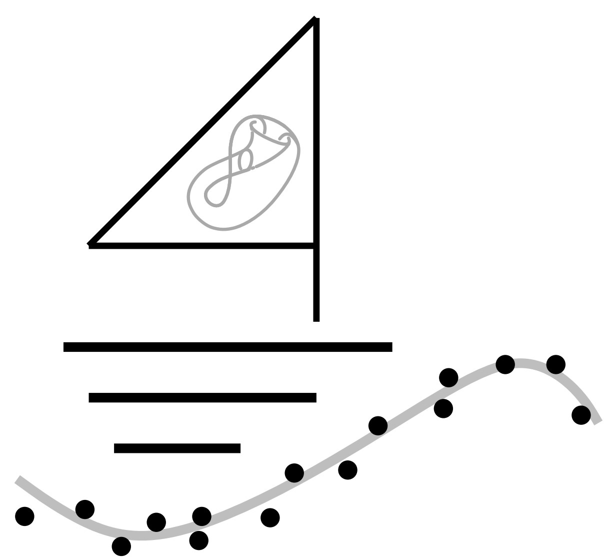
TopSail
Topology meets Statistical Learning



Steve Oudot

Principal Investigator

Current positions:

Inria (research scientist) / **École polytechnique** (part-time professor)

Research interests:

Computational geometry and topology, **Topological Data Analysis (TDA)** (since 2007)

Contributions:

Pioneering work on key aspects of TDA: foundational theorems, efficient algorithms & data structures, connections to statistical learning, applications

Author of one of the 2 reference books on TDA (AMS Monographs series)

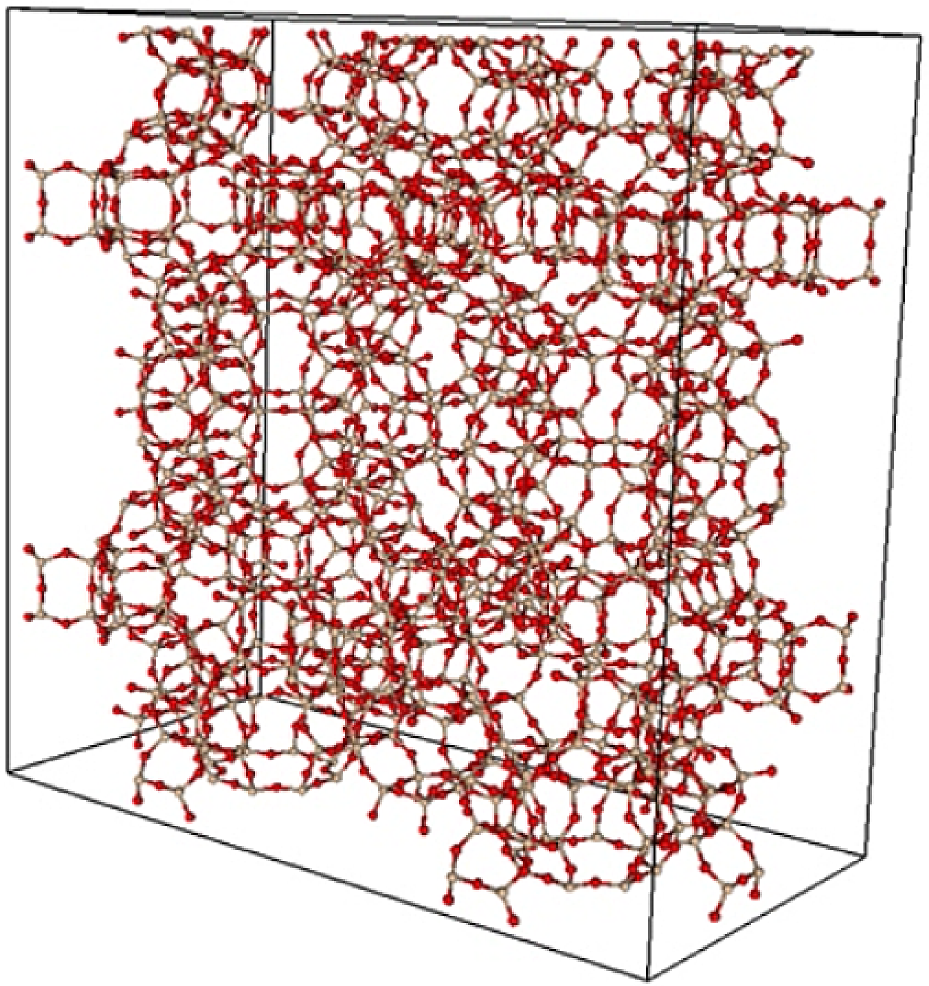
Recent highlights:

- notable publications (2017):
  - **SoCG** (flagship conference in Computational Geometry)
  - **ICML** (flagship conference in Machine Learning)
  - **J. FoCM** (rank A journal in Computational Mathematics)
- notable invitations:
  - plenary speaker at the **Abel Symposium** (2018)
  - senior fellow at **ICERM** (**Brown University**) (Fall 2016)

Context

Modern data sets often have non-trivial underlying structures...

...and these structures (shapes) matter



Example:

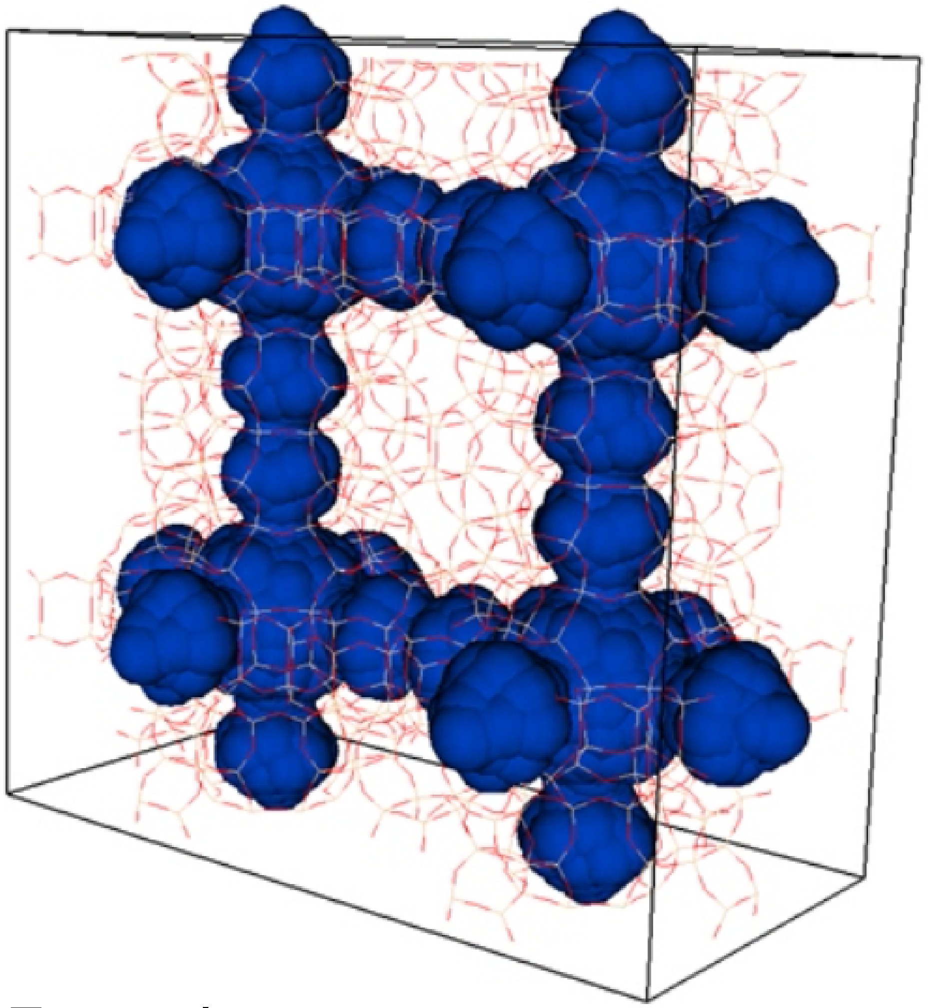
cavities in nanoporous materials
(e.g. zeolites for nanofilters)
determine their physical properties

[Y. Lee et al.: "Quantifying similarity of pore-geometry in nanoporous materials", 2017]

Context

Modern data sets often have non-trivial underlying structures...

...and these structures (shapes) matter



Example:

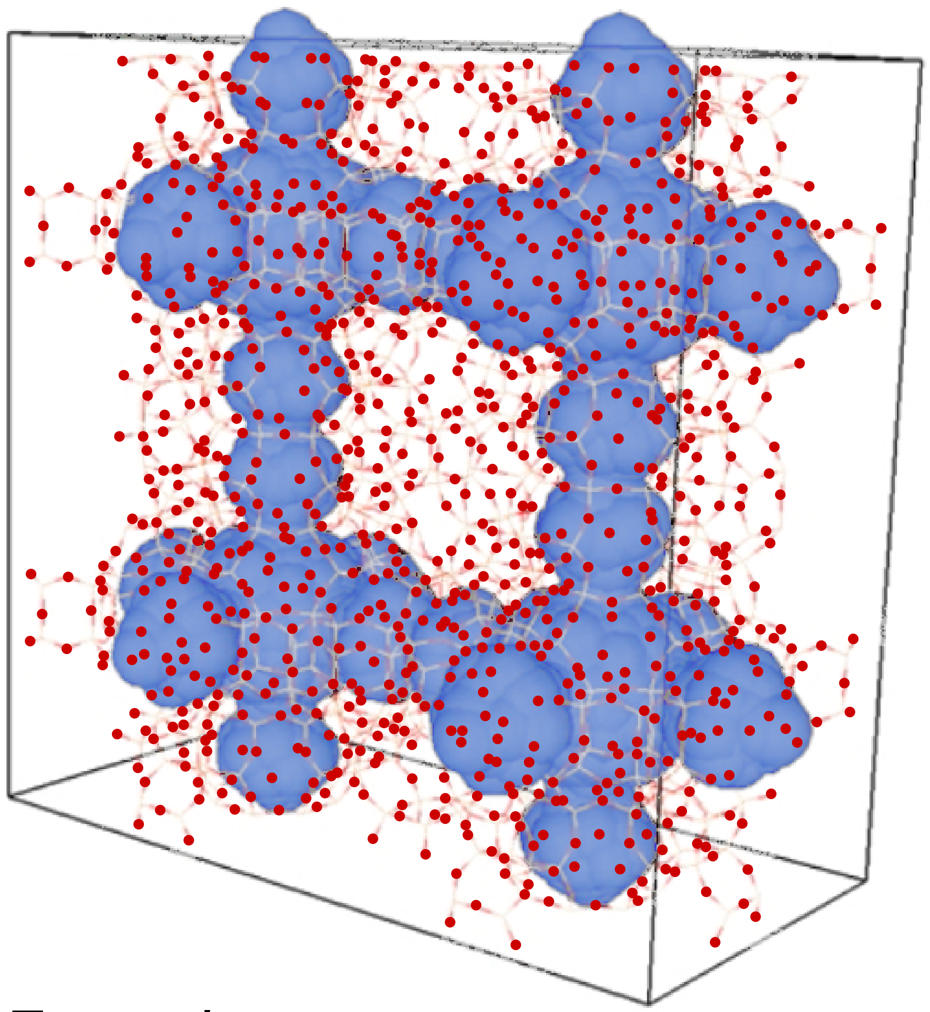
cavities in nanoporous materials
(e.g. zeolites for nanofilters)
determine their physical properties

[Y. Lee et al.: "Quantifying similarity of pore-geometry in nanoporous materials", 2017]

Context

Modern data sets often have non-trivial underlying structures...

...and these structures (shapes) matter



Example:

cavities in nanoporous materials
(e.g. zeolites for nanofilters)
determine their physical properties

$\rightsquigarrow$ need descriptors that can:

capture shapes hidden in data

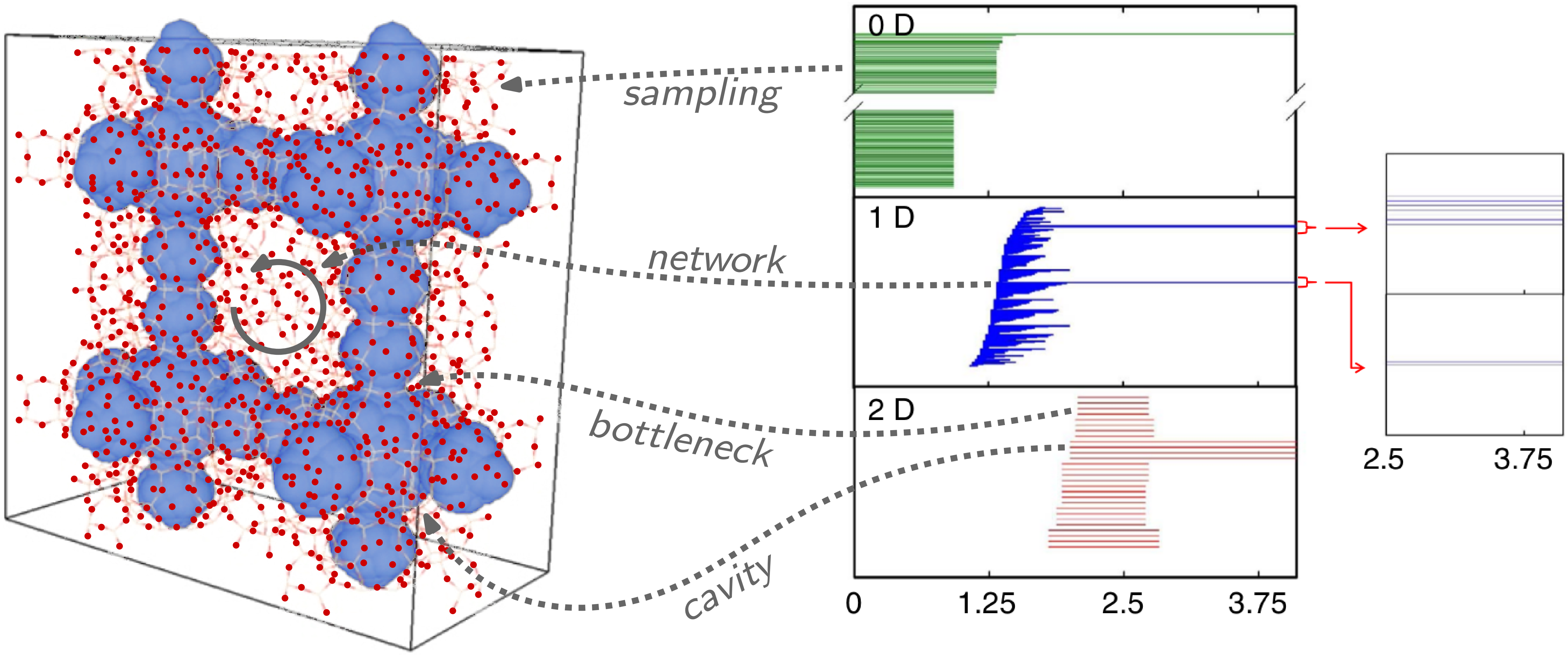
reveal these shapes to users

use tools from **Topology** (*Study of Shapes*)

[Y. Lee et al.: "Quantifying similarity of pore-geometry in nanoporous materials", 2017]

Topological Data Analysis (TDA)

Provides descriptors (barcodes) with distinguished properties



combine **geometry** & **topology**

general and **versatile**

provably stable

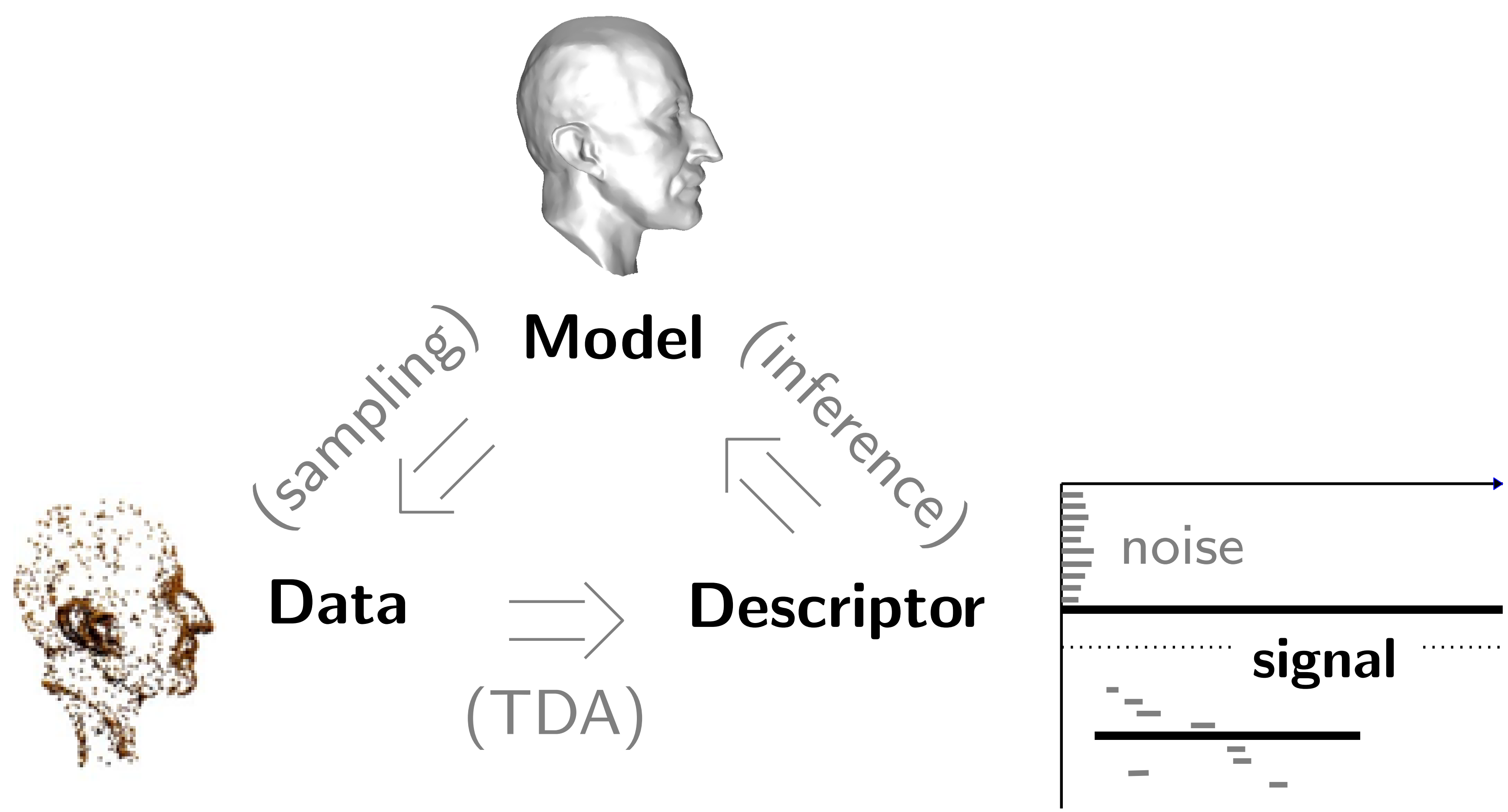
Vision & State of the Art

Topological descriptors: beyond a heuristic visualization tool?

model inference with guarantees

statistics on descriptors

learning with descriptors as features



Vision & State of the Art

Topological descriptors: beyond a heuristic visualization tool?

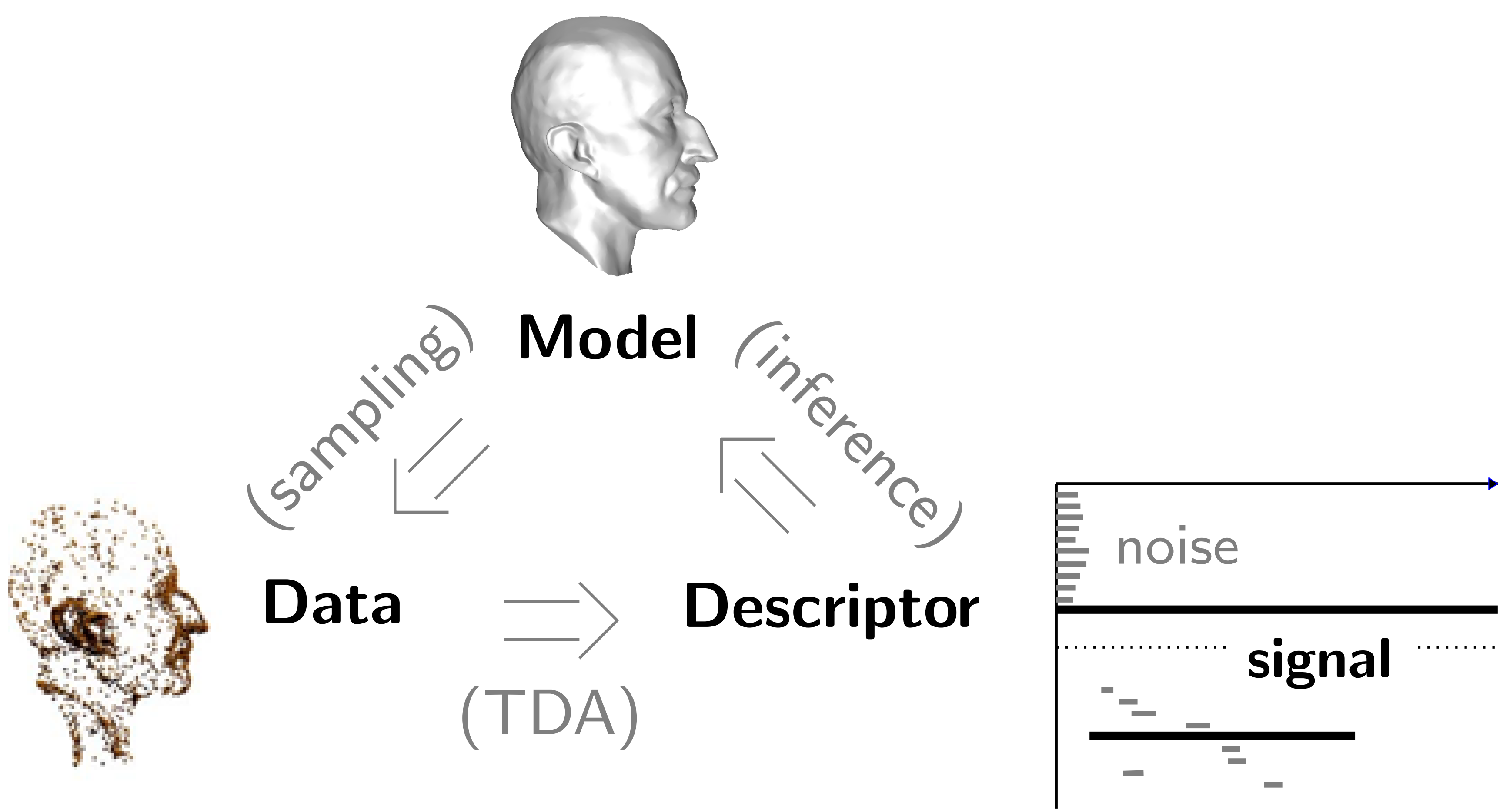
model inference with guarantees

statistics on descriptors

learning with descriptors as features

barcodes **lack sensitivity**

barcode space **not well-suited** for statistics and learning



Vision & State of the Art

Topological descriptors: beyond a heuristic visualization tool?

model inference with guarantees

statistics on descriptors

learning with descriptors as features

barcodes **lack sensitivity**

barcode space **not well-suited** for statistics and learning

State of the art (persistence theory):

structure theorems (existence of barcodes)

algorithms (computation of barcodes)

stability theorems (barcodes as stable descriptors)

+ limited attempts to do statistics/learning

Grand Challenge of TopSaiL

Revisit the **theoretical foundations of TDA** and combine them with **Optimal Transport** to fully **enable inference, statistics and learning** with topological descriptors

Objective 1: Richer / more informative topological descriptors

Objective 2: Framework for statistics and learning in descriptor space

Objective 3: Algorithms and practical applications

Objective 1: Richer topological descriptors

Goal: design new families of topological descriptors that carry richer information and are more informative.

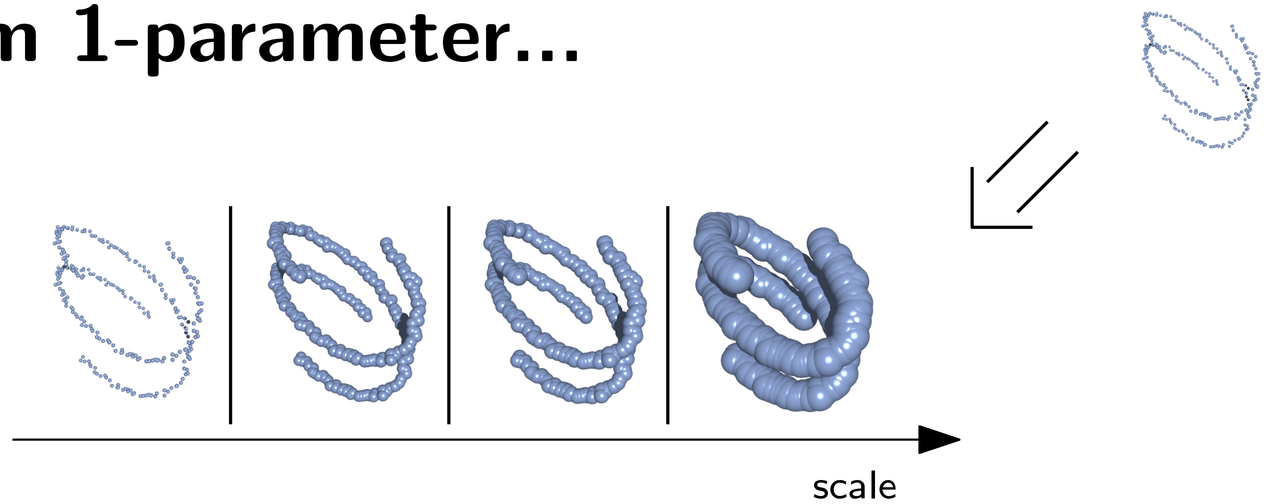
Directions:

Generalized barcodes via multi-parameter persistence

Topological (hyper-)graphs and their approximations

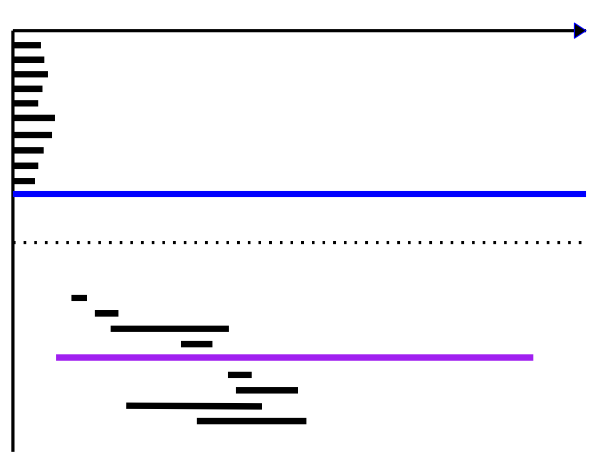
Objective 1: Richer topological descriptors

from 1-parameter...



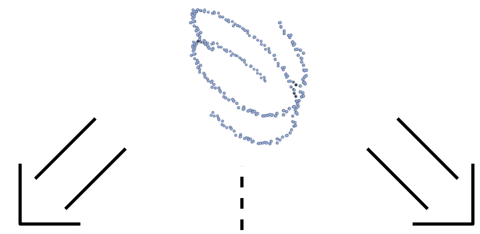
$\boldsymbol{k}[t]$-Mod:

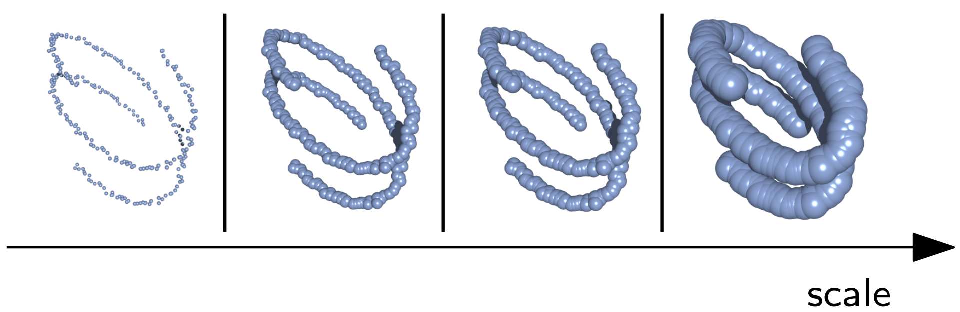
$$M \simeq \bigoplus_{i\in I} t^{b_i}\,\boldsymbol{k}[t] \quad \oplus \quad \bigoplus_{j\in J} t^{b_j}\,\frac{\boldsymbol{k}[t]}{t^{s_j}\,\boldsymbol{k}[t]}$$



Objective 1: Richer topological descriptors

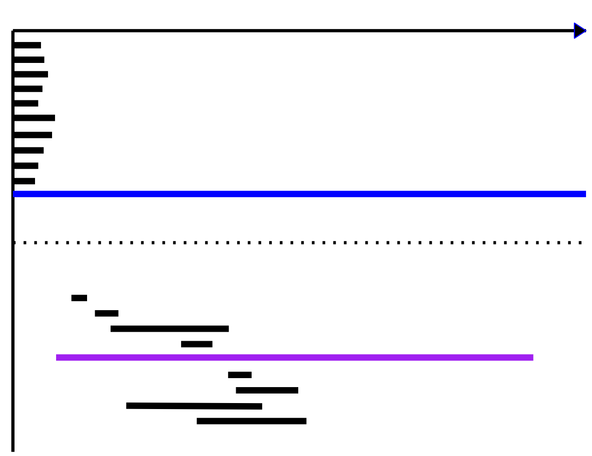
from 1-parameter...



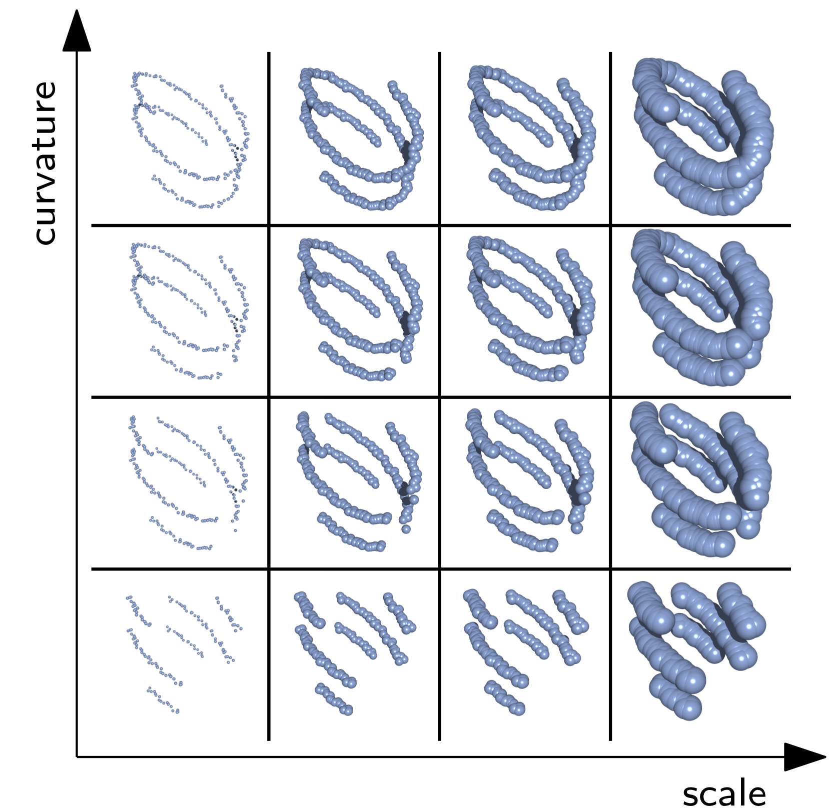


$\boldsymbol{k}[t]$-Mod:

$$M \simeq \bigoplus_{i\in I} t^{b_i}\boldsymbol{k}[t] \quad \oplus \quad \bigoplus_{j\in J} t^{b_j}\frac{\boldsymbol{k}[t]}{t^{s_j}\boldsymbol{k}[t]}$$



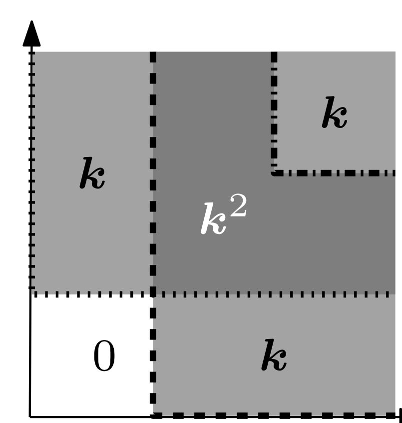
...to multi-parameter



$\boldsymbol{k}[t_1, \cdots, t_n]$-Mod:

$$M \simeq \bigoplus_{i\in I} M_i$$

$M_i \simeq$

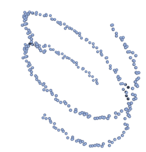


Problem:

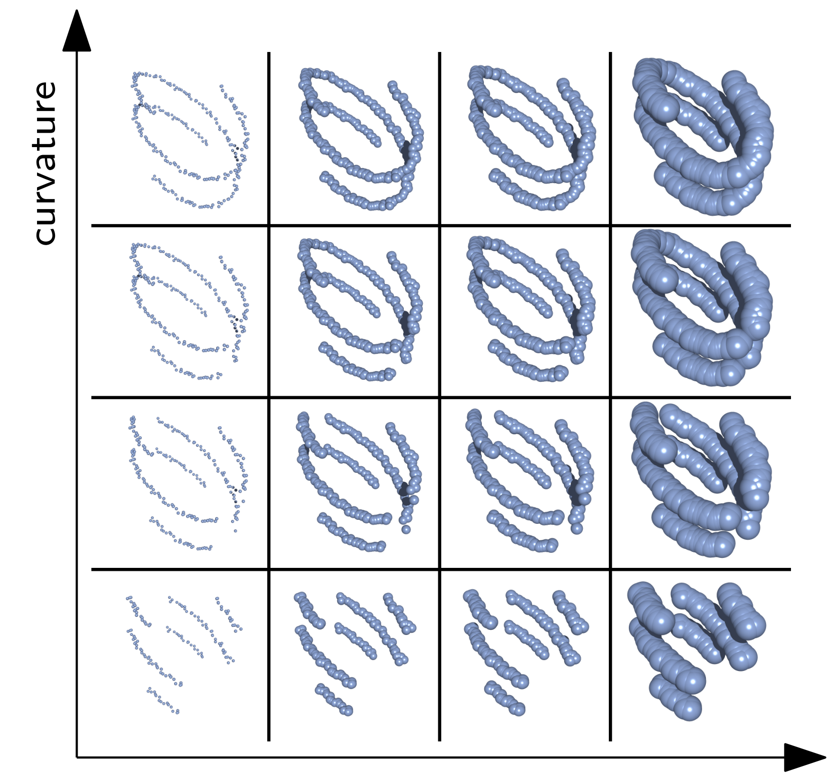
M_i can be arbitrarily 'complex'

[G. Carlsson, A. Zomorodian: "The theory of multidimensional persistence", 2009]

Objective 1: Richer topological descriptors



...to multi-parameter

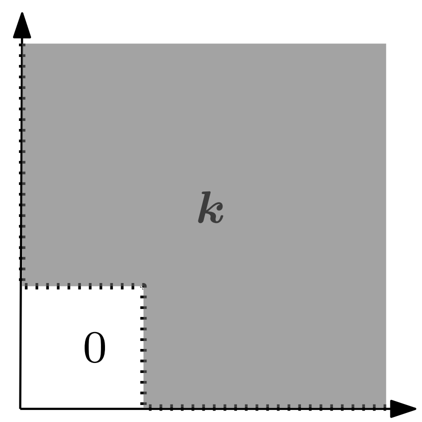


New approach:

approximation by ‘simpler’ (thin) modules

$$M \simeq \bigoplus_{i \in I} M_i$$

$M_i \simeq$



Hard questions:

approximation power of thin modules

stability theory for thin modules

effective computation of thin summands

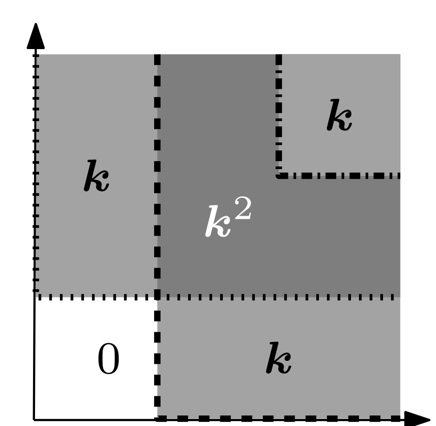
[preliminary results in the case $n = 2$

[J. Cochoy, S. Oudot:"Decomposition of exact pfd persistence bimodules", 2016]]

$\boldsymbol{k}[t_1, \cdots, t_n]$-Mod:

$$M \simeq \bigoplus_{i \in I} M_i$$

$M_i \simeq$



Problem:

M_i can be arbitrarily ‘complex’

[G. Carlsson, A. Zomorodian: "The theory of multidimensional persistence", 2009]

Objective 2: Stats and learning with descriptors

Goal: build a new framework for doing statistics and learning
in the space of topological descriptors

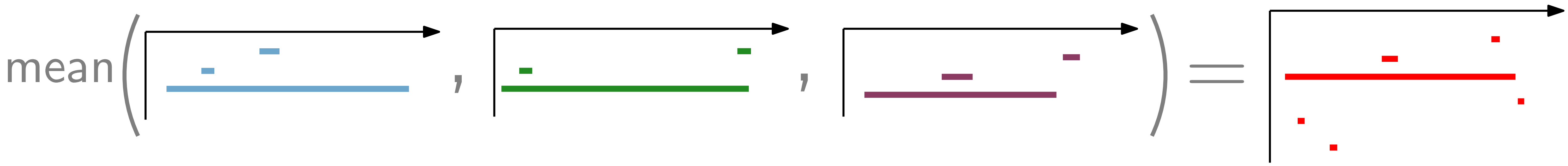
Directions:

$\left.\begin{array}{l}\textbf{Means and barycenters} \text{ for topological descriptors} \\ \\ \textbf{Kernels and vectorizations} \text{ for topological descriptors}\end{array}\right\}$ connection to Optimal Transport

Objective 2: Stats and learning with descriptors

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

Given barcodes $B_1, \cdots, B_n$:

$$\bar{B} \in \operatorname*{argmin}_B \ \frac{1}{n} \sum_i \mathrm{d}_{\mathrm{barcode}}(B,\, B_i)^2$$

Problem: non-convex energy, highly curved space

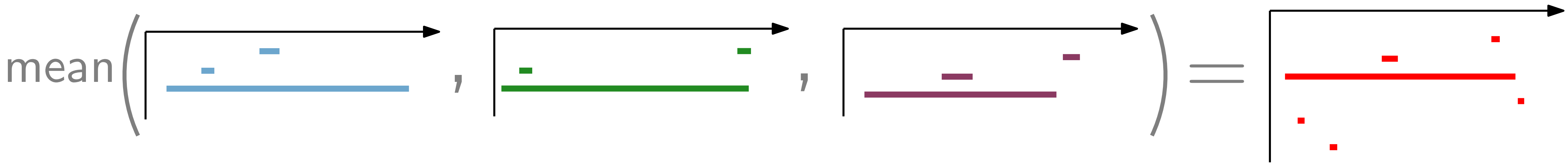
$\Rightarrow$ argmin not unique, local minima, numerical issues

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

Objective 2: Stats and learning with descriptors

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

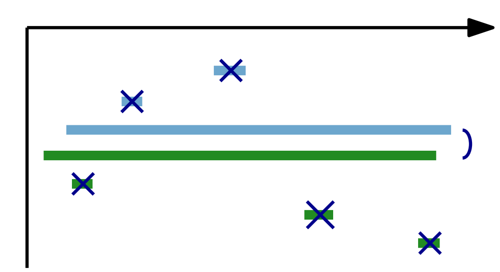
Given barcodes $B_1, \cdots, B_n$:

$$\bar{B} \in \operatorname*{argmin}_{B} \ \frac{1}{n} \sum_i \mathrm{d}_{\mathrm{barcode}}(B,\ B_i)^2$$

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

Problem: non-convex energy, highly curved space

$\Rightarrow$ argmin not unique, local minima, numerical issues

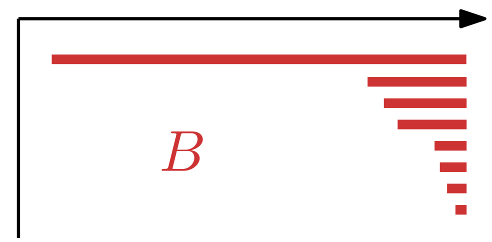


[K. Turner et al.: "Fréchet means for distributions of persistence diagrams", 2012]

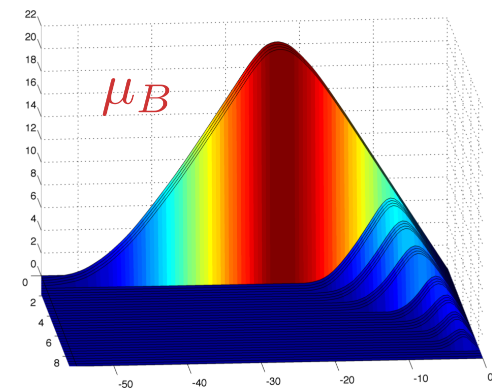
Objective 2: Stats and learning with descriptors

New approach: recast problem in measure space

$B \mapsto \mu_B$



$\Longrightarrow$



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

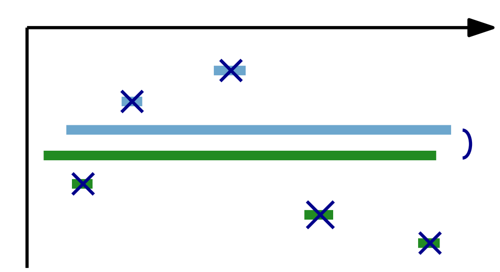
[M. Cuturi, A. Doucet: "Fast computation of Wasserstein barycenters", 2014]

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

Hard questions:

find the 'right' embedding (allows to go back)

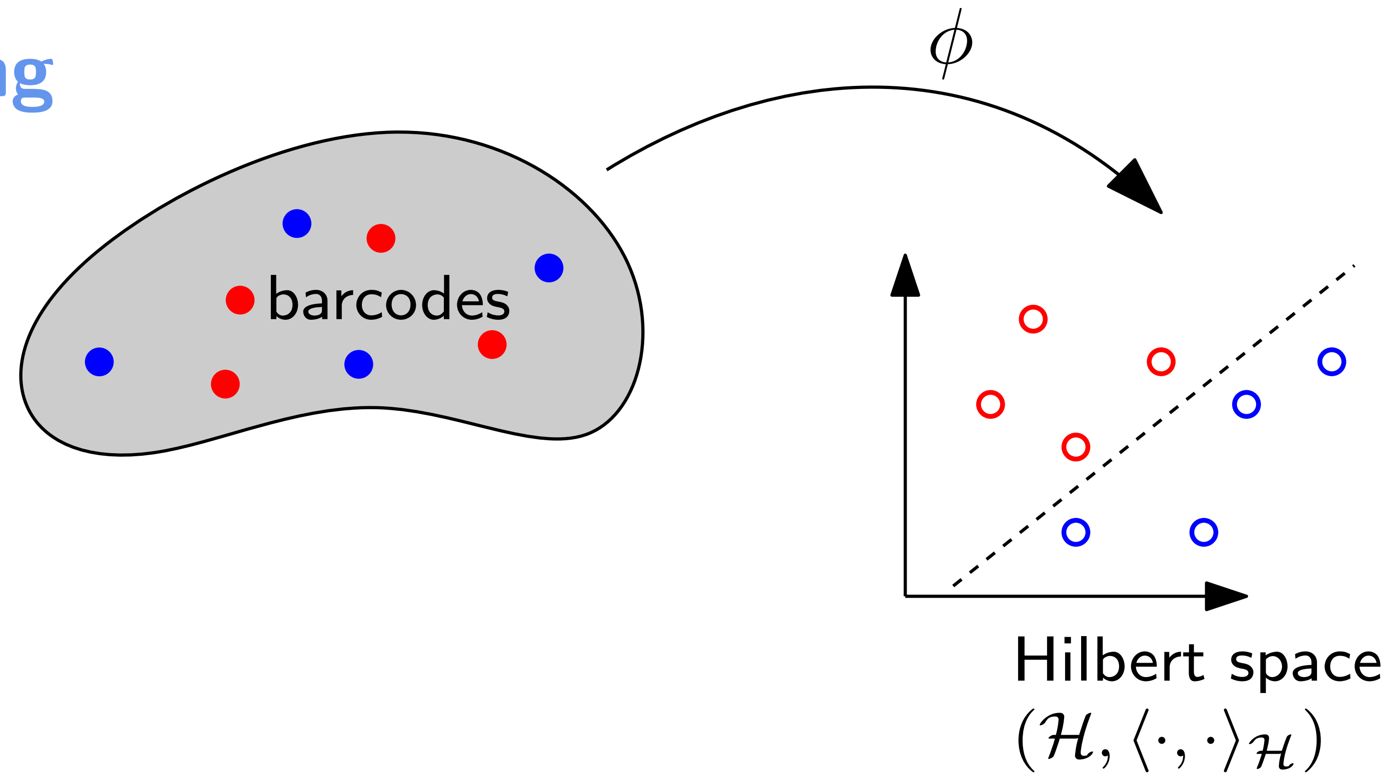
adapt/extend relaxations (highly unbalanced OT)



Objective 2: Stats and learning with descriptors

Kernels as a gateway to learning

features for learning algorithms,
combination with other features



vectorization: compute ϕ

kernel: compute $k = \langle \phi(\cdot), \phi(\cdot) \rangle_{\mathcal{H}}$

Problems:

structure of barcode space not preserved

data overfitting

[R. Kwitt et al.: "Statistical topological data analysis - A kernel perspective", 2015]

Objective 2: Stats and learning with descriptors

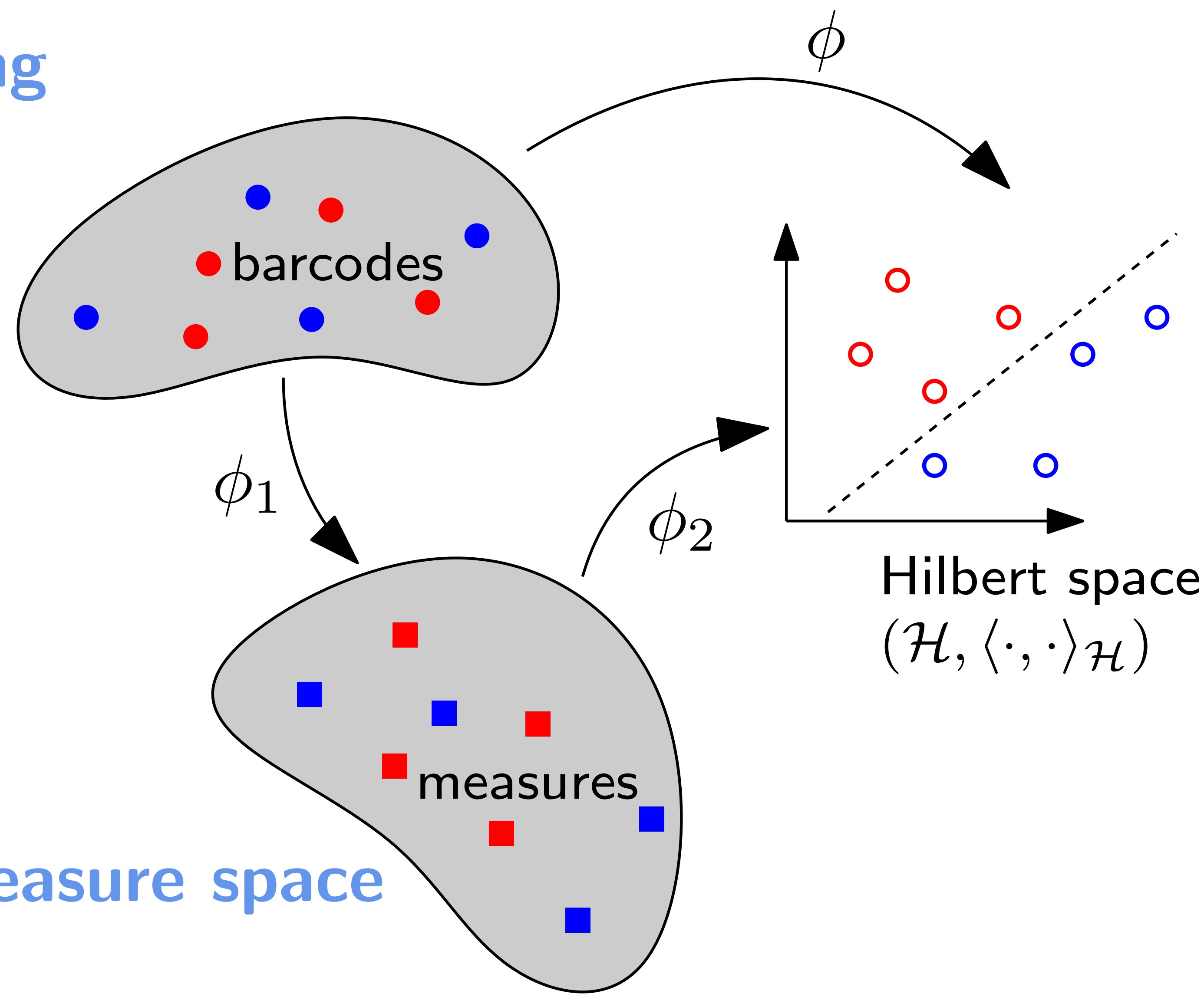
Kernels as a gateway to learning

features for learning algorithms,

combination with other features

vectorization: compute ϕ

kernel: compute $k = \langle \phi(\cdot), \phi(\cdot) \rangle_{\mathcal{H}}$



New approach: go through measure space

Hard questions:

preserve metric via kernels for measures

control rank of Gram matrix (no overfitting)

preliminary results with OT kernels

[M. Carrière, M. Cuturi, S. Oudot:"Sliced Wasserstein kernel for persistence diagrams", 2017]

Objective 3: Algorithms and applications

Software toolbox (open source library) with new-generation TDA descriptors & tools

Interaction with experts on specific data types:

image/geometry processing (M. Ovsjanikov)

genomics (C. Brisken, R. Rabadan)

dynamical systems (K. Mischaikow)

materials sciences (B. Smit)

TopSail at a glance

Refoundation of TDA and combination with **Optimal Transport**
to fully **enable statistical learning** with topological descriptors

Risk:

- at the interface between TDA and OT
- new angle on long-standing questions
  (e.g. multi-parameter persistence)

Impact:

- new light on the foundations of TDA
- diversified uses of topological descriptors
  (e.g. model inference, features for learning)

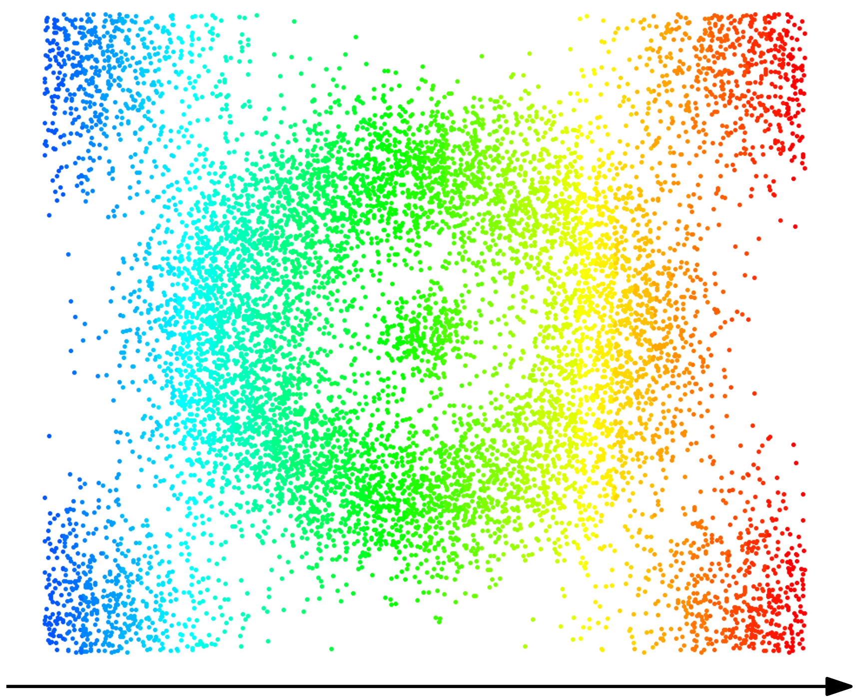
The TopSail research group:

- consolidates the PI's current group with (ERC funded) 1 early career researcher,
  3 PhD students, 2 postdocs, 1 engineer
- inspiring scientific environment: Paris-Saclay, TU Munich, Princeton, Columbia, Duke

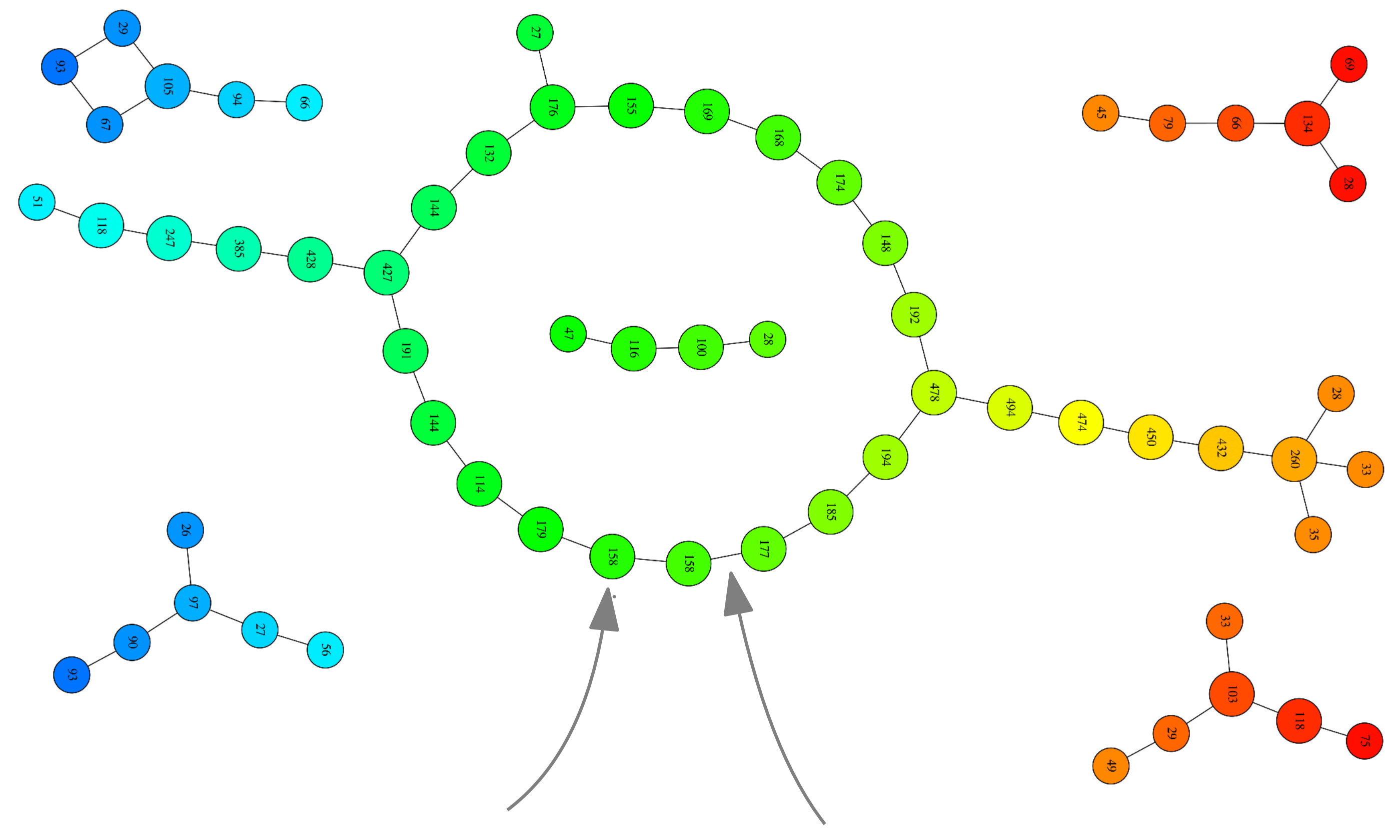
extra slides...

Topological graphs and generalizations

Principle: build an abstraction of data through the fibers of a map



$\Longrightarrow$



provides insight into:

- topology through global structure
- geometry through connection rules

$\Rightarrow$ **easier interpretation**

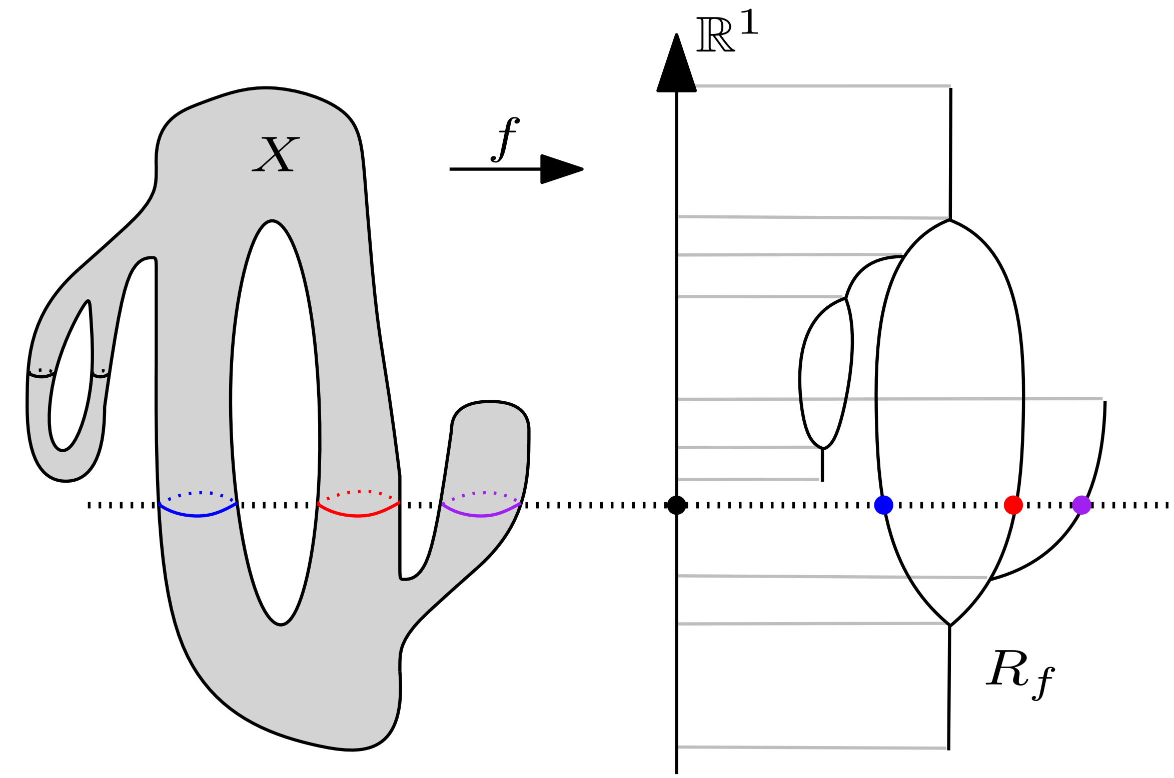
a

Topological graphs and generalizations

Mathematical model: Reeb space

Given $f : X \to \mathbb{R}^k$, $\quad \mathbf{R_f} := \mathbf{X}/\sim$

$x \sim y \iff \exists a \mid x, y \in$ same cc of $f^{-1}(\{a\})$



Approximations from point cloud data

- Mapper
- Joint Contour Net

[G. Singh, F. Mémoli, G. Carlsson: "Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition", 2007]

Connection model $\leftrightarrow$ approximations not well understood

recent progress in 1-d $(k = 1) \rightsquigarrow$ connection to Morse theory & 1-parameter TDA

[M. Carrière, S. Oudot: "Structure and stability of the 1-dimensional Mapper", 2016]

hard questions in higher dimensions $(k > 1) \rightsquigarrow$ connection to singularity theory

[O. Saeki: "Topology of singular fibers of differentiable maps", 2004]

Topological graphs and generalizations

New approach:

complete descriptors + metrics for topological hypergraphs via **sheaf theory**

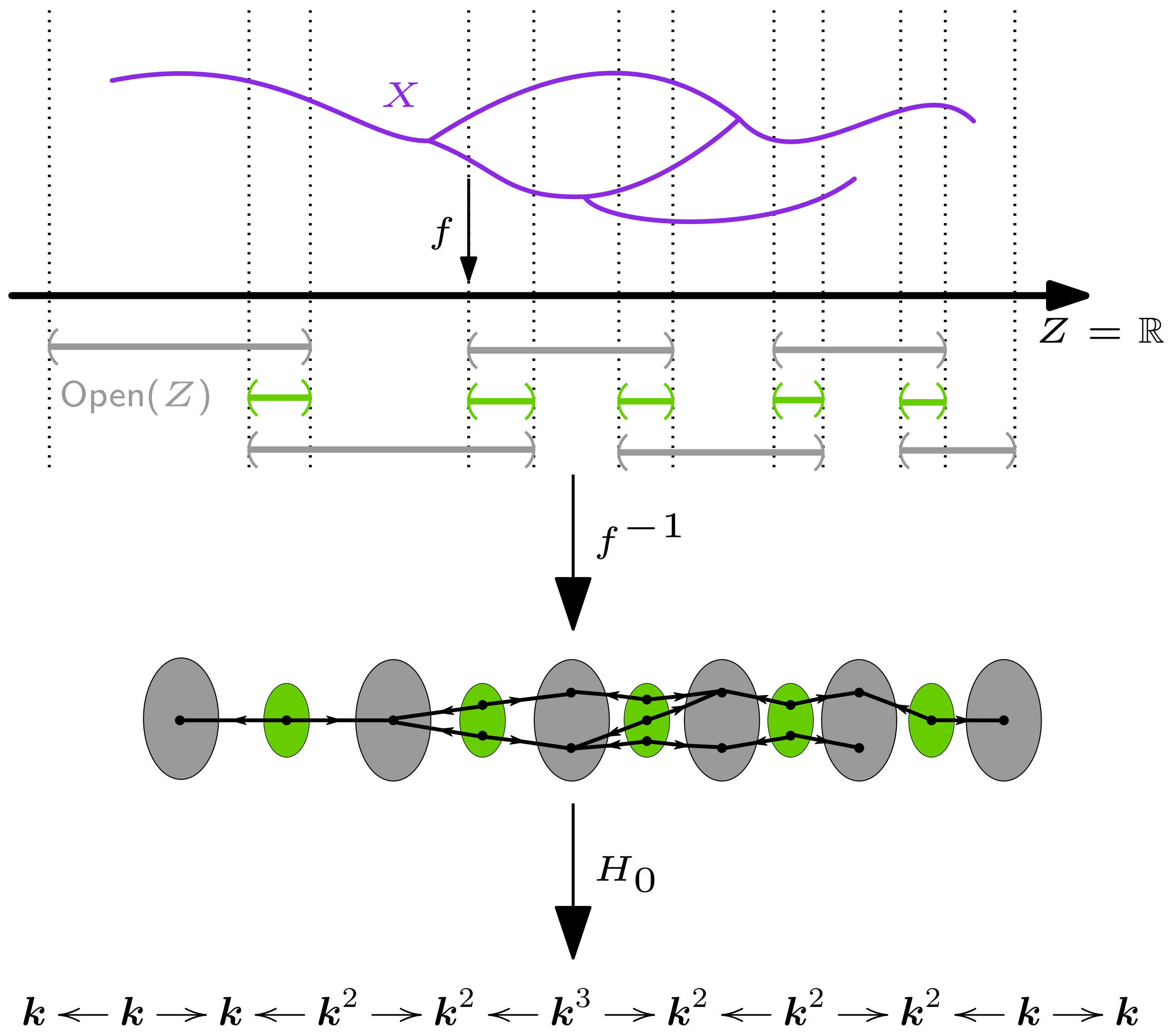
explore the **intrinsic metric space** of hypergraphs

Given $f : X \to Z$, take the **Leray cosheaf**:

$$F : \begin{array}{ccc} \mathrm{Open}(Z) & \longrightarrow & \mathrm{Vect} \\ U & \longmapsto & H_0(f^{-1}(U)) \end{array}$$

decomposition of $F \rightsquigarrow$ descriptor

interleaving of cosheaves $\rightsquigarrow$ metric



[J. Curry: "Sheaves, cosheaves and applications", 2014]

Hard questions:

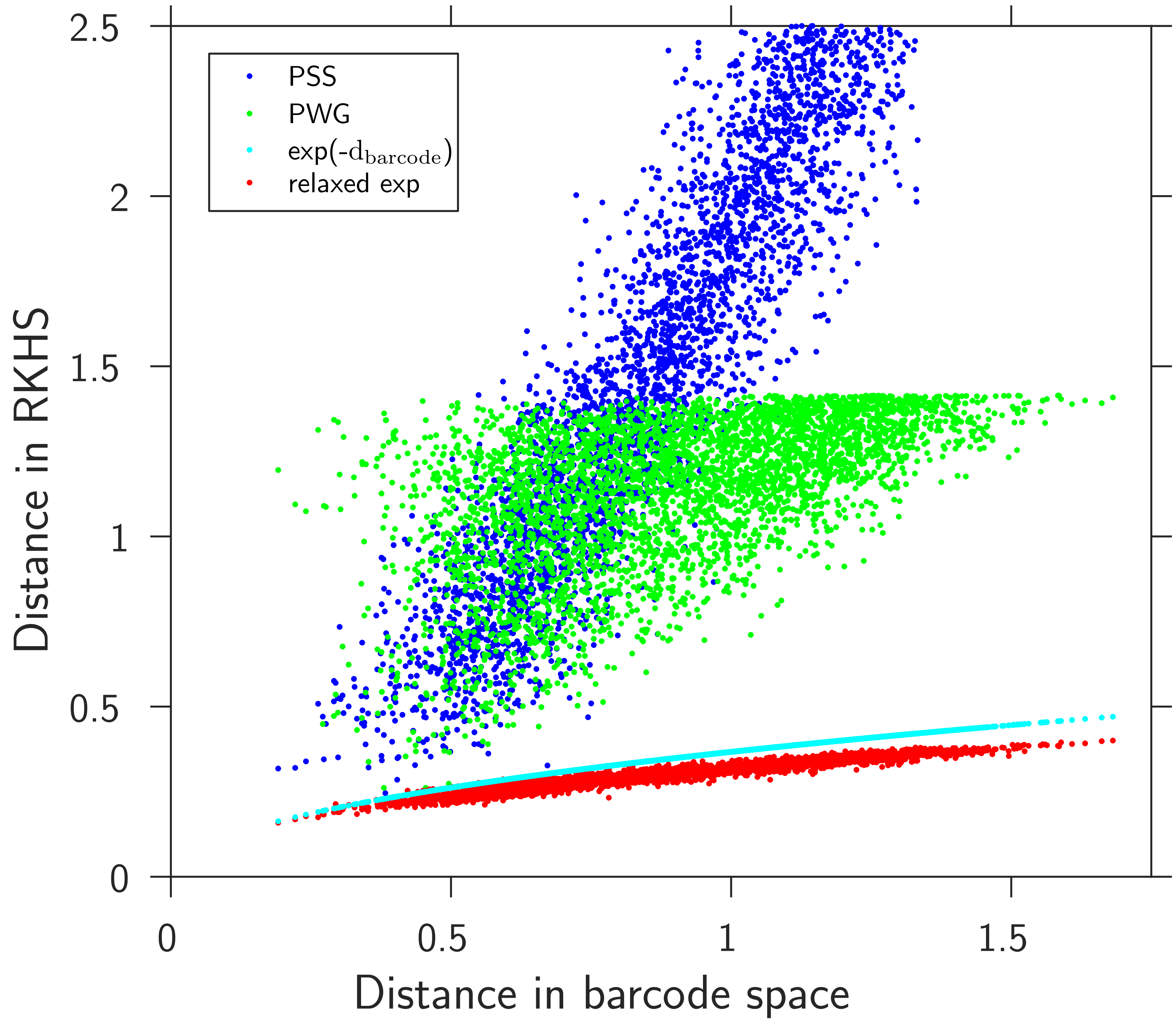
stability theory

effective computation of summands

(varying stratification / quiver)

a

Vectorizations & metric distortion

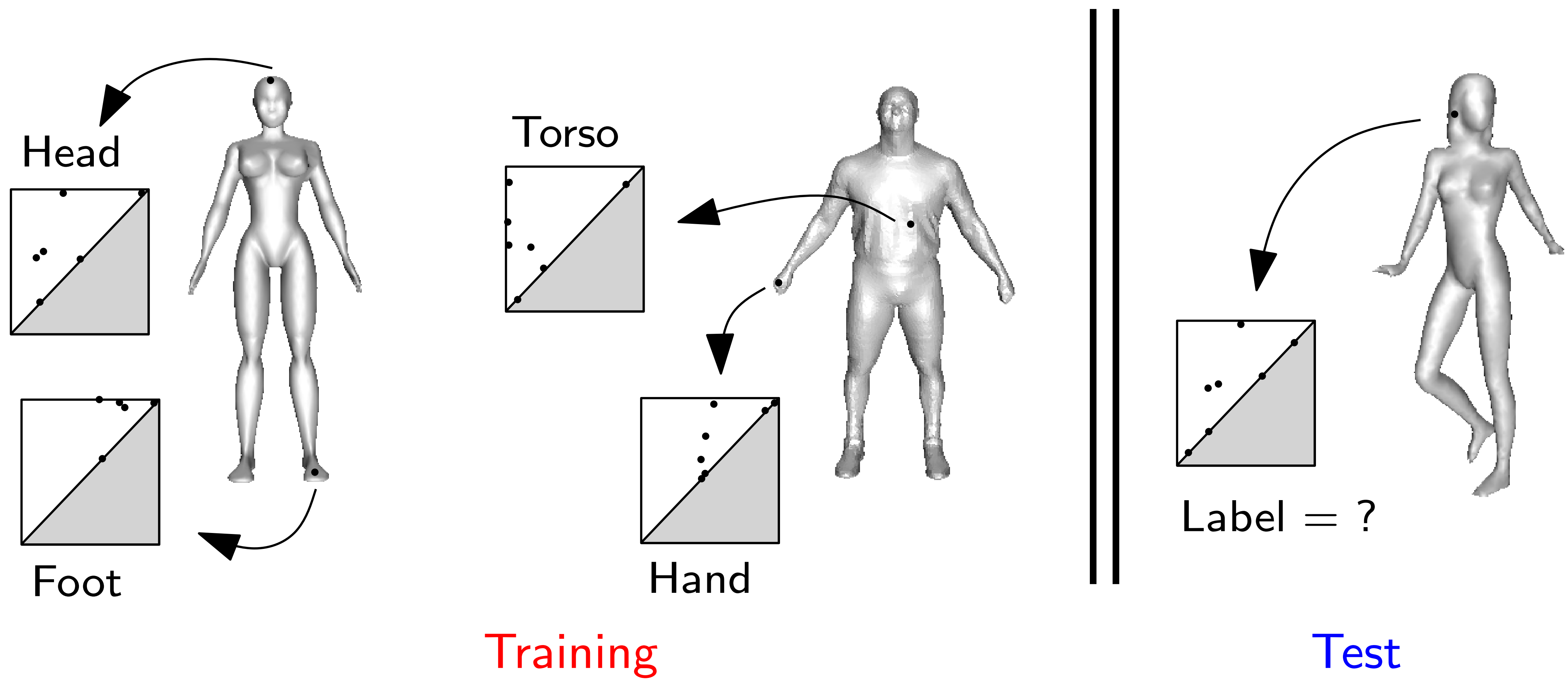


TDA for geometry processing

Task: example-based (supervised) segmentation of 3d shapes

Approach:

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



[M. Carrière, S. Oudot, M. Ovsjanikov: "Stable Topological Signatures for Points on 3D Shapes", 2015]

TDA for geometry processing

Task: example-based (supervised) segmentation of 3d shapes

Approach:

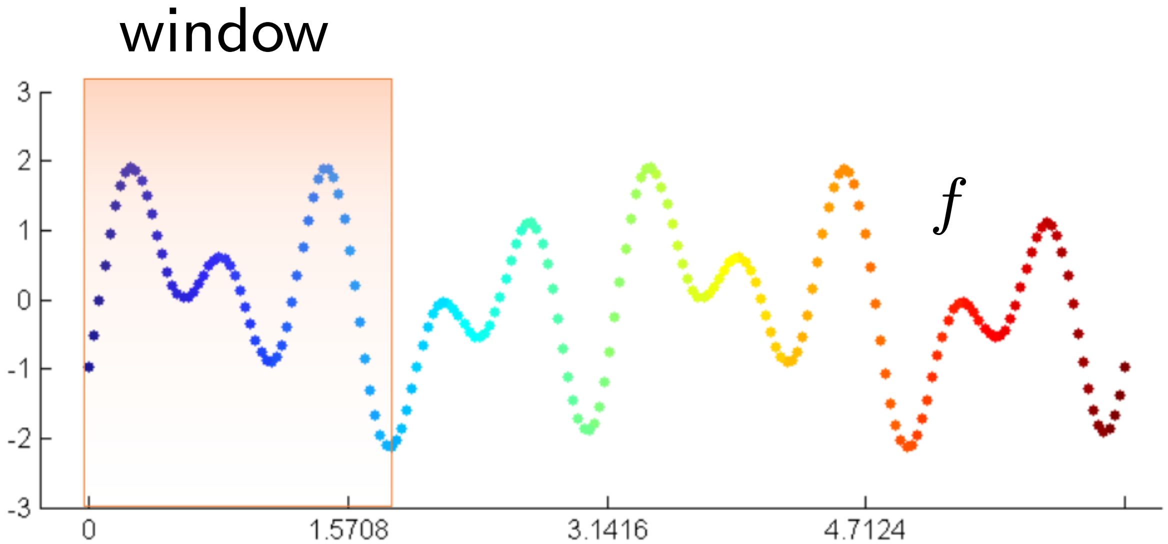
- train a (multiclass) classifier on barcodes extracted from training shapes
- apply classifier to barcodes extracted from query shape

Accuracies (%) using geometric vs TDA descriptors (5 training shapes):

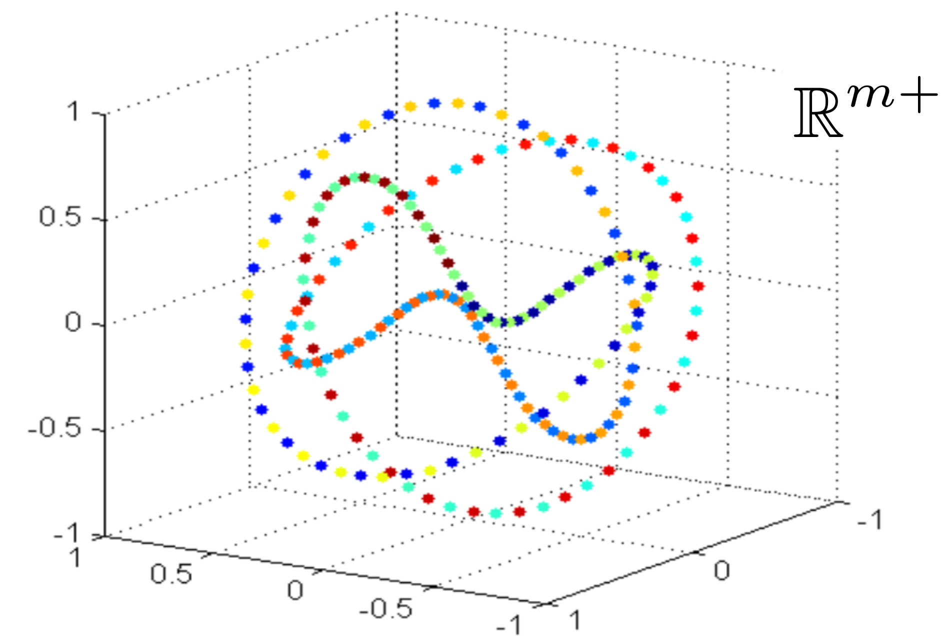
	TDA	geometry	TDA + geometry
Human	74.0	78.7	**88.7**
Airplane	72.6	81.3	**90.7**
Ant	92.3	90.3	**98.5**
FourLeg	73.0	74.4	**84.2**
Octopus	85.2	94.5	**96.6**
Bird	72.0	75.2	**86.5**
Fish	79.6	79.1	**92.3**

[M. Carrière, S. Oudot, M. Ovsjanikov: "Stable Topological Signatures for Points on 3D Shapes", 2015]

TDA for time series modeling & 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

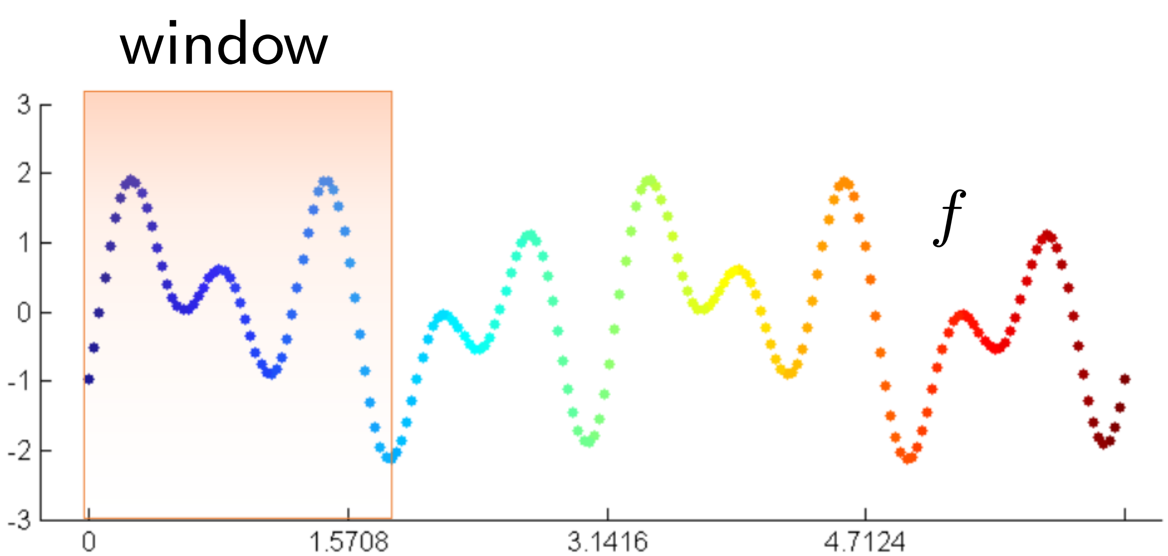
$m\tau$: window size

$m+1$: 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]

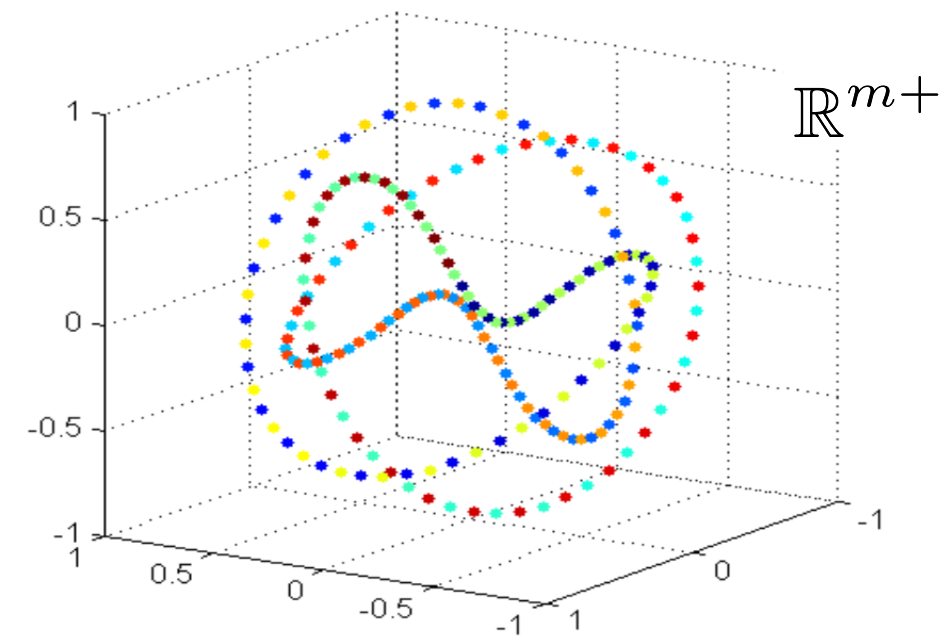
TDA for time series modeling & analysis



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

$\Longrightarrow$

(time-delay embedding)



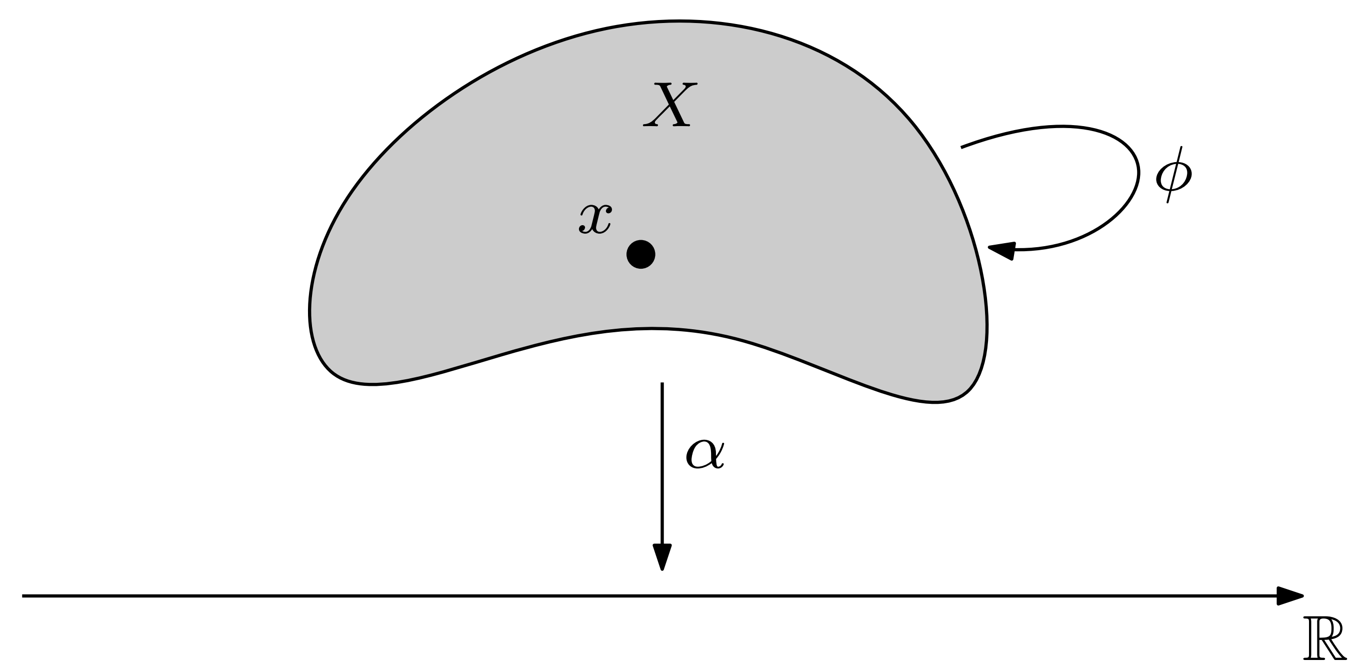
Contributions of TDA:

inference of:

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

no Fourier transform needed

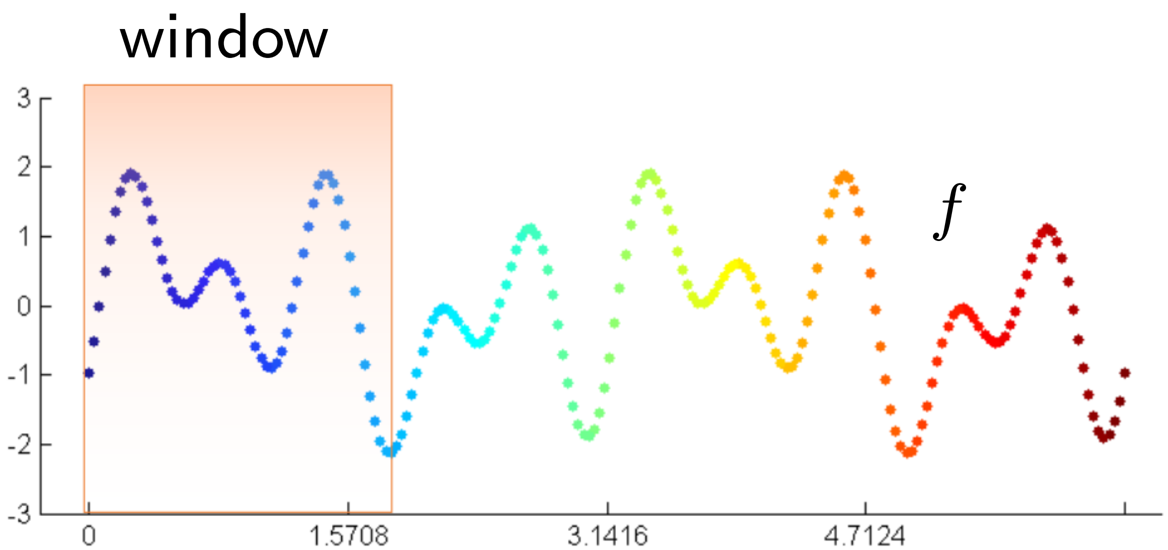
Dynamical system:



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

d

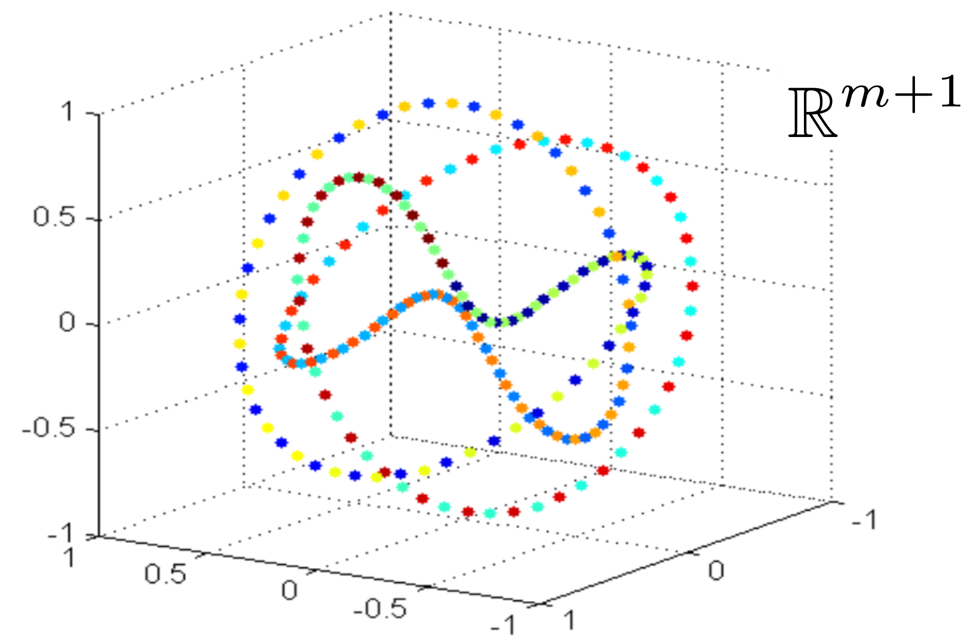
TDA for time series modeling & analysis



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

$\Longrightarrow$

(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]