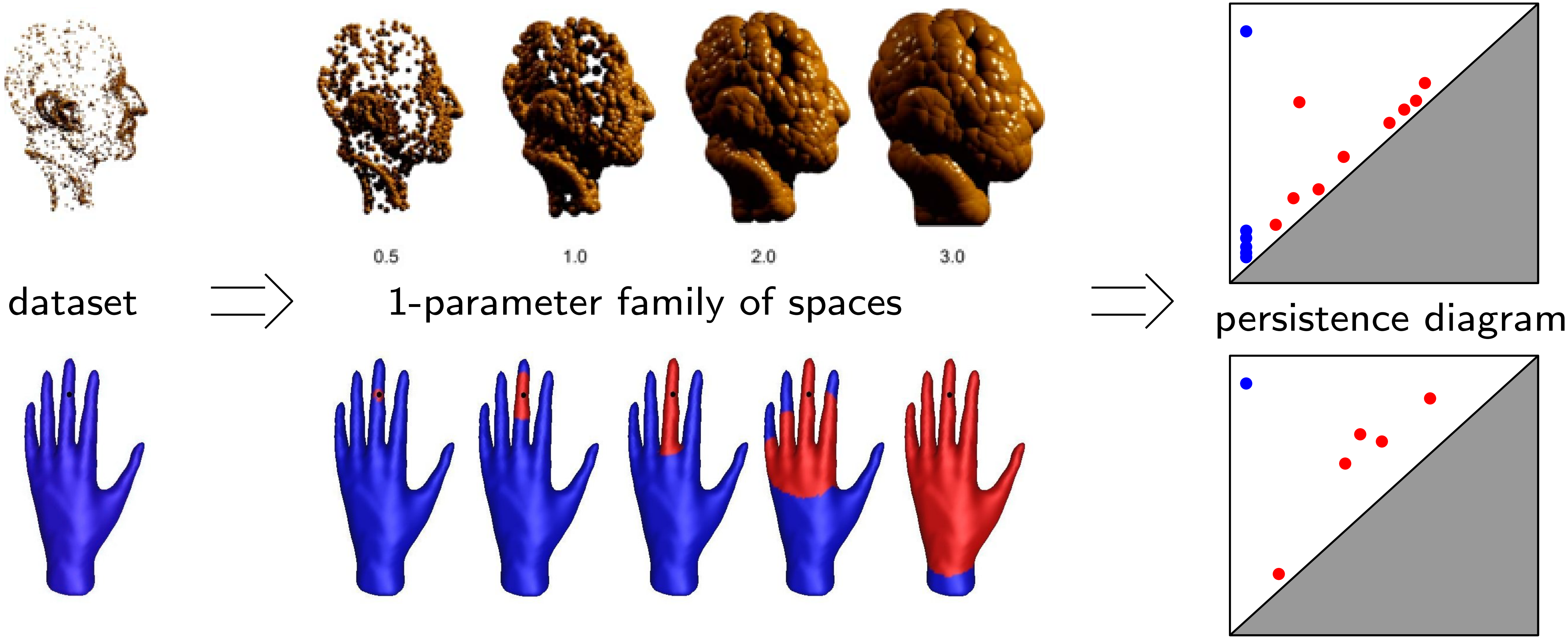
Learning with Persistence Diagrams

Persistence diagrams as descriptors for data



Pros:

- strong invariance and stability:

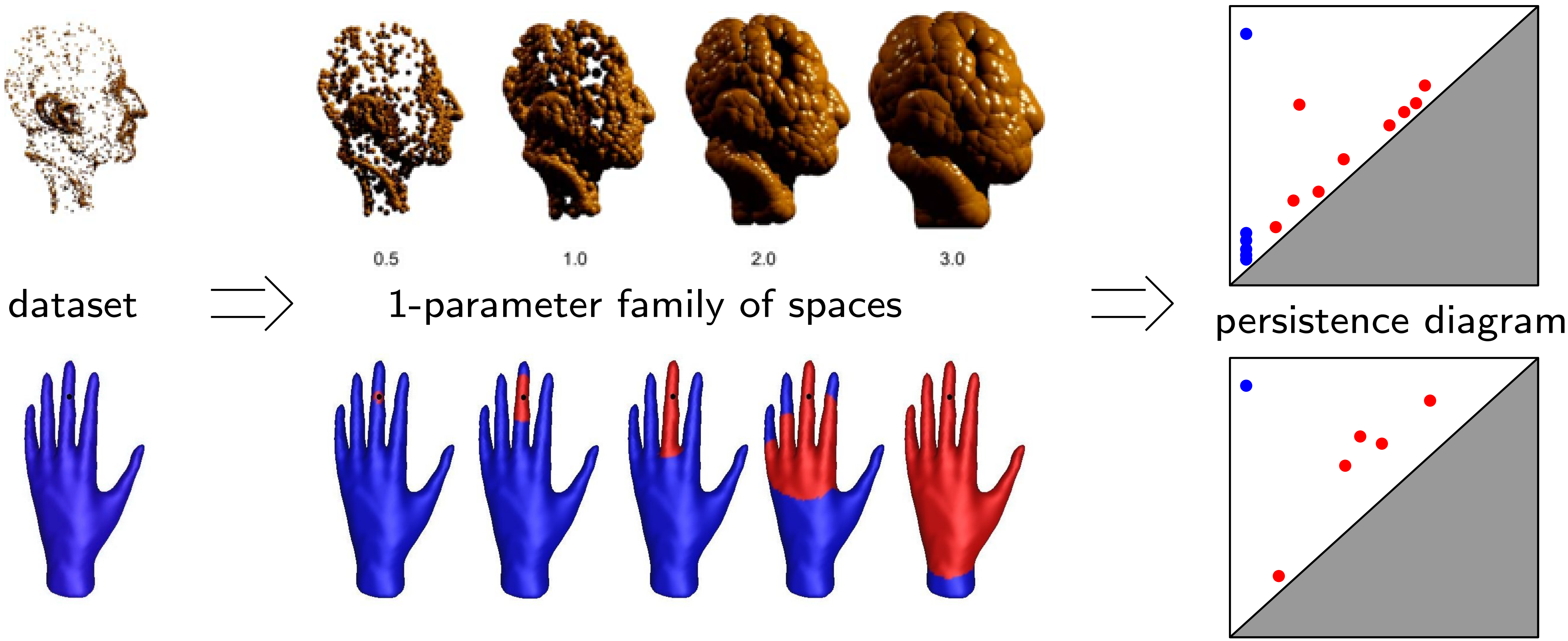
  $\mathrm{d}_p(\mathrm{dgm}\, X, \mathrm{dgm}\, Y) \leq \mathrm{cst}\, \mathrm{d}_{\mathrm{GH}}(X, Y)$

- information of a different nature
- flexible and versatile

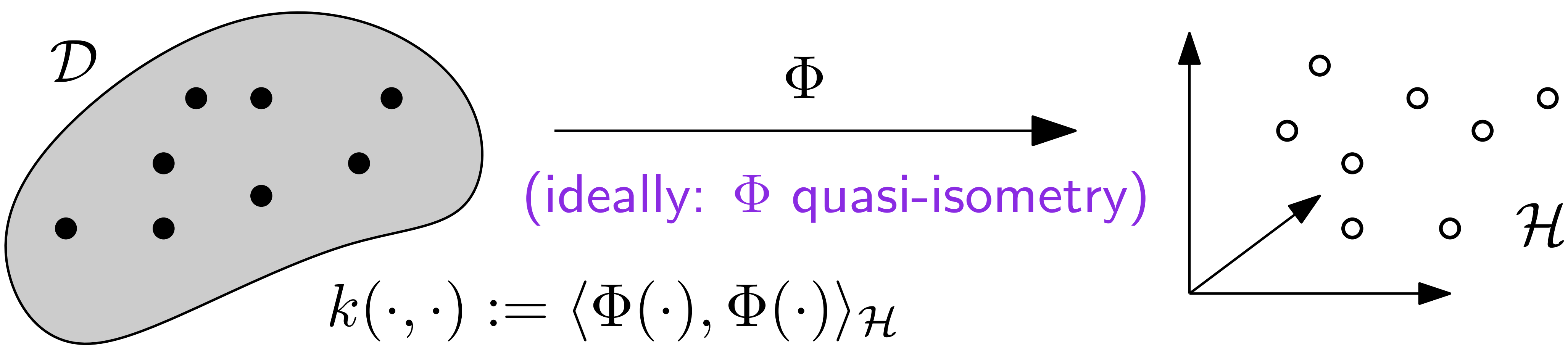
Cons:

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

Persistence diagrams as descriptors for data



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



Reproducing Kernel Hilbert Space

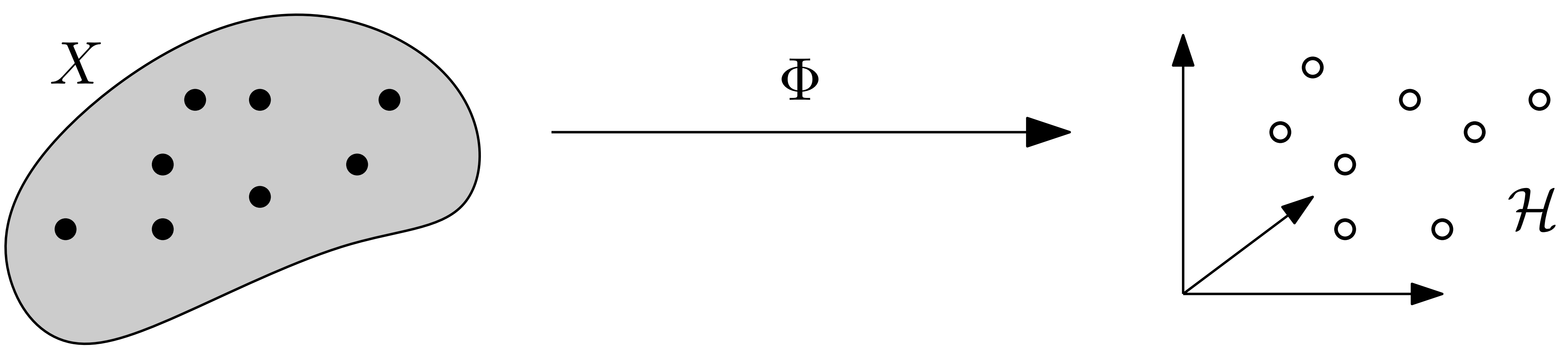
Def: Let $\mathcal{H} \subset \mathbb{R}^X$ Hilbert, with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$
Then, $\mathcal{H}$ is a **RKHS** on X if $\exists \Phi : X \to \mathcal{H}$ s.t.:

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

reproducing property

Terminology:

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



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

Thm: [Moore 1950] $k : X \times X \to \mathbb{R}$ is a kernel iff it is *positive (semi-)definite*, i.e. $\forall n \in \mathbb{N}$, $\forall x_1, \cdots, x_n \in X$, 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: $k(x, y) = (1 + \langle x, y \rangle)^N = \sum_{n_1 + \cdots + n_d = N} \binom{N}{n_1, \cdots, n_d} \underbrace{x_1^{n_1} \cdots x_d^{n_d}}_{\propto \Phi(x)} y_1^{n_1} \cdots y_d^{n_d}$
- Gaussian: $k(x, y) = \exp\left(-\frac{\|x-y\|_2^2}{2\sigma^2}\right)$, $\sigma > 0$. $\quad \mathcal{H} \subset L_2(\mathbb{R}^d)$

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

Thm: [Moore 1950] $k : X \times X \to \mathbb{R}$ is a kernel iff it is *positive (semi-)definite*, i.e. $\forall n \in \mathbb{N}$, $\forall x_1, \cdots, x_n \in X$, the Gram matrix $(k(x_i, x_j))_{i,j}$ is positive semi-definite.

Thm: (Representer) [Schölkopf et al 2001]
Given RKHS $\mathcal{H}$ with kernel k, any function $f^* \in \mathcal{H}$ minimizing

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

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

Kernels for persistence diagrams

Three approaches:

- build kernel from kernels (algebraic operations)

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

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

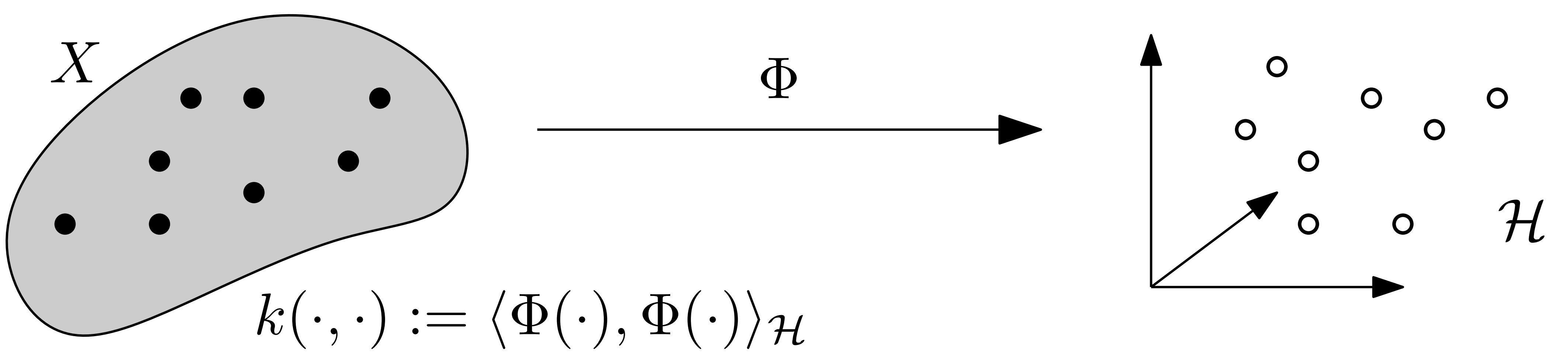
- **product of kernels** $\longleftrightarrow$ **tensor product of feature spaces**

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

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

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: [Kimeldorf, Wahba 1971]

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

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

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

Kernels for persistence diagrams

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: [Kimeldorf, Wahba 1971]

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

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

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:

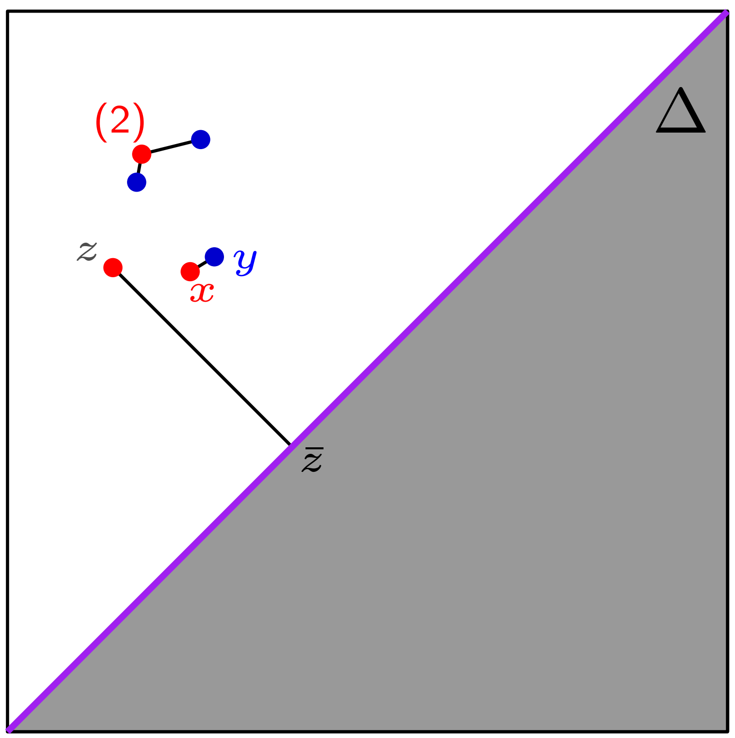
$$c_p(M) := \left(\sum_{(x,y) \text{ matched}} c_p(x,y) + \sum_{z \text{ unmatched}} c_p(z) \right)^{1/p}$$

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

$$\mathrm{d}_p(X,Y) := \inf_{M:X\leftrightarrow Y} c_p(M)$$

Def: bottleneck distance:

$$\mathrm{d}_\infty(X,Y) := \lim_{p\to\infty} \mathrm{d}_p(X,Y)$$



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 $(x, y) \in M$: $c_p(x, y)$

cost of an unmatched point $z \in X \sqcup Y$: $c_p($

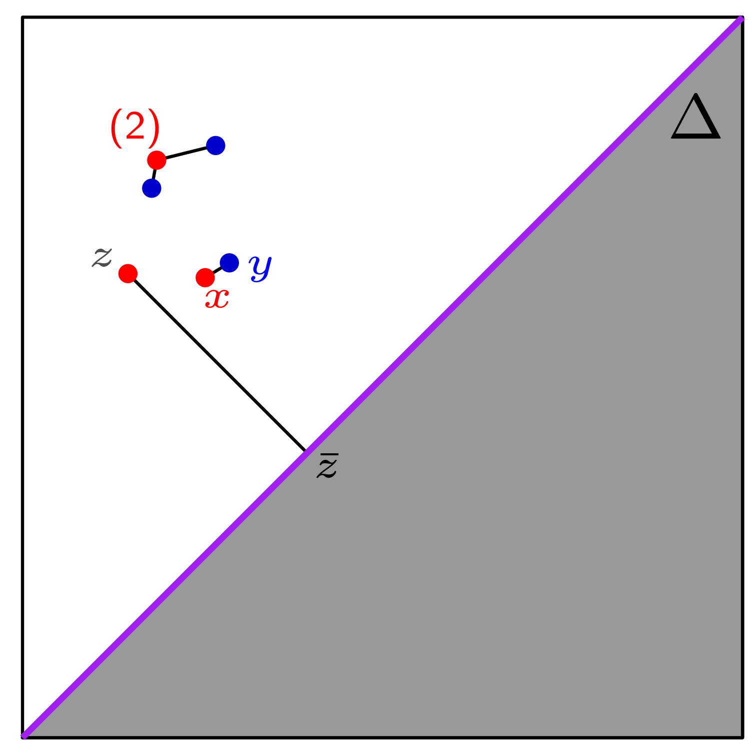
unbalanced optimal transport

d_p is **NOT** cnsd, $\forall p \in \mathbb{R}_{>0} \cup \{\infty\}$

$\Rightarrow$ previous theorem is not applicable

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):

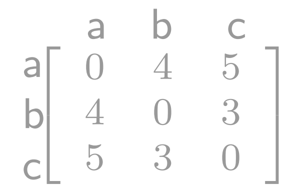
$$\mathrm{d}_p(X, Y) := \inf_{M : X \leftrightarrow Y} c_p(M)$$

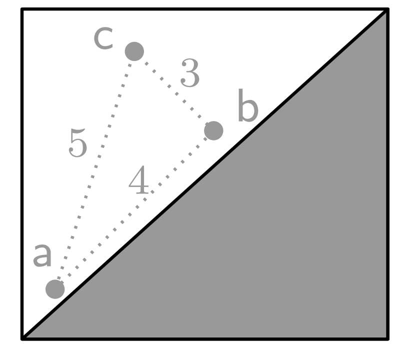
Def: bottleneck distance:

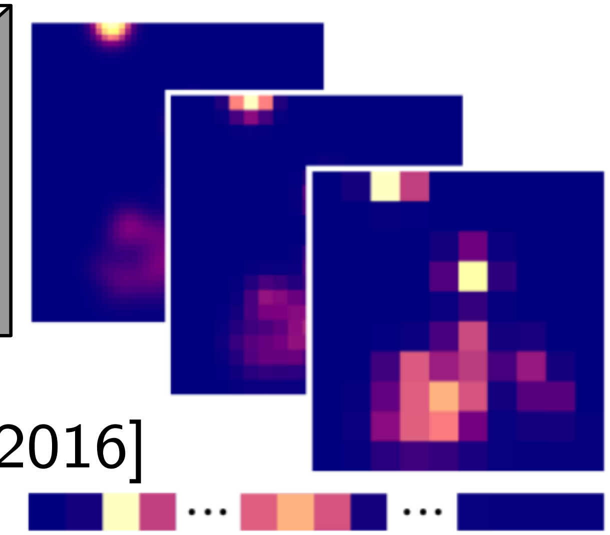
$$\mathrm{d}_\infty(X, Y) := \lim_{p \to \infty} \mathrm{d}_p(X, Y)$$

Kernels for persistence diagrams

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



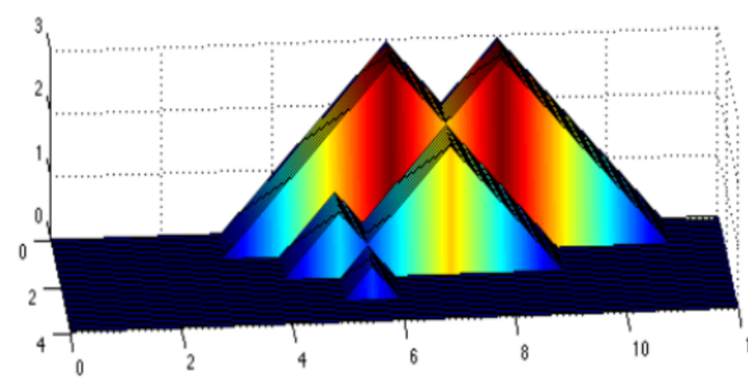




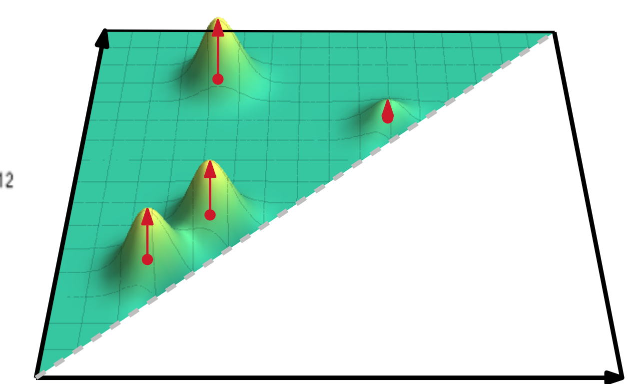
- images [Adams et al. 2015]
- finite metric spaces [Carrière, O., Ovsjanikov 2015]
- polynomial roots or evaluations [Di Fabio, Ferri 2015] [Kališnik 2016]

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

- 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]

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	✓	✓	✓	✓	✓
$\|\phi(\cdot) - \phi(\cdot)\|_{\mathcal{H}} \leq C\, \mathrm{d}_p$	✓	✓	✓	✓	✓
$\|\phi(\cdot) - \phi(\cdot)\|_{\mathcal{H}} \geq c\, \mathrm{d}_p$	✗	✗	✗	✗	✗
injectivity	✗	✗	✓	✓	✓
universality	✗	✗	✗	✗	✓
algorithmic cost	f. map: $O(n^2)$ kernel: $O(d)$	f. map: $O(n^2)$ kernel: $O(d)$	f. map: $O(nd)$ kernel: $O(d)$	$O(n^2)$	$O(n^2)$

Kernels for persistence diagrams

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

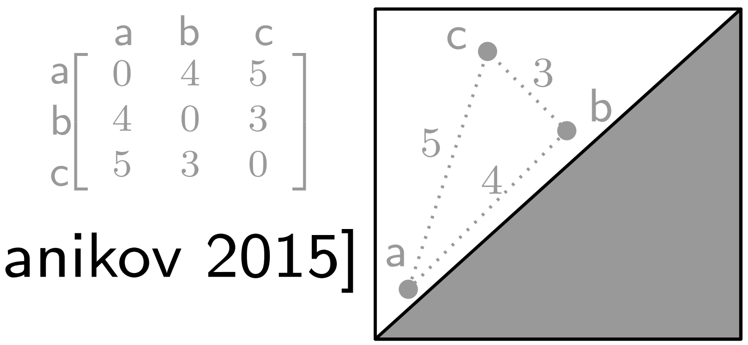
Kernels for persistence diagrams

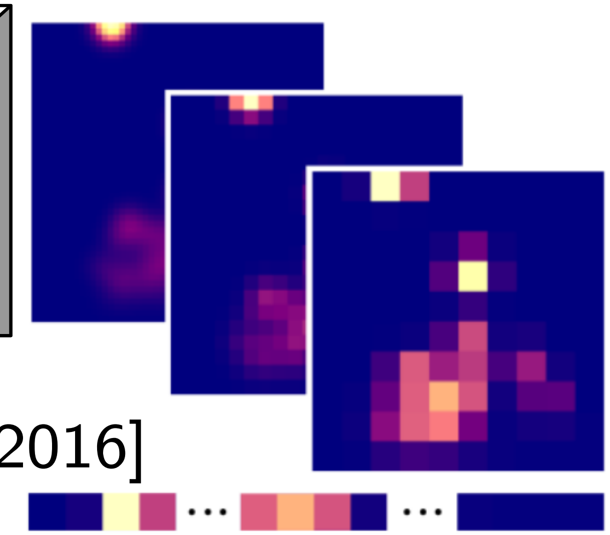
	images	metric spaces	polynomials	landscapes	discrete measures
ambient Hilbert space	$(\mathbb{R}^d, \Vert.\Vert_2)$	$(\mathbb{R}^d, \Vert.\Vert_2)$	$\ell_2(\mathbb{R})$	$L_2(\mathbb{N} \times \mathbb{R})$	$L_2(\mathbb{R}^2)$
positive (semi-)definiteness	✓	✓	✓	✓	✓
$\Vert\phi(\cdot) - \phi(\cdot)\Vert_{\mathcal{H}} \leq C\,\mathrm{d}_p$	✓	✓	✓	✓	✓
$\Vert\phi(\cdot) - \phi(\cdot)\Vert_{\mathcal{H}} \geq c\,\mathrm{d}_p$	✗	✗	✗	✗	✗
injectivity	✗	✗	✓	✓	✓
universality	✗	✗	✗	✗	✓
algorithmic cost	f. map: $O(n^2)$ kernel: $O(d)$	f. map: $O(n^2)$ kernel: $O(d)$	f. map: $O(nd)$ kernel: $O(d)$	$O(n^2)$	$O(n^2)$

Kernels for persistence diagrams

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

- images [Adams et al. 2015]



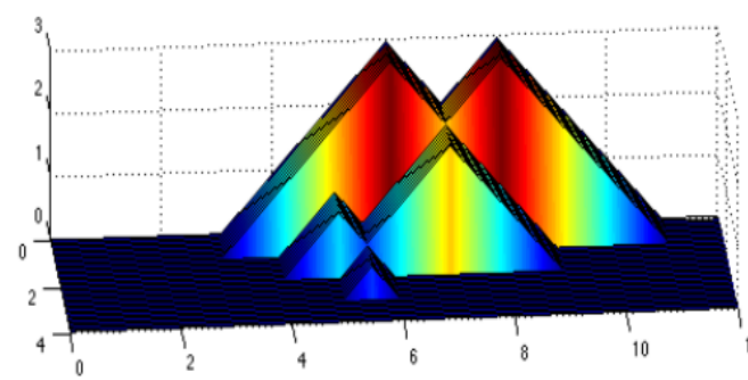


- finite metric spaces [Carrière, O., Ovsjanikov 2015]

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

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

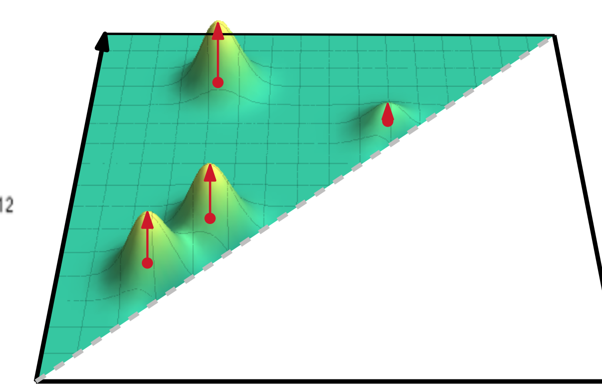
- landscapes [Bubenik 2012] [Bubenik, Dłotko 2015]



- discrete measures:

  → histogram [Bendich et al. 2014]

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

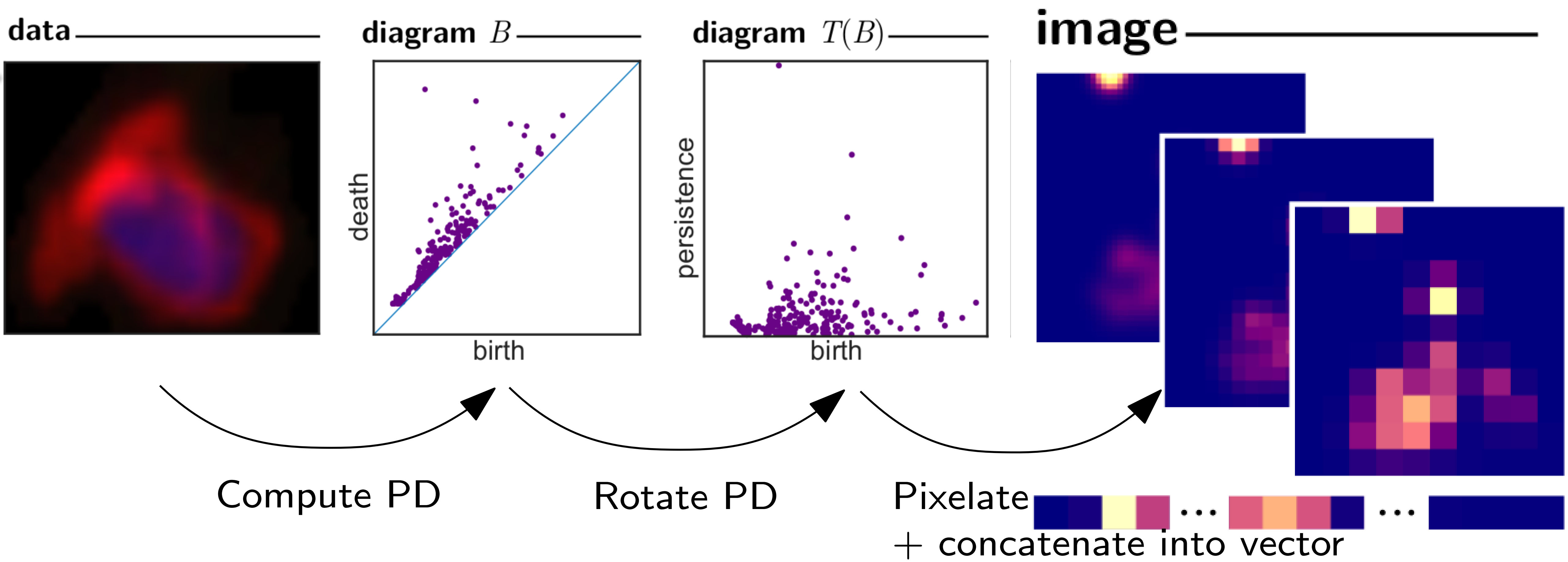


  → convolution with weighted kernel [Kusano, Fukumisu, Hiraoka 2016-17]

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

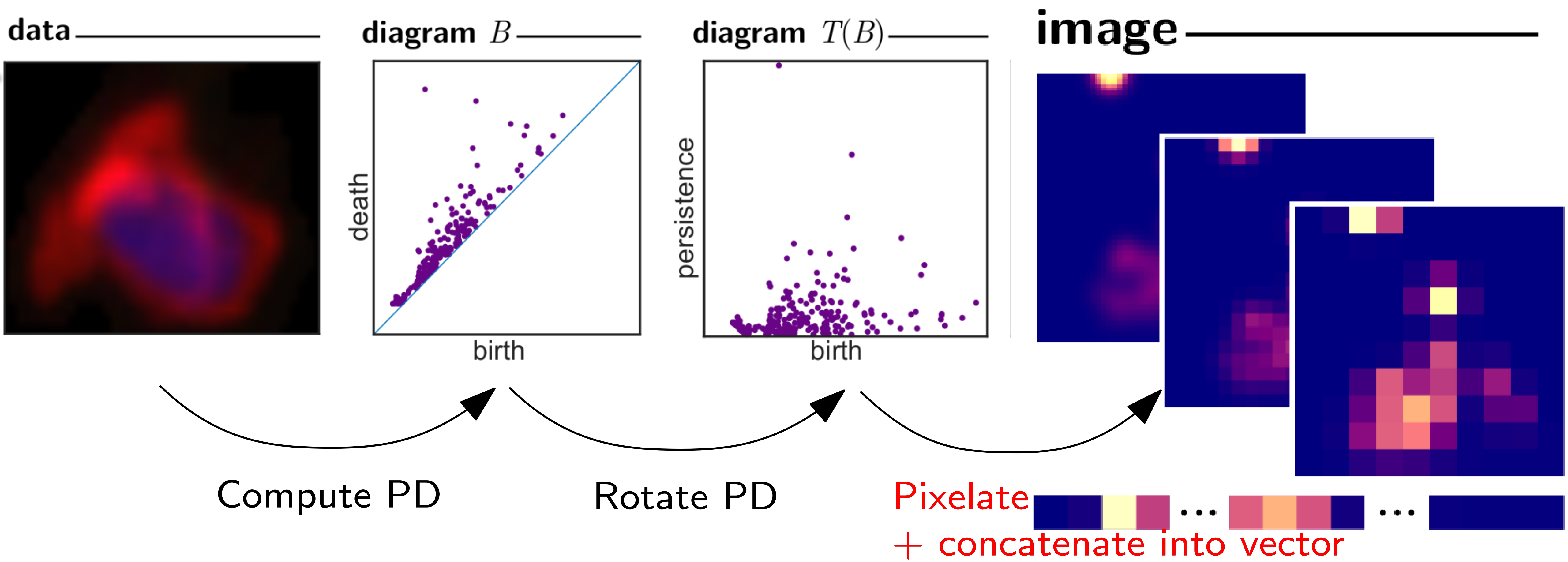
Explicit Feature Map in $\mathbb{R}^d$

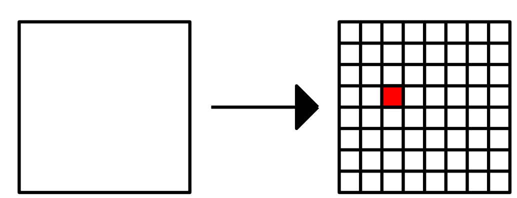
Persistence Images [Adams et al. 2017]



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

Persistence Images [Adams et al. 2017]



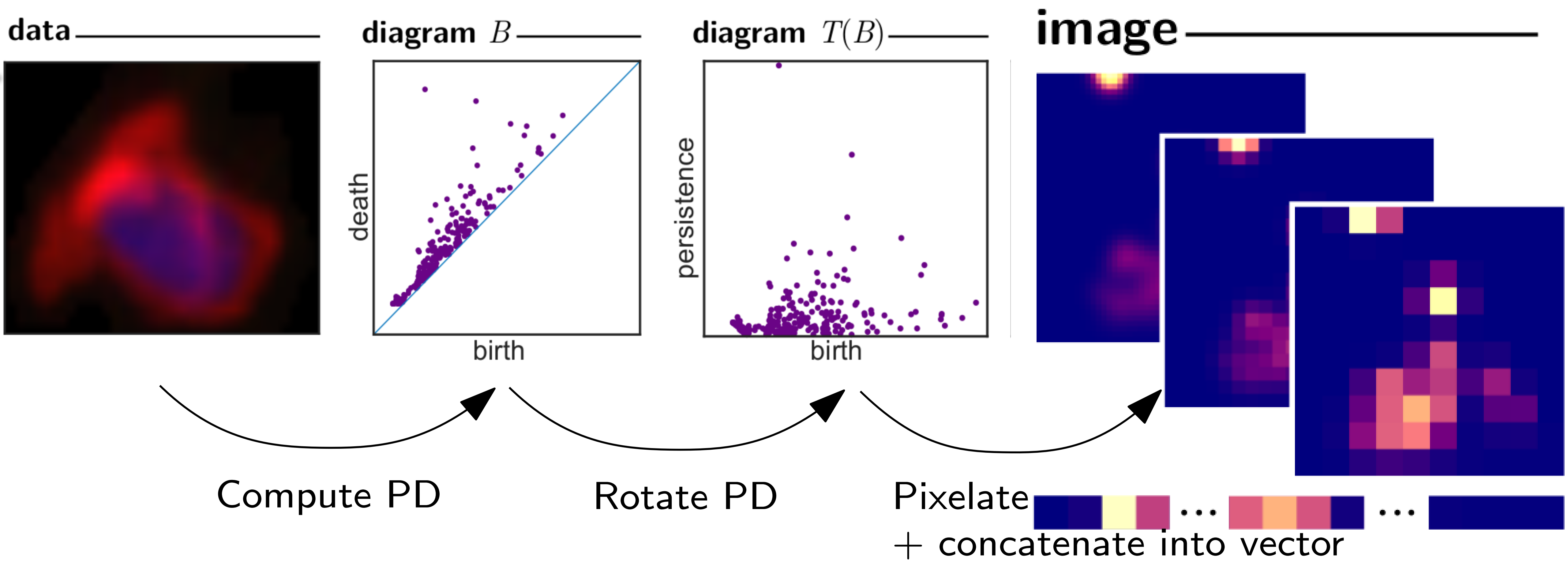
Discretize plane into one or several grid(s): 

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

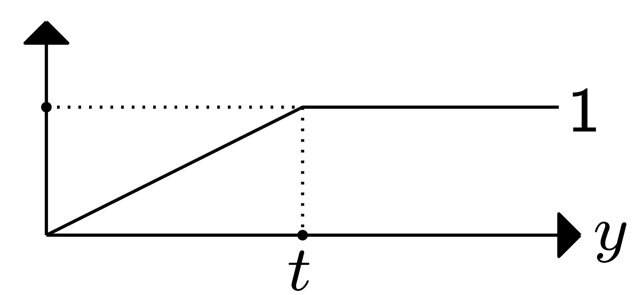
Concatenate all $I(P)$ into a single vector $\mathrm{PI}(\mathrm{dgm})$

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

Persistence Images [Adams et al. 2017]



Stability → weigh points: $w_t(x, y) =$

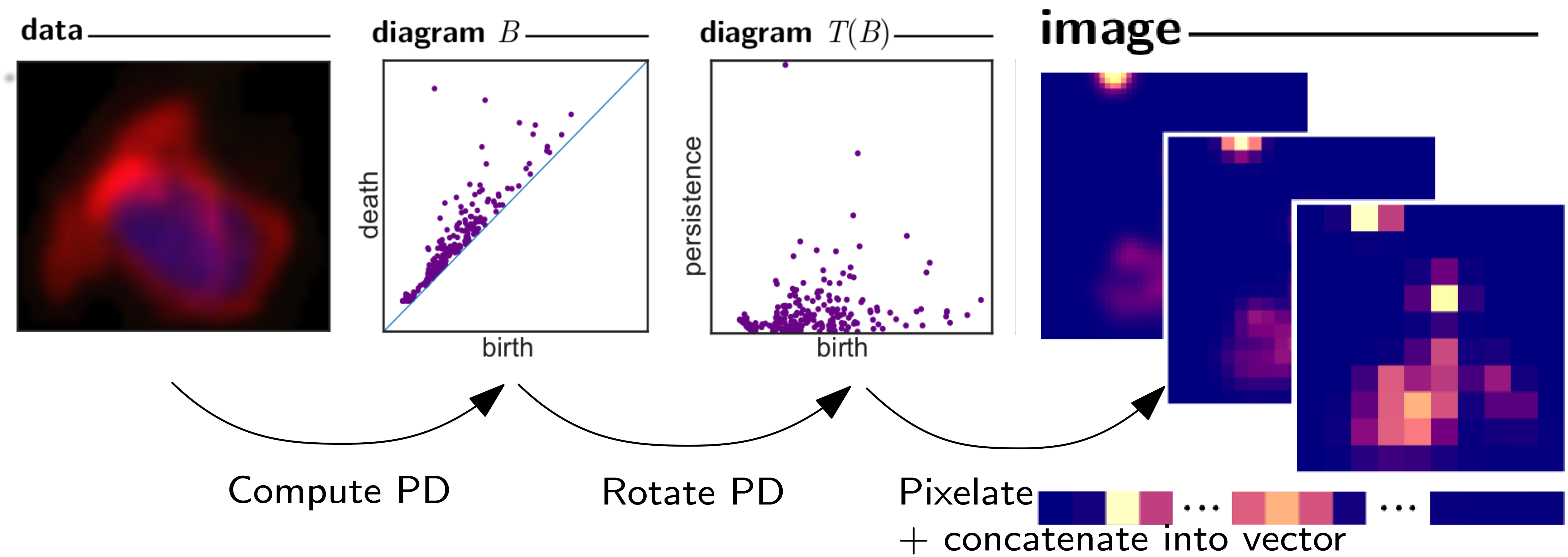


→ blur image

(convolve with Gaussian)

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

Persistence Images [Adams et al. 2017]

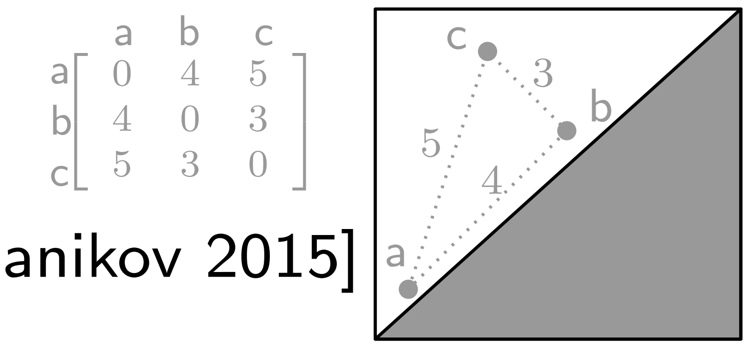


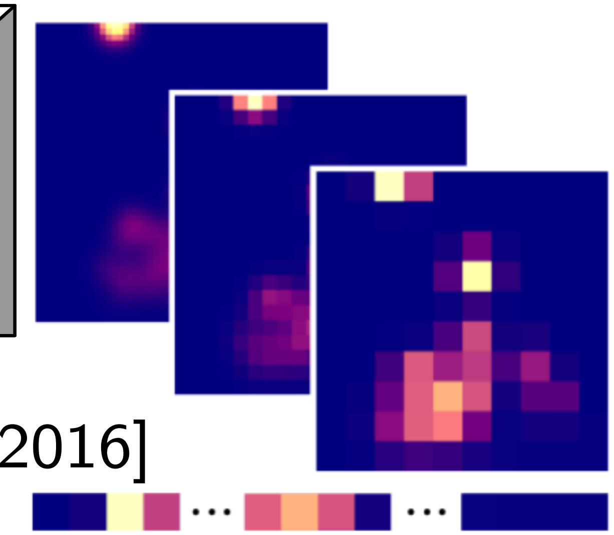
Prop: [Adams et al. 2017]

- $\|\mathrm{PI}(\mathrm{dgm}) - \mathrm{PI}(\mathrm{dgm}')\|_\infty \leq C(w, \phi_p)\, \mathrm{d}_1(\mathrm{dgm}, \mathrm{dgm}')$
- $\|\mathrm{PI}(\mathrm{dgm}) - \mathrm{PI}(\mathrm{dgm}')\|_2 \leq \sqrt{d}C(w, \phi_p)\, \mathrm{d}_1(\mathrm{dgm}, \mathrm{dgm}')$

Kernels for persistence diagrams

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

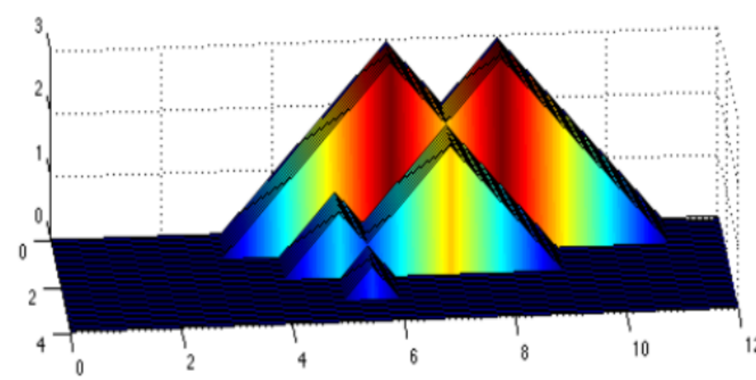




- images [Adams et al. 2015]
- finite metric spaces [Carrière, O., Ovsjanikov 2015]
- polynomial roots or evaluations [Di Fabio, Ferri 2015] [Kališnik 2016]

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

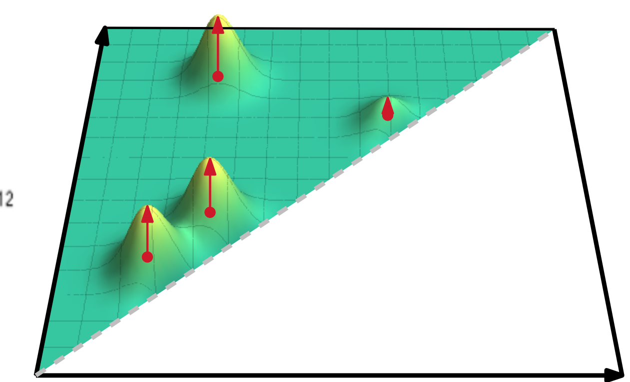
- landscapes [Bubenik 2012] [Bubenik, Dłotko 2015]



- discrete measures:

  → histogram [Bendich et al. 2014]

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

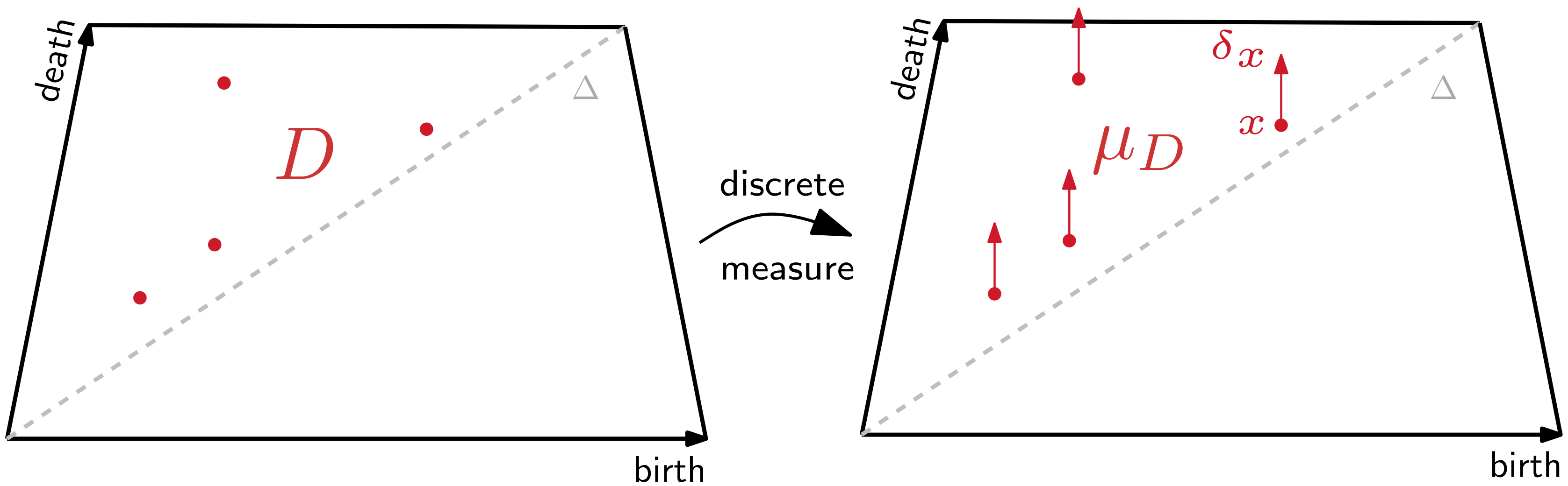


  → convolution with weighted kernel [Kusano, Fukumisu, Hiraoka 2016-17]

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

Explicit feature map

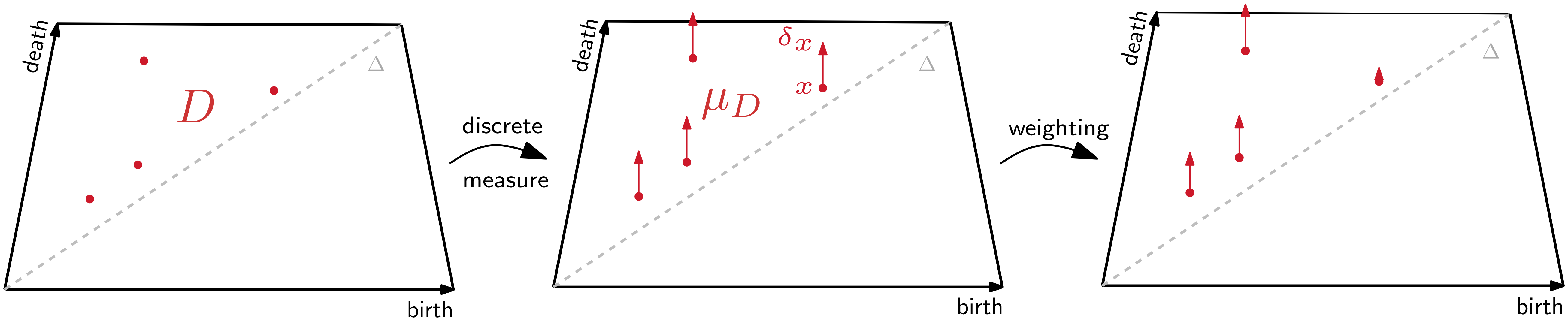
Persistence diagrams as discrete measures:



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

Explicit feature map

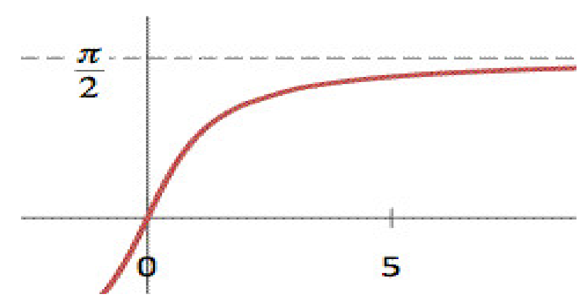
Persistence diagrams as discrete measures:



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

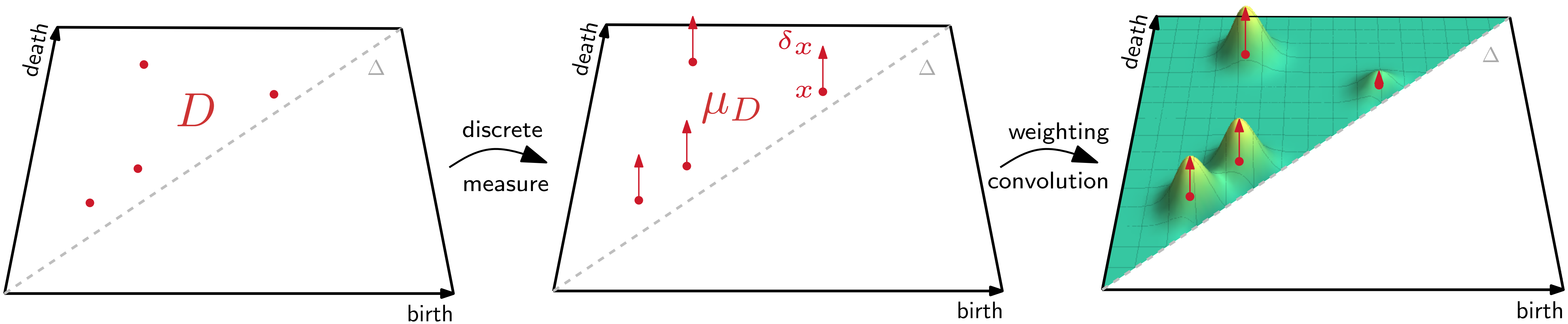
Pb: μ_D is unstable (points on diagonal disappear)

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



Explicit feature map

Persistence diagrams as discrete measures:



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

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

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

Pb: μ_D is unstable (points on diagonal disappear)

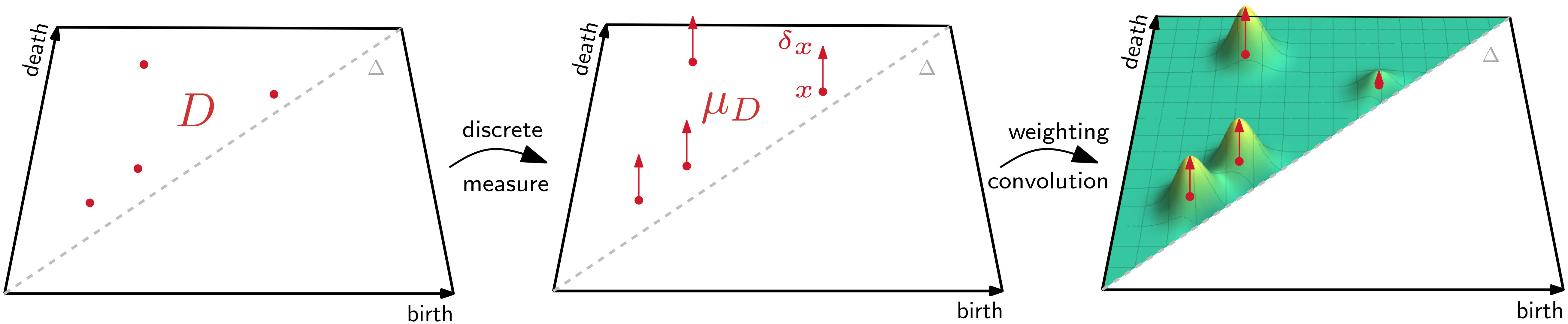
$$w(x) := \arctan\left(c\, \mathrm{d}(x, \Delta)^r\right),\; c, r > 0$$

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

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

Explicit feature map

Persistence diagrams as discrete measures:



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

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

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

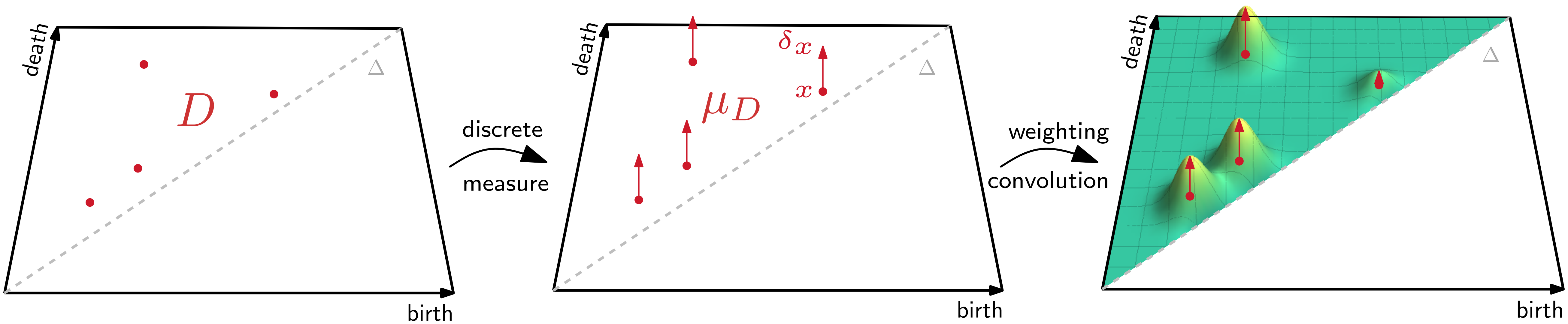
Prop.: [Kusano, Fukumisu, Hiraoka 2016-17]

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

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

Explicit feature map

Persistence diagrams as discrete measures:



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

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

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

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

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

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

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

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	✓	✓	✓	✓	✓
$\|\phi(\cdot) - \phi(\cdot)\|_{\mathcal{H}} \leq C\, \mathrm{d}_p$	✓	✓	✓	✓	✓
$\|\phi(\cdot) - \phi(\cdot)\|_{\mathcal{H}} \geq c\, \mathrm{d}_p$	✗	✗	✗	✗	✗
injectivity	✗	✗	✓	✓	✓
universality	✗	✗	✗	✗	✓
algorithmic cost	f. map: $O(n^2)$ kernel: $O(d)$	f. map: $O(n^2)$ kernel: $O(d)$	f. map: $O(nd)$ kernel: $O(d)$	$O(n^2)$	$O(n^2)$

One kernel to rule them all...

Sliced Wasserstein Kernel [Carrière, Cuturi, O. 2017]

No feature map

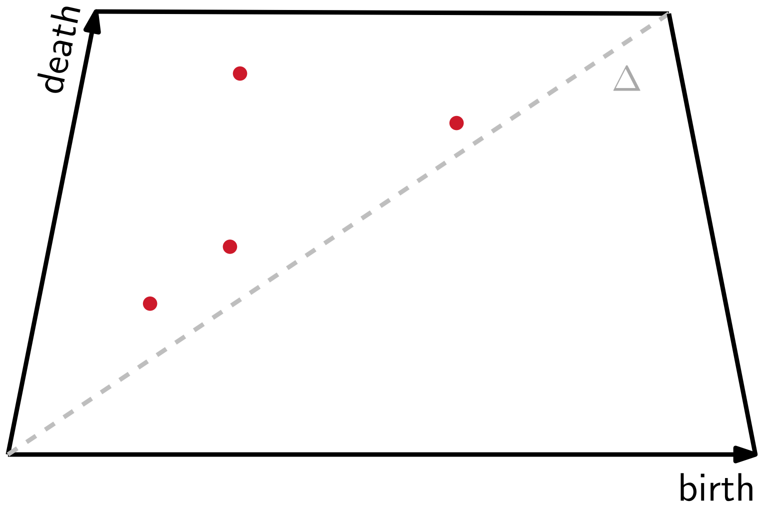
Provably stable

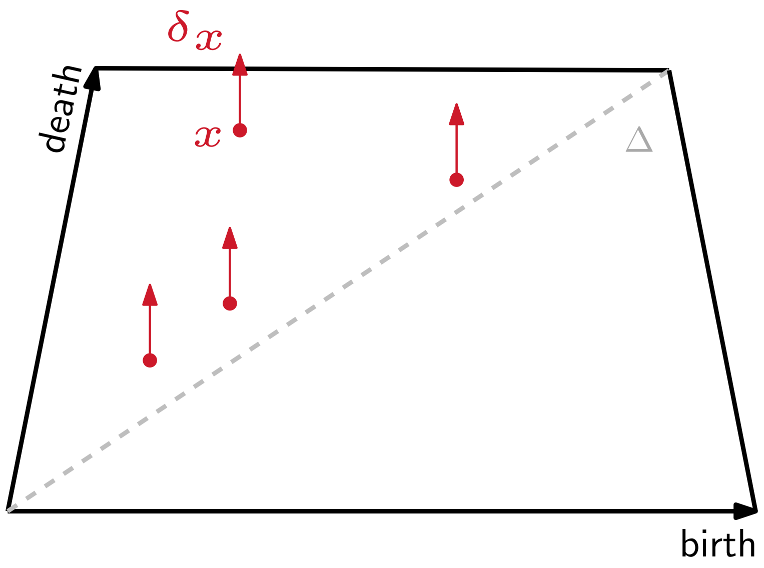
Provably discriminative

Mimicks the Gaussian kernel

View diagrams as discrete measures w/o density functions

Persistence diagrams as discrete measures (II)

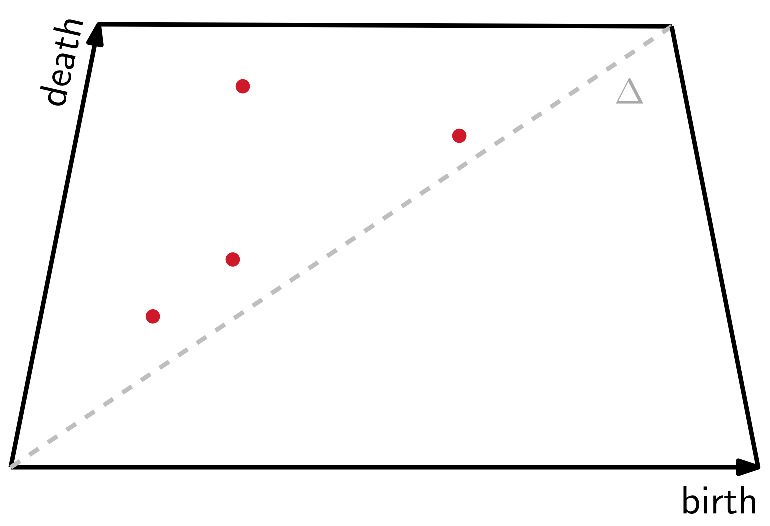


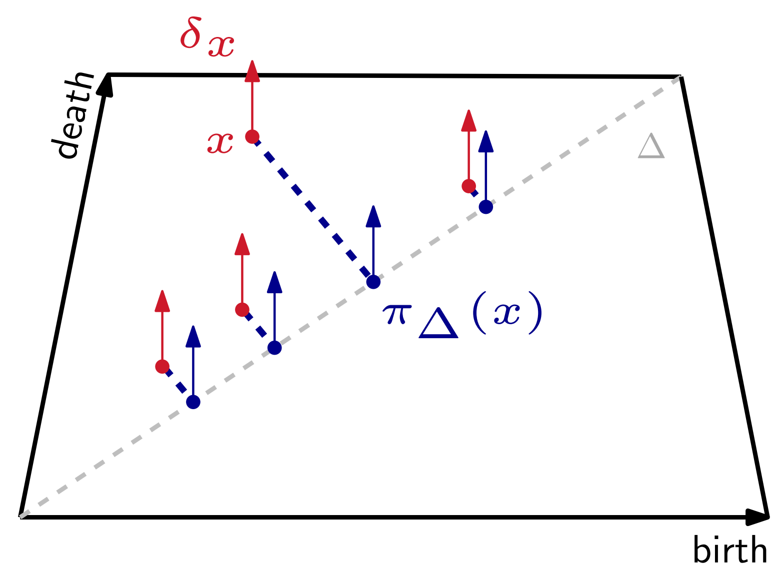


$\mu_D := \sum_{x \in D} \delta_x$

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

Persistence diagrams as discrete measures (II)





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

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

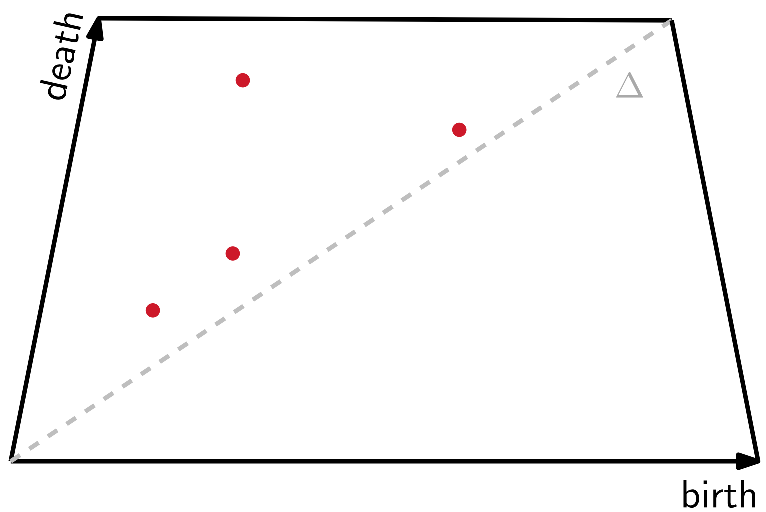
$\rightarrow$ given D, D', let

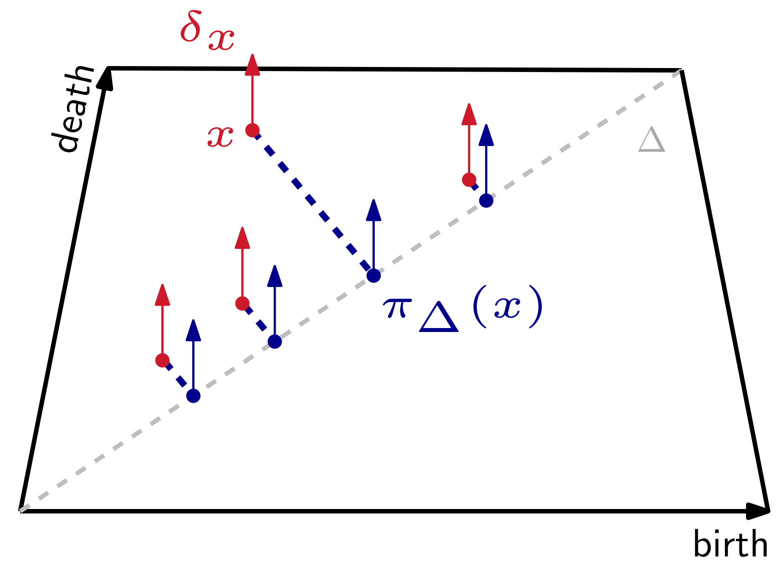
$$\bar\mu_D := \sum_{x\in D}\delta_x + \sum_{y\in D'}\delta_{\pi_\Delta(y)}$$

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

Then, $\mathrm{d}_p(D,D') \leq W_p(\bar\mu_D,\bar\mu_{D'}) \leq 2\,\mathrm{d}_p(D,D')$

Persistence diagrams as discrete measures (II)





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

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

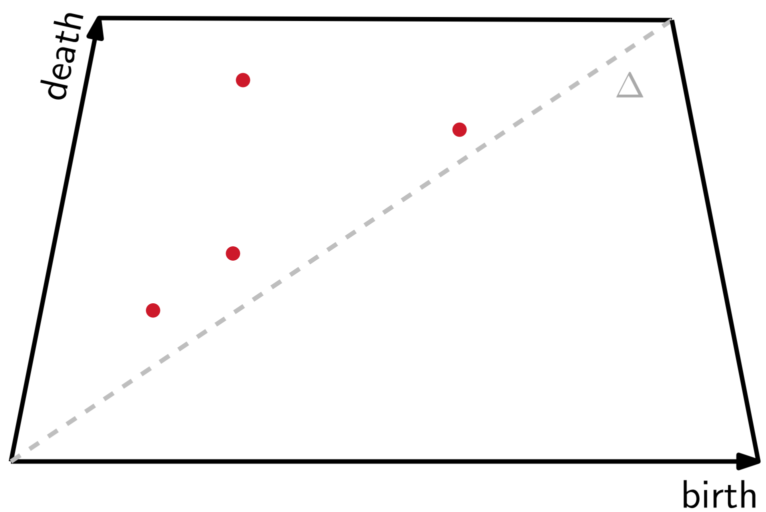
$\rightarrow$ given D, D', let

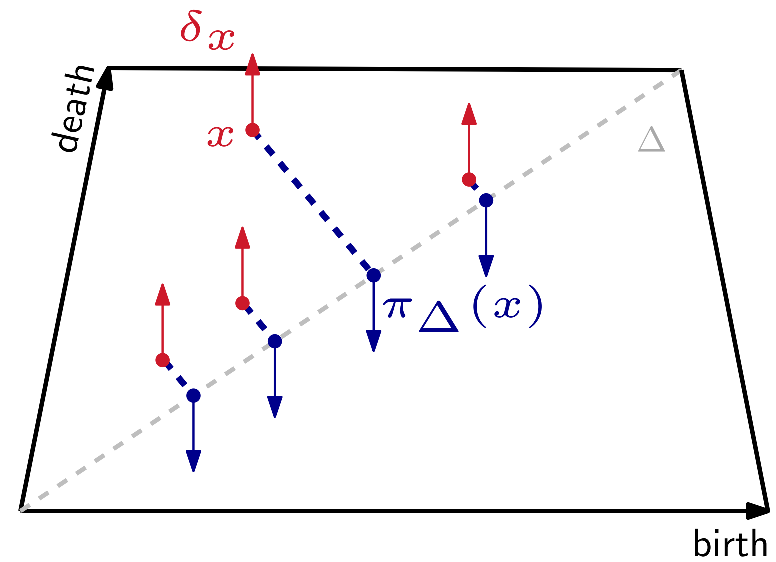
$$\bar{\mu}_D := \sum_{x \in D} \delta_x + \sum_{y \in D'} \delta_{\pi_\Delta(y)}$$

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

Then, $\mathrm{d}_p(D, D') \leq W_p(\bar{\mu}_D, \bar{\mu}_{D'}) \leq 2\, \mathrm{d}_p(D, D')$ **Pb:** $\bar{\mu}_D$ depends on D'

Persistence diagrams as discrete measures (II)





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

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

Solution: transfer mass negatively:

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

$\rightarrow$ signed discrete measure of total mass zero

Kantorovich norm: $\|\tilde{\mu}_D\|_K = W_1(\mu_D, (\pi_\Delta)_* \, \mu_D)$

A Wasserstein Gaussian kernel for PDs?

Thm.: [Kimeldorf, Wahba 1971]
If $d : X \times X \to \mathbb{R}_+$ symmetric is *conditionally negative semidefinite*, i.e.:

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

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)

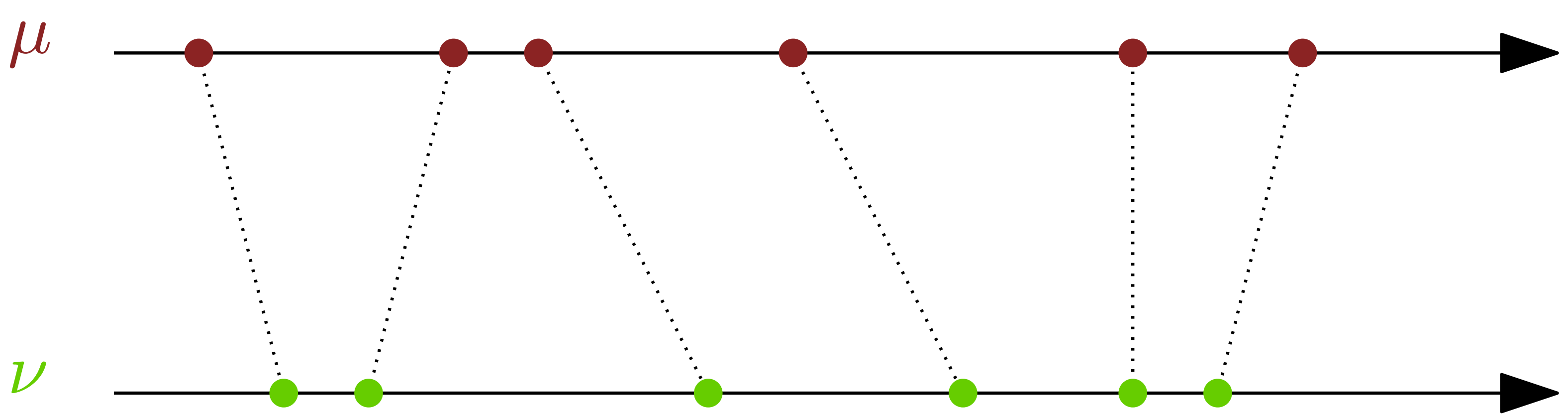
Sliced Wasserstein metric

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

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

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: $W_1(\mu, \nu) = \sum_{i=1}^{m} |x_i - y_i| = \|(x_1, \cdots, x_m) - (y_1, \cdots, y_m)\|_1$



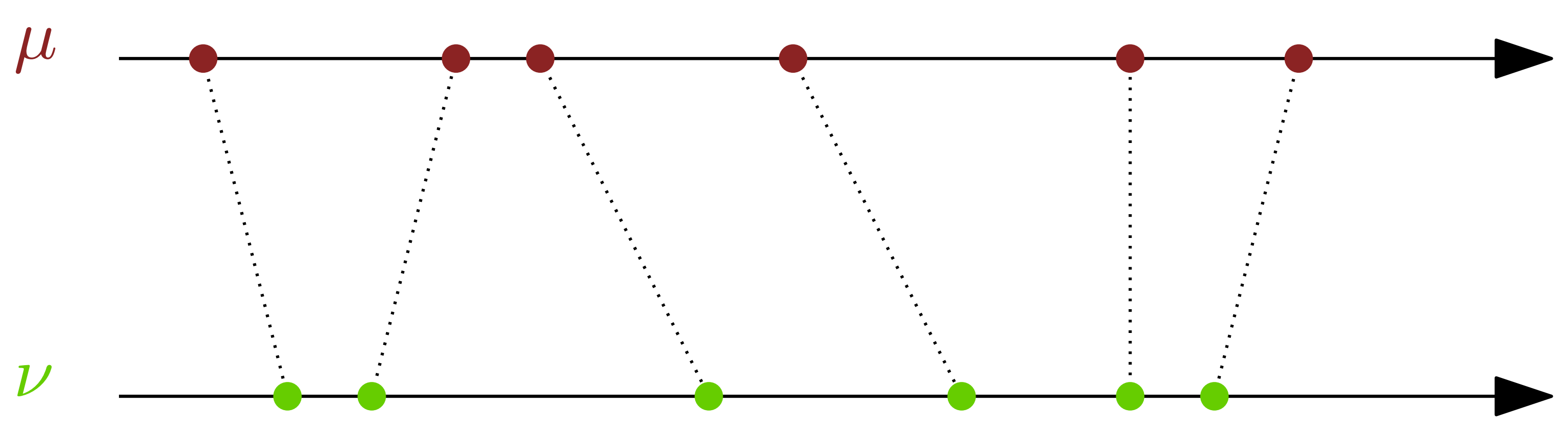
Sliced Wasserstein metric

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

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

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: $W_1(\mu, \nu) = \sum_{i=1}^{m} |x_i - y_i| = \|(x_1, \cdots, x_m) - (y_1, \cdots, y_m)\|_1$



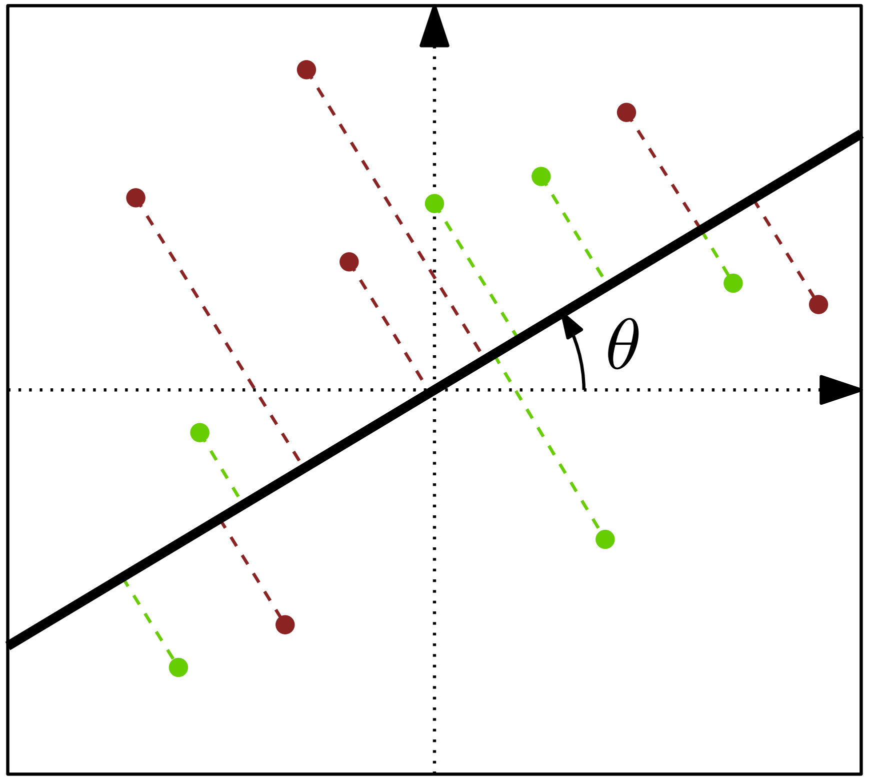
→ W_1 is cnsd and easy to compute (same with $\|\cdot\|_K$ for signed measures)

Sliced Wasserstein metric

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 π_θ = orthogonal projection onto line passing through origin with angle θ.



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

Sliced Wasserstein metric

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 $\mathbb{R}$) [Rabin, Peyré, Delon, Bernot 2011]

- satisfies the axioms of a metric

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

- conditionally negative semidefinite

Sliced Wasserstein kernel

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

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

Corollary: [Kolouri, Zou, Rohde]
k_{SW} is positive semidefinite.

Sliced Wasserstein kernel

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

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

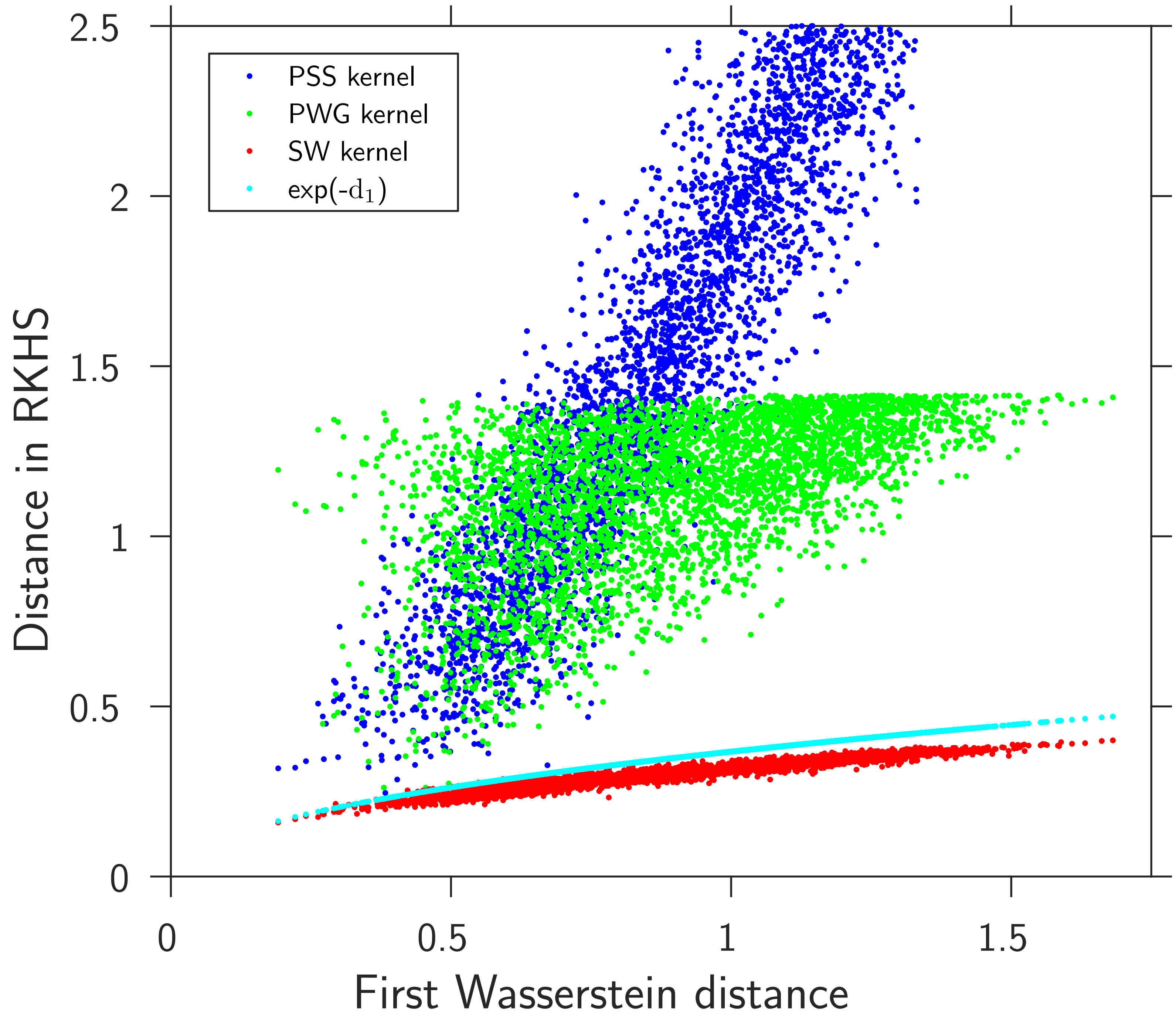
Corollary: [Kolouri, Zou, Rohde]
k_{SW} is positive semidefinite.

Thm.: [Carrière, Cuturi, O. 2017]
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 $D, D' \in \mathcal{D}_N$,

$$\frac{1}{2+4N(2N-1)}\, d_1(D,D') \;\leq\; SW_1(D,D') \;\leq\; 2\sqrt{2}\, d_1(D,D')$$

Corollary: 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

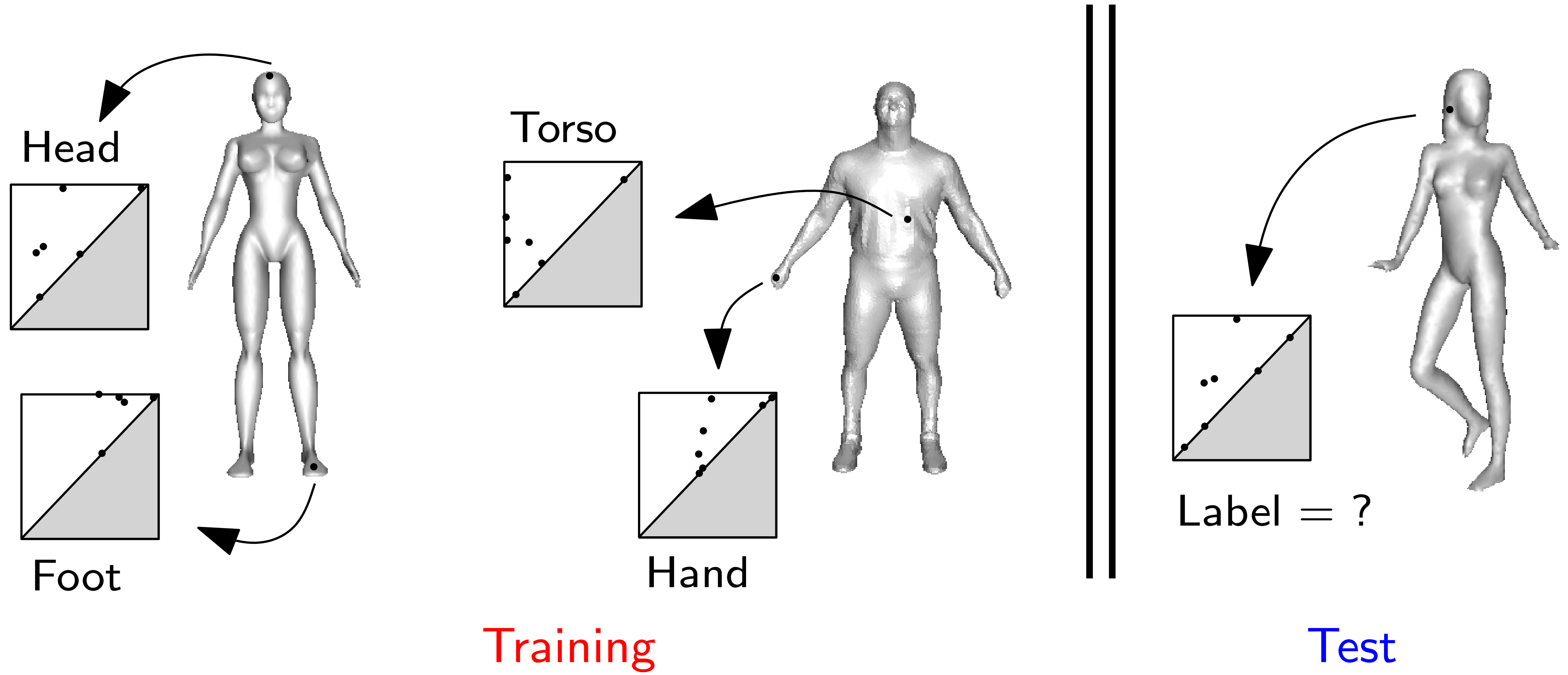


Application to supervised shape segmentation

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

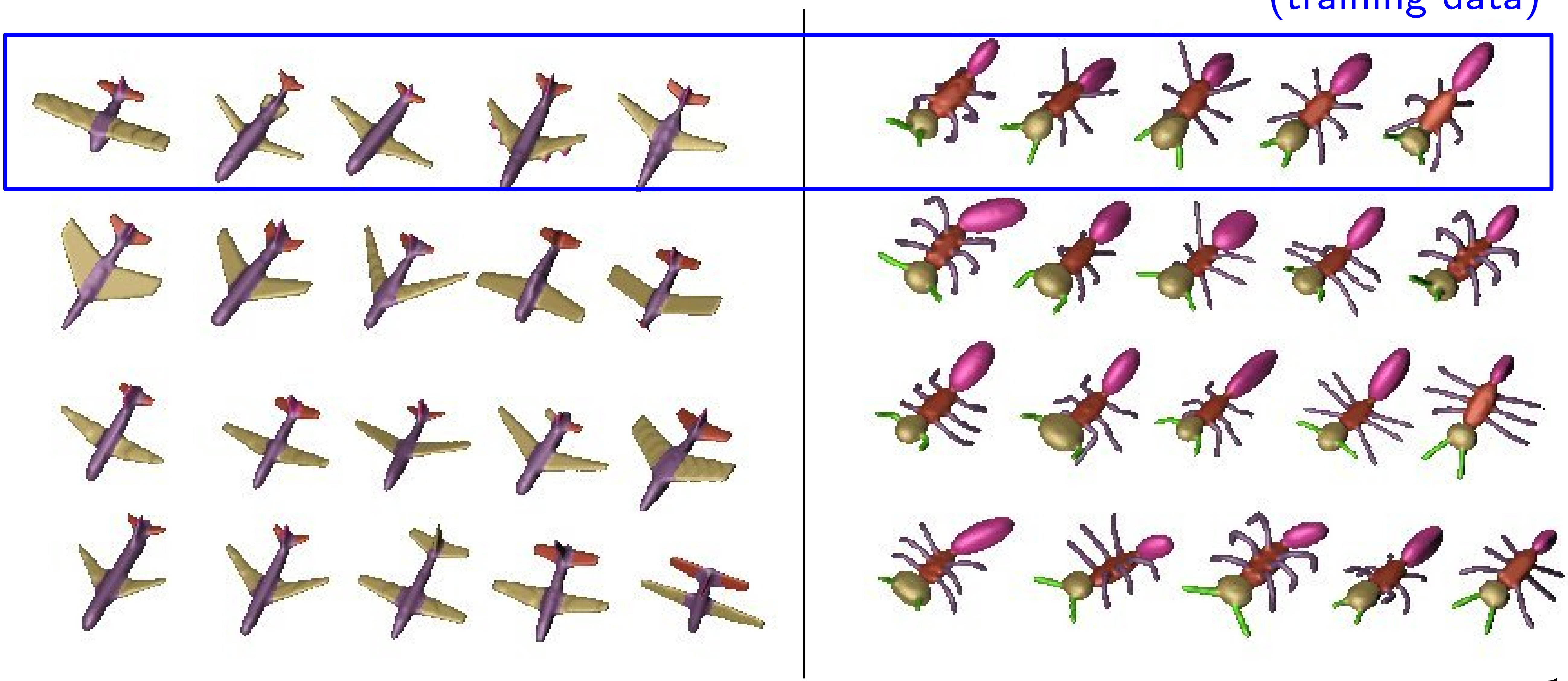


Application to supervised shape segmentation

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

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

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

Application to non-rigid shape matching

Approach: use framework of *functional maps* [Ovsjanikov et al. 2012]

Given a point-to-point map $m : X \to Y$ (seen as measured spaces), consider

the **linear map** $m^* : L^2(Y) \to L^2(X)$ induced by composition with m

- compute an optimal linear map that best preserves a set of signatures (vectors)

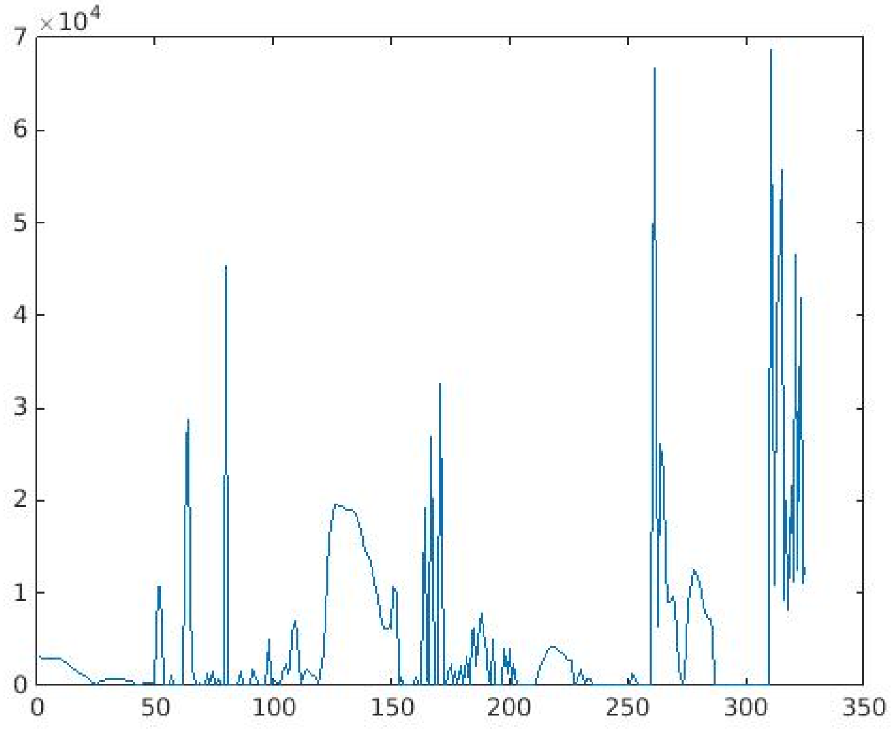
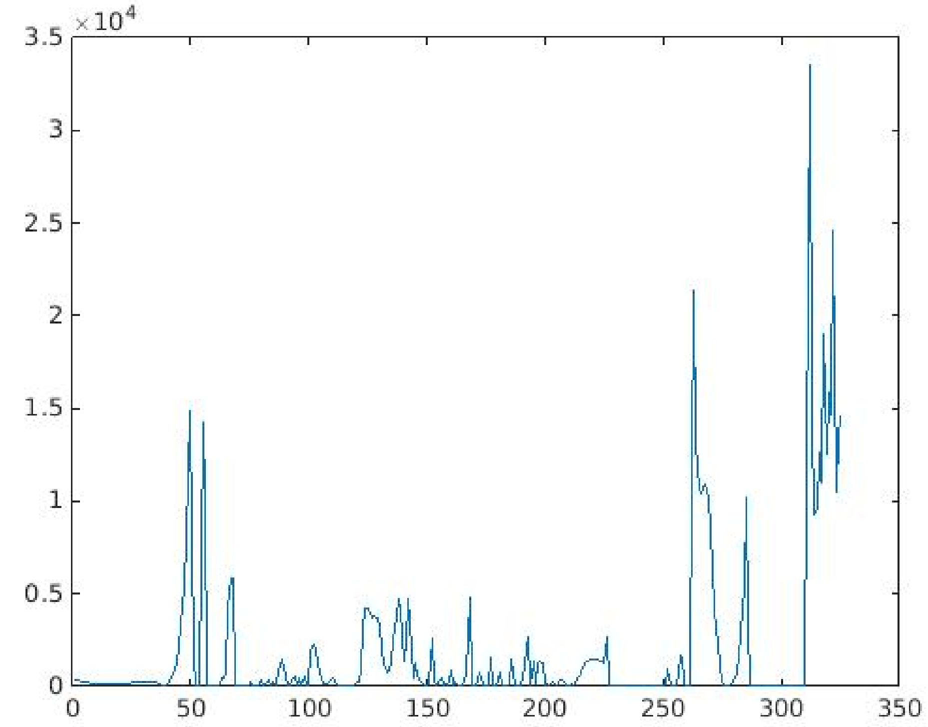
- derive a point-to-point correspondence from this map (via indicator functions)

- evaluate the quality of the correspondence

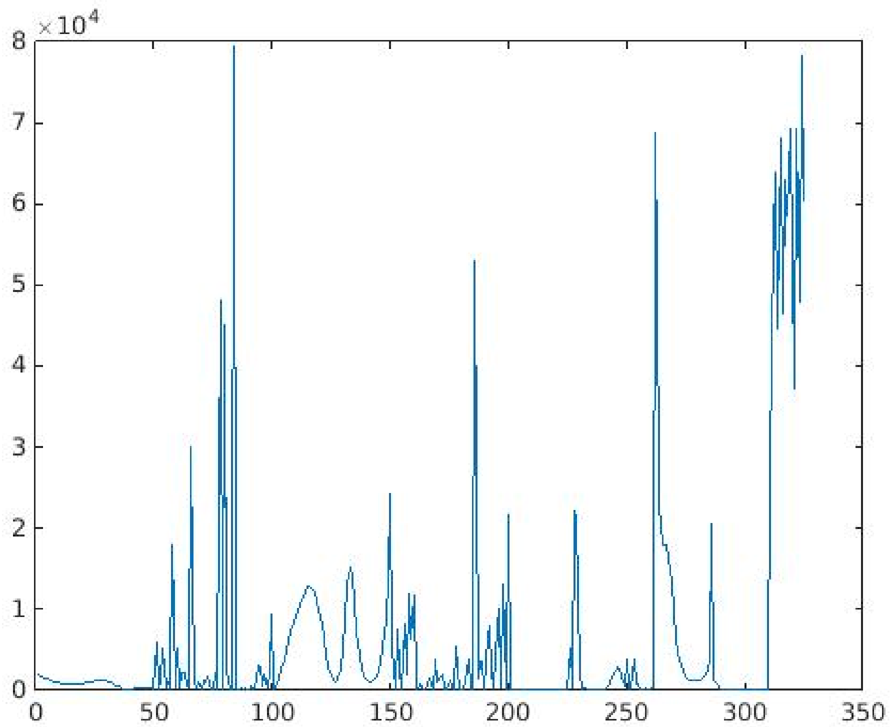
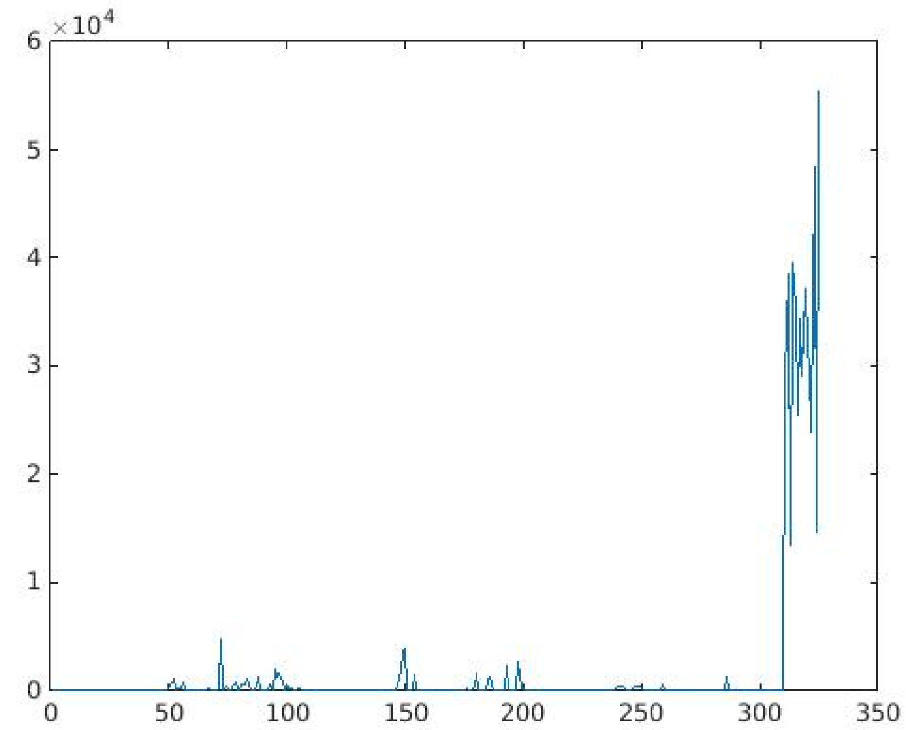
- reduce the dimensionality by taking the first k eigenfunctions
  of the Laplace-Beltrami operator

Application to non-rigid shape matching

Approach: use framework of *functional maps* [Ovsjanikov et al. 2012]

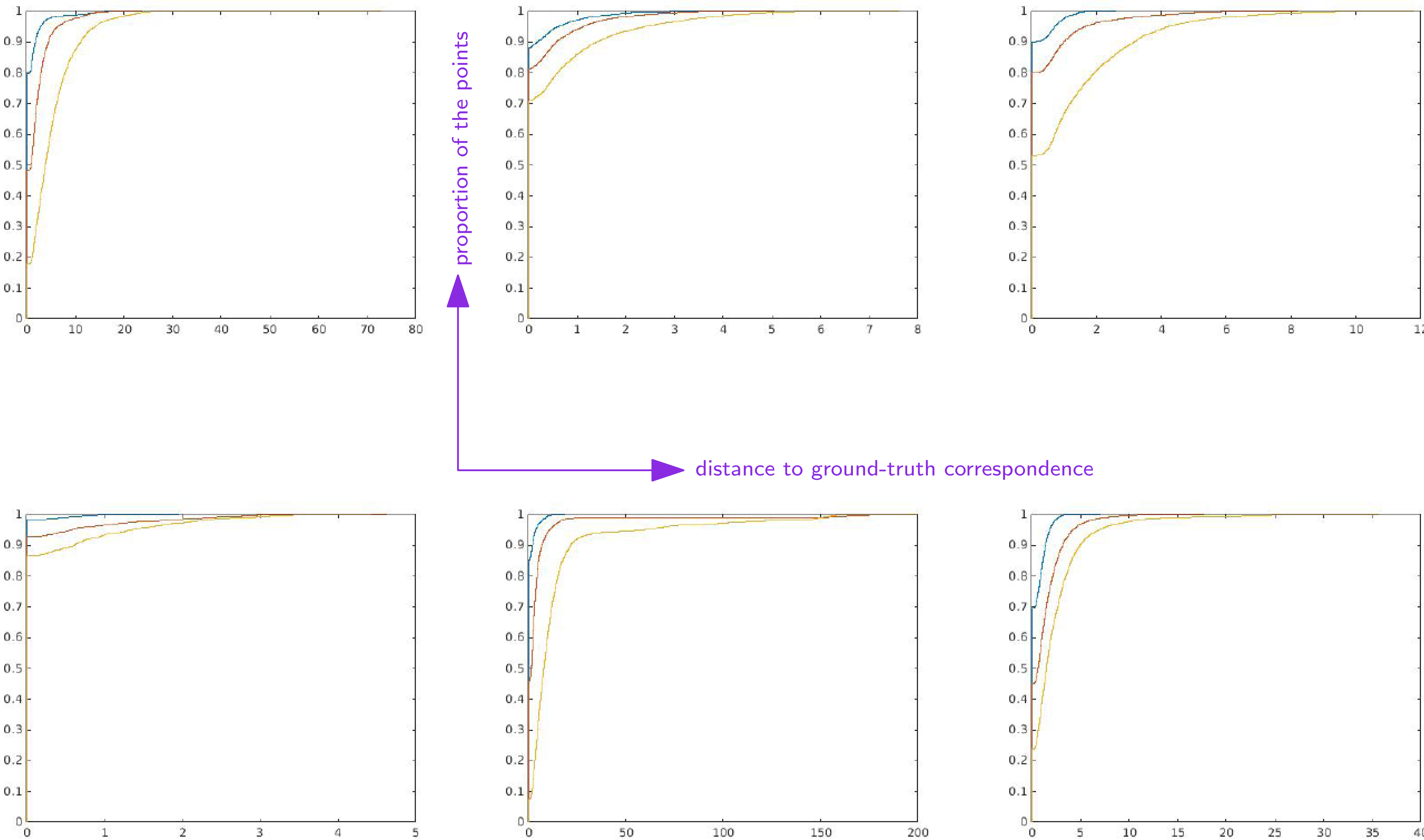



topological signatures (last 30 indices) have a high influence on the choice of optimal map

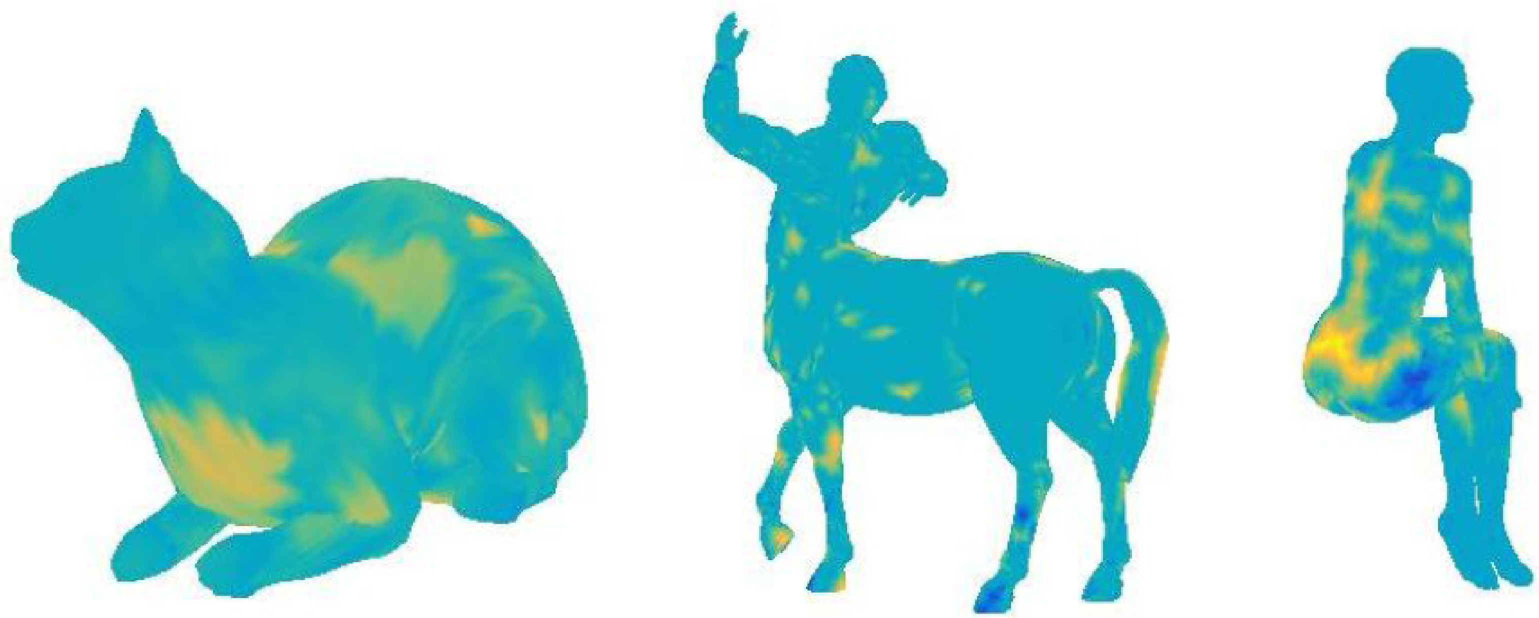
Application to non-rigid shape matching

Approach: use framework of *functional maps* [Ovsjanikov et al. 2012]

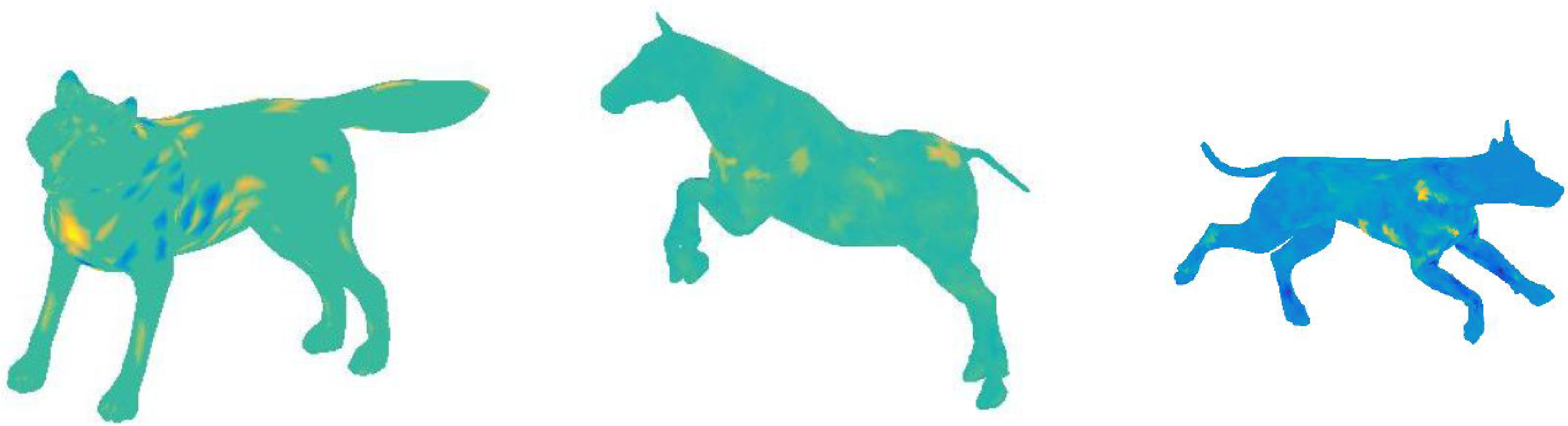


Application to non-rigid shape matching

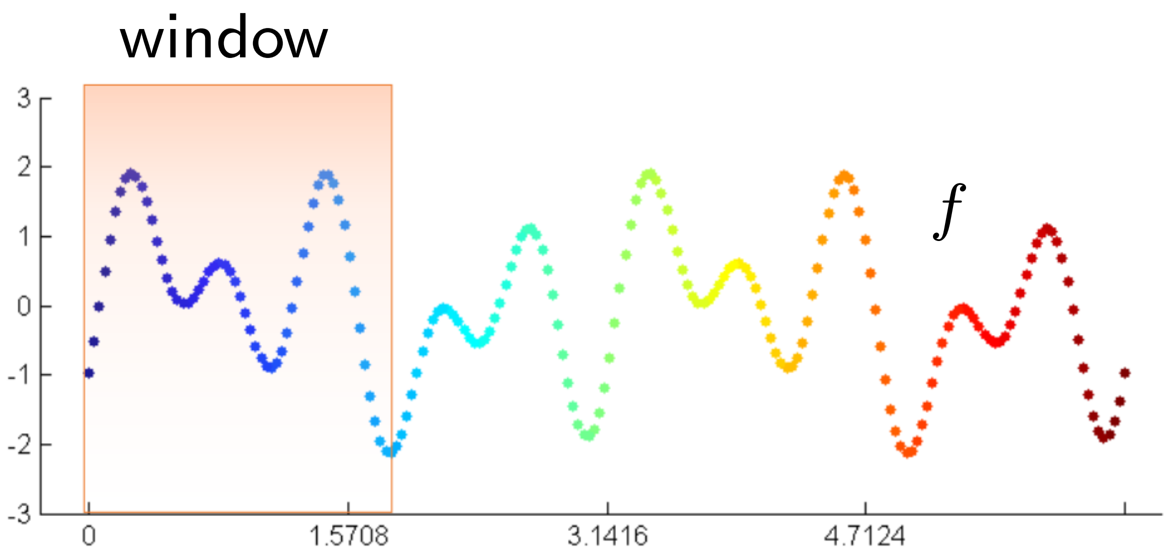
Approach: use framework of *functional maps* [Ovsjanikov et al. 2012]



correspondences in flat regions are improved by topological signatures

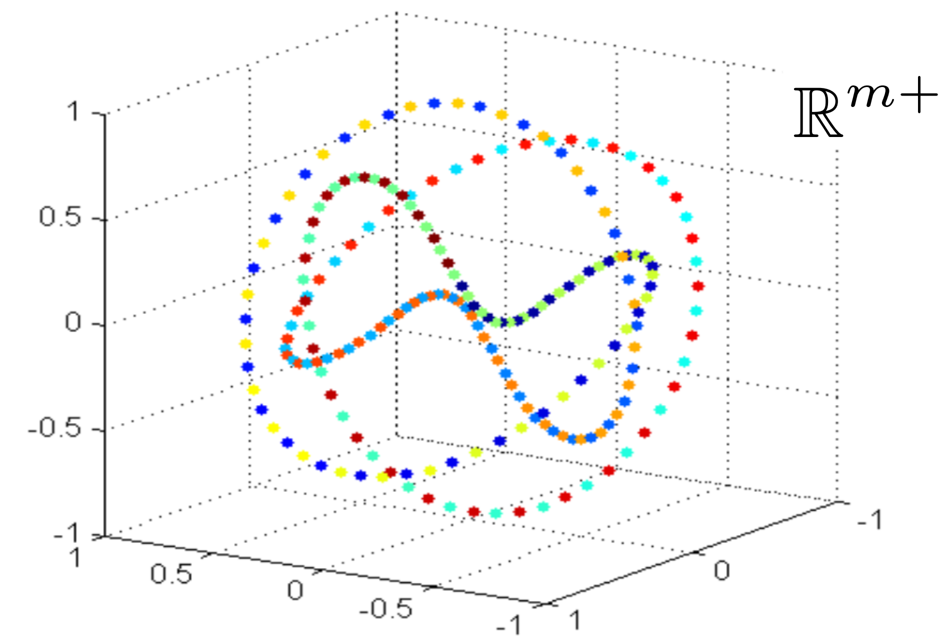


Application to supervised time series analysis



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

(time-delay embedding)



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

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

τ: step / delay

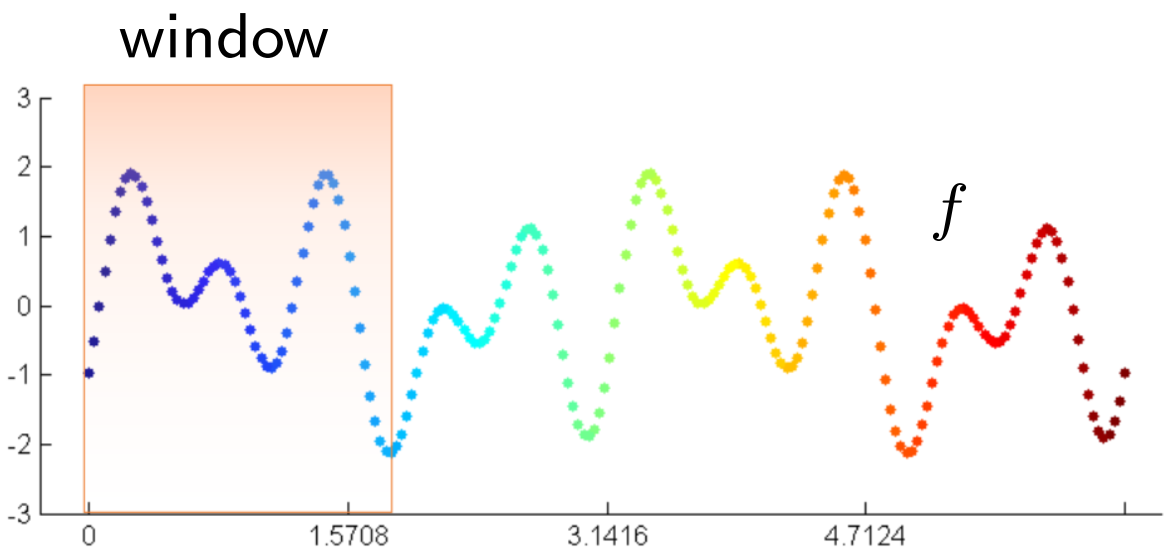
$m\tau$: window size

$m+1$: embedding dimension

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

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

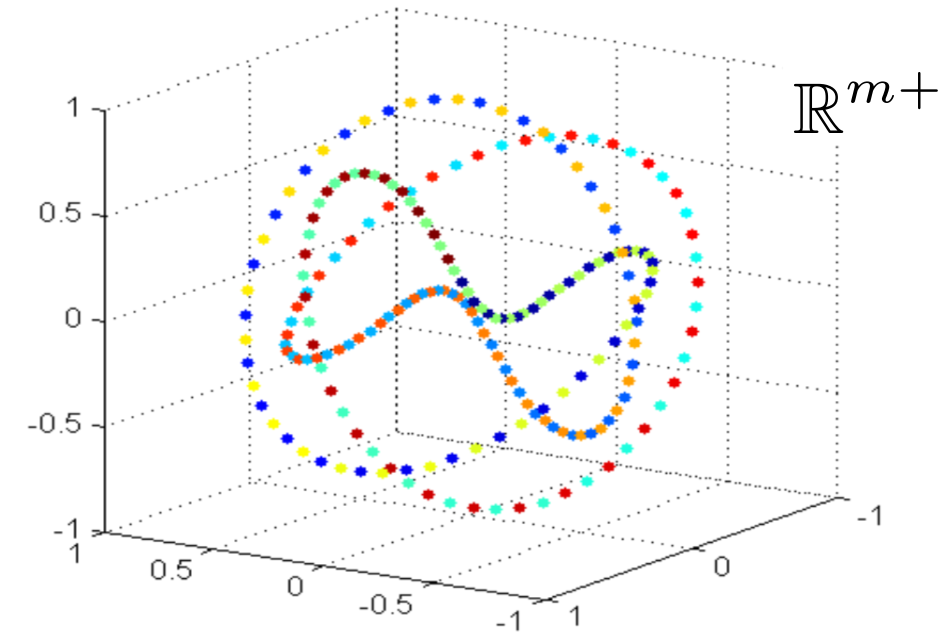
Application to supervised time series analysis



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

$\Longrightarrow$

(time-delay embedding)



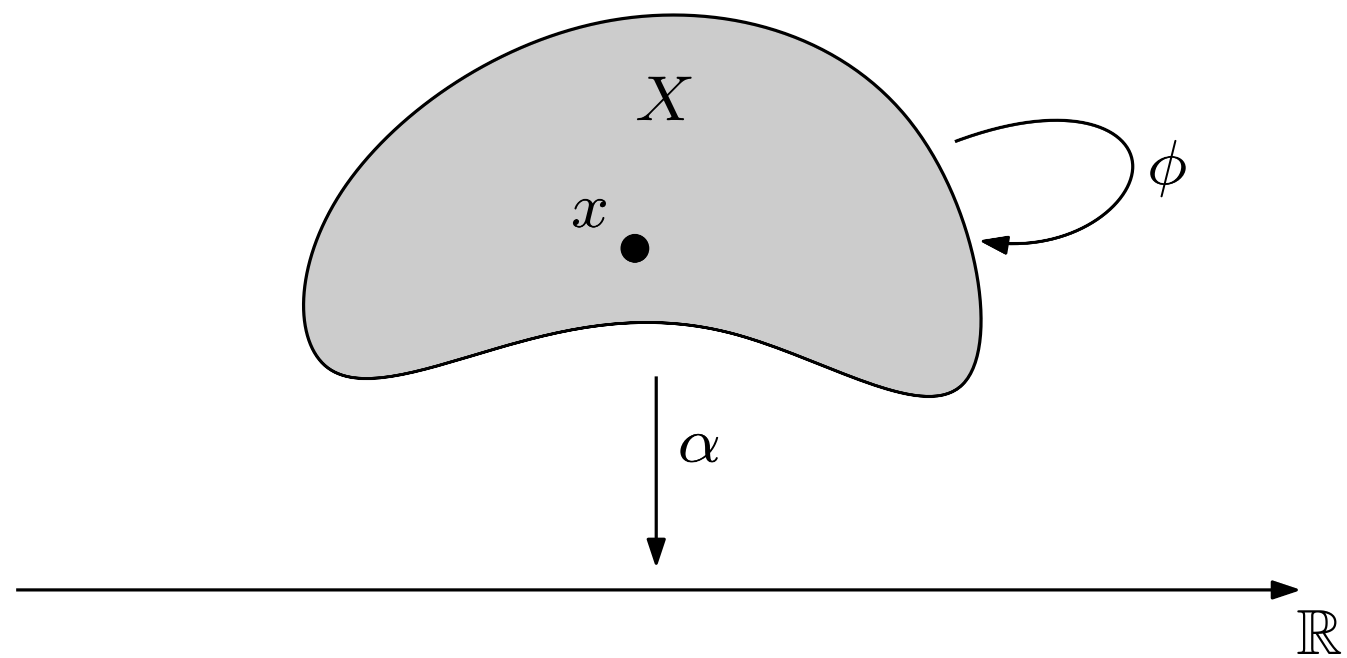
Contributions of TDA:

inference of:

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

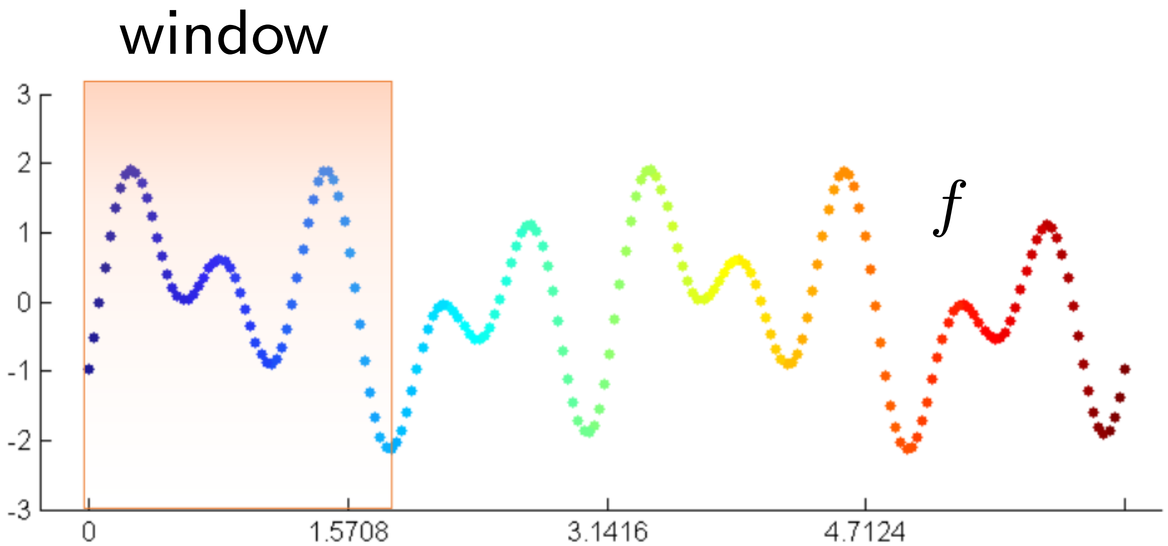
no Fourier transform needed

Dynamical system:



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

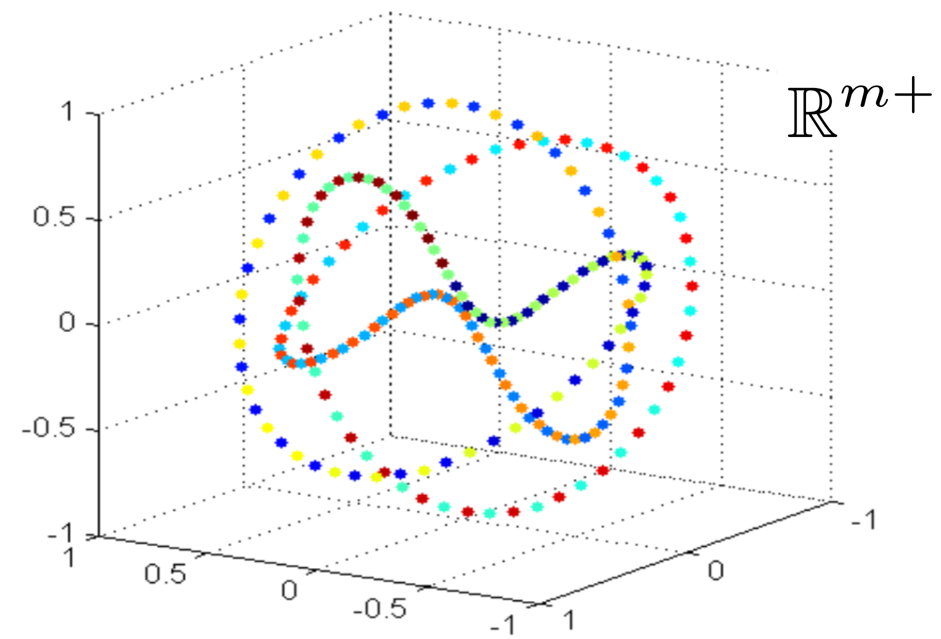
Application to supervised time series analysis



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

$\Longrightarrow$

(time-delay embedding)



Contributions of TDA:

inference of:

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

no Fourier transform needed

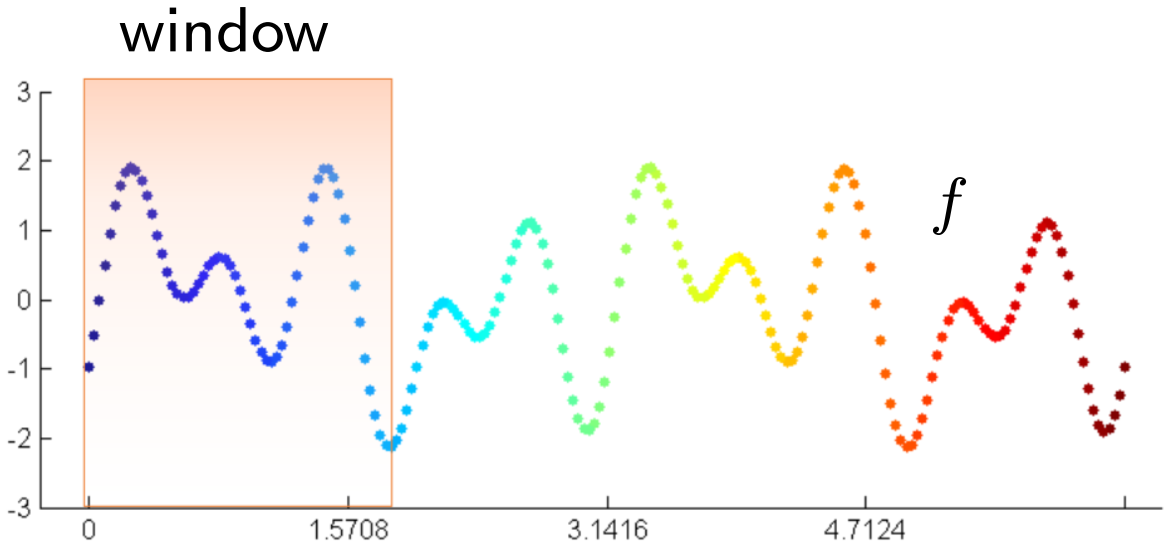
Dynamical system:

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

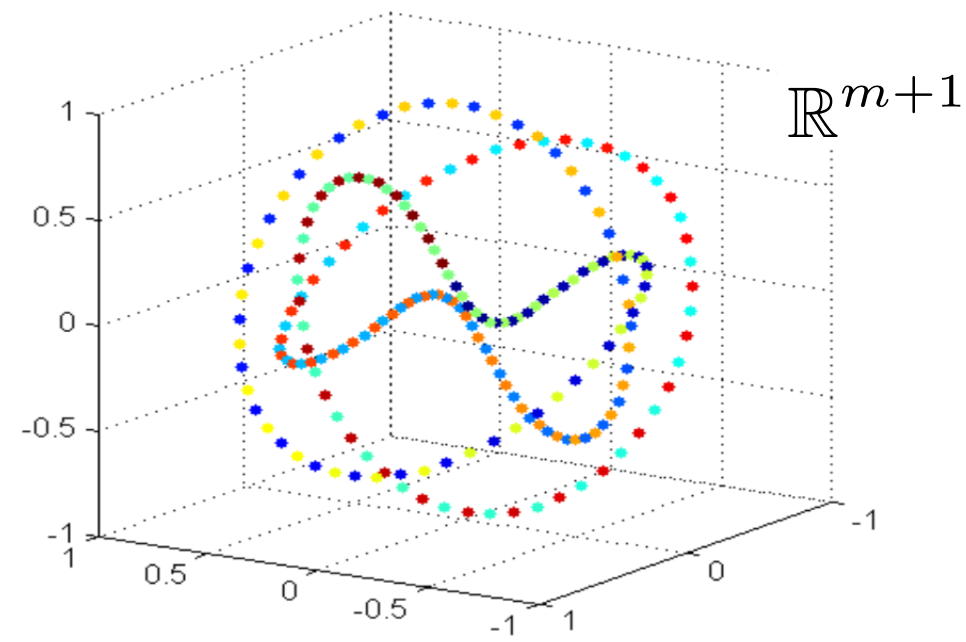
$$X \to \mathbb{R}^{m+1}$$
$$x \mapsto (\alpha(x), \alpha \circ \phi(x), \cdots, \alpha \circ \phi^m(x))$$

is an embedding.

Application to supervised time series analysis



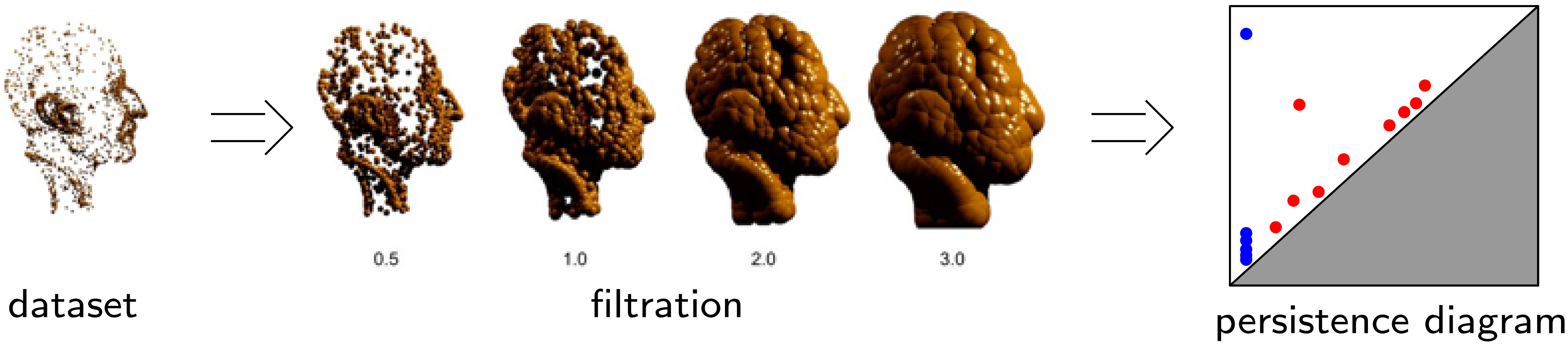
$\mathrm{TD}_{m,\tau}$
$\Longrightarrow$
(time-delay embedding)

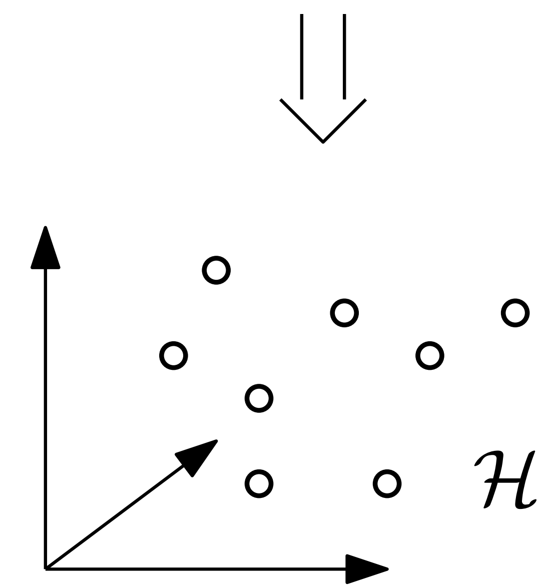


method / dataset	Gyro sensor	EEG dataset	EMG dataset
SVM + statistical features	67.6 ± 4.7	44.4 ± 19.8	15.0 ± 10.0
SVM + Betti sequence	63.5 ± 11.3	66.7 ± 5.6	49.6 ± 18.2
1-d CNN + dynamic time warping	6.4 ± 5.1	72.4 ± 6.1	15.0 ± 10.0
imaging CNN	18.9 ± 5.2	48.9 ± 4.2	10.0 ± 0.0
1-d CNN + Betti sequence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

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

Wrap'up





- kernels for persistence diagrams:
  - stable
  - discriminative
  - easy to compute (closed-form expr., finite-dim. vectors)
  - additive, universal, etc.

- other topic: integration of TDA into learning methods (clustering, NNs, etc.)