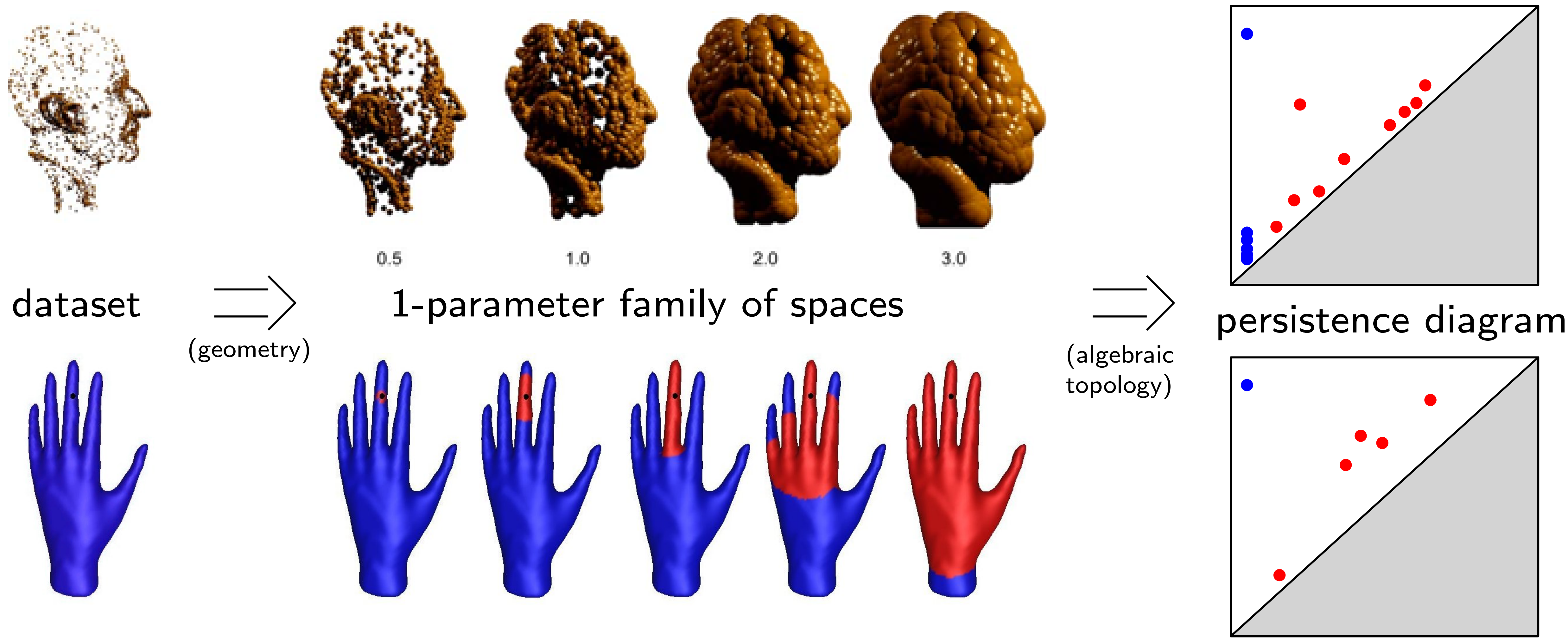
2nd Workshop on Topological Methods in Data Analysis
4th - 6th October 2021, Heidelberg University

Topological Descriptors for Data Science and Machine Learning

Mathieu Carrière
INRIA Sophia-Antipolis
mathieu.carriere@inria.fr

informatics mathematics

Persistence diagrams as descriptors for data



Pros:

- strong invariance and stability:

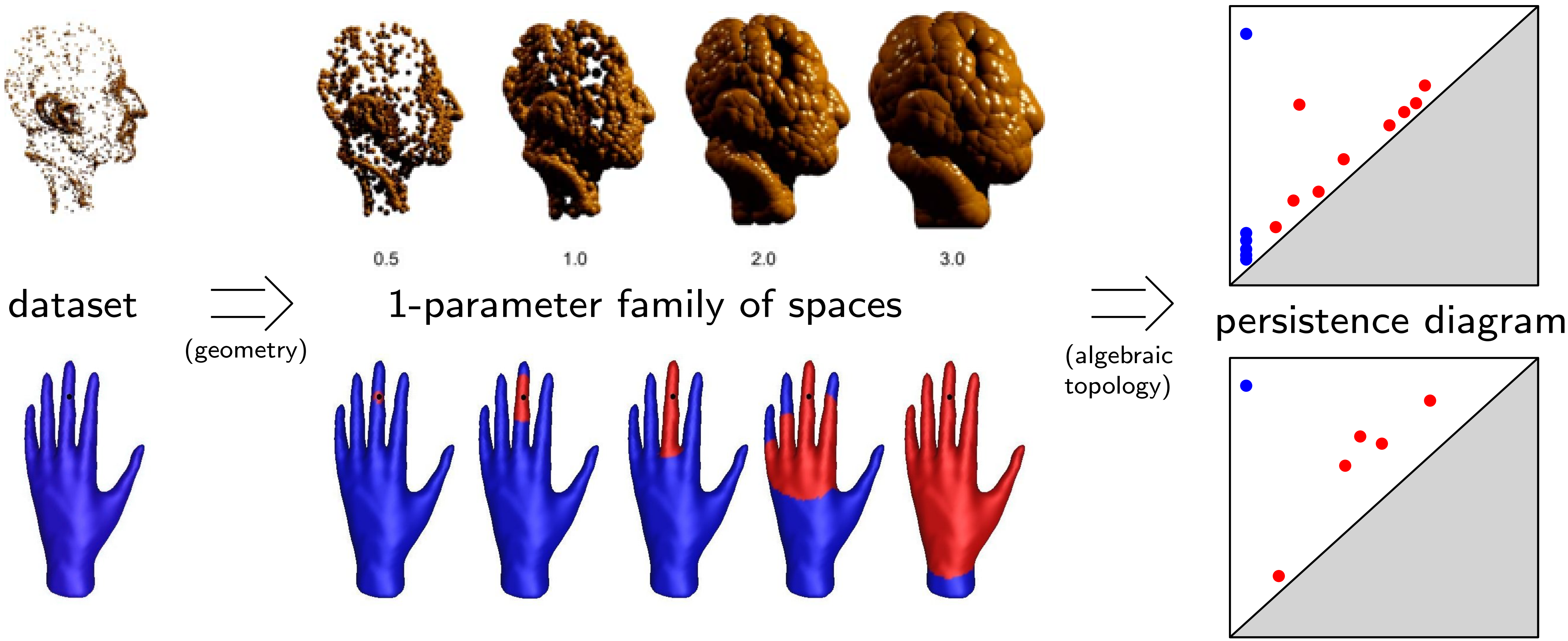
$$d_b(\mathrm{dgm}(R(X)), \mathrm{dgm}(R(Y))) \leq d_{GH}(X, Y)$$

- information of a different nature
- flexible and versatile

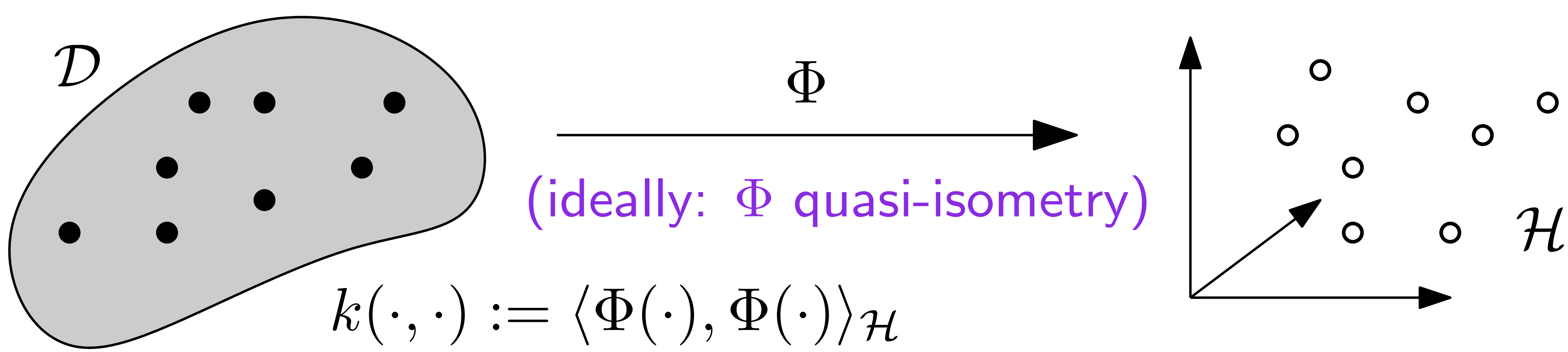
Cons:

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

Persistence diagrams as descriptors for data



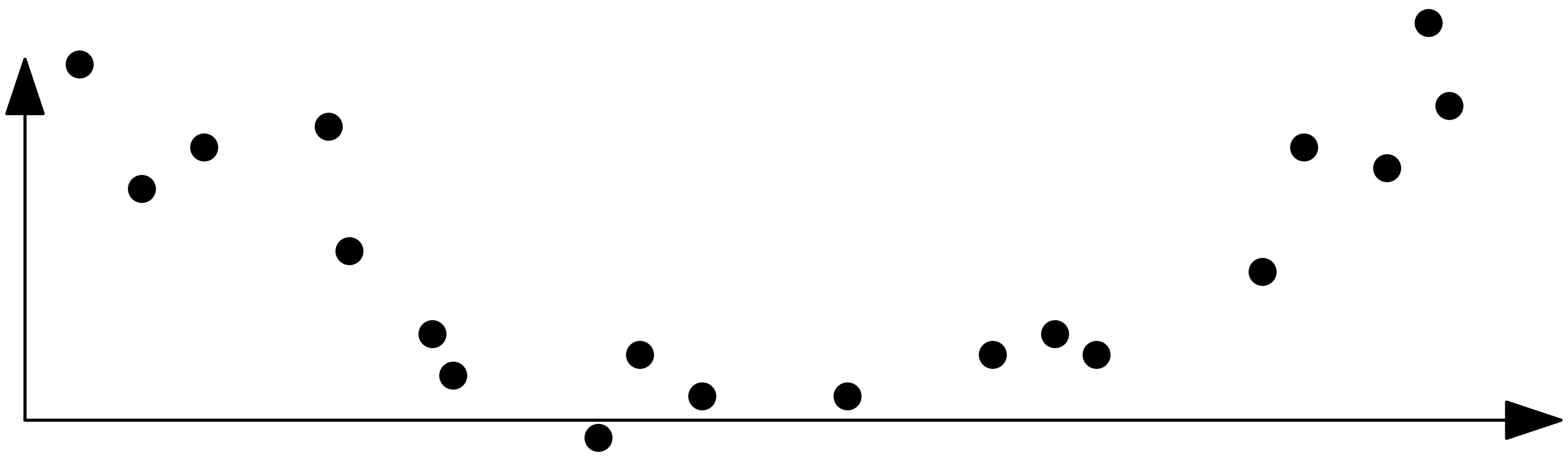
A solution: map diagrams to Hilbert space and use kernel trick



Supervised Machine Learning

Input: n observations + responses $(x_1, y_1), \ldots, (x_n, y_n) \in X \times Y$

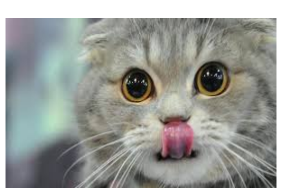
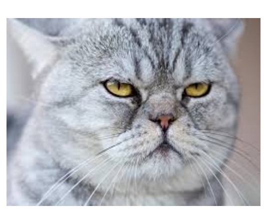
$Y = \mathbb{R}$

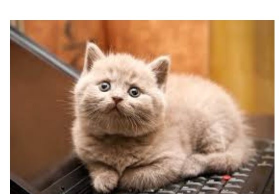


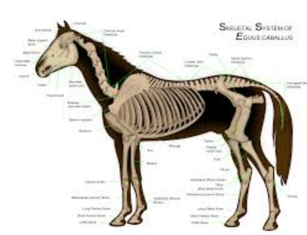
$X = \mathbb{R}$

regression

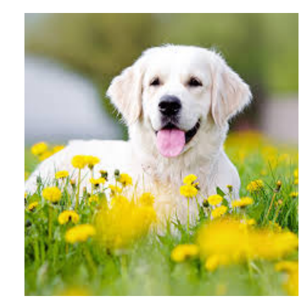
classification

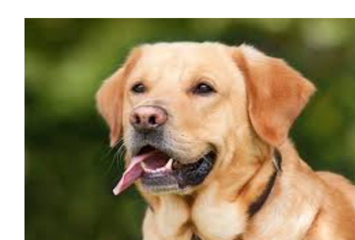
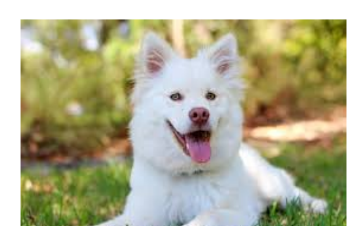










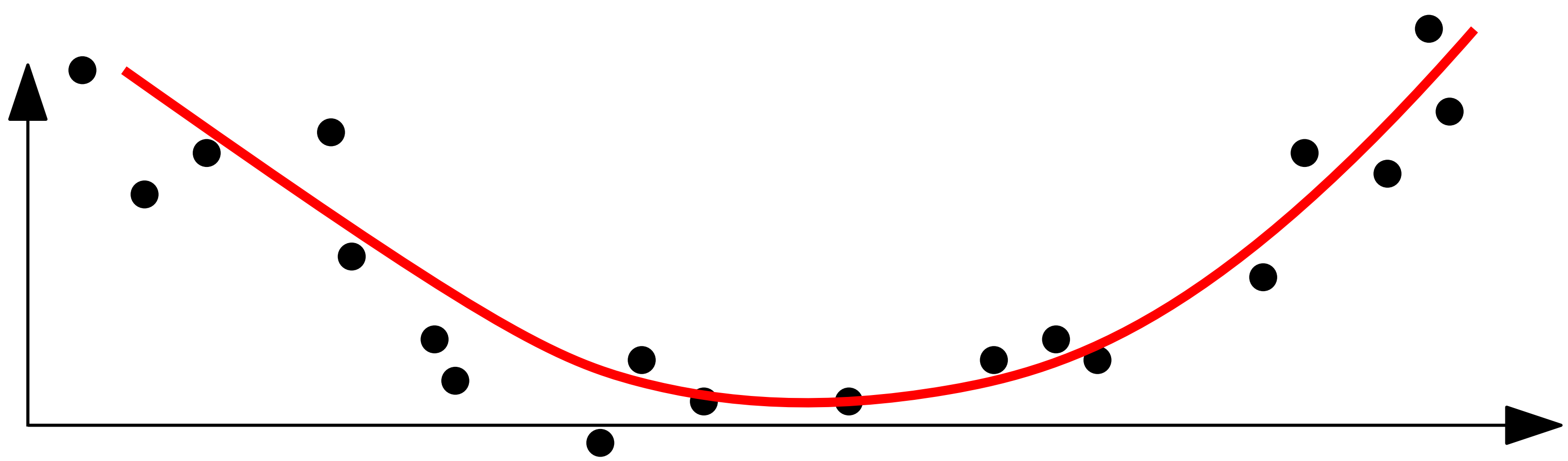



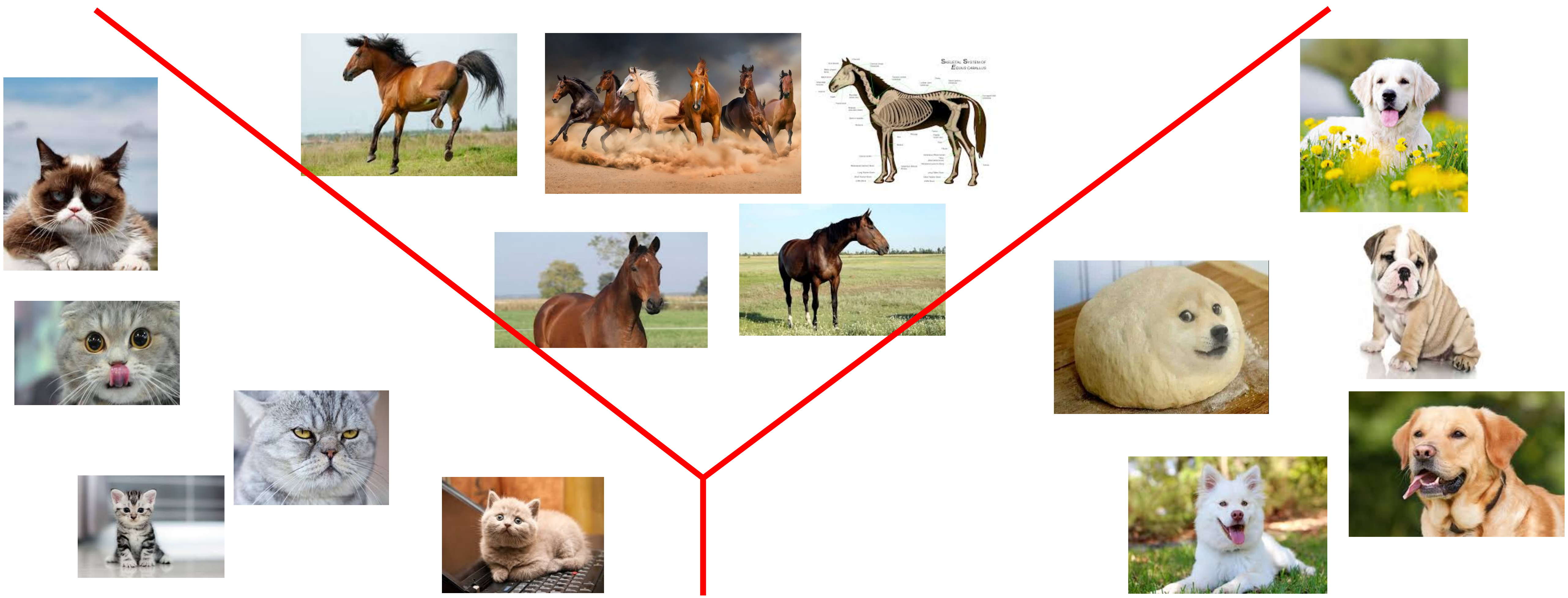
$X =$ images,
$Y = \{$cat, dog, horse$\}$

Supervised Machine Learning

Input: n observations + responses $(x_1, y_1), \ldots, (x_n, y_n) \in X \times Y$

Goal: build a predictor $f : X \to Y$ from $(x_1, y_1), \ldots, (x_n, y_n)$





Empirical Risk Minimization

Optimization problem (supervised regression / classification):

$$f^* = \underset{f \in \mathcal{F}}{\operatorname{argmin}} \; \frac{1}{n} \sum_{i=1}^{n} L(y_i, f(x_i)) + \Omega(f)$$

$\mathcal{F}$ is the **class of predictors**

$L : X \times X \to \mathbb{R}$ is the **loss function**

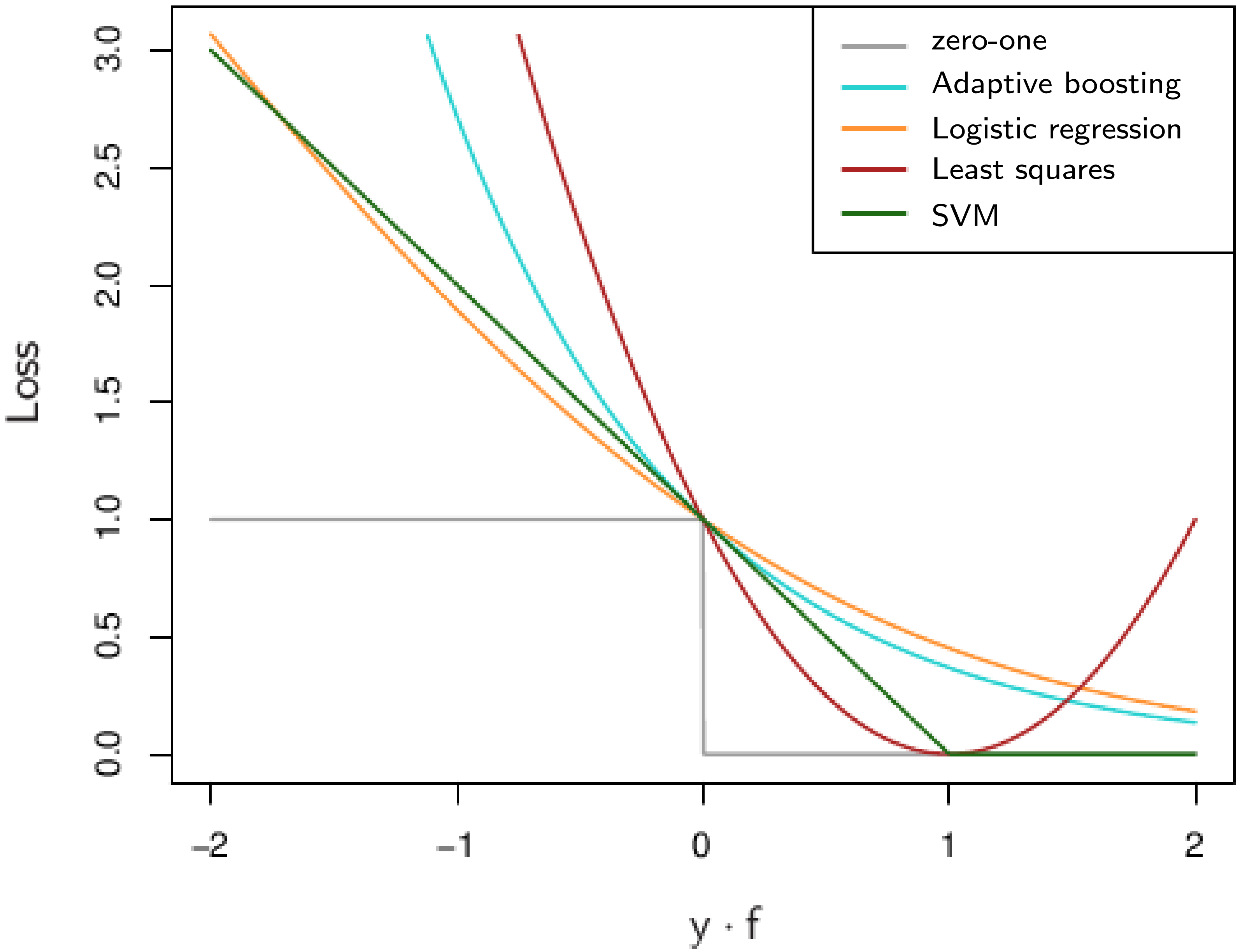
$\Omega : \mathcal{F} \to \mathbb{R}$ is the **regularizer**

$L(y_i, f(x_i))$	Name	
$\mathbb{1}_{y_i \neq f(x_i)}$	zero-one	$\to$ Bayes
$\max\{0, 1 - y_i f(x_i)\}$	hinge	$\to$ Support Vector Machines
$\exp(-y_i f(x_i))$	exponential	$\to$ Adaptive boosting
$\log(1 + \exp(-y_i f(x_i)))$	logistic	$\to$ Logistic regression
$(y_i - f(x_i))^2$	squared	$\to$ Least squares

Empirical Risk Minimization

Optimization problem (supervised regression / classification):

$$f^* = \operatorname*{argmin}_{f \in \mathcal{F}} \ \frac{1}{n} \sum_{i=1}^{n} L(y_i, f(x_i)) + \Omega(f)$$



Empirical Risk Minimization

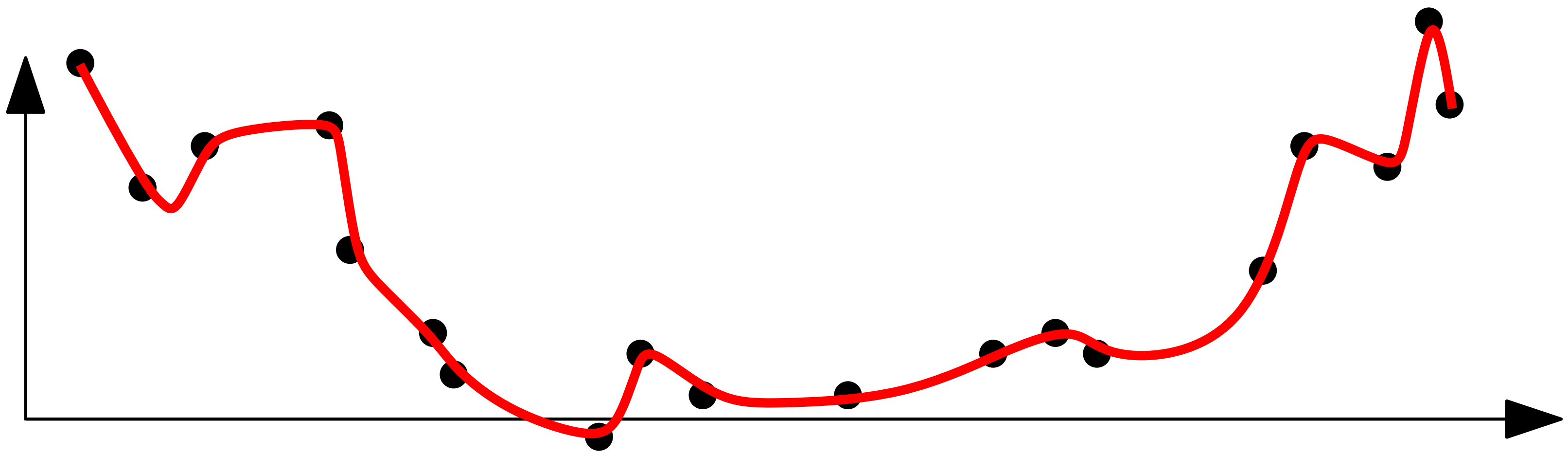
Optimization problem (supervised regression / classification):

$$f^* = \operatorname*{argmin}_{f \in \mathcal{F}} \; \frac{1}{n} \sum_{i=1}^{n} L(y_i, f(x_i)) + \Omega(f)$$

$\mathcal{F}$ is the **class of predictors**

$L : X \times X \to \mathbb{R}$ is the **loss function**

$\Omega : \mathcal{F} \to \mathbb{R}$ is the **regularizer**



$\to$ use regularizer to avoid overfitting

Empirical Risk Minimization

Optimization problem (supervised regression / classification):

$$f^* = \underset{f \in \mathcal{F}}{\operatorname{argmin}} \ \frac{1}{n} \sum_{i=1}^{n} L(y_i, f(x_i)) + \Omega(f)$$

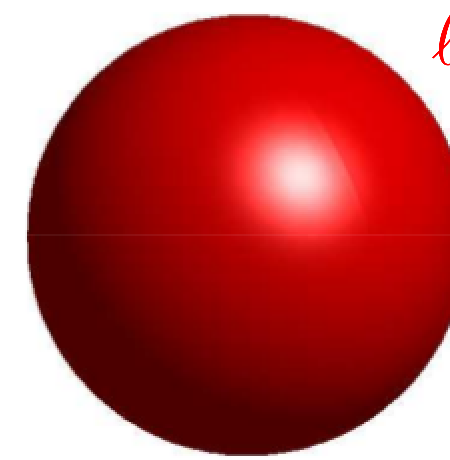
$\mathcal{F}$ is the **class of predictors**

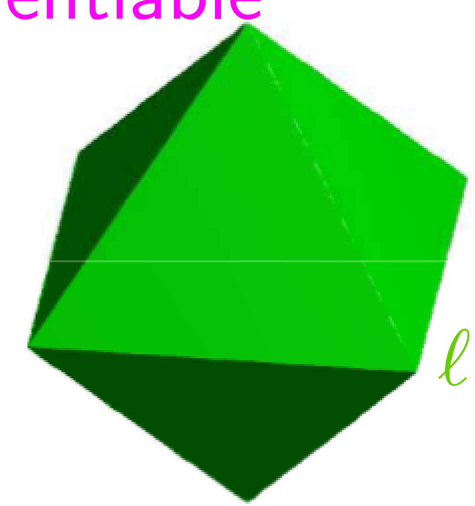
$L : X \times X \to \mathbb{R}$ is the **loss function**

$\Omega : \mathcal{F} \to \mathbb{R}$ is the **regularizer**

$$\mathcal{F} = \{f_w : w \in \mathbb{R}^d\}$$

$\Omega(w)$	Name
$\|w\|_2^2$	ℓ_2 (Tikhonov) $\to$ differentiable
$\|w\|_1$	ℓ_1 (LASSO) $\to$ sparse
$\alpha\|w\|_2^2 + (1-\alpha)\|w\|_1$	elastic net


ℓ_2


ℓ_1

Empirical Risk Minimization

Optimization problem (supervised regression / classification):

$$f^* = \operatorname*{argmin}_{f \in \mathcal{F}} \; \frac{1}{n} \sum_{i=1}^{n} L(y_i, f(x_i)) + \Omega(f)$$

$\mathcal{F}$ is the **class of predictors**

$L : X \times X \to \mathbb{R}$ is the **loss function**

$\Omega : \mathcal{F} \to \mathbb{R}$ is the **regularizer**

Complexity of the minimization grows with the one of $\mathcal{F}$

Easy to control when $\mathcal{F}$ is a Reproducing Kernel Hilbert Space

Reproducing Kernel Hilbert Space

Def: Let $\mathcal{H} \subset \mathbb{R}^X$ Hilbert, with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$
Then, $\mathcal{H}$ is a **RKHS** on X if $\exists \Phi : X \to \mathcal{H}$ s.t.:

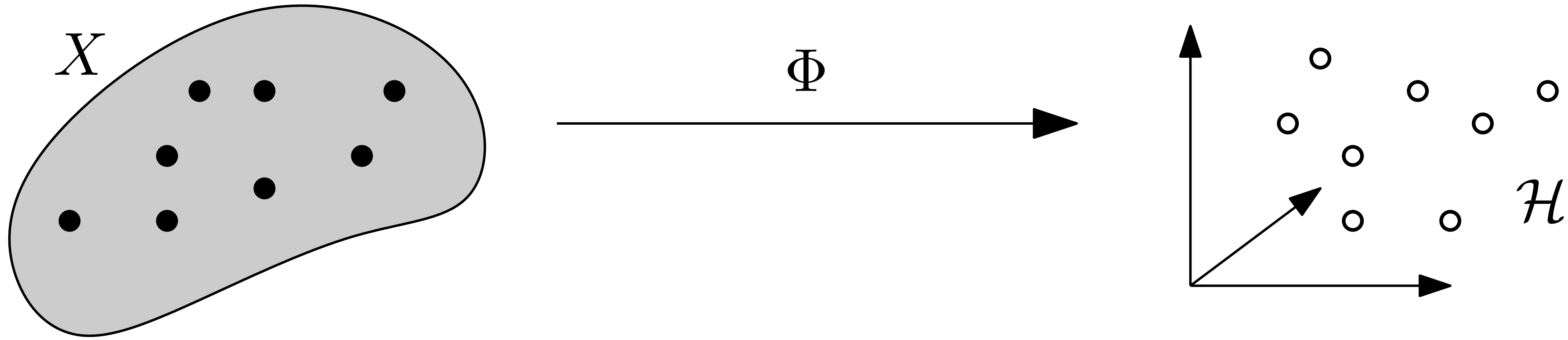
$$\forall x \in X, \forall f \in \mathcal{H}, \; f(x) = \langle f, \Phi(x) \rangle_{\mathcal{H}}$$

reproducing property

Terminology:

- **feature space** $\mathcal{H}$, **feature map** Φ
- **feature vector** $\Phi(x)$
- **kernel** $k = \langle \Phi(\cdot), \Phi(\cdot) \rangle_{\mathcal{H}} : X \times X \to \mathbb{R}$

Case X Hilbert space:
$\mathcal{H} = X^*$, $\Phi(x) = \langle x, \cdot \rangle_X$
Φ isometric isomorphism [Riesz]
$\langle \cdot, \cdot \rangle_{\mathcal{H}} := \langle \Phi^{-1}(\cdot), \Phi^{-1}(\cdot) \rangle_X$



Reproducing Kernel Hilbert Space

[*Theory of Reproducing Kernels*, Aronszajn, Trans. Amer. Math. Soc., 1950]

Def: Let $\mathcal{H} \subset \mathbb{R}^X$ Hilbert, with inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}}$
Then, $\mathcal{H}$ is a **RKHS** on X if $\exists \Phi : X \to \mathcal{H}$ s.t.:

$$\forall x \in X,\ \forall f \in \mathcal{H},\ f(x) = \langle f, \Phi(x)\rangle_{\mathcal{H}}$$

reproducing property

Prop: Given X, the kernel of a RKHS on X is unique.
Conversely, k is the kernel of at most one RKHS on X.

Thm: The function $k : X \times X \to \mathbb{R}$ is a kernel iff it is *positive (semi-)definite*, i.e. $\forall n \in \mathbb{N}$, $\forall x_1, \ldots, x_n \in X$, the Gram matrix $(k(x_i, x_j))_{i,j}$ is positive semi-definite.

Examples in $X = (\mathbb{R}^d, \langle\cdot,\cdot\rangle)$:

- linear: $k(x,y) = \langle x, y\rangle$ $\quad \mathcal{H} = (\mathbb{R}^d)^*,\ \Phi(x) = \langle x, \cdot\rangle$
- polynomial: $k(x,y) = (1 + \langle x,y\rangle)^N = \sum_{n_1+\cdots+n_d=N} \binom{N}{n_1,\ldots,n_d} \underbrace{x_1^{n_1}\ldots x_d^{n_d}}_{\propto \Phi(x)} y_1^{n_1}\ldots y_d^{n_d}$
- Gaussian: $k(x,y) = \exp\left(-\frac{\|x-y\|_2^2}{2\sigma^2}\right)$, $\sigma > 0$. $\quad \mathcal{H} \subseteq L_2(\mathbb{R}^d)$

Reproducing Kernel Hilbert Space

[*A correspondence between Bayesian estimation on stochastic processes and smoothing by splines*, Kimeldorf, Wahba, The Annals Math. Stat., 1970]

Def: Let $\mathcal{H} \subset \mathbb{R}^X$ Hilbert, with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$
Then, $\mathcal{H}$ is a **RKHS** on X if $\exists \Phi : X \to \mathcal{H}$ s.t.:

$$\forall x \in X, \forall f \in \mathcal{H}, \; f(x) = \langle f, \Phi(x) \rangle_{\mathcal{H}}$$

reproducing property

Thm: (Representer)
Given RKHS $\mathcal{H}$ with kernel k, any function $f^* \in \mathcal{H}$ minimizing

$$\tfrac{1}{n} \sum_{i=1}^n L(y_i, f(x_i)) + \Omega(\|f\|_{\mathcal{H}})$$

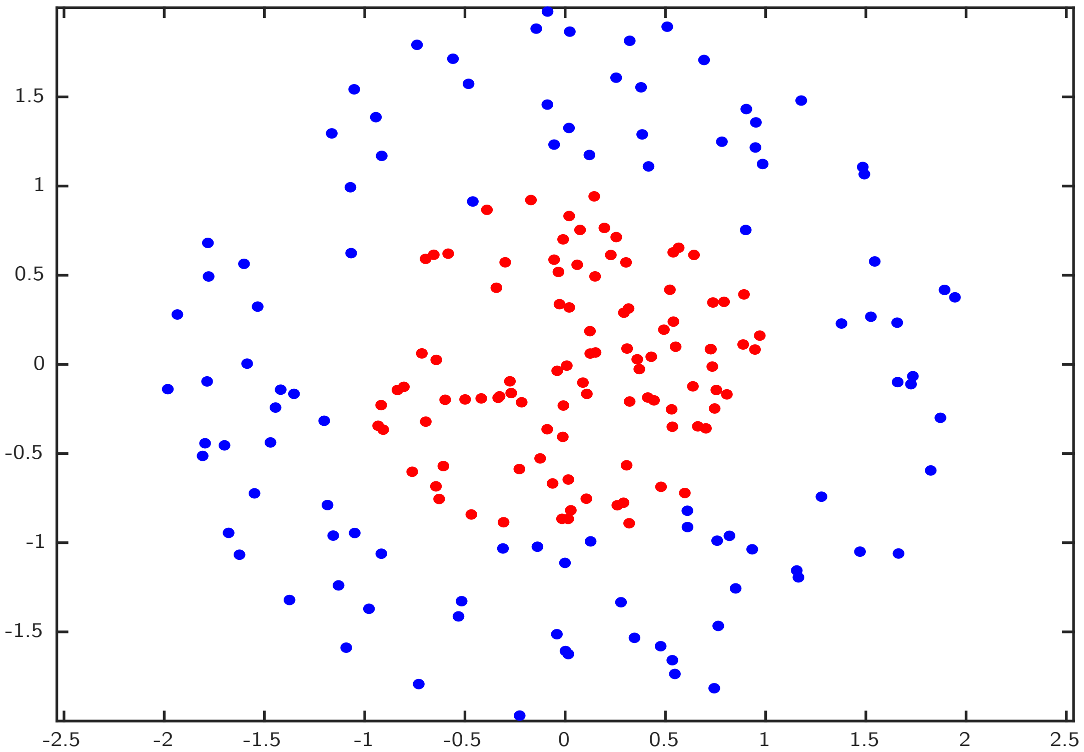
is of the form $f^*(\cdot) = \sum_{j=1}^n \alpha_j k(x_j, \cdot)$, where $\alpha_1, \dots, \alpha_n \in \mathbb{R}$.

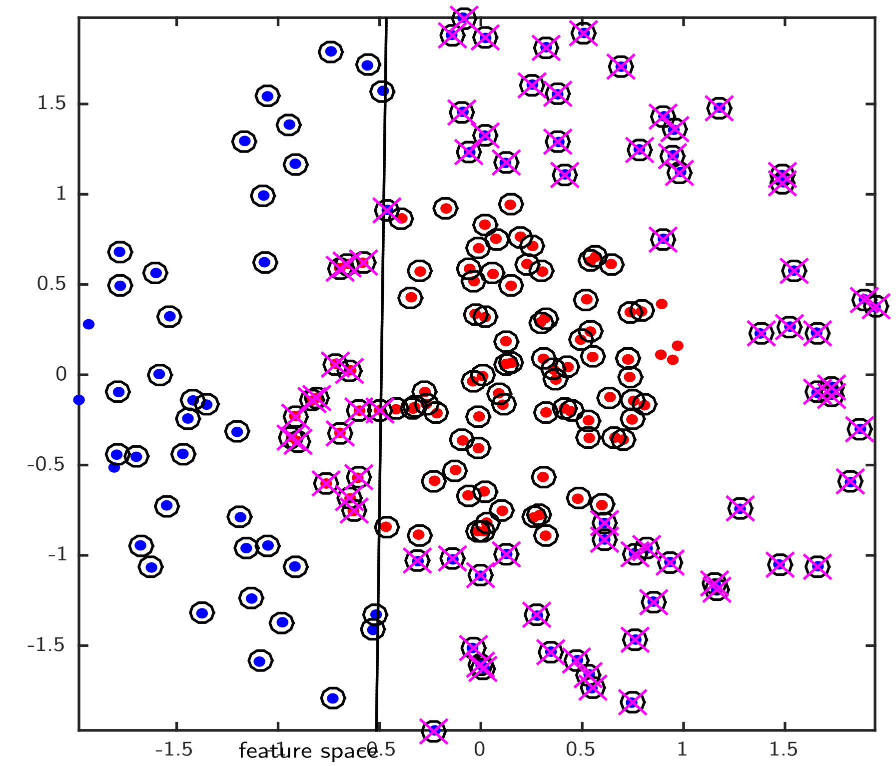
$$\rightsquigarrow \quad \operatorname*{argmin}_{\alpha} \frac{1}{n} \sum_{i=1}^n L\left(y_i, \sum_{j=1}^n \alpha_j k(x_j, x_i) \right) + \Omega\left(\sum_{i,j=1}^n \alpha_i \alpha_j k(x_i, x_j) \right)$$

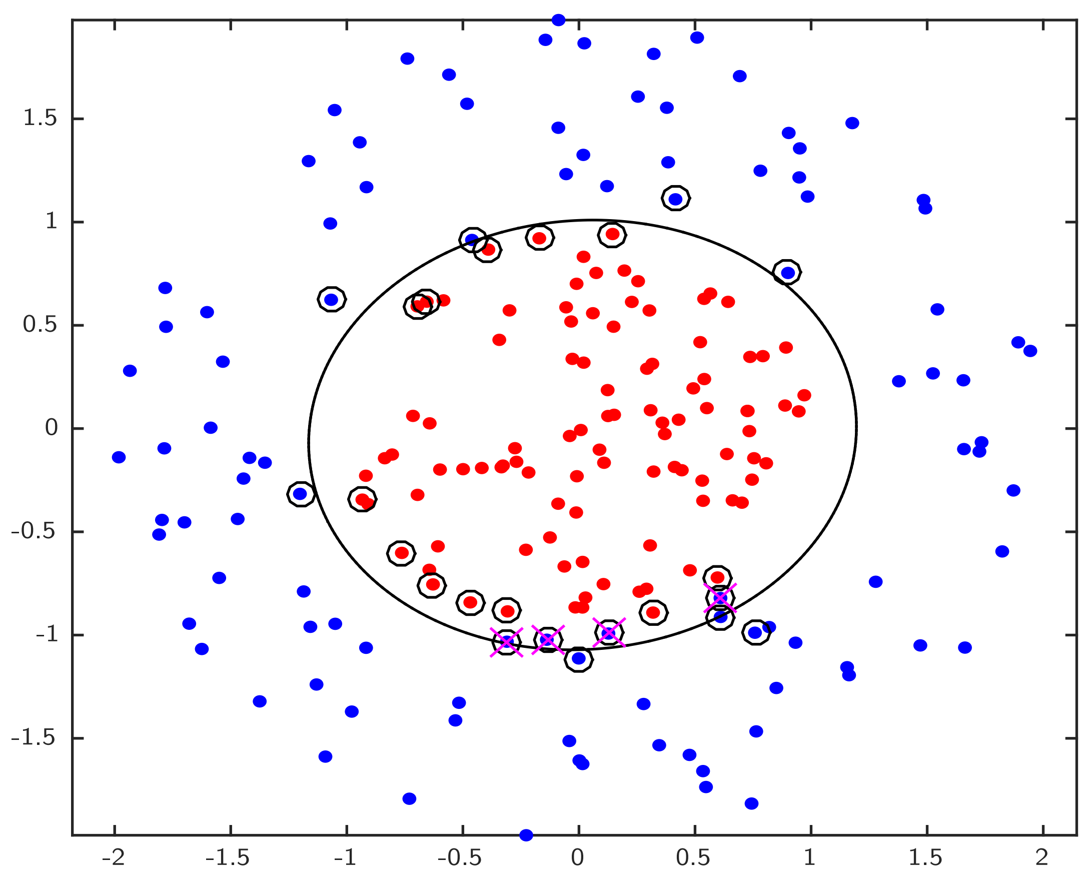
where $\alpha = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}$ and $K = (k(x_i, x_j))_{ij}$

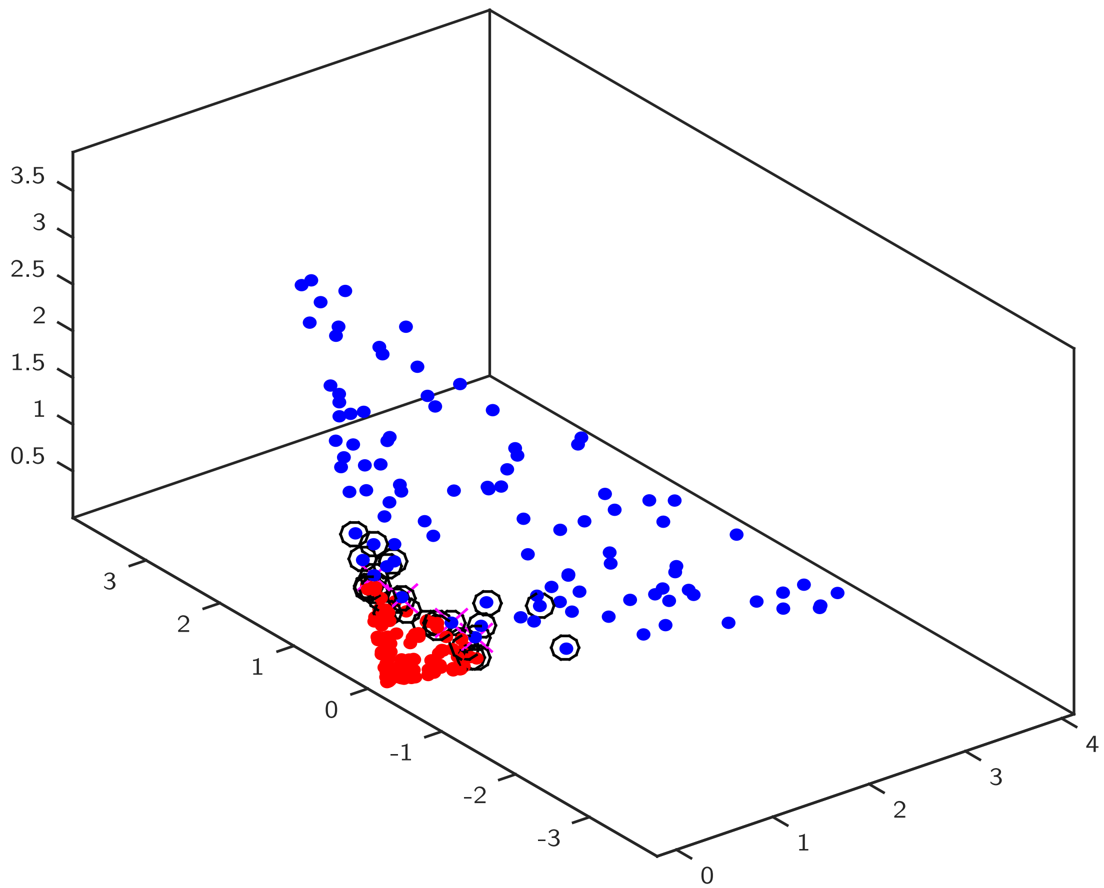
only the $k(x_i, x_j)$ are required to minimize (**kernel trick**)

Kernel Trick









Kernels for persistence diagrams

Three approaches:

- build kernel from kernels (algebraic operations)

- **sum of kernels** $\longleftrightarrow$ **concatenation of feature spaces**

$$k_1(x,y) + k_2(x,y) = \left\langle \begin{pmatrix} \Phi_1(x) \\ \Phi_2(x) \end{pmatrix}, \begin{pmatrix} \Phi_1(y) \\ \Phi_2(y) \end{pmatrix} \right\rangle$$

- **product of kernels** $\longleftrightarrow$ **tensor product of feature spaces**

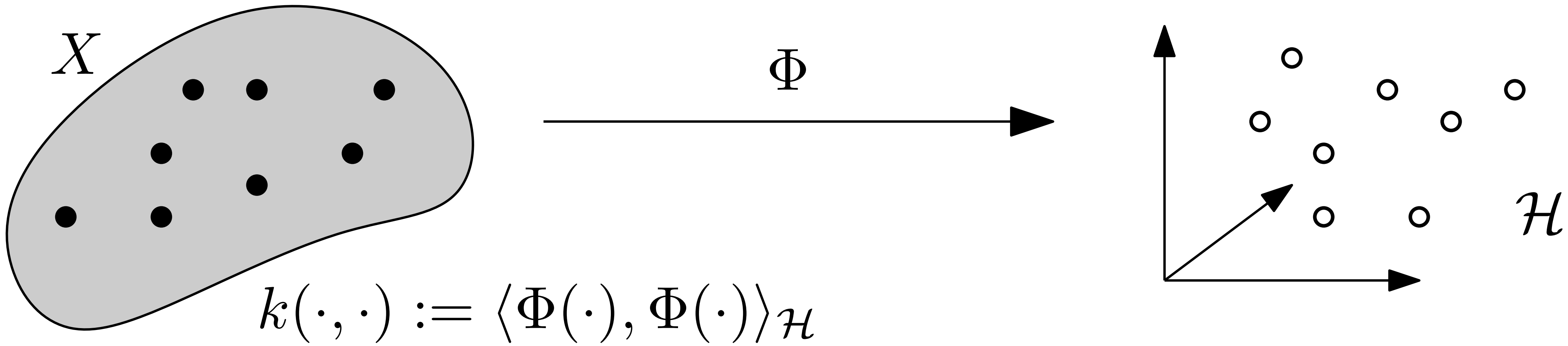
$$k_1(x,y)k_2(x,y) = \left\langle \Phi_1(x)\Phi_2(x)^T, \Phi_1(y)\Phi_2(y)^T \right\rangle$$

Q: prove it.

Kernels for persistence diagrams

Three approaches:

- build kernel from kernels (algebraic operations)
- define explicit feature map $\Phi : X \to \mathcal{H}$ (vectorization)



Kernels for persistence diagrams

[*A correspondence between Bayesian estimation on stochastic processes and smoothing by splines*, Kimeldorf, Wahba, The Annals Math. Stat., 1970]

Three approaches:

- build kernel from kernels (algebraic operations)
- define explicit feature map $\Phi : X \to \mathcal{H}$ (vectorization)
- define kernel from metric via radial basis function

Thm:

If $d : X \times X \to \mathbb{R}_+$ symmetric is *conditionally negative semidefinite*, i.e.:

$$\forall n \in \mathbb{N},\ \forall x_1, \ldots, x_n \in X,\ \sum_{i=1}^{n} \alpha_i = 0 \Longrightarrow \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \, d(x_i, x_j) \leq 0,$$

then $k(x, y) = \exp\left(-\frac{d(x,y)}{2\sigma^2}\right)$ is positive definite for all $\sigma > 0$.

Q: does this apply to persistence diagrams?

Space of persistence diagrams

Persistence diagram $\equiv$ **finite** multiset in the open half-plane $\Delta \times \mathbb{R}_{>0}$

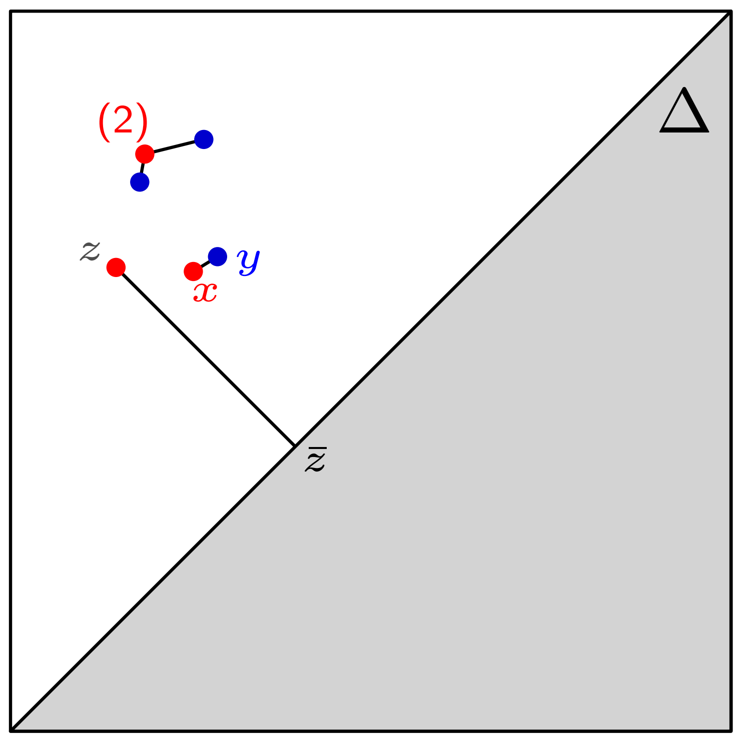
Given a **partial matching** $M : X \leftrightarrow Y$:

cost of a matched pair $(x, y) \in M$: $c_p(x, y) := \|x - y\|_\infty^p$

cost of an unmatched point $z \in X \sqcup Y$: $c_p(z) := \|z - \bar{z}\|_\infty^p$

cost of M:

$$c_p(M) := \left(\sum_{(x,\,y)\text{ matched}} c_p(x, y) \; + \sum_{z\text{ unmatched}} c_p(z) \right)^{1/p}$$



Def: p-th diagram distance (extended metric):

$$d_p(X, Y) := \inf_{M : X \leftrightarrow Y} c_p(M)$$

Def: bottleneck distance:

$$d_b(X, Y) := \lim_{p \to \infty} d_p(X, Y)$$

Space of persistence diagrams

Persistence diagram $\equiv$ **finite** multiset in the open half-plane $\Delta \times \mathbb{R}_{>0}$

Given a **partial matching** $M : X \leftrightarrow Y$:

cost of a matched pair $(x, y) \in M$: $c_p(x, y)$

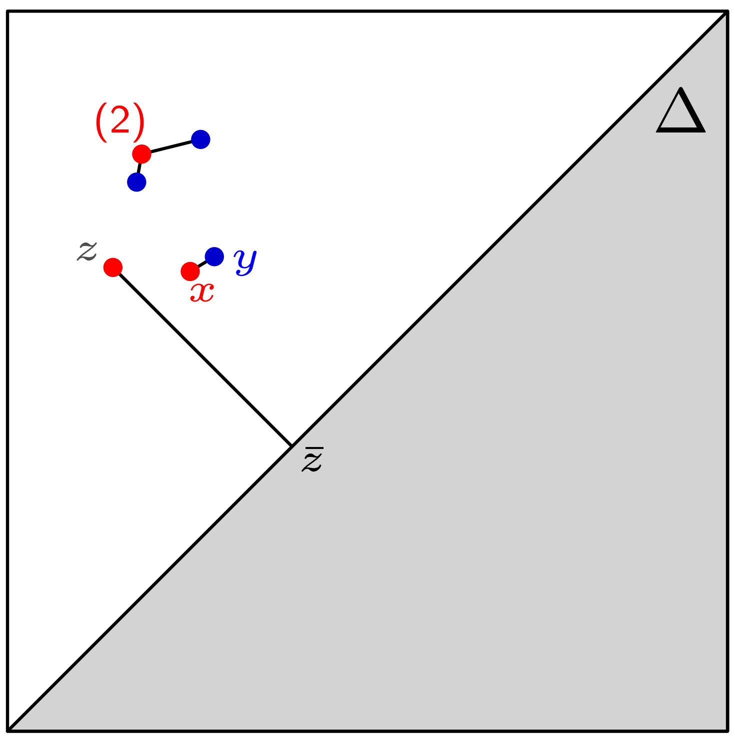
d_p is **NOT** cnsd

$\Rightarrow$ previous theorem is not applicable

cost of an unmatched point $z \in X \sqcup Y$: $c_p(z) := \|z - \bar{z}\|_\infty^p$

cost of M:

$$c_p(M) := \left(\sum_{(x,\,y)\text{ matched}} c_p(x, y) \; + \sum_{z\text{ unmatched}} c_p(z) \right)^{1/p}$$



Def: p-th diagram distance (extended metric):

$$d_p(X, Y) := \inf_{M : X \leftrightarrow Y} c_p(M)$$

Def: bottleneck distance:

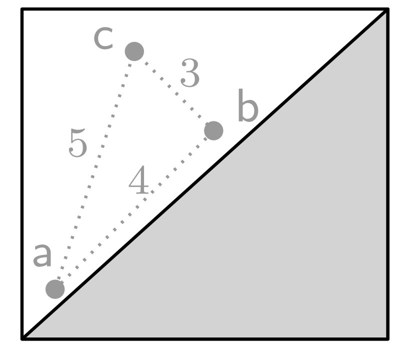
$$d_b(X, Y) := \lim_{p \to \infty} d_p(X, Y)$$

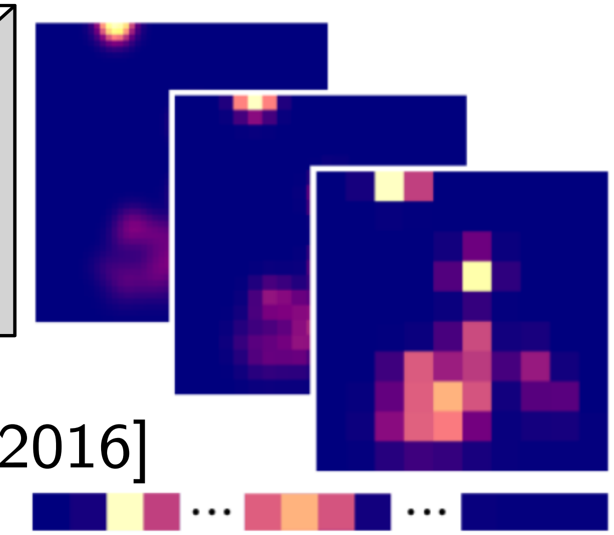
Kernels for persistence diagrams

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

- images [Adams et al. 2015]

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

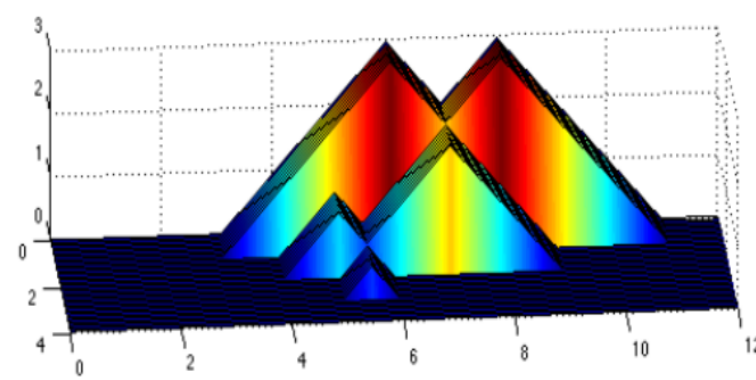




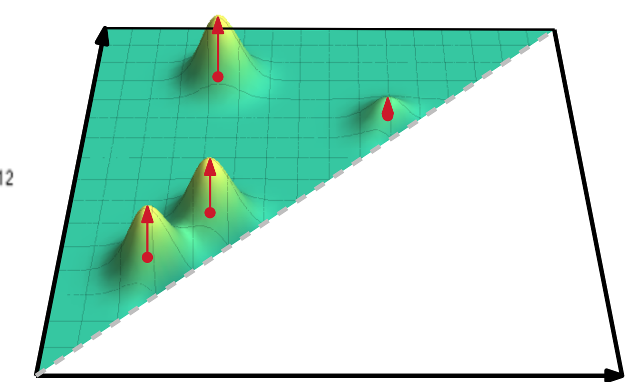
- finite metric spaces [Carrière et al. 2015]
- polynomial roots or evaluations [Di Fabio, Ferri 2015] [Kališnik 2016]

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

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



- discrete measures:
  - → histogram [Bendich et al. 2014]
  - → convolution with fixed kernel [Chepushtanova et al. 2015]
  - → 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	$(\mathbb{R}^d, \|.\|_2)$	$(\mathbb{R}^d, \|.\|_2)$	$\ell_2(\mathbb{R})$	$L_2(\mathbb{N} \times \mathbb{R})$	$L_2(\mathbb{R}^2)$
positive (semi-)definiteness	✓	✓	✓	✓	✓
$\|\phi(\cdot) - \phi(\cdot)\|_{\mathcal{H}} \leq \phi(d_p)$	✓	✓	✓	✓	✓
$\|\phi(\cdot) - \phi(\cdot)\|_{\mathcal{H}} \geq \psi(d_p)$	✗	✗	✗	✗	✗
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

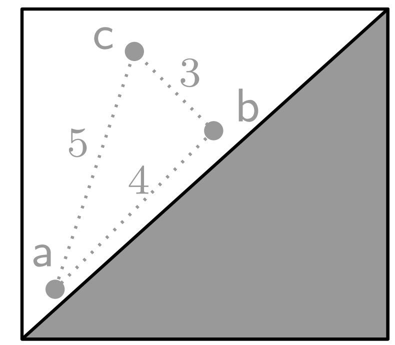
	images	metric spaces	polynomials	landscapes	discrete measures
ambient Hilbert space	$(\mathbb{R}^d, \|.\|_2)$	$(\mathbb{R}^d, \|.\|_2)$	$\ell_2(\mathbb{R})$	$L_2(\mathbb{N} \times \mathbb{R})$	$L_2(\mathbb{R}^2)$
positive (semi-)definiteness	✓	✓	✓	✓	✓
$\|\phi(\cdot) - \phi(\cdot)\|_{\mathcal{H}} \leq \phi(d_p)$	✓	✓	✓	✓	✓
$\boxed{\|\phi(\cdot) - \phi(\cdot)\|_{\mathcal{H}} \geq \psi(d_p)}$	✗	✗	✗	✗	✗
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: $O(d)$	$O(n^2)$	$O(n^2)$

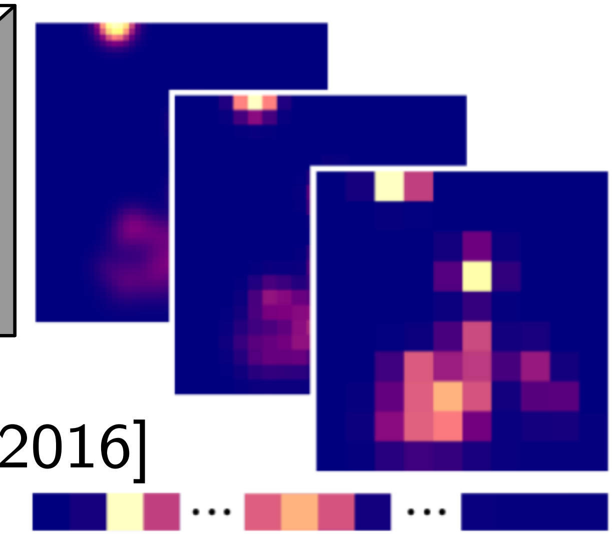
Kernels for persistence diagrams

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

- $\boxed{\text{images [Adams et al. 2015]}}$

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

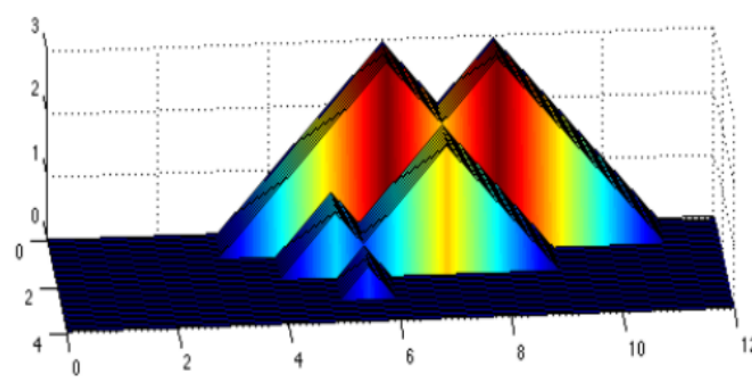




- finite metric spaces [Carrière, O., Ovsjanikov 2015]
- polynomial roots or evaluations [Di Fabio, Ferri 2015] [Kališnik 2016]

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

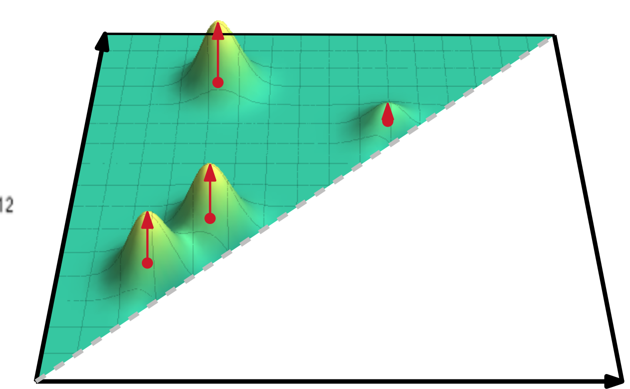
- landscapes [Bubenik 2012] [Bubenik, Dłotko 2015]



- discrete measures:

  $\rightarrow$ histogram [Bendich et al. 2014]

  $\rightarrow$ convolution with fixed kernel [Chepushtanova et al. 2015]

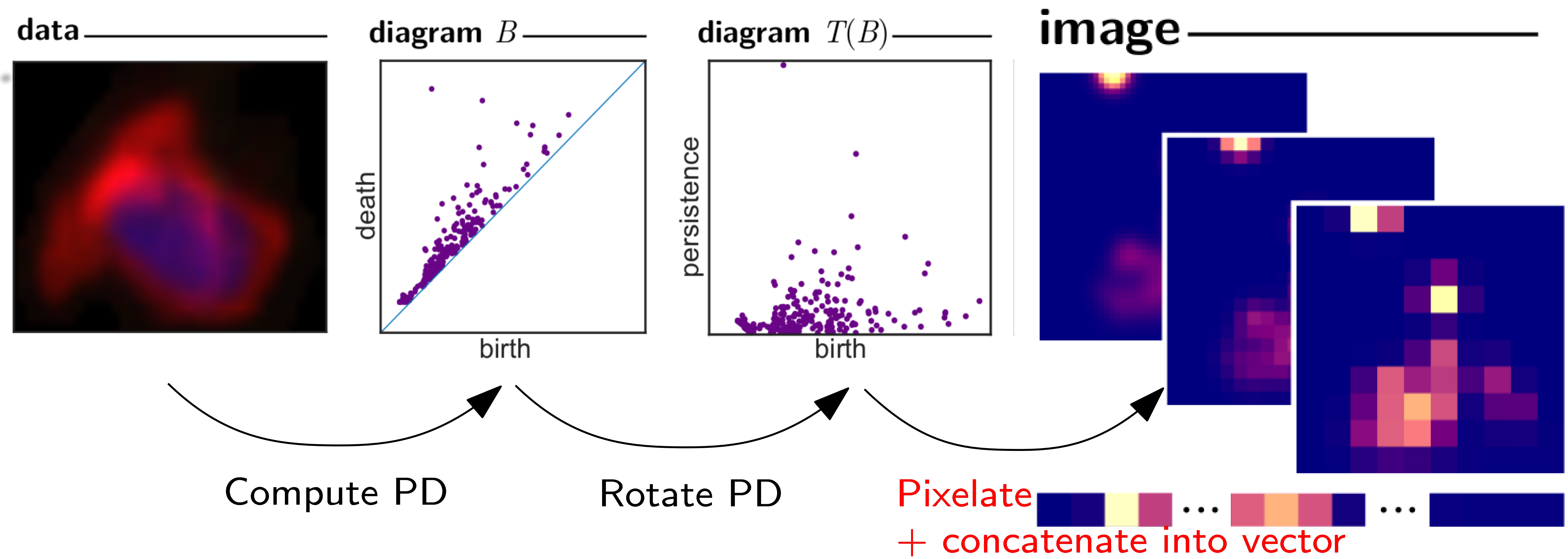


  $\rightarrow$ convolution with weighted kernel [Kusano, Fukumisu, Hiraoka 2016-17]

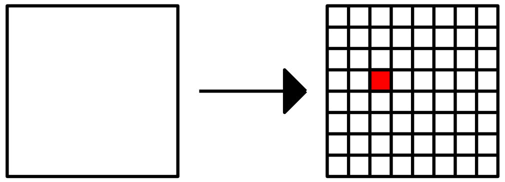
  $\rightarrow$ heat diffusion [Reininghaus et al. 2015] + exponential [Kwit et al. 2015]

Explicit Feature Map in $\mathbb{R}^d$

[*Persistence Images: A Stable Vector Representation of Persistent Homology*, Adams et al., JMLR, 2017]



Discretize plane into one or several grid(s):

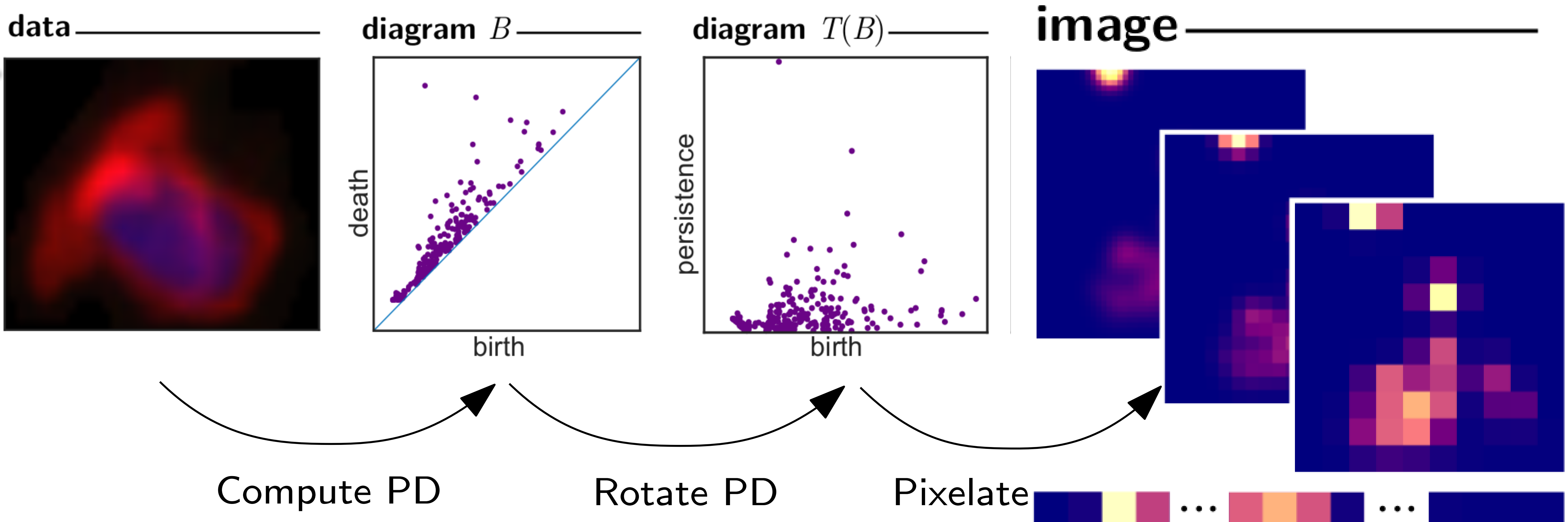


For each pixel P, compute $I(P) = \#\, \mathrm{D} \cap P$

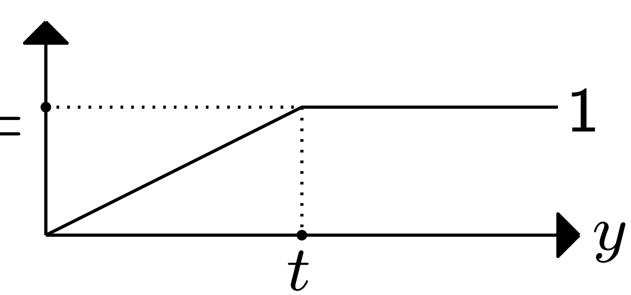
Concatenate all $I(P)$ into a single vector $\mathrm{PI}(\mathrm{D})$

Explicit Feature Map in $\mathbb{R}^d$

[*Persistence Images: A Stable Vector Representation of Persistent Homology*, Adams et al., JMLR, 2017]



Stability $\to$ weight points: $w_t(x, y) =$

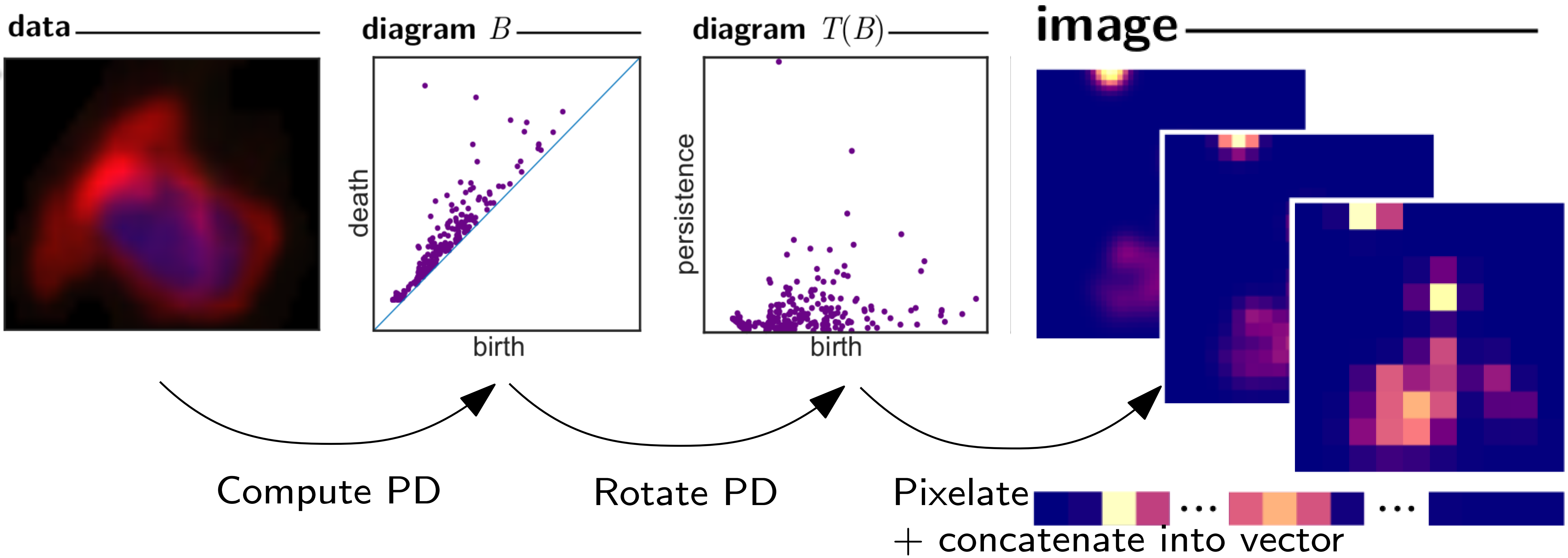


$\to$ blur image

(convolve with Gaussian $\to$ details forthcoming)

Explicit Feature Map in $\mathbb{R}^d$

[*Persistence Images: A Stable Vector Representation of Persistent Homology*, Adams et al., JMLR, 2017]



Prop:

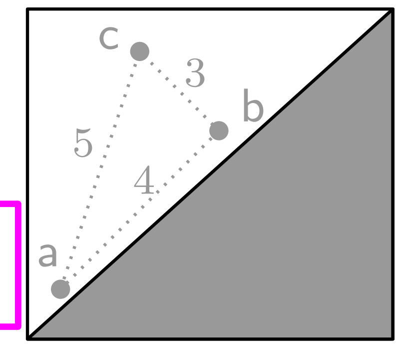
- $\|\mathrm{PI}(\mathrm{D}) - \mathrm{PI}(\mathrm{D}')\|_\infty \leq C(w, \phi_p)\, d_1(\mathrm{D}, \mathrm{D}')$
- $\|\mathrm{PI}(\mathrm{D}) - \mathrm{PI}(\mathrm{D}')\|_2 \leq \sqrt{d}C(w, \phi_p)\, d_1(\mathrm{D}, \mathrm{D}')$

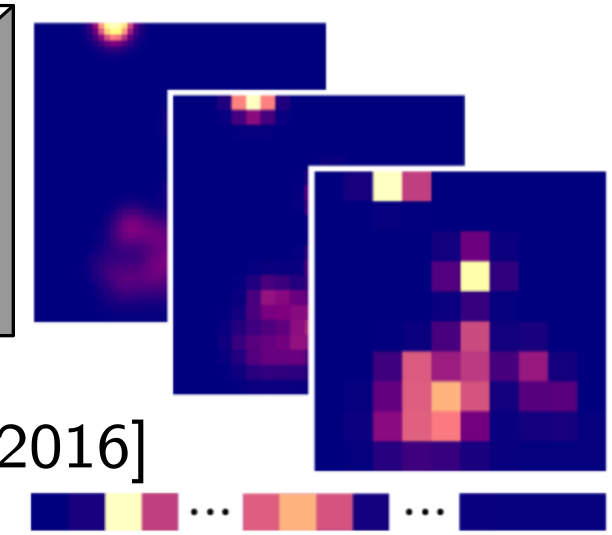
Kernels for persistence diagrams

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

- images [Adams et al. 2015]

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



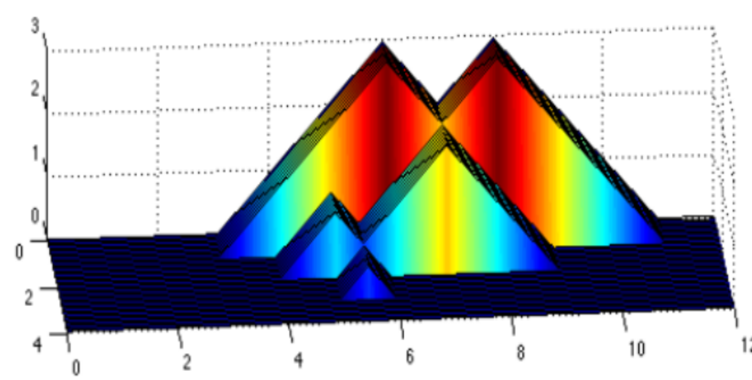


- finite metric spaces [Carrière et al. 2015]

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

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

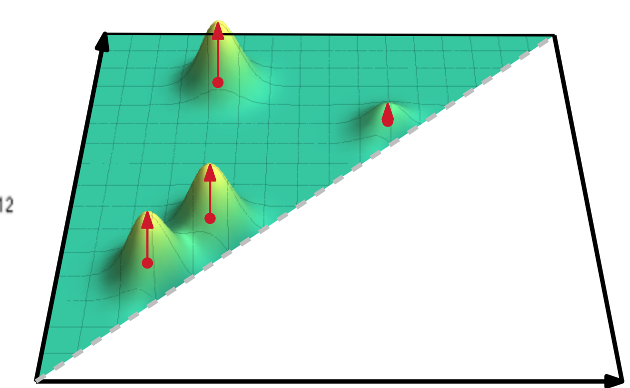
- landscapes [Bubenik 2012] [Bubenik, Dłotko 2015]



- discrete measures:

  → histogram [Bendich et al. 2014]

  → convolution with fixed kernel [Chepushtanova et al. 2015]



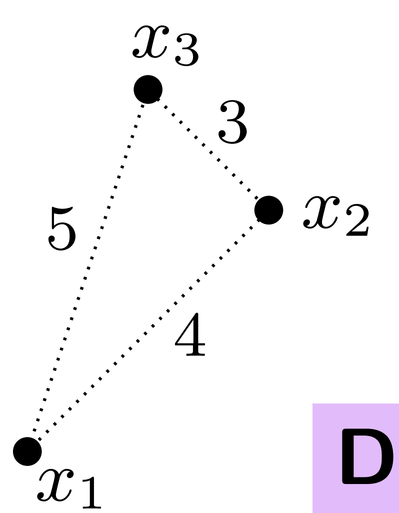
  → convolution with weighted kernel [Kusano, Fukumisu, Hiraoka 2016-17]

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

Explicit Feature Map in $\mathbb{R}^d$

[*Stable topological signatures for points on 3D shapes*, Carrière, Oudot, Ovsjanikov, SGP, 2015]

finite metric space



$\xrightarrow{\phi_1}$

distance matrix

$$\begin{array}{c c} & \begin{array}{ccc} x_1 & x_2 & x_3 \end{array} \\ \begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array} & \begin{bmatrix} 0 & 4 & 5 \\ 4 & 0 & 3 \\ 5 & 3 & 0 \end{bmatrix} \end{array}$$

Def: $\Phi = \Phi_3 \circ \Phi_2 \circ \Phi_1$

$\xrightarrow{\phi_2}$

sorted sequence
with finite support

$$(5, 4, 3, 0, \cdots)$$

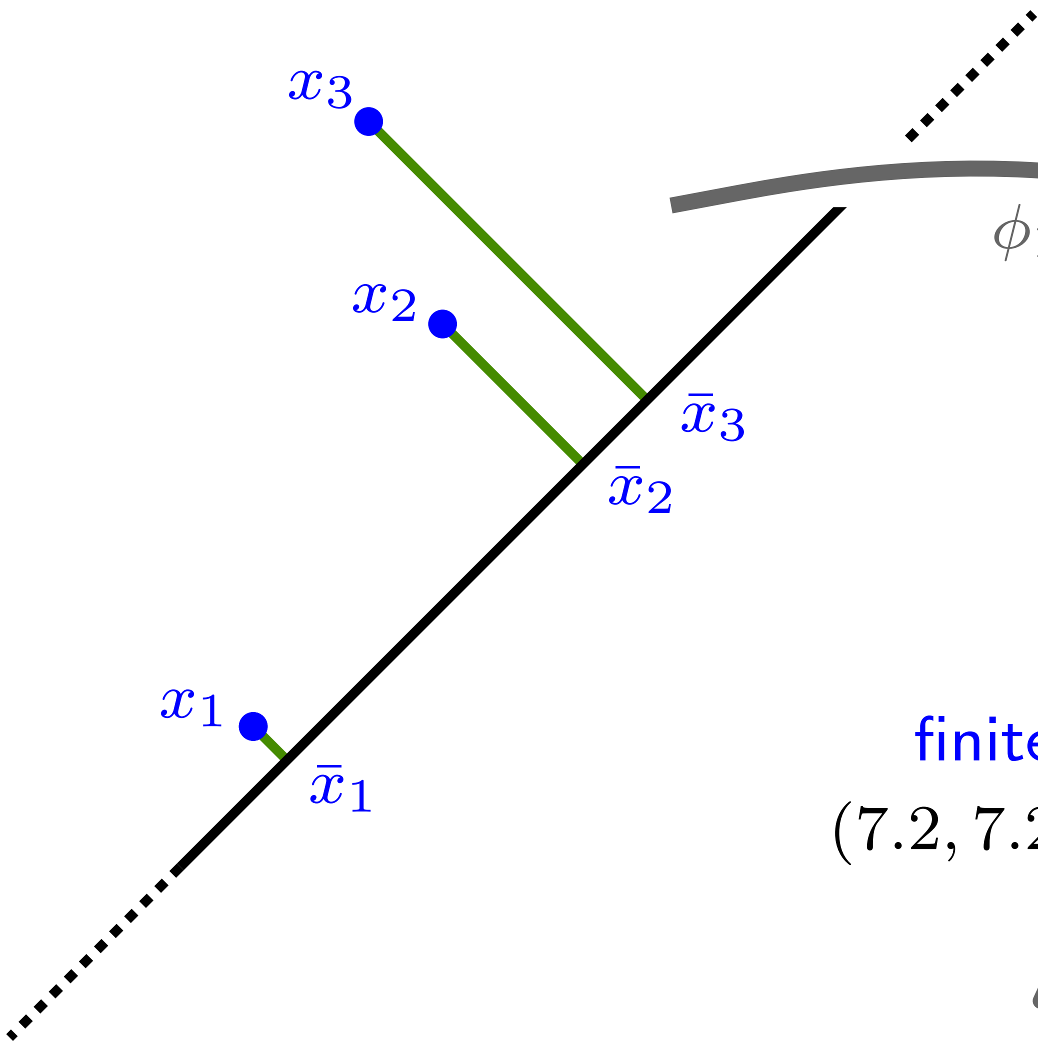
$\xrightarrow{\phi_3}$

$$(5, 4, 3, 0, \cdots, 0)$$

finite-dimensional vector

Explicit Feature Map in $\mathbb{R}^d$

[*Stable topological signatures for points on 3D shapes*, Carrière, Oudot, Ovsjanikov, SGP, 2015]



ϕ_1

distance matrix

$$\left[\begin{array}{ll} D_{ij} = \min\{ & \|x_i - x_j\|_\infty, \\ & \|x_i - \bar{x}_i\|_\infty, \\ & \|x_j - \bar{x}_j\|_\infty\} \end{array}\right]$$

ϕ_2

$$(7.2, 7.2, 5.6, 4, 4, 4, 0, \cdots)$$

sorted sequence
with finite support

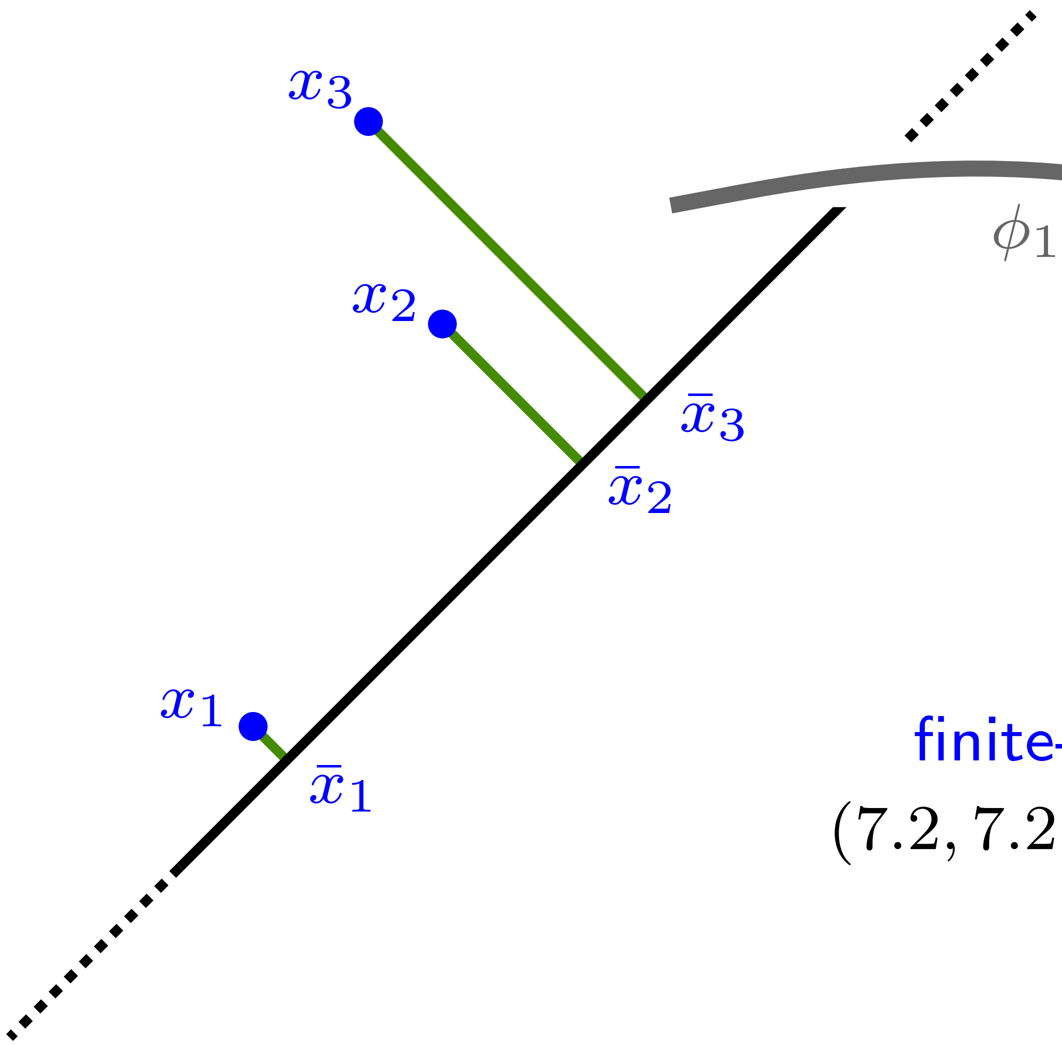
ϕ_3

finite-dimensional vector

$$(7.2, 7.2, 5.6, 4, 4, 4, 0, \cdots, 0)$$

Explicit Feature Map in $\mathbb{R}^d$

[*Stable topological signatures for points on 3D shapes*, Carrière, Oudot, Ovsjanikov, SGP, 2015]



ϕ_1

distance matrix

$$\begin{bmatrix} D_{ij} = \min\{ & \|x_i - x_j\|_\infty, \\ & \|x_i - \bar{x}_i\|_\infty, \\ & \|x_j - \bar{x}_j\|_\infty\} \end{bmatrix}$$

finite-dimensional vector

$$(7.2, 7.2, 5.6, 4, 4, 4, 0, \cdots, 0)$$

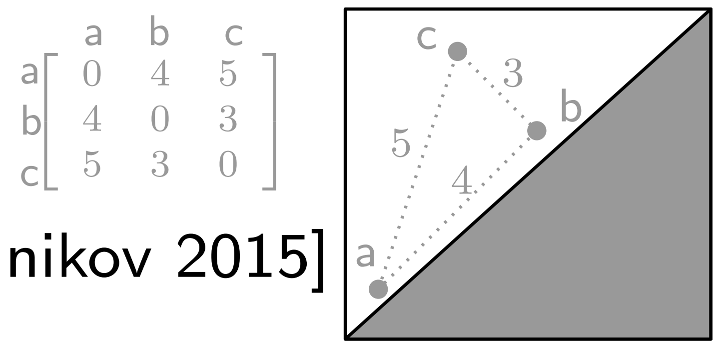
Prop:

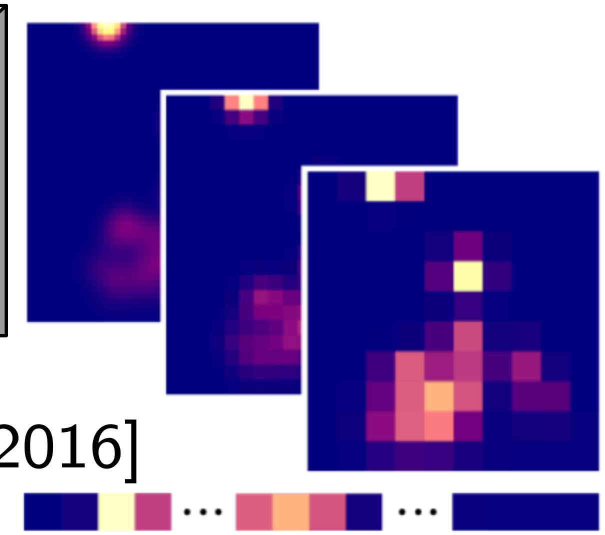
- $\|\Phi(\mathrm{D}) - \Phi(\mathrm{D}')\|_\infty \leq 2\, d_\infty(\mathrm{D}, \mathrm{D}')$
- $\|\Phi(\mathrm{D}) - \Phi(\mathrm{D}')\|_2 \leq 2\sqrt{d}\, d_\infty(\mathrm{D}, \mathrm{D}')$

Kernels for persistence diagrams

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

- images [Adams et al. 2015]

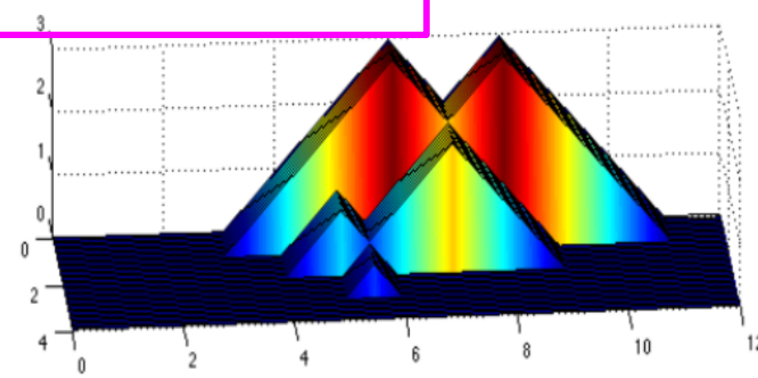




- finite metric spaces [Carrière, O., Ovsjanikov 2015]
- polynomial roots or evaluations [Di Fabio, Ferri 2015] [Kališnik 2016]

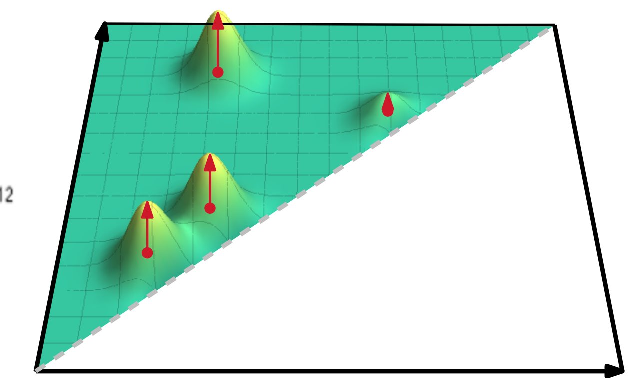
$$\{p_1, \cdots, p_n\} \mapsto (P_1(p_1, \cdots, p_n), \cdots, P_r(p_1, \cdots, p_n), \cdots)$$

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



- discrete measures:

  → histogram [Bendich et al. 2014]



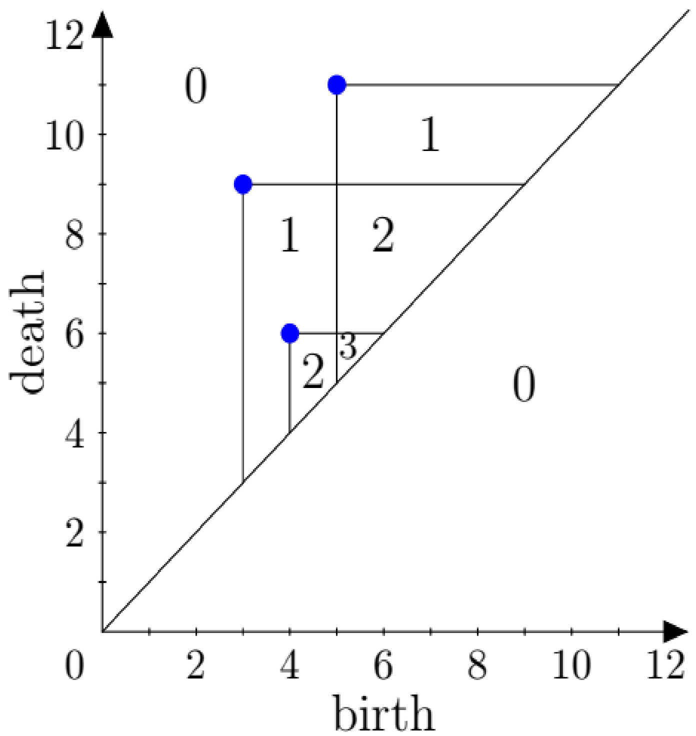
  → convolution with fixed kernel [Chepushtanova et al. 2015]

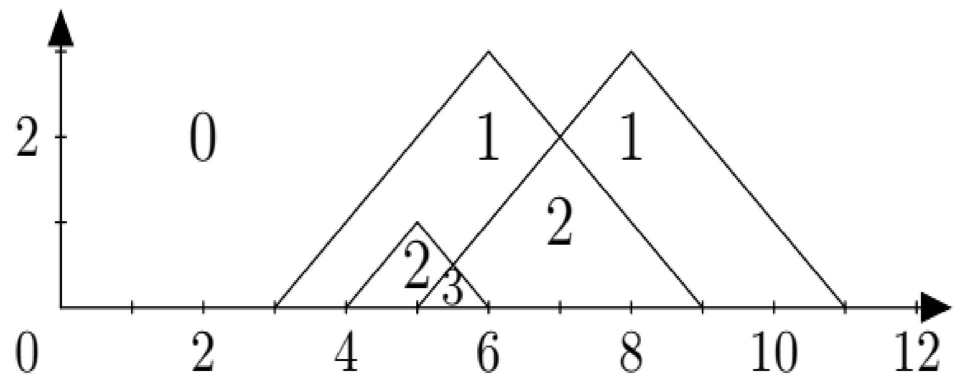
  → convolution with weighted kernel [Kusano, Fukumisu, Hiraoka 2016-17]

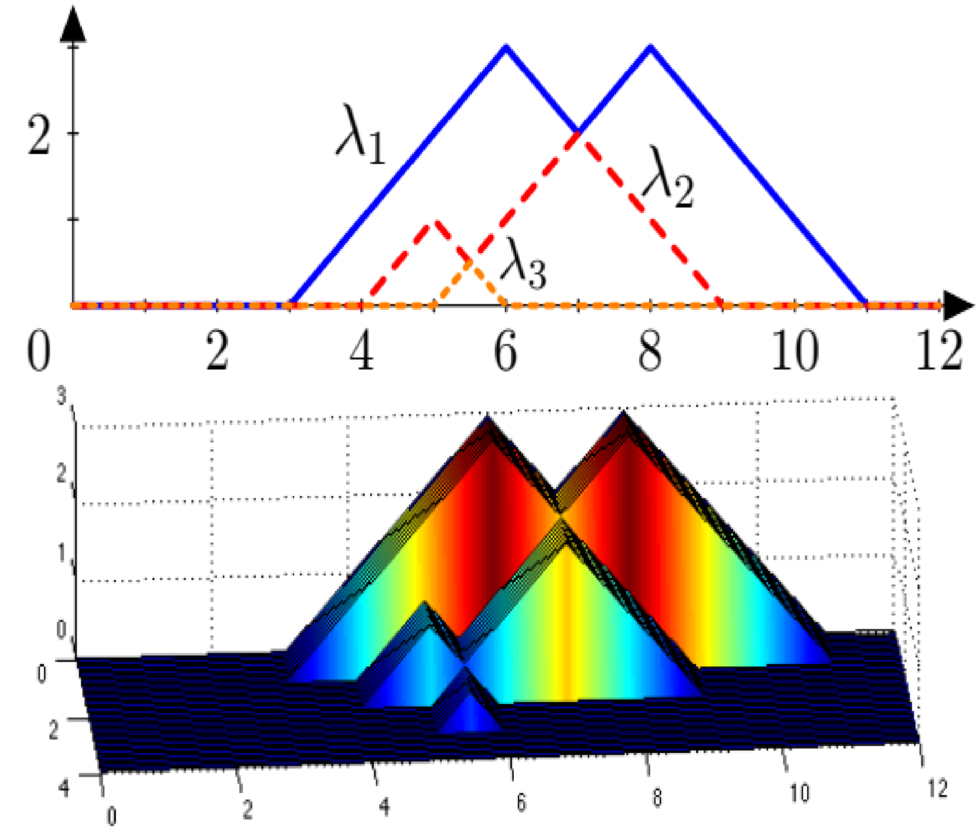
  → heat diffusion [Reininghaus et al. 2015] + exponential [Kwit et al. 2015]

Explicit Feature Map in Function Space

[*Statistical Topological Data Analysis using Persistence Landscapes*, Bubenik, JMLR, 2015]







Rotate PD
Compute rank function

Use boundaries of rank function

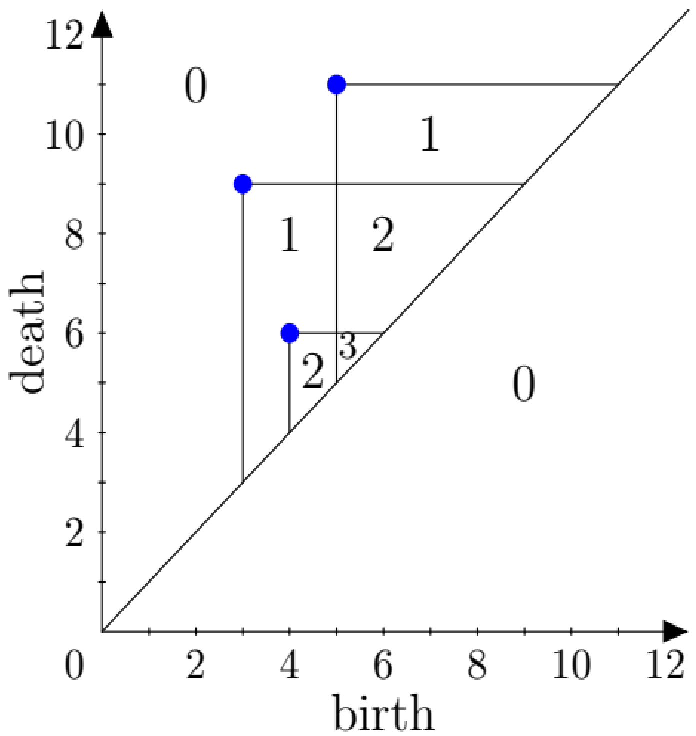
$$x \leq y \Longrightarrow f^{-1}(-\infty, x) \subseteq f^{-1}(-\infty, y)$$

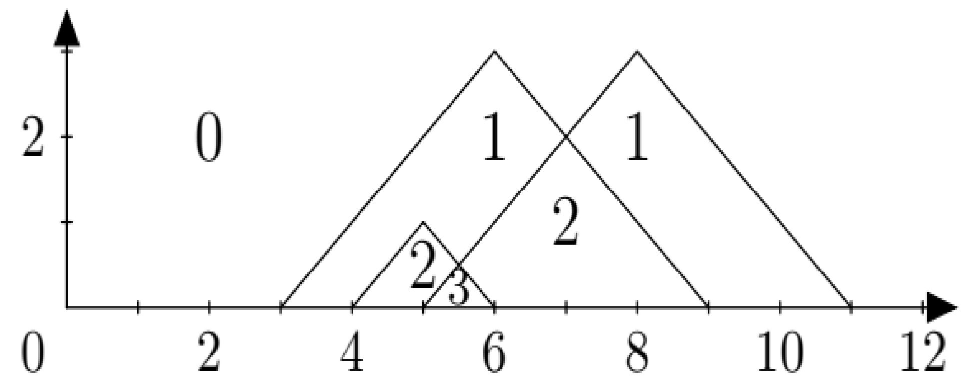
$\iota_x^y : H(f^{-1}(-\infty, x)) \to H(f^{-1}(-\infty, y))$ induced linear map

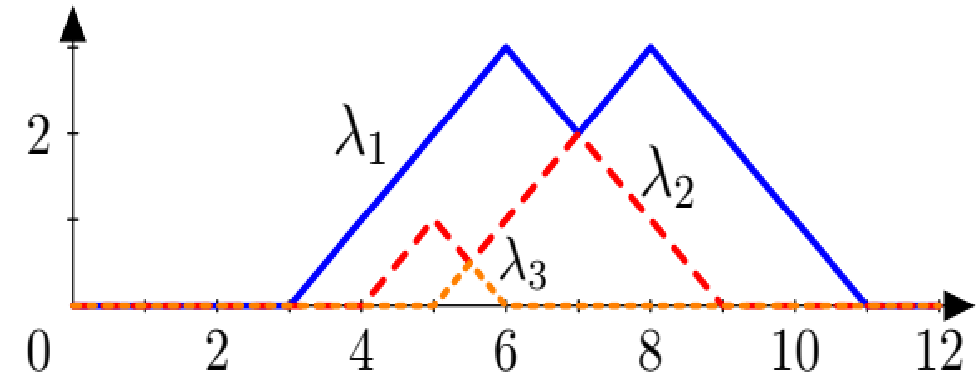
Rank function is defined as $\lambda(x, y) = \operatorname{rank} \iota_x^y$

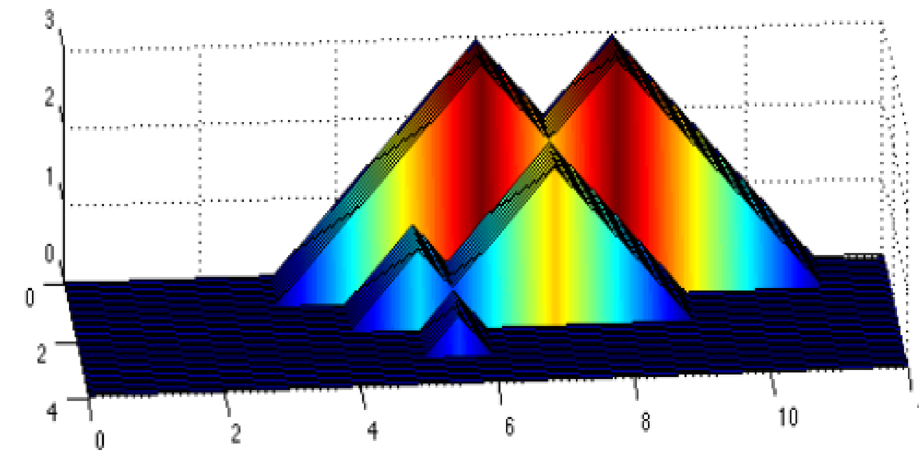
Explicit Feature Map in Function Space

[*Statistical Topological Data Analysis using Persistence Landscapes*, Bubenik, JMLR, 2015]









Rotate PD
Compute rank function

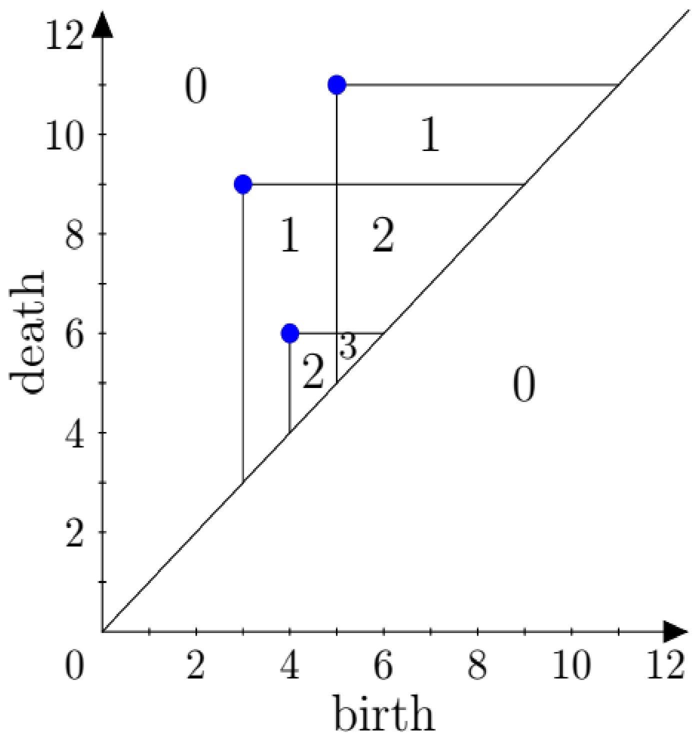
Use boundaries of rank function

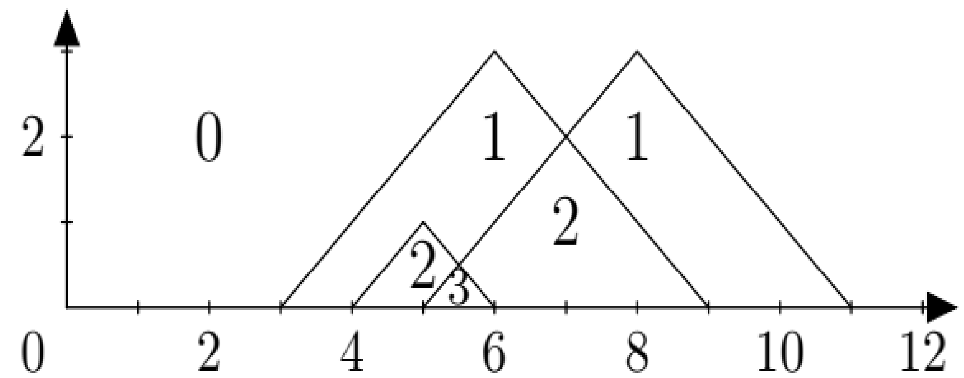
Boundaries of rank function: $\lambda_i(t) = \sup\{s \geq 0 : \lambda(t-s, t+s) \geq i\}$

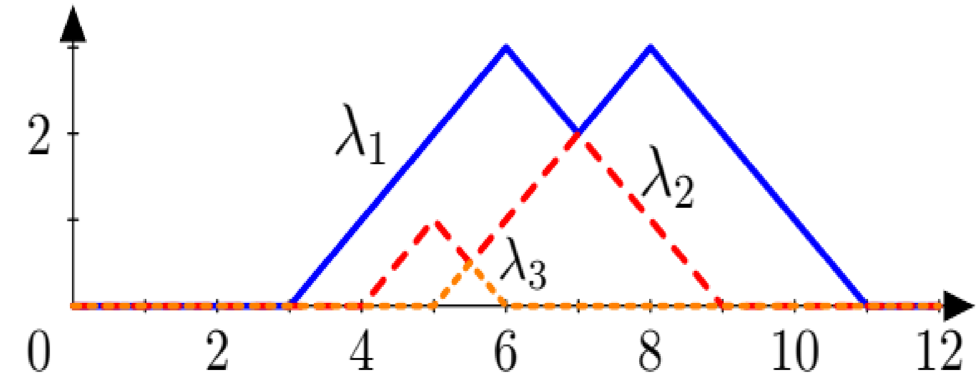
Landscape $\Lambda : \mathbb{R}^2 \to \mathbb{R}$ is defined as: $\Lambda(i,t) = \lambda_{\lfloor i \rfloor}(t)$

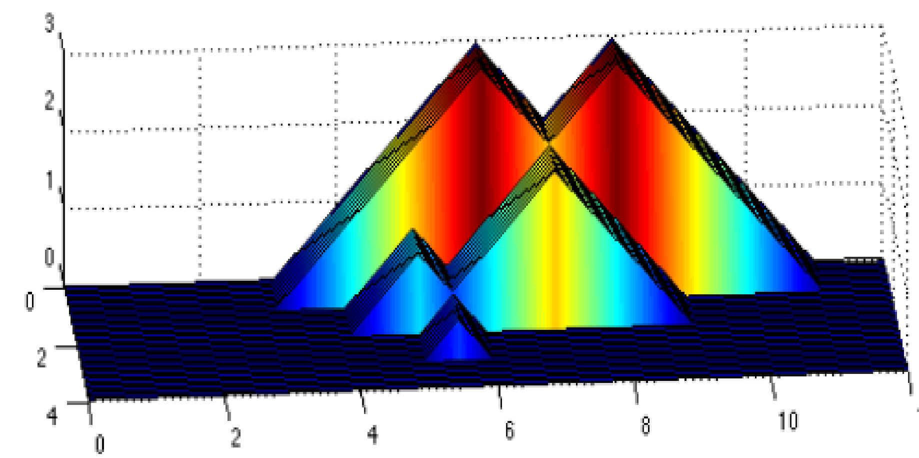
Explicit Feature Map in Function Space

[*Statistical Topological Data Analysis using Persistence Landscapes*, Bubenik, JMLR, 2015]









Rotate PD
Compute rank function

Use boundaries of rank function

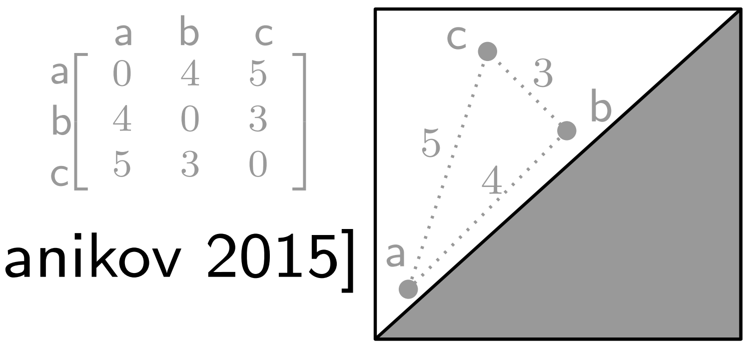
Prop:

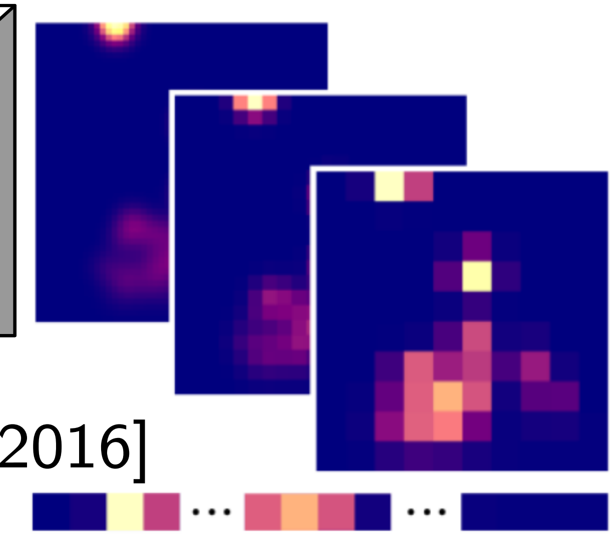
- $\|\Lambda(\mathrm{D}) - \Lambda(\mathrm{D}')\|_\infty \leq d_\infty(\mathrm{D}, \mathrm{D}')$
- $\min\{1, C(\mathrm{D}, \mathrm{D}')\|\Lambda(\mathrm{D}) - \Lambda(\mathrm{D}')\|_2\} \leq d_2(\mathrm{D}, \mathrm{D}')$

Kernels for persistence diagrams

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

- images [Adams et al. 2015]



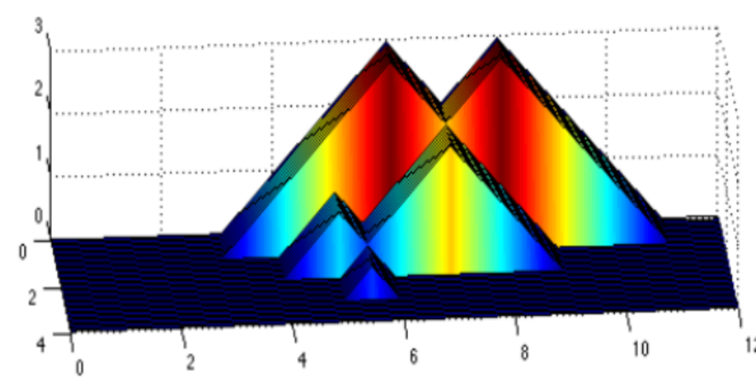


- finite metric spaces [Carrière, O., Ovsjanikov 2015]

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

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

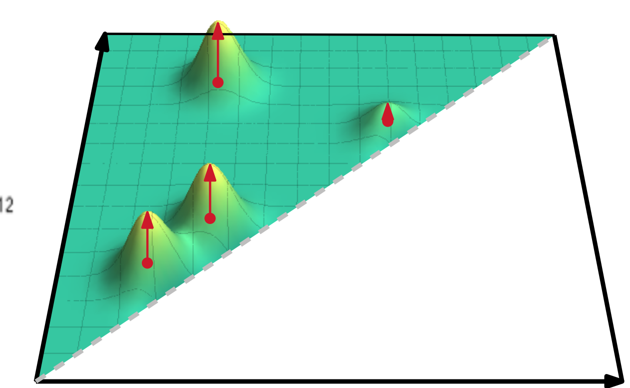
- landscapes [Bubenik 2012] [Bubenik, Dłotko 2015]



- discrete measures:

  → histogram [Bendich et al. 2014]

  → convolution with fixed kernel [Chepushtanova et al. 2015]

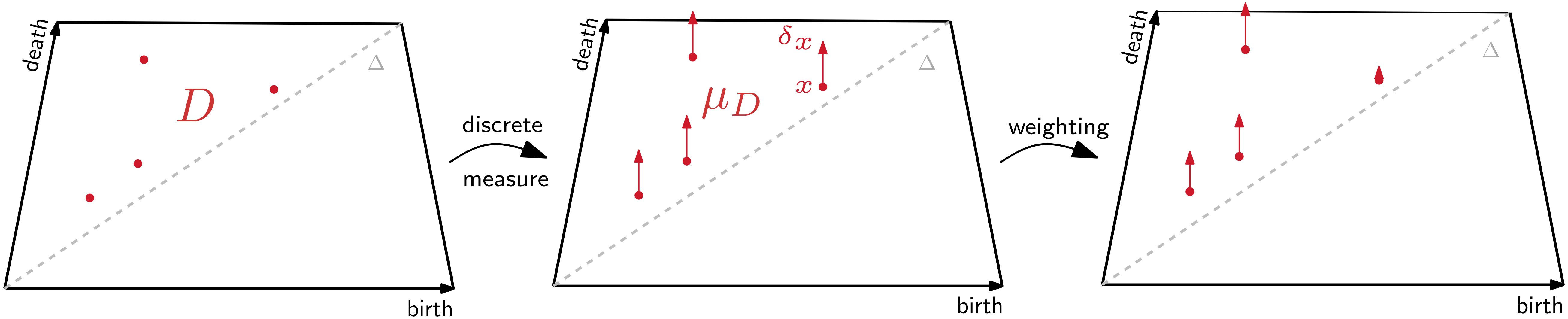


  → convolution with weighted kernel [Kusano, Fukumisu, Hiraoka 2016-17]

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

Explicit Feature Map in Function Space

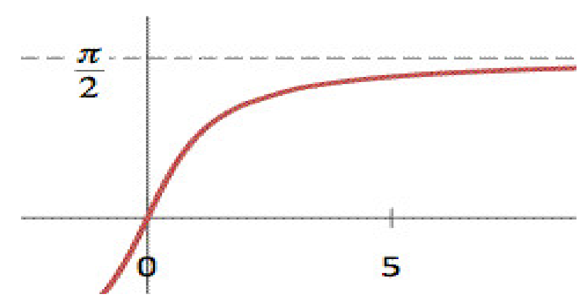
[*Persistence weighted Gaussian kernel for topological data analysis*, Kisano, Hiraoka, Fukumizu, ICML, 2016]



$\mu_{\mathrm{D}} := \sum_{x \in \mathrm{D}} \delta_x$ $\quad$ $\mu_{\mathrm{D}}^{w} := \sum_{x \in \mathrm{D}} w(x) \delta_{\bar{x}}$

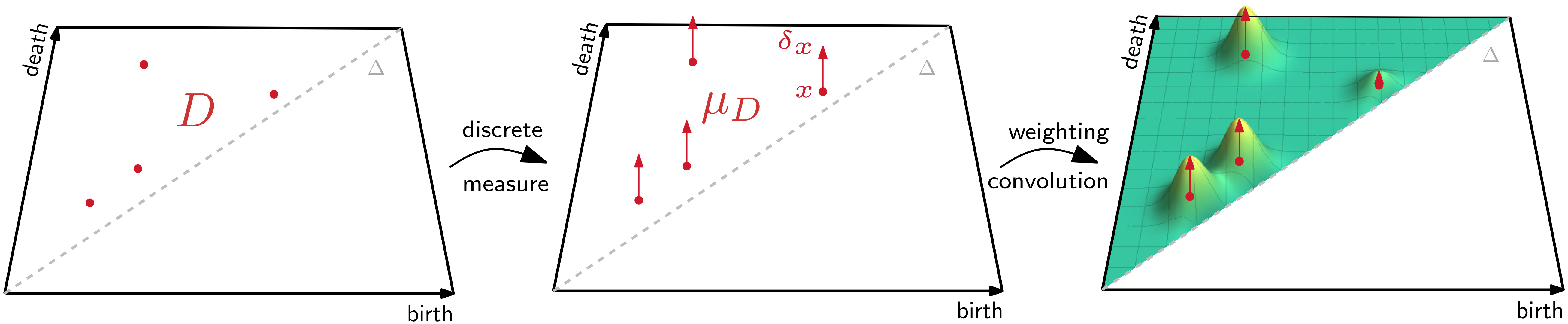
Pb: μ_{D} is unstable (points on diagonal disappear)

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



Explicit Feature Map in Function Space

[*Persistence weighted Gaussian kernel for topological data analysis*, Kisano, Hiraoka, Fukumizu, ICML, 2016]



$\mu_{\mathrm{D}} := \sum_{x \in \mathrm{D}} \delta_x$

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

$\tilde{\mu}_{\mathrm{D}}^w := \mu_{\mathrm{D}}^w * \mathcal{N}(0, \sigma)$

Pb: μ_{D} is unstable (points on diagonal disappear)

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

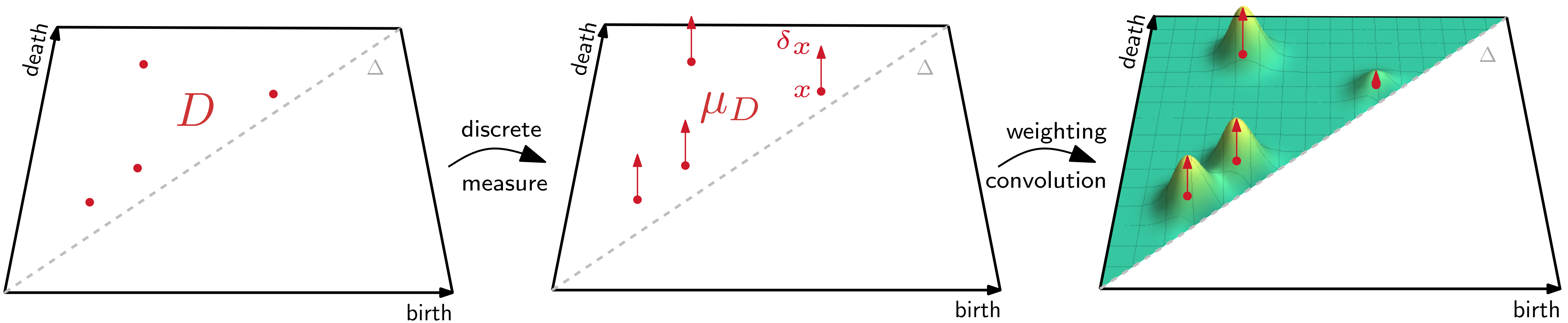
Def: $\phi(\mathrm{D})$ is the density function of $\mu_{\mathrm{D}}^w * \mathcal{N}(0, \sigma)$ w.r.t. Lebesgue measure:

$$\phi(\mathrm{D}) := \frac{1}{\sqrt{2\pi}\sigma} \sum_{x \in \mathrm{D}} \arctan(c\, d(x, \Delta)^r) \exp\left(- \frac{\| \cdot - x \|^2}{2\sigma^2} \right)$$

$$k(\mathrm{D}, \mathrm{D}') := \langle \phi(\mathrm{D}),\ \phi(\mathrm{D}') \rangle_{L_2(\Delta \times \mathbb{R}_+)}$$

Explicit Feature Map in Function Space

[*Persistence weighted Gaussian kernel for topological data analysis*, Kisano, Hiraoka, Fukumizu, ICML, 2016]



$\mu_{\mathrm{D}} := \sum_{x\in\mathrm{D}} \delta_x$

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

$\tilde{\mu}_{\mathrm{D}}^w := \mu_{\mathrm{D}}^w * \mathcal{N}(0,\sigma)$

Prop:

- $\|\phi(\mathrm{D}) - \phi(\mathrm{D}')\|_{\mathcal{H}} \leq \mathrm{cst}\, \mathrm{d}_p(\mathrm{D},\mathrm{D}')$.
- ϕ is injective and $\exp(k)$ is universal

Pb: convolution reduces discriminativity $\rightarrow$ use discrete measure instead

$$\phi(\mathrm{D}) := \frac{1}{\sqrt{2\pi}\sigma} \sum_{x\in\mathrm{D}} \arctan(c\, d(x,\Delta)^r)\, \exp\left(-\frac{\|\cdot - x\|^2}{2\sigma^2}\right)$$

$$k(\mathrm{D},\mathrm{D}') := \langle \phi(\mathrm{D}),\, \phi(\mathrm{D}')\rangle_{L_2(\Delta\times\mathbb{R}_+)}$$

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	✓	✓	✓	✓	✓
$\\|\phi(\cdot) - \phi(\cdot)\\|_{\mathcal{H}} \leq \phi(d_p)$	✓	✓	✓	✓	✓
$\\|\phi(\cdot) - \phi(\cdot)\\|_{\mathcal{H}} \geq \psi(d_p)$	✗	✗	✗	✗	✗
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: $O(d)$	$O(n^2)$	$O(n^2)$

One kernel to rule them all...

[*Sliced Wasserstein Kernel for persistence diagrams*, Carrière, Cuturi, Oudot, ICML, 2017]

Sliced Wasserstein Kernel

No feature map

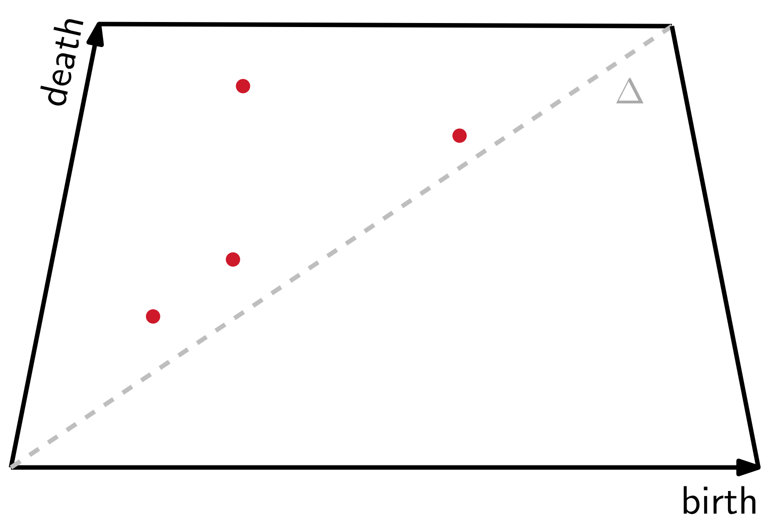
Provably stable

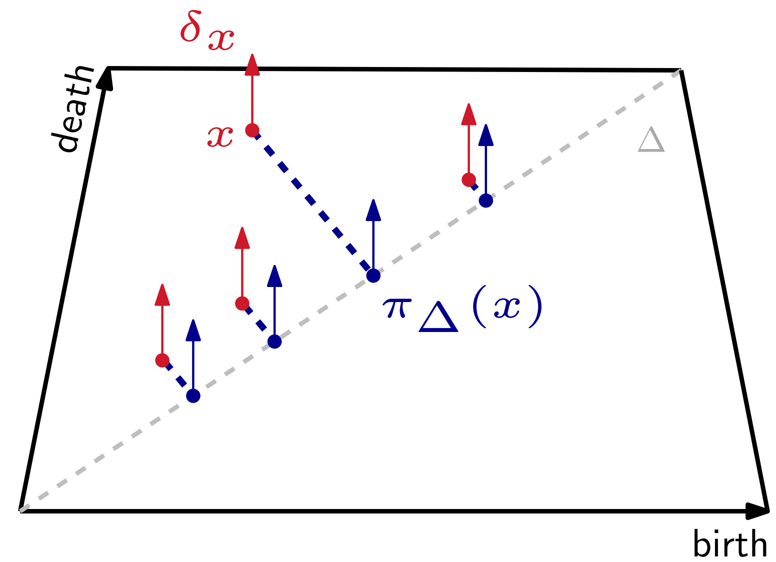
Provably discriminative

Mimicks the Gaussian kernel

View diagrams as discrete measures w/o density functions

Persistence diagrams as discrete measures





$\mu_{\mathrm{D}} := \sum_{x \in \mathrm{D}} \delta_x$

Pb: $d_p(\mathrm{D}, \mathrm{D}') \not\propto W_p(\mu_{\mathrm{D}}, \mu_{\mathrm{D}'})$ (W_p does not even make sense here)

$\rightarrow$ given D, D', let

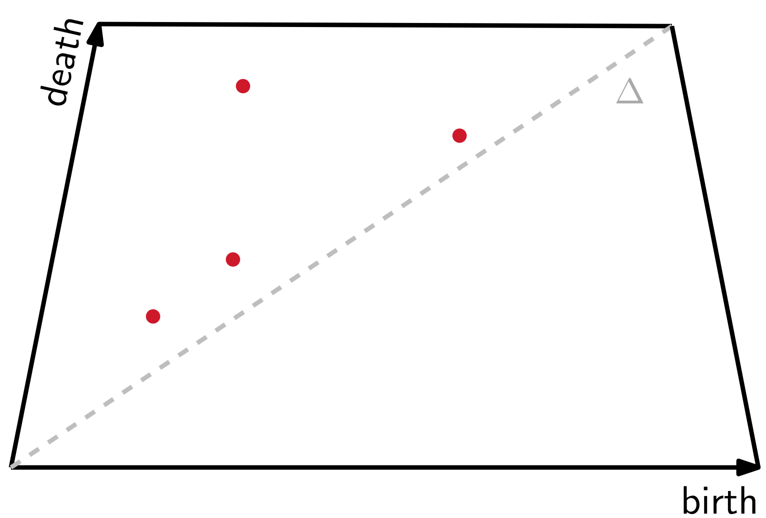
$$\bar{\mu}_{\mathrm{D}} := \sum_{x \in \mathrm{D}} \delta_x + \sum_{y \in \mathrm{D}'} \delta_{\pi_\Delta(y)}$$

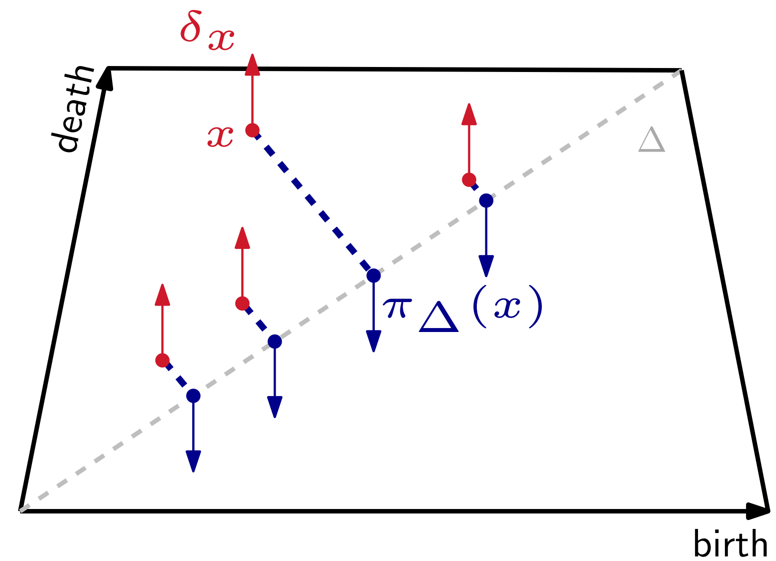
$$\bar{\mu}_{\mathrm{D}'} := \sum_{y \in \mathrm{D}'} \delta_y + \sum_{x \in \mathrm{D}} \delta_{\pi_\Delta(x)}$$

Then, $d_p(\mathrm{D}, \mathrm{D}') \leq W_p(\bar{\mu}_{\mathrm{D}}, \bar{\mu}_{\mathrm{D}'}) \leq 2\, d_p(\mathrm{D}, \mathrm{D}')$

Pb: $\bar{\mu}_{\mathrm{D}}$ depends on D'

Persistence diagrams as discrete measures





$\mu_{\mathrm{D}} := \sum_{x\in\mathrm{D}} \delta_x$

Pb: $d_p(\mathrm{D},\mathrm{D}') \not\propto W_p(\mu_{\mathrm{D}},\mu_{\mathrm{D}'})$ (W_p does not even make sense here)

Solution: transfer mass negatively to μ_D:

$$\tilde{\mu}_{\mathrm{D}} := \sum_{x\in\mathrm{D}} \delta_x - \sum_{x\in\mathrm{D}} \delta_{\pi_\Delta(x)} \quad \in \mathcal{M}_0(\mathbb{R}^2)$$

$\rightarrow$ signed discrete measure of total mass zero

metric: Kantorovich norm $\|\cdot\|_K$

Persistence diagrams as discrete measures

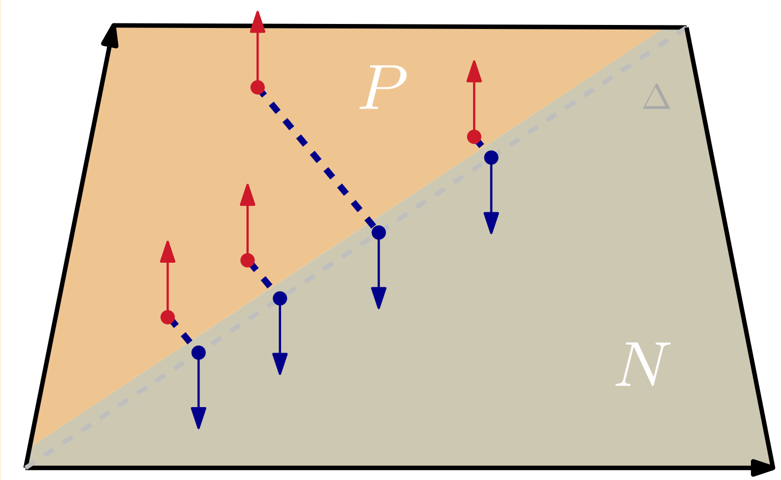
Hahn decomp. thm: For any $\mu \in \mathcal{M}_0(X, \Sigma)$ there exist measurable sets P, N such that:

(i) $P \cup N = X$ and $P \cap N = \emptyset$

(ii) $\mu(B) \geq 0$ for every measureable set $B \subseteq P$

(iii) $\mu(B) \leq 0$ for every measureable set $B \subseteq N$

Moreover, the decomposition is essentially unique.



$\forall B \in \Sigma$, let $\mu^+(B) := \mu(B \cap P)$ and $\mu^-(B) := -\mu(B \cap N) \in \mathcal{M}_+(X)$

Def: $\|\mu\|_K := W_1(\mu^+, \mu^-)$

Prop: $\forall \mu, \nu \in \mathcal{M}_0(X)$, $\quad W_1(\underbrace{\mu^+ + \nu^-}_{\bar{\mu}_{\mathrm{D}}}, \underbrace{\nu^+ + \mu^-}_{\bar{\mu}_{\mathrm{D}'}}) = \|\underset{\tilde{\mu}_{\mathrm{D}}}{\mu} - \underset{\tilde{\mu}_{\mathrm{D}'}}{\nu}\|_K$

for persistence diagrams:

$W_1(\bar{\mu}_{\mathrm{D}}, \bar{\mu}_{\mathrm{D}'}) = \|\tilde{\mu}_{\mathrm{D}} - \tilde{\mu}_{\mathrm{D}'}\|_K$

A Wasserstein Gaussian kernel for PDs?

Thm:
If $d : X \times X \to \mathbb{R}_+$ symmetric is *conditionally negative semidefinite*, i.e.:

$$\forall n \in \mathbb{N},\ \forall x_1, \ldots, x_n \in X,\ \sum_{i=1}^{n} \alpha_i = 0 \Longrightarrow \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \, d(x_i, x_j) \leq 0,$$

then $k(x, y) := \exp\left(-\frac{d(x,y)}{2\sigma^2}\right)$ is positive semidefinite.

Pb: W_1 is not cnsd, neither is d_1

Solutions:

- relax the measures (e.g. convolution)
- relax the metric (e.g. regularization, slicing)

Sliced Wasserstein metric

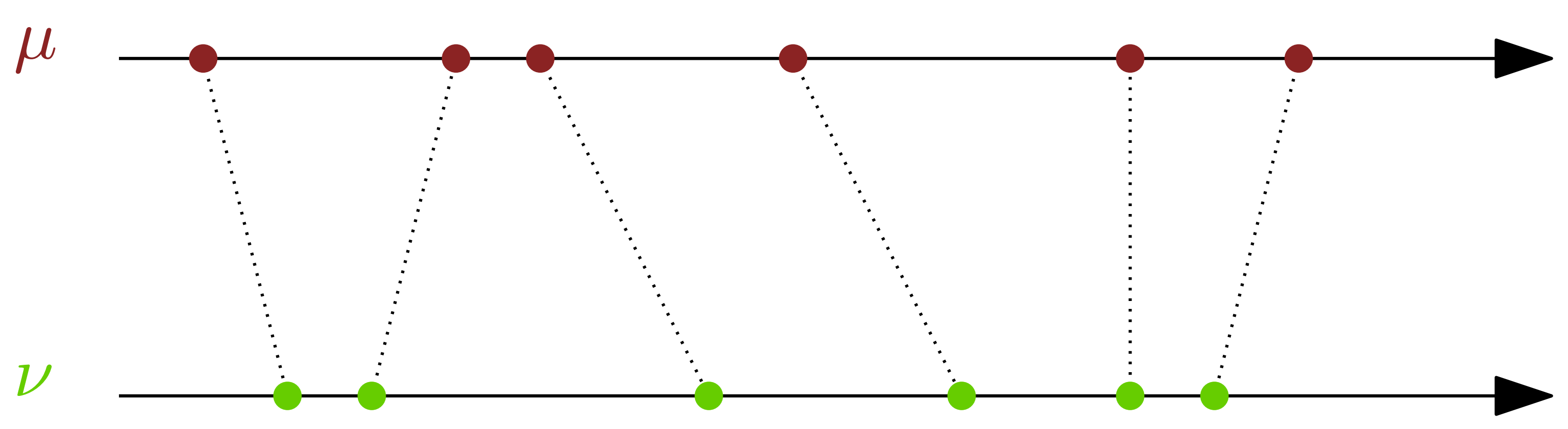
[*Sliced Wasserstein Kernel for persistence diagrams*, Carrière, Cuturi, Oudot, ICML, 2017]

Special case: $X = \mathbb{R}$, μ, ν discrete measures of mass n

$\mu := \sum_{i=1}^n \delta_{x_i}, \;\; \nu := \sum_{i=1}^n \delta_{y_i}$

Sort the atoms of μ, ν along the real line: $x_i \leq x_{i+1}$ and $y_i \leq y_{i+1}$ for all i

Then: $W_1(\mu, \nu) = \sum_{i=1}^n |x_i - y_i| = \|(x_1, \cdots, x_n) - (y_1, \cdots, y_n)\|_1$



$\rightarrow W_1$ is cnsd and easy to compute (same with $\|\cdot\|_K$ for signed measures)

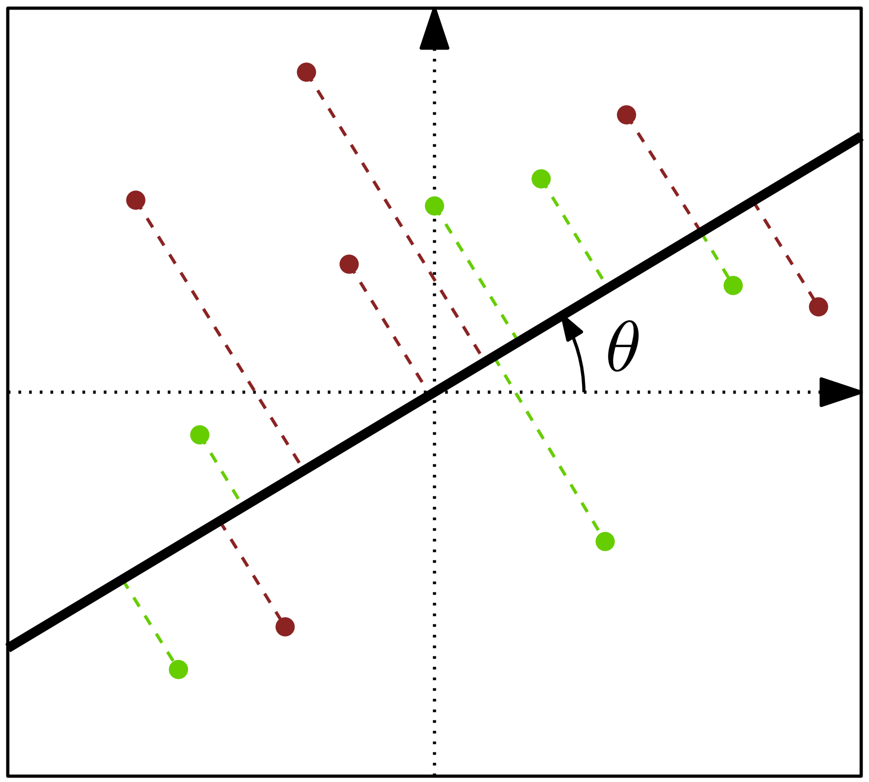
Sliced Wasserstein metric

[*Sliced Wasserstein Kernel for persistence diagrams*, Carrière, Cuturi, Oudot, ICML, 2017]

Def: (sliced Wasserstein distance) for $\mu, \nu \in \mathcal{M}_+(\mathbb{R}^2)$,

$$SW_1(\mu,\nu) := \frac{1}{2\pi}\int_{\theta\in\mathbb{S}^1} W_1(\pi_\theta\#\mu,\, \pi_\theta\#\nu)\, d\theta$$

where $\pi_\theta =$ orthogonal projection onto line passing through origin with angle θ.



$\rightarrow$ from integral geometry: $\int_{\mathrm{Gr}(1,2)} \cdots$

Sliced Wasserstein metric

[*Sliced Wasserstein Kernel for persistence diagrams*, Carrière, Cuturi, Oudot, ICML, 2017]

Def: (sliced Wasserstein distance) for $\mu, \nu \in \mathcal{M}_+(\mathbb{R}^2)$,

$$SW_1(\mu,\nu) := \frac{1}{2\pi} \int_{\theta \in \mathbb{S}^1} W_1(\pi_\theta \# \mu,\, \pi_\theta \# \nu)\, d\theta$$

where $\pi_\theta =$ orthogonal projection onto line passing through origin with angle θ.

Props: (inherited from W_1 over $\mathbb{R}$)

- satisfies the axioms of a metric

- well-defined barycenters, fast to compute via stochastic gradient descent, etc.

- conditionally negative semidefinite

Sliced Wasserstein kernel

[*Sliced Wasserstein Kernel for persistence diagrams*, Carrière, Cuturi, Oudot, ICML, 2017]

Def: Given $\sigma > 0$, for any $\mu, \nu \in \mathcal{M}_+(\mathbb{R}^2)$:

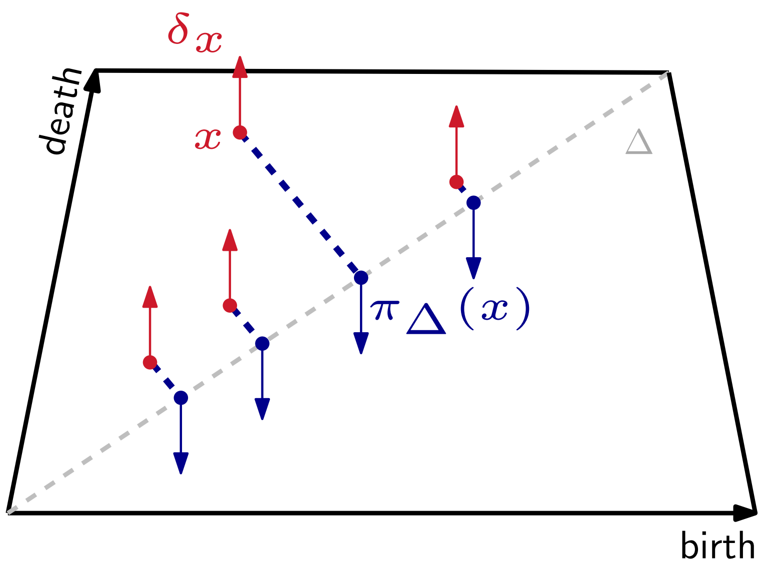
$$k_{SW}(\mu,\nu) := \exp\left(-\frac{SW_1(\mu,\nu)}{2\sigma^2}\right)$$

Cor: (from SW cnsd)
k_{SW} is positive semidefinite.

→ application to persistence diagrams:

$$\mathrm{D} \mapsto \mu_\mathrm{D} := \textstyle\sum_{x\in\mathrm{D}} \delta_x$$

$$\mapsto \tilde{\mu}_\mathrm{D} := \mu_\mathrm{D} - \pi_\Delta \# \mu_\mathrm{D}$$



$$SW_1(\mathrm{D},\mathrm{D}') := \int_{\theta\in\mathcal{S}^1} \|\pi_\theta\#\tilde{\mu}_\mathrm{D} - \pi_\theta\#\tilde{\mu}_{\mathrm{D}'}\|_K \, d\theta$$

$$k_{SW}(\mathrm{D},\mathrm{D}') := \exp\left(-\frac{SW_1(\mathrm{D},\mathrm{D}')}{2\sigma^2}\right)$$

- positive semidefinite

- simple and fast to compute

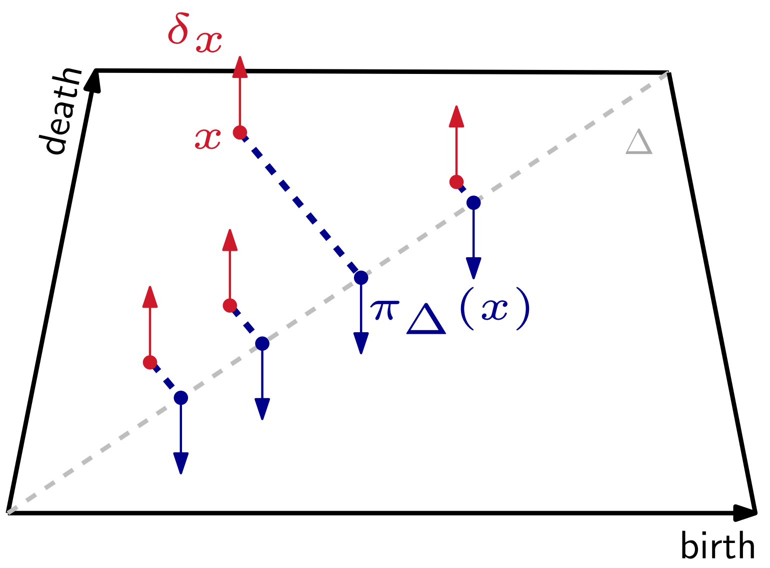
Sliced Wasserstein kernel

[*Sliced Wasserstein Kernel for persistence diagrams*, Carrière, Cuturi, Oudot, ICML, 2017]

Thm:
The metrics d_1 and SW_1 on the space $\mathcal{D}_N$ of persistence diagrams of size bounded by N are strongly equivalent, namely: for $\mathrm{D}, \mathrm{D}' \in \mathcal{D}_N$,

$$\frac{1}{2+4N(2N-1)}\,\mathrm{d}_1(\mathrm{D},\mathrm{D}') \;\leq\; SW_1(\mathrm{D},\mathrm{D}') \;\leq\; 2\sqrt{2}\,\mathrm{d}_1(\mathrm{D},\mathrm{D}')$$

Q: prove it.



$\rightarrow$ application to persistence diagrams:

$$\mathrm{D} \mapsto \mu_\mathrm{D} := \textstyle\sum_{x\in\mathrm{D}} \delta_x$$

$$\mapsto \tilde{\mu}_\mathrm{D} := \mu_\mathrm{D} - \pi_\Delta \# \mu_\mathrm{D}$$

$$SW_1(\mathrm{D},\mathrm{D}') := \int_{\theta\in\mathcal{S}^1} \|\pi_\theta\#\tilde{\mu}_\mathrm{D} - \pi_\theta\#\tilde{\mu}_{\mathrm{D}'}\|_K \, d\theta$$

$$k_{SW}(\mathrm{D},\mathrm{D}') := \exp\left(-\frac{SW_1(\mathrm{D},\mathrm{D}')}{2\sigma^2}\right)$$

Sliced Wasserstein kernel

[*Sliced Wasserstein Kernel for persistence diagrams*, Carrière, Cuturi, Oudot, ICML, 2017]

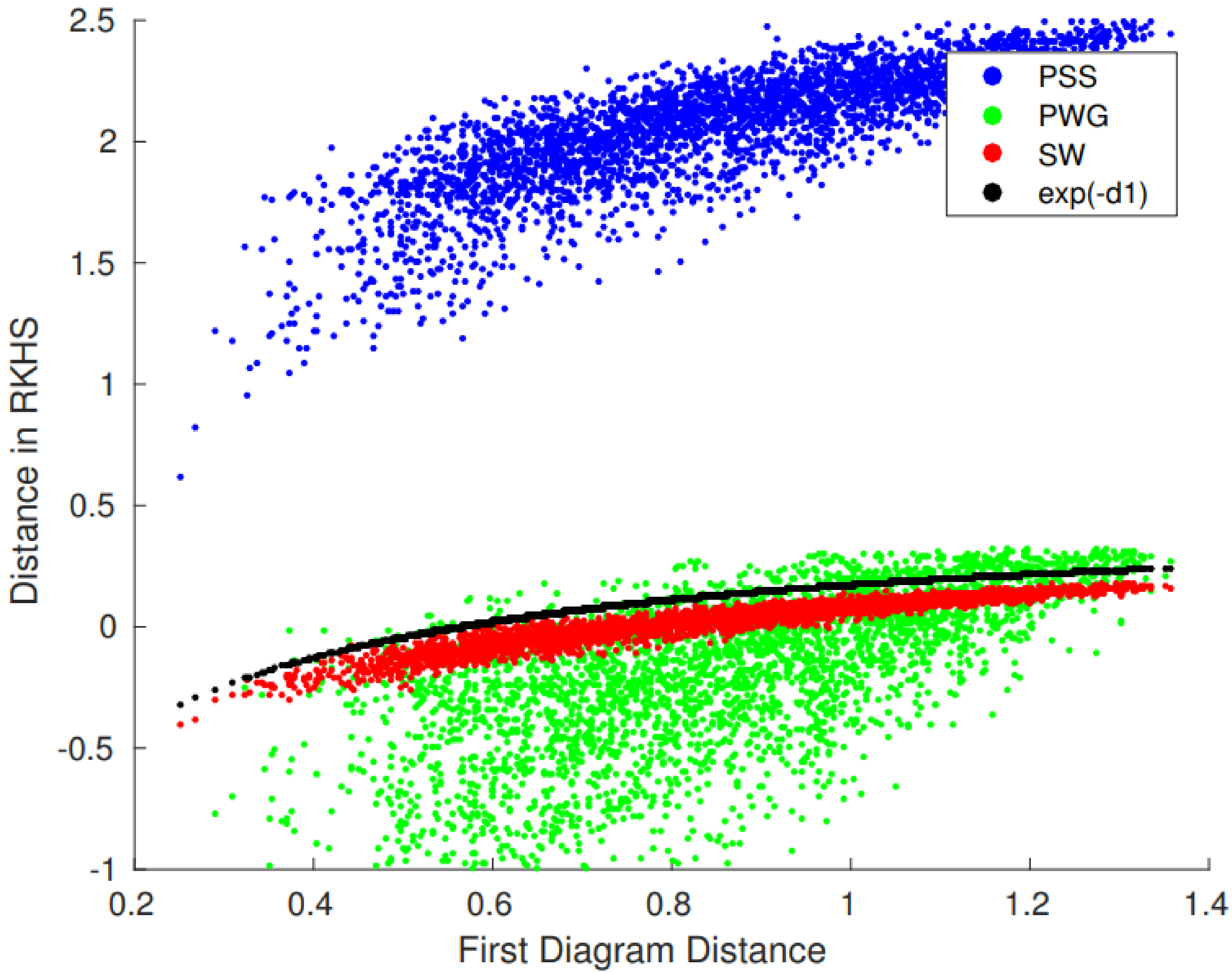
Thm:
The metrics d_1 and SW_1 on the space $\mathcal{D}_N$ of persistence diagrams of size bounded by N are strongly equivalent, namely: for $\mathrm{D}, \mathrm{D}' \in \mathcal{D}_N$,

$$\frac{1}{2+4N(2N-1)}\,\mathrm{d}_1(\mathrm{D},\mathrm{D}') \;\leq\; SW_1(\mathrm{D},\mathrm{D}') \;\leq\; 2\sqrt{2}\,\mathrm{d}_1(\mathrm{D},\mathrm{D}')$$

Q: prove it.

Cor: The feature map ϕ associated with k_{SW} is weakly metric-preserving: $\exists g, h$ nonzero except at 0 such that $g \circ d_1 \leq \|\phi(\cdot) - \phi(\cdot)\|_{\mathcal{H}} \leq h \circ d_1$.

Metric distortion in practice

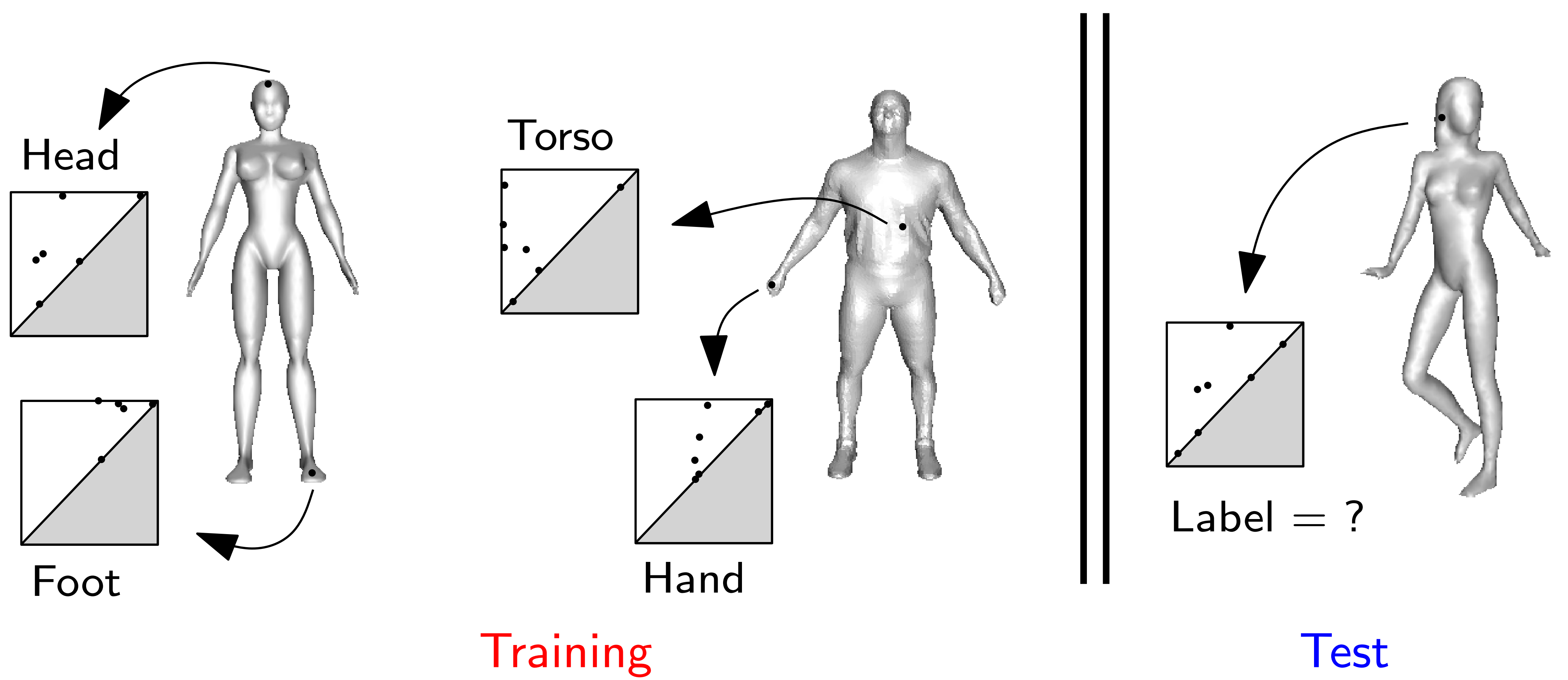


Application to supervised shape segmentation

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

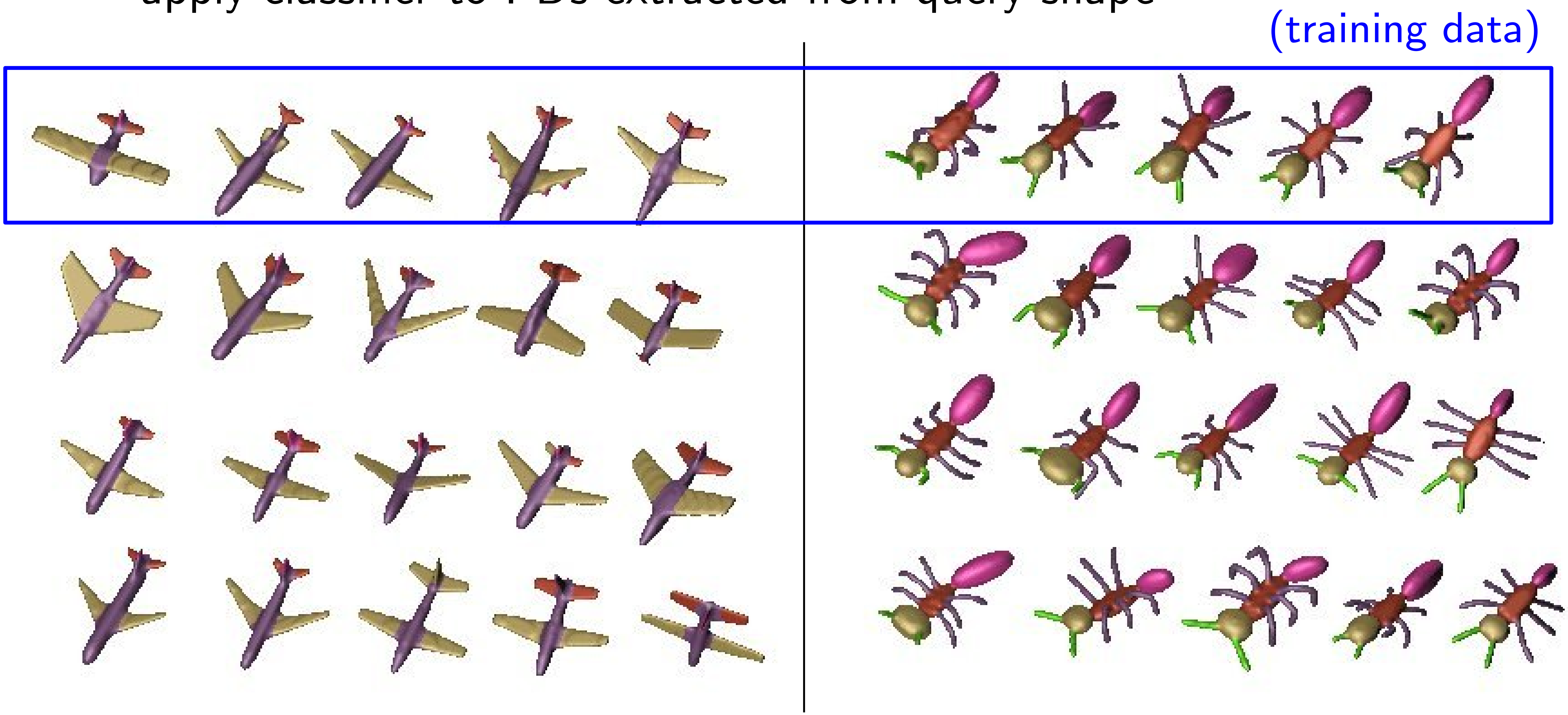


Application to supervised shape segmentation

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

Error rates (%) using TDA descriptors (kernels on barcodes):

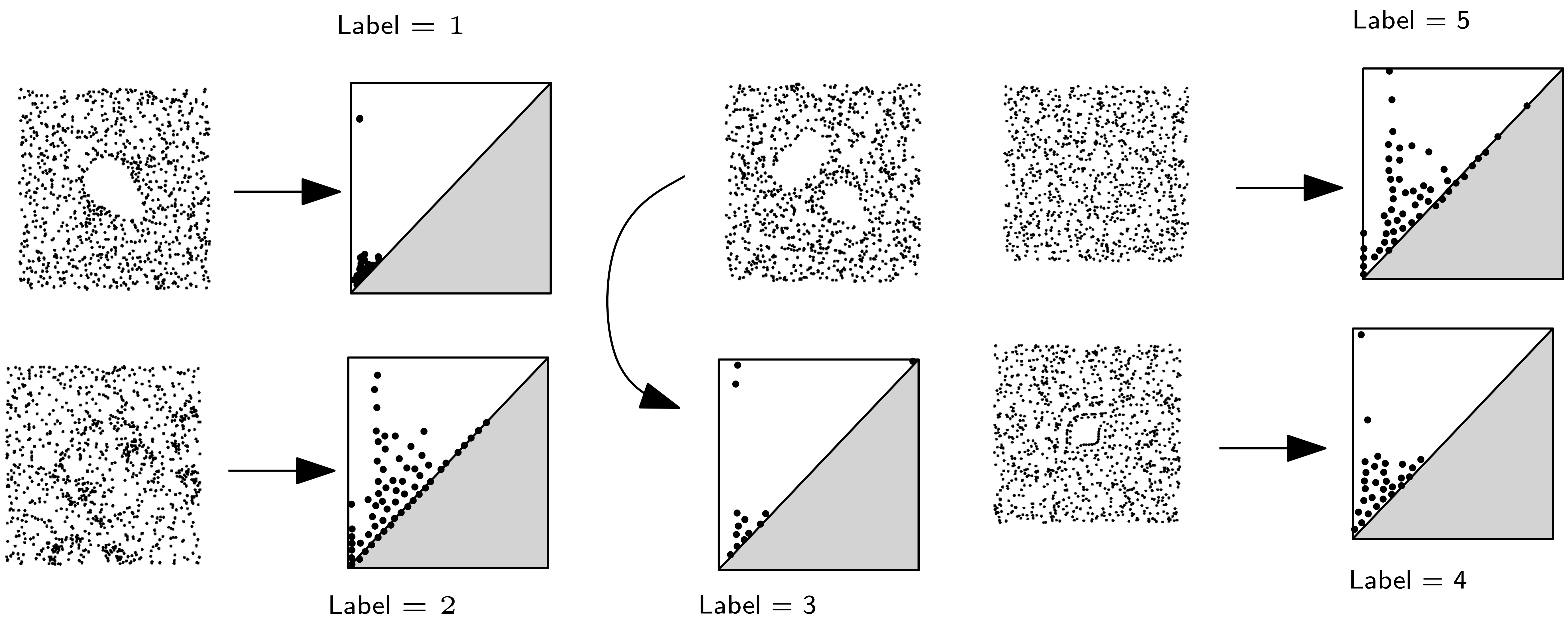
	TDA	geometry/stats	TDA + geometry/stats
Human	26.0	21.3	**11.3**
Airplane	27.4	18.7	**9.3**
Ant	7.7	9.7	**1.5**
FourLeg	27.0	25.6	**15.8**
Octopus	14.8	5.5	**3.4**
Bird	28.0	24.8	**13.5**
Fish	20.4	20.9	**7.7**

Application to supervised orbits classification

Goal: classify orbits of *linked twisted map*, modelling fluid flow dynamics

Orbits described by (depending on parameter r):

$$\begin{cases} x_{n+1} = x_n + r\, y_n(1 - y_n) \mod 1 \\ y_{n+1} = y_n + r\, x_{n+1}(1 - x_{n+1}) \mod 1 \end{cases}$$



Application to supervised orbits classification

Goal: classify orbits of *linked twisted map*, modelling fluid flow dynamics

Orbits described by (depending on parameter r):

$$\begin{cases} x_{n+1} = x_n + r\, y_n(1 - y_n) \mod 1 \\ y_{n+1} = y_n + r\, x_{n+1}(1 - x_{n+1}) \mod 1 \end{cases}$$

Accuracies ($\%$) using only TDA descriptors (kernels on barcodes):

	k_{PSS}	k_{PWG}	k_{SW}
Orbit	64.0 ± 0.0	78.7 ± 0.0	$\mathbf{83.7} \pm 1.1$

(PDs as discrete measures)

Running times (in seconds on N-sized parameter space from 100 orbits):

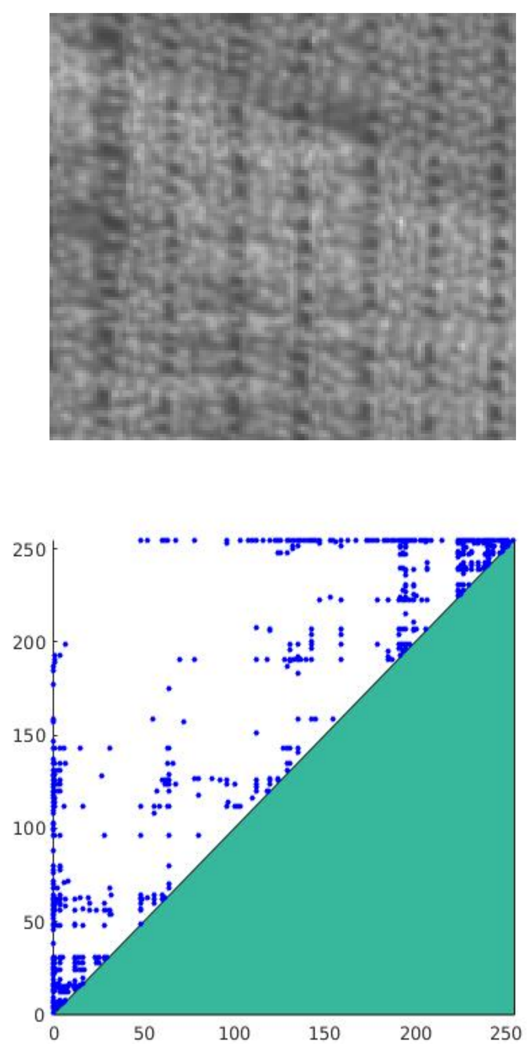
	k_{PSS}	k_{PWG}	k_{SW}
Orbit	$N \times 9183.4 \pm 65.6$	$N \times 69.2 \pm 0.9$	$385.8 \pm 0.2 + NC$

Application to supervised texture classification

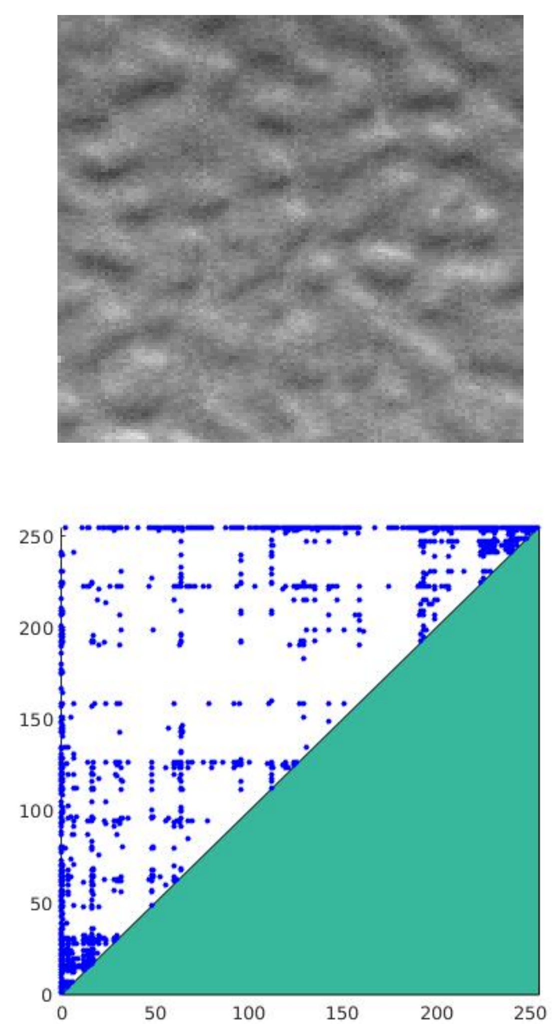
Goal: classify textures from the OUTEX00000 database

Textures described by CLBP (Compound Local Binary Pattern)

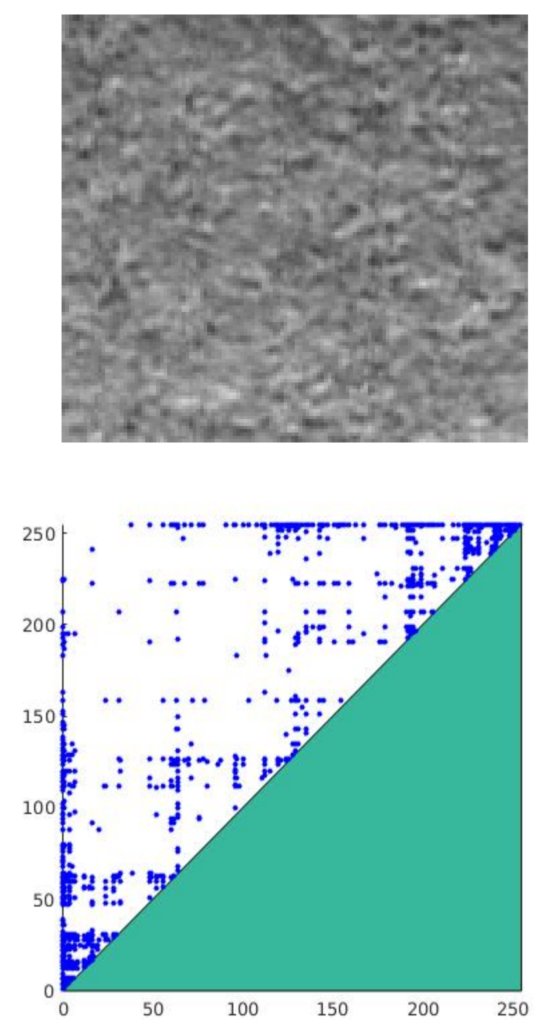
→ apply degree-0 persistence on 1st sign component



Label = Canvas



Label = Carpet



Label = Tile

Application to supervised texture classification

Goal: classify textures from the OUTEX00000 database

Textures described by CLBP (Compound Local Binary Pattern)

$\rightarrow$ apply degree-0 persistence on 1st sign component

Accuracies ($\%$) using only TDA descriptors (kernels on barcodes):

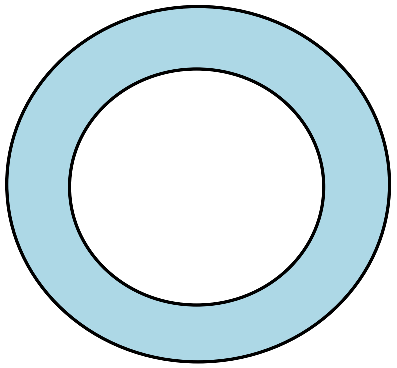
	k_{PSS}	k_{PWG}	k_{SW}
Orbit	$\mathbf{98.7} \pm 0.06$	96.7 ± 0.4	96.1 ± 0.1

(PDs as discrete measures)

Running times (in seconds on N-sized parameter space from 100 orbits):

	k_{PSS}	k_{PWG}	k_{SW}
Orbit	$N \times 10337.4 \pm 140.5$	$N \times 45.9 \pm 0.6$	$126.4 \pm 0.2 + NC$

Statistics on Persistence Diagrams

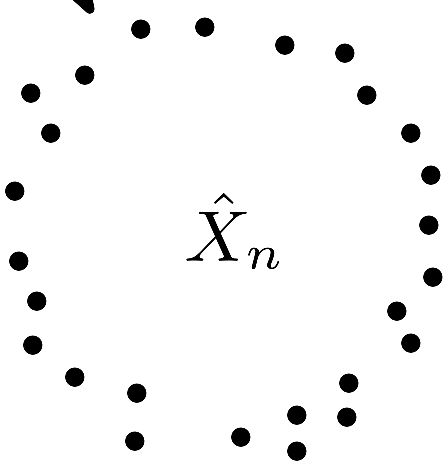


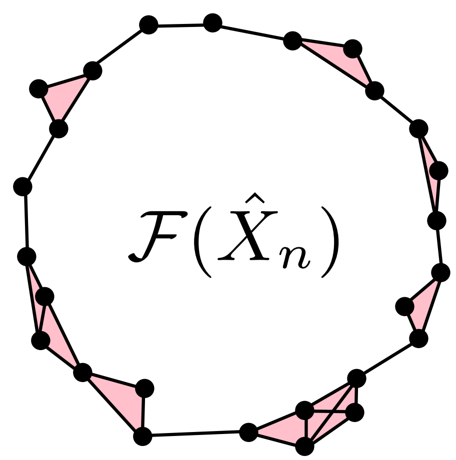
(X, d) metric space

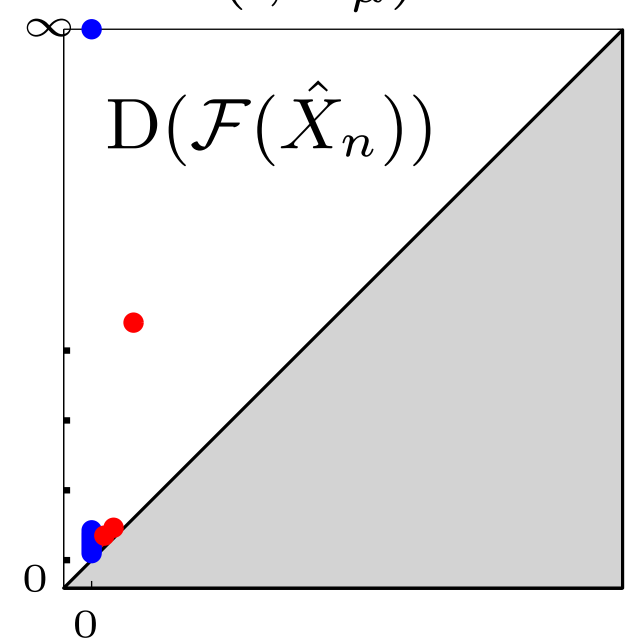
μ probability measure with compact support X_μ

Examples:
- $\mathcal{F}(\hat{X}_n) = \mathrm{Rips}(\hat{X}_n)$
- $\mathcal{F}(\hat{X}_n) = \check{\mathrm{C}}\mathrm{ech}(\hat{X}_n)$
- $\mathcal{F}(\hat{X}_n) =$ sublevelset filtration of $d(., X_\mu)$.

Sample n points according to μ.



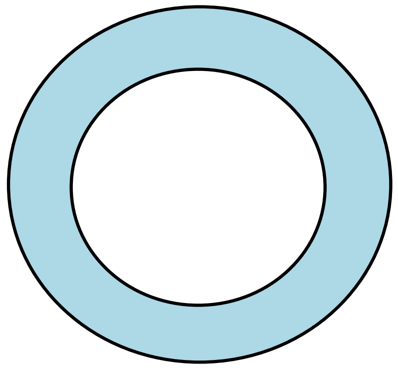




Questions:

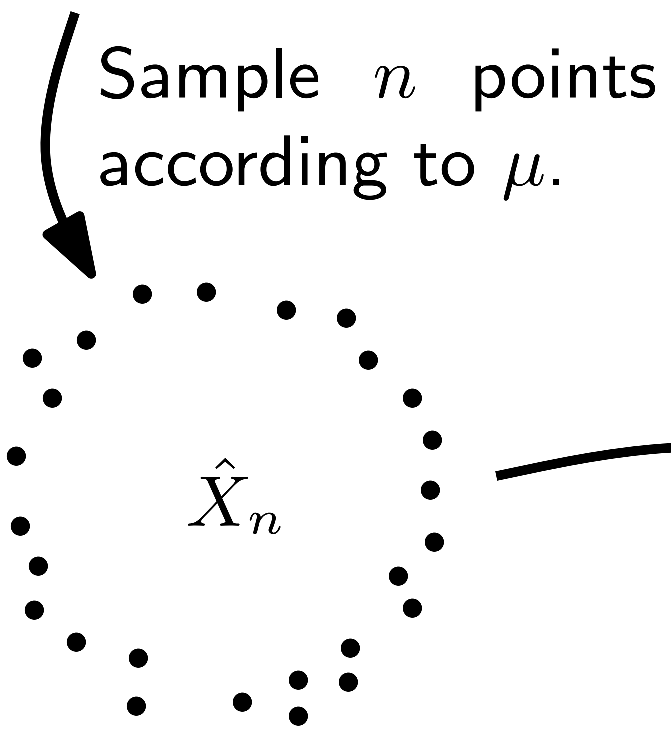
- Statistical properties of $\mathrm{D}(\mathcal{F}(\hat{X}_n))$? $\mathrm{D}(\mathcal{F}(\hat{X}_n)) \to ?$ as $n \to +\infty$?
- Can we do more statistics with persistence diagrams?

Statistics on Persistence Diagrams



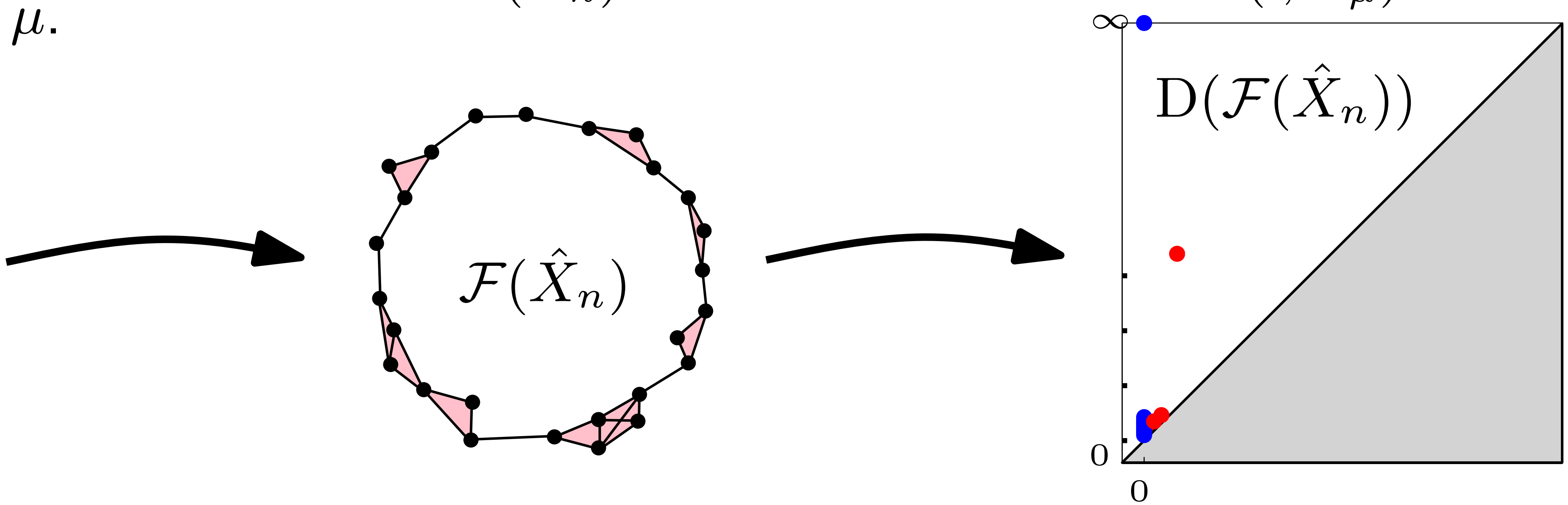
(X, d) metric space

μ probability measure with compact support X_μ



Examples:

- $\mathcal{F}(\hat{X}_n) = \mathrm{Rips}(\hat{X}_n)$
- $\mathcal{F}(\hat{X}_n) = \check{\mathrm{C}}\mathrm{ech}(\hat{X}_n)$
- $\mathcal{F}(\hat{X}_n) =$ sublevelset filtration of $d(., X_\mu)$.



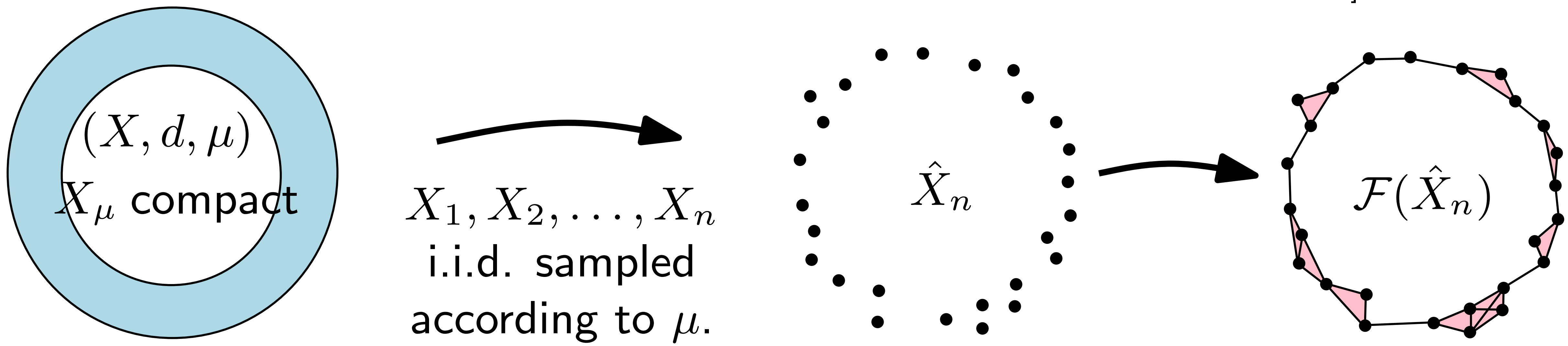
Stability thm: $d_b(\mathrm{D}(\mathcal{F}(X_\mu)), \mathrm{D}(\mathcal{F}(\hat{X}_m))) \leq 2 d_{GH}(X_\mu, \hat{X}_n)$

So, for any $\varepsilon > 0$,

$$P\left(d_b\left(\mathrm{D}(\mathcal{F}(X_\mu)), \mathrm{D}(\mathcal{F}(\hat{X}_n))\right) > \varepsilon\right) \leq P\left(d_{GH}(X_\mu, \hat{X}_n) > \frac{\varepsilon}{2}\right)$$

Deviation inequality

[*Convergence rates for persistence diagram estimation in Topological Data Analysis*, Chazal, Glisse, Labruère, Michel ICML, 2014]



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

Thm: If μ satisfies the (a, b)-standard assumption, then for any $\varepsilon > 0$:

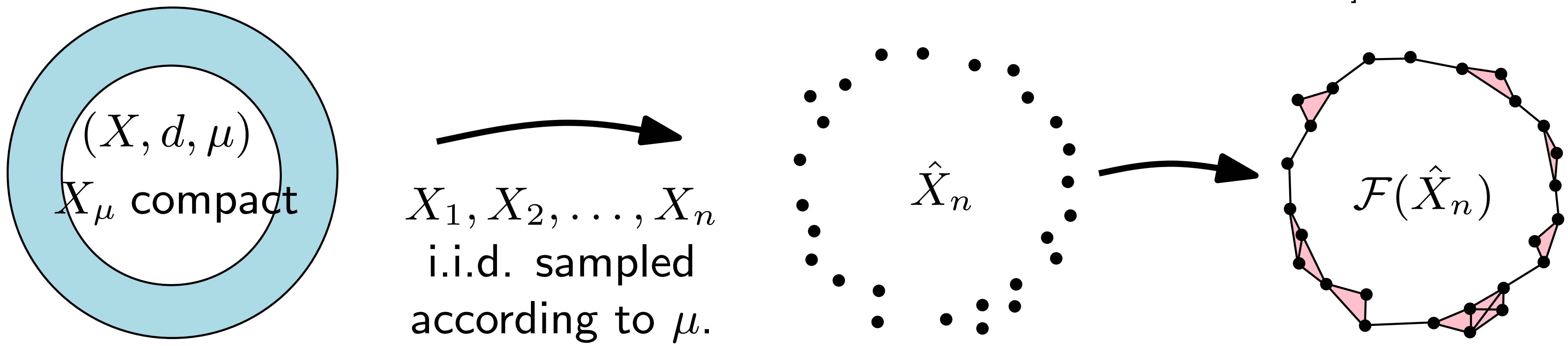
$$P\left(d_b\left(\mathrm{D}(\mathcal{F}(X_\mu)), \mathrm{D}(\mathcal{F}(\hat{X}_n))\right) > \varepsilon\right) \leq \min\left\{\frac{8^b}{a\varepsilon^b}\exp\left(-na\varepsilon^b\right), 1\right\}.$$

Moreover $\displaystyle\lim_{n\to\infty} P\left(d_b\left(\mathrm{D}(\mathcal{F}(X_\mu)), \mathrm{D}(\mathcal{F}(\hat{X}_n))\right) \leq C_1\left(\frac{\log n}{n}\right)^{1/b}\right) = 1.$

where C_1 is a constant only depending on a and b.

Deviation inequality

[*Convergence rates for persistence diagram estimation in Topological Data Analysis*, Chazal, Glisse, Labruère, Michel ICML, 2014]



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

Sketch of proof:

1. Upperbound $P\left(d_H(X_\mu, \hat{X}_n) > \frac{\varepsilon}{2}\right)$.
2. (a, b) standard assumption $\Rightarrow$ an explicit upper bound for the covering number of X_μ (by balls of radius $\varepsilon/2$).
3. Apply "union bound" argument.

$$C(\varepsilon) \leq P(\varepsilon/2)$$

$$+\ \mu(B(x, \varepsilon/2)) \geq a(\varepsilon/2)^b$$

Minimax rate of convergence

[*Convergence rates for persistence diagram estimation in Topological Data Analysis*, Chazal, Glisse, Labruère, Michel ICML, 2014]

Let $\mathcal{P}(a, b, X)$ be the set of all the probability measures on the metric space (X, d) satisfying the (a, b)-standard assumption on X:

Thm: Let $\mathcal{P}(a, b, X)$ be the set of (a, b)-standard proba measures on X. Then:

$$\sup_{\mu \in \mathcal{P}(a,b,X)} \mathbb{E}\left[d_b(\mathrm{D}(\mathcal{F}(X_\mu)), \mathrm{D}(\mathcal{F}(\hat{X}_n)))\right] \leq C \left(\frac{\log n}{n}\right)^{1/b}$$

where the constant C only depends on a and b (not on X!). Assume moreover that there exists a non isolated point x in X and let x_m be a sequence in $X \setminus \{x\}$ such that $d(x, x_n) \leq (an)^{-1/b}$. Then for any estimator $\hat{\mathrm{D}}_n$ of $\mathrm{D}(\mathcal{F}(X_\mu))$:

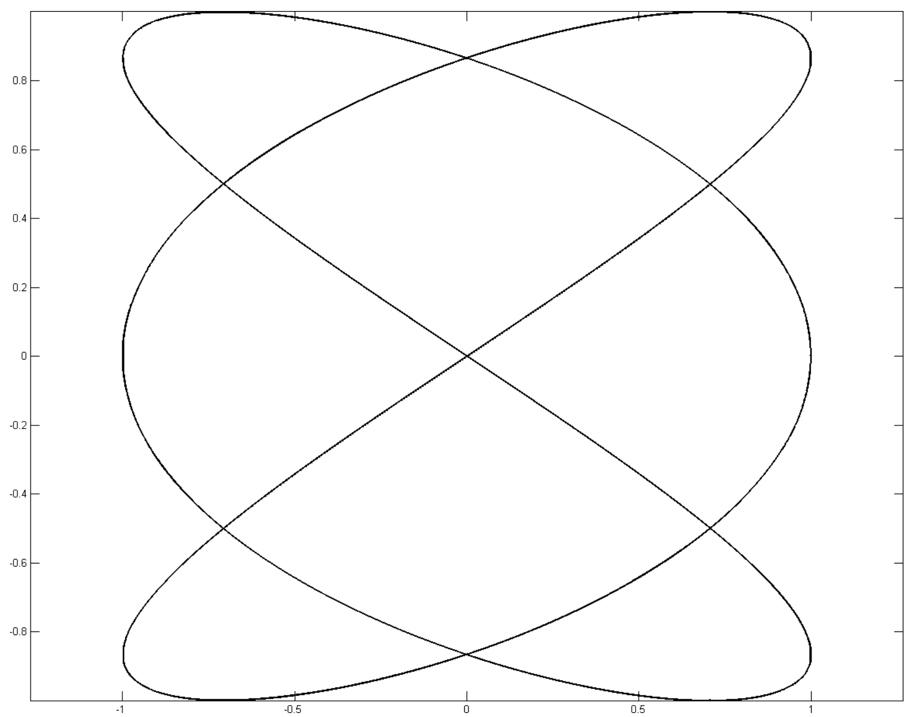
$$\liminf_{n \to \infty} d(x, x_n)^{-1} \sup_{\mu \in \mathcal{P}(a,b,X)} \mathbb{E}\left[d_b(\mathrm{D}(\mathcal{F}(X_\mu)), \hat{\mathrm{D}}_n)\right] \geq C'$$

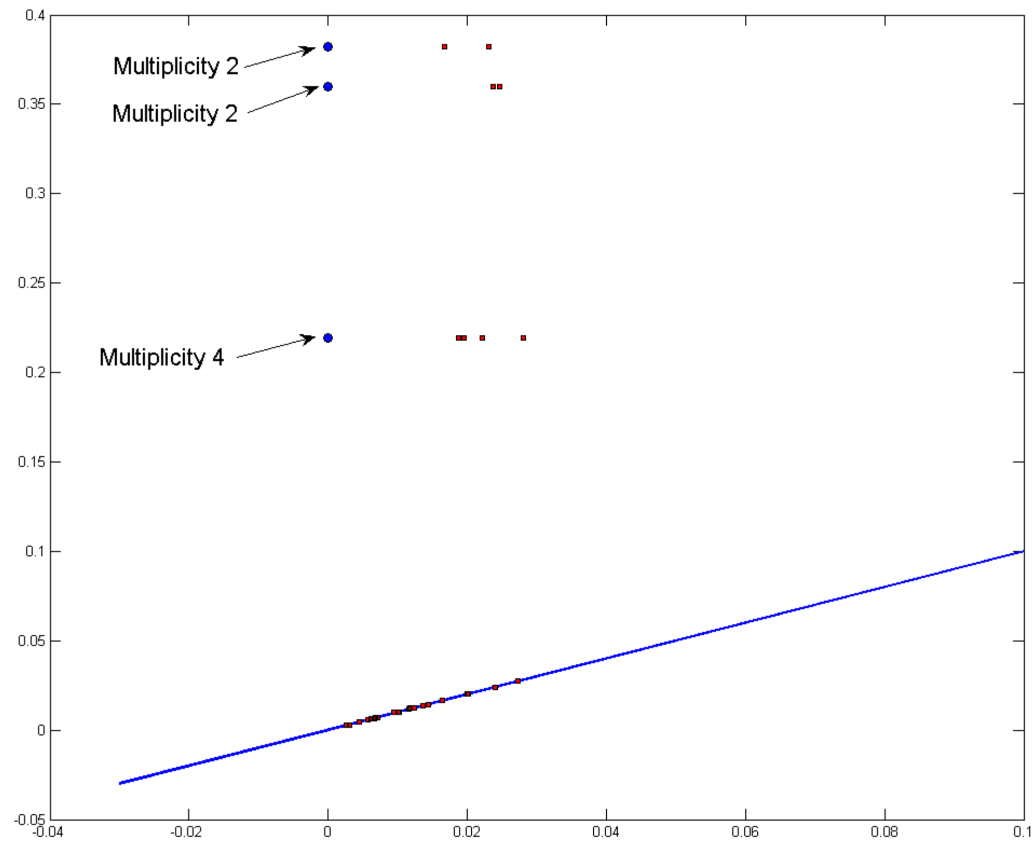
where C' is an absolute constant.

Rem: we can obtain slightly better bounds if X_μ is a submanifold of $\mathbb{R}^D$.

Numerical illustrations

[*Convergence rates for persistence diagram estimation in Topological Data Analysis*, Chazal, Glisse, Labruère, Michel ICML, 2014]

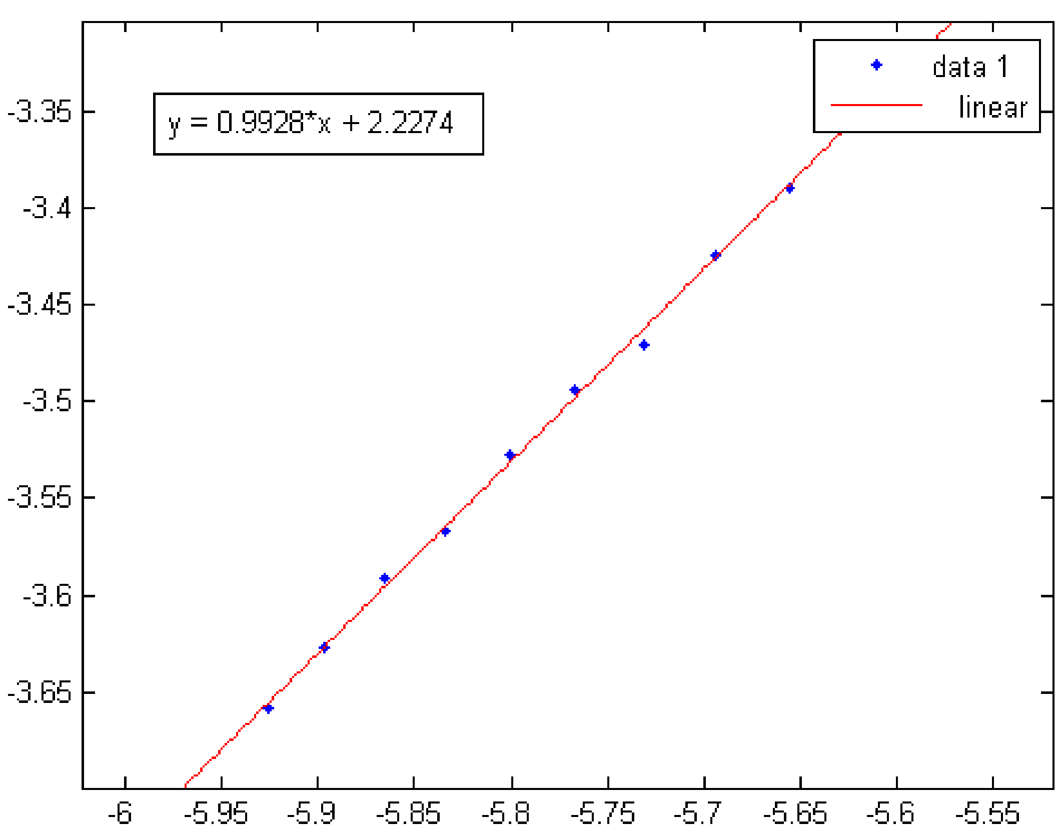




- μ: unif. measure on Lissajous curve X_μ.
- $\mathcal{F}$: distance to X_μ in $\mathbb{R}^2$.
- sample $k = 300$ sets of n points for $n = [2100 : 100 : 3000]$.
- compute

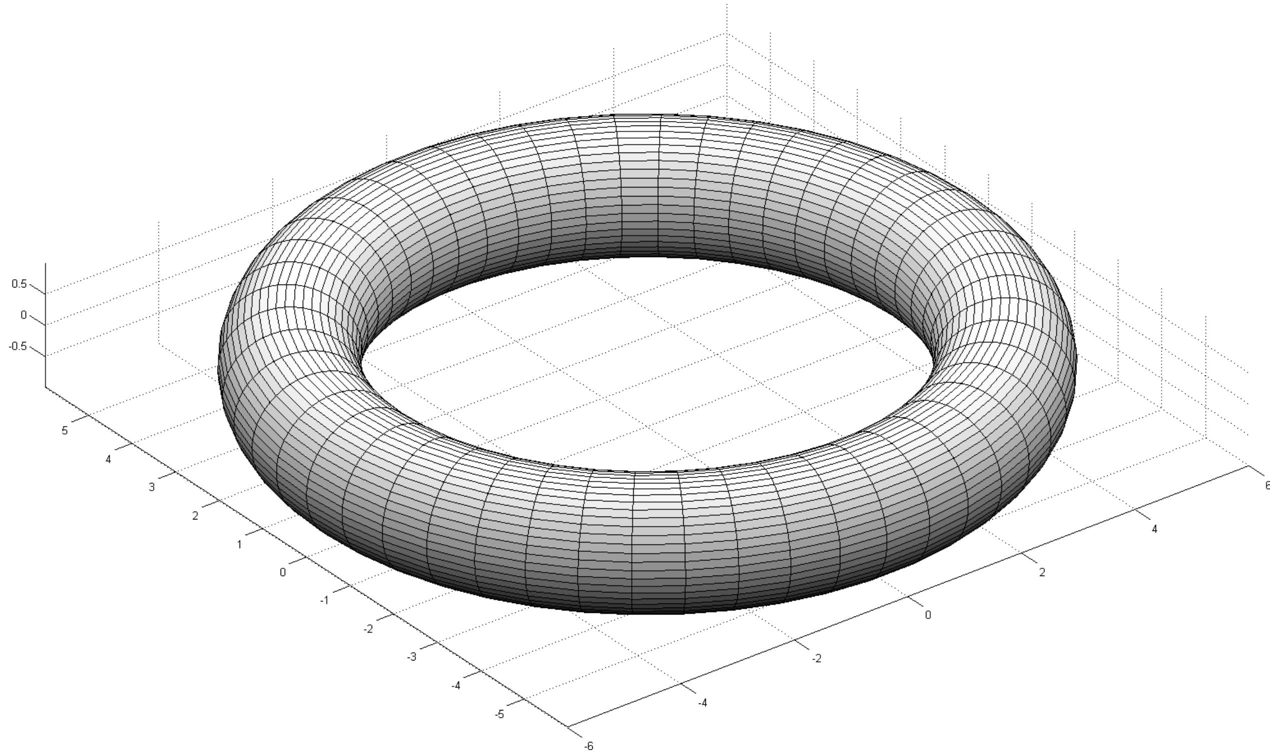
$$\hat{\mathbb{E}}_n = \hat{\mathbb{E}}[d_b(\mathrm{D}(\mathcal{F}(X_\mu)), \mathrm{D}(\mathcal{F}(\hat{X}_n)))].$$

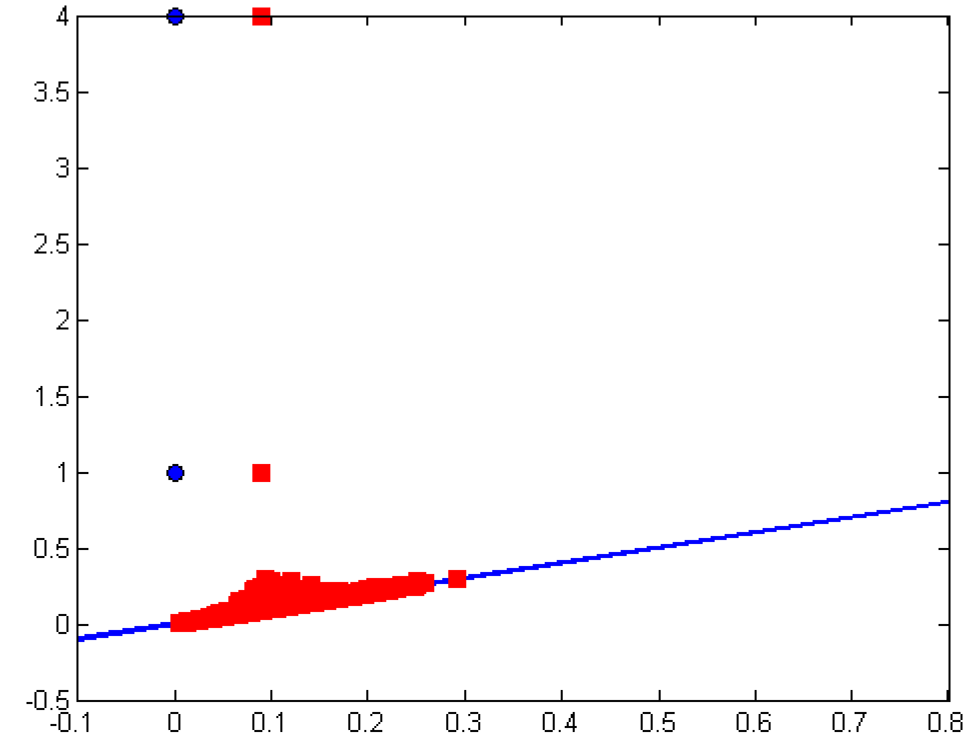
- plot $\log(\hat{\mathbb{E}}_n)$ as a function of $\log(\log(n)/n)$.



Numerical illustrations

[*Convergence rates for persistence diagram estimation in Topological Data Analysis*, Chazal, Glisse, Labruère, Michel ICML, 2014]

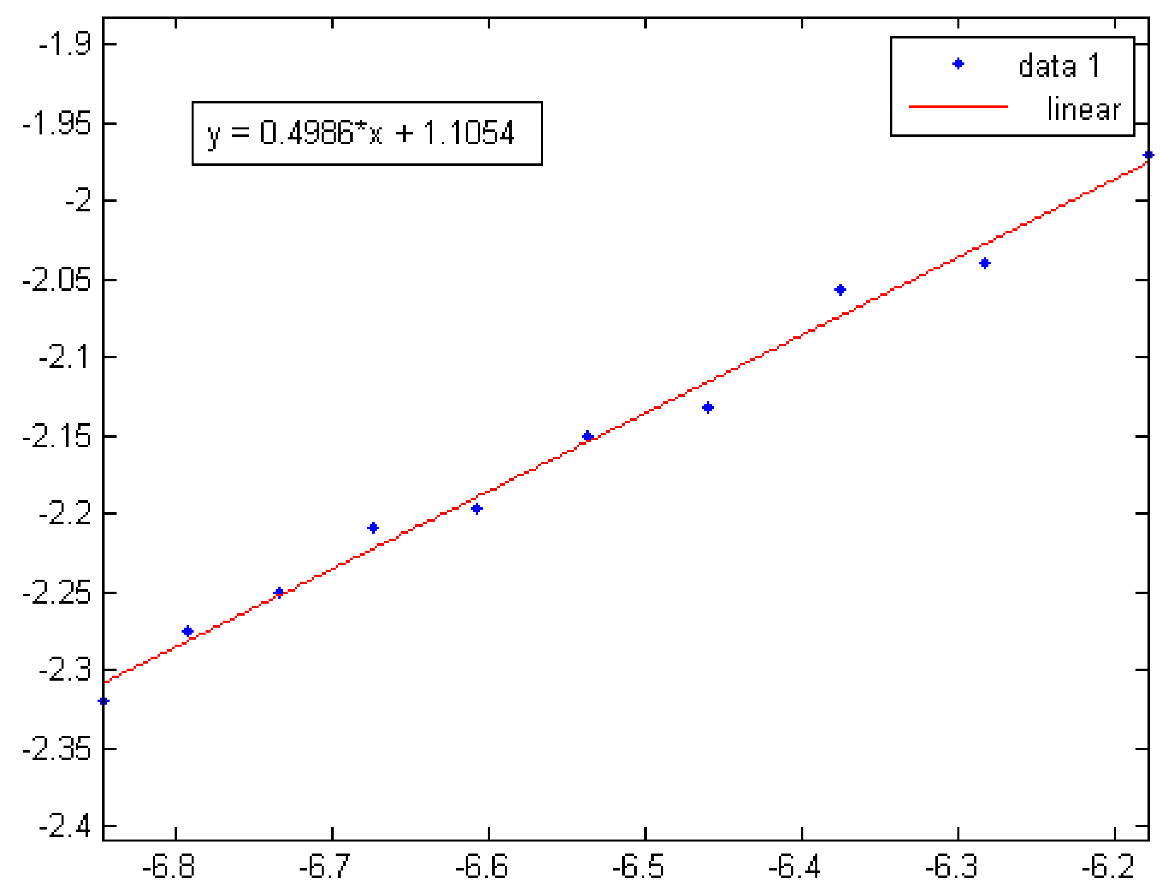




- μ: unif. measure on a torus X_μ.
- $\mathcal{F}$: distance to X_μ in $\mathbb{R}^3$.
- sample $k = 300$ sets of n points for $n = [12000 : 1000 : 21000]$.
- compute

$$\hat{\mathbb{E}}_n = \hat{\mathbb{E}}[d_b(\mathrm{D}(\mathcal{F}(X_\mu)), \mathrm{D}(\mathcal{F}(\hat{X}_n)))].$$

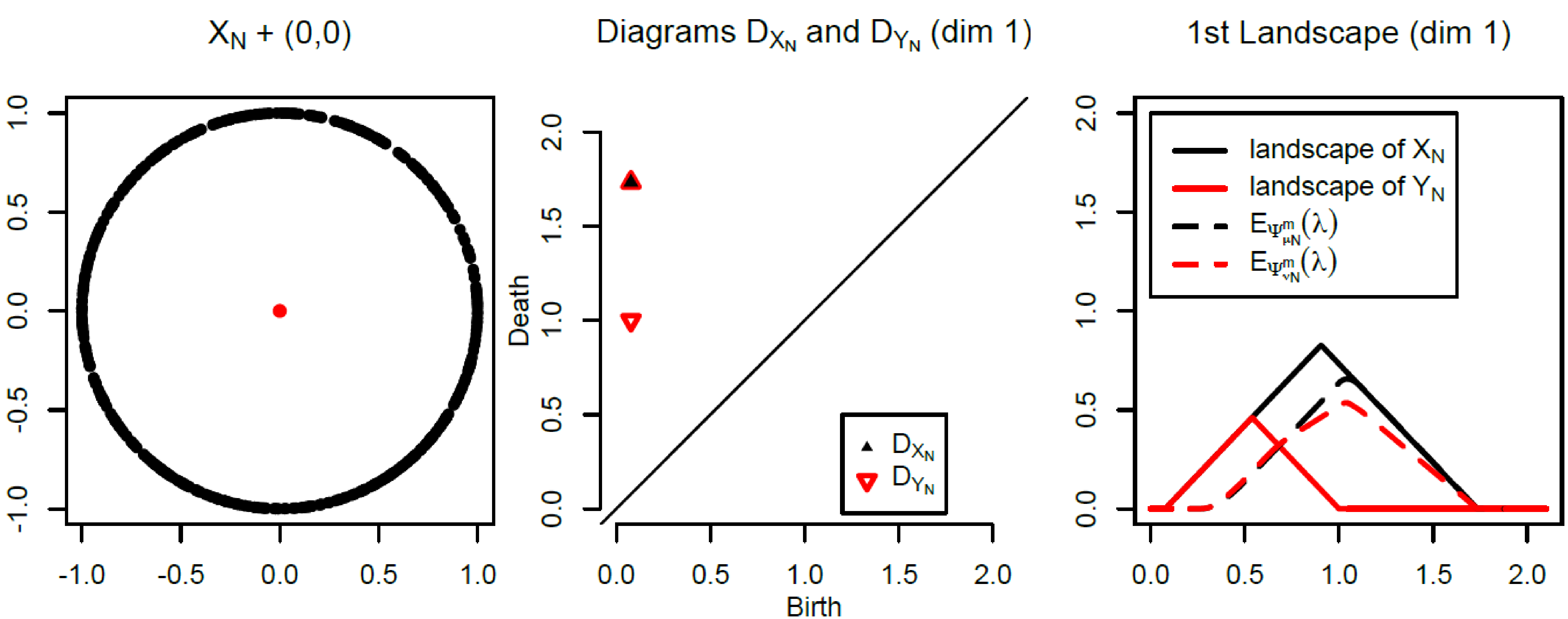
- plot $\log(\hat{\mathbb{E}}_n)$ as a function of $\log(\log(n)/n)$.



Numerical illustrations: confidence for landscapes

[*On the Bootstrap for Persistence Diagrams and Landscapes*, Chazalet al., Model. Anal. Inform. Sist., 2013]

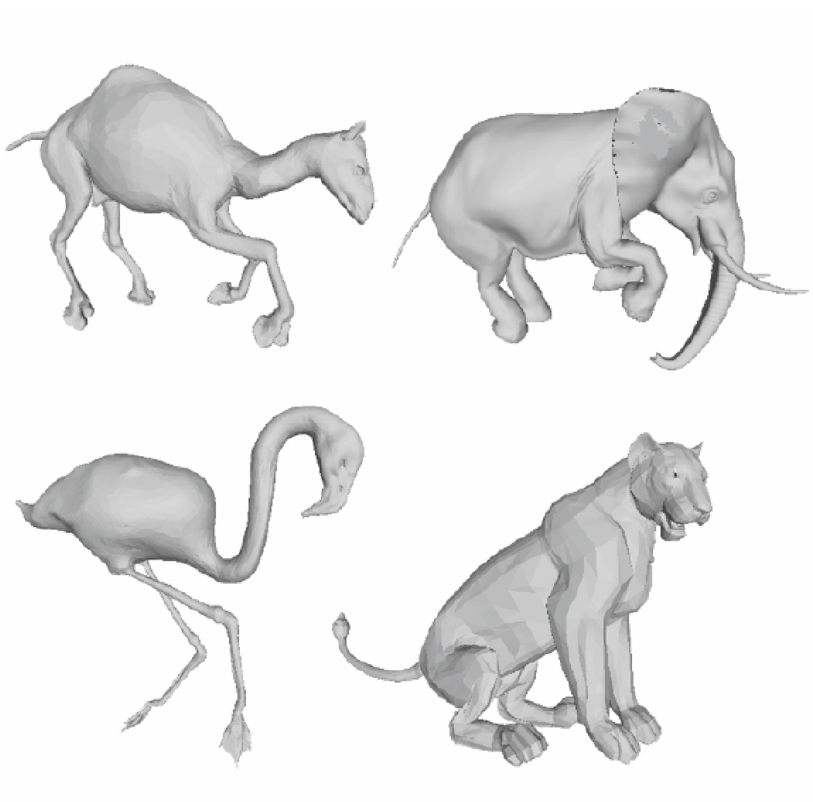
Example: Circle with one outlier.

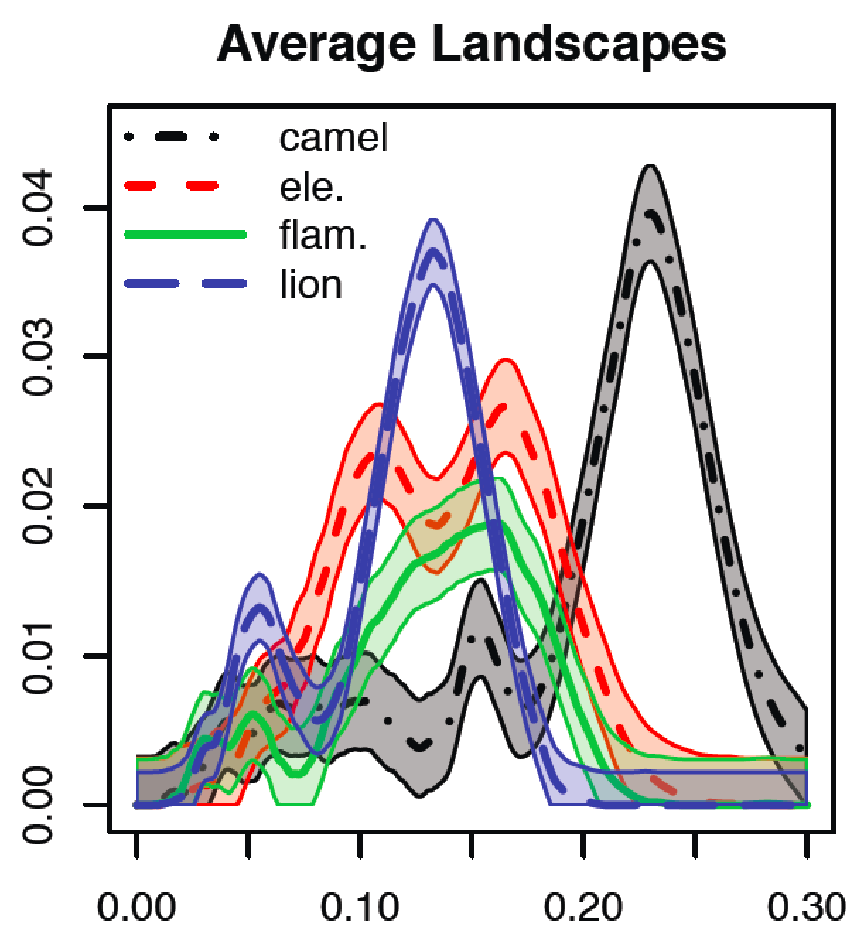


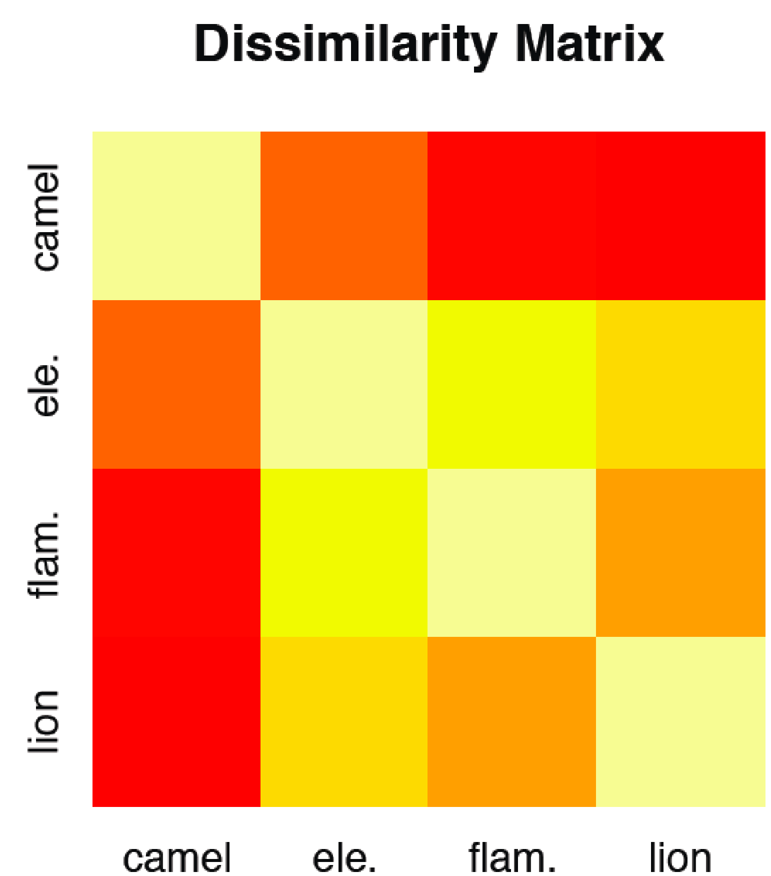
Numerical illustrations: confidence for landscapes

[*On the Bootstrap for Persistence Diagrams and Landscapes*, Chazalet al., Model. Anal. Inform. Sist., 2013]

Example: 3D shapes





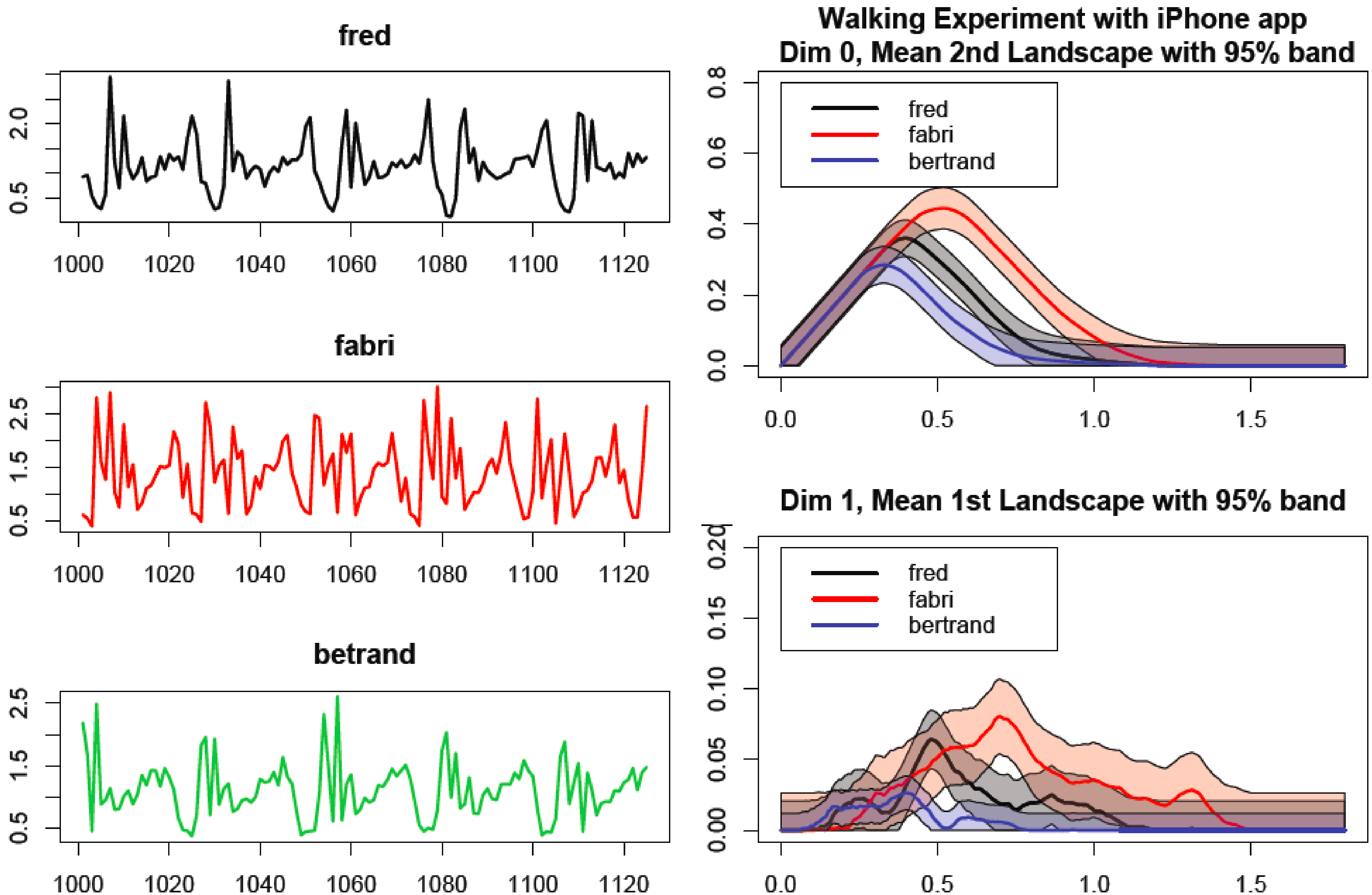


From $k = 100$ subsamples of size $n = 300$

Numerical illustrations: confidence for landscapes

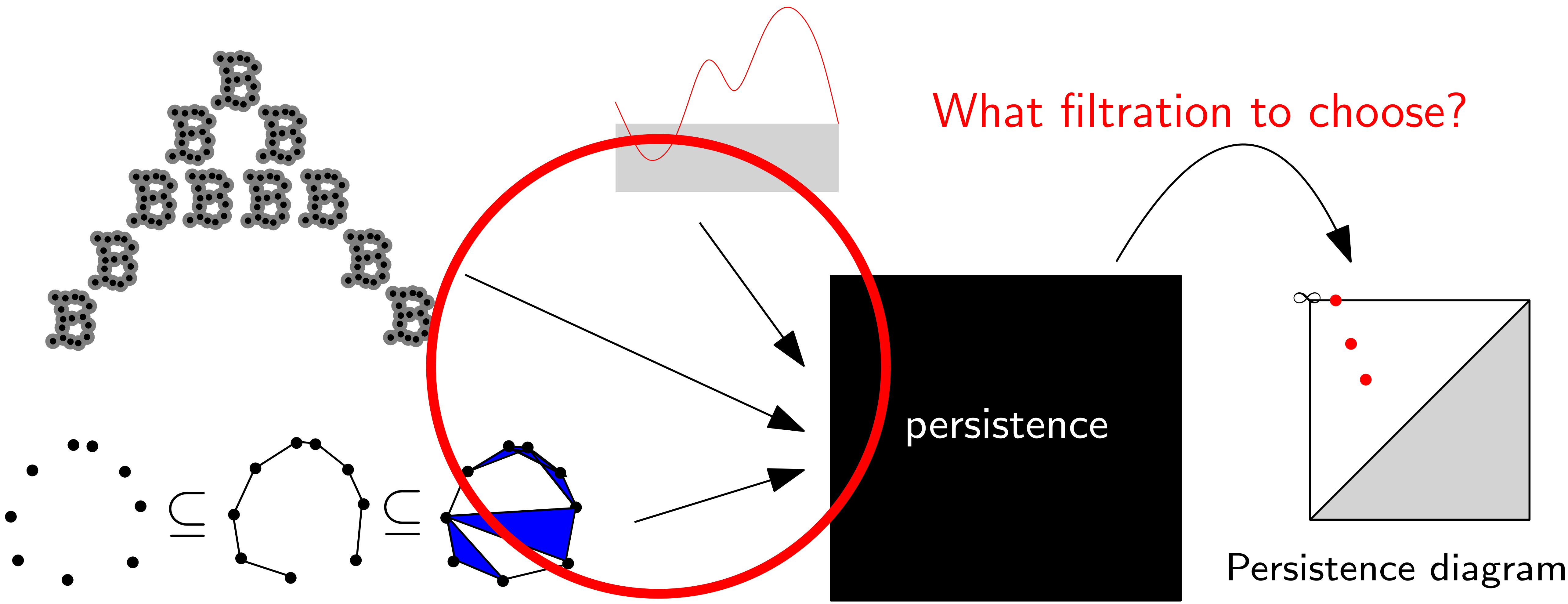
(Toy) Example: Accelerometer data from smartphone.

[*On the Bootstrap for Persistence Diagrams and Landscapes*, Chazalet al., Model. Anal. Inform. Sist., 2013]

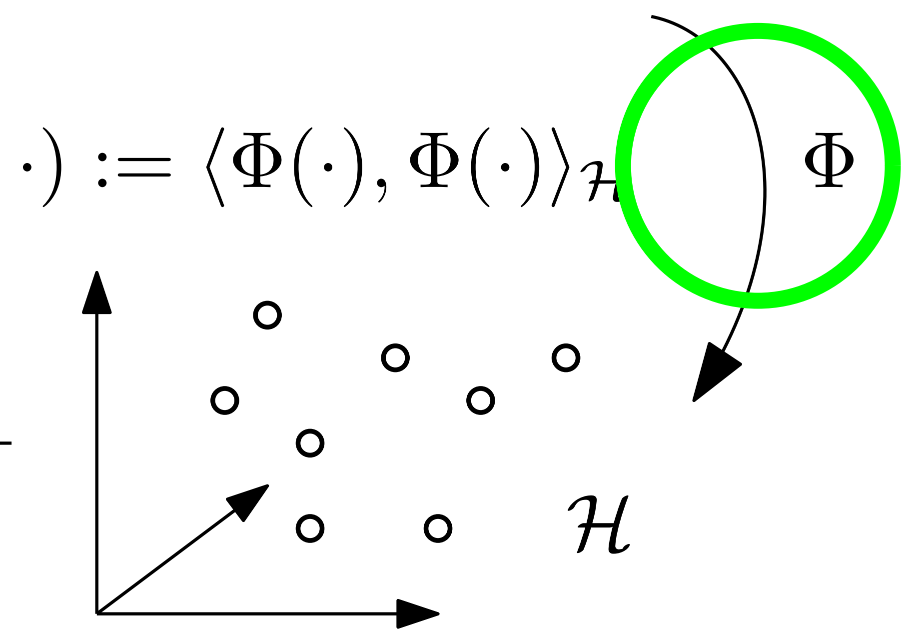


- spatial time series (accelerometer data from the smarphone of users).
- no registration/calibration preprocessing step needed to compare!

Persistence Diagrams and Machine Learning



$$k(\cdot,\cdot) := \langle \Phi(\cdot), \Phi(\cdot) \rangle_{\mathcal{H}}$$



- Classifier (RF, SVM, NN etc.)
- Dim. red. (PCA, MDS, UMAP, t-SNE)
- Clustering (DBSCAN, K-means, etc.)

Etc.

What linearization to choose?

The space of persistence diagrams

[*On the Metric Distortion of Embedding Persistence Diagrams into separable Hilbert Spaces*, Bauer, Carrière, SoCG, 2019]

Q: What happens in general when one embeds PDs in Hilbert?

Def: Two metrics d, d' are *equivalent* if

$$\exists 0 < A, B < +\infty \text{ s.t. } A\, d(\cdot,\cdot) \leq d'(\cdot,\cdot) \leq B\, d(\cdot,\cdot)$$

Prop: $\mathcal{H}$ Hilbert with dot product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and distance $\|\cdot\|_{\mathcal{H}}$. Assume $d_{\mathcal{H}}$ and d_∞ or d_p are equivalent.

(i) $\mathcal{H} = \mathbb{R}^d \Rightarrow$ **Impossible**

even if the PDs are included in $[-L, L]^2$ and have less than N points

(ii) $\mathcal{H}$ separable, $p = 1 \Rightarrow$ either $A \to 0$ or $B \to +\infty$

when $L, N \to +\infty$

Q: prove (ii).

The space of persistence diagrams

[*On the Metric Distortion of Embedding Persistence Diagrams into separable Hilbert Spaces*, Bauer, Carrière, SoCG, 2019]

Q: What happens in general when one embeds PDs in Hilbert?

Def: Two metrics d, d' are *equivalent* if

$$\exists 0 < A, B < +\infty \text{ s.t. } A\, d(\cdot,\cdot) \leq d'(\cdot,\cdot) \leq B\, d(\cdot,\cdot)$$

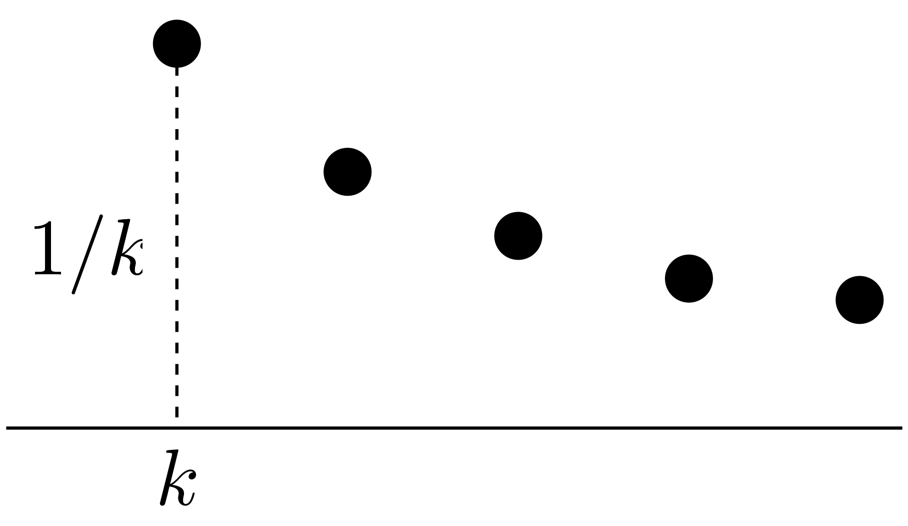
Proof:

(ii) The space of PDs with possibly infinite number of points is not separable with respect to d_1

Consider $S = \{D_u\}_{u \in \{0,1\}^{\mathbb{N}}}$

where $D_u = \left\{ \left(k, k + \frac{1}{k}\right) \,:\, u_k = 1 \right\}$

S is not countable with d_1



The space of persistence diagrams

[*On the Metric Distortion of Embedding Persistence Diagrams into separable Hilbert Spaces*, Bauer, Carrière, SoCG, 2019]

Q: What happens in general when one embeds PDs in Hilbert?

Def: Two metrics d, d' are *equivalent* if

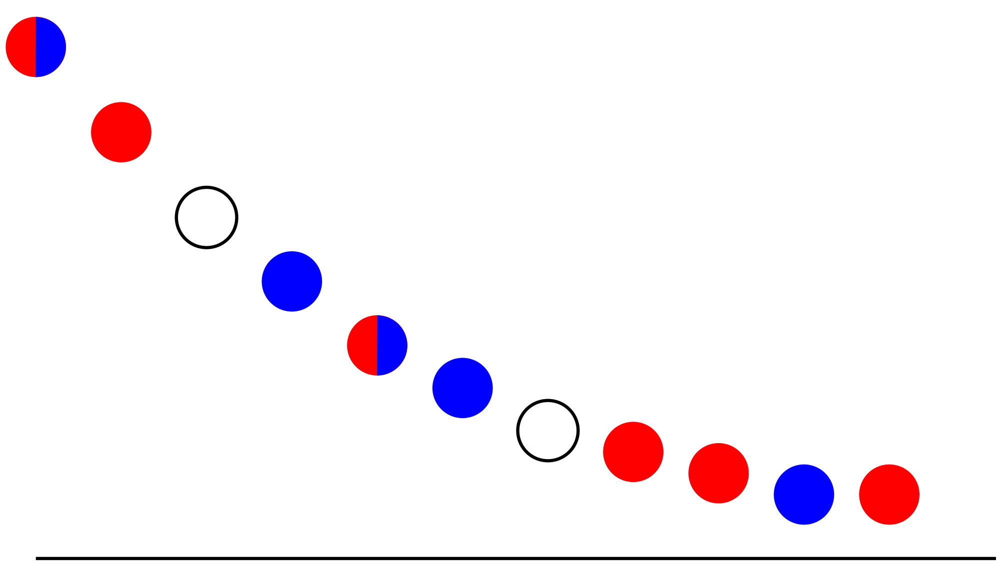
$$\exists 0 < A, B < +\infty \text{ s.t. } A\, d(\cdot,\cdot) \leq d'(\cdot,\cdot) \leq B\, d(\cdot,\cdot)$$

Proof:

$$S = \{D_u\}_{u \in \{0,1\}^{\mathbb{N}}}$$

Indeed, let $S' \subseteq S$ be a dense set and $\epsilon > 0$

$$\forall D_u \in S,\ \exists D_{u'} \in S' : \mathrm{d}_1(D_u, D_{u'}) \leq \epsilon$$



Supports of u' and u must differ on a finite number of terms only

$\Rightarrow$ card$(S') \geq \boxed{\text{card}(S/\sim)}$ uncountable!

where $D_u \sim D_v \Leftrightarrow \text{supp}(u) \triangle \text{supp}(v) < \infty$

The space of persistence diagrams

[*On the Metric Distortion of Embedding Persistence Diagrams into separable Hilbert Spaces*, Bauer, Carrière, SoCG, 2019]

Q: What happens in general when one embeds PDs in Hilbert?

Def: Two metrics d, d' are *equivalent* if

$$\exists 0 < A, B < +\infty \text{ s.t. } A\, d(\cdot,\cdot) \leq d'(\cdot,\cdot) \leq B\, d(\cdot,\cdot)$$

Ex: Persistence surface

$$\Phi(\mathrm{D}) = \sum_{p\in\mathrm{D}} w(p)\cdot\exp\left(-\frac{\|\cdot - p\|_2^2}{2\sigma^2}\right)$$

where $w((x,y)) = \arctan\left(C|y-x|^{\alpha}\right)$ with $C, \alpha > 0$

If $\alpha \geq 2$, S is in the domain of Φ

The space of persistence diagrams

[*On the Metric Distortion of Embedding Persistence Diagrams into separable Hilbert Spaces*, Bauer, Carrière, SoCG, 2019]

Q: What happens in general when one embeds PDs in Hilbert?

Def: Two metrics d, d' are *equivalent* if

$$\exists 0 < A, B < +\infty \text{ s.t. } A\, d(\cdot,\cdot) \leq d'(\cdot,\cdot) \leq B\, d(\cdot,\cdot)$$

Proof:

(i) is a little more tricky

Def: Let (X, d) be a metric space. Given a subset $E \subset X$ and $r > 0$, let $N_r(E)$ be the least number of open balls of radius $\leq r$ that can cover E. The *Assouad dimension* of (X, d) is:

$$\dim_A(X,d) = \inf\{\alpha : \exists C \text{ s.t. } \sup_x N_{\beta r}(B(x,r)) \leq C\beta^{-\alpha},\ 0 < \beta \leq 1\}$$

$\dim_A$ is preserved for equivalent metrics

$\dim_A(\mathcal{D}, d_p) = +\infty$ whereas $\dim_A(\mathbb{R}^d) = d$

The space of persistence diagrams

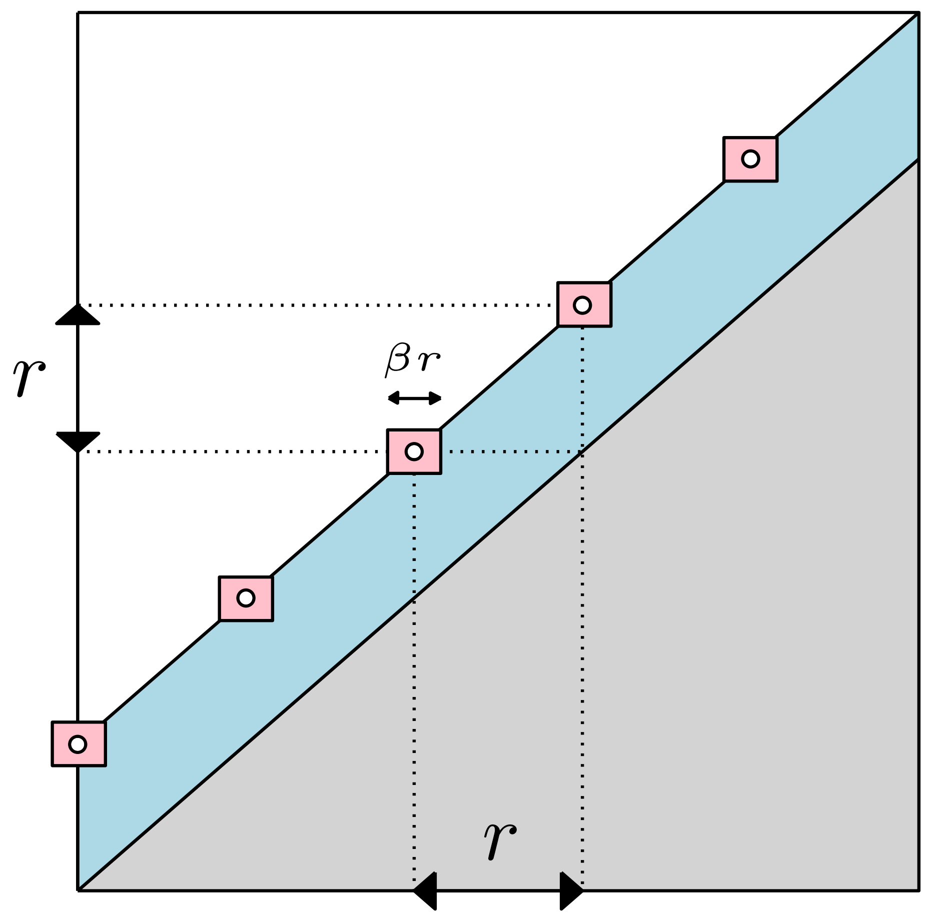
[*On the Metric Distortion of Embedding Persistence Diagrams into separable Hilbert Spaces*, Bauer, Carrière, SoCG, 2019]

Q: What happens in general when one embeds PDs in Hilbert?

Def: Two metrics d, d' are *equivalent* if

$$\exists 0 < A, B < +\infty \text{ s.t. } A\, d(\cdot,\cdot) \leq d'(\cdot,\cdot) \leq B\, d(\cdot,\cdot)$$

Proof:



Idea: Consider the ball of radius r around the empty diagram and diagrams with single points at distance r from Δ and from each other

The number of such diagrams increases to $+\infty$ as β goes to 0

$\dim_A$ is preserved for equivalent metrics

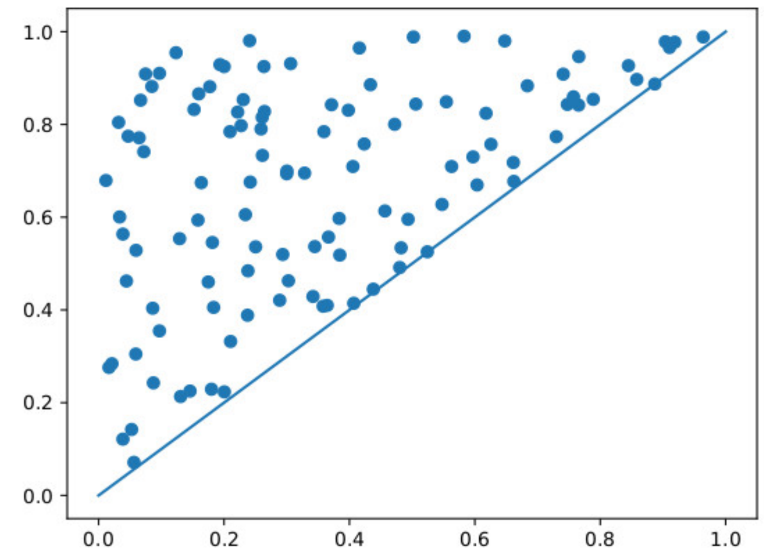
$\dim_A(\mathcal{D}, d_p) = +\infty$ whereas $\dim_A(\mathbb{R}^d) = d$

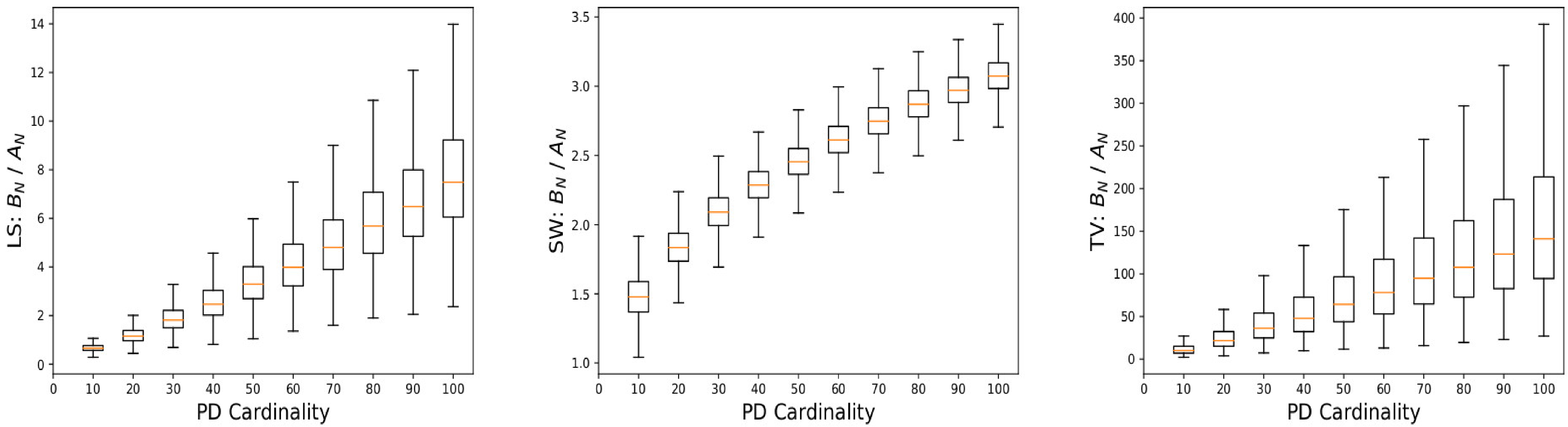
The space of persistence diagrams

[*On the Metric Distortion of Embedding Persistence Diagrams into separable Hilbert Spaces*, Bauer, Carrière, SoCG, 2019]

Illustrations:

We generate diagrams by uniformly sampling into the upper unit half-square



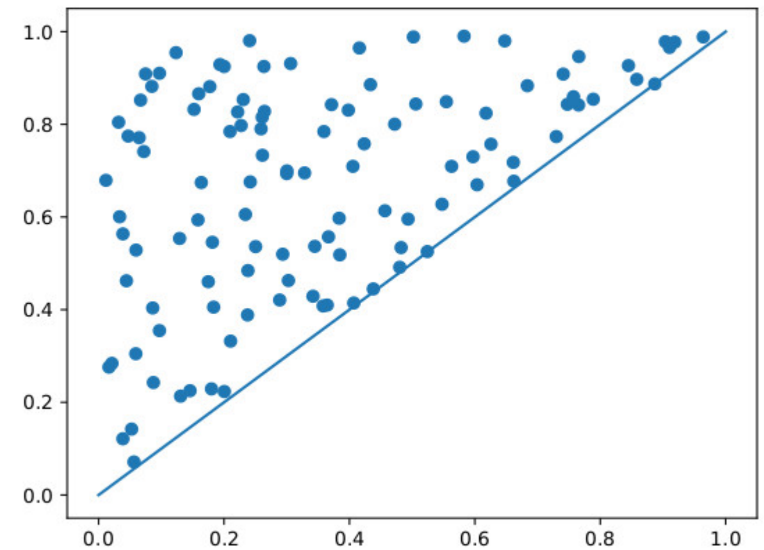


The space of persistence diagrams

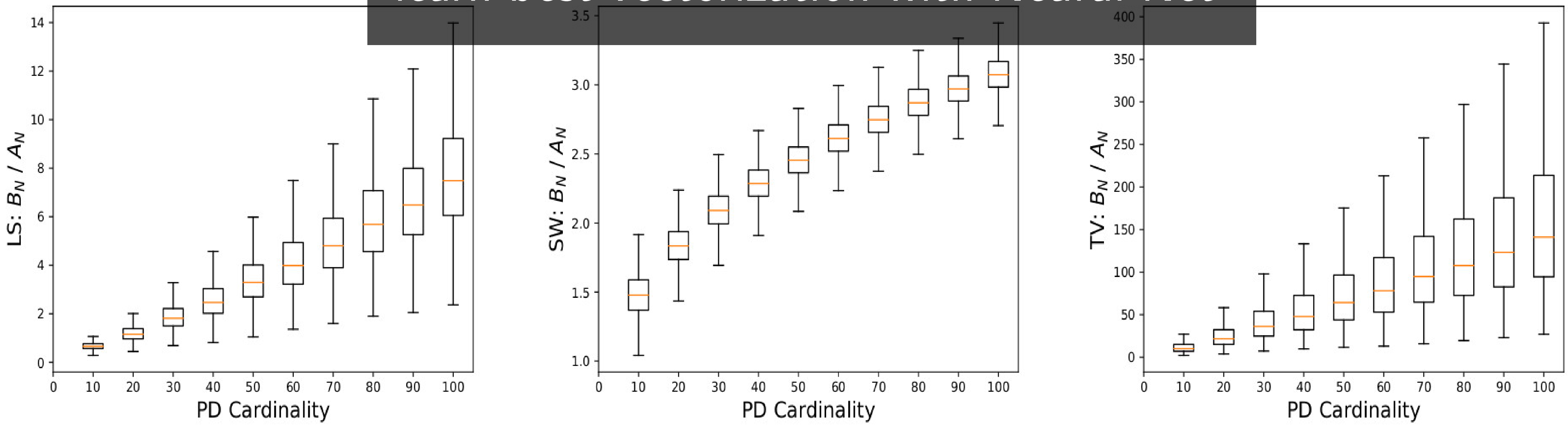
[*On the Metric Distortion of Embedding Persistence Diagrams into separable Hilbert Spaces*, Bauer, Carrière, SoCG, 2019]

Illustrations:

We generate diagrams by uniformly sampling into the upper unit half-square



Idea: Stay in Euclidean space $\mathbb{R}^d$ but *learn* best vectorization with Neural Net



The Deep Set architecture

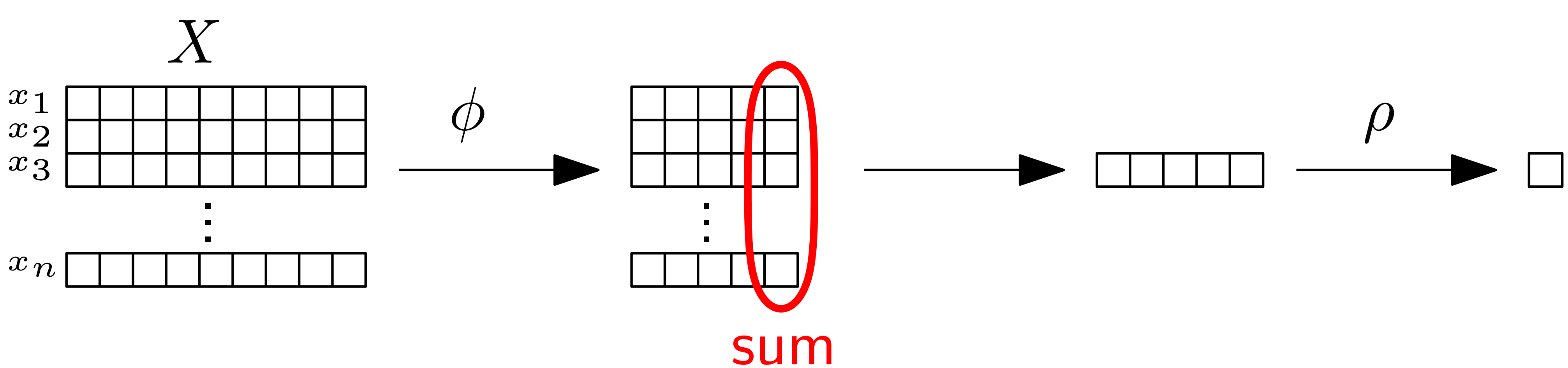
[*Deep Sets*, Zaheer, Kottur, Ravanbakhsh, Poczos, Salakhutdinov, Smola, NeurIPS, 2017]

Deep Set is a novel neural net architecture that is able to handle sets instead of finite dimensional vectors

Input: $\{x_1, ..., x_n\} \subset \mathbb{R}^d$ instead of $x \in \mathbb{R}^d$

Network is *permutation invariant*: $F(X) = \rho\left(\sum_i \phi(x_i)\right)$

$$\Rightarrow F(\{x_1, ..., x_n\}) = F(\{x_{\sigma(1)}, ..., x_{\sigma(n)}\}),\ \forall \sigma$$



In practice: $\phi(x_i) = W \cdot x_i + b$

The Deep Set architecture

[*Deep Sets*, Zaheer, Kottur, Ravanbakhsh, Poczos, Salakhutdinov, Smola, NeurIPS, 2017]

Deep Set is a novel neural net architecture that is able to handle sets instead of finite dimensional vectors

Input: $\{x_1, ..., x_n\} \subset \mathbb{R}^d$ instead of $x \in \mathbb{R}^d$

Network is *permutation invariant*: $F(X) = \rho\left(\sum_i \phi(x_i)\right)$

Universality theorem

Thm:

A function f is permutation invariant iif $f(X) = \rho\left(\sum_i \phi(x_i)\right)$ for some ρ and ϕ, whenever X is included in a *countable* space

Application to PDs

Application to PDs

[*PersLay: A Neural Network Layer for Persistence Diagrams and New Graph Topological Signatures*, Carrière, Chazal, Ike, Lacombe, Royer, Umeda, AISTATS, 2019]

Permutation invariant layers generalize several TDA approaches

$\rightarrow$ persistence images $\rightarrow$ landscapes $\rightarrow$ Betti curves

[*Time Series Classification via Topological Data Analysis*, Umeda, Trans. Jap. Soc. for AI, 2017]

But not all of them since $\mathbb{R}^2$ is not countable

Using any permutation invariant operation (such as max, min, kth largest value) allows to generalize other TDA approaches

$$\boxed{\mathrm{PersLay}(\mathrm{D}) = \rho\left(\mathsf{op}\{w(p)\cdot\phi(p)\}_{p\in\mathrm{D}}\right)}$$

Permutation-invariant operation

Weight function

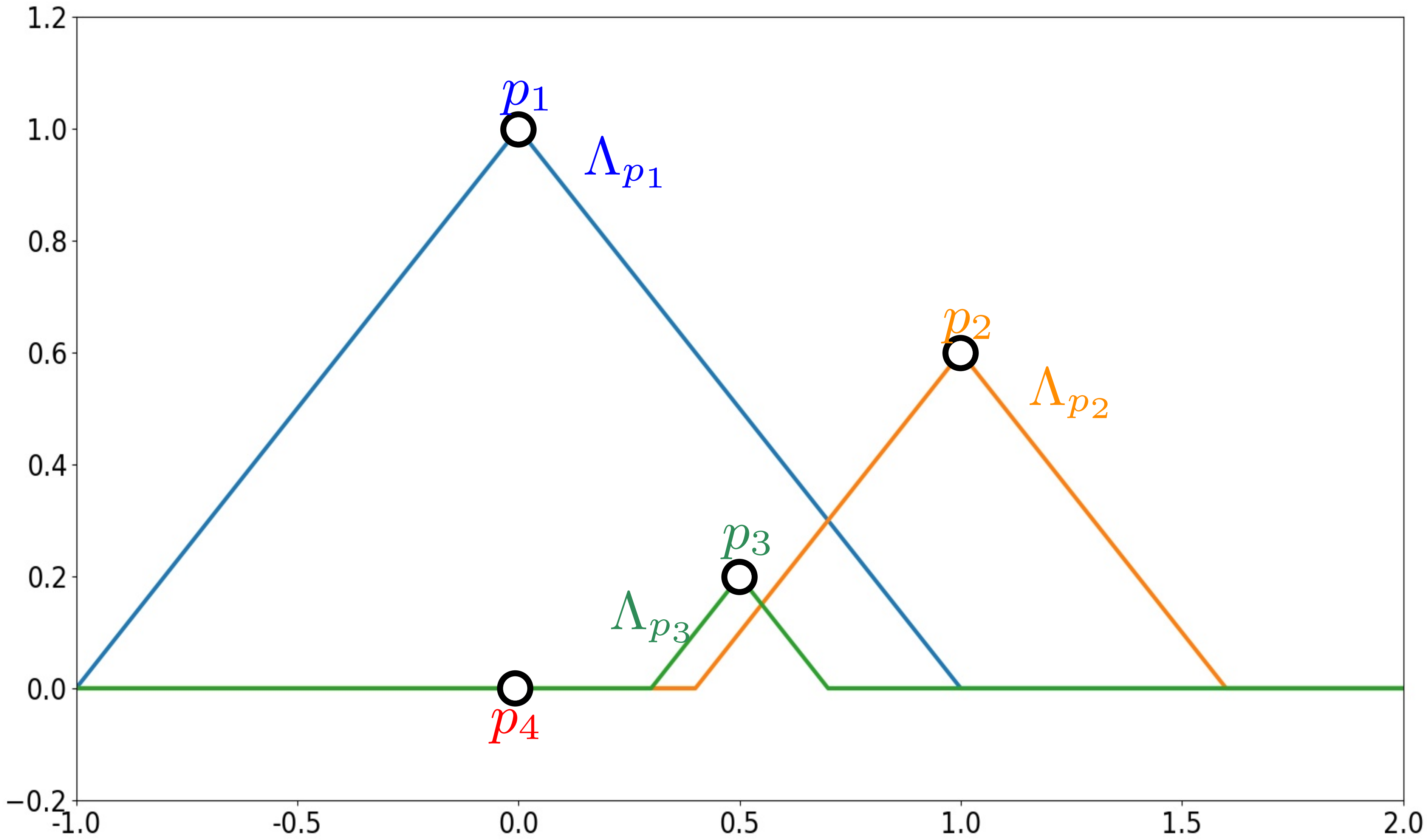
Point transformation

Application to PDs

[*PersLay: A Neural Network Layer for Persistence Diagrams and New Graph Topological Signatures*, Carrière, Chazal, Ike, Lacombe, Royer, Umeda, AISTATS, 2019]

Parameters $t_1, \cdots, t_q \in \mathbb{R}$

$$w(p) = 1 \qquad \phi_\Lambda : p \mapsto \begin{bmatrix} \Lambda_p(t_1) \\ \Lambda_p(t_2) \\ \vdots \\ \Lambda_p(t_q) \end{bmatrix} \qquad \text{op} = \text{top-}k$$

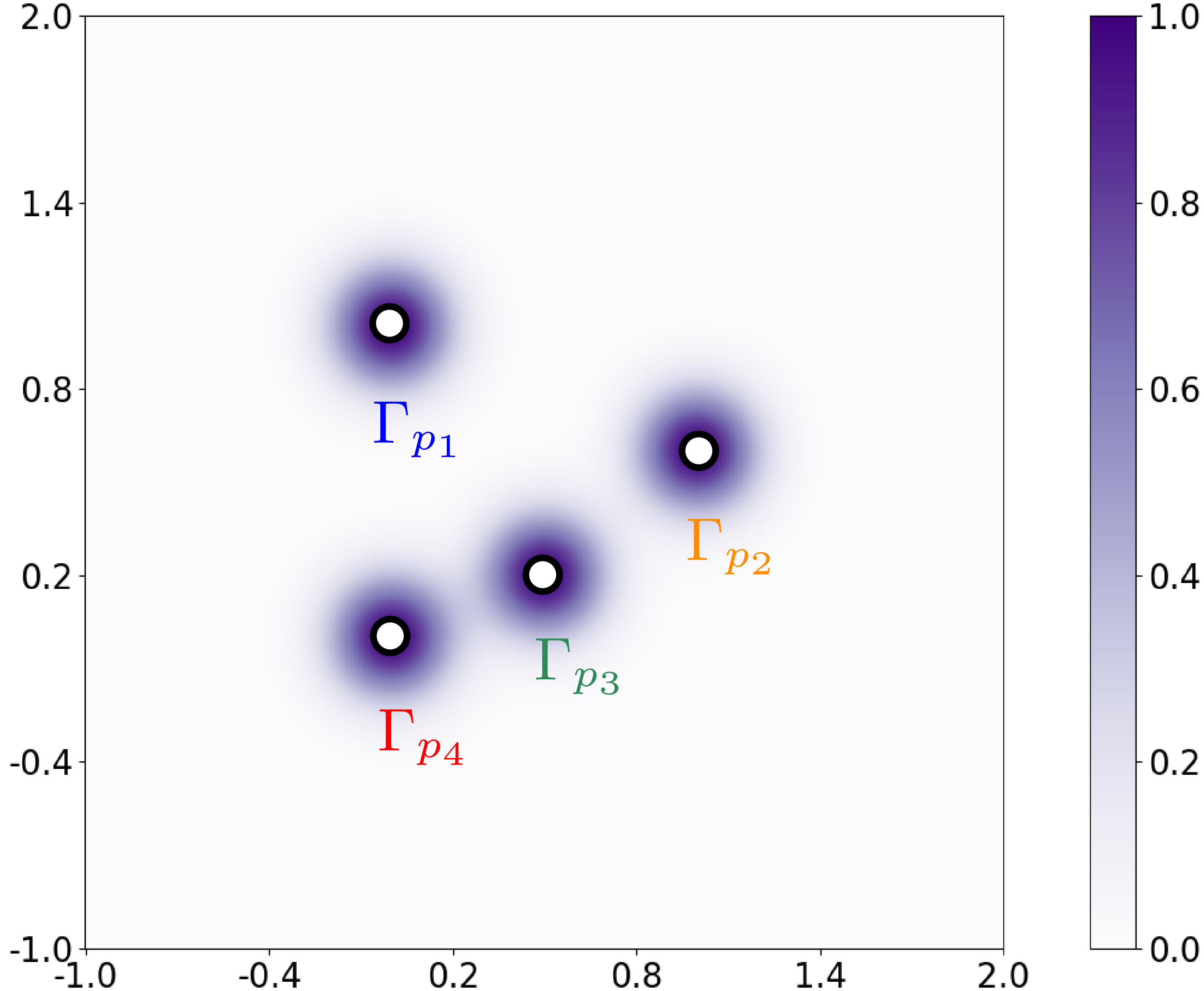


Application to PDs

[*PersLay: A Neural Network Layer for Persistence Diagrams and New Graph Topological Signatures*, Carrière, Chazal, Ike, Lacombe, Royer, Umeda, AISTATS, 2019]

Parameters $t_1, \cdots, t_q \in \mathbb{R}^2$

$$w(p) = w_t((x, y)) \qquad \phi_\Gamma : p \mapsto \begin{bmatrix} \Gamma_p(t_1) \\ \Gamma_p(t_2) \\ \vdots \\ \Gamma_p(t_q) \end{bmatrix} \qquad \text{op} = \text{sum}$$



Application to PDs

[*PersLay: A Neural Network Layer for Persistence Diagrams and New Graph Topological Signatures*, Carrière, Chazal, Ike, Lacombe, Royer, Umeda, AISTATS, 2019]

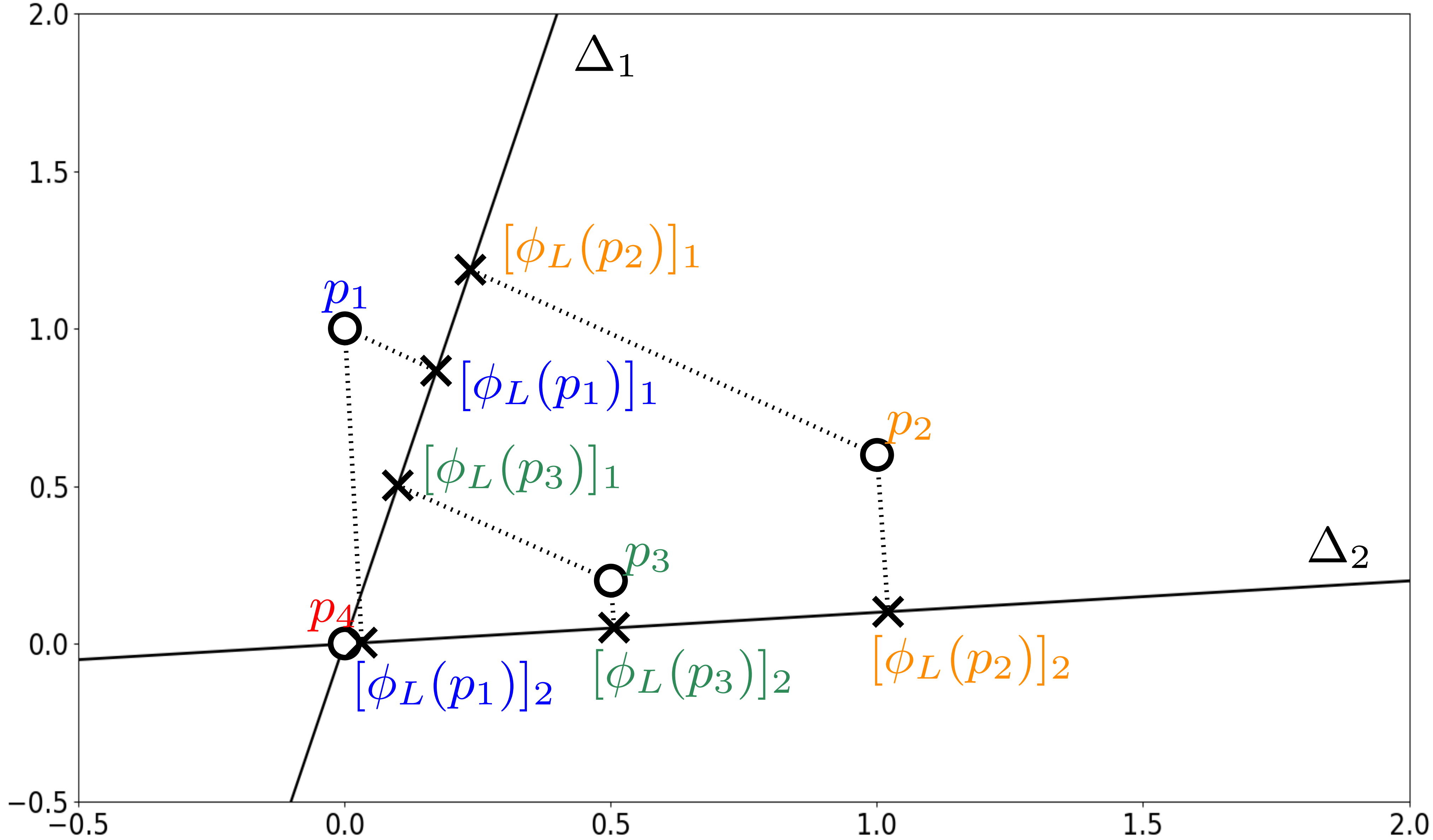
Parameters $\Delta_1, \cdots, \Delta_q \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$

$b_{\Delta_1}, \cdots, b_{\Delta_q} \in \mathbb{R}$

$$\phi_L : p \mapsto \begin{bmatrix} \langle p, e_{\Delta_1} \rangle + b_{\Delta_1} \\ \langle p, e_{\Delta_2} \rangle + b_{\Delta_2} \\ \vdots \\ \langle p, e_{\Delta_q} \rangle + b_{\Delta_q} \end{bmatrix}$$

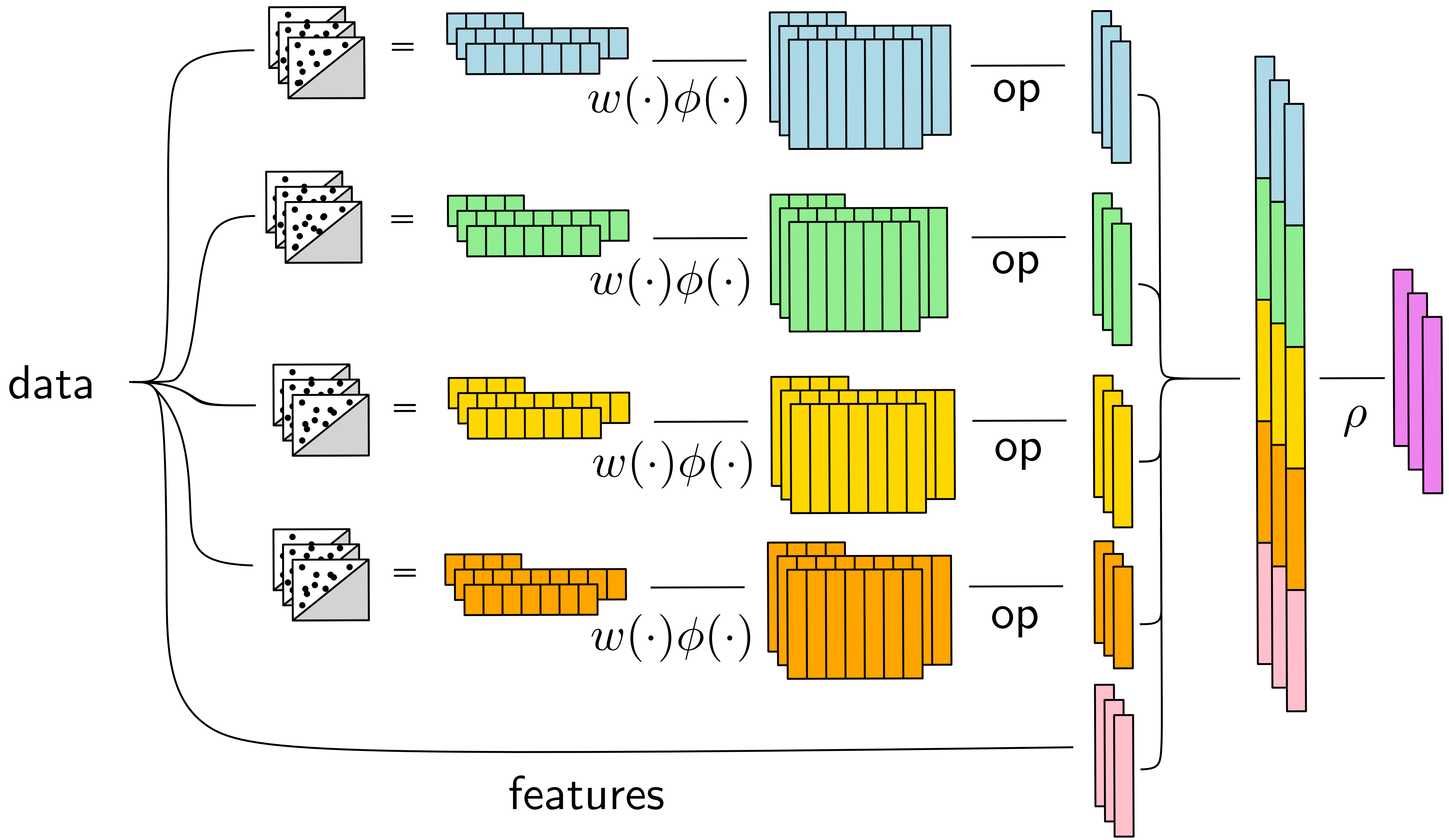
$w(p) = 1$

op $=$ top-k



Application to PDs

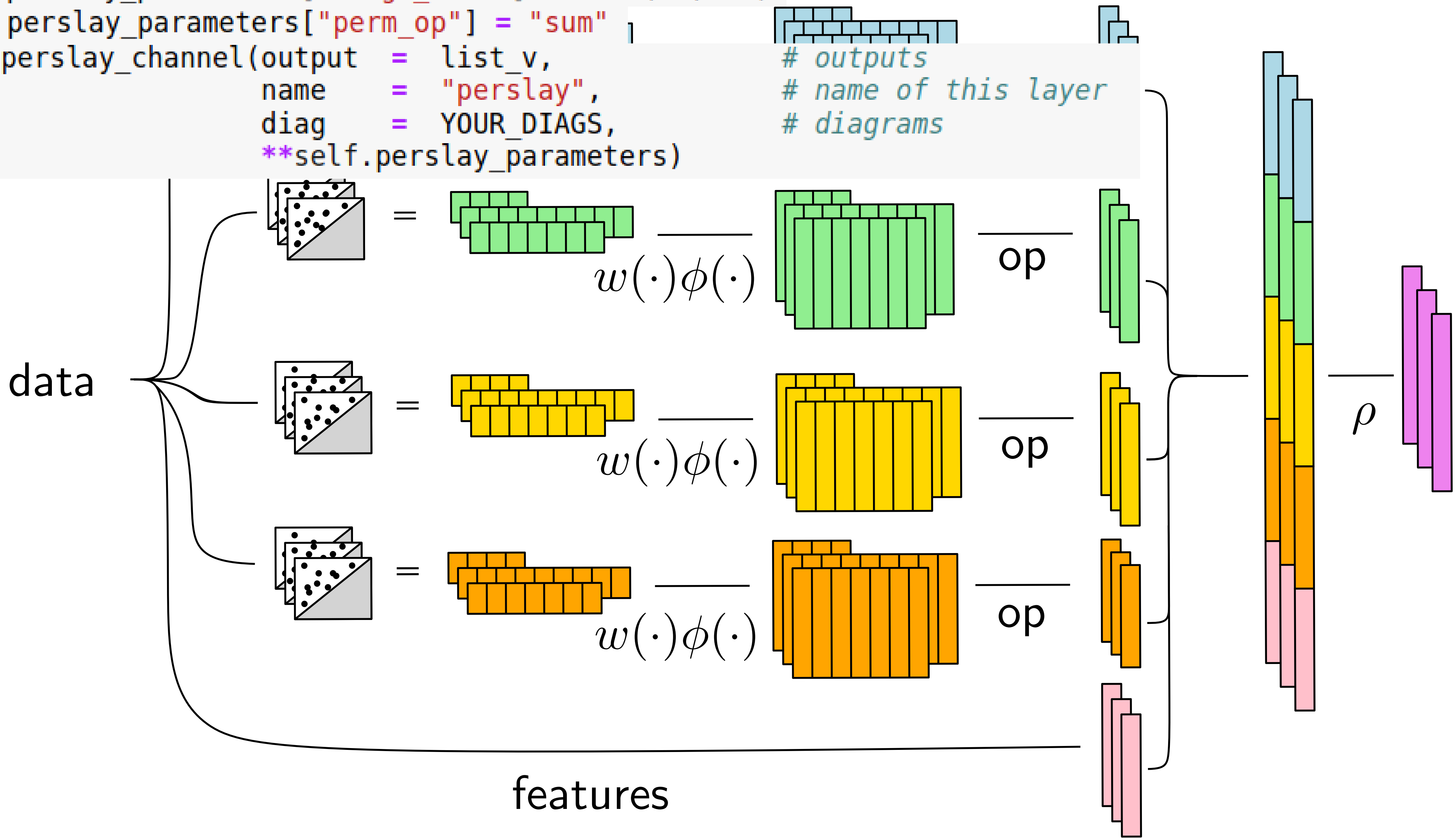
[*PersLay: A Neural Network Layer for Persistence Diagrams and New Graph Topological Signatures*, Carrière, Chazal, Ike, Lacombe, Royer, Umeda, AISTATS, 2019]



Application to PDs

[*PersLay: A Neural Network Layer for Persistence Diagrams and New Graph Topological Signatures*, Carrière, Chazal, Ike, Lacombe, Royer, Umeda, AISTATS, 2019]

```python
from perslay.perslay import perslay_channel
perslay_parameters["layer"]          = "im"
perslay_parameters["image_size"]     = (20, 20)
perslay_parameters["perm_op"] = "sum"
perslay_channel(output  =  list_v,                # outputs
                name    =  "perslay",             # name of this layer
                diag    =  YOUR_DIAGS,            # diagrams
                **self.perslay_parameters)
```



Application to graph classification

[*PersLay: A Neural Network Layer for Persistence Diagrams and New Graph Topological Signatures*, Carrière, Chazal, Ike, Lacombe, Royer, Umeda, AISTATS, 2019]

Let $G = (V, E)$ be a graph, A its adjacency matrix

D its degree matrix

and $L_w(G) = I - D^{-1/2} A D^{-1/2}$ its normalized Laplacian.

$L_w(G)$ decomposes on a orthonormal basis $\phi_1 \dots \phi_n$

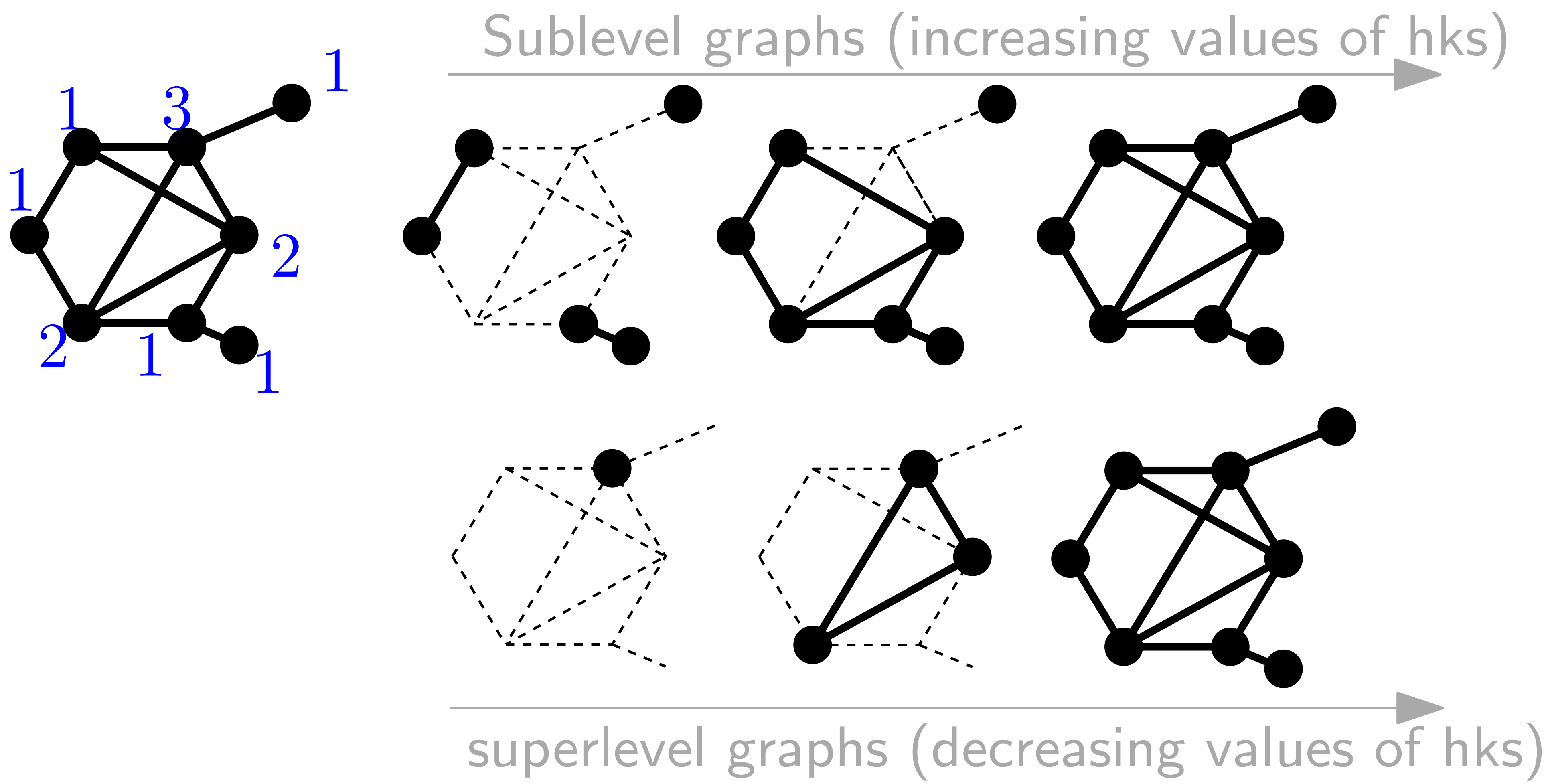
with eigenvalues $0 \leq \lambda_1 \leq \cdots \leq \lambda_n \leq 2$

Def: Let $t \geq 0$, and define the *Heat Kernel Signature* of param t:

$$\mathrm{hks}_{G,t} : v \mapsto \sum_{k=1}^{n} \exp(-\lambda_k t) \phi_k(v)^2$$

Application to graph classification

[*PersLay: A Neural Network Layer for Persistence Diagrams and New Graph Topological Signatures*, Carrière, Chazal, Ike, Lacombe, Royer, Umeda, AISTATS, 2019]



Def: Let $t \geq 0$, and define the *Heat Kernel Signature* of param t:

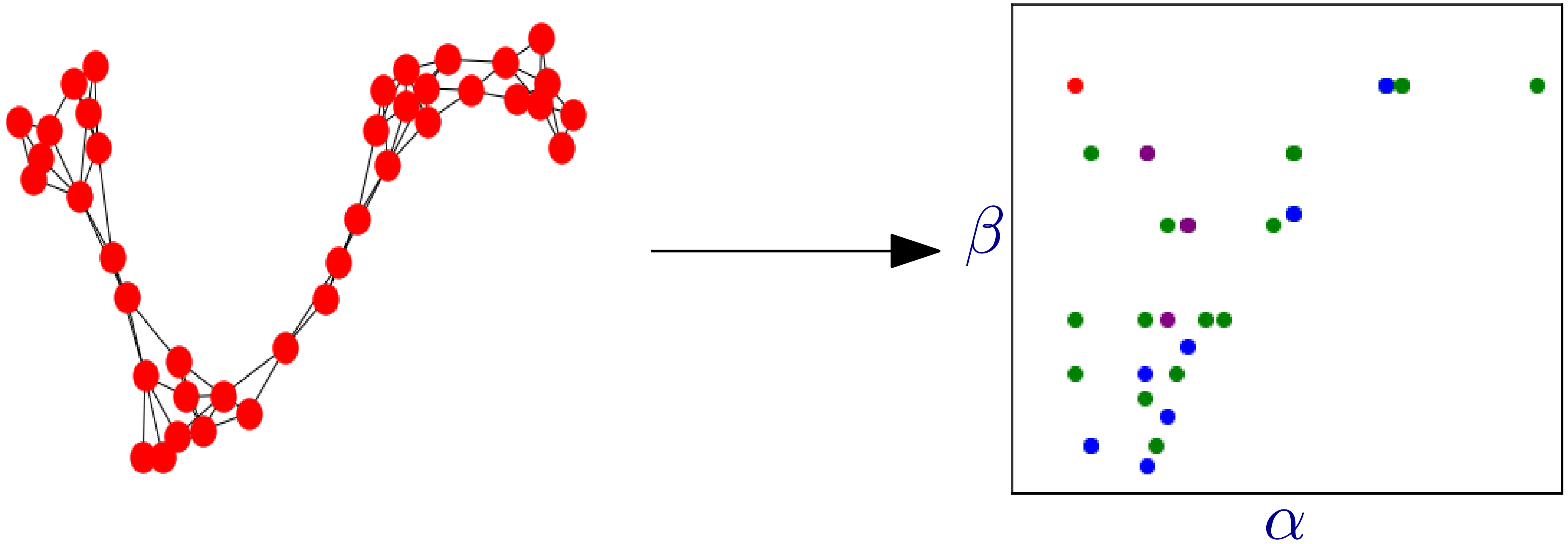
$$\mathrm{hks}_{G,t} : v \mapsto \sum_{k=1}^{n} \exp(-\lambda_k t)\phi_k(v)^2$$

Application to graph classification

[*PersLay: A Neural Network Layer for Persistence Diagrams and New Graph Topological Signatures*, Carrière, Chazal, Ike, Lacombe, Royer, Umeda, AISTATS, 2019]

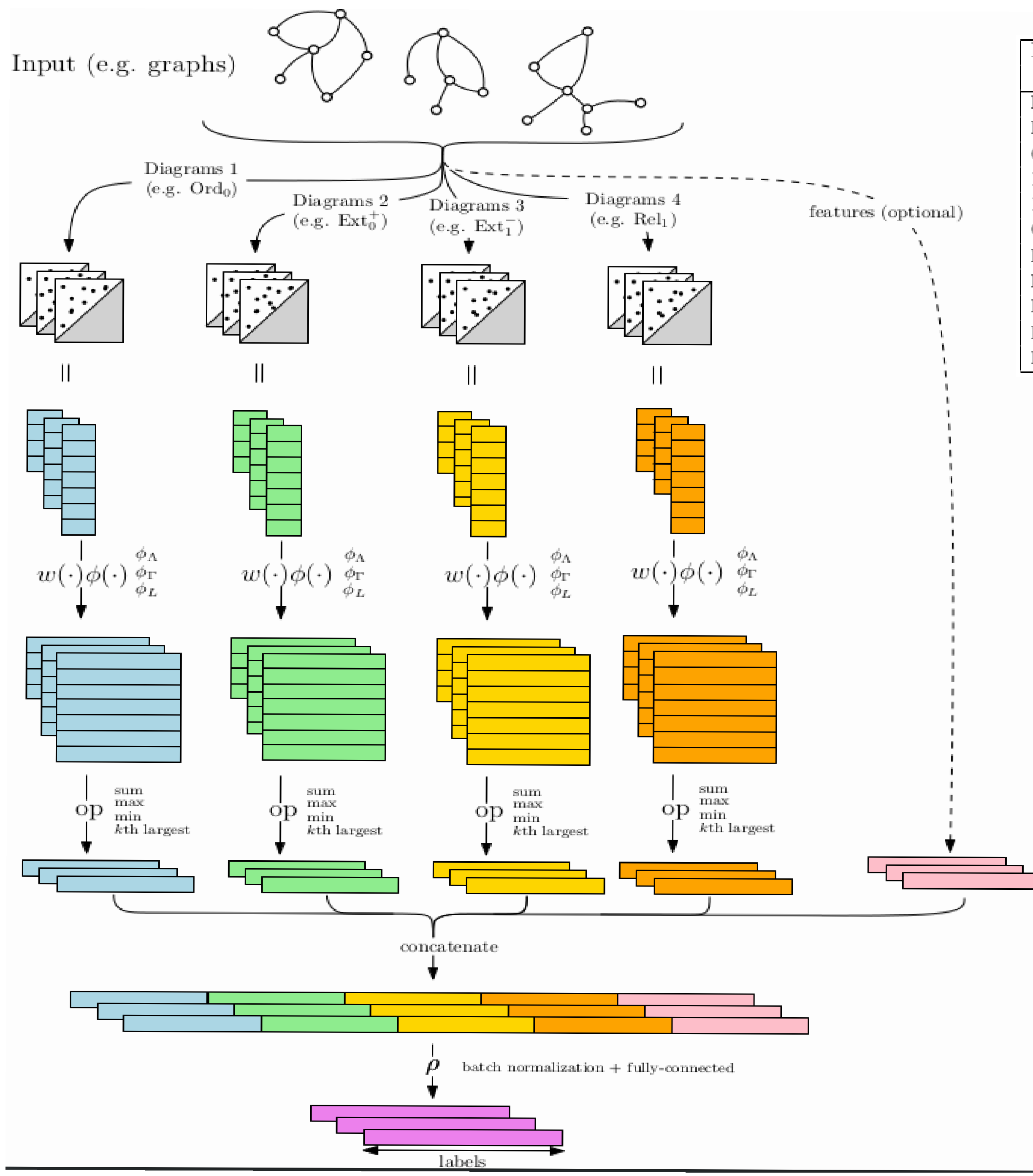
Graph from the `PROTEINS` dataset

Corresponding extended persistence diagram



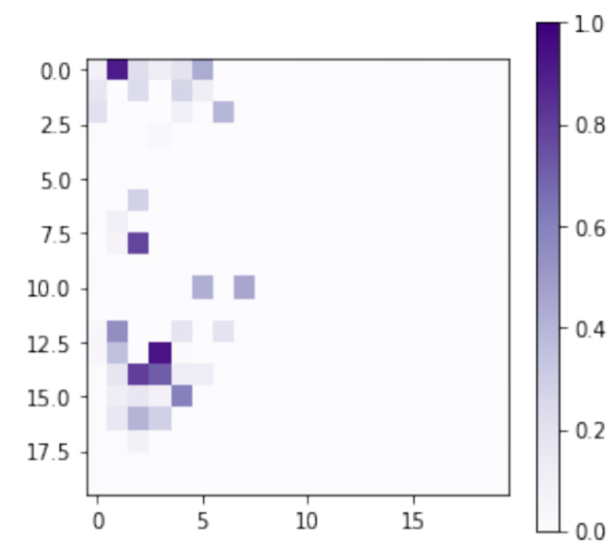
Application to graph classification

[*PersLay: A Neural Network Layer for Persistence Diagrams and New Graph Topological Signatures*, Carrière, Chazal, Ike, Lacombe, Royer, Umeda, AISTATS, 2019]



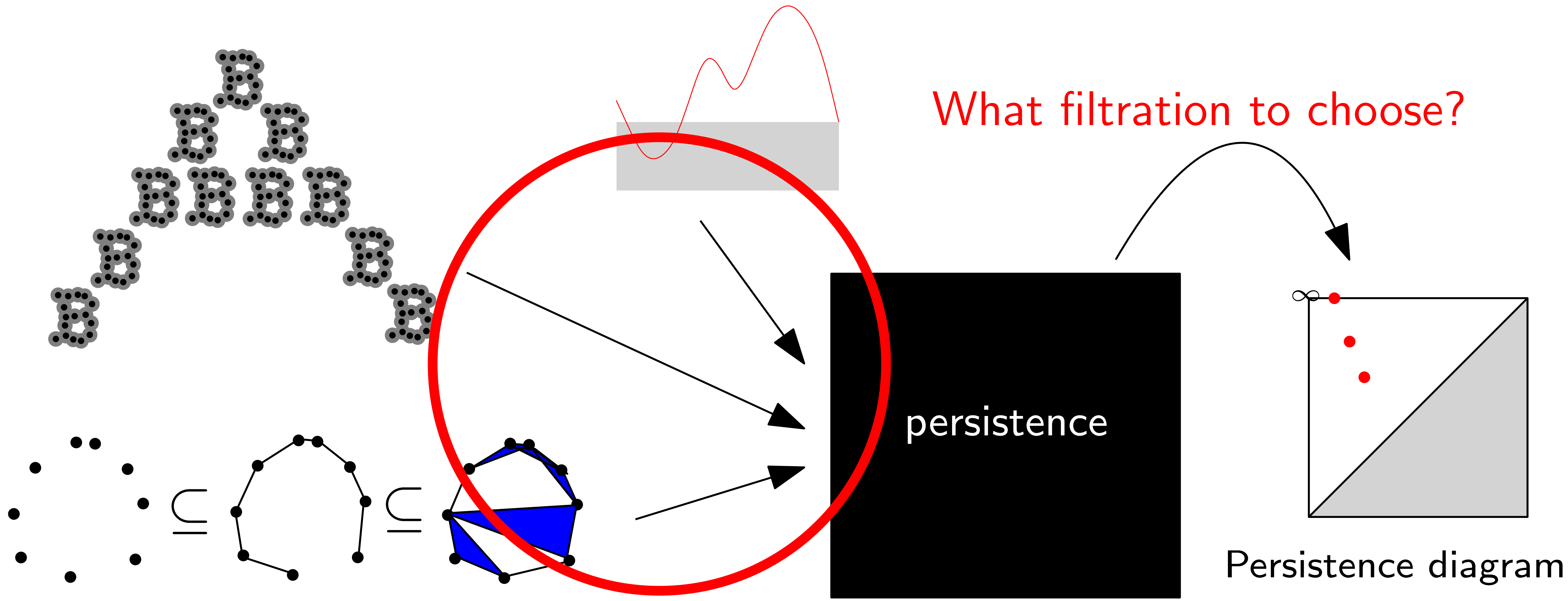
Dataset	SV[1]	RetGK* [2]	FGSD [3]	GCNN [4]	GIN [5]	PersLay	
						Mean	Max
REDDIT5K	—	56.1	47.8	52.9	57.0	55.6	56.5
REDDIT12K	—	48.7	—	46.6	—	47.7	49.1
COLLAB	—	81.0	80.0	79.6	80.1	76.4	78.0
IMDB-B	72.9	71.9	73.6	73.1	74.3	71.2	72.6
IMDB-M	50.3	47.7	52.4	50.3	52.1	48.8	52.2
COX2*	78.4	80.1	—	—	—	80.9	81.6
DHFR*	78.4	81.5	—	—	—	80.3	80.9
MUTAG*	88.3	90.3	92.1	86.7	89.0	89.8	91.5
PROTEINS*	72.6	75.8	73.4	76.3	75.9	74.8	75.9
NCI1*	71.6	84.5	79.8	78.4	82.7	73.5	74.0
NCI109*	70.5	—	78.8	—	—	69.5	70.1

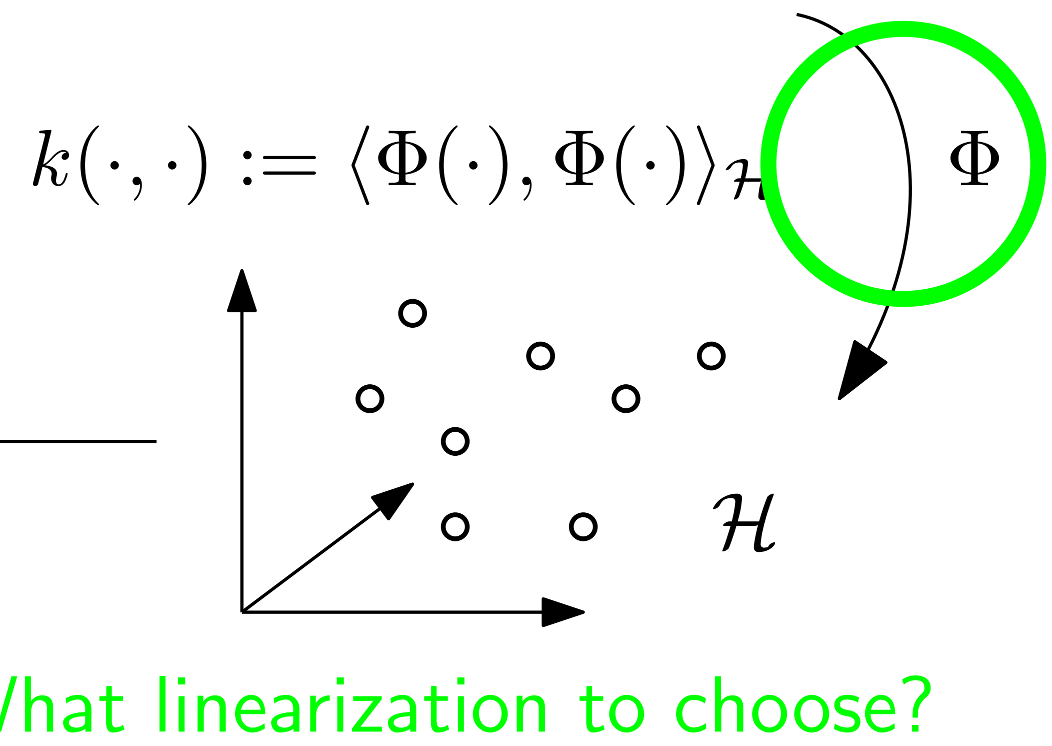
Weight function learnt



(after training on the `MUTAG` dataset)

Persistence diagrams and optimization





- Classifier (RF, SVM, NN etc.)
- Dim. red. (PCA, MDS, UMAP, t-SNE)
- Clustering (DBSCAN, K-means, etc.)

Etc.

Problem setting

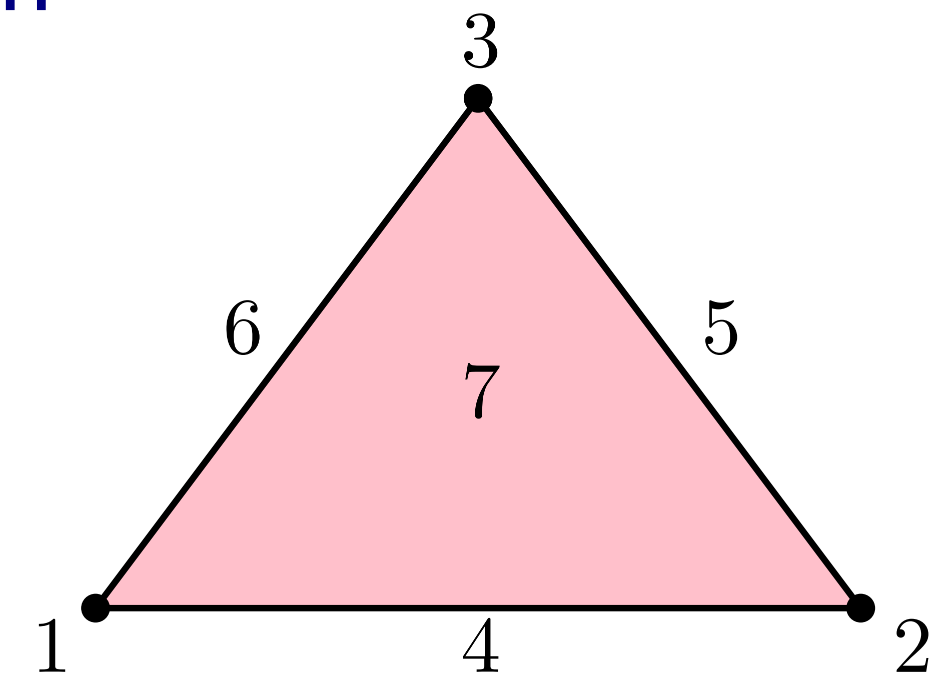
Q: How to define $\nabla \mathrm{D}$?

Q: Given a parameterized family of functions $\mathcal{F} = \{f_\theta \ : \ \theta \in \Theta\}$, how to define $\nabla_\theta \mathrm{D}_k(f_\theta)$?

Q: Given a point cloud $X \subseteq \mathbb{R}^d$, how to define $\nabla_X \mathrm{D}_k(\mathrm{Rips}(X))$?

Idea: Let's go back to the PD construction...

Computation with matrix reduction

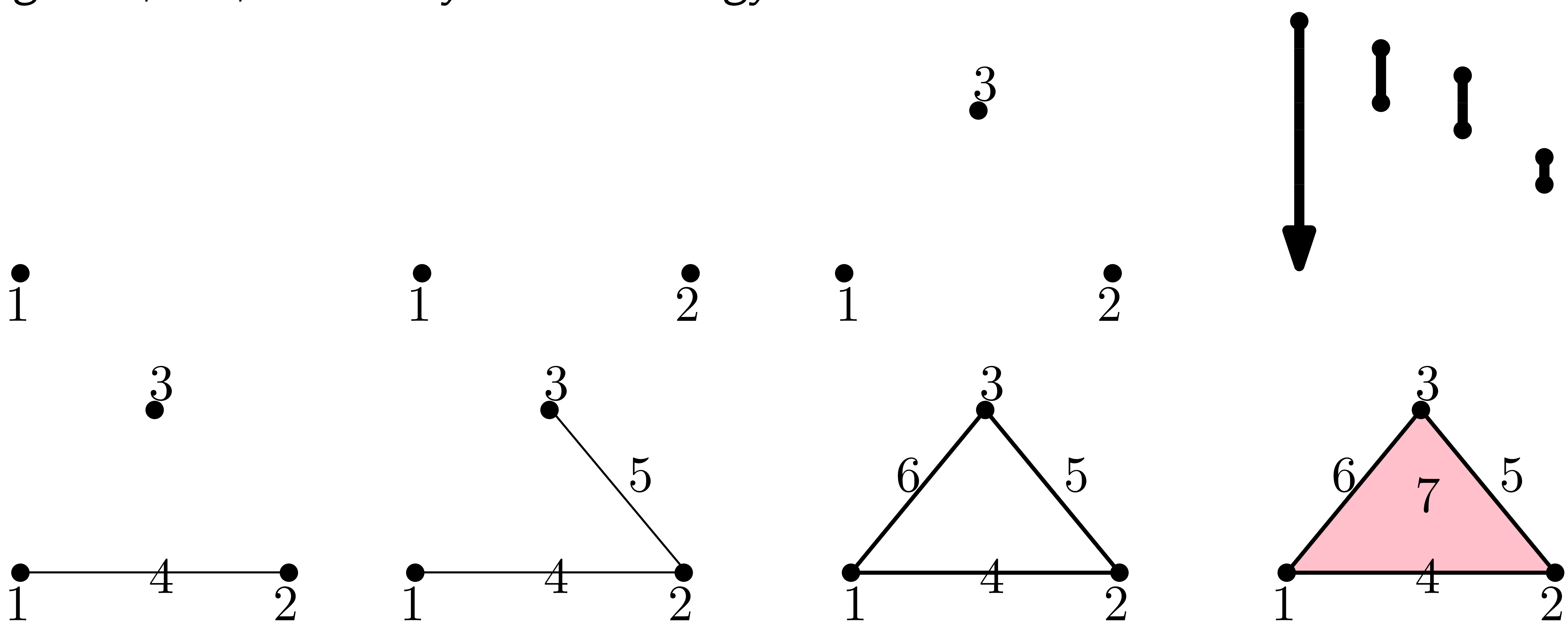


Input: simplicial filtration

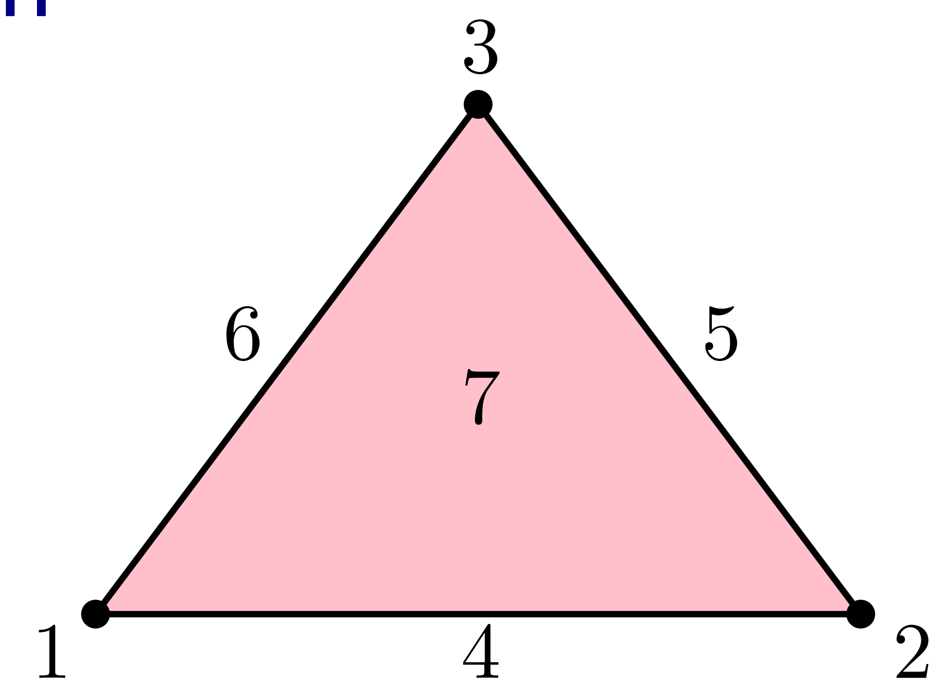
(Persistent) homology can be computed by using the fact that each simplex is either:

positive, i.e., it *creates a new homology class*

negative, i.e., it *destroys an homology class*



Computation with matrix reduction



Input: simplicial filtration

Output: boundary matrix
reduced to column-echelon form

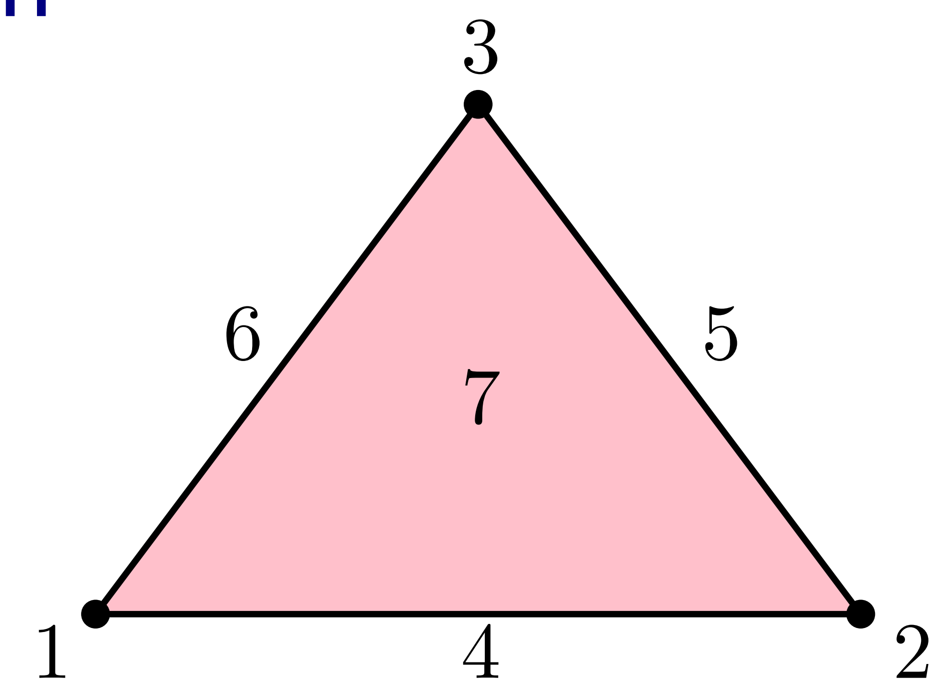
simplex pairs give finite intervals:
$[2,4)$, $[3,5)$, $[6,7)$

unpaired simplices give infinite intervals: $[1,+\infty)$

	1	2	3	4	5	6	7
1				$*$		$*$	
2				$*$	$*$		
3					$*$	$*$	
4							$*$
5							$*$
6							$*$
7							

	1	2	3	4	5	6	7
$\boxed{1}$				$*$			
2				1	$*$		
3					1		
4							$*$
5							$*$
6							1
7							

Computation with matrix reduction



Input: simplicial filtration

Output: boundary matrix reduced to column-echelon form

simplex pairs give finite intervals:
$[2,4)$, $[3,5)$, $[6,7)$

unpaired simplices give infinite intervals: $[1,+\infty)$

A persistence diagram D is made of all $(\mathcal{F}(\sigma_+),\mathcal{F}(\sigma_-)) \in \mathbb{R}^2$ where σ_+ (resp. σ_-) is positive (resp. negative), and $\mathcal{F}$ is the filtration function.

Thus we can define the gradient of a point $p = (\mathcal{F}(\sigma_+),\mathcal{F}(\sigma_-)) \in \mathrm{D}$ as

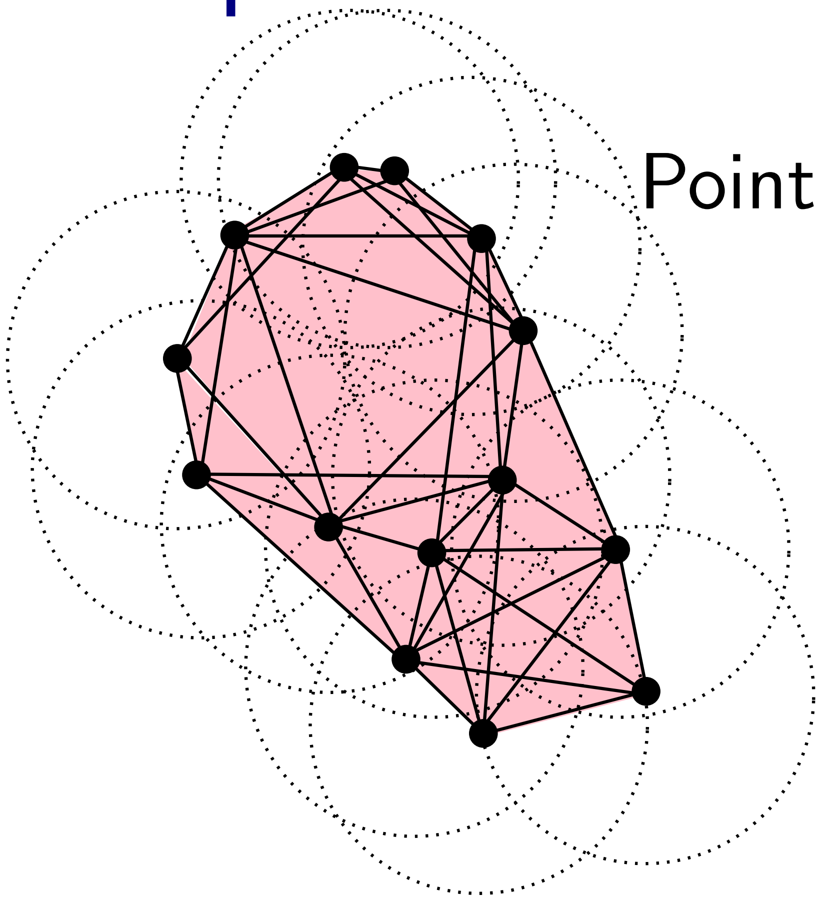
$$\nabla p = [\nabla \mathcal{F}(\sigma_+), \nabla \mathcal{F}(\sigma_-)]$$

	1	2	3	4	5	6	7
1				*			
2				1	*		
3					1		
4							*
5							*
6							1
7							

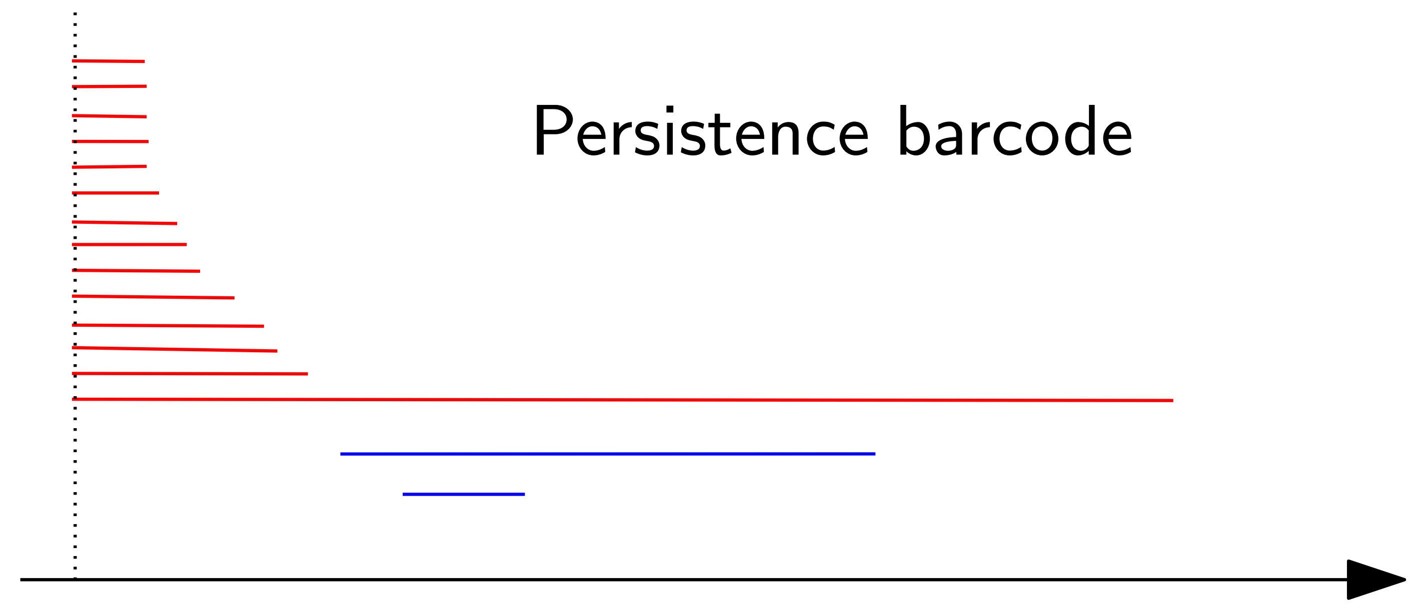
Example: Vietoris-Rips gradient

Q: Define and compute Vietoris-Rips gradient?

Example: Vietoris-Rips gradient



Point cloud $\hat{X}_n$



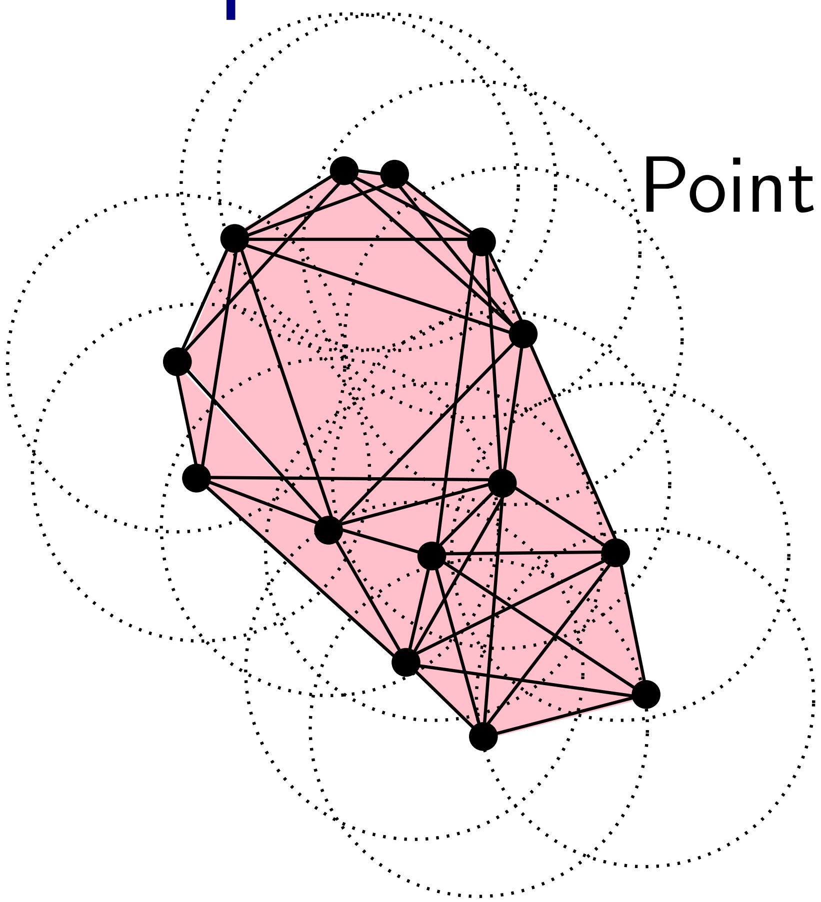
Given k-dim. simplex $\sigma = [v_0, \ldots, v_k]$, one has

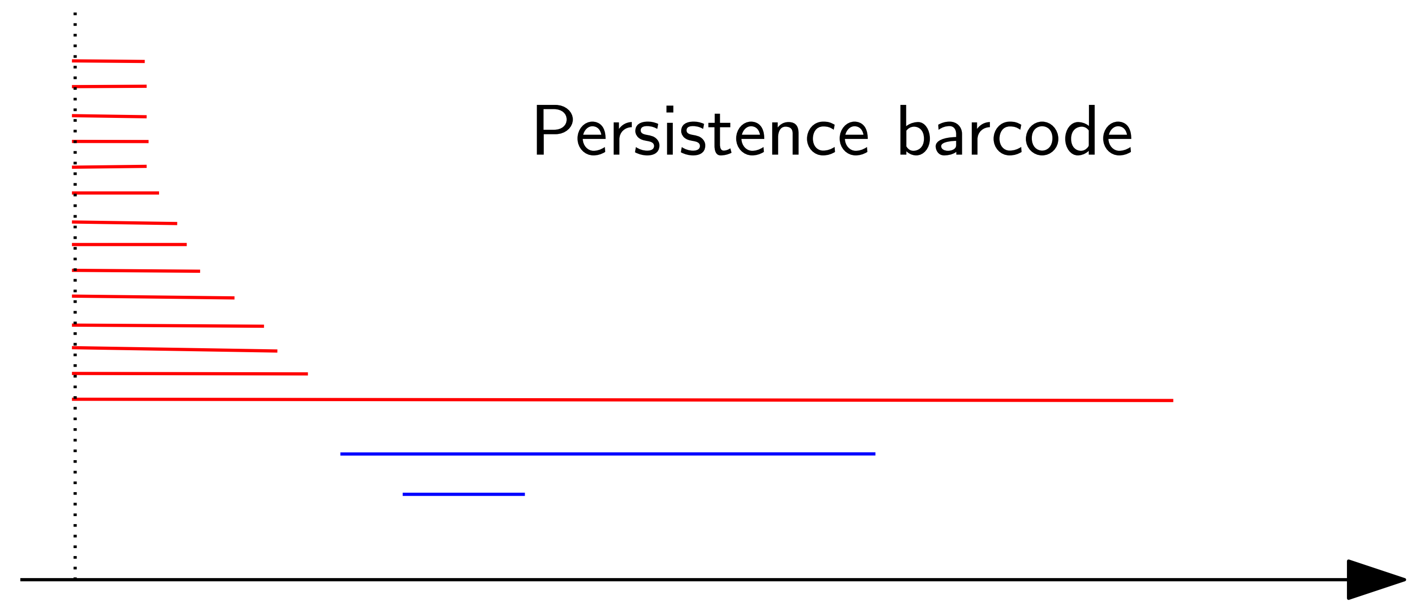
$$\mathcal{F}(\sigma) = \max_{i,j} \|v_i - v_j\|$$

Let $p = (\mathcal{F}(\sigma_+), \mathcal{F}(\sigma_-)) \in \mathrm{D}_k(\mathrm{Rips}(X))$

with $\sigma_+ = \{v_0, \ldots, v_k\}$ and $\sigma_- = \{w_0, \ldots, w_{k+1}\}$

Example: Vietoris-Rips gradient



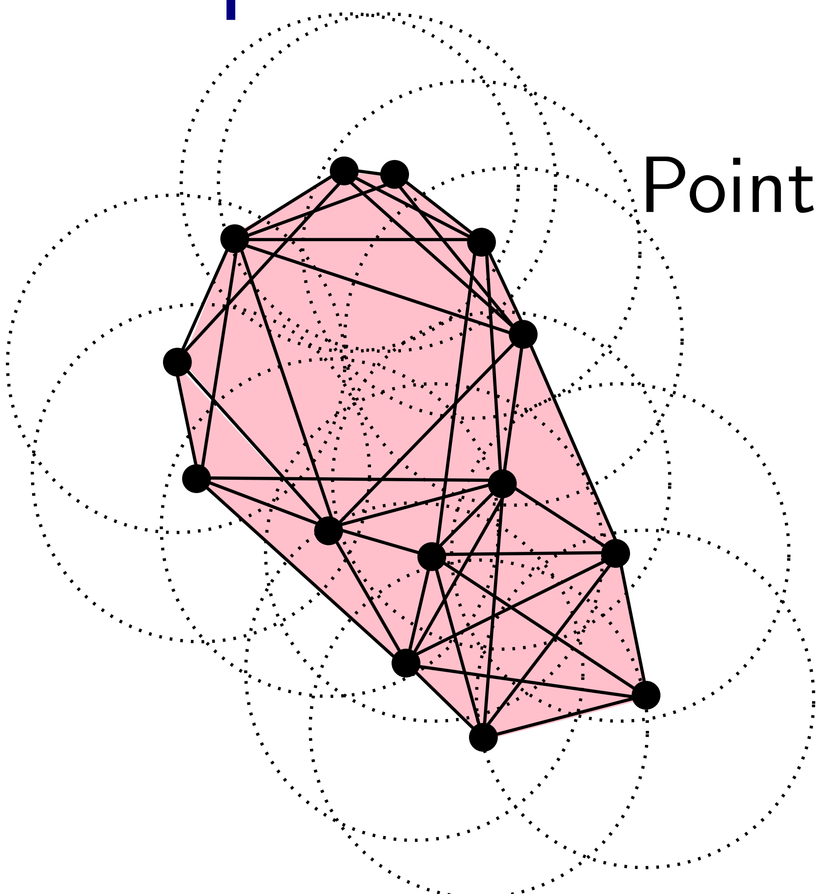


$$\nabla_X p = \left[\frac{\partial}{\partial X}\|v_{i^*} - v_{j^*}\|, \frac{\partial}{\partial X}\|w_{a^*} - w_{b^*}\|\right]$$

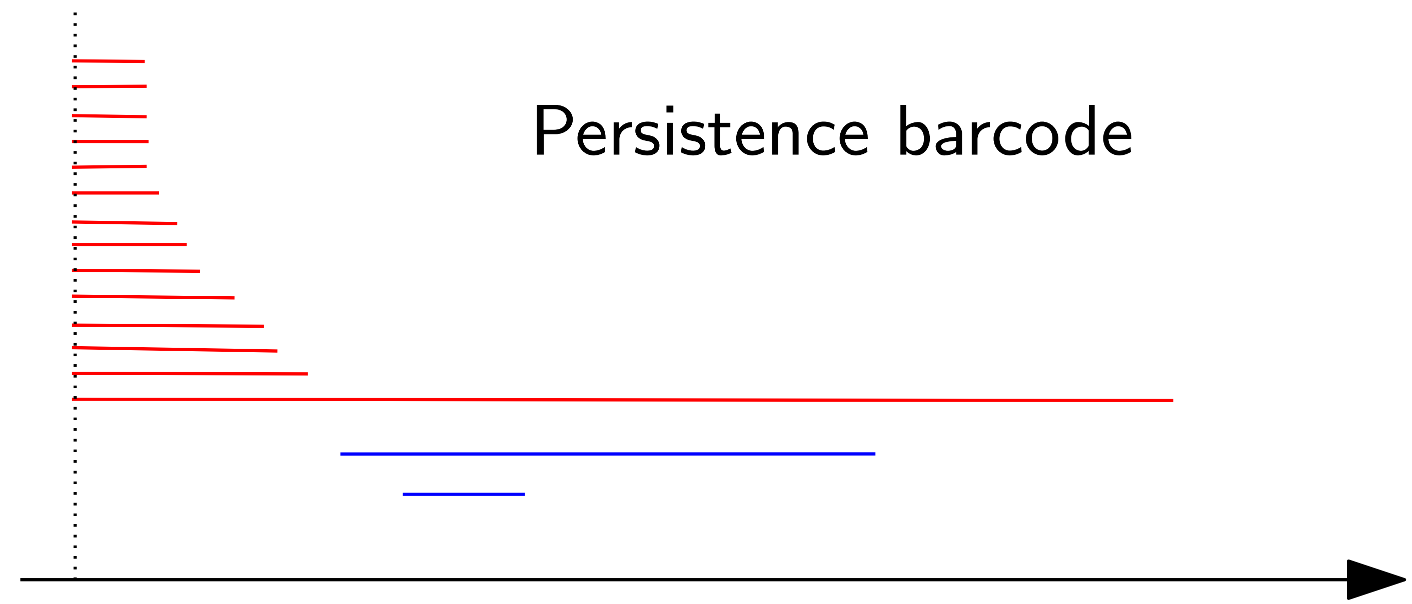
$\frac{\partial}{\partial v_i^{(d)}}\|v_{i^*} - v_{j^*}\| = (-)\frac{1}{\|v_{i^*} - v_{j^*}\|}(v_{i^*}^{(d)} - v_{j^*}^{(d)})$ if $i = i^*$ (j^*) and 0 otherwise

With this gradient rule, one can do gradient descent with any function of persistence!

Example: Vietoris-Rips gradient



Point cloud $\hat{X}_n$

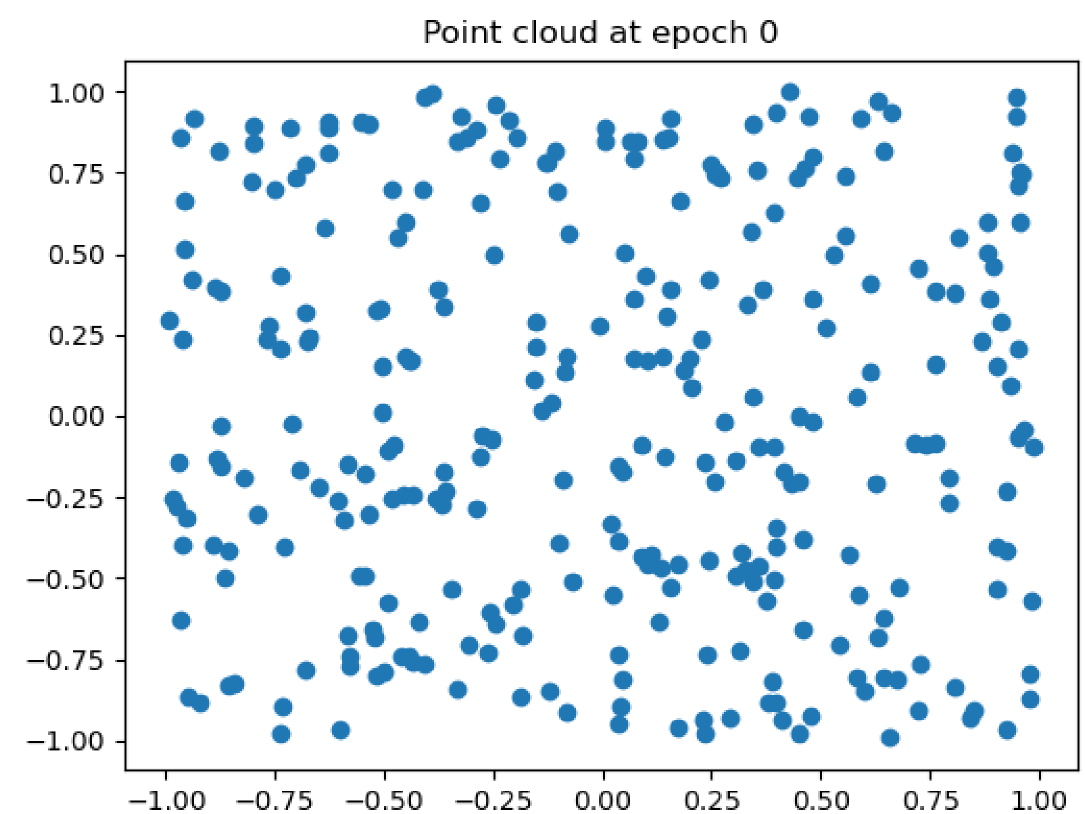


Let's say we want to maximize the number of holes in that point cloud.

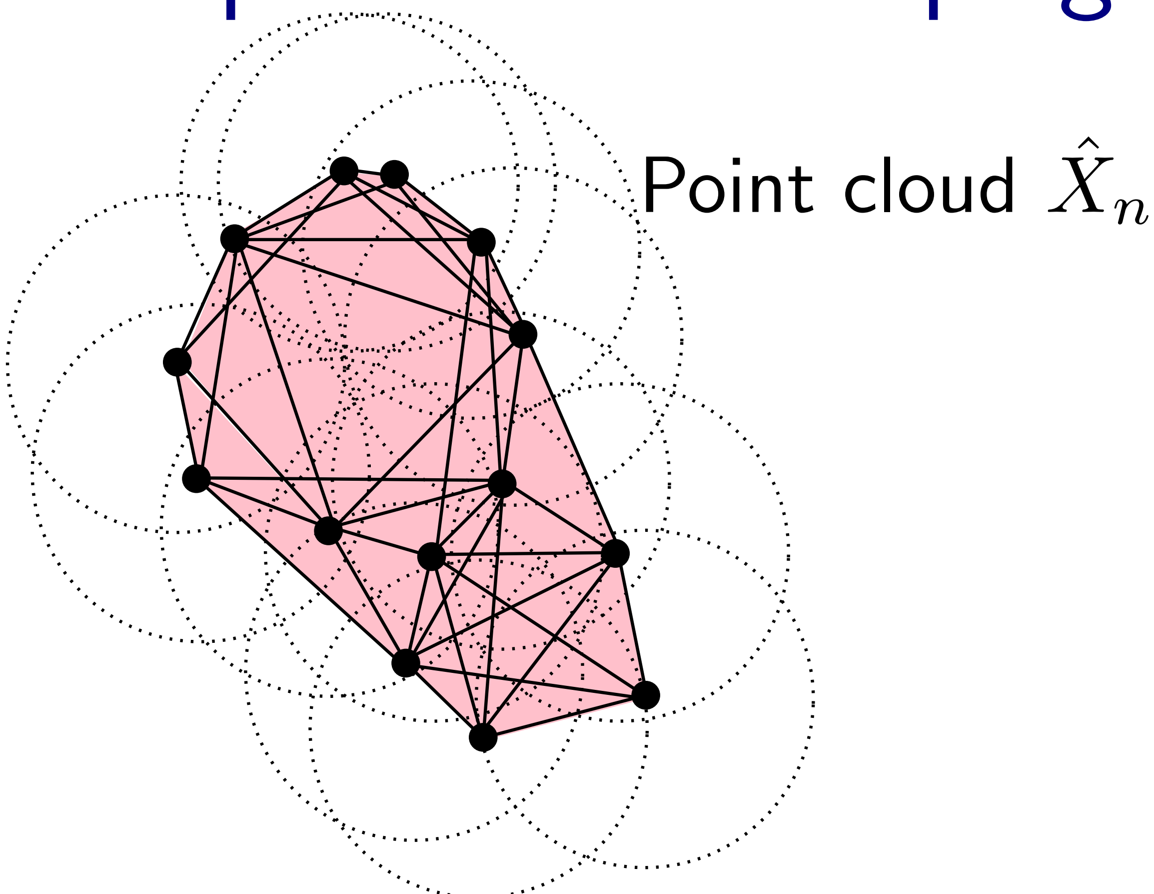
We can use gradient descent to minimize loss

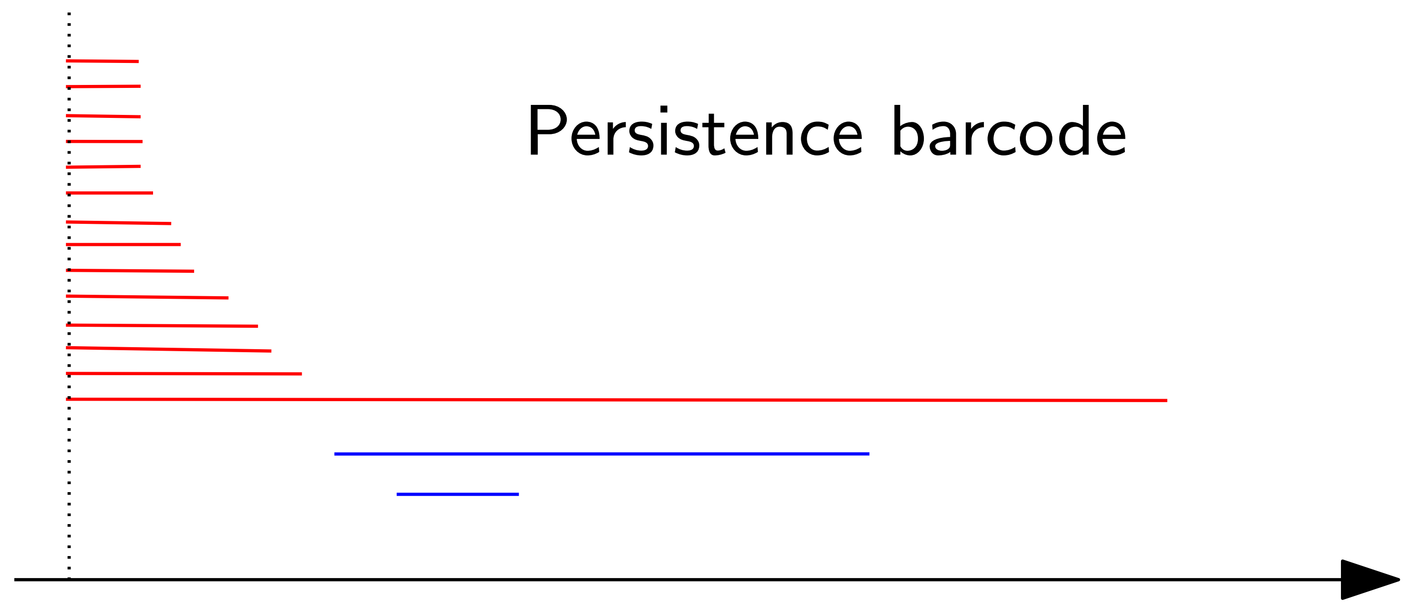
$$\mathcal{L}(X) = -\sum_{p} \|p\|_2^2,$$

with $p \in \mathrm{D}_1(\mathrm{Rips}(X))$



Example: Vietoris-Rips gradient



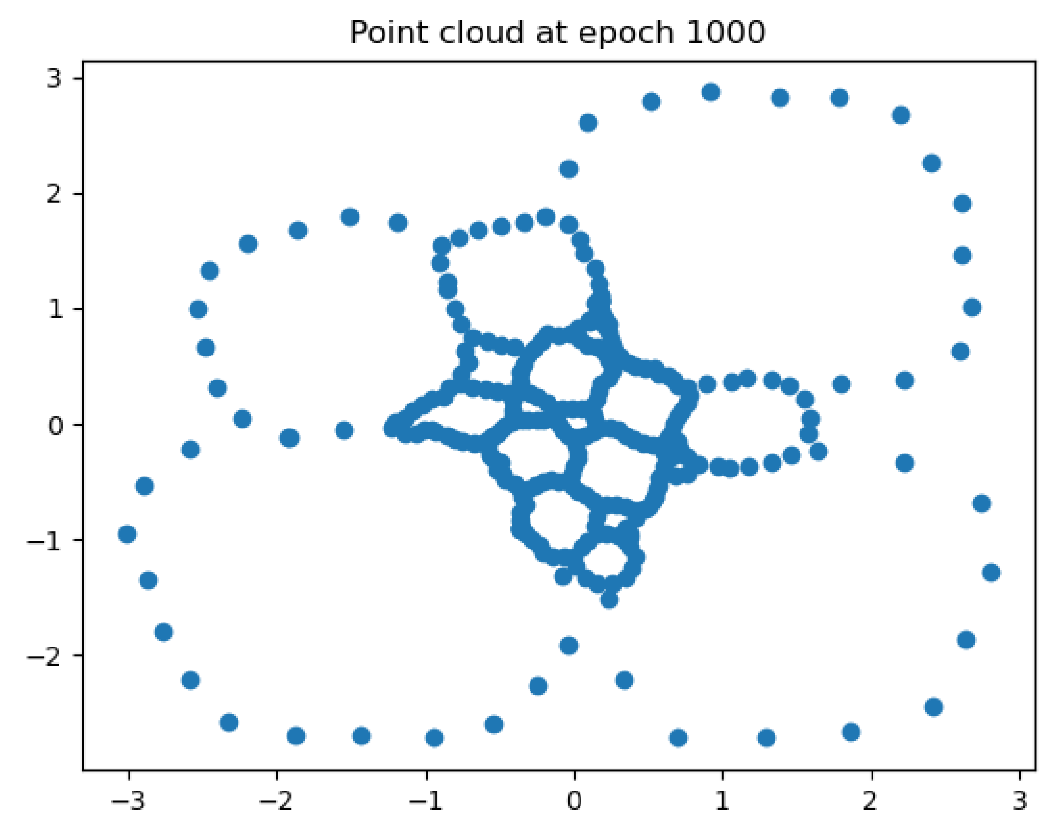


Let's say we want to maximize the number of holes in that point cloud.

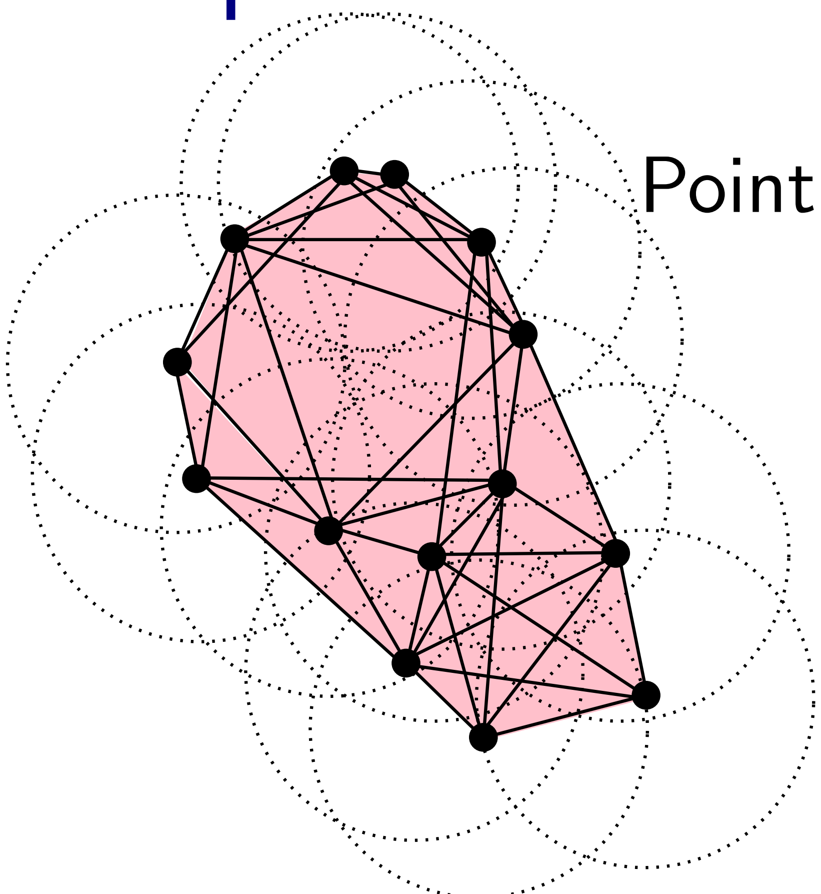
We can use gradient descent to minimize loss

$$\mathcal{L}(X) = -\sum_p \|p\|_2^2,$$

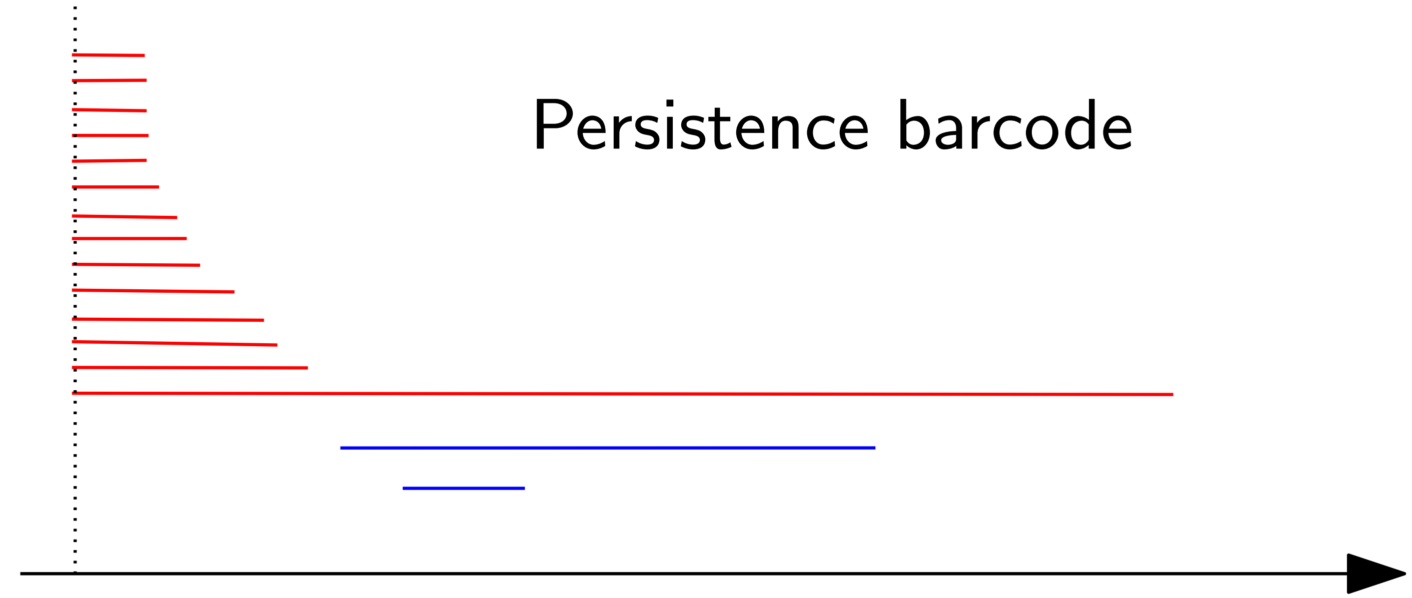
with $p \in \mathrm{D}_1(\mathrm{Rips}(X))$



Example: Vietoris-Rips gradient



Point cloud $\hat{X}_n$

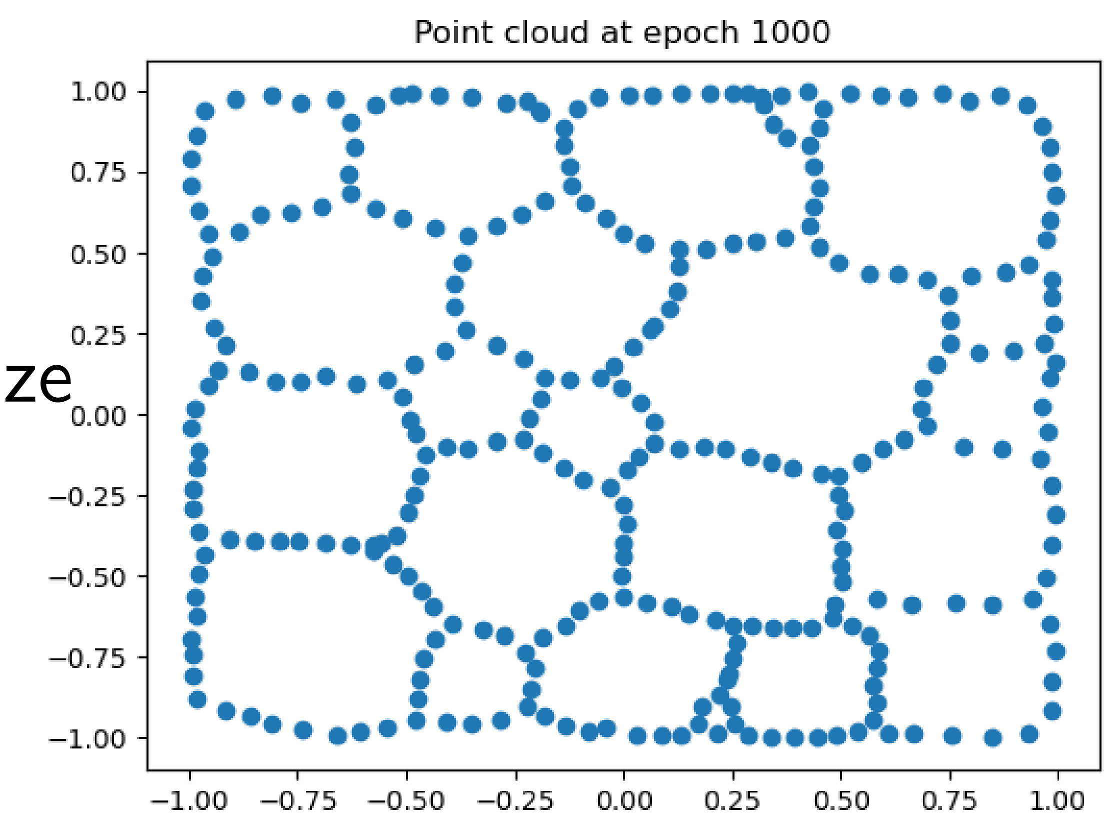


Let's say we want to maximize the number of holes in that point cloud.

We can use gradient descent to minimize loss

$$\mathcal{L}(X) = -\sum_{p} \|p\|_2^2 + d(X, C),$$

with $p \in \mathrm{D}_1(\mathrm{Rips}(X))$ and C unit square



Example: Sublevel sets

Given k-dim. simplex $\sigma = [v_0, \dots, v_k]$, one has

$$\mathcal{F}(\sigma) = \max_i \, f_\theta(v_i)$$

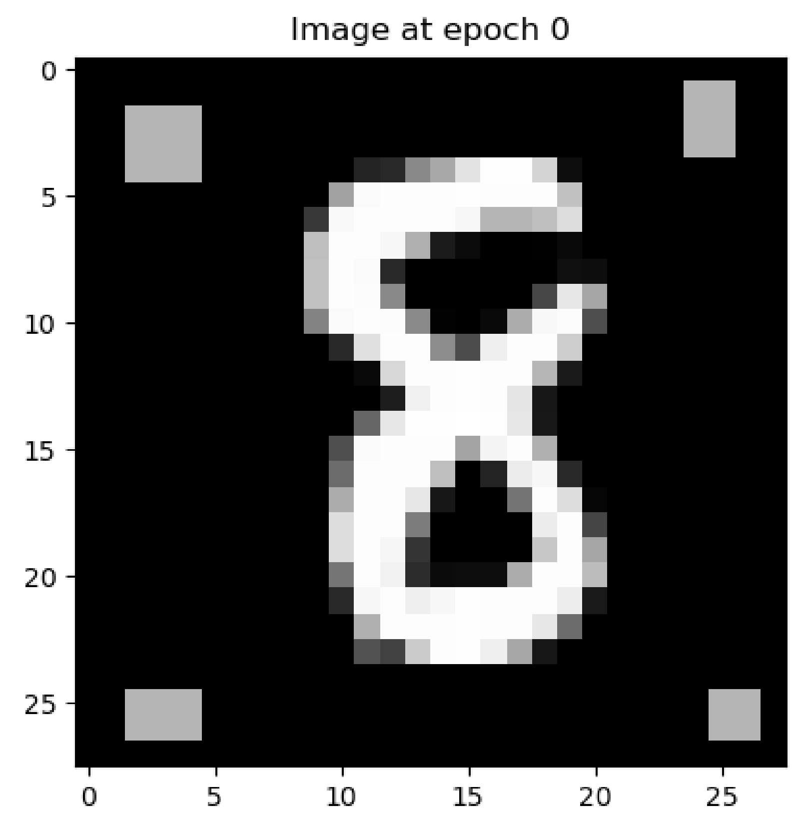
$$\nabla_\theta p = \left[\tfrac{\partial}{\partial \theta} f_\theta(v_{i^*}), \tfrac{\partial}{\partial \theta} f_\theta(w_{a^*}) \right]$$

Let's say we want to remove the stains in that image.

We can use gradient descent to minimize loss

$$\mathcal{L}(X) = \sum_p \|p\|_2^2,$$

with $p \in \mathrm{D}_0(I)$



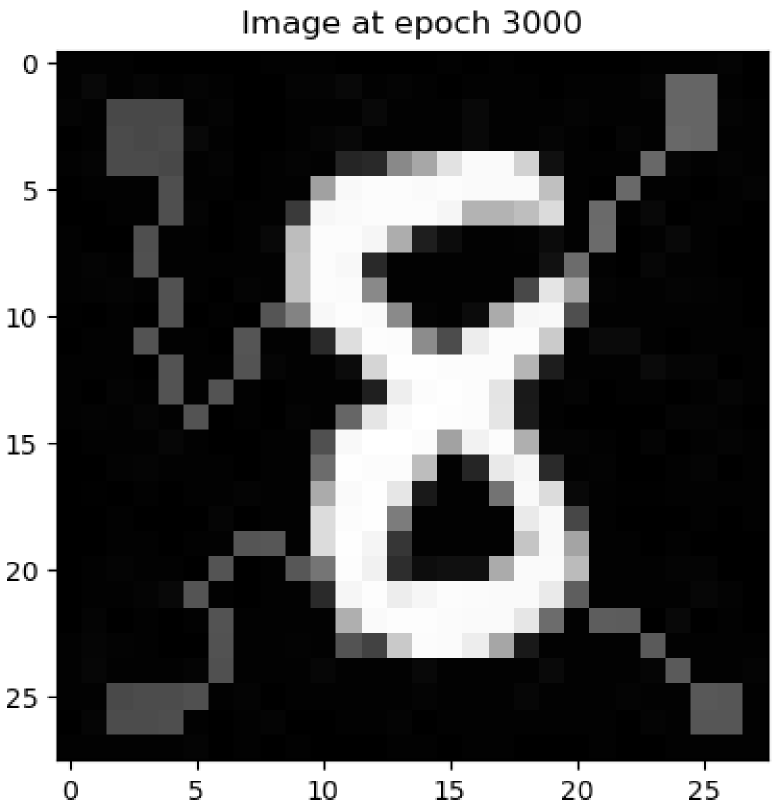
Example: Sublevel sets

Given k-dim. simplex $\sigma = [v_0, \ldots, v_k]$, one has

$$\mathcal{F}(\sigma) = \max_i \ f_\theta(v_i)$$

$$\nabla_\theta p = \left[\tfrac{\partial}{\partial\theta} f_\theta(v_{i^*}), \tfrac{\partial}{\partial\theta} f_\theta(w_{a^*})\right]$$

Let's say we want to remove the stains in that image.



We can use gradient descent to minimize loss

$$\mathcal{L}(X) = \sum_p \|p\|_2^2,$$

with $p \in \mathrm{D}_0(I)$

Example: Sublevel sets

Given k-dim. simplex $\sigma = [v_0, \ldots, v_k]$, one has

$$\mathcal{F}(\sigma) = \max_i \; f_\theta(v_i)$$

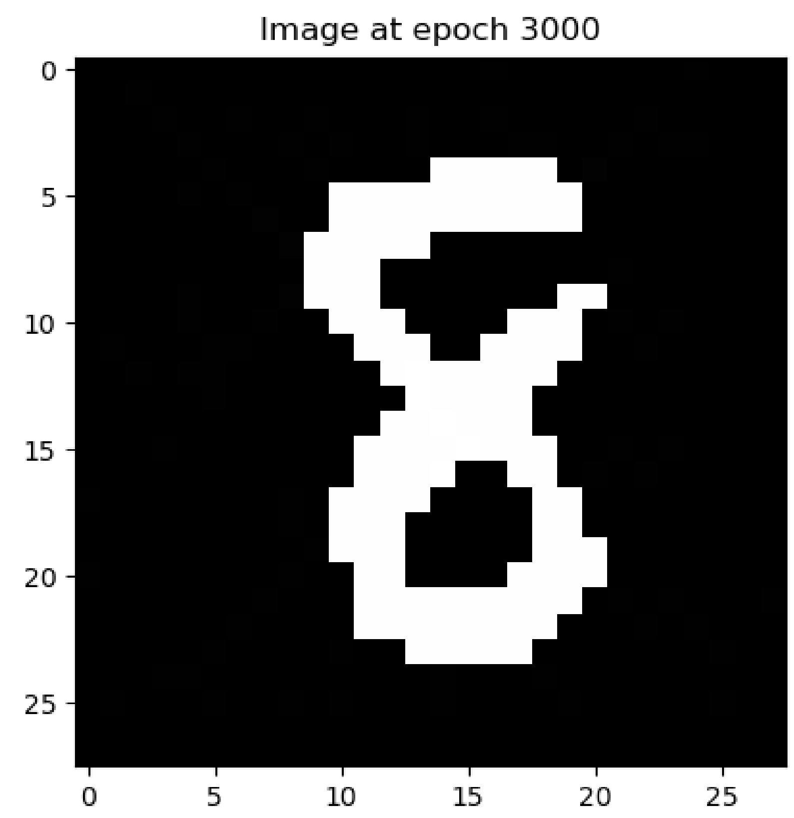
$$\nabla_\theta p = \left[\tfrac{\partial}{\partial\theta} f_\theta(v_{i^*}), \tfrac{\partial}{\partial\theta} f_\theta(w_{a^*})\right]$$

Let's say we want to remove the stains in that image.

We can use gradient descent to minimize loss

$$\mathcal{L}(X) = \sum_p \|p\|_2^2 + \sum_{P \in I} \max\{|P|, |1-P|\},$$

with $p \in \mathrm{D}_0(I)$



Topological gradient descent

[*Optimizing persistent homology based functions*, Carrière, Chazal, Glisse, Ike, Kanna, Umeda, ICML, 2021]

For a fixed ordering of the simplices in a simplicial complex K, the corresponding persistence diagram always has the same number of points: its gradient is well-defined!

If the ordering changes, the boundary matrix can have a new reduced form and the persistence diagram can have a new, different number of points.

Prop: Let K be a simplicial complex and let $\Phi : A \to \mathbb{R}^{|K|}$ a (parameterized) filtration of K. There exists a partition $A = S \sqcup O_1 \sqcup \cdots \sqcup O_k$ s.t. all the restrictions $\Phi : O_i \to \mathbb{R}^{|K|}$ are differentiable.

The O_i's are the parts of A where the ordering of the simplices of K is preserved, and S is the boundaries of all O_i's.

Q: What is S for Vietoris-Rips? Sublevel sets?

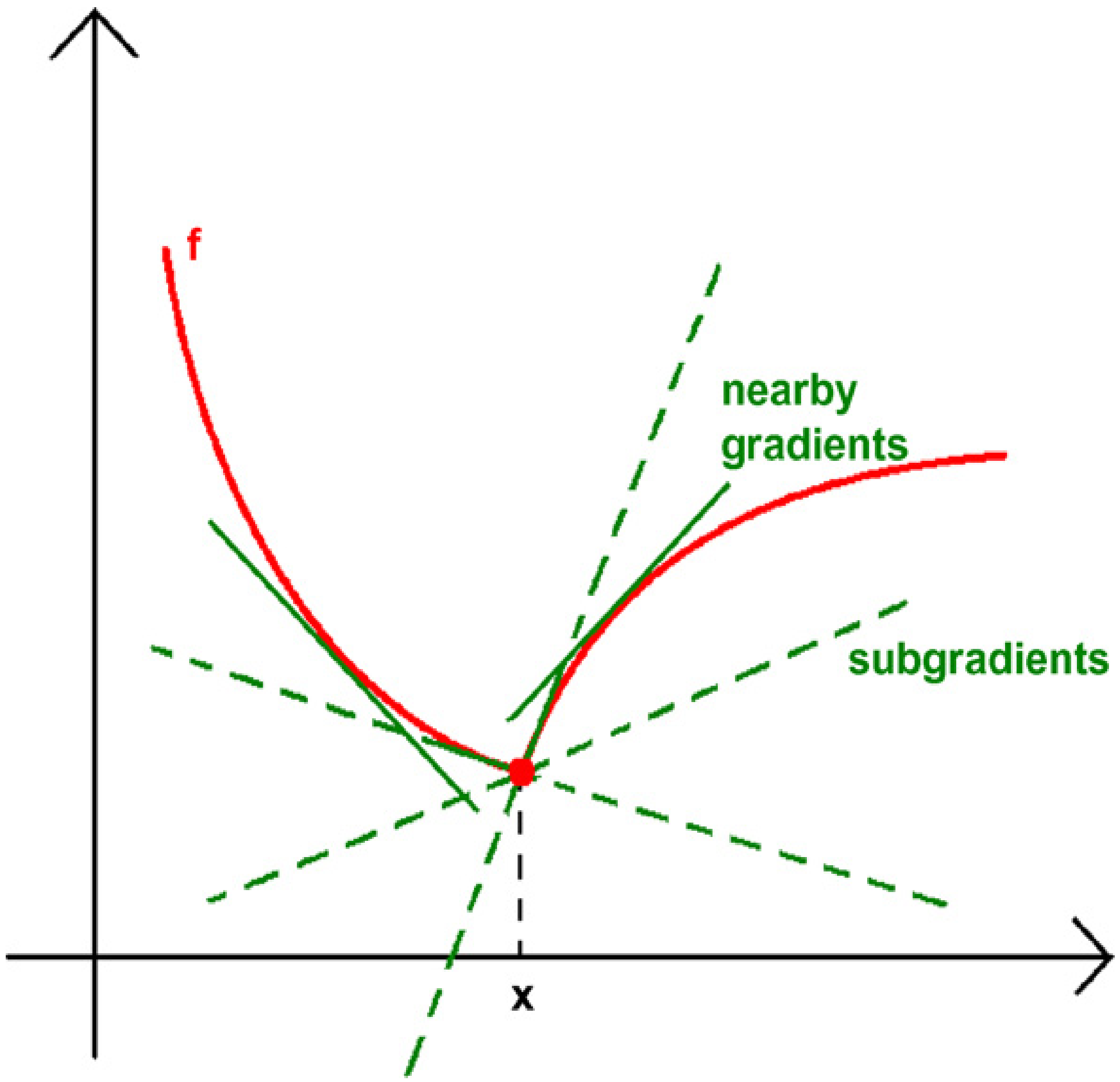
Topological gradient descent

[*Optimizing persistent homology based functions*, Carrière, Chazal, Glisse, Ike, Kanna, Umeda, ICML, 2021]

Def: The *Clarke subdifferential* $\partial\mathcal{L}$ of $\mathcal{L}$ is the set:

$$\partial_x\mathcal{L} = \mathrm{conv}\{\lim_{x_i\to x}\nabla\mathcal{L}(x_i) : \mathcal{L} \text{ is diff. at } x_i\},$$

where conv denotes the convex hull.



Topological gradient descent

[*Optimizing persistent homology based functions*, Carrière, Chazal, Glisse, Ike, Kanna, Umeda, ICML, 2021]

Let $\{\alpha_k\}_k$, $\{\zeta_k\}_k$ s.t.

$\alpha_k \geq 0$, $\sum_k \alpha_k = +\infty$ and $\sum_k \alpha_k^2 < +\infty$

ζ_k random variables s.t. $E[\zeta_k] = 0$ and $E[\|\zeta_k\|^2] < C$ for some $C > 0$

Thm: As long as $\mathcal{L} \circ \mathrm{Pers} \circ \Phi$ is locally Lipschitz, the sequence

$$a_{k+1} = a_k - \alpha_k(g_k + \zeta_k),$$

where $g_k \in \partial_{a_k}(\mathcal{L} \circ \mathrm{Pers} \circ \Phi)$, converges to a critical point of $\mathcal{L} \circ \mathrm{Pers} \circ \Phi$.

Q: Does this result apply to d_b and d_p? What is the gradient?

Topological stratified gradient descent

[*A gradient sampling algorithm for stratified maps with applications to topological data analysis*, Leygonie, Carrière, Lacombe, Oudot, 2021]

Better guarantees can be obtained by smoothing the gradient definition.

Def: The *smoothed topological gradient* of $\mathrm{Pers} \circ \Phi$ is defined as:

$$\tilde{\nabla}_a = \mathrm{argmin}\{\|g\| \, : \, g \in \mathrm{conv}(S_a)\}$$

where $S_a = \{\nabla_{a'} \, : \, a' \in O_i, O_i \in \mathcal{N}(O_a)\}$, where O_a is the stratum associated to a, and $\mathcal{N}(O_a)$ is the set of strata that are close to O_a.

Intuitively, close strata means that their corresponding orderings are very similar, e.g., they differ by single swaps, or their distance is bounded by $\epsilon > 0$.

Thm: Let $\epsilon > 0$. As long as $\mathcal{L} \circ \mathrm{Pers} \circ \Phi$ is Lipschitz, the sequence

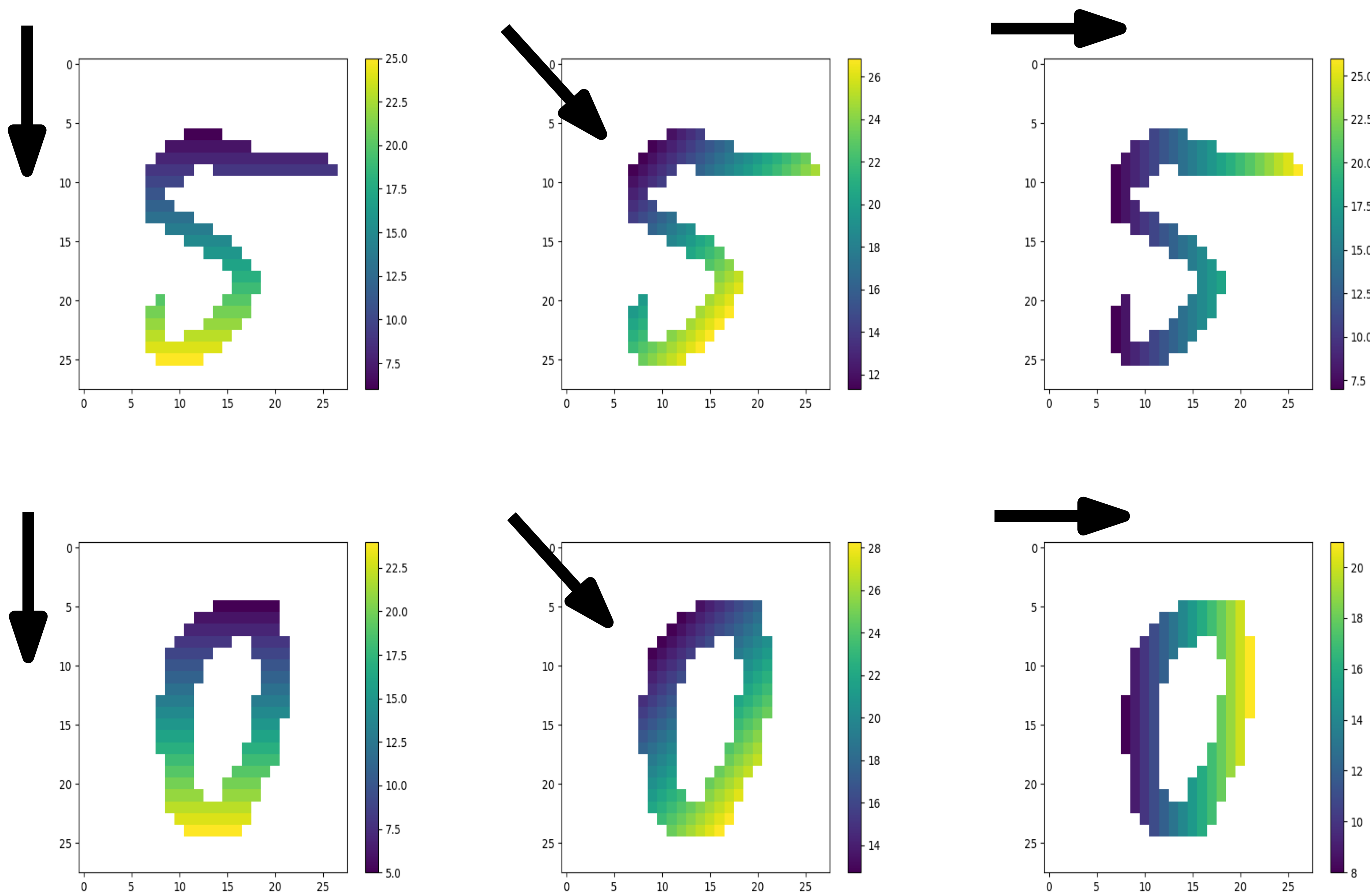
$$a_{k+1} = a_k - \epsilon \cdot \tilde{\nabla}_{a_k} / \|\tilde{\nabla}_{a_k}\|,$$

converges in **finitely many** iterations to $\tilde{a}$ s.t. $\exists \bar{a} \, : \, \tilde{\nabla}_{\bar{a}} = 0$ and $\|\tilde{a} - \bar{a}\| \leq \epsilon$.

Example: filter selection

Assume we have a supervised classification task. The goal is to find a filtration from a family $\mathcal{F}$ such that the corresponding persistence diagrams give the best classification score.

Ex: images filtered by a direction parameterized by angle.



Example: filter selection

Assume we have a supervised classification task. The goal is to find a filtration from a family $\mathcal{F}$ such that the corresponding persistence diagrams give the best classification score.

Idea: minimize:

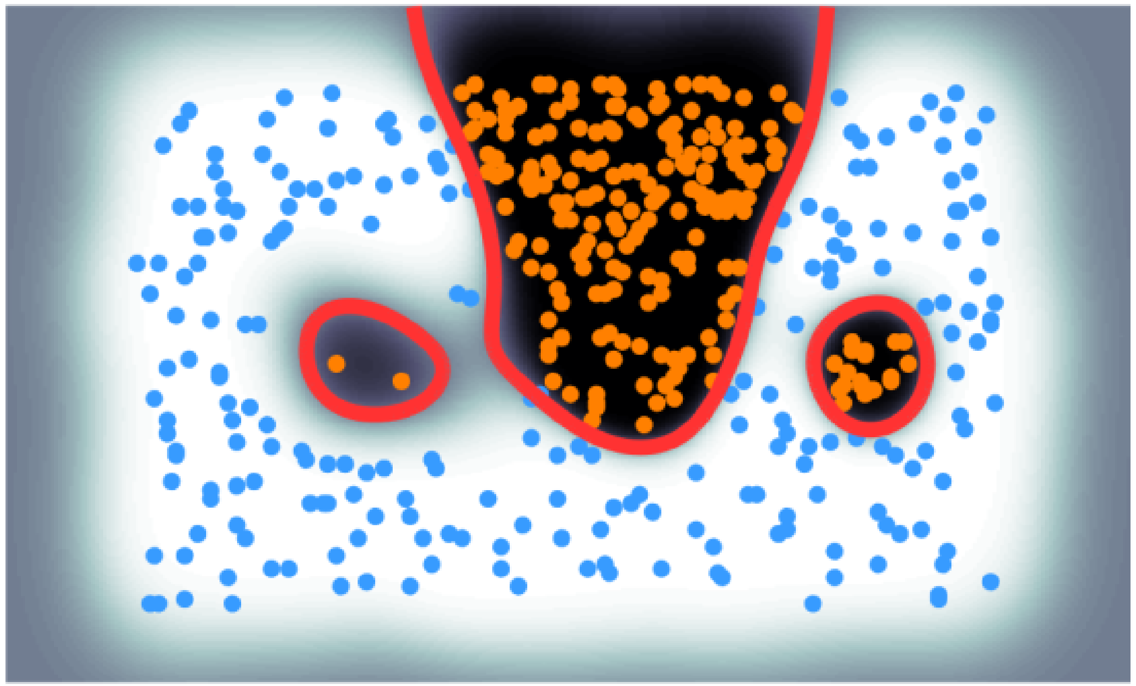
$$\mathcal{L}(f) = \sum_l \frac{\sum_{y_i=y_j=l} \mathrm{d}_p(\mathrm{D}_f(x_i), \mathrm{D}_f(x_j))}{\sum_{y_i=l} \mathrm{d}_p(\mathrm{D}_f(x_i), \mathrm{D}_f(x_j))},$$

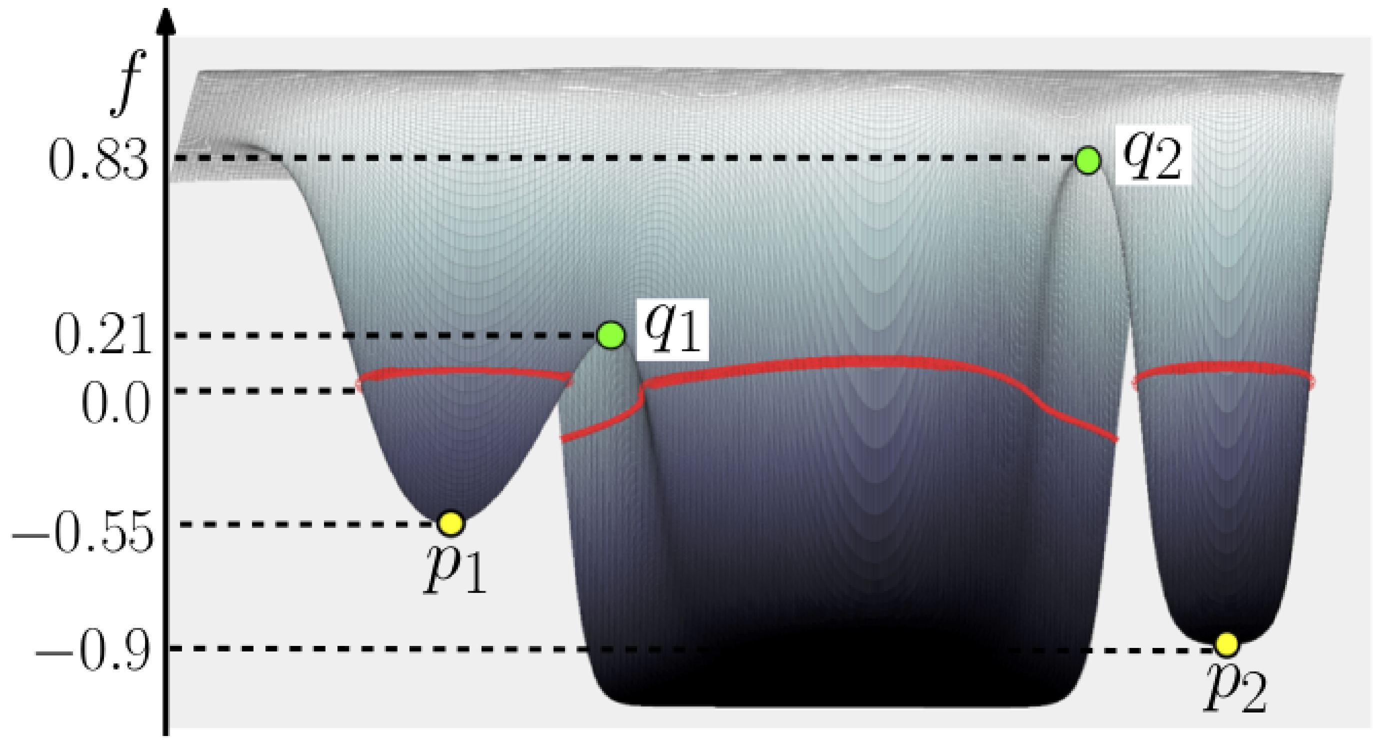
one can also use Sliced Wasserstein for speedup.

Dataset	Baseline	Before	After	Difference	Dataset	Baseline	Before	After	Difference
vs01	100.0	61.3	99.0	**+37.6**	vs26	99.7	98.8	98.2	-0.6
vs02	99.4	98.8	97.2	-1.6	vs28	99.1	96.8	96.8	0.0
vs06	99.4	87.3	98.2	**+10.9**	vs29	99.1	91.6	98.6	**+7.0**
vs09	99.4	86.8	98.3	**+11.5**	vs34	99.8	99.4	99.1	-0.3
vs16	99.7	89.0	97.3	**+8.3**	vs36	99.7	99.3	99.3	-0.1
vs19	99.6	84.8	98.0	**+13.2**	vs37	98.9	94.9	97.5	**+2.6**
vs24	99.4	98.7	98.7	0.0	vs57	99.7	90.5	97.2	**+6.7**
vs25	99.4	80.6	97.2	**+16.6**	vs79	99.1	85.3	96.9	**+11.5**

More examples

[*A Topological Regularizer for Classifiers via Persistent Homology*, Chen, Ni, Bai, Wang, AISTATS, 2019]





[*Topological autoencoders*, Moor, Horn, Rieck, Borgwardt, ICML, 2020]