Supervised Machine Learning, Kernel Methods and Persistence Diagrams

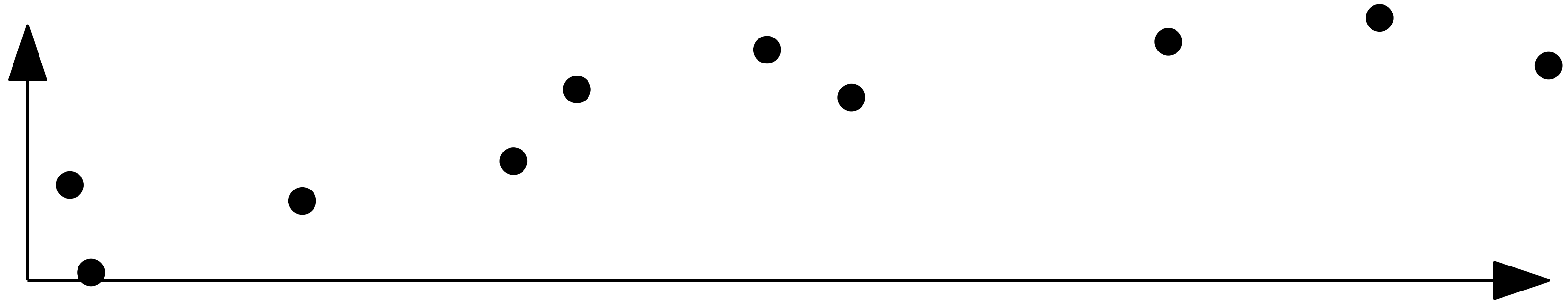
Mathieu Carrière

Reading Group
Inria Saclay
12/09/2017

Supervised Machine Learning

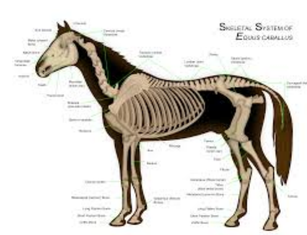
n observations at hand $(x_1, y_1), \cdots, (x_n, y_n) \in X \times Y$

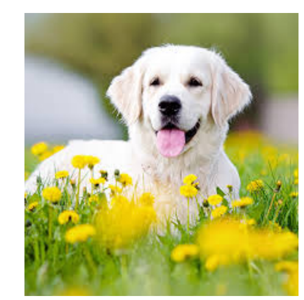
$Y = \mathbb{R}$

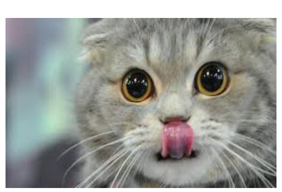


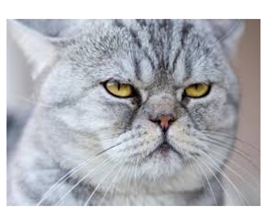
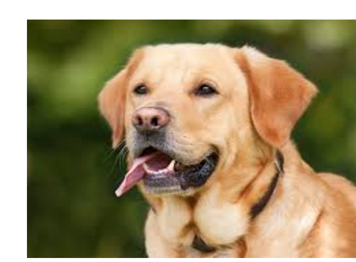
$X = \mathbb{R}$



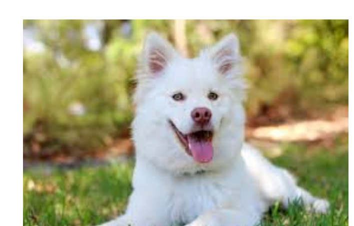





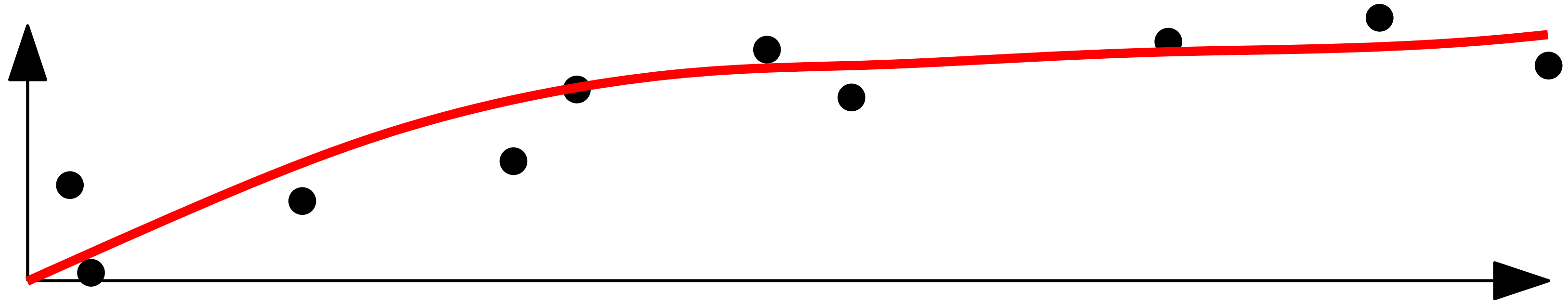
X = images,
Y = {cat, dog, horse}

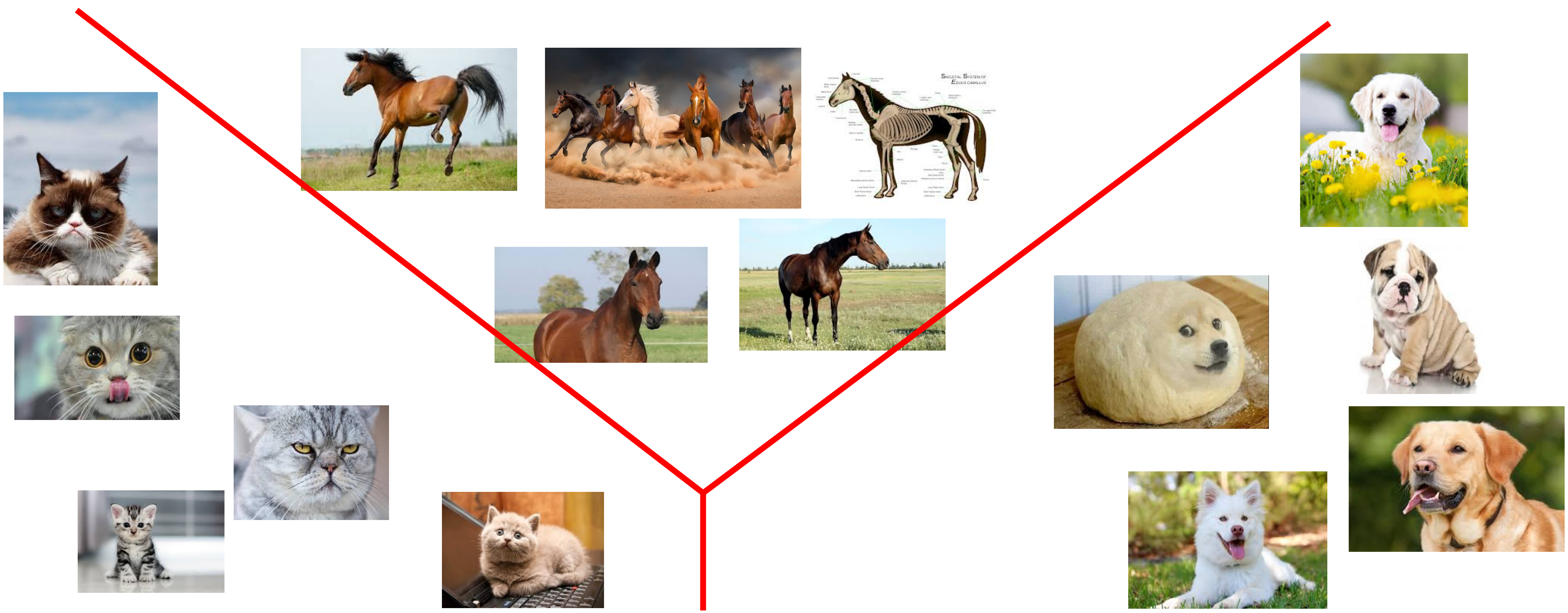


Supervised Machine Learning

n observations at hand $(x_1, y_1), \cdots, (x_n, y_n) \in X \times Y$

Goal: produce classifier $f : X \to Y$, where $f = f((x_1, y_1), \cdots, (x_n, y_n))$





Empirical Risk Minimization

Common formalization:

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

Empirical Risk Minimization

Common formalization:

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

$\mathcal{F}$ 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*

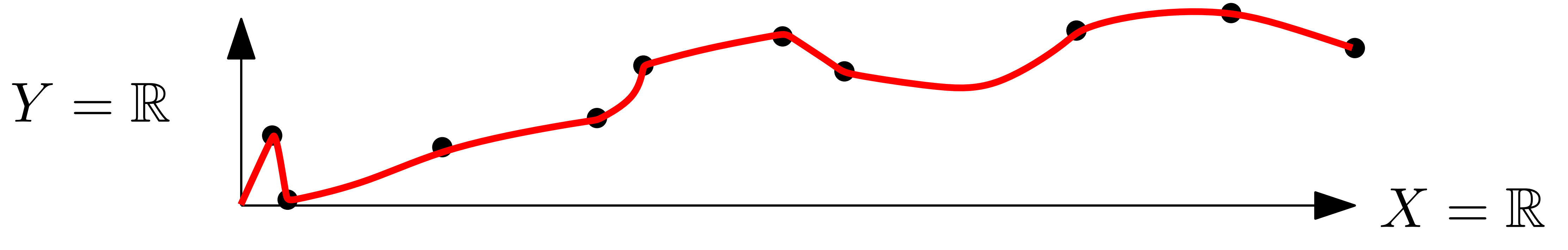
Empirical Risk Minimization

Common formalization:

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

Why is there a regularizer?

$\rightarrow$ avoid overfitting



Empirical Risk Minimization

Common formalization:

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

$\mathcal{F}$ 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*

$L(y_i, f(x_i))$	Name
$\delta_{y_i = f(x_i)}$	zero-one
$\max\{0, 1 - y_i f(x_i)\}$	hinge $\to$ SVM
$\log(1 + \exp(-y_i f(x_i)))$	exponential $\to$ Adaboost
$(y_i - f(x_i))^2$	squared $\to$ Least squares

Empirical Risk Minimization

Common formalization:

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

$\mathcal{F}$ 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*

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

$\Omega(w)$	Name
$\|w\|_2^2$	$\ell_2 \to$ convex and differentiable
$\|w\|_1$	$\ell_1 \to$ sparse
$\alpha\|w\|_2^2 + (1-\alpha)\|w\|_1$	elastic net

Empirical Risk Minimization

Common formalization:

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

$\mathcal{F}$ 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 when $\mathcal{F}$ is a *Reproducing Kernel Hilbert Space*

Reproducing Kernel Hilbert Space

Def: Let $\mathcal{H} \subset \mathbb{R}^X$ Hilbert, with dot product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$.
$\mathcal{H}$ is a RKHS if $\exists\, k : X \times X \to \mathbb{R}$ s.t.:

(i) $\{k_x \,:\, x \in X\} \subset \mathcal{H}$, where $k_x : x \mapsto k(x, \cdot)$,

(ii) $f(x) = \langle f, k_x \rangle_{\mathcal{H}}$, $\forall x \in X$ and $\forall f \in \mathcal{H}$.

Reproducing Kernel Hilbert Space

Def: Let $\mathcal{H} \subset \mathbb{R}^X$ Hilbert, with dot product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$.
$\mathcal{H}$ is a RKHS if $\exists\, k : X \times X \to \mathbb{R}$ s.t.:

(i) $\{k_x : x \in X\} \subset \mathcal{H}$, where $k_x : x \mapsto k(x, \cdot)$,

(ii) $f(x) = \langle f, k_x \rangle_{\mathcal{H}}$, $\forall x \in X$ and $\forall f \in \mathcal{H}$.

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

Reproducing Kernel Hilbert Space

Def: Let $\mathcal{H} \subset \mathbb{R}^X$ Hilbert, with dot product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$.
$\mathcal{H}$ is a RKHS if $\exists\ k : X \times X \to \mathbb{R}$ s.t.:

(i) $\{k_x\ :\ x \in X\} \subset \mathcal{H}$, where $k_x : x \mapsto k(x, \cdot)$,

(ii) $f(x) = \langle f, k_x \rangle_{\mathcal{H}}$, $\forall x \in X$ and $\forall f \in \mathcal{H}$.

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

Th: [Moore 1950] $k : X \times X \to \mathbb{R}$ is a kernel iif if it is positive definite, i.e. $\sum_{i,j} a_i a_j k(x_i, x_j) \geq 0$ for any $a_1, \cdots, a_n \in \mathbb{R}$ and $x_1, \cdots, x_n \in X$.

Reproducing Kernel Hilbert Space

Def: Let $\mathcal{H} \subset \mathbb{R}^X$ Hilbert, with dot product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$.
$\mathcal{H}$ is a RKHS if $\exists\, k : X \times X \to \mathbb{R}$ s.t.:

(i) $\{k_x \ : \ x \in X\} \subset \mathcal{H}$, where $k_x : x \mapsto k(x, \cdot)$,

(ii) $f(x) = \langle f, k_x \rangle_{\mathcal{H}}$, $\forall x \in X$ and $\forall f \in \mathcal{H}$.

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

Th: [Moore 1950] $k : X \times X \to \mathbb{R}$ is a kernel iif if it is positive definite,
i.e. $\sum_{i,j} a_i a_j k(x_i, x_j) \geq 0$ for any $a_1, \cdots, a_n \in \mathbb{R}$ and $x_1, \cdots, x_n \in X$.

Ex:

- linear: $k(x, y) = \langle x, y \rangle$
- polynomial: $k(x, y) = (\alpha \langle x, y \rangle + 1)^\beta$, $\alpha, \beta \in \mathbb{R}$,
- Gaussian: $k(x, y) = \exp\left(-\frac{\|x-y\|_2^2}{2\sigma^2}\right)$, $\sigma > 0$.

Reproducing Kernel Hilbert Space

Def: Let $\mathcal{H} \subset \mathbb{R}^X$ Hilbert, with dot product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$.
$\mathcal{H}$ is a RKHS if $\exists\, k : X \times X \to \mathbb{R}$ s.t.:

(i) $\{k_x \,:\, x \in X\} \subset \mathcal{H}$, where $k_x : x \mapsto k(x, \cdot)$,

(ii) $f(x) = \langle f, k_x \rangle_{\mathcal{H}}$, $\forall x \in X$ and $\forall f \in \mathcal{H}$.

Th: [Scholkopf et al 2001]
$\mathcal{H}$ RKHS with kernel k. Any function $f^* \in \mathcal{H}$ minimizing

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

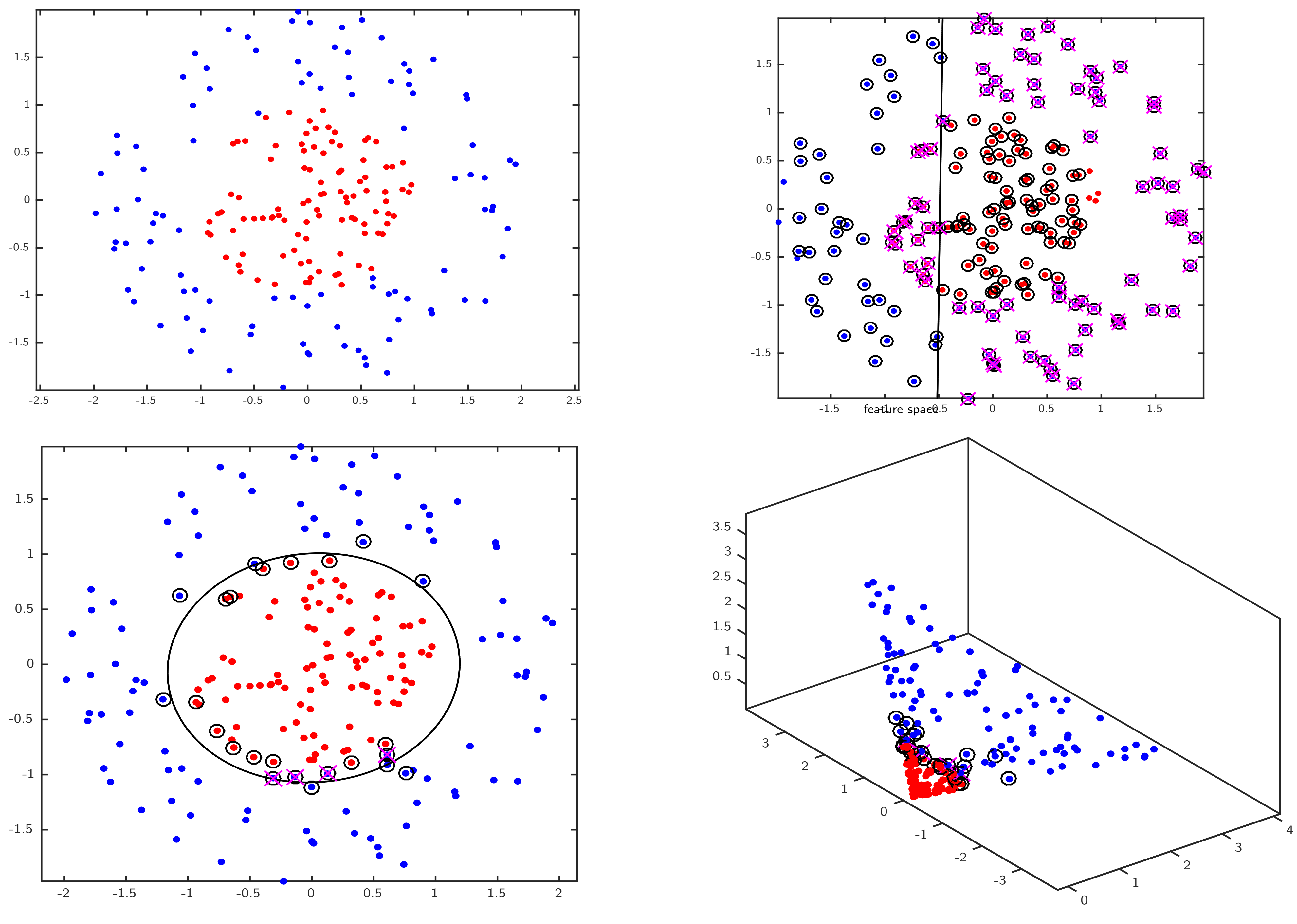
is of the form $f^*(\cdot) = \sum_{i=1}^{n} \alpha_i k(x_i, \cdot)$, where $\alpha_1, \cdots, \alpha_n \in \mathbb{R}$.

Only the $k(x_i, x_j)$ are required to minimize!!

Kernel Trick

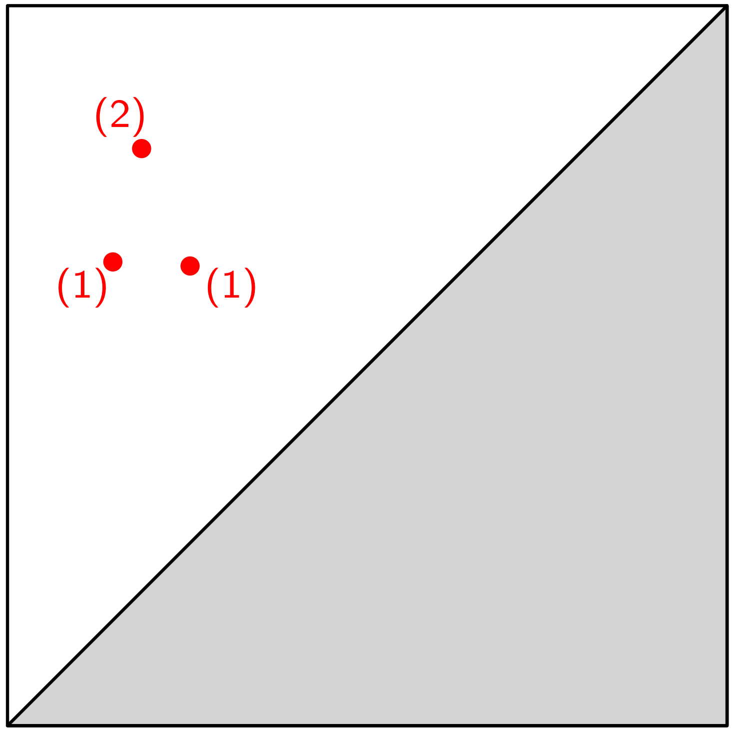
$k : X \times X \to \mathbb{R}$ is positive definite, with RKHS $\mathcal{H}_k$. Then:

$$k(x, y) = \langle k_x, k_y \rangle_{\mathcal{H}_k} = \langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}_k}$$



Persistence diagrams

Def: A finite multiset in the open half-plane $\Delta \times \mathbb{R}_+$



Persistence diagrams

Def: A finite multiset in the open half-plane $\Delta \times \mathbb{R}_+$

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

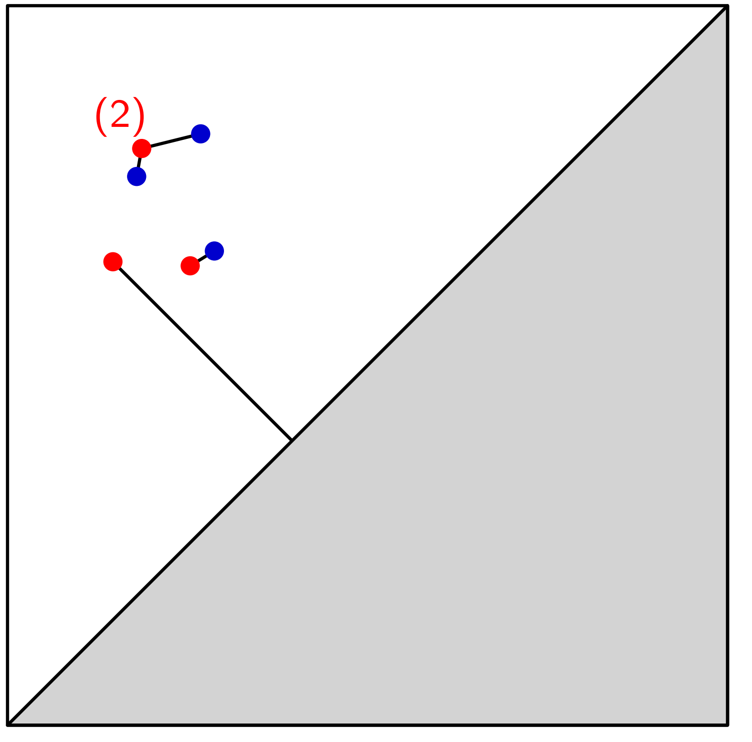
cost of M: $c_p(M) = \left(\sum_{(x,\,y)\text{ matched}} c_p(x, y) \; + \sum_{z\text{ unmatched}} c_p(z)\right)^{1/p}$

p-th diagram distance (extended metric):

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

bottleneck distance:

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



Kernels for persistence diagrams

Two methods:

- define explicit feature map $\Phi : \mathcal{D} \rightarrow \mathcal{H}$
- define Gaussian kernel with Berg theorem

Kernels for persistence diagrams

Two methods:

- define explicit feature map $\Phi : \mathcal{D} \to \mathcal{H}$
- define Gaussian kernel with Berg theorem

Th: [Berg, Christensen, Ressel 1984]
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

Two methods:

- define explicit feature map $\Phi : \mathcal{D} \to \mathcal{H}$
- define Gaussian kernel with Berg theorem

Th: [Berg, Christensen, Ressel 1984]
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?

Pb: d_p is not cnsd, for any $p > 0$

Kernels for persistence diagrams

View persistence diagrams as:

- landscapes (collections of 1-d functions) [Bubenik 2012] [Bubenik Dłotko 2015]
- discrete measures:
  - $\rightarrow$ 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]
  - $\rightarrow$ sliced on lines [C. Oudot Cuturi 2017]
- finite metric spaces [C. Oudot Ovsjanikov 2015]
- polynomial roots or evaluations [Di Fabio Ferri 2015] [Kališnik 2016]

Kernels for persistence diagrams

View persistence diagrams as:

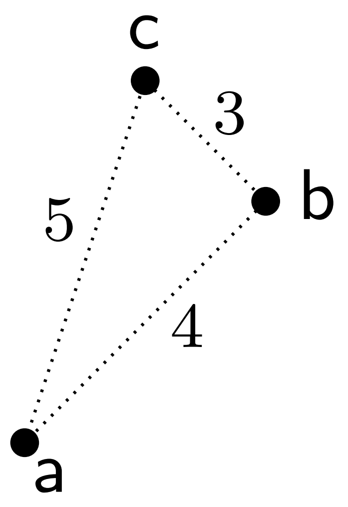
- landscapes (collections of 1-d functions) [Bubenik 2012] [Bubenik Dłotko 2015]
- discrete measures:
  - $\rightarrow$ convolution with fixed kernel [Chepushtanova et al. 2015]
  - $\rightarrow$ convolution with weighted kernel [Kusano Fukumisu Hiraoka 2016-17]
  - $\rightarrow$ heat diffusion [Reininghaus et al. 2015] $+$ exponential [Kwit et al. 2015]
  - $\rightarrow$ sliced on lines [C. Oudot Cuturi 2017]
- finite metric spaces [C. Oudot Ovsjanikov 2015]
- polynomial roots or evaluations [Di Fabio Ferri 2015] [Kališnik 2016]

Kernels for persistence diagrams

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

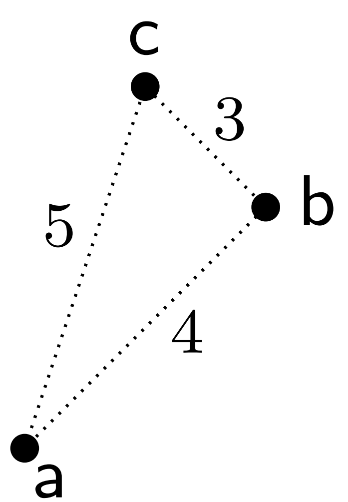
Persistence diagrams as metric spaces

finite metric space



Persistence diagrams as metric spaces

finite metric space



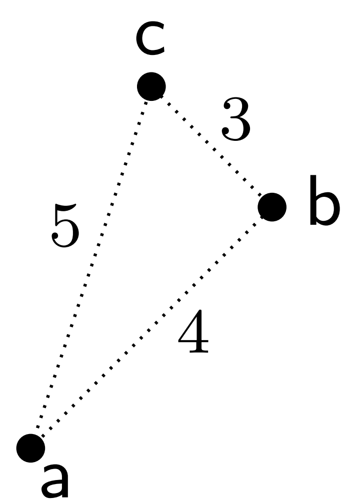
$$\xrightarrow{\phi_1}$$

distance matrix

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

Persistence diagrams as metric spaces

finite metric space



$\xrightarrow{\phi_1}$

distance matrix

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

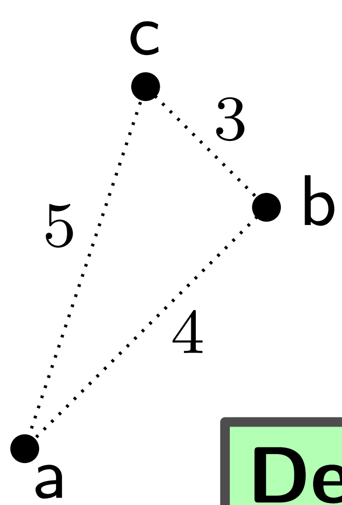
$\xrightarrow{\phi_2}$

sorted sequence
with finite support

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

Persistence diagrams as metric spaces

finite metric space



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

$\xrightarrow{\phi_1}$

distance matrix

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

$\xrightarrow{\phi_2}$

sorted sequence
with finite support

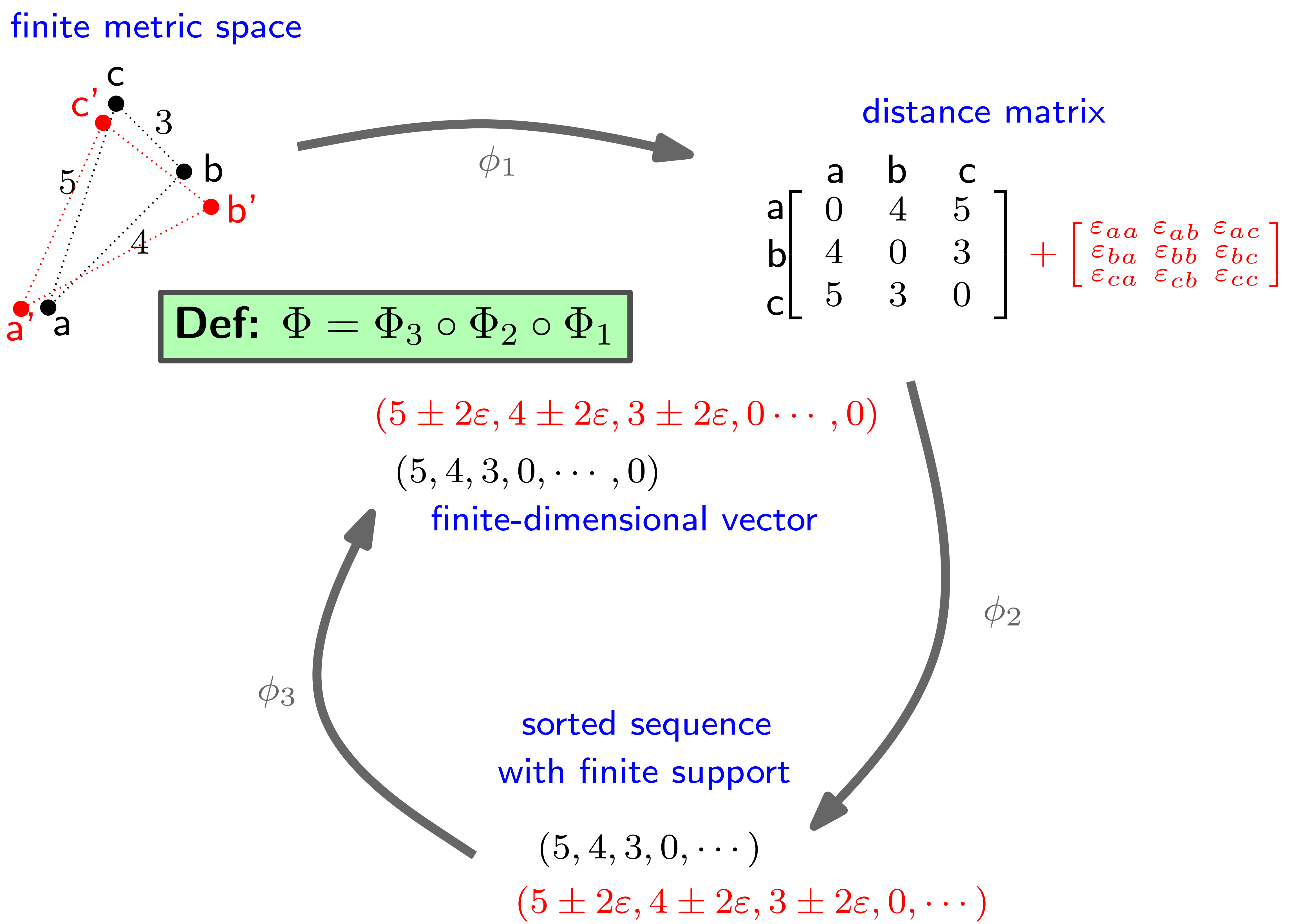
$$(5, 4, 3, 0, \cdots)$$

$\xrightarrow{\phi_3}$

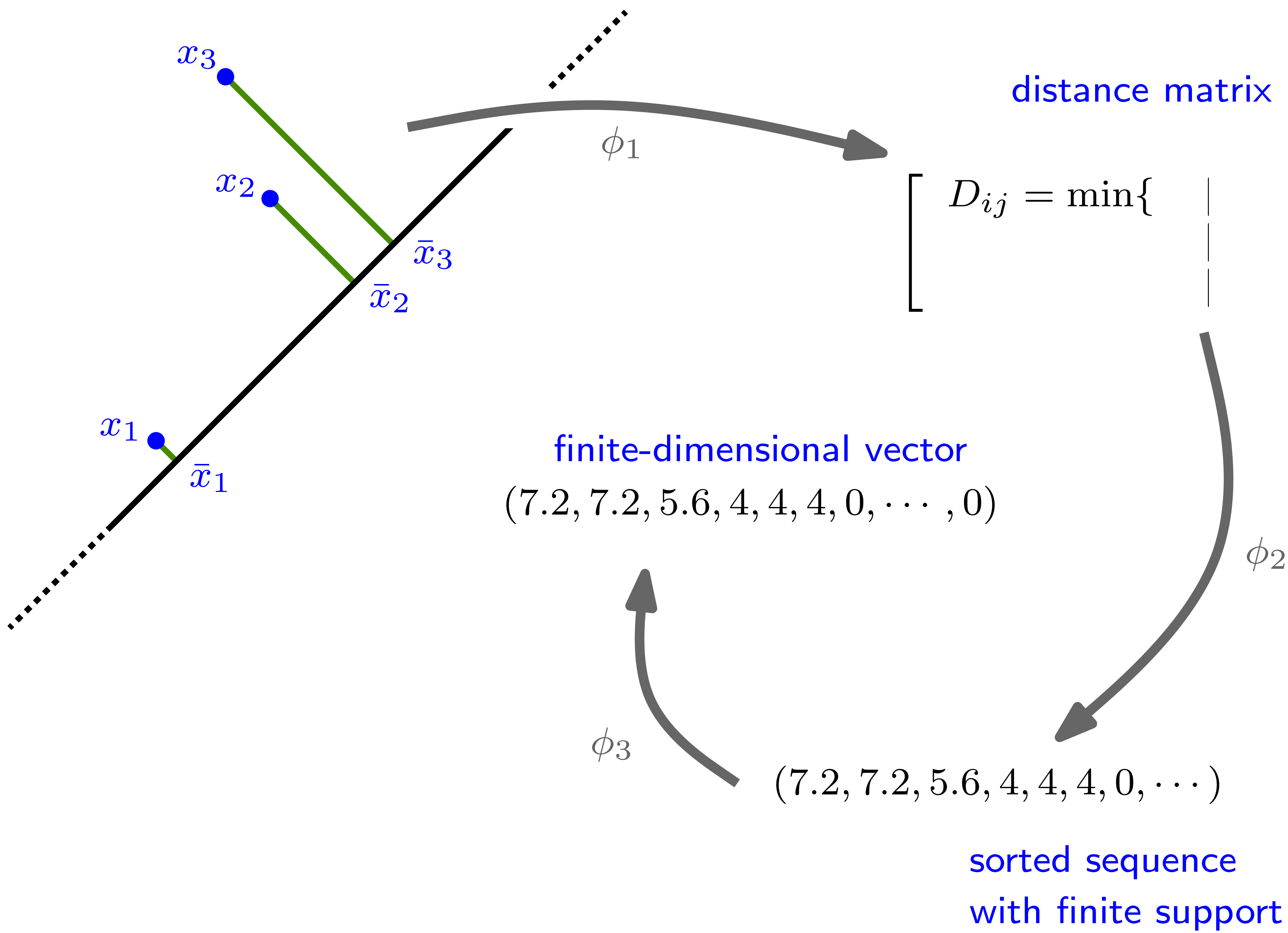
$$(5, 4, 3, 0, \cdots, 0)$$

finite-dimensional vector

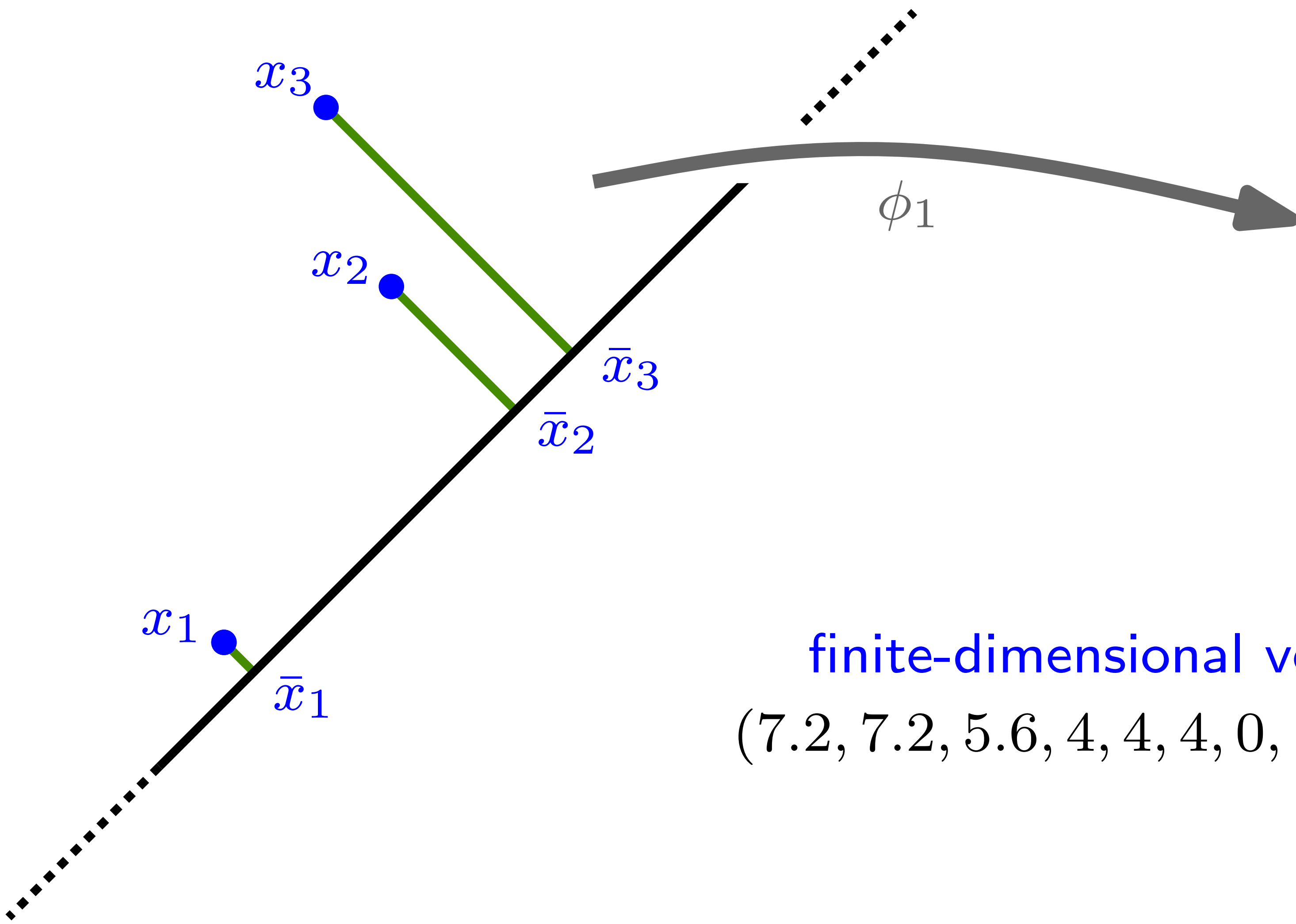
Stability of feature map



Stability of feature map



Stability of feature map



distance matrix

$$\left[\; D_{ij} = \min\{ \quad \right|$$

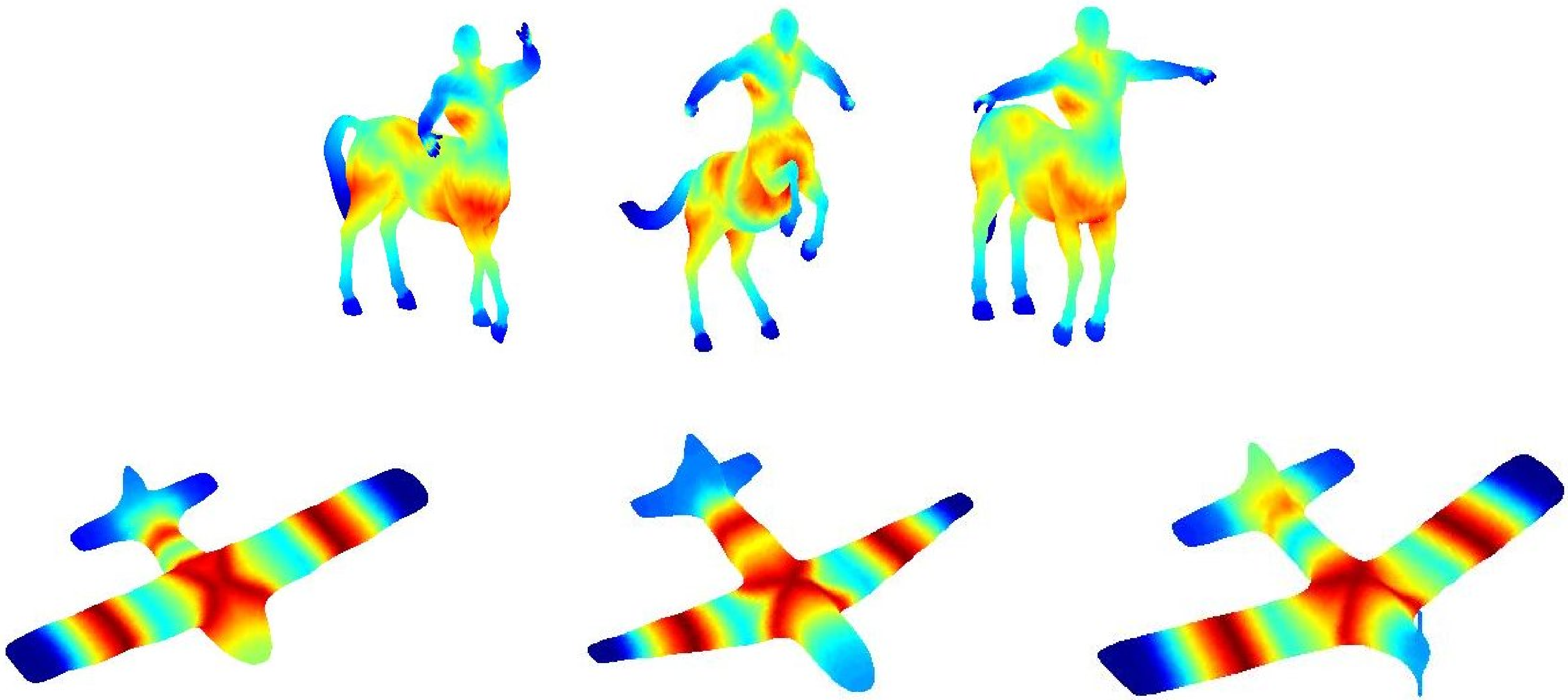
finite-dimensional vector

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

Prop: [C. Oudot Ovsjanikov 2015]

- $\|\Phi(D) - \Phi(D')\|_\infty \leq 2\,\mathrm{d}_\infty(D, D')$
- $\|\Phi(D) - \Phi(D')\|_p \leq 2D^{-\frac{1}{p}}\,\mathrm{d}_\infty(D, D')$

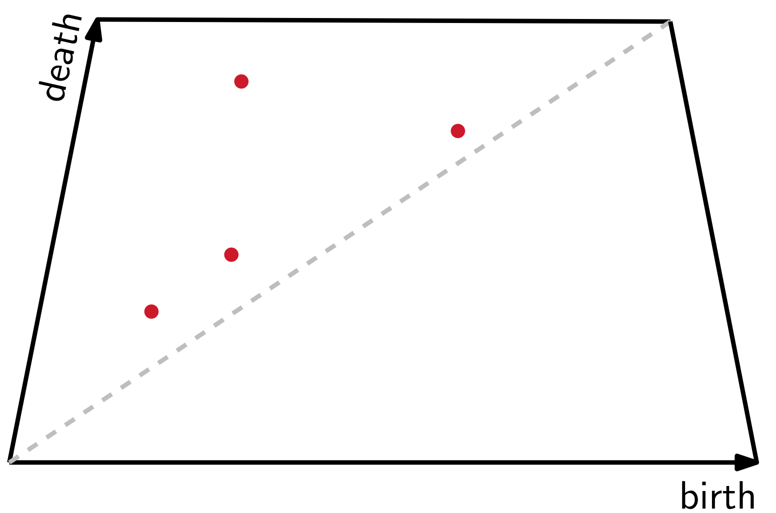
Stability of feature map

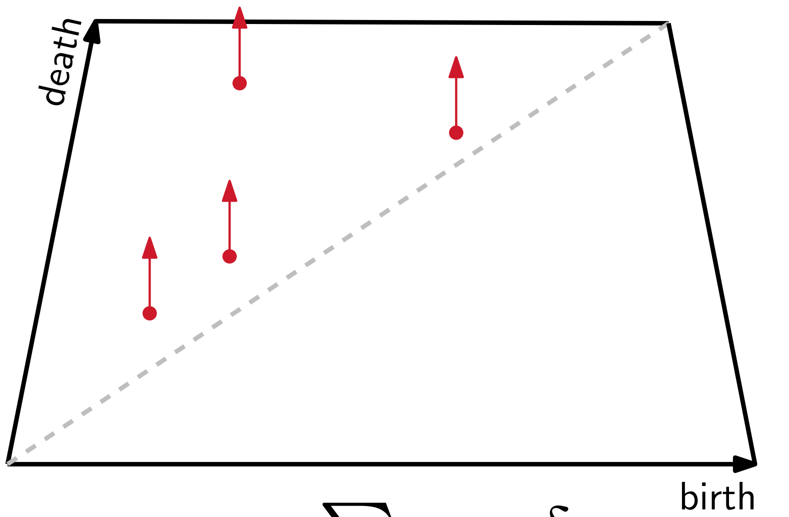


Prop: [C. Oudot Ovsjanikov 2015]

- $\|\Phi(D) - \Phi(D')\|_\infty \leq 2\,\mathrm{d}_\infty(D, D')$
- $\|\Phi(D) - \Phi(D')\|_p \leq 2D^{-\frac{1}{p}}\,\mathrm{d}_\infty(D, D')$

Persistence diagrams as discrete measures





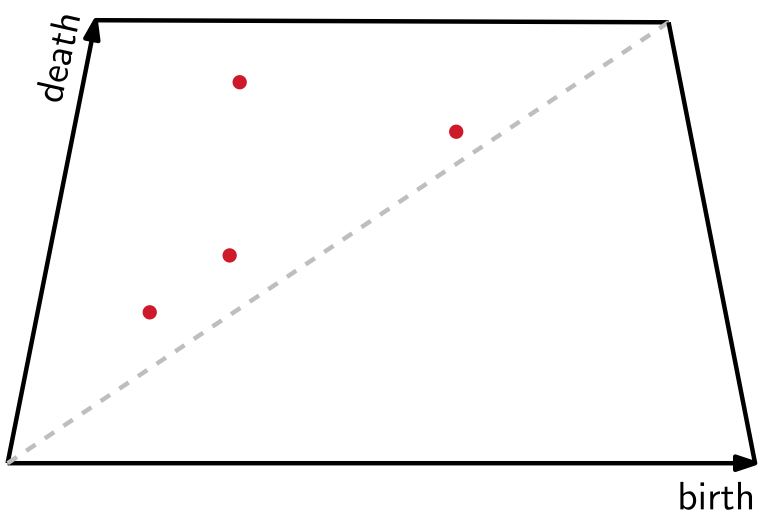
$$\mu_D = \sum_{x \in D} \delta_x$$

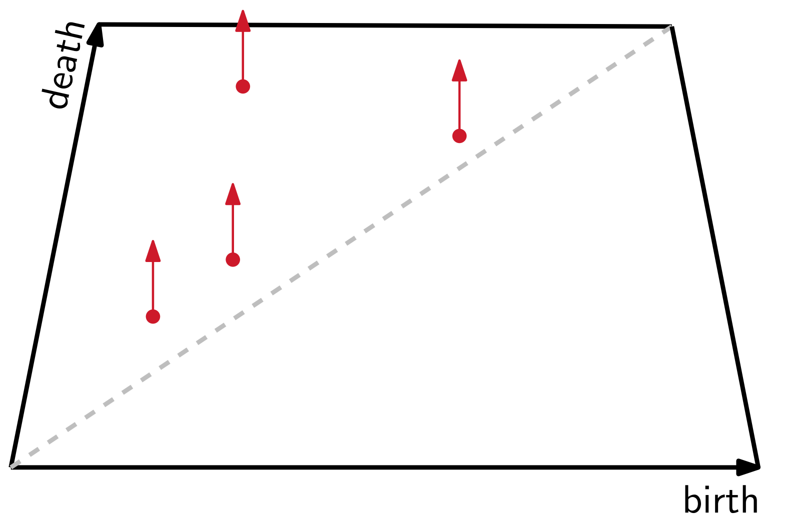
- discrete measures accurately represent persistence diagrams
- discrete measures can be compared with 1-Wasserstein distance which is (almost) cnsd and looks like the diagram distance d_1

Def: Let $\mu = \sum_{i=1}^{n} \delta_{x_i}$ and $\nu = \sum_{i=1}^{n} \delta_{y_i}$

$$W_1(\mu, \nu) = \inf_\pi \ \sum_{i=1}^{n} \|x_i - y_{\pi(i)}\|$$

Persistence diagrams as discrete measures

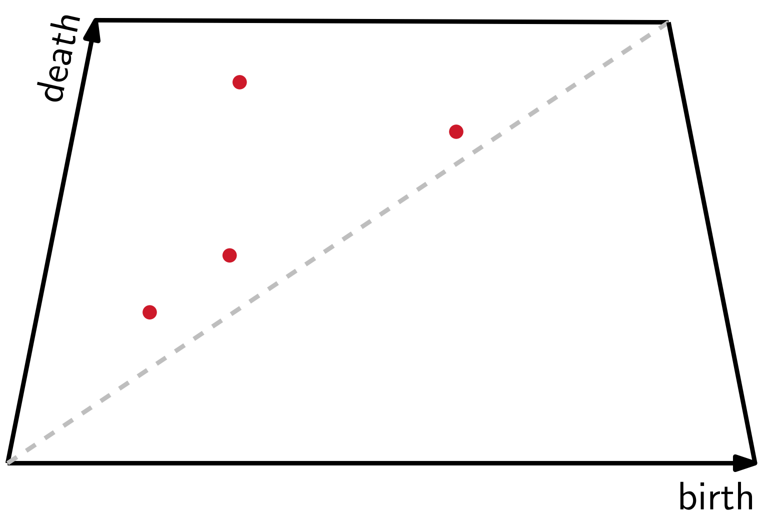


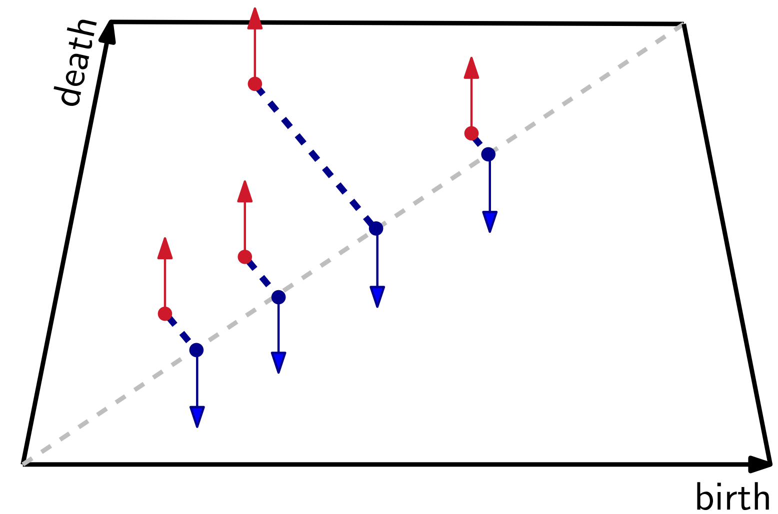


$\mu_D = \sum_{x \in D} \delta_x$

Pb: $d_1(D, D') \neq W_1(\mu_D, \mu_{D'})$ (W_1 does not even make sense here)

Persistence diagrams as discrete measures





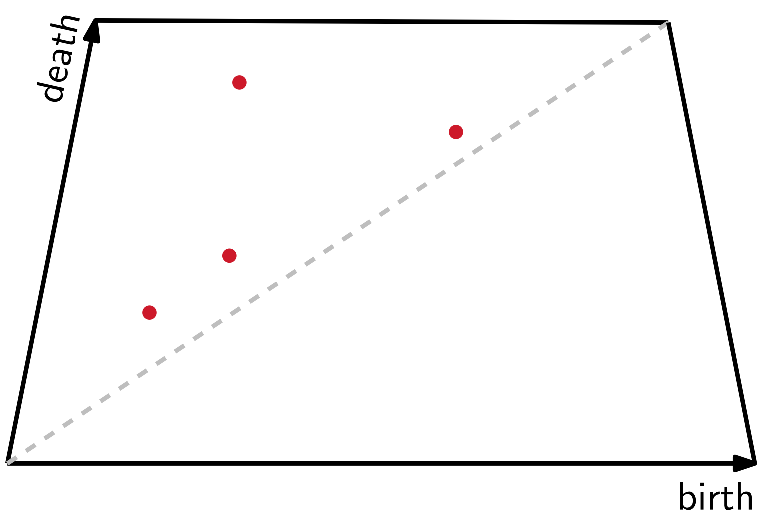
$\mu_D = \sum_{x\in D} \delta_x$

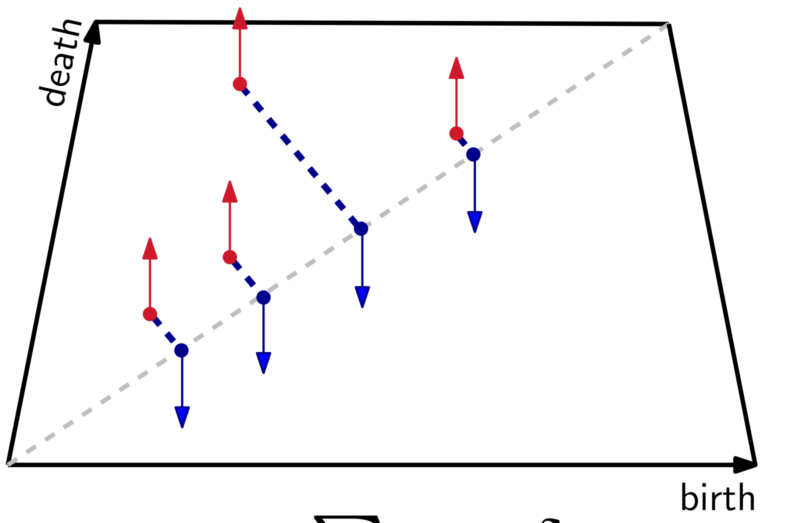
Pb: $d_1(D, D') \neq W_1(\mu_D, \mu_{D'})$ (W_1 does not even make sense here)

Solution: use projections onto the diagonal

$$\mu_D^+ = \sum_{x\in D} \delta_x \qquad \mu_D^- = \sum_{x\in D} \delta_{\pi_\Delta(x)}$$

Persistence diagrams as discrete measures





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

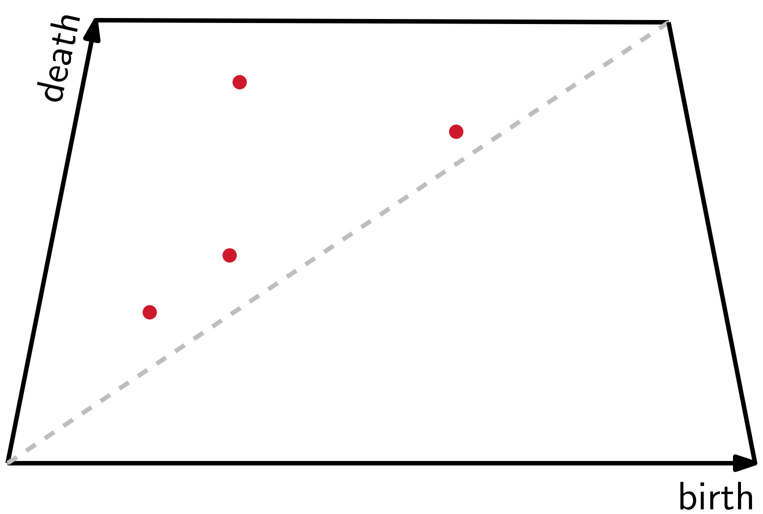
Pb: $d_1(D, D') \neq W_1(\mu_D, \mu_{D'})$ (W_1 does not even make sense here)

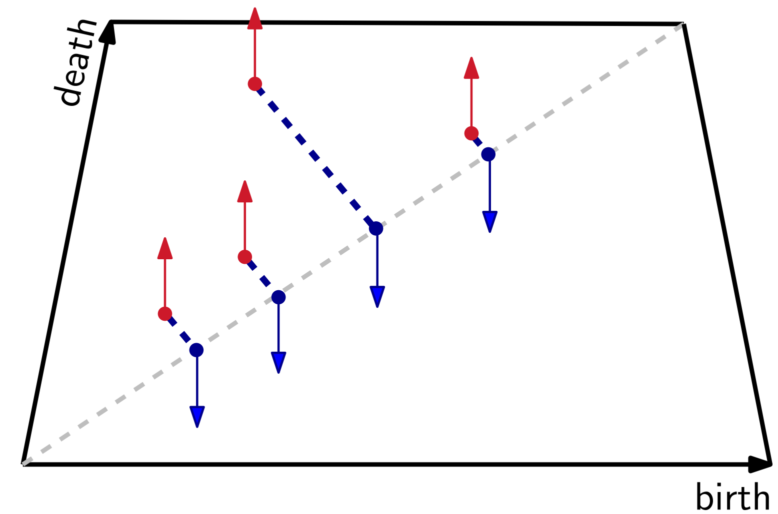
Solution: use projections onto the diagonal

$\mu_D^+ = \sum_{x \in D} \delta_x \qquad \mu_D^- = \sum_{x \in D} \delta_{\pi_\Delta(x)}$

Prop: $d_1(D, D') \leq W_1(\mu_D^+ + \mu_{D'}^-, \mu_{D'}^+ + \mu_D^-) \leq 2 d_1(D, D')$

Persistence diagrams as discrete measures





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

Pb: $d_1(D, D') \neq W_1(\mu_D, \mu_{D'})$ (W_1 does not even make sense here)

Solution: use projections onto the diagonal

$$\mu_D^+ = \sum_{x\in D} \delta_x \qquad \mu_D^- = \sum_{x\in D} \delta_{\pi_\Delta(x)}$$

Prop: $d_1(D, D') \le W_1(\mu_D^+ + \mu_{D'}^-, \, \mu_{D'}^+ + \mu_D^-) \le 2 d_1(D, D')$

Pb: W_1 is not cnsd, neither is d_1

Solution: relax the metric with slicing

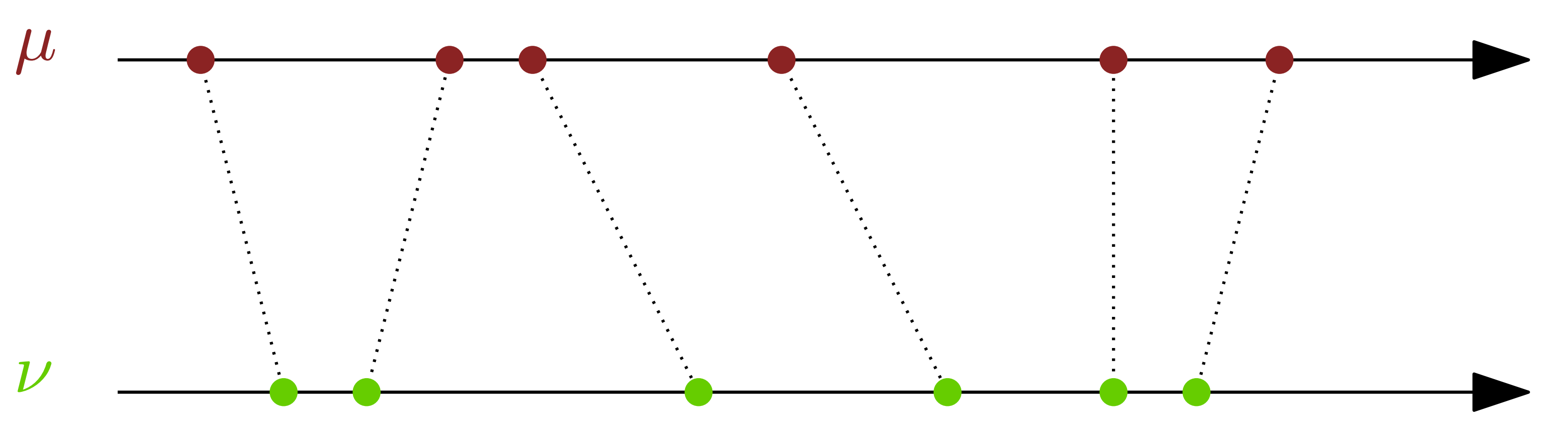
Sliced Wasserstein metric

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



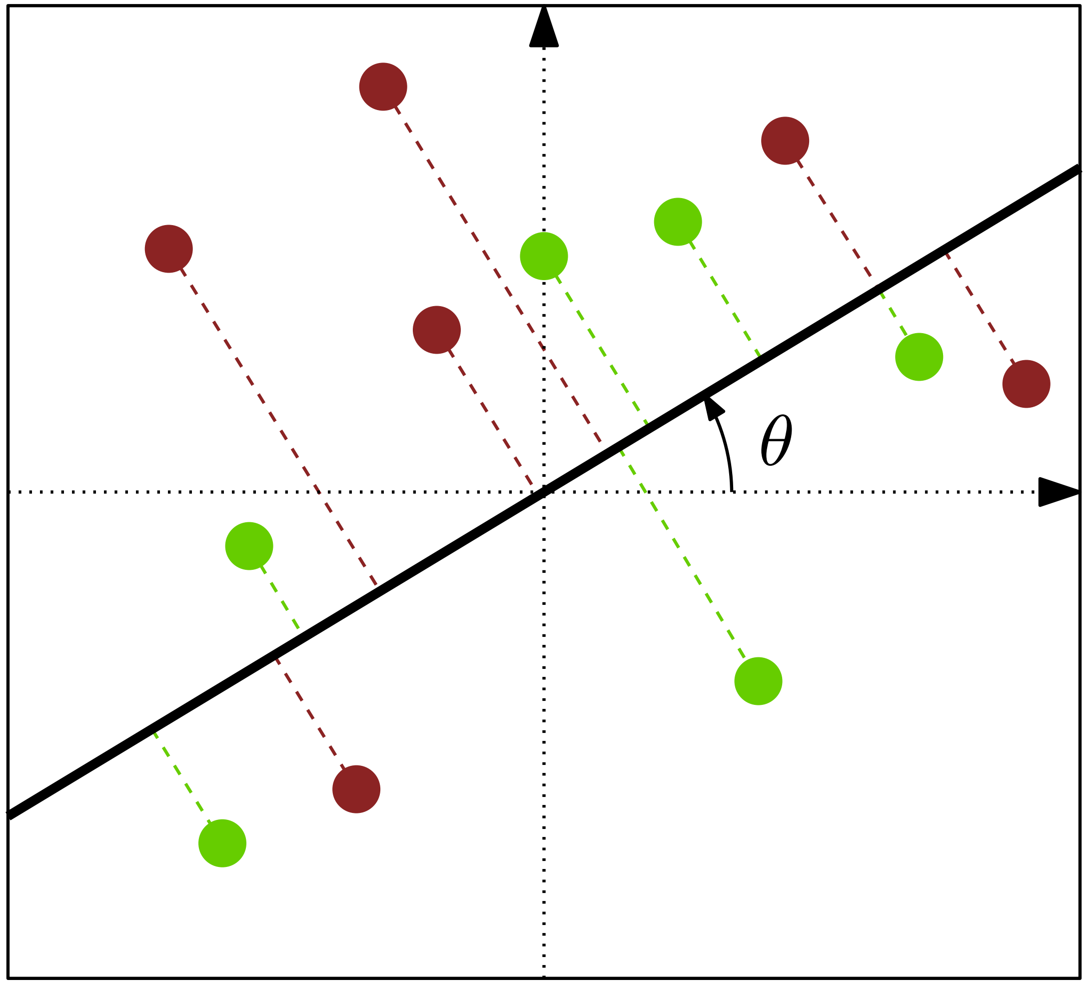
$\rightarrow W_1$ is cnsd and easy to compute

Sliced Wasserstein metric

Def (sliced Wasserstein distance): for D, D',

$$SW_1(D, D') = \frac{1}{2\pi} \int_{\theta \in \mathbb{S}^1} W_1(\pi_\theta \#(\mu_D^+ + \mu_{D'}^-), \pi_\theta \#(\mu_{D'}^+ + \mu_{D'}^-))\, d\theta$$

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



Sliced Wasserstein metric

Def (sliced Wasserstein distance): for D, D',

$$SW_1(D, D') = \frac{1}{2\pi} \int_{\theta \in \mathbb{S}^1} W_1(\pi_\theta \#(\mu_D^+ + \mu_{D'}^-), \pi_\theta \#(\mu_{D'}^+ + \mu_{D'}^-))\, d\theta$$

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

Prop: (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 SGD, etc.

- conditionally negative semidefinite

Sliced Wasserstein metric

Def (sliced Wasserstein distance): for D, D',

$$SW_1(D, D') = \frac{1}{2\pi} \int_{\theta \in \mathbb{S}^1} W_1(\pi_\theta \#(\mu_D^+ + \mu_{D'}^-), \pi_\theta \#(\mu_{D'}^+ + \mu_{D'}^-))\, d\theta$$

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

Corollary: k_{SW} is positive semidefinite.

Sliced Wasserstein metric

Def (sliced Wasserstein distance): for D, D',

$$SW_1(D,D') = \frac{1}{2\pi} \int_{\theta \in \mathbb{S}^1} W_1(\pi_\theta \#(\mu_D^+ + \mu_{D'}^-), \pi_\theta \#(\mu_{D'}^+ + \mu_{D'}^-))\, d\theta$$

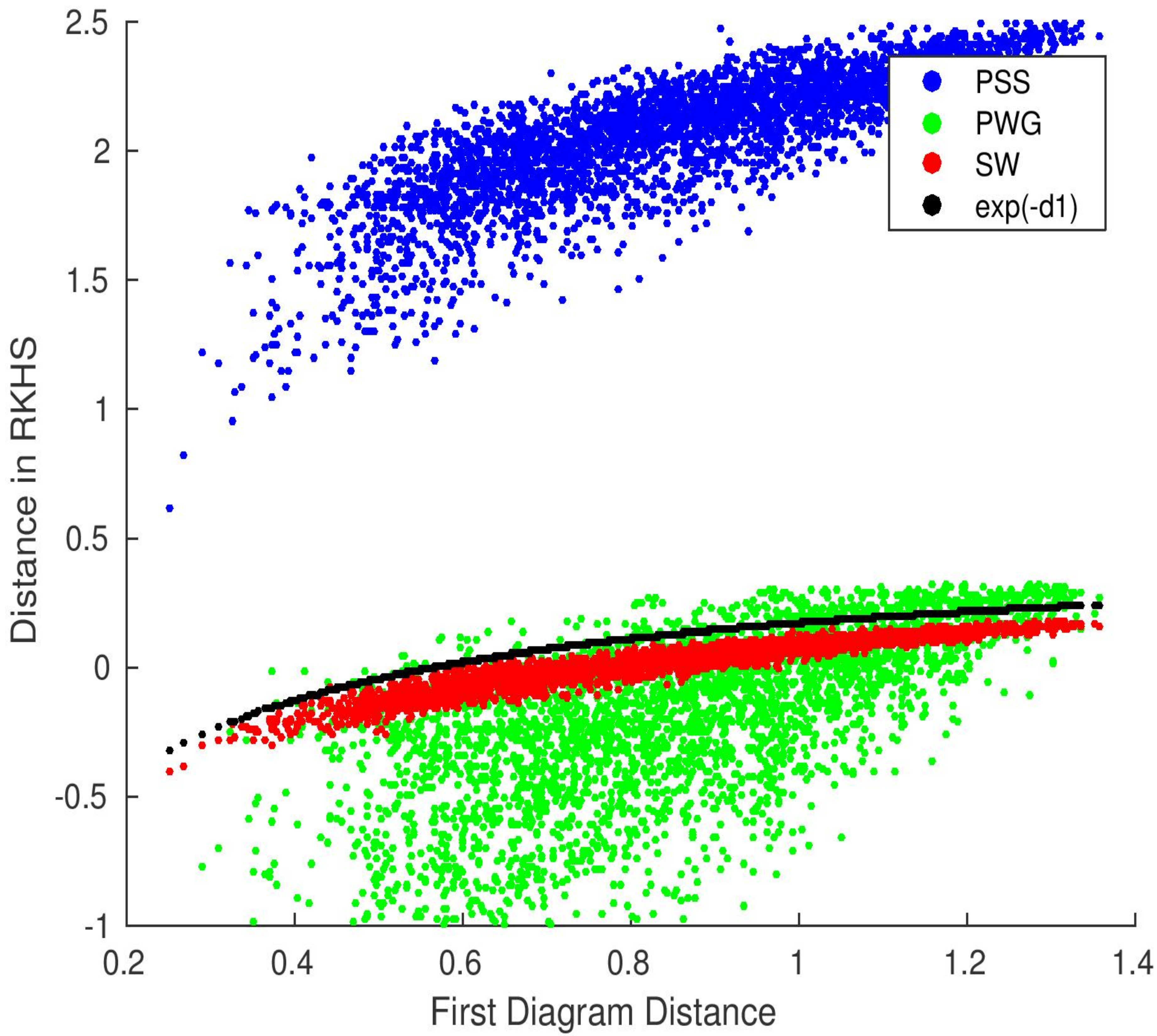
where $\pi_\theta =$ orthogonal projection onto line passing through origin with angle θ.

Corollary: k_{SW} is positive semidefinite.

Th: 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 $D, D' \in \mathcal{D}_N$,

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

Metric distortion in practice

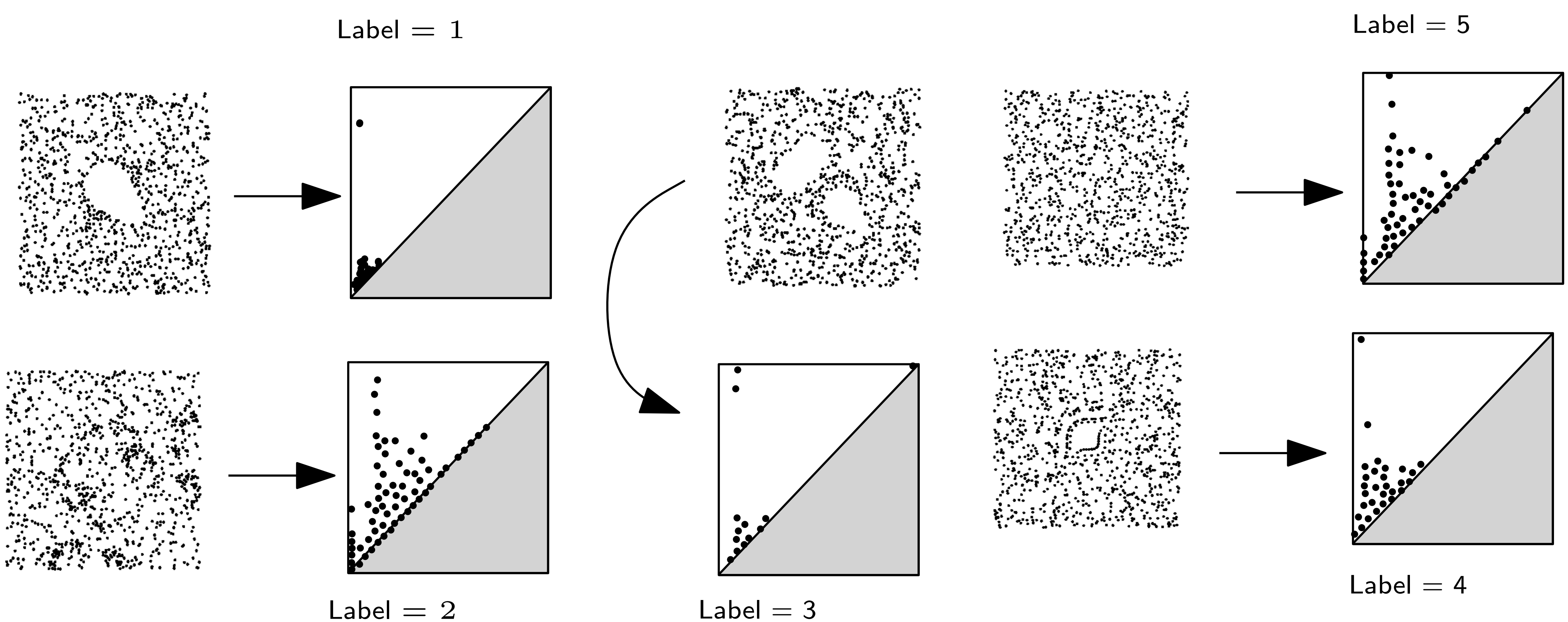


Application to supervised orbits classification

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)

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	$N \times 9183.4 \pm 65.6$	$N \times 69.2 \pm 0.9$	$385.8 \pm 0.2 + NC$

($\phi(\cdot)$ recomputed for each σ)

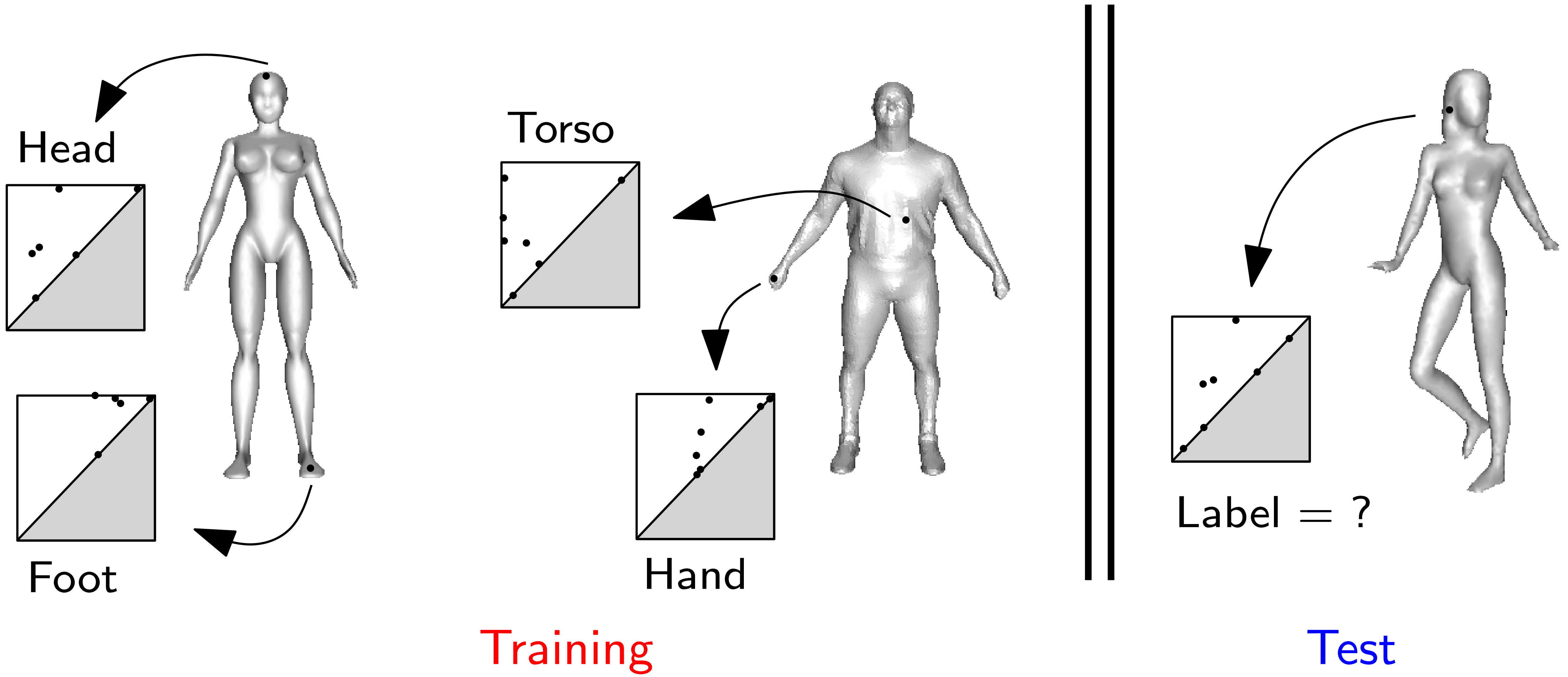
(SW_1 computed only once)

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 only TDA descriptors (kernels on barcodes):

	k_{PSS}	k_{PWG}	k_{SW}
Human	68.5 ± 2.0	64.2 ± 1.2	$\mathbf{74.0} \pm 0.2$
Airplane	55.4 ± 2.4	61.3 ± 2.9	$\mathbf{72.6} \pm 0.2$
Ant	86.3 ± 1.0	87.4 ± 0.5	$\mathbf{92.3} \pm 0.2$
FourLeg	67.0 ± 2.5	64.0 ± 0.6	$\mathbf{73.0} \pm 0.4$
Octopus	77.6 ± 1.0	78.6 ± 1.3	$\mathbf{85.2} \pm 0.5$
Bird	67.6 ± 1.8	$\mathbf{72.0} \pm 1.2$	67.0 ± 0.5
Fish	76.1 ± 1.6	$\mathbf{79.6} \pm 0.5$	75.0 ± 0.4

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 only TDA descriptors (kernels on barcodes):

	k_{PSS}	k_{PWG}	k_{SW}
Human	68.5 ± 2.0	64.2 ± 1.2	$\mathbf{74.0} \pm 0.2$
Airplane	55.4 ± 2.4	61.3 ± 2.9	$\mathbf{72.6} \pm 0.2$
Ant	86.3 ± 1.0	87.4 ± 0.5	$\mathbf{92.3} \pm 0.2$
FourLeg	67.0 ± 2.5	64.0 ± 0.6	$\mathbf{73.0} \pm 0.4$
Octopus	77.6 ± 1.0	78.6 ± 1.3	$\mathbf{85.2} \pm 0.5$
Bird	67.6 ± 1.8	$\mathbf{72.0} \pm 1.2$	67.0 ± 0.5
Fish	76.1 ± 1.6	$\mathbf{79.6} \pm 0.5$	75.0 ± 0.4

Metric Properties of Kernel Embeddings

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

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

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

Metric Properties of Kernel Embeddings

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

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

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

Prop: [C. Bauer 201?]
$\mathcal{H}$ Hilbert with dot product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and distance $\| \cdot \|_{\mathcal{H}}$. Assume $\mathrm{d}_{\mathcal{H}}$ and d_∞ or d_p are equivalent.

- $\mathcal{H} = \mathbb{R}^d \Rightarrow$ **Impossible**

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

- $\mathcal{H}$ separable $\Rightarrow$ either $A \to 0$ or $B \to +\infty$ when $L, N \to +\infty$

Metric Properties of Kernel Embeddings

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

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

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

Th: [C. Bauer 201?]

$\mathcal{H}$ RKHS of $k(D, D') = \exp\left(-\frac{\mathrm{d}(D,D')}{2\sigma}\right)$ with d cnsd. Assume d and d_∞ or d_p are equivalent.

Then either $A \to 0$ or $B \to +\infty$ when L or $N \to +\infty$.

Metric Properties of Kernel Embeddings

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

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

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

Th: [C. Bauer 201?]
$\mathcal{H}$ RKHS of $k(D, D') = \exp\left(-\frac{\mathrm{d}(D,D')}{2\sigma}\right)$ with d cnsd. Assume d and d_∞ or d_p are equivalent.

Then either $A \to 0$ or $B \to +\infty$ when L or $N \to +\infty$.

Proof?

Metric Properties of Kernel Embeddings

Let $\mathcal{D} = \{\oplus_{i=1}^{n(n+1)/2}(x, x+\alpha) = D_n : n \in \mathbb{N}^*\}$ complete.

We are going to show that $\mathcal{D}$ is compact, which is not true.

Metric Properties of Kernel Embeddings

Let $\mathcal{D} = \{\oplus_{i=1}^{n(n+1)/2}(x, x+\alpha) = D_n : n \in \mathbb{N}^*\}$ complete.

We are going to show that $\mathcal{D}$ is compact, which is not true.

$\forall n \in \mathbb{N}^*$, $\|\Phi(D_n)\|_{\mathcal{H}} = \sqrt{k(D_n, D_n)} = 1.$

Metric Properties of Kernel Embeddings

Let $\mathcal{D} = \{\oplus_{i=1}^{n(n+1)/2}(x, x+\alpha) = D_n : n \in \mathbb{N}^*\}$ complete.

We are going to show that $\mathcal{D}$ is compact, which is not true.

$\forall n \in \mathbb{N}^*$, $\|\Phi(D_n)\|_{\mathcal{H}} = \sqrt{k(D_n, D_n)} = 1.$

Alaoglu th: $\exists \psi : \Phi(D_{\psi(n)})$ weakly converges to $f \in \mathcal{H}$,

i.e. $\forall v \in \mathcal{H}$, $\langle v, \Phi(D_{\psi(n)})\rangle_{\mathcal{H}} \to \langle v, f\rangle_{\mathcal{H}}$

Metric Properties of Kernel Embeddings

Let $\mathcal{D} = \{\oplus_{i=1}^{n(n+1)/2}(x, x+\alpha) = D_n : n \in \mathbb{N}^*\}$ complete.

We are going to show that $\mathcal{D}$ is compact, which is not true.

$\forall n \in \mathbb{N}^*$, $\|\Phi(D_n)\|_{\mathcal{H}} = \sqrt{k(D_n, D_n)} = 1.$

Alaoglu th: $\exists \psi : \Phi(D_{\psi(n)})$ weakly converges to $f \in \mathcal{H}$,

i.e. $\forall v \in \mathcal{H}$, $\langle v, \Phi(D_{\psi(n)})\rangle_{\mathcal{H}} \to \langle v, f\rangle_{\mathcal{H}}$

Mercer th: $\exists \{e_i\}_{i\in\mathbb{N}^*}$ orthobasis of $\mathcal{H}$ s.t.

$k(D_p, D_q) = \sum_i \lambda_i e_i(D_p) e_i(D_q)$ + convergence is uniform

Metric Properties of Kernel Embeddings

Let $\mathcal{D} = \{\oplus_{i=1}^{n(n+1)/2}(x, x+\alpha) = D_n : n \in \mathbb{N}^*\}$ complete.

We are going to show that $\mathcal{D}$ is compact, which is not true.

$\forall n \in \mathbb{N}^*$, $\|\Phi(D_n)\|_{\mathcal{H}} = \sqrt{k(D_n, D_n)} = 1.$

Alaoglu th: $\exists \psi : \Phi(D_{\psi(n)})$ weakly converges to $f \in \mathcal{H}$,

i.e. $\forall v \in \mathcal{H}$, $\langle v, \Phi(D_{\psi(n)})\rangle_{\mathcal{H}} \to \langle v, f\rangle_{\mathcal{H}}$

Mercer th: $\exists \{e_i\}_{i \in \mathbb{N}^*}$ orthobasis of $\mathcal{H}$ s.t.

$k(D_p, D_q) = \sum_i \lambda_i e_i(D_p) e_i(D_q)$ + convergence is uniform

Hence $1 = \lim_{n\to+\infty} \|D_{\psi(n)}\|_{\mathcal{H}}^2 = \lim_{n\to+\infty} \sum_i \lambda_i e_i(D_{\psi(n)})^2$

Metric Properties of Kernel Embeddings

Let $\mathcal{D} = \{\oplus_{i=1}^{n(n+1)/2}(x, x+\alpha) = D_n : n \in \mathbb{N}^*\}$ complete.

We are going to show that $\mathcal{D}$ is compact, which is not true.

$\forall n \in \mathbb{N}^*$, $\|\Phi(D_n)\|_{\mathcal{H}} = \sqrt{k(D_n, D_n)} = 1.$

Alaoglu th: $\exists \psi : \Phi(D_{\psi(n)})$ weakly converges to $f \in \mathcal{H}$,

i.e. $\forall v \in \mathcal{H}$, $\langle v, \Phi(D_{\psi(n)})\rangle_{\mathcal{H}} \to \langle v, f\rangle_{\mathcal{H}}$

Mercer th: $\exists \{e_i\}_{i\in\mathbb{N}^*}$ orthobasis of $\mathcal{H}$ s.t.

$k(D_p, D_q) = \sum_i \lambda_i e_i(D_p) e_i(D_q)$ + convergence is uniform

Hence $1 = \lim_{n\to+\infty} \|D_{\psi(n)}\|^2_{\mathcal{H}} = \lim_{n\to+\infty} \sum_i \lambda_i e_i(D_{\psi(n)})^2$

$= \sum_i \lambda_i \left(\lim_{n\to+\infty} e_i(D_{\psi(n)})^2\right) = \sum_i \lambda_i \left(\lim_{n\to+\infty} \langle e_i, \Phi(D_{\psi(n)})\rangle^2_{\mathcal{H}}\right)$

RKHS

Metric Properties of Kernel Embeddings

Let $\mathcal{D} = \{\oplus_{i=1}^{n(n+1)/2}(x, x+\alpha) = D_n : n \in \mathbb{N}^*\}$ complete.

We are going to show that $\mathcal{D}$ is compact, which is not true.

$\forall n \in \mathbb{N}^*$, $\|\Phi(D_n)\|_{\mathcal{H}} = \sqrt{k(D_n, D_n)} = 1.$

Alaoglu th: $\exists \psi : \Phi(D_{\psi(n)})$ weakly converges to $f \in \mathcal{H}$,

i.e. $\forall v \in \mathcal{H}$, $\langle v, \Phi(D_{\psi(n)})\rangle_{\mathcal{H}} \to \langle v, f\rangle_{\mathcal{H}}$

Mercer th: $\exists \{e_i\}_{i\in\mathbb{N}^*}$ orthobasis of $\mathcal{H}$ s.t.

$k(D_p, D_q) = \sum_i \lambda_i e_i(D_p) e_i(D_q)$ + convergence is uniform

Hence $1 = \lim_{n\to+\infty} \|D_{\psi(n)}\|_{\mathcal{H}}^2 = \lim_{n\to+\infty} \sum_i \lambda_i e_i(D_{\psi(n)})^2$

$= \sum_i \lambda_i \left(\lim_{n\to+\infty} e_i(D_{\psi(n)})^2\right) = \sum_i \lambda_i \left(\lim_{n\to+\infty} \langle e_i, \Phi(D_{\psi(n)})\rangle_{\mathcal{H}}^2\right)$

$= \sum_i \lambda_i \langle e_i, f\rangle_{\mathcal{H}}^2 = \|f\|_{\mathcal{H}}^2 \implies \|\Phi(D_{\psi(n)}) - f\|_{\mathcal{H}} \to 0.$

Metric Properties of Kernel Embeddings

Let $\mathcal{D} = \{\oplus_{i=1}^{n(n+1)/2}(x, x+\alpha) = D_n : n \in \mathbb{N}^*\}$ complete.

We are going to show that $\mathcal{D}$ is compact, which is not true.

$\forall n \in \mathbb{N}^*$, $\|\Phi(D_n)\|_{\mathcal{H}} = \sqrt{k(D_n, D_n)} = 1$.

Alaoglu th: $\exists \psi : \Phi(D_{\psi(n)})$ weakly converges to $f \in \mathcal{H}$,

i.e. $\forall v \in \mathcal{H}$, $\langle v, \Phi(D_{\psi(n)})\rangle_{\mathcal{H}} \to \langle v, f\rangle_{\mathcal{H}}$

Mercer th: $\exists \{e_i\}_{i\in\mathbb{N}^*}$ orthobasis of $\mathcal{H}$ s.t.

$k(D_p, D_q) = \sum_i \lambda_i e_i(D_p) e_i(D_q)$ + convergence is uniform

Hence $1 = \lim_{n\to+\infty} \|D_{\psi(n)}\|_{\mathcal{H}}^2 = \lim_{n\to+\infty} \sum_i \lambda_i e_i(D_{\psi(n)})^2$

$= \sum_i \lambda_i \left(\lim_{n\to+\infty} e_i(D_{\psi(n)})^2\right) = \sum_i \lambda_i \left(\lim_{n\to+\infty} \langle e_i, \Phi(D_{\psi(n)})\rangle_{\mathcal{H}}^2\right)$

$= \sum_i \lambda_i \langle e_i, f\rangle_{\mathcal{H}}^2 = \|f\|_{\mathcal{H}}^2 \implies \|\Phi(D_{\psi(n)}) - f\|_{\mathcal{H}} \to 0.$

Since $\Phi(\mathcal{D})$ is complete because $\mathcal{D}$ is, $f \in \Phi(\mathcal{D})$ i.e. $f = \Phi(D_N)$.

Finally $D_{\psi(n)} \to D_N$ because of metric equivalence.

Thank you!!