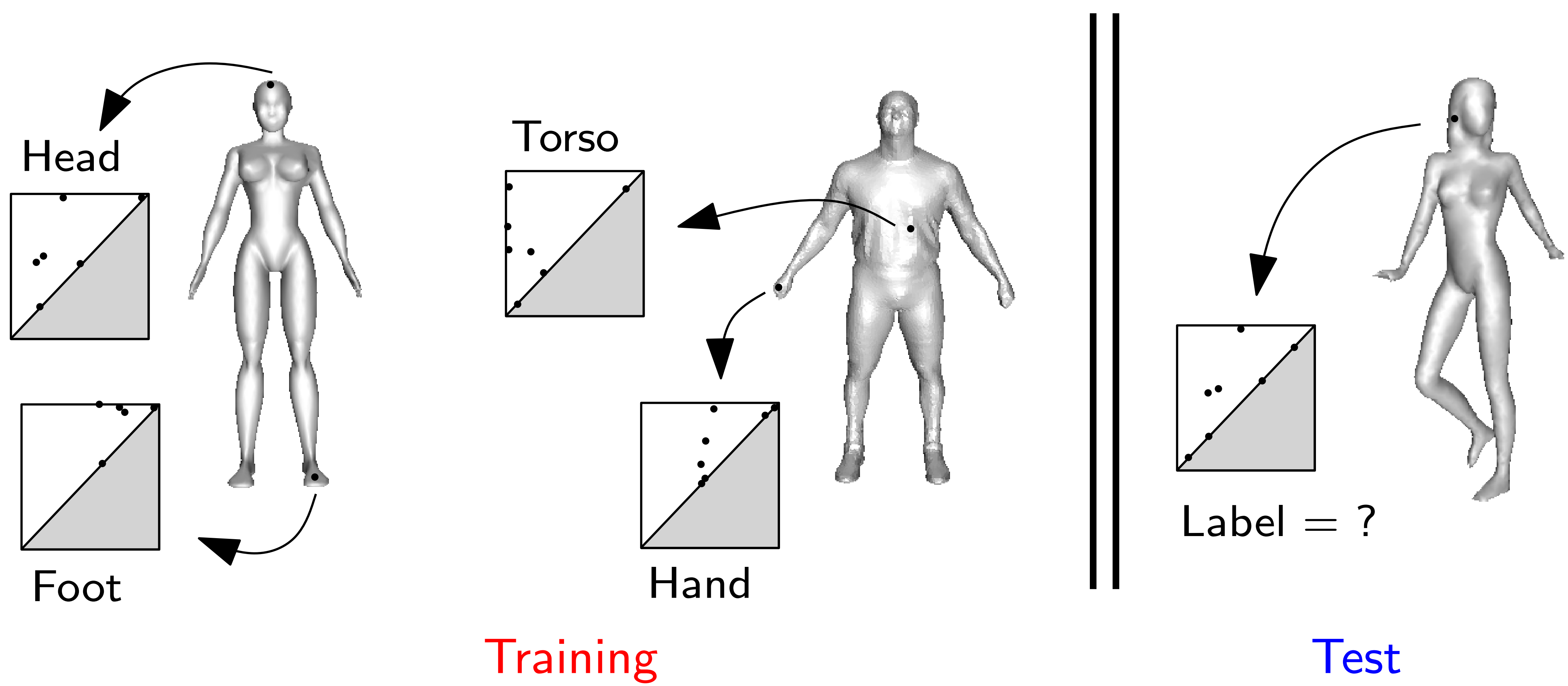


Application to supervised shape segmentation

Goal: segment 3d shapes based on examples

Approach:

- train a (multiclass) classifier on PDs extracted from the training shapes
- apply classifier to PDs extracted from query shape

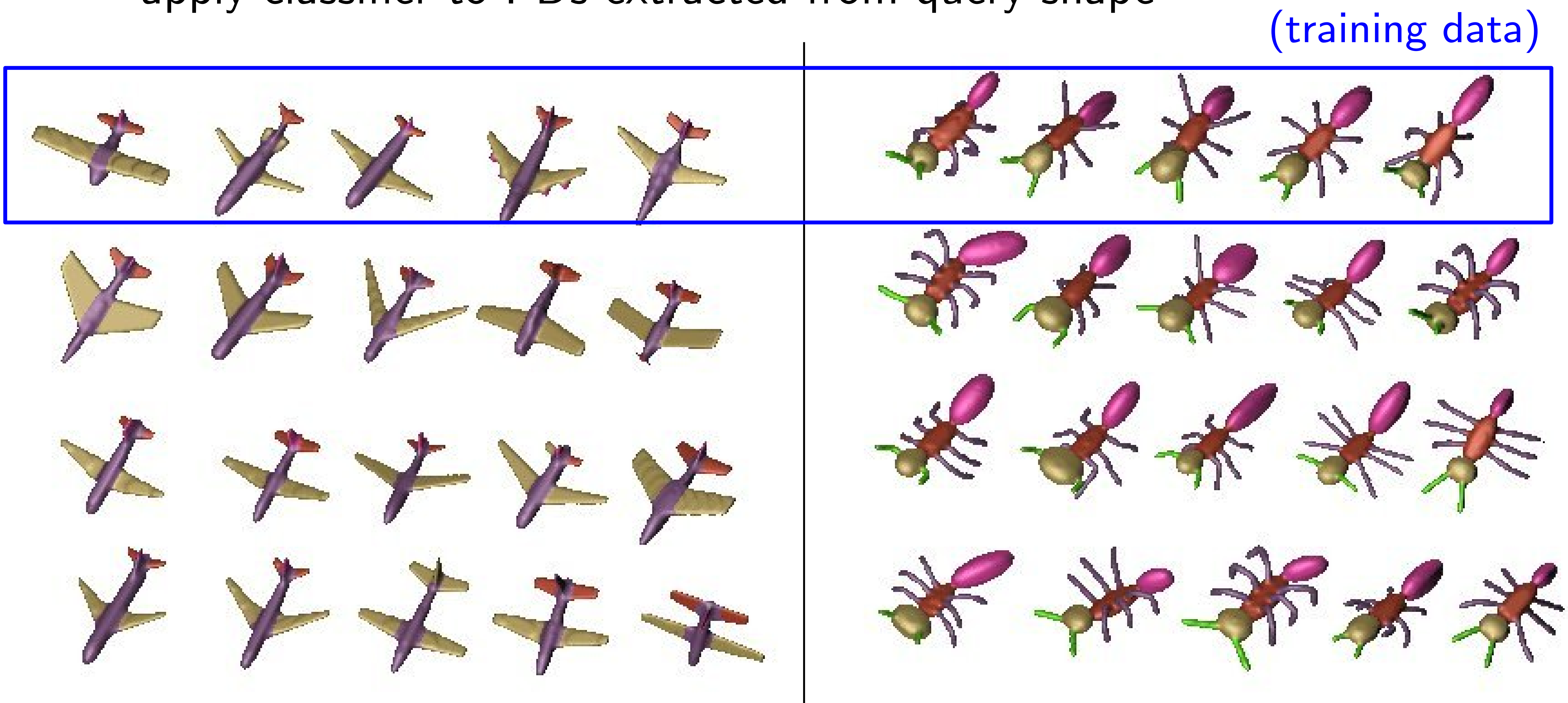


Application to supervised shape segmentation

Goal: segment 3d shapes based on examples

Approach:

- train a (multiclass) classifier on PDs extracted from the training shapes
- apply classifier to PDs extracted from query shape



Application to supervised shape segmentation

Goal: segment 3d shapes based on examples

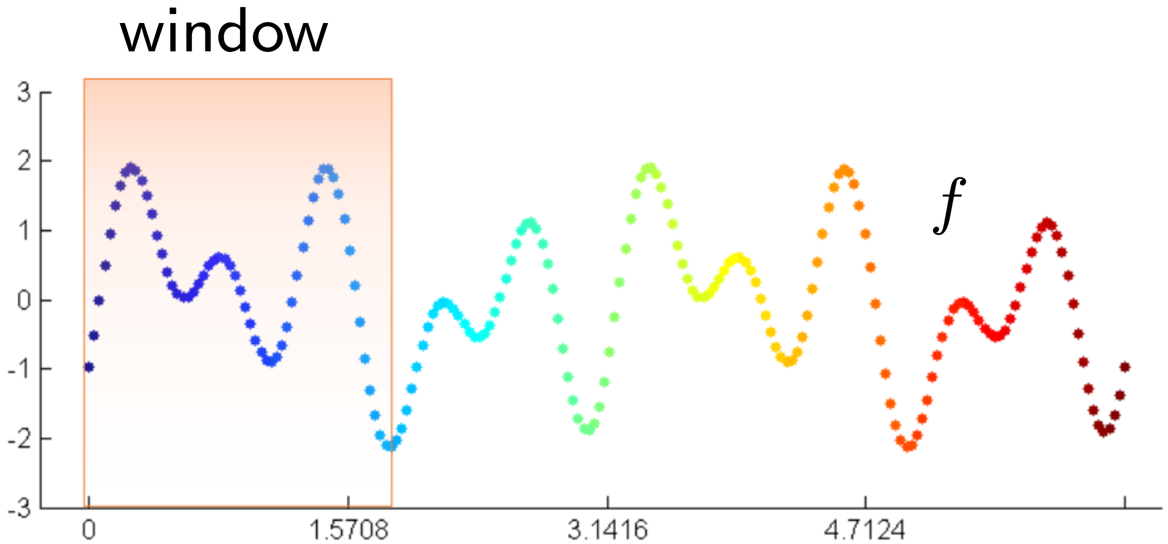
Approach:

- train a (multiclass) classifier on PDs extracted from the training shapes
- apply classifier to PDs extracted from query shape

Accuracies (%) using TDA descriptors (kernels on barcodes):

	TDA	geometry	TDA + geometry
Human	74.0	78.7	**88.7**
Airplane	72.6	81.3	**90.7**
Ant	92.3	90.3	**98.5**
FourLeg	73.0	74.4	**84.2**
Octopus	85.2	94.5	**96.6**
Bird	72.0	75.2	**86.5**
Fish	79.6	79.1	**92.3**

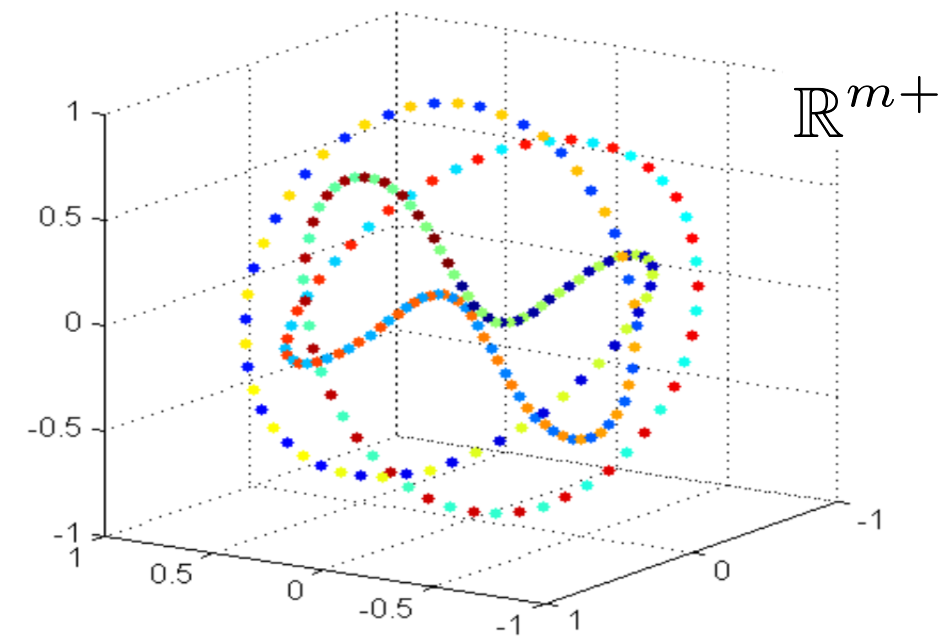
Application to supervised time series analysis



$\mathrm{TD}_{m,\tau}$

$\Longrightarrow$

(time-delay embedding)



$f : \mathbb{N} \to \mathbb{R}$

$$\mathrm{TD}_{m,\tau}(f) := \begin{bmatrix} f(t) \\ f(t+\tau) \\ \vdots \\ f(t+m\tau) \end{bmatrix}$$

τ: step / delay

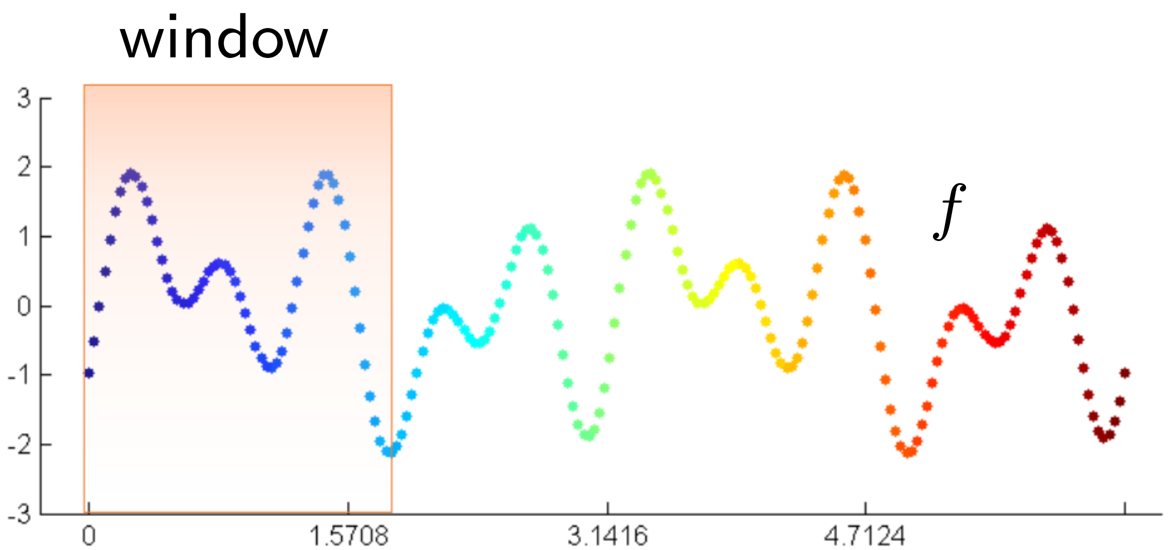
$m\tau$: window size

$m+1$: embedding dimension

signal	**embedded data**
periodicity	circularity
# prominent harmonics (N)	min. ambient dimension ($m \geq 2N$)
# non-commensurate freq.	intrinsic dimension ($\mathbb{S}^1 \times \cdots \times \mathbb{S}^1$)

[J. Perea et al.:"SW1PerS: Sliding windows and 1-persistence scoring", 2015]

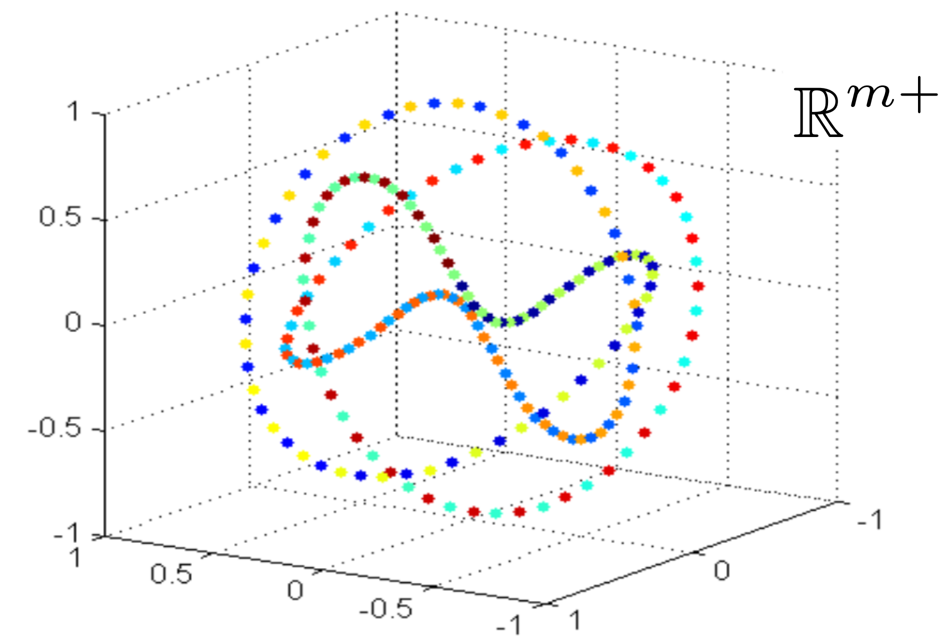
Application to supervised time series analysis



$\mathrm{TD}_{m,\tau}$

$\Longrightarrow$

(time-delay embedding)



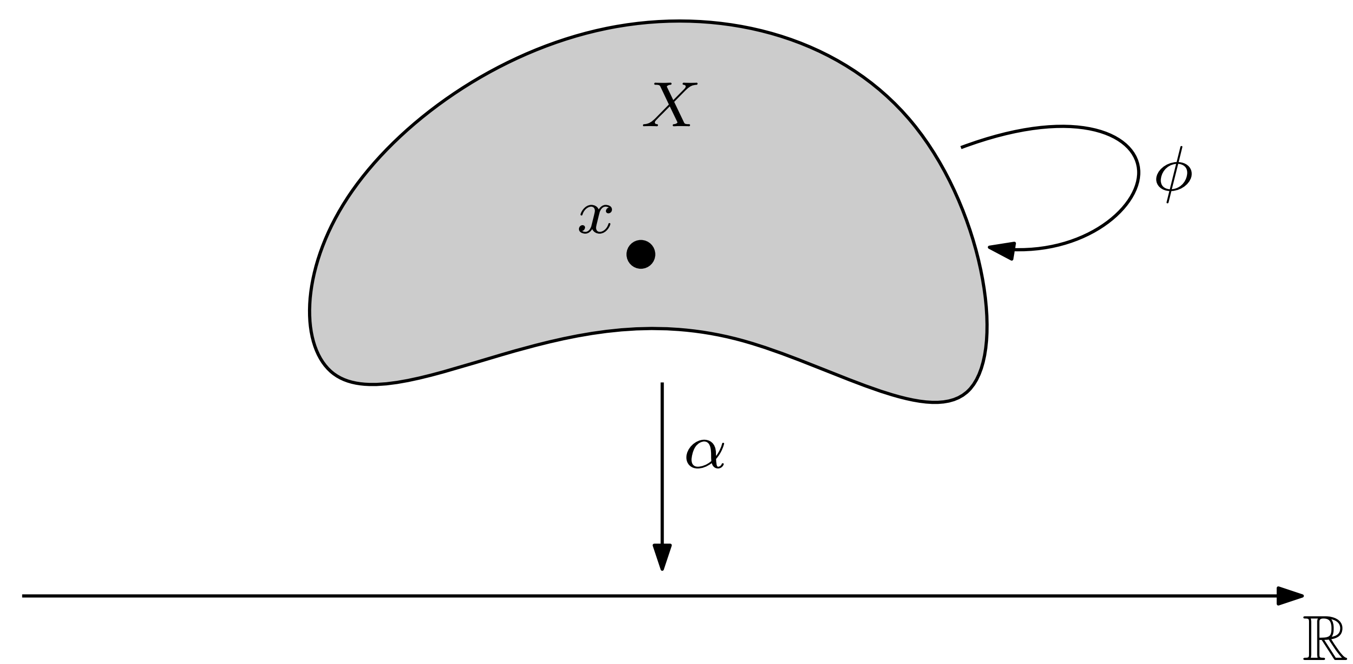
Contributions of TDA:

inference of:

- periodicity
- harmonics
- non-commensurate freq.
- underlying state space

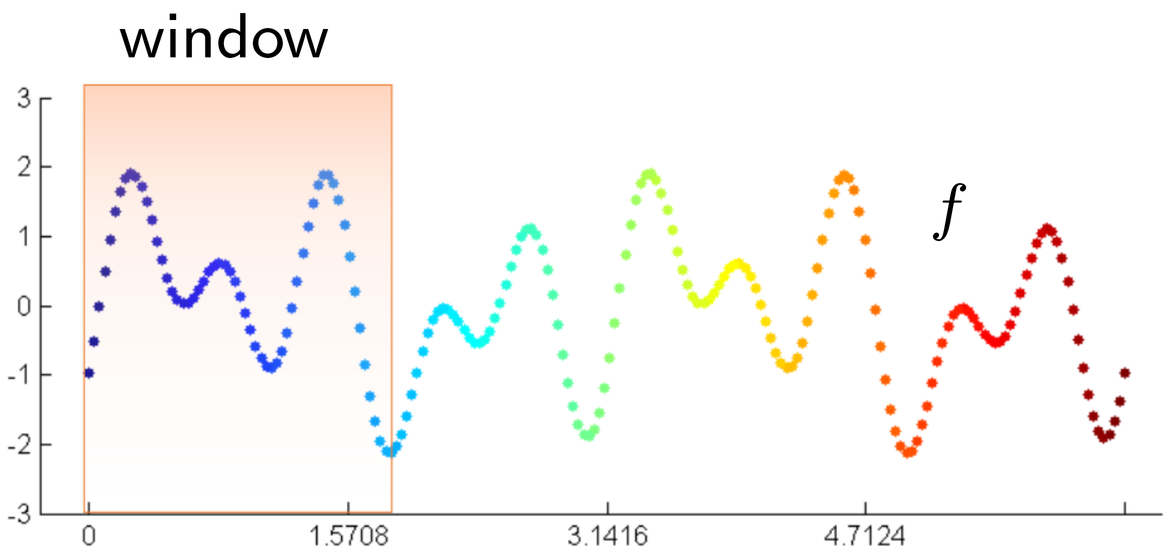
no Fourier transform needed

Dynamical system:



$$f_x(n) := \alpha(\phi^n(x))$$

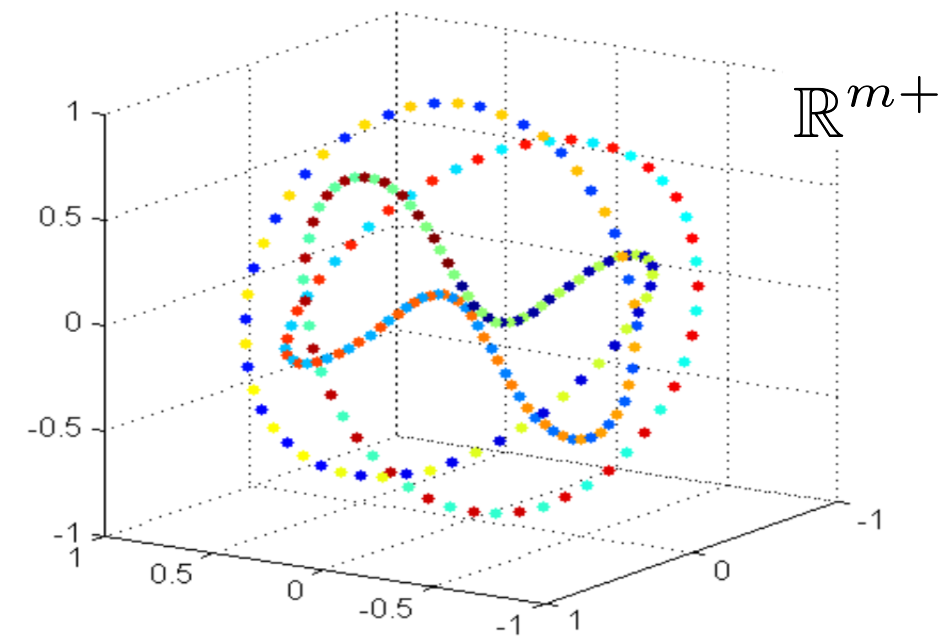
Application to supervised time series analysis



$\mathrm{TD}_{m,\tau}$

$\Longrightarrow$

(time-delay embedding)



Contributions of TDA:

inference of:

- periodicity
- harmonics
- non-commensurate freq.
- underlying state space

no Fourier transform needed

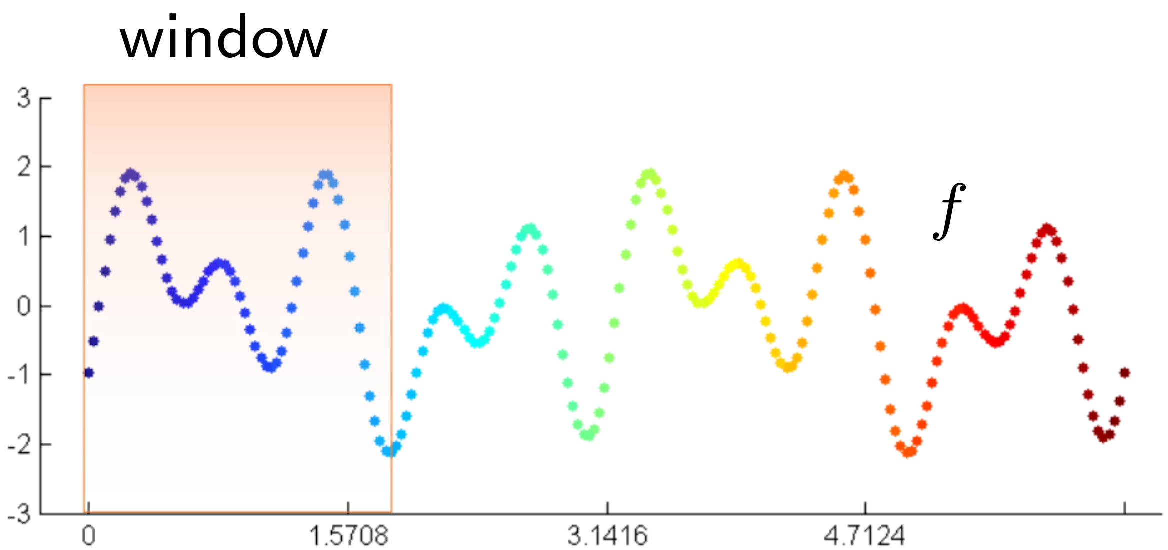
Dynamical system:

Thm: [Nash, Takens]
Given a Riemannian manifold X of dimension $\frac{m}{2}$, it is a **generic property** of $\phi \in \mathrm{Diff}_2(X)$ and $\alpha \in C^2(X, \mathbb{R})$ that

$$X \to \mathbb{R}^{m+1}$$
$$x \mapsto (\alpha(x), \alpha \circ \phi(x), \cdots, \alpha \circ \phi^m(x))$$

is an embedding.

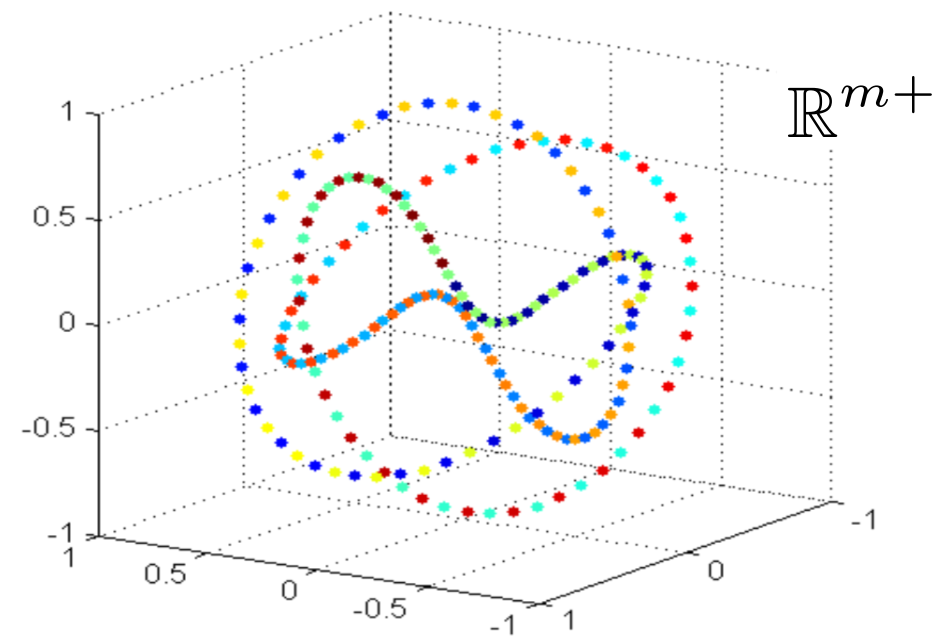
Application to supervised time series analysis



$\mathrm{TD}_{m,\tau}$

$\Longrightarrow$

(time-delay embedding)



method / dataset	Gyro sensor	EEG dataset	EMG dataset
SVM + statistical features	67.6 ± 4.7	44.4 ± 19.8	15.0 ± 10.0
SVM + Betti sequence	63.5 ± 11.3	66.7 ± 5.6	49.6 ± 18.2
1-d CNN + dynamic time warping	6.4 ± 5.1	72.4 ± 6.1	15.0 ± 10.0
imaging CNN	18.9 ± 5.2	48.9 ± 4.2	10.0 ± 0.0
1-d CNN + Betti sequence	$\mathbf{79.8} \pm 5.0$	$\mathbf{75.38} \pm 5.7$	$\mathbf{74.4} \pm 10.6$

[Y. Umeda:"Time Series Classification via Topological Data Analysis", 2017]

Outline

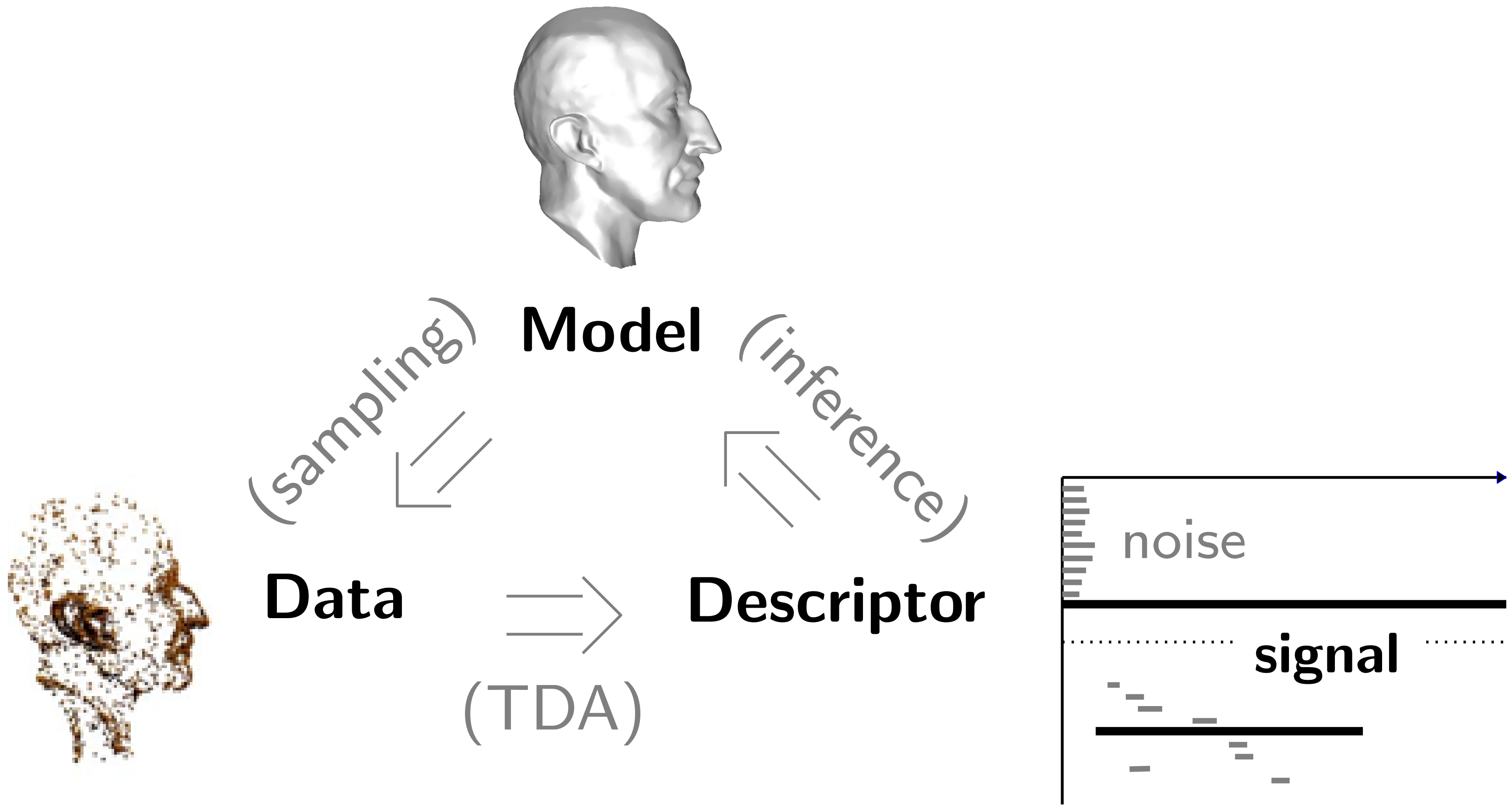
1. Descriptors and stability

2. Vectorizations and kernels

3. Statistics

4. Discrimination power

Statistics for persistence diagrams



Statistics:

- signal vs noise discrimination
- convergence rates
- confidence indices/intervals, principal components, etc.

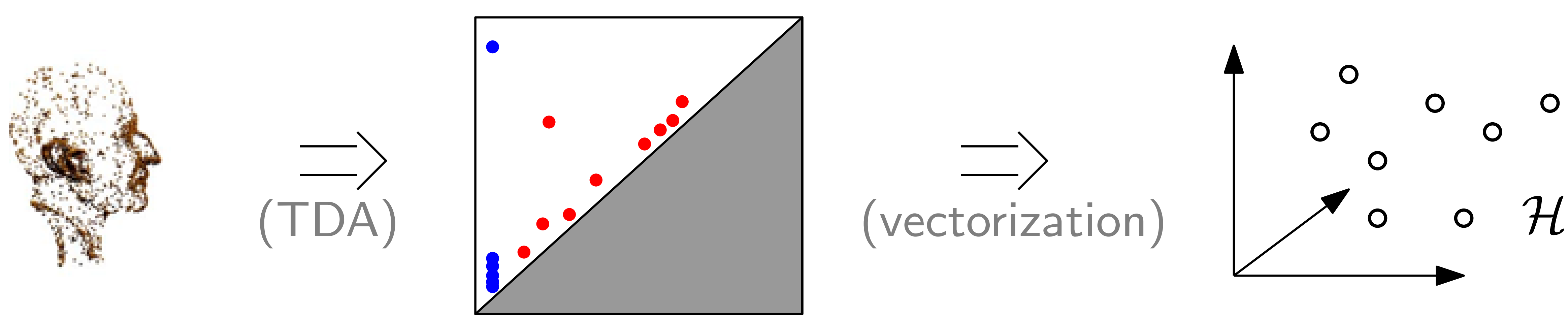
Statistics for persistence diagrams

3 approaches for statistics:

- Fréchet means in diagrams space

- embedding into Hilbert spaces

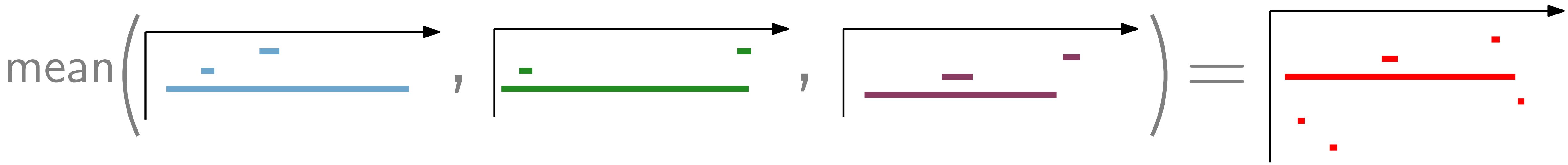
- push-forwards from data space

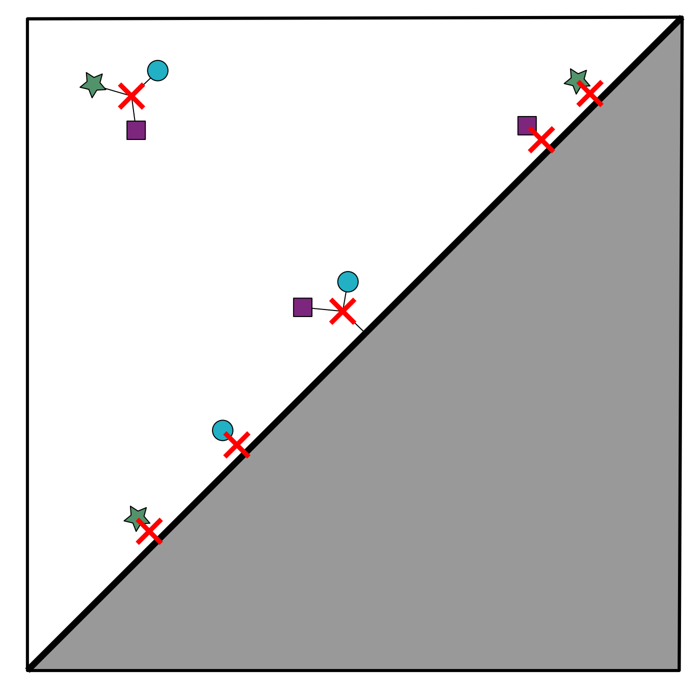


1. Fréchet means in diagrams space

Means as a gateway to statistical analysis

central limit theorems, confidence intervals, geodesic PCA,
clustering (k-means, EM, Mean-Shift, etc.)

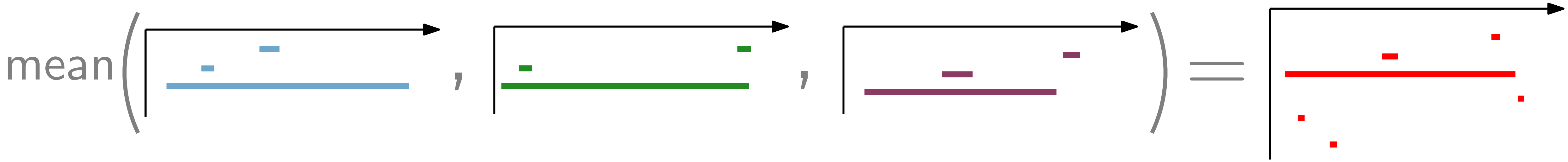




1. Fréchet means in diagrams space

Means as a gateway to statistical analysis

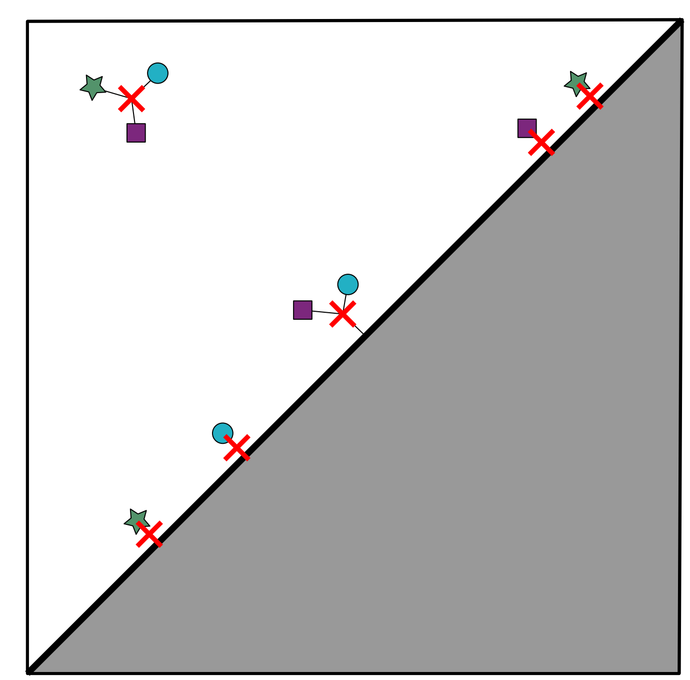
central limit theorems, confidence intervals, geodesic PCA,
clustering (k-means, EM, Mean-Shift, etc.)



No coordinates $\rightsquigarrow$ means as minimizers of variance (Fréchet means)

Given diagrams $D_1, \cdots, D_n$:

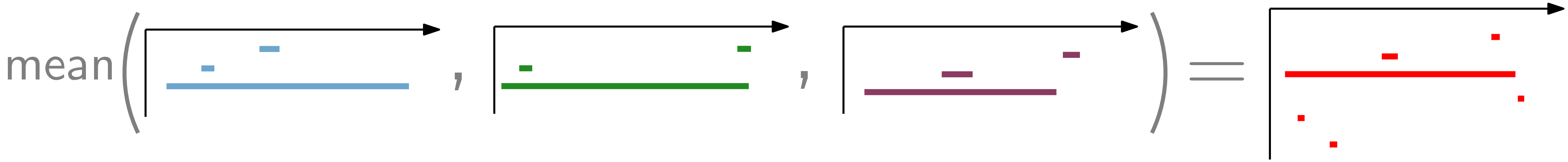
$$\bar{D} \in \arg\min_D \frac{1}{n} \sum_i \mathrm{d}_p(D,\, D_i)^2$$



1. Fréchet means in diagrams space

Means as a gateway to statistical analysis

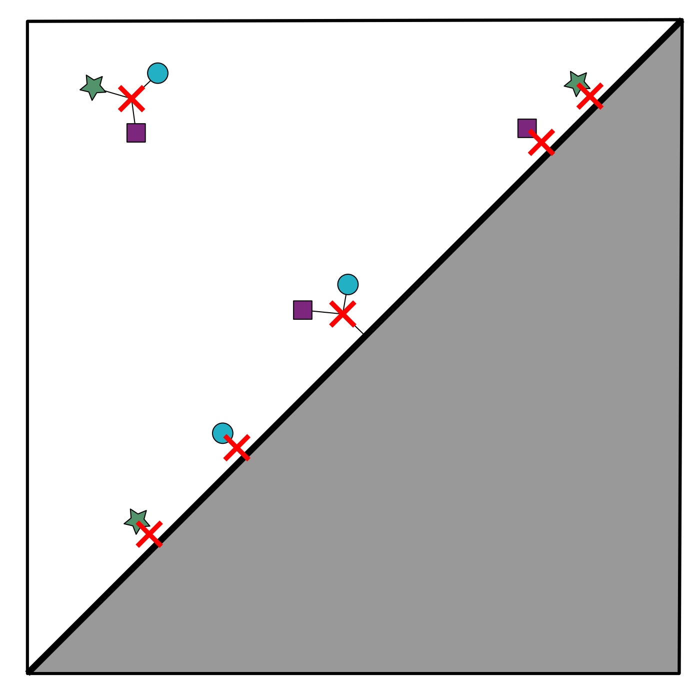
central limit theorems, confidence intervals, geodesic PCA,
clustering (k-means, EM, Mean-Shift, etc.)



No coordinates $\rightsquigarrow$ means as minimizers of variance (Fréchet means)

Given diagrams $D_1, \cdots, D_n$:

$$\bar{D} \in \arg\min_{D} \frac{1}{n} \sum_i \mathrm{d}_p(D, D_i)^2$$



Problem: non-convex energy, highly curved space

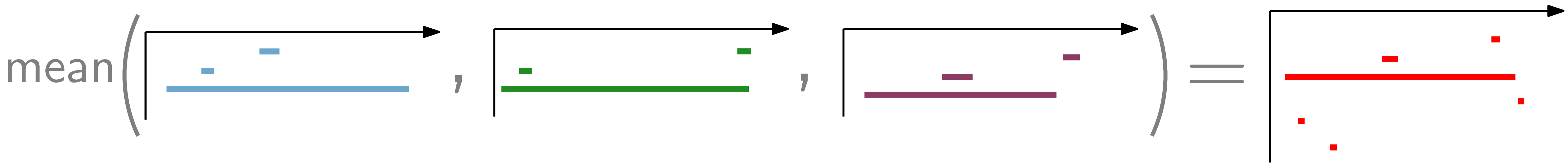
$\Rightarrow$ $\arg\min$ not unique, local minima, numerical issues

[K. Turner et al.: "Fréchet means for distributions of persistence diagrams", 2012]

1. Fréchet means in diagrams space

Means as a gateway to statistical analysis

central limit theorems, confidence intervals, geodesic PCA,
clustering (k-means, EM, Mean-Shift, etc.)



No coordinates $\rightsquigarrow$ means as minimizers of variance (Fréchet means)

Given diagrams $D_1, \cdots, D_n$:

$$\bar{D} \in \arg\min_{D} \frac{1}{n} \sum_i \mathrm{d}_p(D,\, D_i)^2$$

barcode distance is a transportation type distance $\rightsquigarrow$ connection to Optimal Transport

Problem: non-convex energy, highly curved space

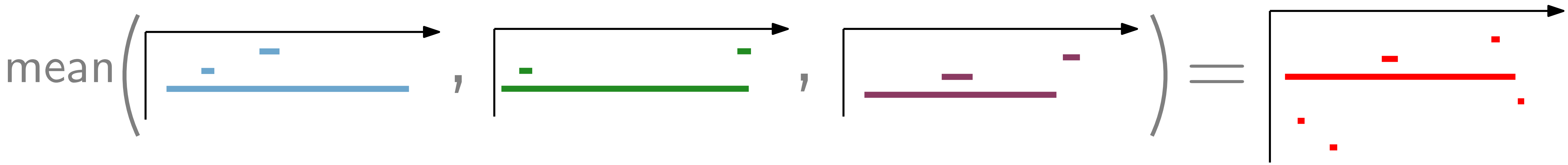
$\Rightarrow$ $\arg\min$ not unique, local minima, numerical issues

[K. Turner et al.: "Fréchet means for distributions of persistence diagrams", 2012]

1. Fréchet means in diagrams space

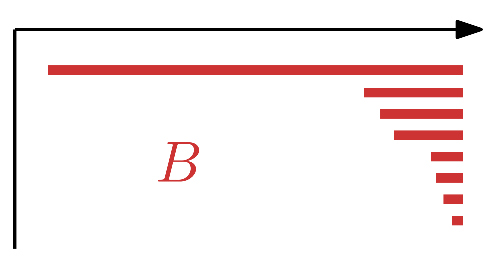
Means as a gateway to statistical analysis

central limit theorems, confidence intervals, geodesic PCA, clustering (k-means, EM, Mean-Shift, etc.)

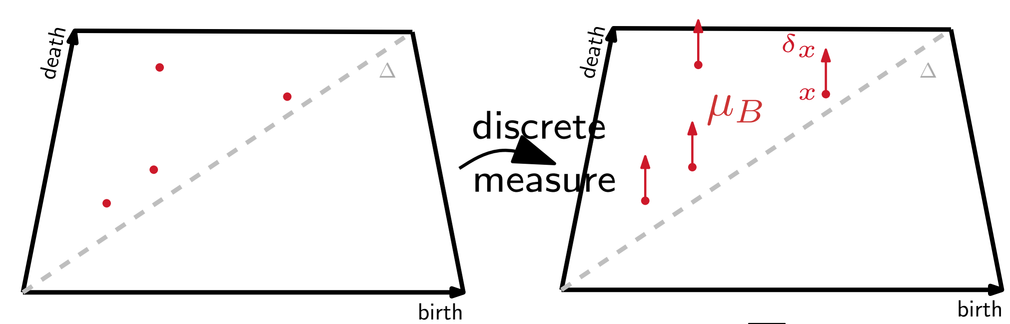


New approach: recast problem in measure space

$B \;\mapsto\; \mu_B$



$\Longrightarrow$



$\mu_D := \sum_{x \in D} \delta_x$

$\rightsquigarrow$ use relaxations from Optimal Transport (OT):

measures: $\mu_B \mapsto \mu_B * \mathcal{U}_{[0,\varepsilon]^2}$

[M. Agueh, G. Carlier: "Barycenters in the Wasserstein Space", 2011]

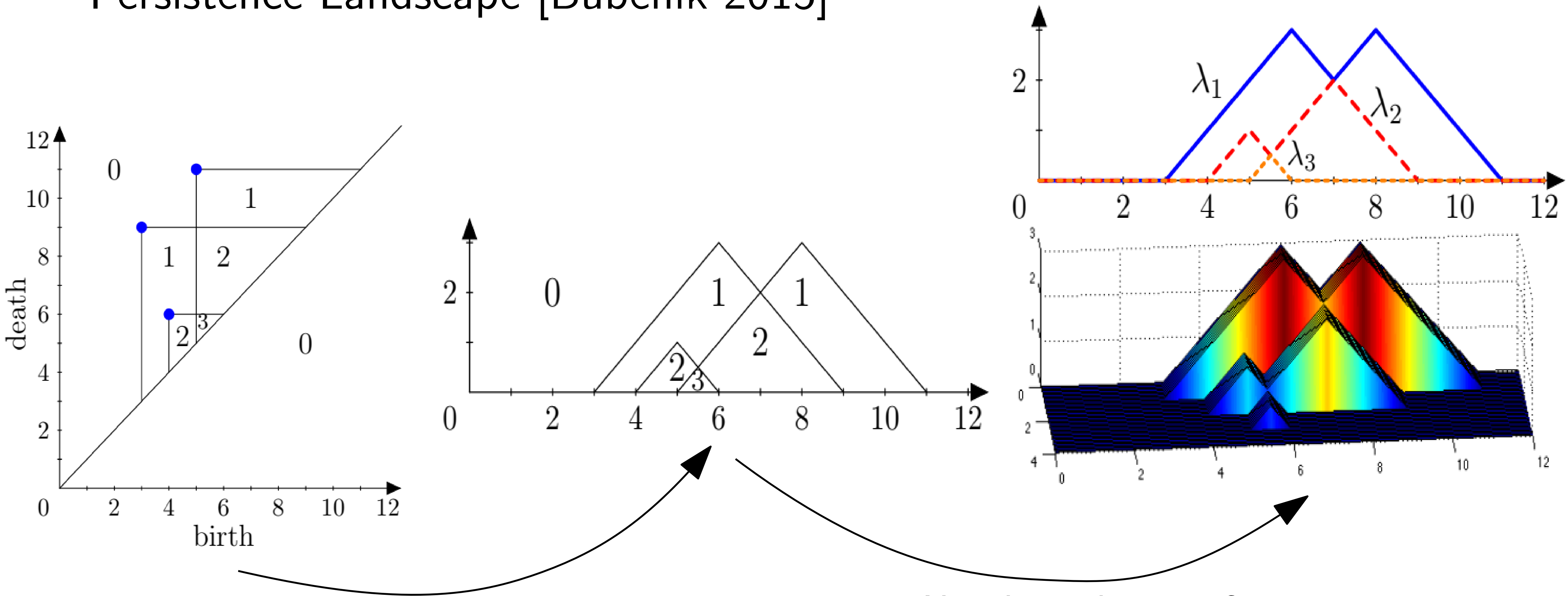
metric: $W_{2,\gamma}(\mu_{B_i}, \mu_{B_j})^2 := \inf_{\nu} \int \|x-y\|^2 d\nu(x,y) + \gamma\, H(\nu)$

[M. Cuturi, A. Doucet: "Fast computation of Wasserstein barycenters", 2014]

strictly convex problem
$\Rightarrow$ unique mean
easy to compute

2. Hilbert space embedding

Persistence Landscape [Bubenik 2015]

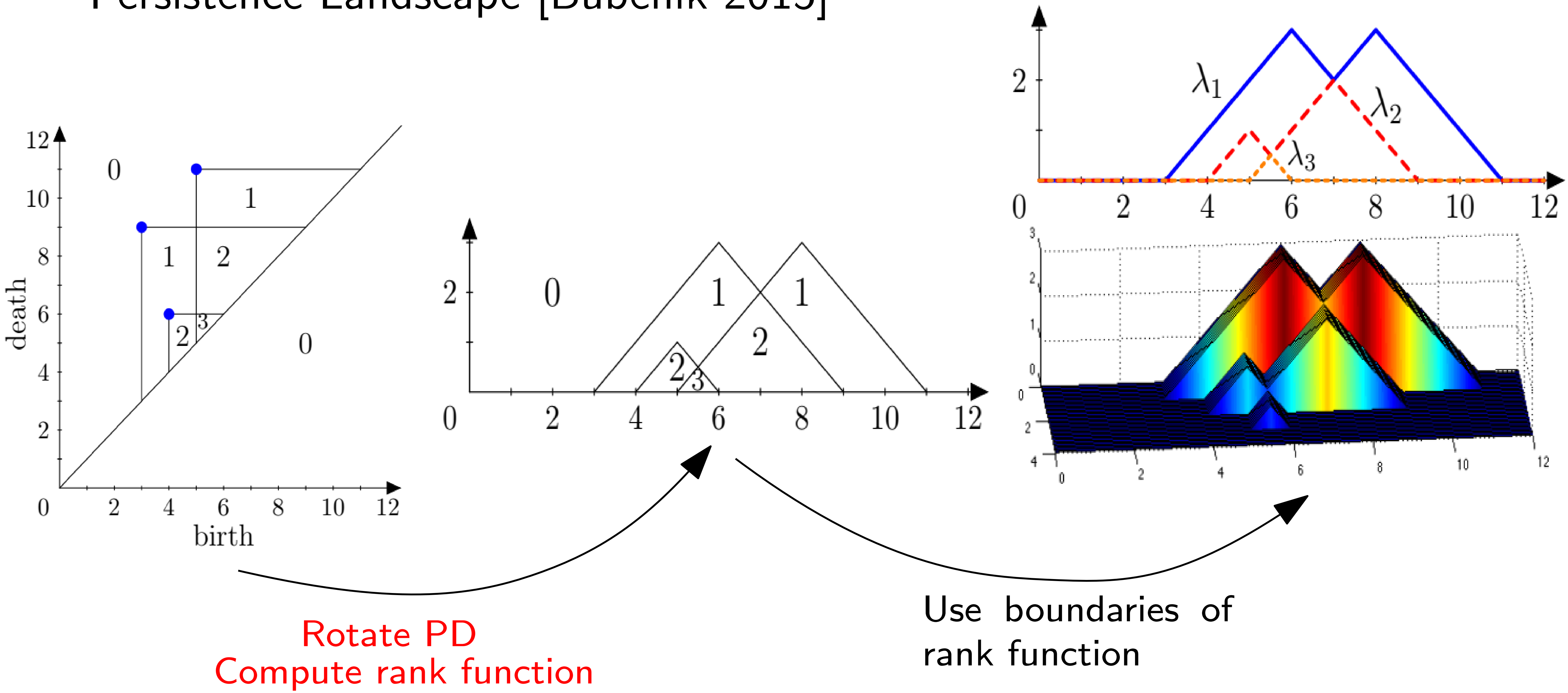


Rotate PD
Compute rank function

Use boundaries of rank function

2. Hilbert space embedding

Persistence Landscape [Bubenik 2015]



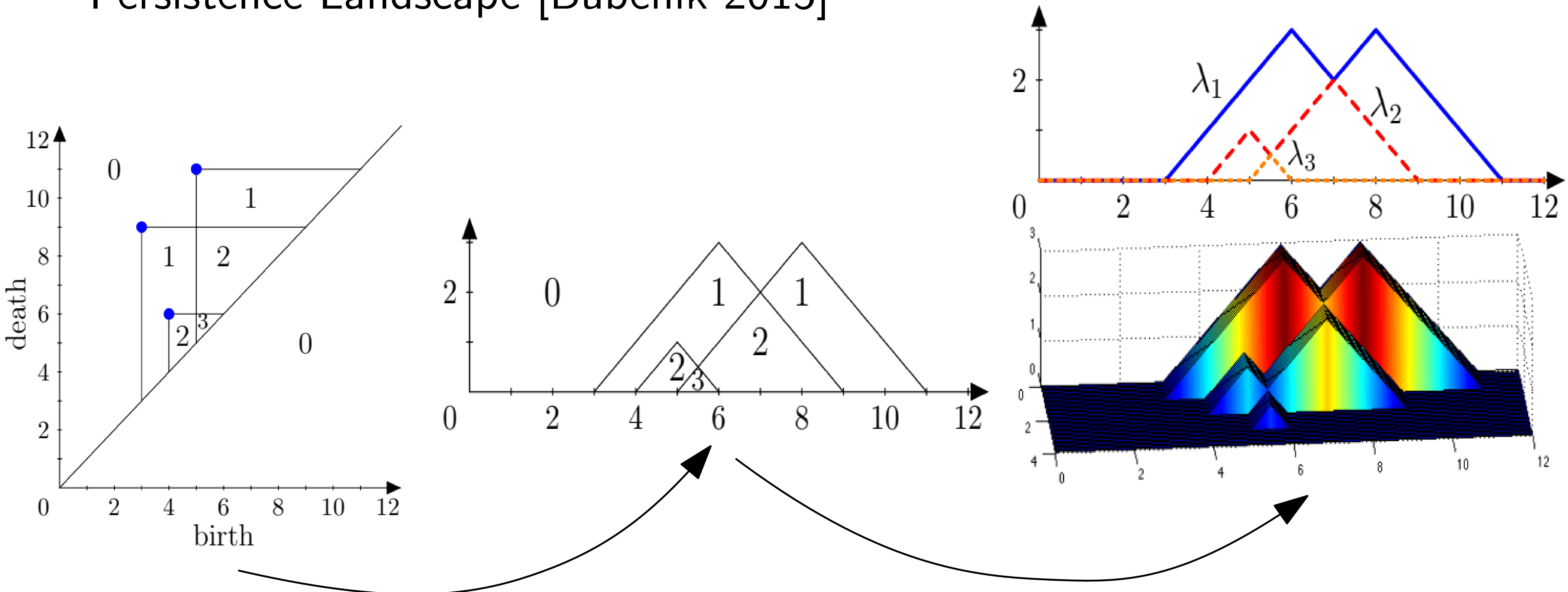
$$x \le y \Longrightarrow f^{-1}(-\infty, x) \subseteq f^{-1}(-\infty, y)$$

$\iota_x^y : H(f^{-1}(-\infty, x)) \to H(f^{-1}(-\infty, y))$ induced linear map

Rank function is defined as $\lambda(x, y) = \operatorname{rank} \iota_x^y$

2. Hilbert space embedding

Persistence Landscape [Bubenik 2015]



Rotate PD
Compute rank function

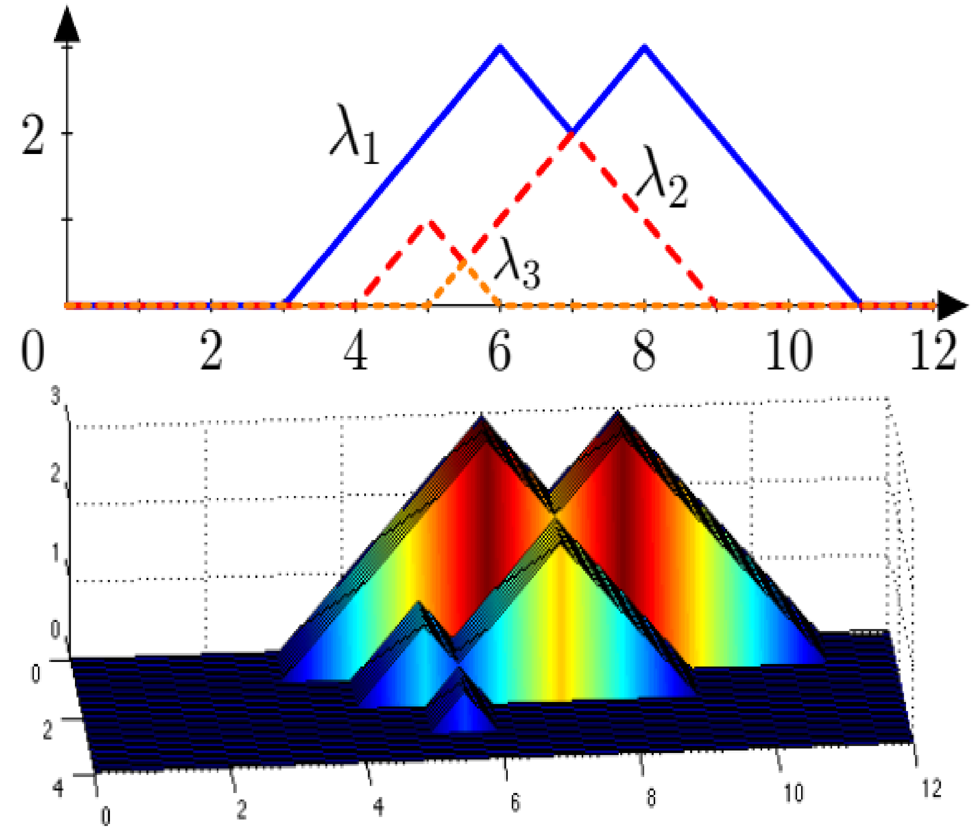
Use boundaries of rank function

Boundaries of rank function: $\lambda_i(t) = \sup\{s \geq 0 : \lambda(t-s, t+s) \geq i\}$

Landscape $\Lambda : \mathbb{R}^2 \to \mathbb{R}$ is defined as: $\Lambda(i,t) = \lambda_{\lfloor i \rfloor}(t)$

2. Hilbert space embedding

Persistence Landscape [Bubenik 2015]



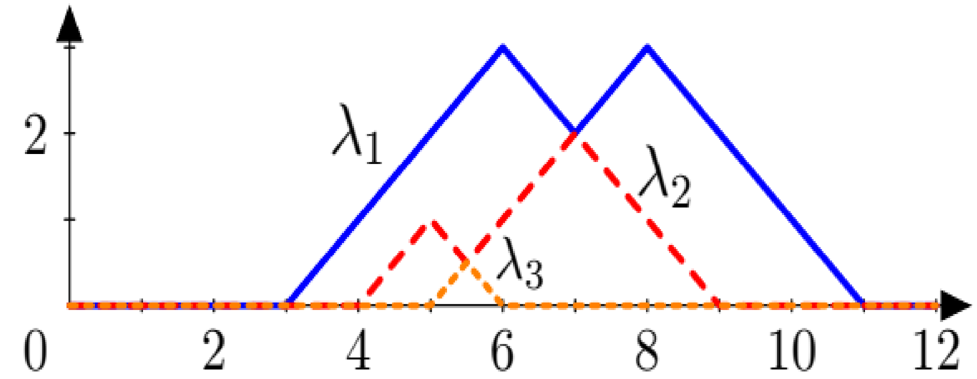
Prop: [Bubenik 2015]

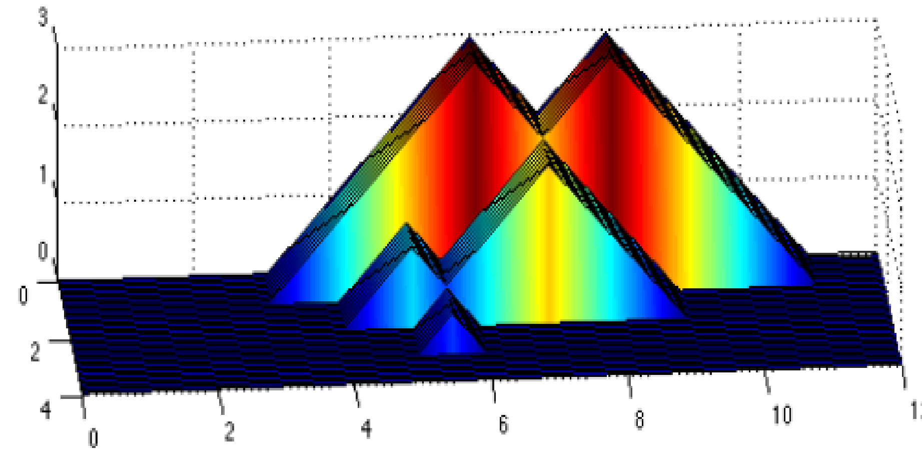
$$\|\Lambda(\mathrm{dgm}) - \Lambda(\mathrm{dgm}')\|_\infty \leq \mathrm{d}_\infty(\mathrm{dgm}, \mathrm{dgm}')$$

$\Longrightarrow$ Λ is Lipschitz hence Borel measurable

2. Hilbert space embedding

Persistence Landscape [Bubenik 2015]





Prop: [Bubenik 2015]

$\|\Lambda(\mathrm{dgm}) - \Lambda(\mathrm{dgm}')\|_\infty \leq \mathrm{d}_\infty(\mathrm{dgm}, \mathrm{dgm}')$

$\Longrightarrow$ Λ is Lipschitz hence Borel measurable

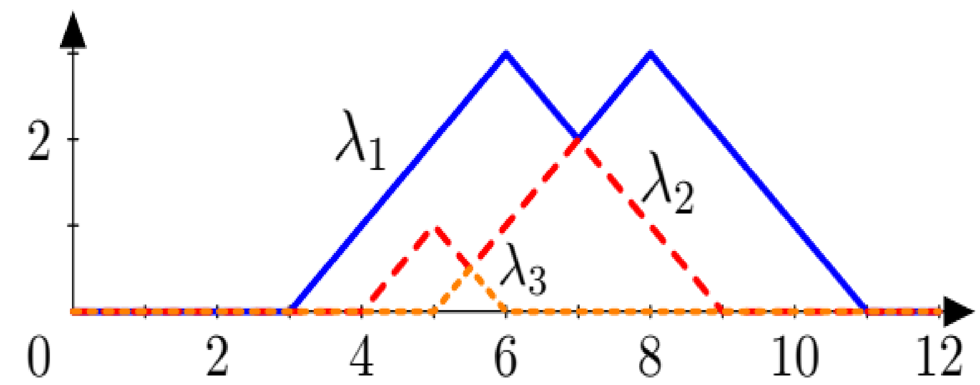
Given $D_1, \cdots, D_n \sim \mu$ iid, let $\bar{\Lambda}^n = \frac{1}{n}\sum_{i=1}^n \Lambda(D_i)$

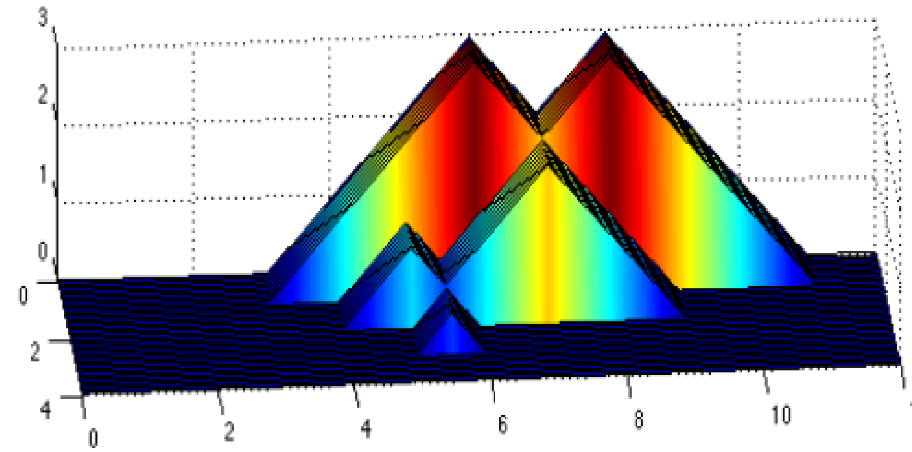
Thm: (strong law of large numbers) [Bubenik 2015]

If $E(\|\Lambda(\mu)\|) < +\infty$, then $\bar{\Lambda}^n \xrightarrow{\text{a.s.}} E(\Lambda(\mu))$.

2. Hilbert space embedding

Persistence Landscape [Bubenik 2015]





Prop: [Bubenik 2015]

$$\|\Lambda(\mathrm{dgm}) - \Lambda(\mathrm{dgm}')\|_\infty \leq \mathrm{d}_\infty(\mathrm{dgm}, \mathrm{dgm}')$$

$\Longrightarrow$ Λ is Lipschitz hence Borel measurable

Given $D_1, \cdots, D_n \sim \mu$ iid, let $\bar{\Lambda}^n = \frac{1}{n}\sum_{i=1}^n \Lambda(D_i)$

Thm: (strong law of large numbers) [Bubenik 2015]

If $E(\|\Lambda(\mu)\|) < +\infty$, then $\bar{\Lambda}^n \xrightarrow{\text{a.s.}} E(\Lambda(\mu))$.

Thm: (central limit theorem) [Bubenik 2015]

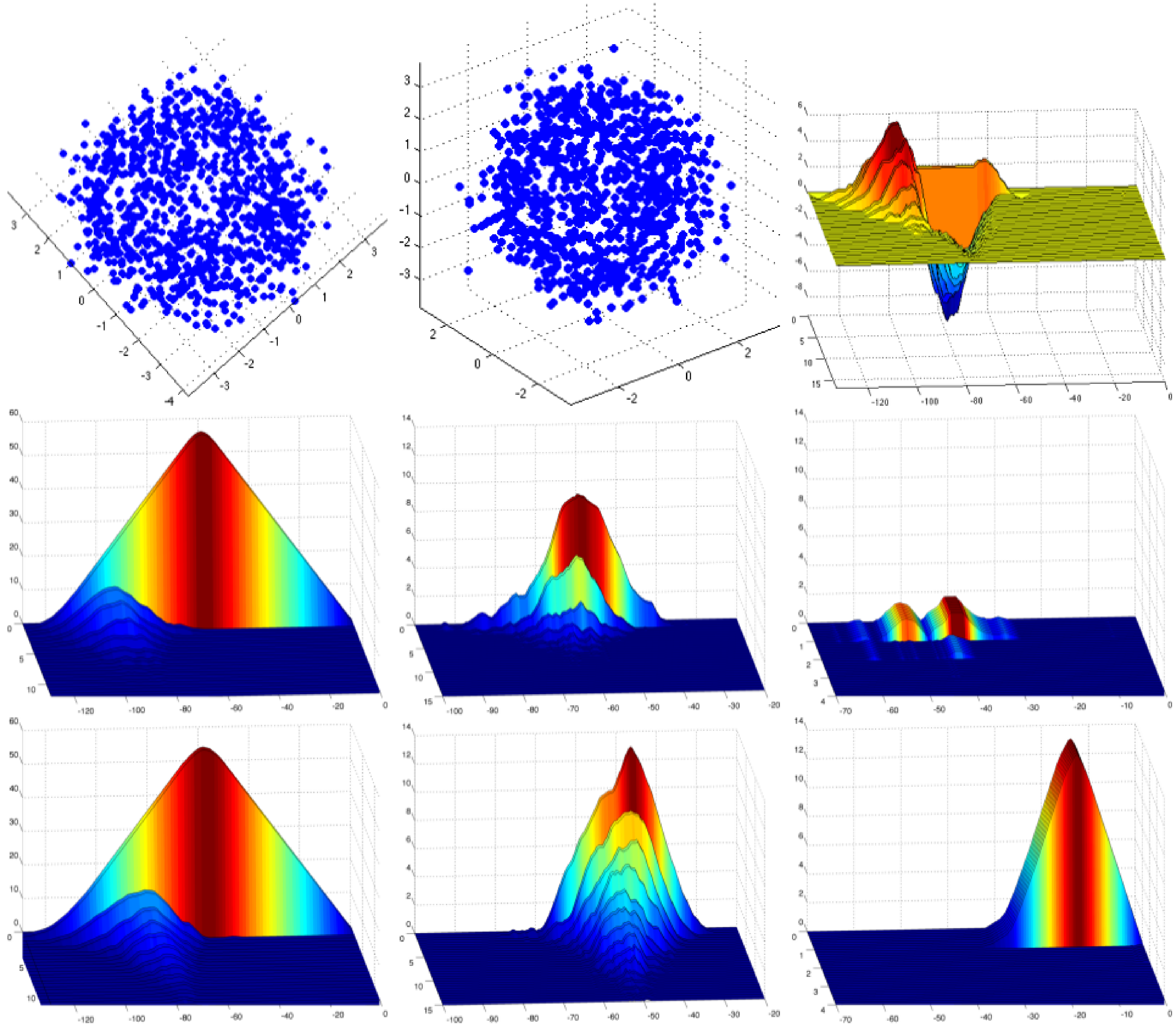
If $E(\|\Lambda(\mu)\|) < +\infty$ and $E(\|\Lambda(\mu)\|^2) < +\infty$, then

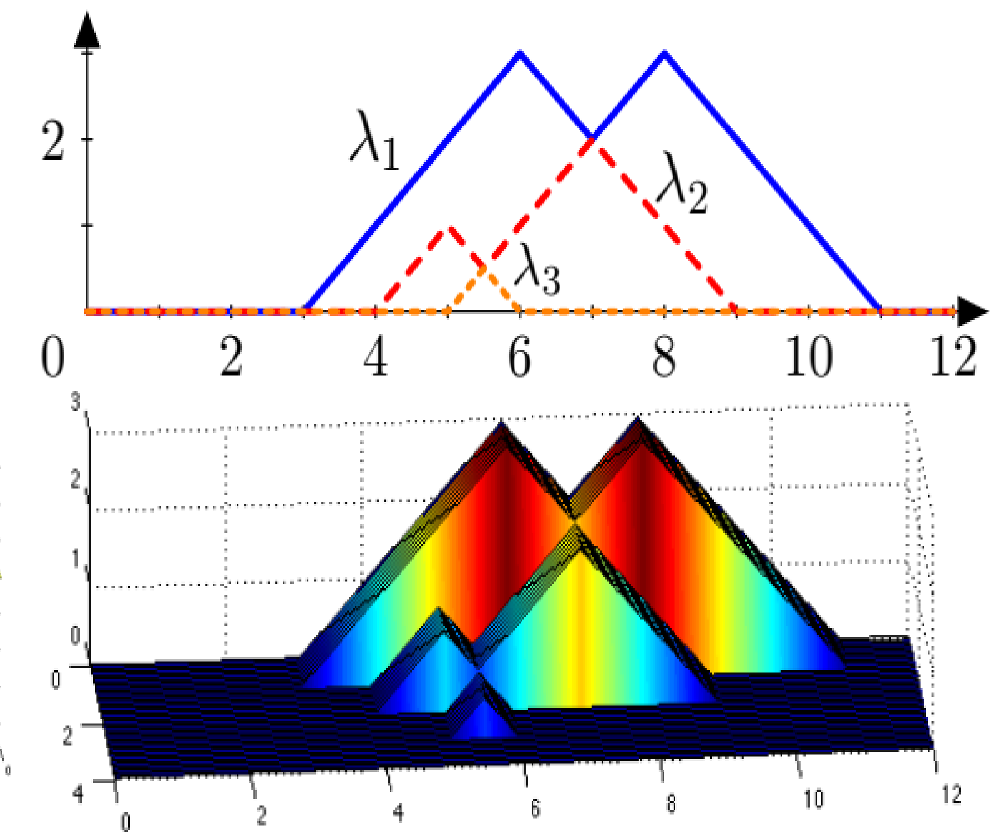
$$\sqrt{n}\left(\bar{\Lambda}^n - E(\Lambda(\mu))\right) \xrightarrow{d} \mathcal{N}(0, \Sigma(\Lambda(\mu))).$$

2. Hilbert space embedding

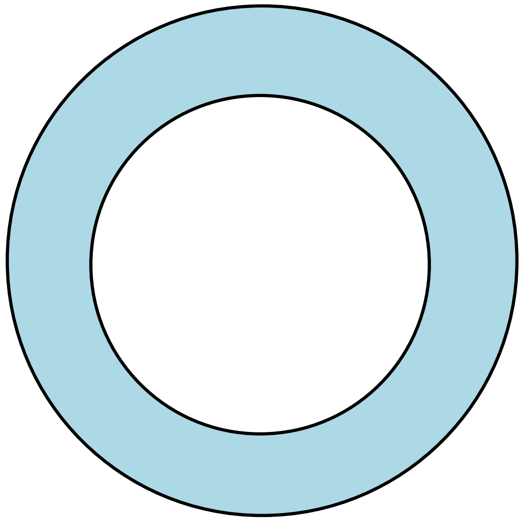
Persistence Landscape [Bubenik 2015]

Problem: mean landscape is not a landscape





3. Push-forwards from data space



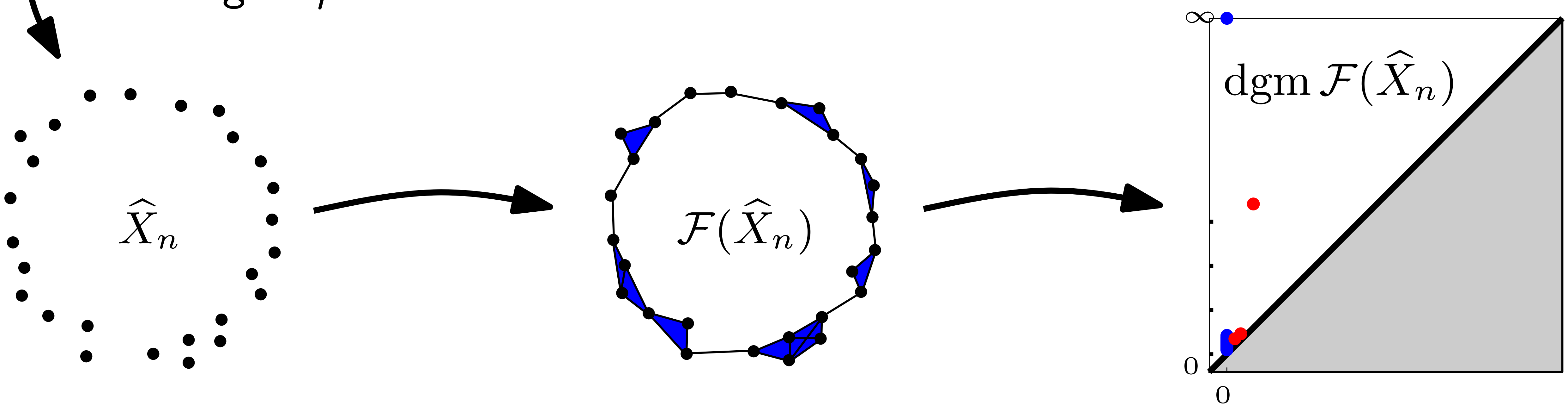
(X, d_X) compact metric space

μ probability measure supported on X $(\operatorname{supp} \mu = X)$

Sample n points iid according to μ.

Examples:

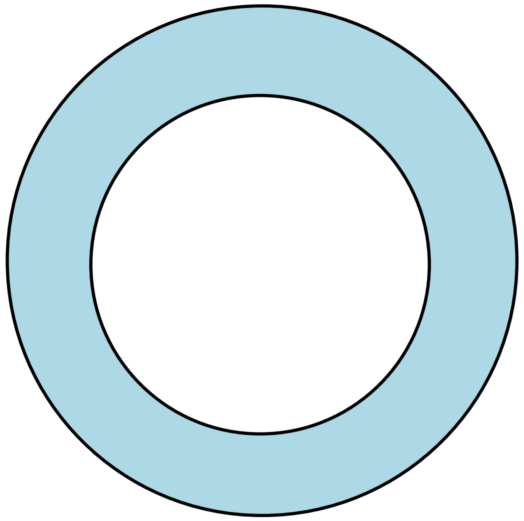
- $\mathcal{F}(\widehat{X}_n) = \mathcal{R}(\widehat{X}_n, \mathrm{d}_X)$
- $\mathcal{F}(\widehat{X}_n) = \mathcal{C}(\widehat{X}_n, \mathrm{d}_X)$
- $\cdots$



Questions:

- Statistical properties of the estimator $\operatorname{dgm} \mathcal{F}(\widehat{X}_n)$?
- Convergence to the ground truth $\operatorname{dgm} \mathcal{F}(X)$? Deviation bounds?

3. Push-forwards from data space



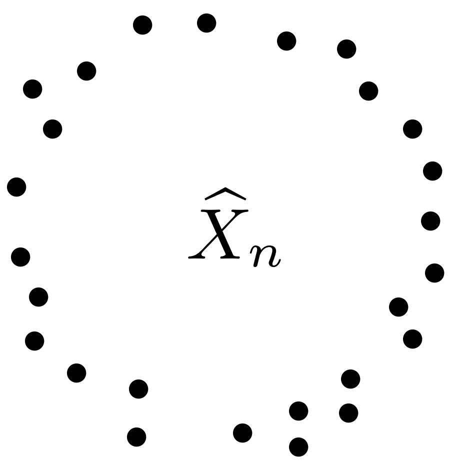
(X, d_X) compact metric space

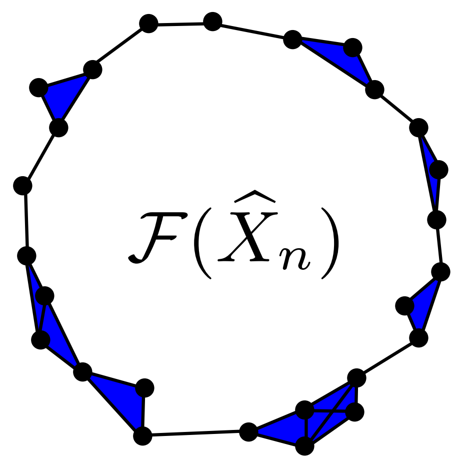
μ probability measure supported on X $(\operatorname{supp}\mu = X)$

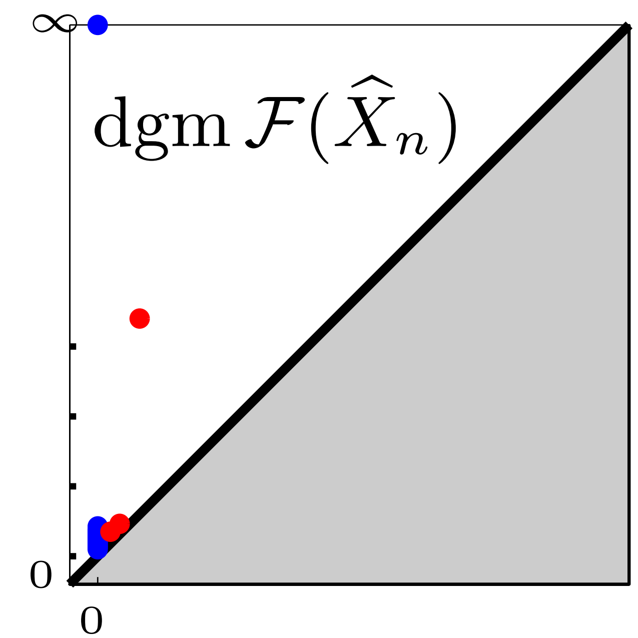
Sample n points iid according to μ.

Examples:

- $\mathcal{F}(\widehat{X}_n) = \mathcal{R}(\widehat{X}_n, \mathrm{d}_X)$
- $\mathcal{F}(\widehat{X}_n) = \mathcal{C}(\widehat{X}_n, \mathrm{d}_X)$
- $\cdots$





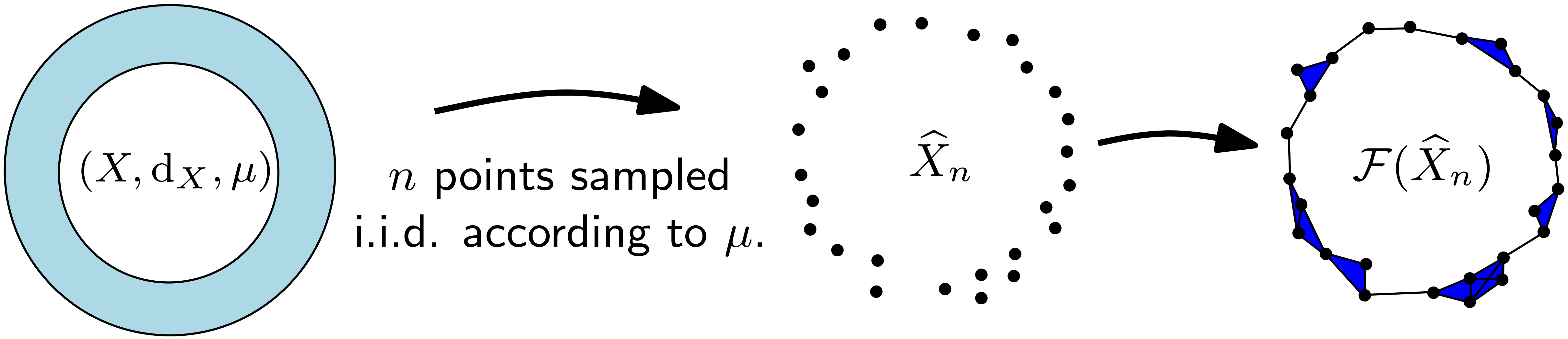


$\boxed{\textbf{Stability thm: } \mathrm{d}_\infty(\operatorname{dgm}\mathcal{F}(\widehat{X}_n), \operatorname{dgm}\mathcal{F}(X)) \leq 2\mathrm{d}_\mathrm{H}(\widehat{X}_n, X)}$

$\Rightarrow$ for any $\varepsilon > 0$,

$$\mathbb{P}\left(\mathrm{d}_\infty\left(\operatorname{dgm}\mathcal{F}(\widehat{X}_n), \operatorname{dgm}\mathcal{F}(X),\right) > \varepsilon\right) \leq \mathbb{P}\left(\mathrm{d}_\mathrm{H}(\widehat{X}_n, X) > \frac{\varepsilon}{2}\right)$$

Deviation inequality



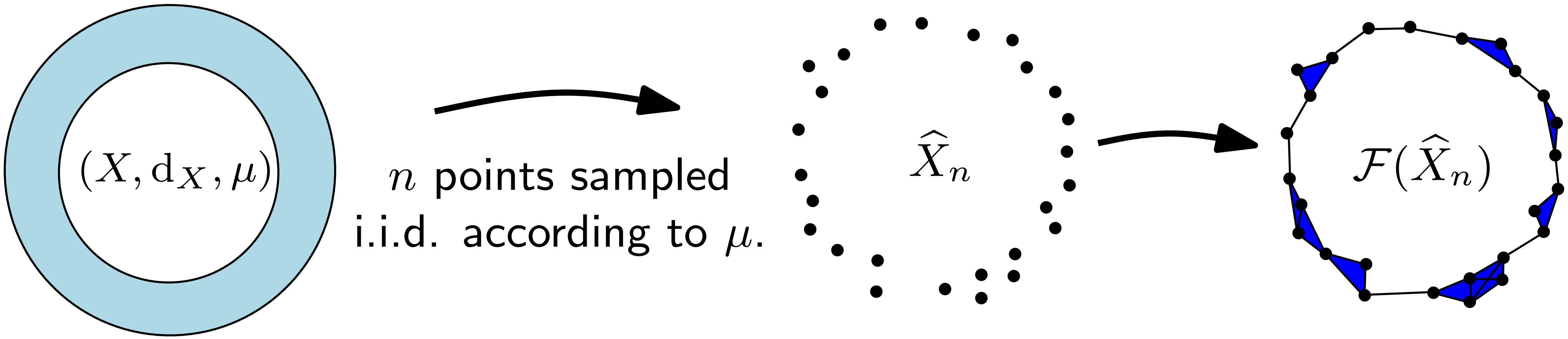
For $a, b > 0$, μ satisfies the (a, b)-**standard** assumption if for any $x \in X$ and any $r > 0$, we have $\mu(B(x, r)) \geq \min(ar^b, 1)$.

Theorem [Chazal, Glisse, Labruère, Michel 2014-15]:

If μ is (a, b)-standard then for any $\varepsilon > 0$:

$$\mathbb{P}\left(\mathrm{d}_\infty\left(\mathrm{dgm}\,\mathcal{F}(\widehat{X}_n), \mathrm{dgm}\,\mathcal{F}(X)\right) > \varepsilon\right) \leq \frac{8^b}{a\varepsilon^b}\exp(-na\varepsilon^b)$$

Deviation inequality / rate of convergence



For $a, b > 0$, μ satisfies the (a, b)-**standard** assumption if for any $x \in X$ and any $r > 0$, we have $\mu(B(x, r)) \geq \min(ar^b, 1)$.

Theorem [Chazal, Glisse, Labruère, Michel 2014-15]:

If μ is (a, b)-standard then for any $\varepsilon > 0$:

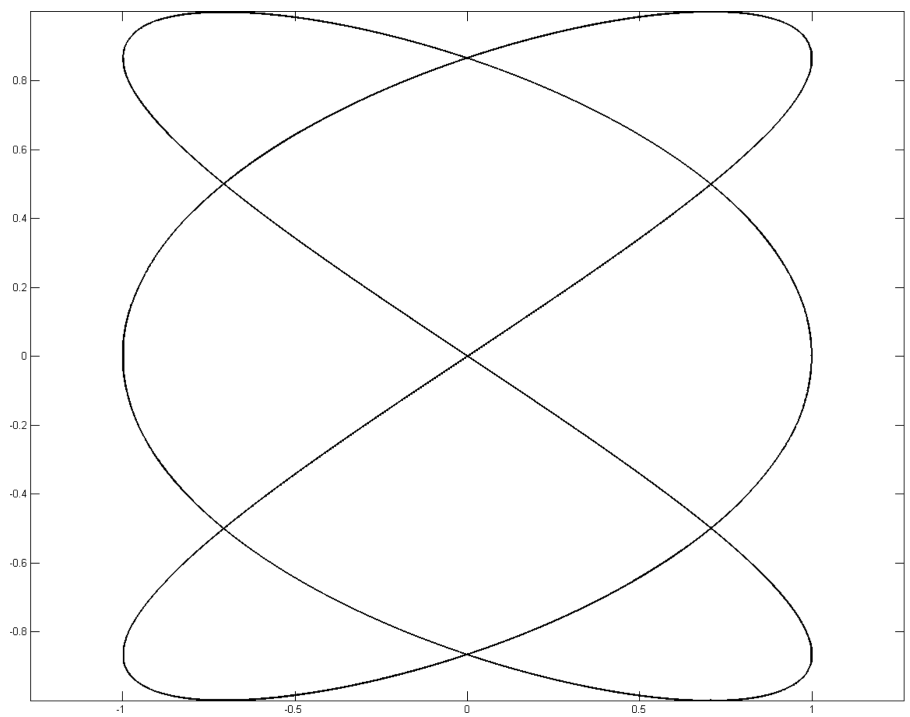
$$\mathbb{P}\left(\mathrm{d}_\infty\left(\mathrm{dgm}\,\mathcal{F}(\widehat{X}_n), \mathrm{dgm}\,\mathcal{F}(X)\right) > \varepsilon\right) \leq \frac{8^b}{a\varepsilon^b}\exp(-na\varepsilon^b)$$

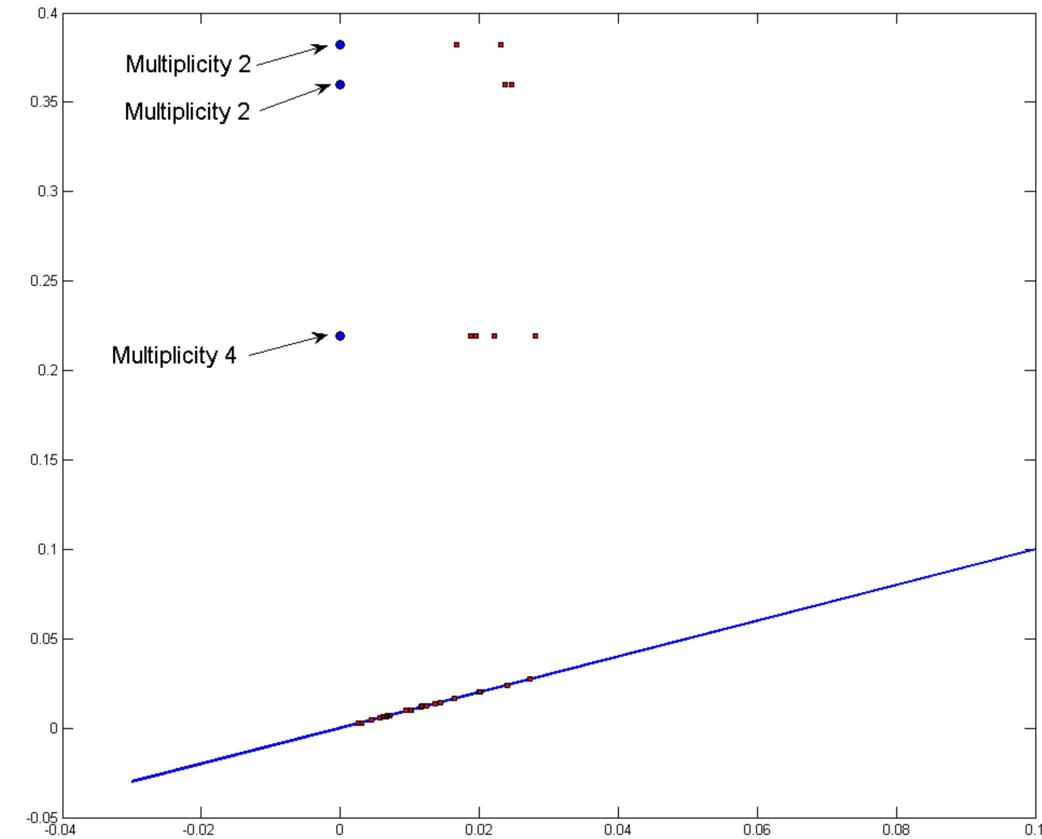
Corollary [Chazal, Glisse, Labruère, Michel 2014-15]:

$$\sup_{\mu \in \mathcal{P}} \mathbb{E}\left[\mathrm{d}_\infty\left(\mathrm{dgm}\,\mathcal{F}(\widehat{X}_n),\ \mathrm{dgm}\,\mathcal{F}(X)\right)\right] \leq C\left(\frac{\log n}{n}\right)^{1/b},$$

where C depends only on a, b. Moreover, the estimator $\mathrm{dgm}\,\mathcal{F}(\widehat{X}_n)$ is **minimax optimal** (up to a $\log n$ factor) on the space $\mathcal{P}$ of (a, b)-standard probability measures on X.

Numerical illustrations

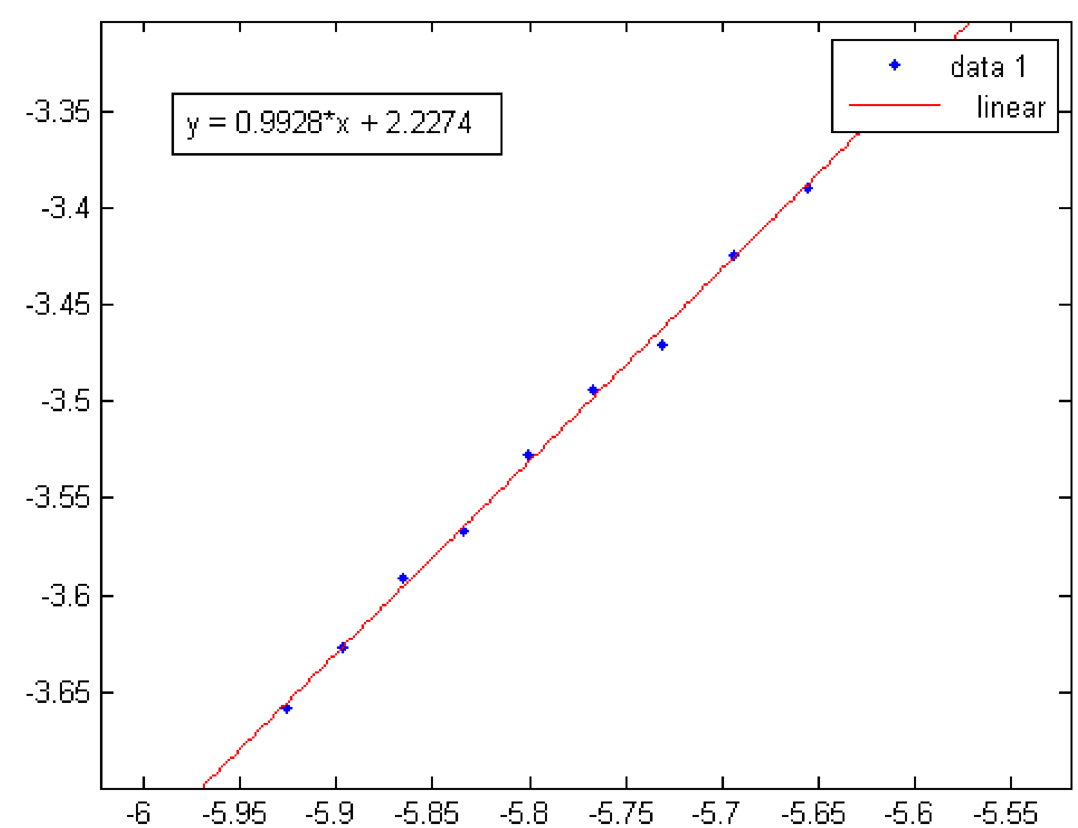




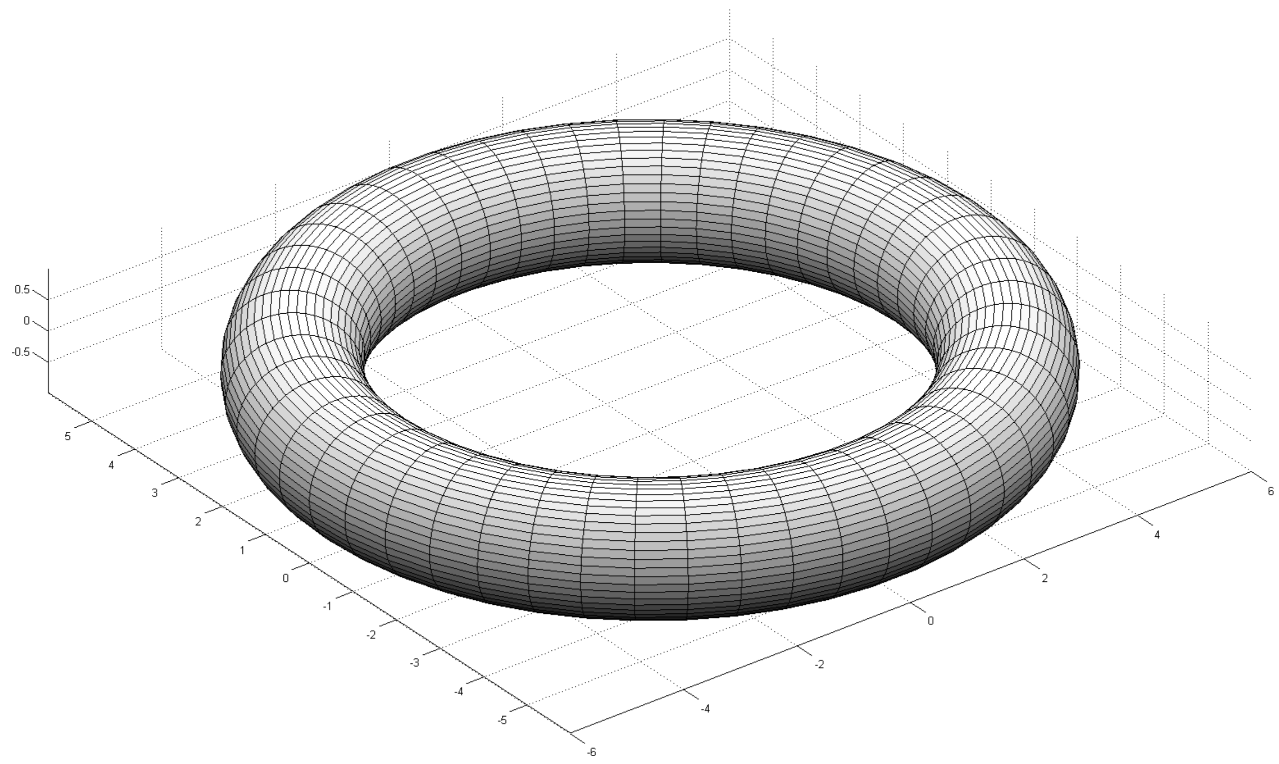
- μ: unif. measure on Lissajous curve X.
- $\mathcal{F}$: distance to X in $\mathbb{R}^2$.
- sample $k = 300$ sets of n points for $n = [2100 : 100 : 3000]$.
- compute

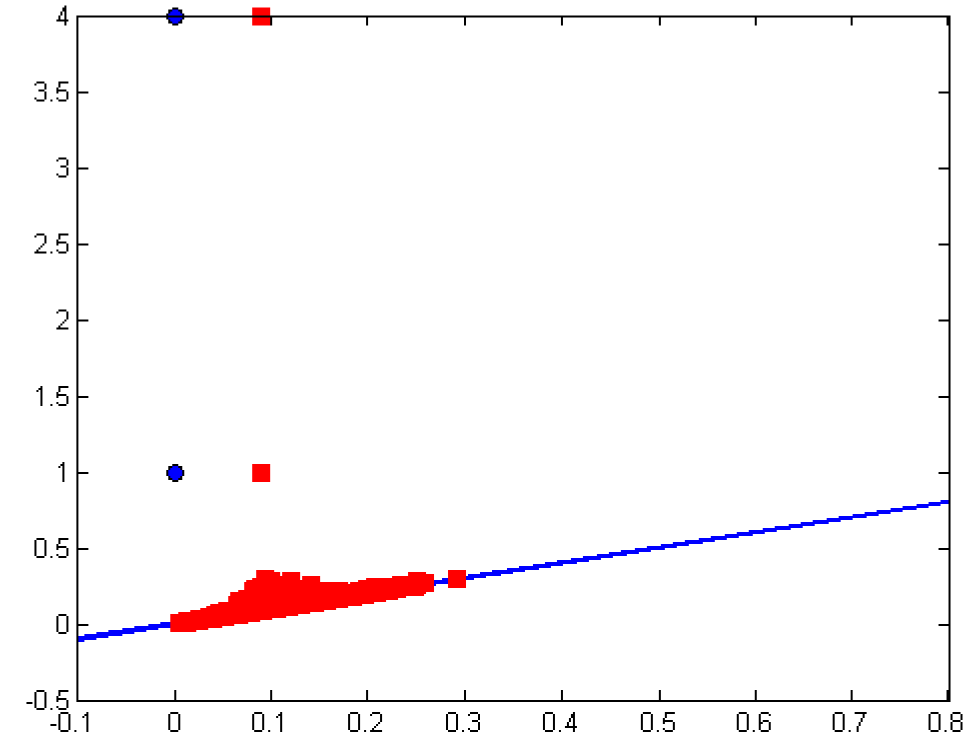
$$\widehat{\mathbb{E}}_n = \widehat{\mathbb{E}}[\mathrm{d}_\infty(\mathrm{dgm}\,\mathcal{F}(\widehat{X_n}), \mathrm{dgm}\,\mathcal{F}(X))].$$

- plot $\log(\widehat{\mathbb{E}}_n)$ as a function of $\log(\log(n)/n)$.



Numerical illustrations

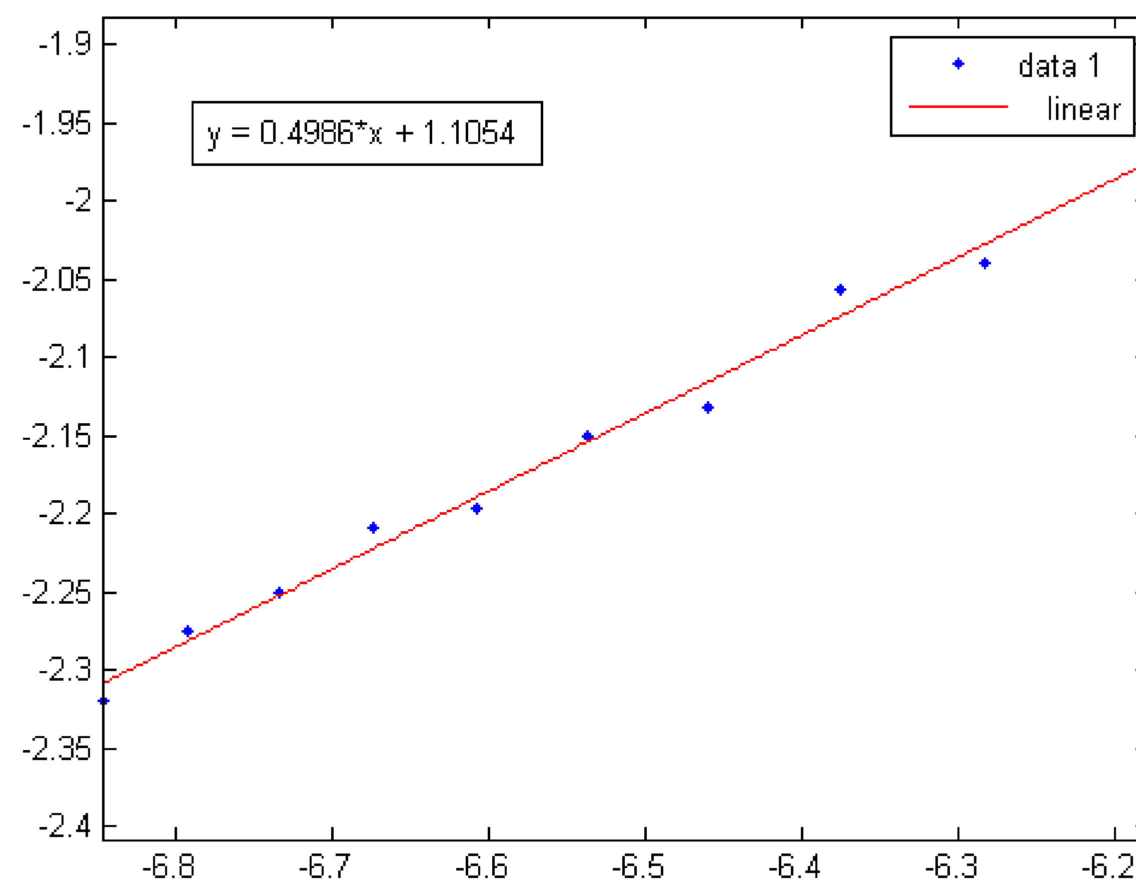




- μ: unif. measure on a torus X.
- $\mathcal{F}$: distance to X in $\mathbb{R}^3$.
- sample $k = 300$ sets of n points for $n = [12000 : 1000 : 21000]$.
- compute

$$\widehat{\mathbb{E}}_n = \widehat{\mathbb{E}}[\mathrm{d}_\infty(\mathrm{dgm}\,\mathcal{F}(\widehat{X_n}), \mathrm{dgm}\,\mathcal{F}(X))].$$

- plot $\log(\widehat{\mathbb{E}}_n)$ as a function of $\log(\log(n)/n)$.



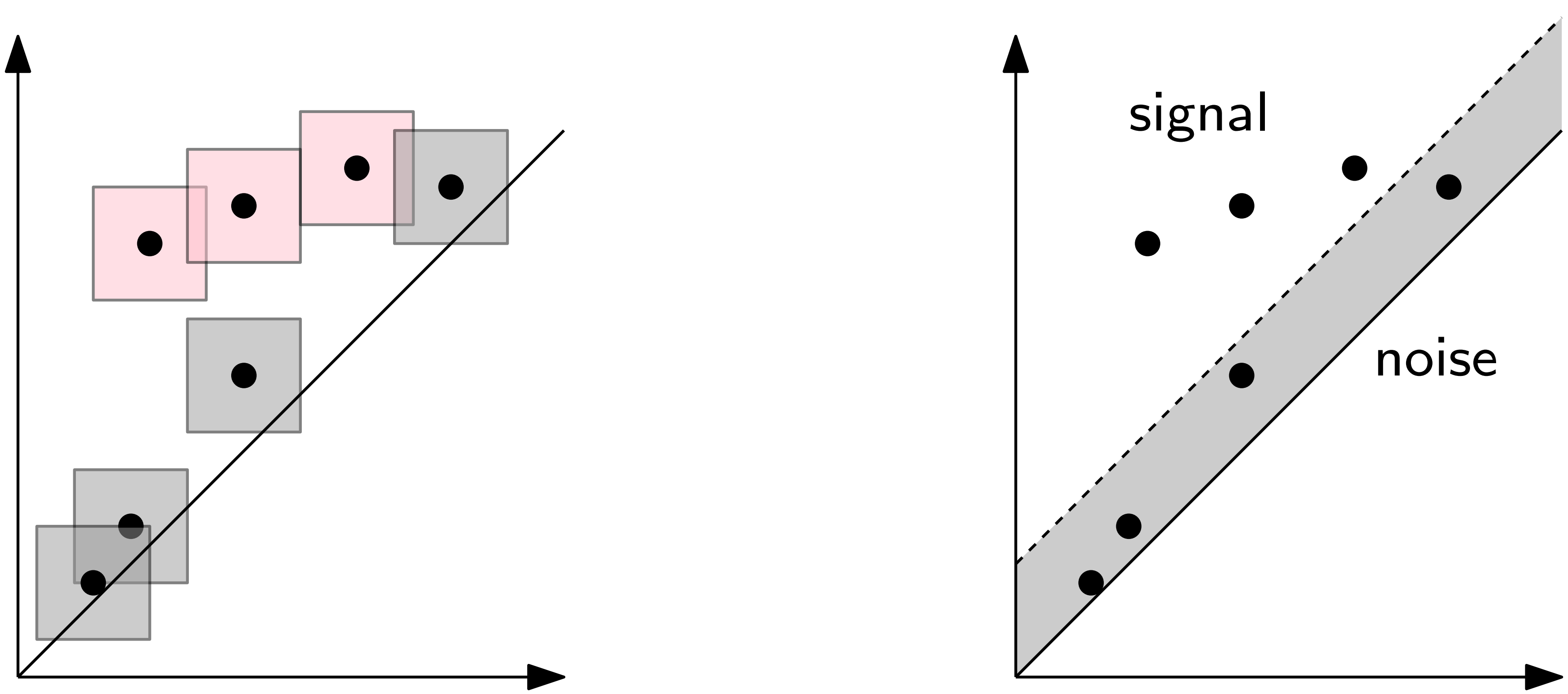
Confidence regions

Setup: $(X, \mathrm{d}_X, \mu) \to \widehat{X}_n \to \mathcal{F}(\widehat{X}_n) \to \operatorname{dgm} \mathcal{F}(\widehat{X}_n)$

Goal: given $\alpha \in (0,1)$, estimate $c_n(\alpha) \geq 0$ such that

$$\limsup_{n\to\infty} \mathbb{P}\left(\mathrm{d}_\infty\left(\operatorname{dgm}\mathcal{F}(\widehat{X}_n), \operatorname{dgm}\mathcal{F}(X)\right) > c_n(\alpha)\right) \leq \alpha$$

$\to$ confidence region: d_∞-ball of radius $c_n(\alpha)$ around $\operatorname{dgm}\mathcal{F}(\widehat{X}_n)$



Confidence regions

Setup: $(X, \mathrm{d}_X, \mu) \to \widehat{X}_n \to \mathcal{F}(\widehat{X}_n) \to \mathrm{dgm}\,\mathcal{F}(\widehat{X}_n)$

Goal: given $\alpha \in (0,1)$, estimate $c_n(\alpha) \geq 0$ such that

$$\limsup_{n\to\infty} \mathbb{P}\left(\mathrm{d}_\infty\left(\mathrm{dgm}\,\mathcal{F}(\widehat{X}_n), \mathrm{dgm}\,\mathcal{F}(X)\right) > c_n(\alpha)\right) \leq \alpha$$

Note: we already have an inequality of this kind but...

$$\mathbb{P}\left(\mathrm{d}_\infty\left(\mathrm{dgm}\,\mathcal{F}(\widehat{X}_n), \mathrm{dgm}\,\mathcal{F}(X)\right) > \varepsilon\right) \leq \frac{8^b}{a\varepsilon^b}\exp(-na\varepsilon^b)$$

unknown

Confidence regions

Setup: $(X, \mathrm{d}_X, \mu) \to \widehat{X}_n \to \mathcal{F}(\widehat{X}_n) \to \operatorname{dgm} \mathcal{F}(\widehat{X}_n)$

Goal: given $\alpha \in (0,1)$, estimate $c_n(\alpha) \geq 0$ such that

$$\limsup_{n\to\infty} \mathbb{P}\left(\mathrm{d}_\infty\left(\operatorname{dgm}\mathcal{F}(\widehat{X}_n), \operatorname{dgm}\mathcal{F}(X)\right) > c_n(\alpha)\right) \leq \alpha$$

Bootstrap: (ideally)

- draw $X^* = X_1^*, \cdots, X_n^*$ iid from $\mu_{\widehat{X}_n}$ (empirical measure on $\widehat{X}_n$)
- compute $d^* = \mathrm{d}_\infty\left(\operatorname{dgm}\mathcal{F}(X^*), \operatorname{dgm}\mathcal{F}(\widehat{X}_n)\right)$
- repeat N times to get $d_1^*, \cdots, d_N^*$
- let $c_n(\alpha)$ be the $(1-\alpha)$ quantile of $\frac{1}{N}\sum_{i=1}^N I(d_i^* \geq t)$

Principle [Efron 1979]: variations of $\operatorname{dgm}\mathcal{F}(X^*)$ around $\operatorname{dgm}\mathcal{F}(\widehat{X}_n)$ are same as variations of $\operatorname{dgm}\mathcal{F}(\widehat{X}_n)$ around $\operatorname{dgm}\mathcal{F}(X)$.

Note: requires some conditions on (X, d_X, μ) and diagram space

Confidence regions

Setup: $(X, \mathrm{d}_X, \mu) \to \widehat{X}_n \to \mathcal{F}(\widehat{X}_n) \to \operatorname{dgm}\mathcal{F}(\widehat{X}_n)$

Goal: given $\alpha \in (0,1)$, estimate $c_n(\alpha) \geq 0$ such that

$$\limsup_{n\to\infty} \mathbb{P}\left(\mathrm{d}_\infty\left(\operatorname{dgm}\mathcal{F}(\widehat{X}_n), \operatorname{dgm}\mathcal{F}(X)\right) > c_n(\alpha)\right) \leq \alpha$$

Bootstrap: (in fact)

- draw $X^* = X_1^*, \cdots, X_n^*$ iid from $\mu_{\widehat{X}_n}$ (empirical measure on $\widehat{X}_n$)
- compute $d^* = \cancel{\mathrm{d}_\infty\left(\operatorname{dgm}\mathcal{F}(X^*), \operatorname{dgm}\mathcal{F}(\widehat{X}_n)\right)}$ $\mathrm{d}_{\mathrm{H}}(X^*, \hat{X}_n)$
- repeat N times to get $d_1^*, \cdots, d_N^*$
- let $c_n(\alpha)$ be the $(1-\alpha)$ quantile of $\frac{1}{N}\sum_{i=1}^N I(d_i^* \geq t)$

Theorem [Balakrishnan et al. 2013] + [Chazal et al. 2014]:

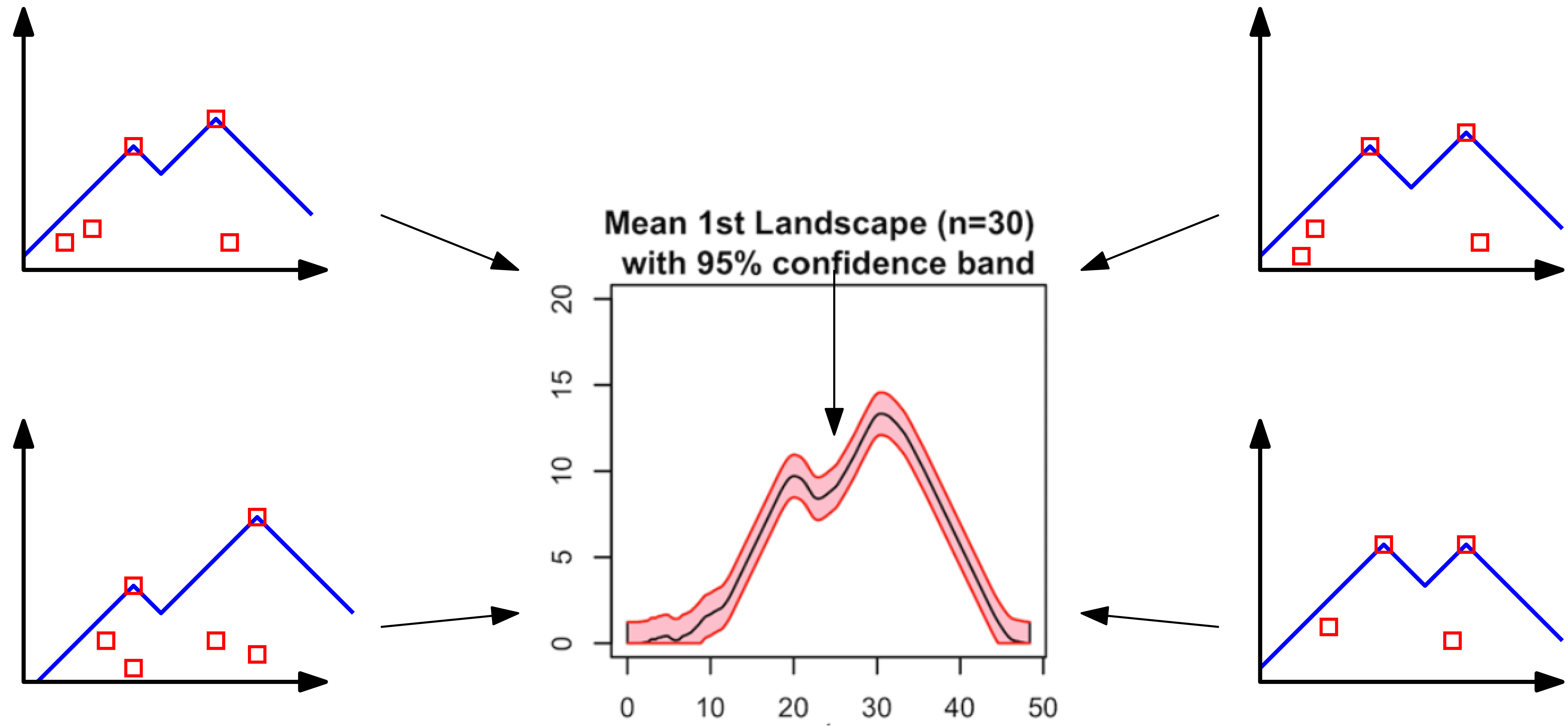
$$\limsup_{n\to\infty} \mathbb{P}\left(\mathrm{d}_\infty\left(\operatorname{dgm}\mathcal{F}(\widehat{X}_n), \operatorname{dgm}\mathcal{F}(X)\right) > c_n(\alpha)\right) \leq \alpha.$$

Confidence regions

Setup: $(X, \mathrm{d}_X, \mu) \to \widehat{X}_n^1, \cdots, \widehat{X}_n^m \to \phi_k(D_n^1), \cdots, \phi_k(D_n^m)$

$$\downarrow$$

empirical mean feature vector $\longrightarrow$ $\bar{v} = \frac{1}{m}\sum_{i=1}^m \phi_k(D_n^i)$



Confidence regions

Setup: $(X, \mathrm{d}_X, \mu) \to \widehat{X}_n^1, \cdots, \widehat{X}_n^m \to \phi_k(D_n^1), \cdots, \phi_k(D_n^m)$

$$\downarrow$$

empirical mean feature vector $\longrightarrow$ $\bar{v} = \frac{1}{m} \sum_{i=1}^m \phi_k(D_n^i)$

Goal: given $\alpha \in (0,1)$, estimate $c_n(\alpha) \geq 0$ such that

$$\limsup_{n \to \infty} \mathbb{P}\left(\left\| \bar{v} - \underbrace{\mathbb{E}_{(\phi_k \circ \mathrm{dgm} \circ \mathcal{F})^*(\mu^{\otimes n})}[v]} \right\|_{\mathcal{H}_k} > c_n(\alpha) \right) \leq \alpha$$

mean feature vector according to the measure induced by $\mu^{\otimes n}$

(call it $\Lambda_{\mu,n}$ for landscapes)

Confidence regions

Setup: $(X, \mathrm{d}_X, \mu) \to \widehat{X}_n^1, \cdots, \widehat{X}_n^m \to \Lambda(D_n^1), \cdots, \Lambda(D_n^m)$

$$\downarrow$$

$$\bar{\Lambda} = \frac{1}{m} \sum_{i=1}^m \Lambda(D_n^i)$$

Bootstrap with landscapes:

- draw $\Lambda_1^*, \cdots, \Lambda_m^*$ iid from $\frac{1}{m} \sum_{i=1}^m \delta_{\Lambda(D_n^i)}$
- compute $\bar{\Lambda}^* = \frac{1}{m} \sum_{i=1}^m \Lambda_i^*$ and $d^* = \|\bar{\Lambda}^* - \bar{\Lambda}\|_\infty$
- repeat N times to get $d_1^*, \cdots, d_N^*$
- let $c_n(\alpha)$ be the $(1-\alpha)$ quantile of $\frac{1}{N} \sum_{i=1}^N I(\, d_i^* \geq t)$

Theorem [Chazal et al. 2014]:

$$\limsup_{m \to \infty} \mathbb{P}\left(\left\|\bar{\Lambda} - \Lambda_{\mu,n}\right\|_\infty > c_n(\alpha)\right) \leq \alpha.$$

Confidence regions

Setup: $(X, \mathrm{d}_X, \mu) \to \widehat{X}_n^1, \cdots, \widehat{X}_n^m \to \Lambda(D_n^1), \cdots, \Lambda(D_n^m)$

$$\downarrow$$

$$\bar{\Lambda} = \frac{1}{m}\sum_{i=1}^m \Lambda(D_n^i)$$

Bootstrap with landscapes:

- draw $\Lambda_1^*, \cdots, \Lambda_m^*$ iid from $\frac{1}{m}\sum_{i=1}^m \delta_{\Lambda(D_n^i)}$
- compute $\bar{\Lambda}^* = \frac{1}{m}\sum_{i=1}^m \Lambda_i^*$ and $d^* = \|\bar{\Lambda}^* - \bar{\Lambda}\|_\infty$
- repeat N times to get $d_1^*, \cdots, d_N^*$ $\quad |\bar{\Lambda}^*(t) - \bar{\Lambda}(t)|$
- let $c_n(\alpha)$ be the $(1-\alpha)$ quantile of $\frac{1}{N}\sum_{i=1}^N I(\, d_i^* \geq t)$

Theorem [Chazal et al. 2014]:

$$\limsup_{m\to\infty} \mathbb{P}\left(\left\|\bar{\Lambda} - \Lambda_{\mu,n}\right\|_\infty > c_n(\alpha)\right) \leq \alpha.$$

$$|\bar{\Lambda}(t) - \Lambda_{\mu,n}(t)|$$

Note: can be done for a fixed t

Confidence regions

Setup: $(X, \mathrm{d}_X, \mu) \to \widehat{X}_n^1, \cdots, \widehat{X}_n^m \to \Lambda(D_n^1), \cdots, \Lambda(D_n^m)$

$$\downarrow$$

$$\bar{\Lambda} = \frac{1}{m}\sum_{i=1}^m \Lambda(D_n^i)$$

Bootstrap with landscapes:

- draw $\Lambda_1^*, \cdots, \Lambda_m^*$ iid from $\frac{1}{m}\sum_{i=1}^m \delta_{\Lambda(D_n^i)}$
- compute $\bar{\Lambda}^* = \frac{1}{m}\sum_{i=1}^m \Lambda_i^*$ and $d^* = \|\bar{\Lambda}^* - \bar{\Lambda}\|_\infty$
- repeat N times to get $d_1^*, \cdots, d_N^*$
- let $c_n(\alpha)$ be the $(1-\alpha)$ quantile of $\frac{1}{N}\sum_{i=1}^N I(\, d_i^* \geq t)$

Theorem [Chazal et al. 2015]:

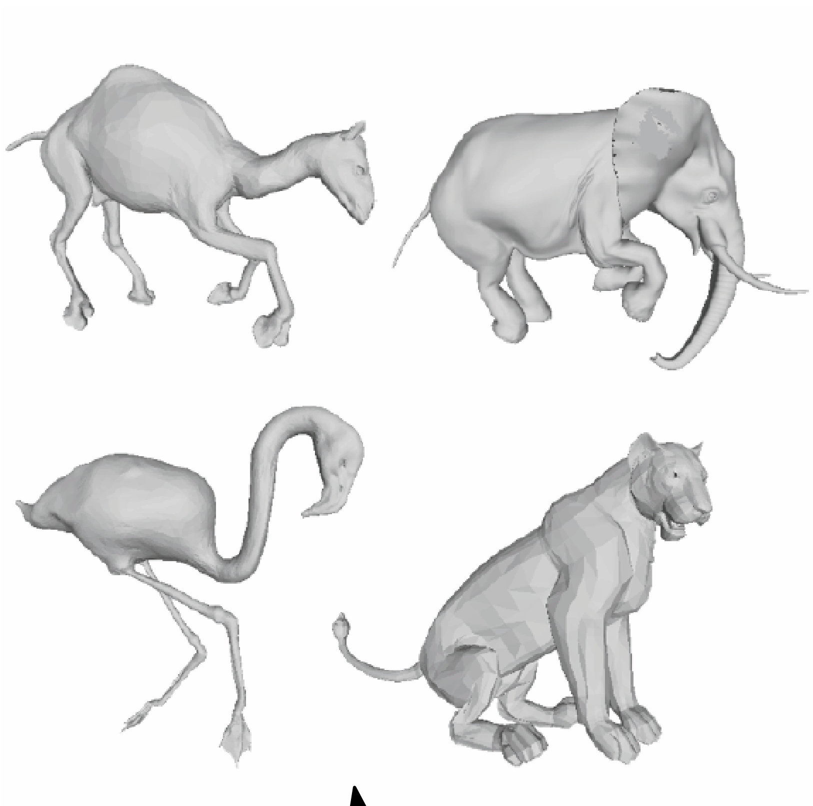
$$\big\|\bar{\Lambda} - \Lambda(\operatorname{dgm} \mathcal{F}(X))\big\|_\infty \leq \underbrace{\big\|\bar{\Lambda} - \Lambda_{\mu,n}\big\|_\infty}_{\text{variance term}} + \underbrace{\|\Lambda_{\mu,n} - \Lambda(\operatorname{dgm} \mathcal{F}(X))\|_\infty}_{\text{bias term} \leq C\left(\frac{\log n}{an}\right)^{1/b} \text{ when } \mu \text{ is } (a,b)\text{-standard}}$$

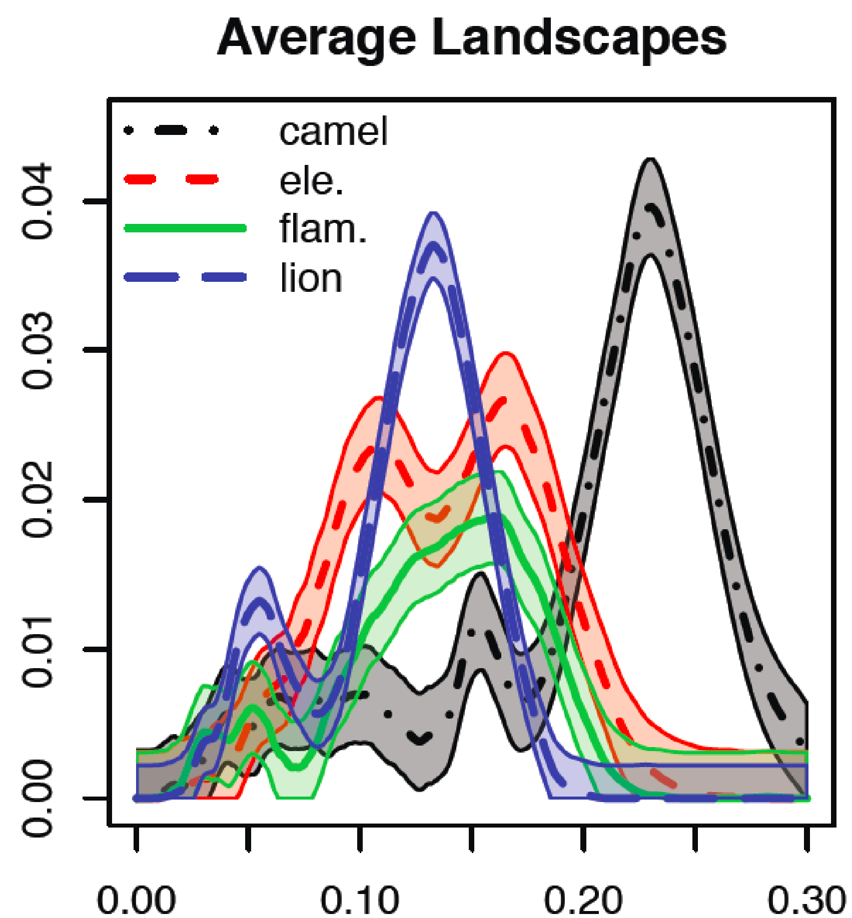
variance term

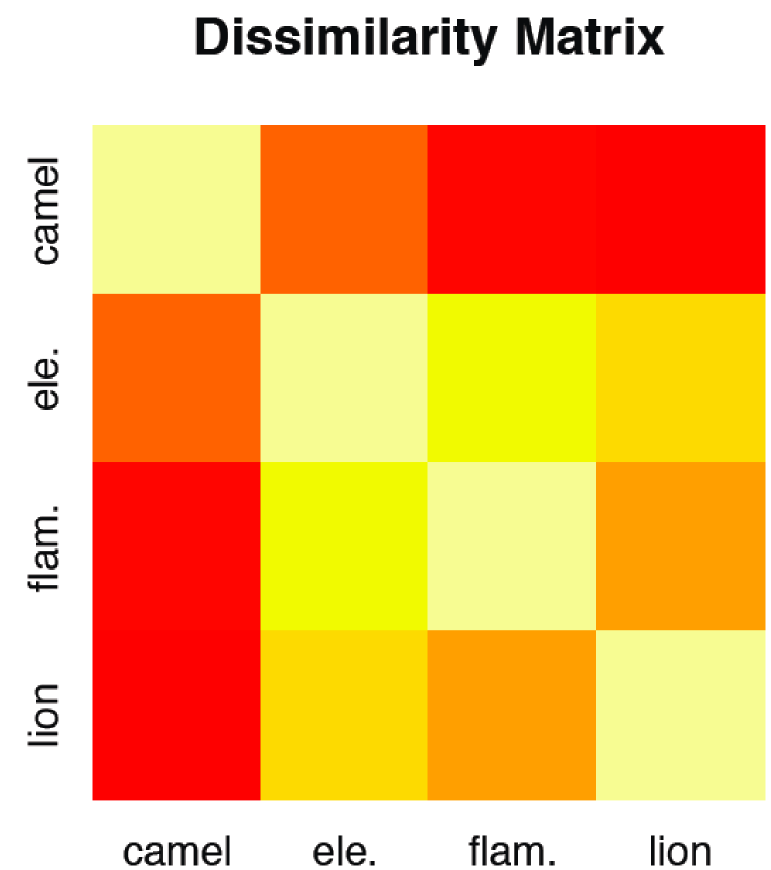
bias term $\leq C\left(\frac{\log n}{an}\right)^{1/b}$ when μ is (a,b)-standard

Some applications

Application 1: 3D shapes classification





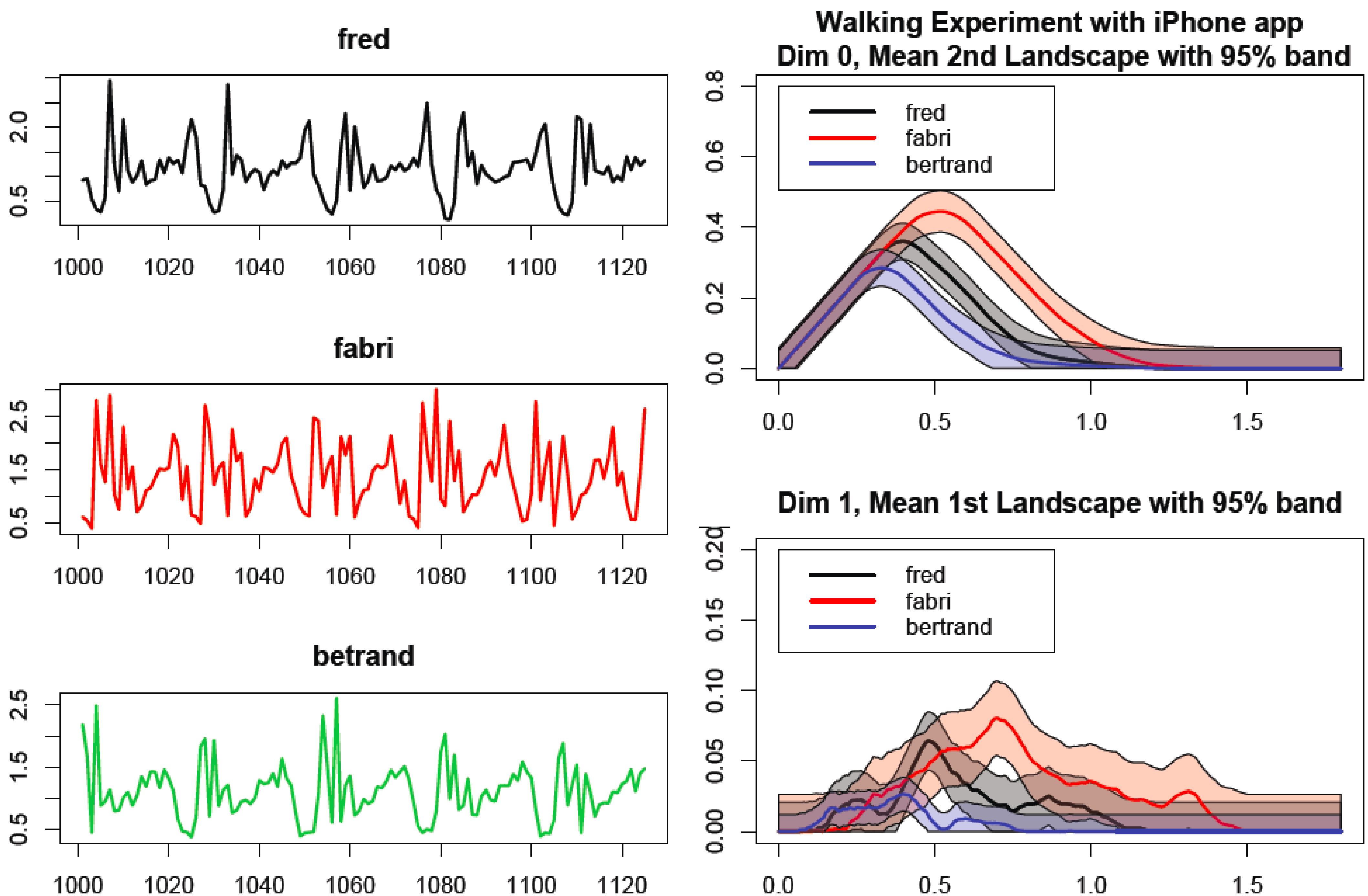


each mesh has 7K to 40K vertices

From $m = 100$ subsamples of size $n = 300$

Some applications

Application 2: walking behaviors classification from smartphone accelerometer data



- spatial time series (accelerometer data from the smarphone of users).
- no registration/calibration preprocessing step needed to compare!

Outline

1. Descriptors and stability

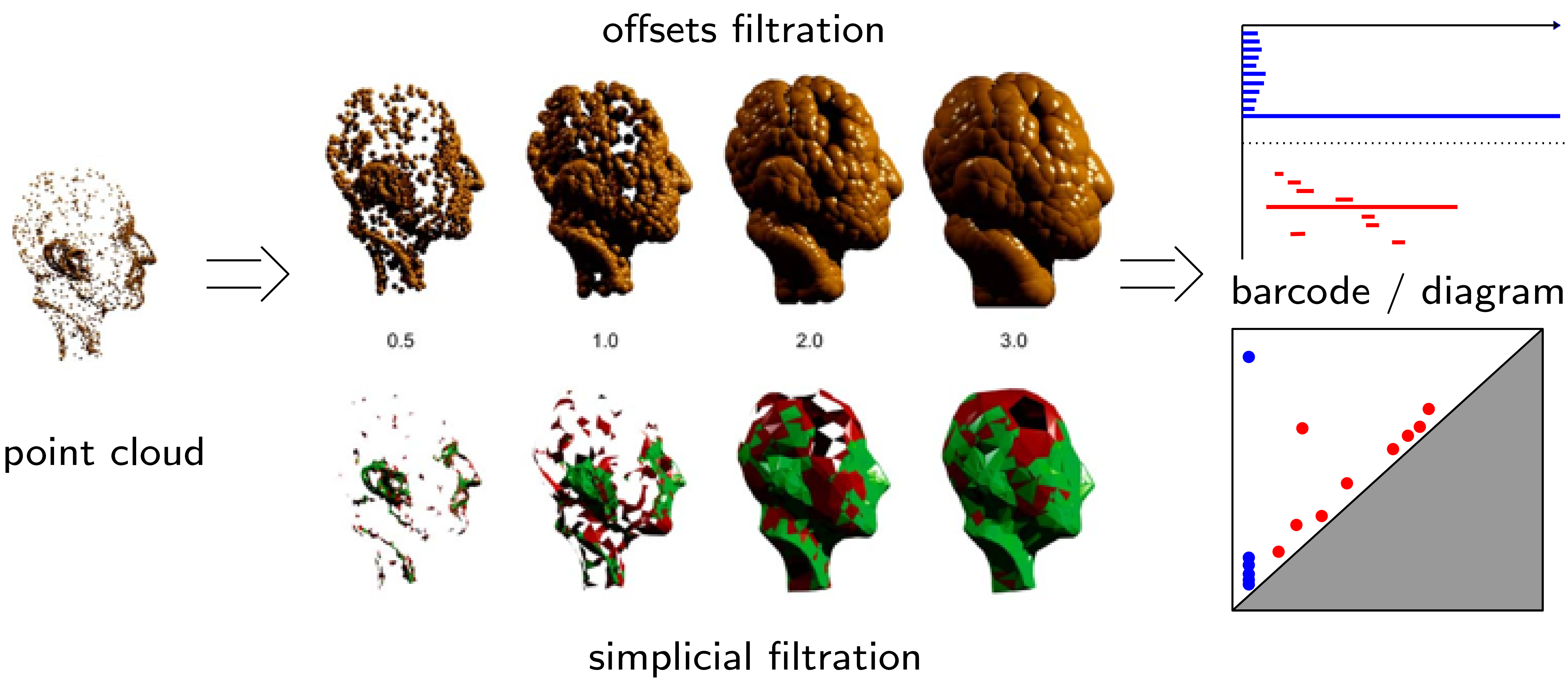
2. Vectorizations and kernels

3. Statistics

4. Discrimination power

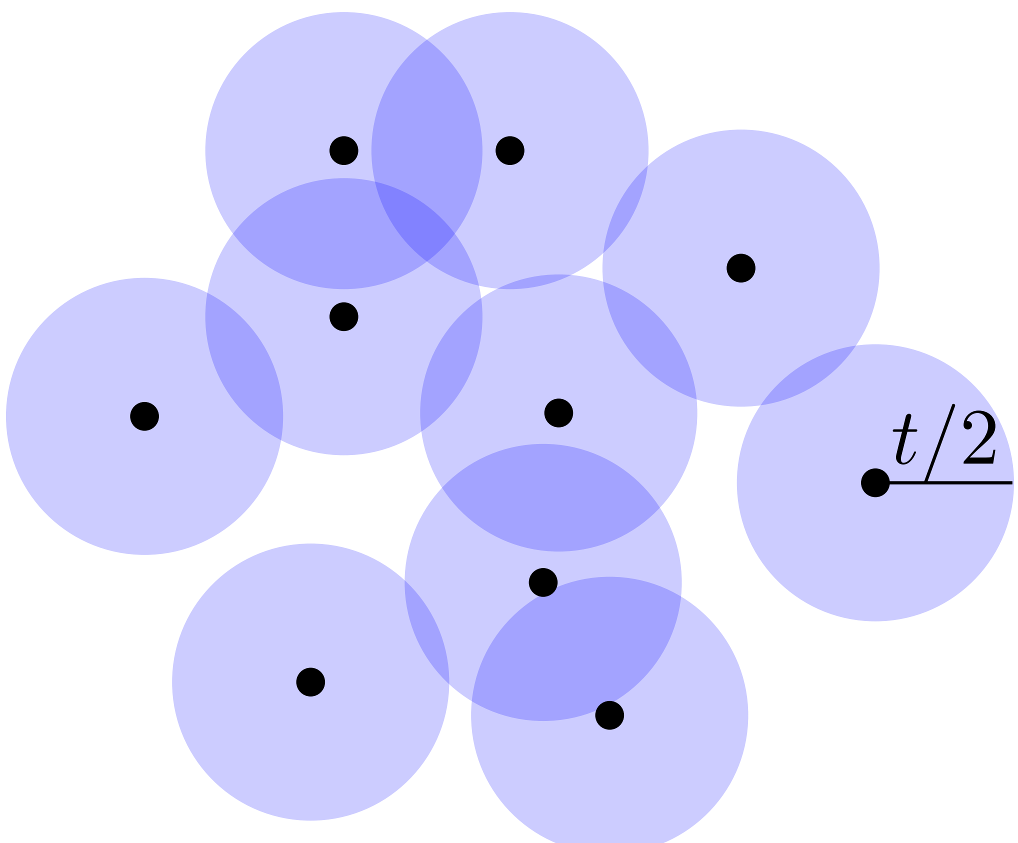
How discriminative are persistence diagrams?

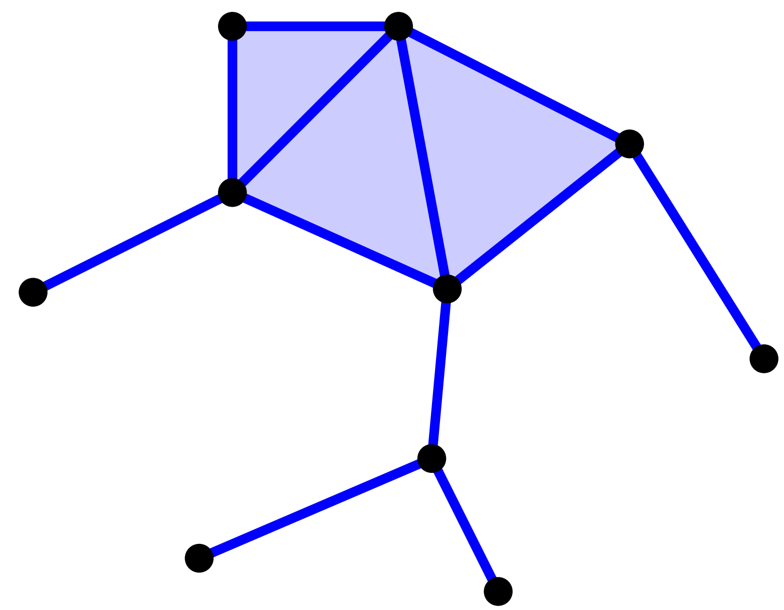
- Unions of balls — Vietoris-Rips filtrations



How discriminative are persistence diagrams?

- Unions of balls — Vietoris-Rips filtrations



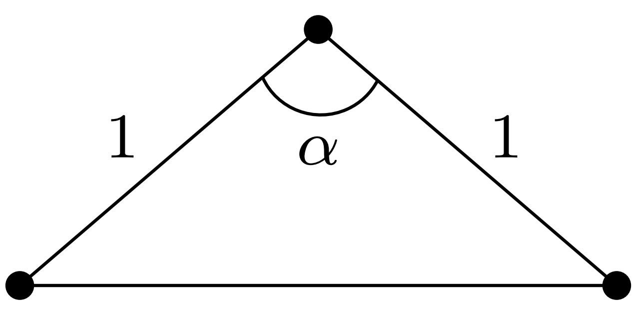


$\mathrm{R}_t(X, \mathrm{d}_X)$

$$\{x_0, \cdots, x_r\} \in R_t(X, \mathrm{d}_X) \quad \iff \quad t \geq \max_{i,j} \mathrm{d}_X(x_i, x_j)$$

How discriminative are persistence diagrams?

- Unions of balls — Vietoris-Rips filtrations



$$\operatorname{dgm} \mathcal{R}(P, \ell_2) = \{(0, +\infty)\} \sqcup \{(0, 1)\} \sqcup \{(0, 1)\}$$

$\Rightarrow$ diagrams for different values of α are indistinguishable

How discriminative are persistence diagrams?

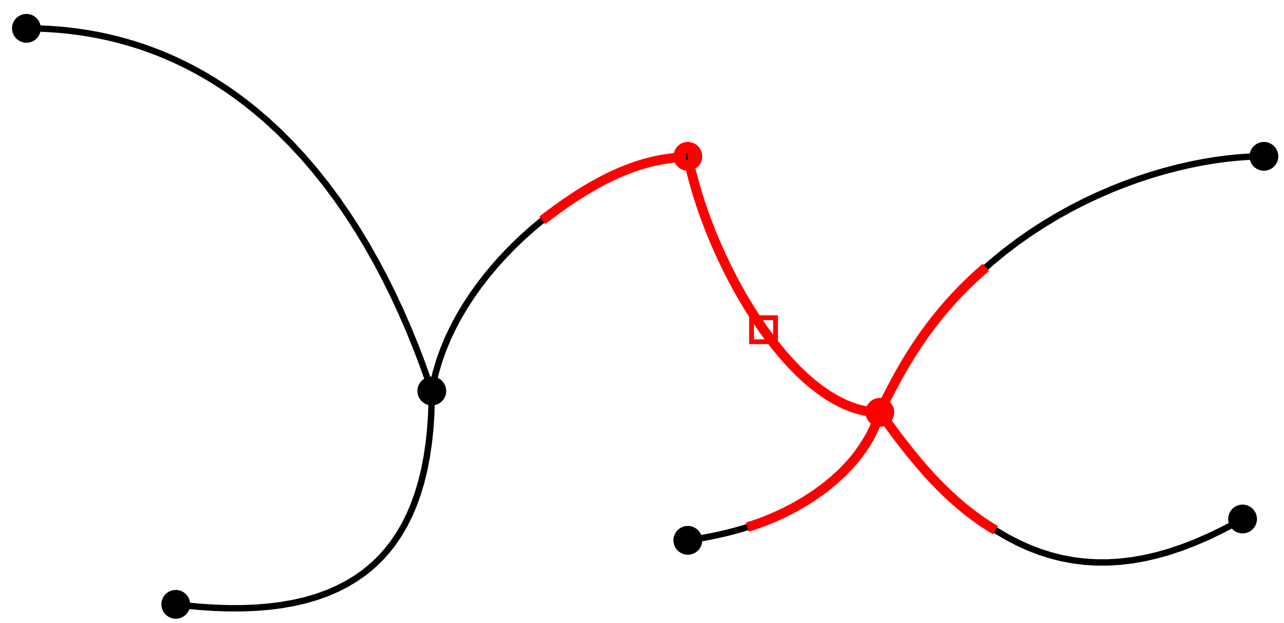
- Unions of balls — Vietoris-Rips filtrations

Prop: [Folklore]

For any *metric tree* (X, d_X):

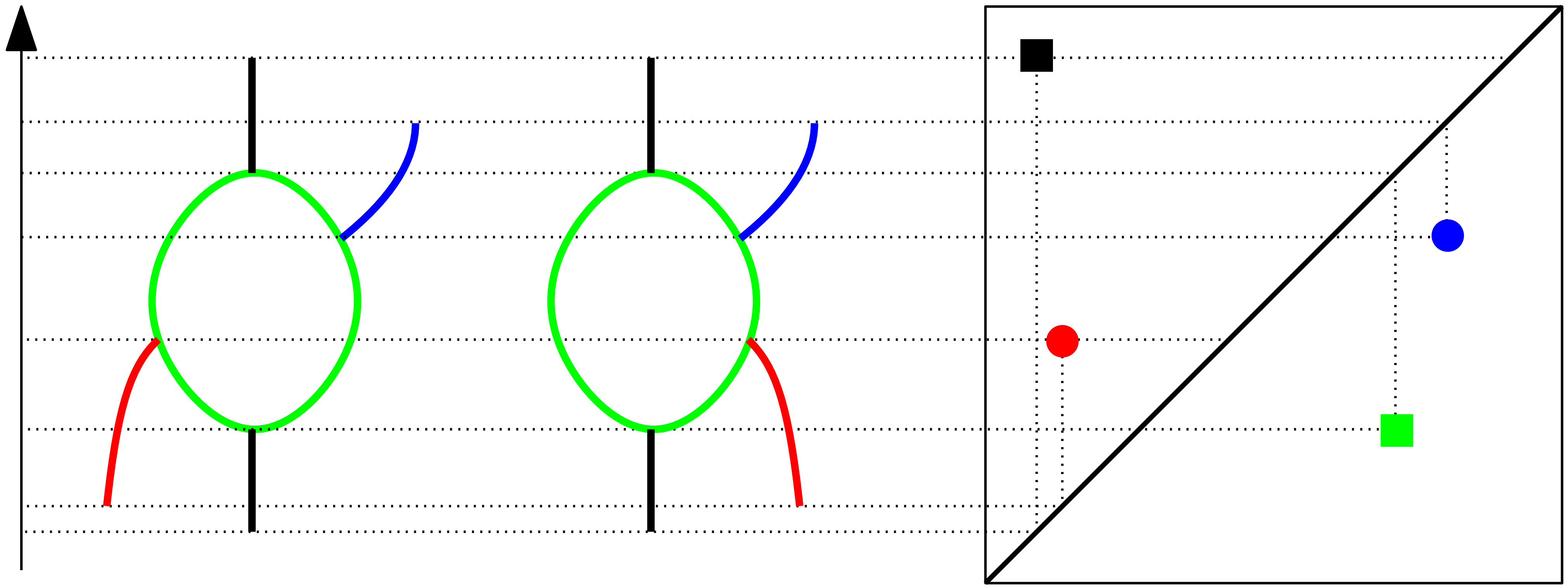
$$\mathrm{dgm}\, \mathcal{R}(X, \mathrm{d}_X) = \{(0, +\infty)\}$$

$\Rightarrow$ no information on the metric



How discriminative are persistence diagrams?

- Unions of balls — Vietoris-Rips filtrations
- Reeb graphs



⇒ Reeb graphs are indistinguishable from their diagrams

How discriminative are persistence diagrams?

- Unions of balls — Vietoris-Rips filtrations
- Reeb graphs
- Real-valued functions

Prop: [Folklore]
Given $f : X \to \mathbb{R}$ and $h : Y \to X$ homeomorphism,

$$\operatorname{dgm} f \circ h = \operatorname{dgm} f$$

$\Rightarrow$ Persistence is invariant under reparametrizations

How discriminative are persistence diagrams?

- Unions of balls — Vietoris-Rips filtrations
- Reeb graphs
- Real-valued functions

possible solutions:

- richer topological invariants (e.g. persistent homotopy)
- use several filter functions (**concatenation** vs **multipersistence**)

How discriminative are persistence diagrams?

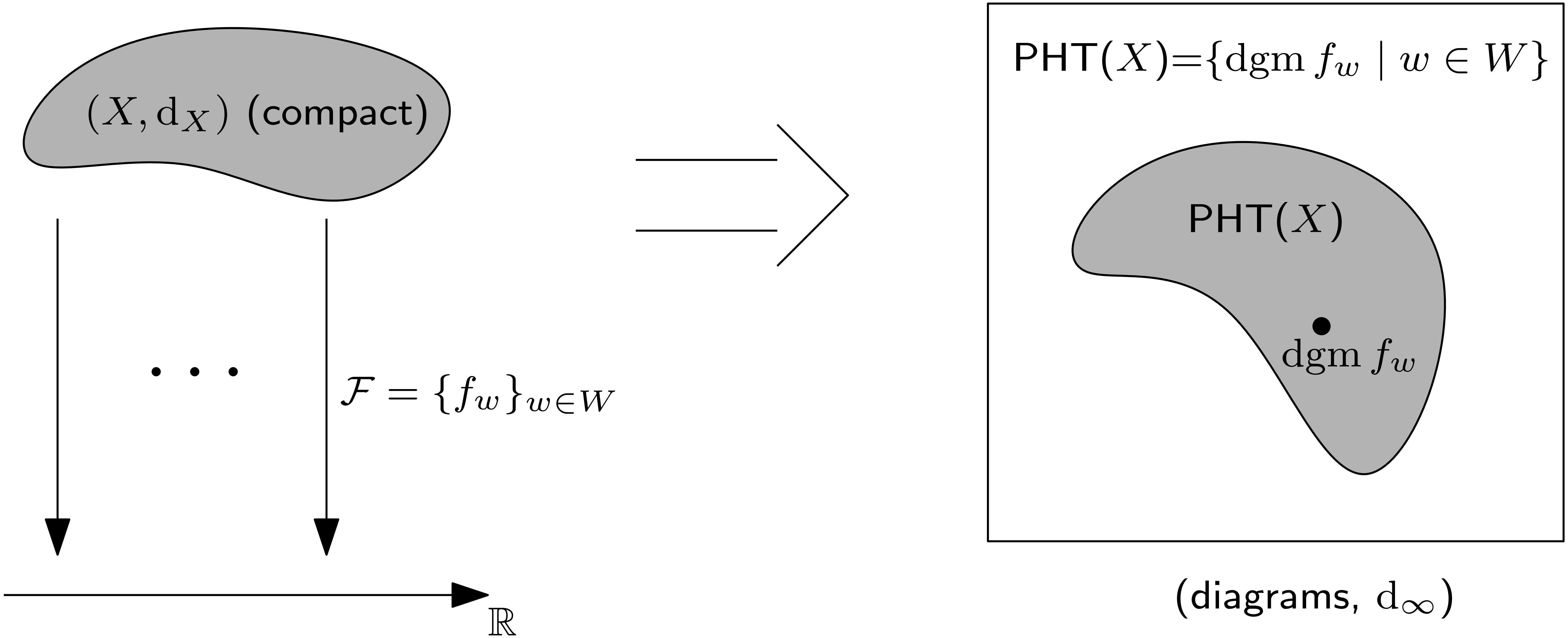
- Unions of balls — Vietoris-Rips filtrations
- Reeb graphs
- Real-valued functions

possible solutions:

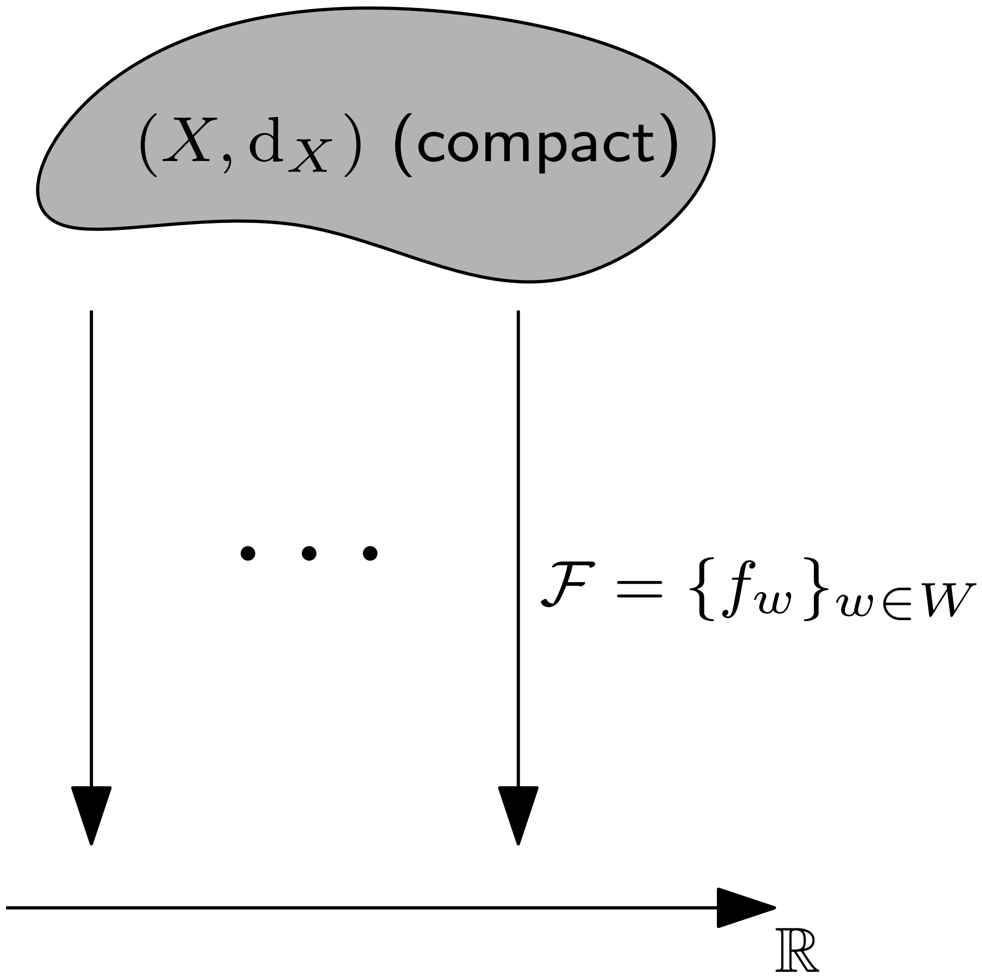
- richer topological invariants (e.g. persistent homotopy)
- use several filter functions (**concatenation** vs **multipersistence**)

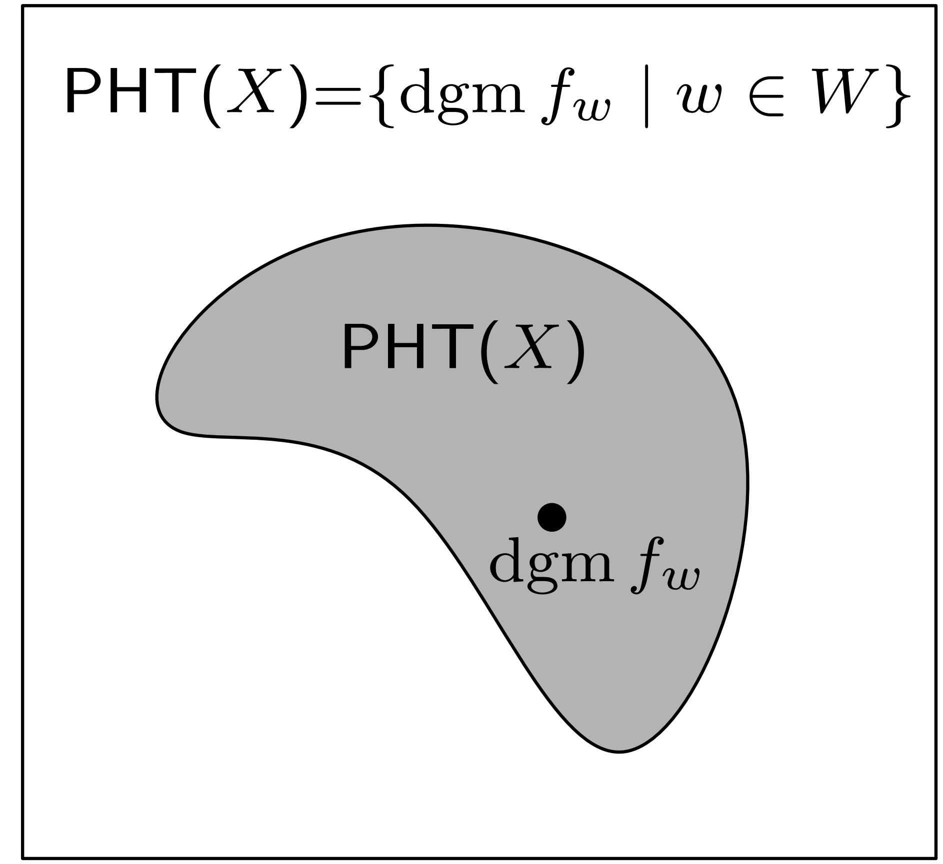
Steve Nicolas

Persistent Homology Transform (PHT)



Persistent Homology Transform (PHT)





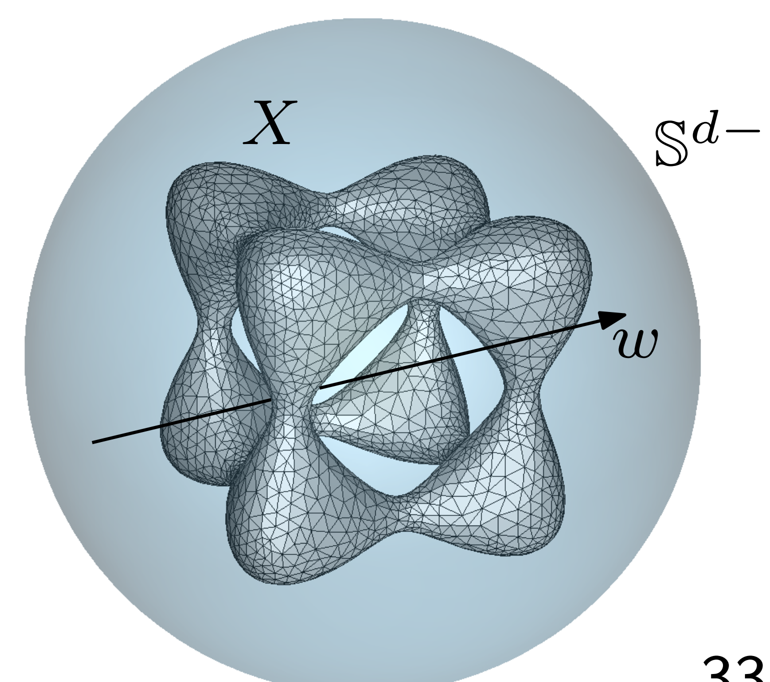
(diagrams, d_∞)

Thm: [Turner, Mukherjee, Boyer 2014]

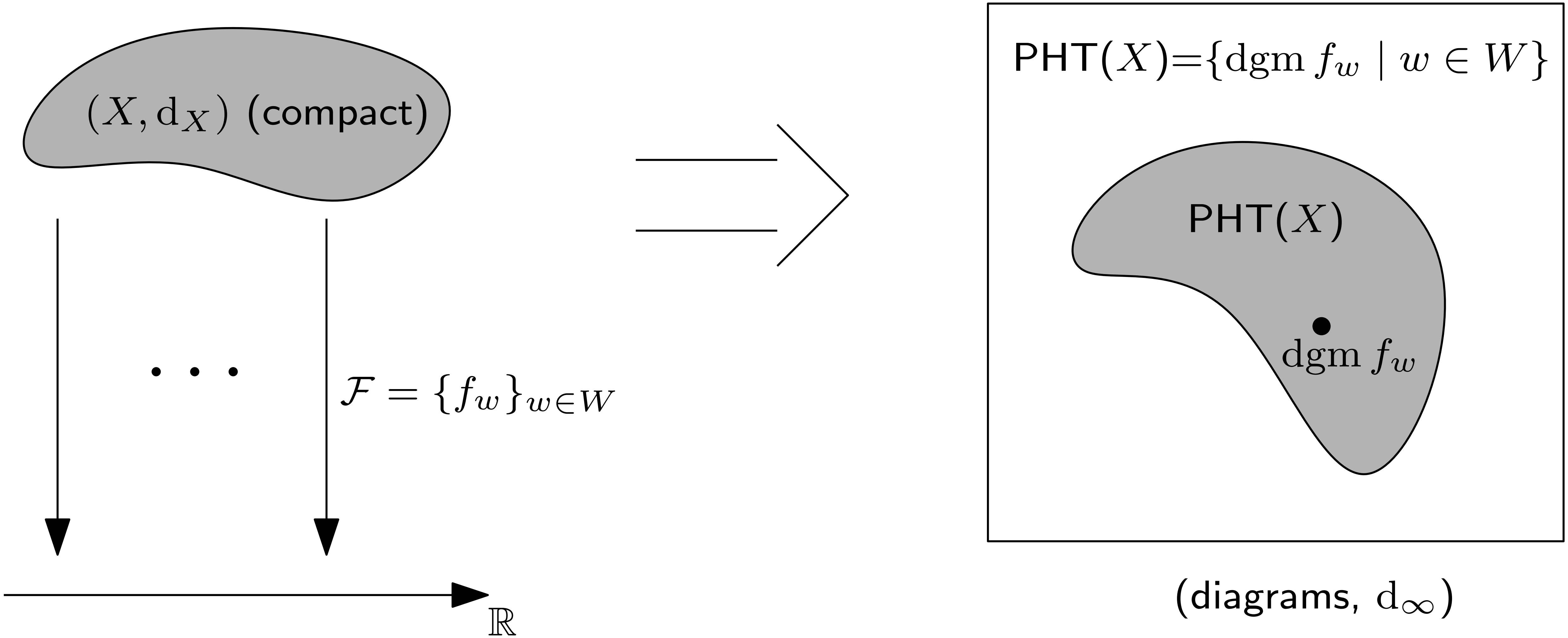
Let $\mathcal{F} = \{\langle \cdot, w \rangle\}_{w \in \mathbb{S}^{d-1}}$, where $d = 2, 3$ is fixed. Then, PHT is injective on the set of linear embeddings of compact simplicial complexes in $\mathbb{R}^d$.

Extension: [Turner et al., in progress]

True for arbitrary d and semialgebraic compact sets.



Persistent Homology Transform (PHT)

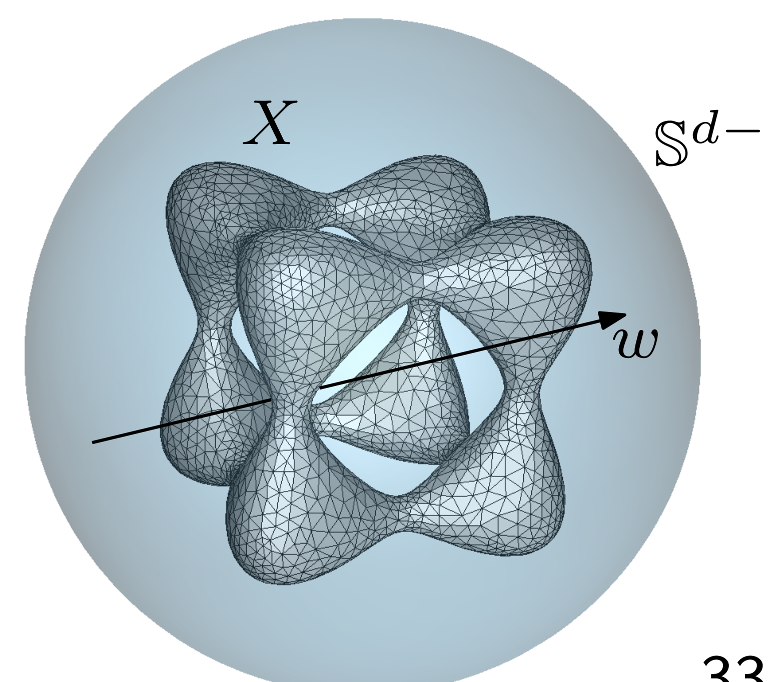


Thm: [Turner, Mukherjee, Boyer 2014]

Let $\mathcal{F} = \{\langle \cdot, w \rangle\}_{w \in \mathbb{S}^{d-1}}$, where $d = 2, 3$ is fixed. Then, PHT is injective on the set of linear embeddings of compact simplicial complexes in $\mathbb{R}^d$.

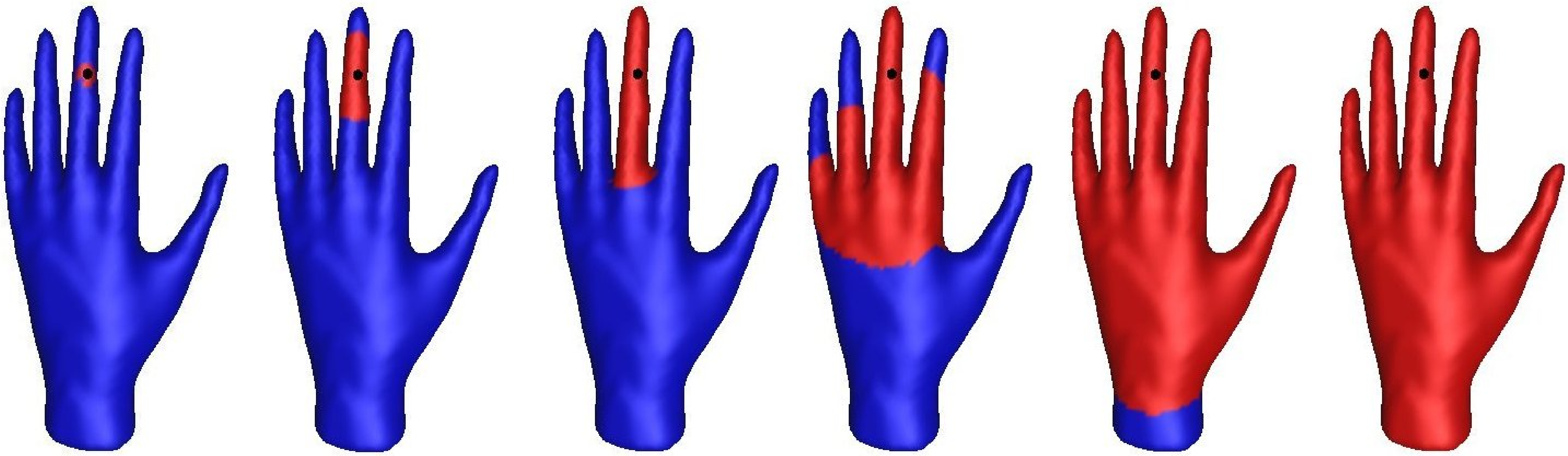
Corollary: PHT is a sufficient statistic for such sets

$\Rightarrow$ parametric inference



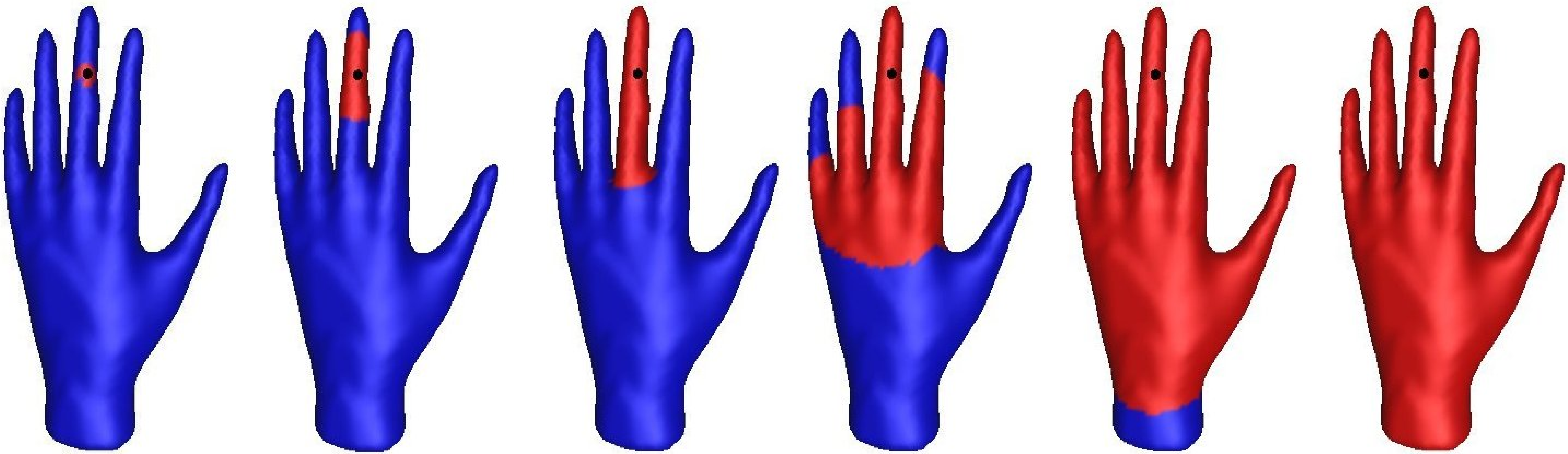
PHT for intrinsic metrics

Given (X, d_X) compact length space, take $\mathcal{F} = \{\mathrm{d}_X(\cdot, x)\}_{x \in X}$



PHT for intrinsic metrics

Given (X, d_X) compact length space, take $\mathcal{F} = \{\mathrm{d}_X(\cdot, x)\}_{x \in X}$



Thm: [O., Solomon 2017]

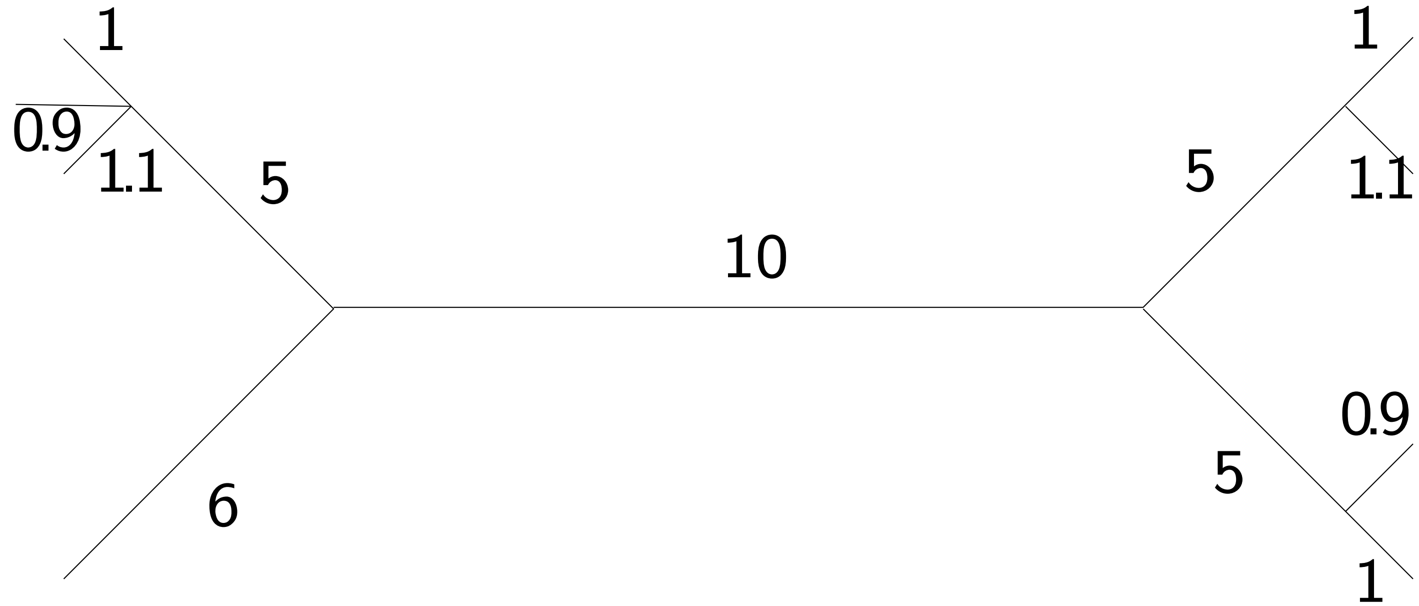
There is a Gromov-Hausdorff dense subset of the compact length spaces on which the intrinsic PHT is injective.

Generic injectivity

Generative model:

metric graph $\equiv$ combinatorial graph (V, E) + edge weights $E \to \mathbb{R}_+$

mixture (proba. mass function , proba. measure with density on $\mathbb{R}_+^{|E|}$)



Generic injectivity

Generative model:

metric graph $\equiv$ combinatorial graph (V, E) + edge weights $E \to \mathbb{R}_+$

mixture (proba. mass function , proba. measure with density on $\mathbb{R}_+^{|E|}$)

Thm: [O., Solomon 2017]
Under this model, there is a full-measure subset of the metric graphs on which the intrinsic PHT is injective.

Generic injectivity

Generative model:

metric graph $\equiv$ combinatorial graph (V, E) + edge weights $E \to \mathbb{R}_+$

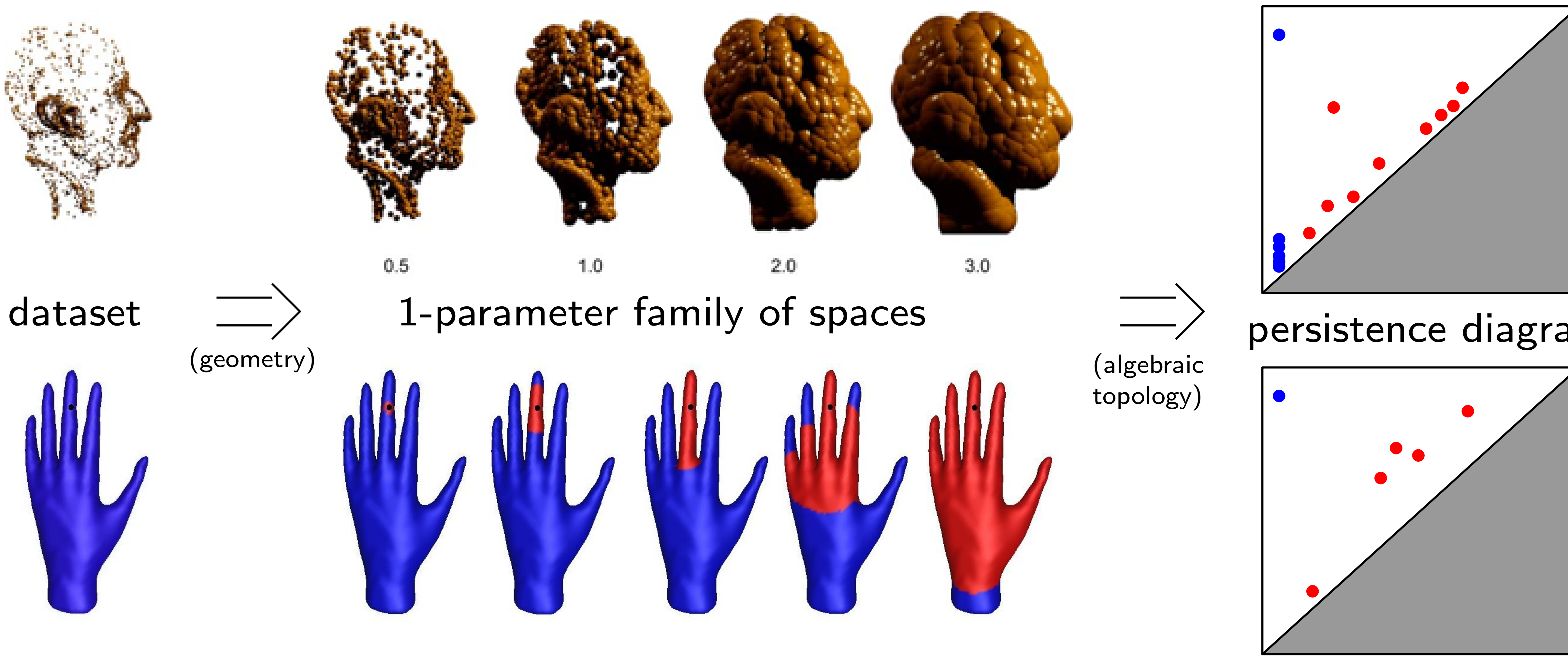
mixture (proba. mass function , proba. measure with density on $\mathbb{R}_+^{|E|}$)

Thm: [O., Solomon 2017]

Under this model, there is a full-measure subset of the metric graphs on which the intrinsic PHT is injective.

Aim: PHT as a sufficient statistic for metric graphs
$\Rightarrow$ parametric inference

Persistence diagrams as descriptors for data



Pros:

- strong invariance and stability:

  $d_\infty(\operatorname{dgm} X, \operatorname{dgm} Y) \leq \operatorname{cst} d_{\mathrm{GH}}(X, Y)$

- information of a different nature
- flexible and versatile

Pros:

- quasi-isometric maps to Hilbert space
- kernel trick
- provable discriminativity on certain classes of spaces
- (statistics via push-forwards/pull-backs)