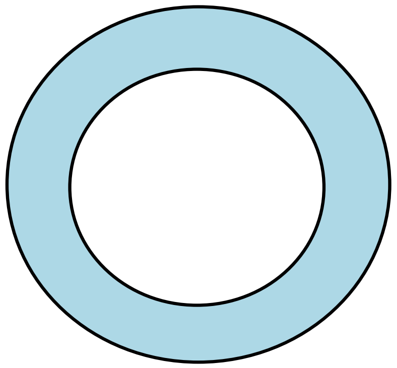
Topological Machine Learning (I): Statistics and Representations

1. Topological Inference
2. Persistence Representations
3. Learning Representations

Topological Machine Learning (I): Statistics and Representations

1. **Topological Inference**
2. Persistence Representations
3. Learning Representations

Topological inference setting



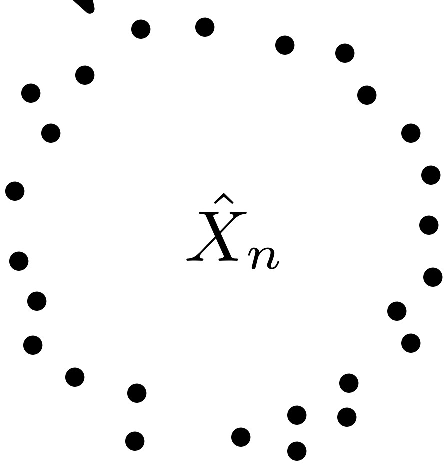
(X, d) metric space

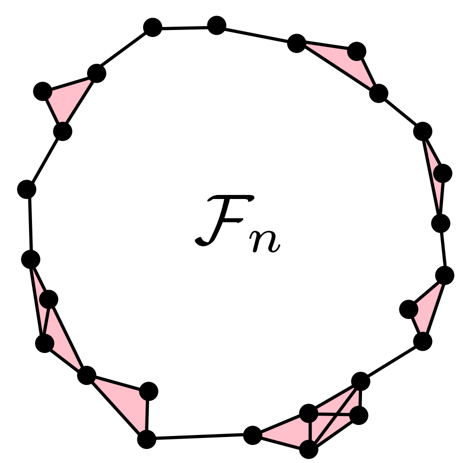
μ probability measure with compact support X_μ

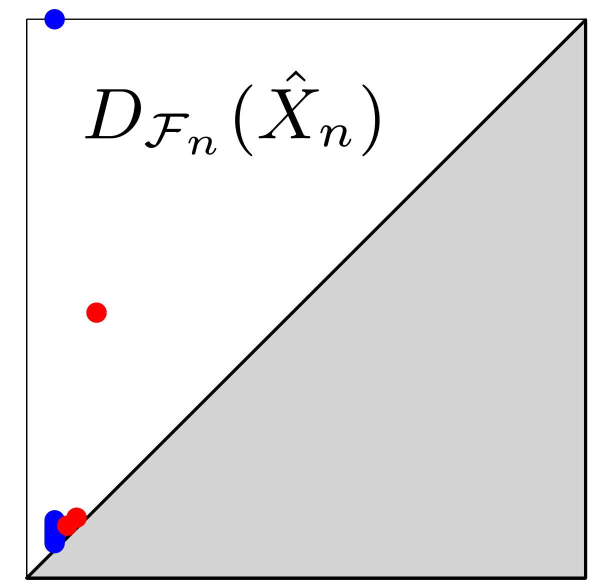
Examples:

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

Sample n points according to μ.

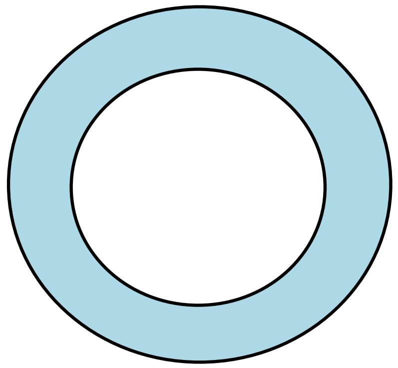






Questions: Statistical properties of $D_{\mathcal{F}_n}(\hat{X}_n)$? $D_{\mathcal{F}_n}(\hat{X}_n) \to ?$ as $n \to +\infty$?

Topological inference setting



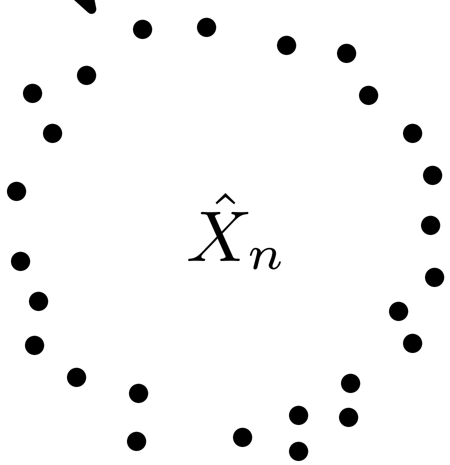
(X, d) metric space

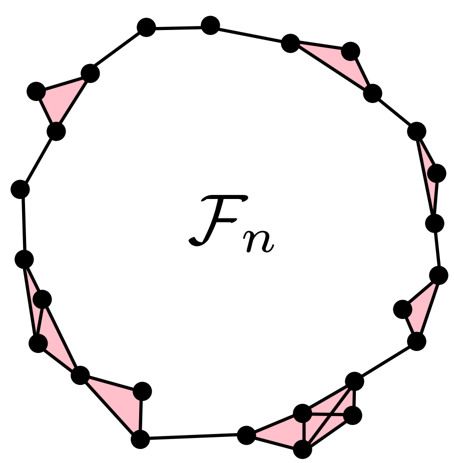
μ probability measure with compact support X_μ

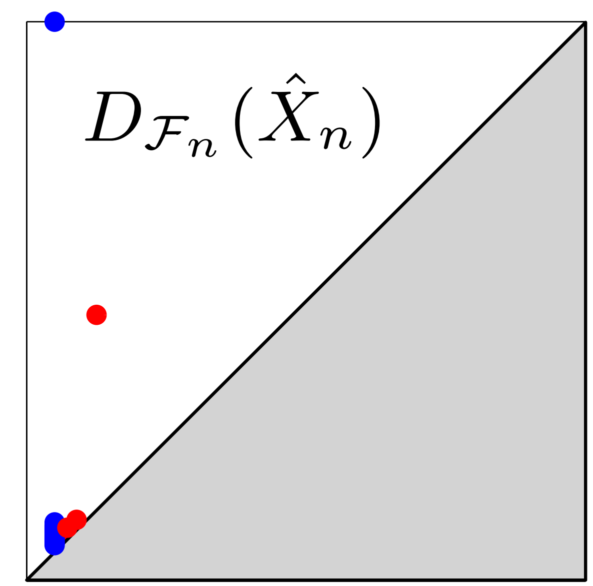
Sample n points according to μ.

Examples:

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







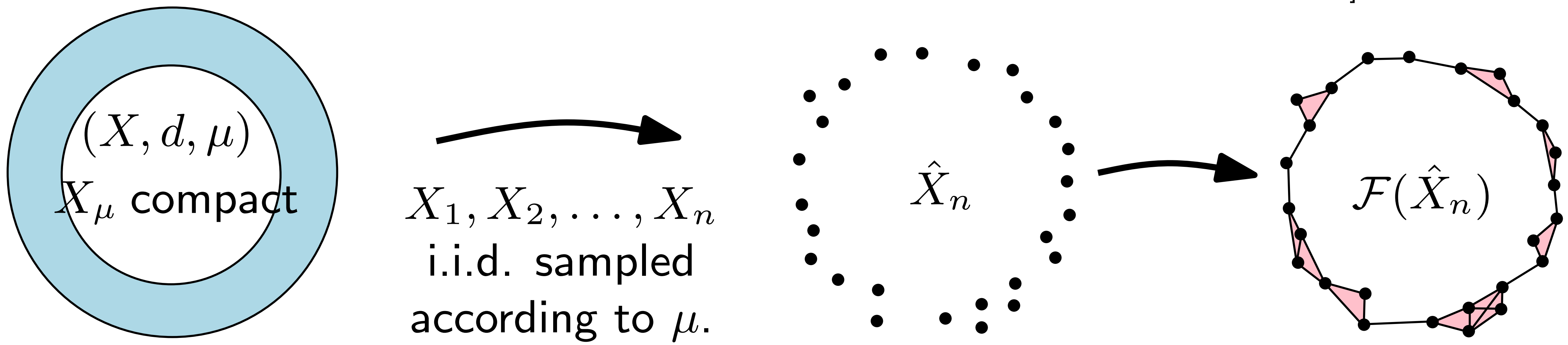
Stability thm: $\mathrm{d}_b(D_{\mathcal{F}}(X_\mu), D_{\mathcal{F}_n}(\hat{X}_n)) \leq d_H(X_\mu, \hat{X}_n)$

So, for any $\varepsilon > 0$,

$$\mathbb{P}\left(\mathrm{d}_b\left(D_{\mathcal{F}}(X_\mu), D_{\mathcal{F}_n}(\hat{X}_n)\right) > \varepsilon\right) \leq \mathbb{P}\left(d_H(X_\mu, \hat{X}_n) > \varepsilon\right)$$

Deviation inequality

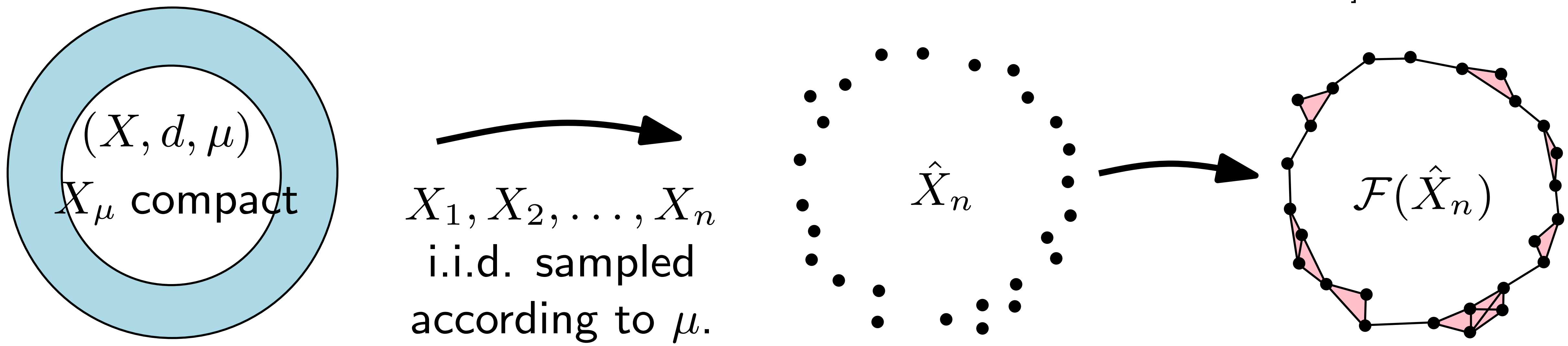
[*Convergence rates for persistence diagram estimation in Topological Data Analysis*, Chazal, Glisse, Labruère, Michel ICML, 2014]



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

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 $\varepsilon > 0$:

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

Moreover, $\lim_{n\to\infty} \mathbb{P}\left(\mathrm{d}_b(D_{\mathcal{F}}(X_\mu), D_{\mathcal{F}_n}(\hat{X}_n)) \leq C\left(\frac{\log n}{n}\right)^{1/b}\right) = 1$, where C is a constant that only depends on a and b.

Minimax rate of convergence

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

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

Minimax rate of convergence

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

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

Thm: One has the following:

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

where the constant C depends only on a and b. Assume moreover that there exists a non isolated point x in X and let $\{x_n\}_n$ be a sequence in $X \setminus \{x\}$ such that $d(x, x_n) \sim (an)^{-1/b}$. Then for any estimator $\hat{D}_n$ of $D_{\mathcal{F}}(X_\mu)$:

$$\sup_{\mu \in \mathcal{P}} \mathbb{E}\left[\mathrm{d}_b(D_{\mathcal{F}}(X_\mu), \hat{D}_n)\right] \geq C' d(x, x_n)$$

where C' is an absolute constant.

Minimax rate of convergence

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

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

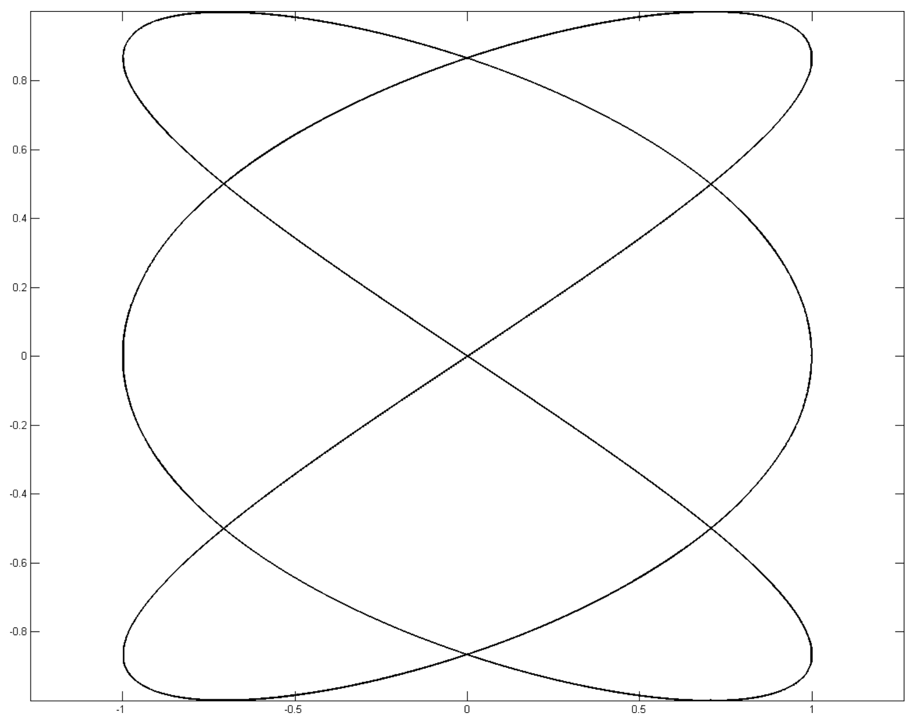
Proof: Apply Le Cam's lemma:

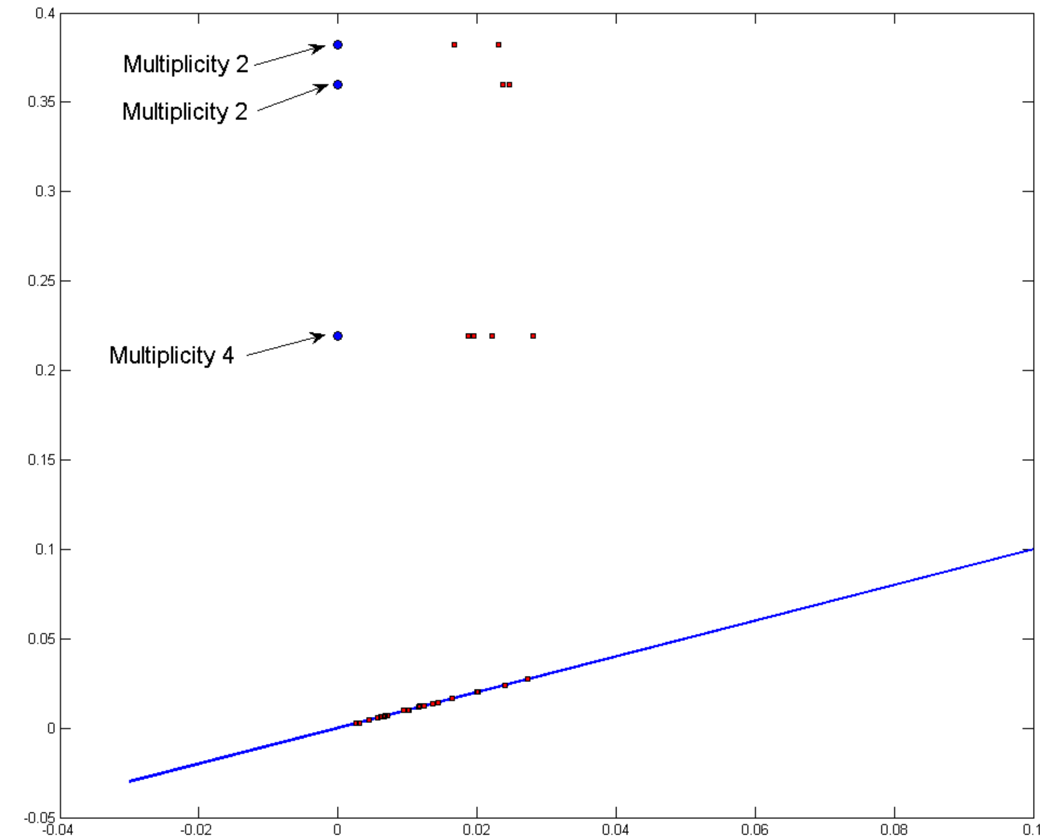
$$\sup_{\mu \in \mathcal{P}} \mathbb{E}\left[\mathrm{d}_b(D_{\mathcal{F}}(X_\mu), \hat{D}_n)\right] \geq \frac{1}{8}\mathrm{d}_b(D_{\mathcal{F}}(X_{\mu_0}), D_{\mathcal{F}}(X_{\mu_1}))(1 - \mathrm{TV}(\mu_0, \mu_1))^{2n}$$

with $\mu_0 = \delta_x$ and $\mu_1 = \frac{1}{n}\delta_{x_n} + (1 - \frac{1}{n})\delta_x$.

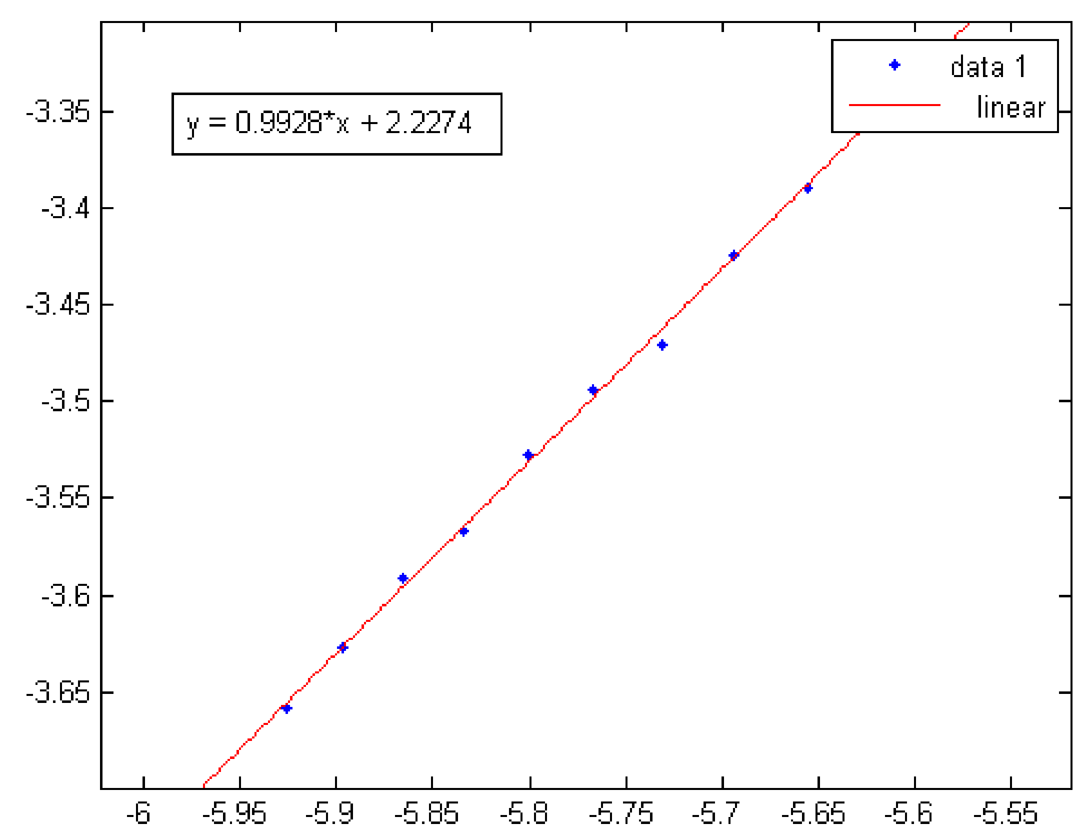
Numerical illustrations

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



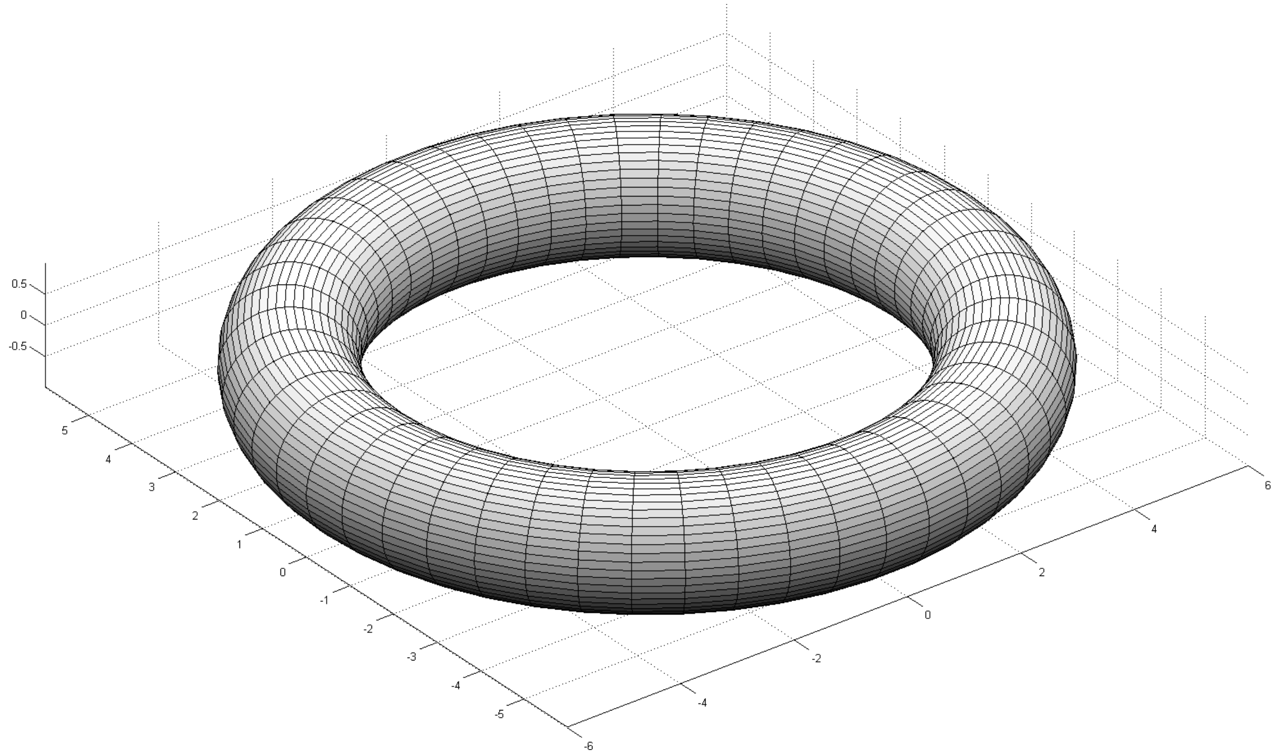


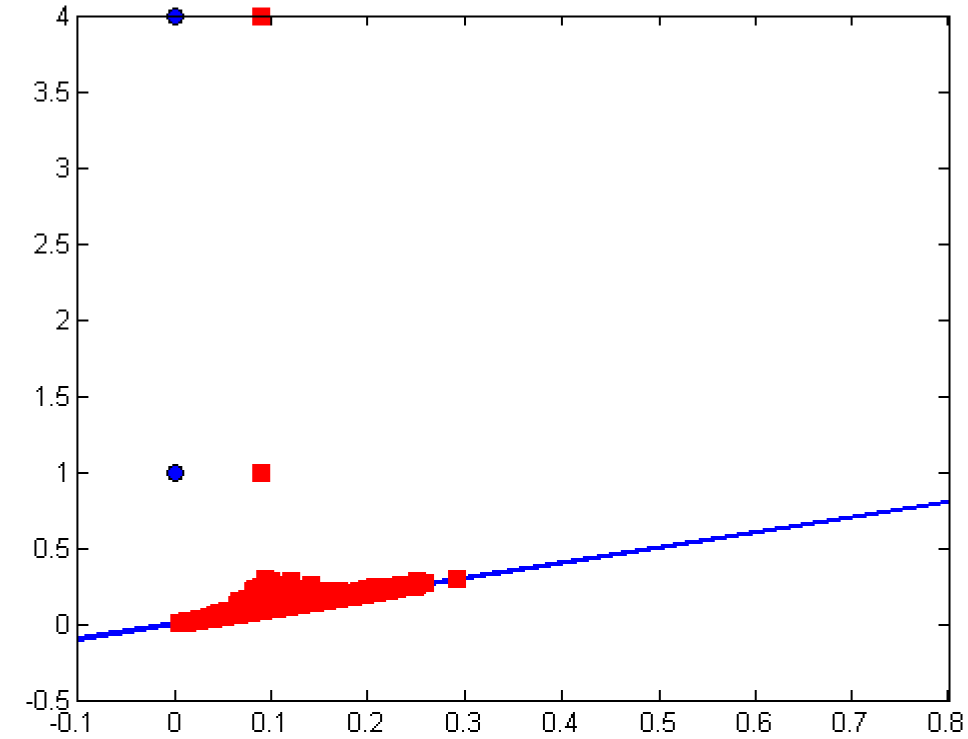
- μ: unif. measure on Lissajous curve X_μ.
- $\mathcal{F}$: distance to X_μ in $\mathbb{R}^2$.
- sample 300 sets of n points for various n.
- compute $\hat{\mathbb{E}}_n = \hat{\mathbb{E}}[\mathrm{d}_b(D_{\mathcal{F}}(X_\mu), D_{\mathcal{F}_n}(\hat{X}_n))]$.
- plot $\log(\hat{\mathbb{E}}_n)$ as a function of $\log(\log(n)/n)$.



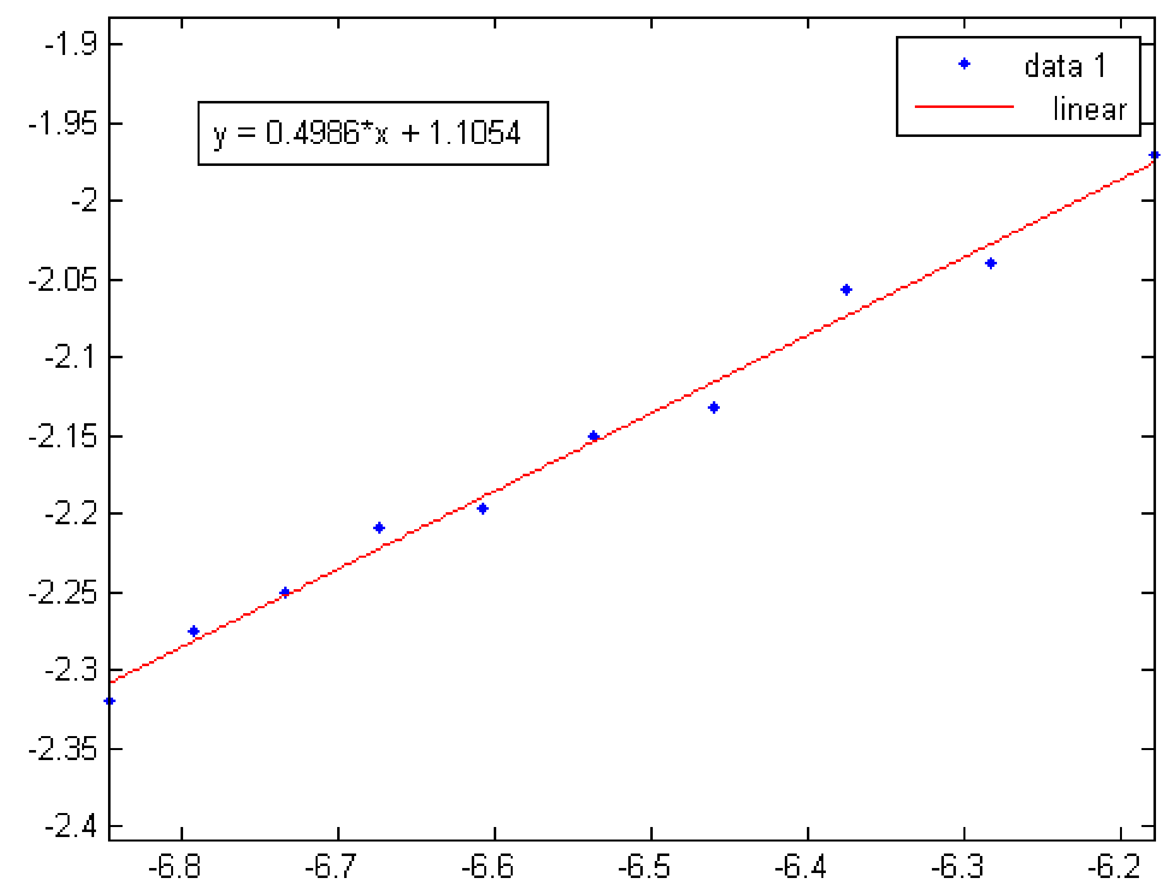
Numerical illustrations

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





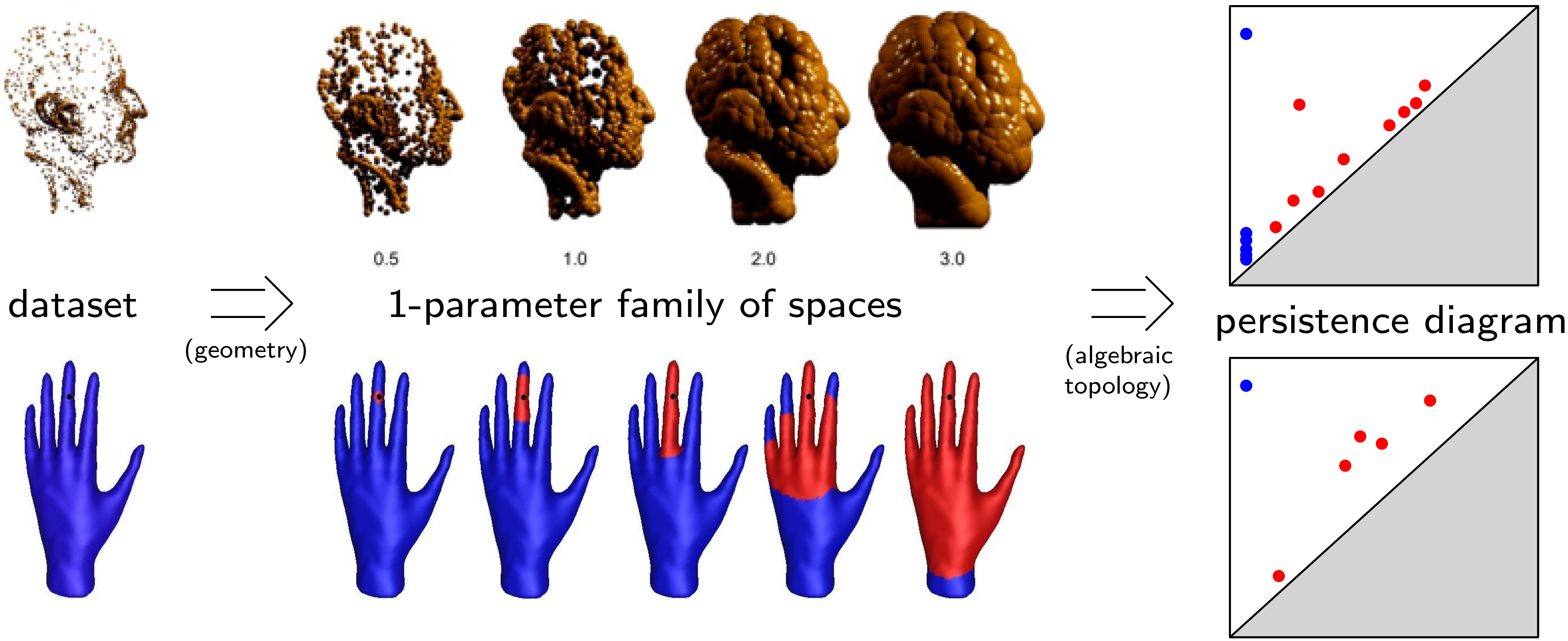
- μ: unif. measure on a torus X_μ.
- $\mathcal{F}$: distance to X_μ in $\mathbb{R}^3$.
- sample 300 sets of n points for various n.
- compute $\hat{\mathbb{E}}_n = \hat{\mathbb{E}}[\mathrm{d}_b(D_{\mathcal{F}}(X_\mu), D_{\mathcal{F}_n}(\hat{X}_n))]$.
- plot $\log(\hat{\mathbb{E}}_n)$ as a function of $\log(\log(n)/n)$.



Topological Machine Learning (I): Statistics and Representations

1. Topological Inference
2. **Persistence Representations**
3. Learning Representations

Persistence diagrams as data descriptors



Pros:

- strong invariance and stability:

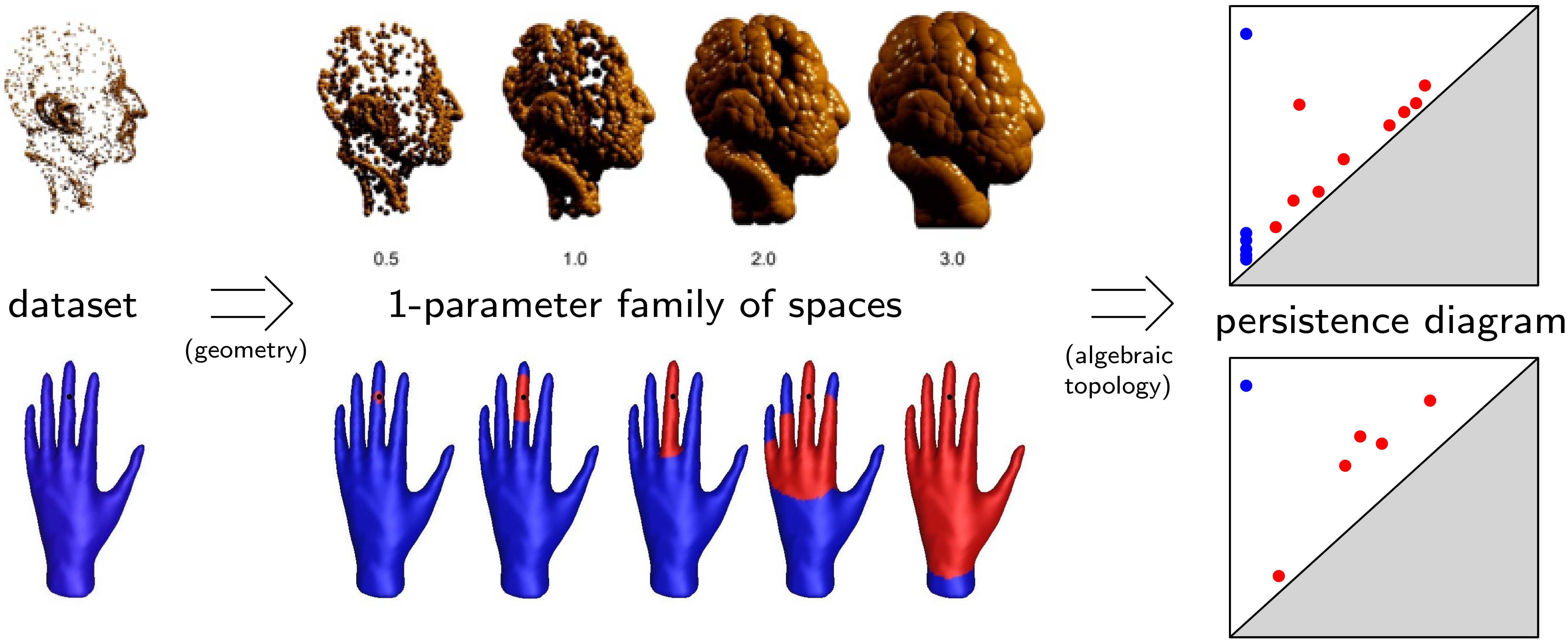
$$\mathrm{d}_b(D_{\text{Rips}}(X), D_{\text{Rips}}(Y)) \leq d_{GH}(X, Y)$$

- information of a different nature
- flexible and versatile

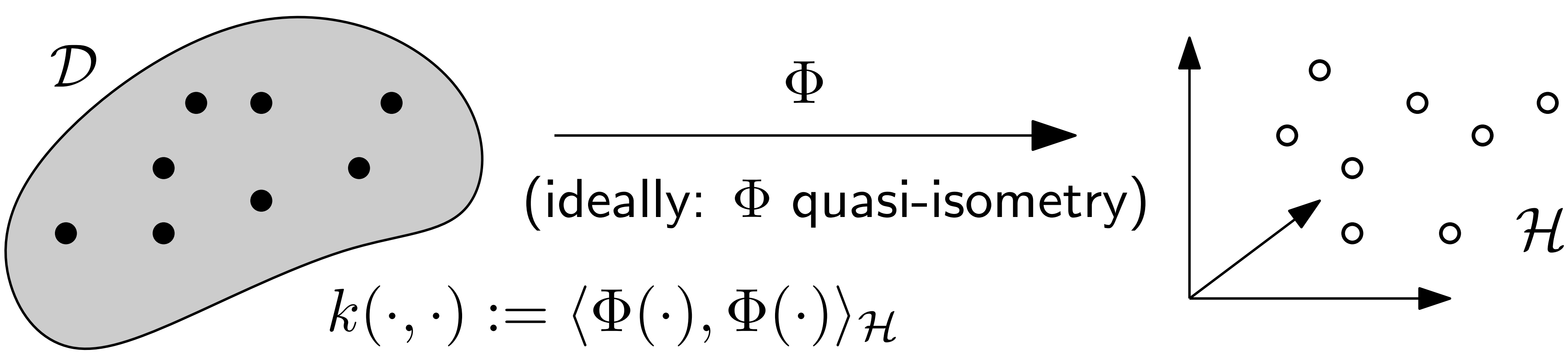
Cons:

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

Persistence diagrams as data descriptors



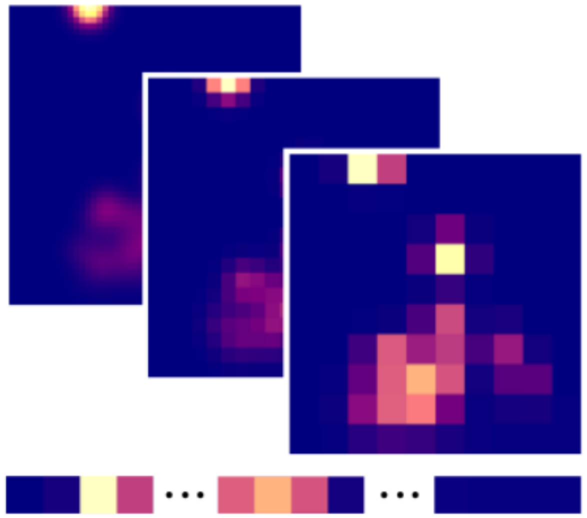
Solution: use **representations** = embeddings of PDs into Hilbert space



Representations of persistence diagrams

State of the Art: define ϕ via:

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

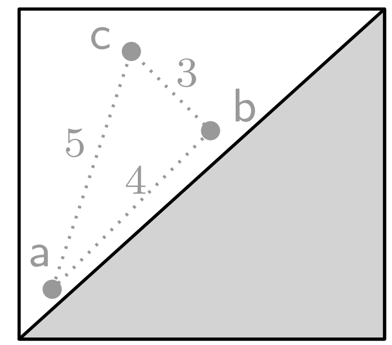


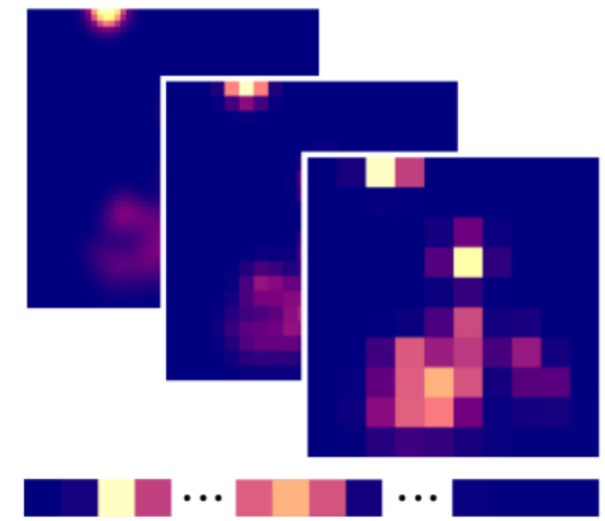
Representations of persistence diagrams

State of the Art: define ϕ via:

- images — [*Persistence Images: A Stable Vector Representation of Persistent Homology*, Adams et al., JMLR, 2017]
- finite metric spaces — [*Stable topological signatures for points on 3D shapes*, C., Oudot, Ovsjanikov, SGP, 2015]

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





Representations of persistence diagrams

State of the Art: define ϕ via:

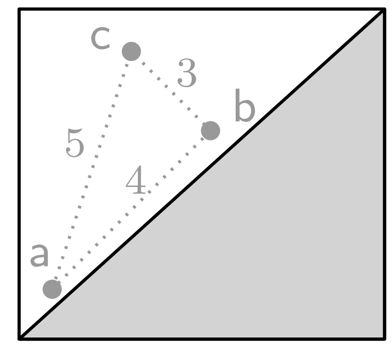
- images

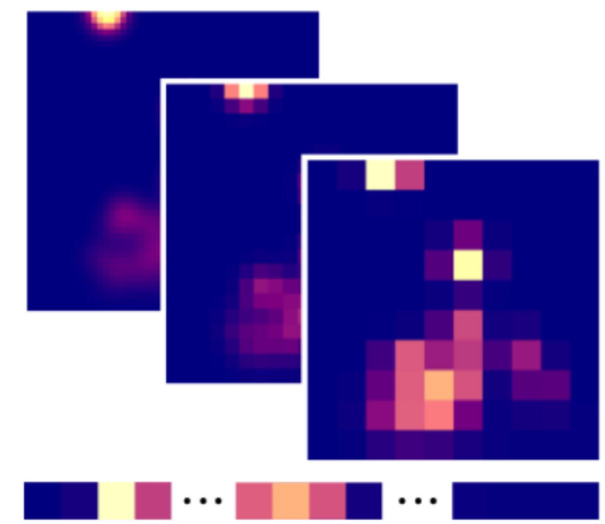
  [*Persistence Images: A Stable Vector Representation of Persistent Homology*, Adams et al., JMLR, 2017]

- finite metric spaces

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

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





- polynomial roots or evaluations

  [*Tropical coordinates on the space of persistence barcodes*, Kalisnik, FoCM, 2018]

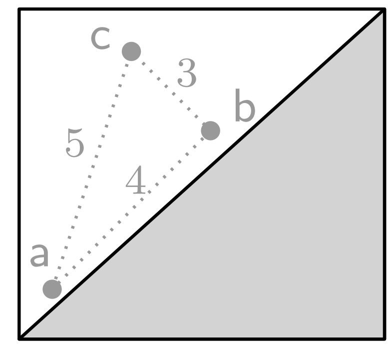
$$\{p_1, \ldots, p_n\} \mapsto (P_1(p_1, \ldots, p_n), \ldots, P_r(p_1, \ldots, p_n), \ldots)$$

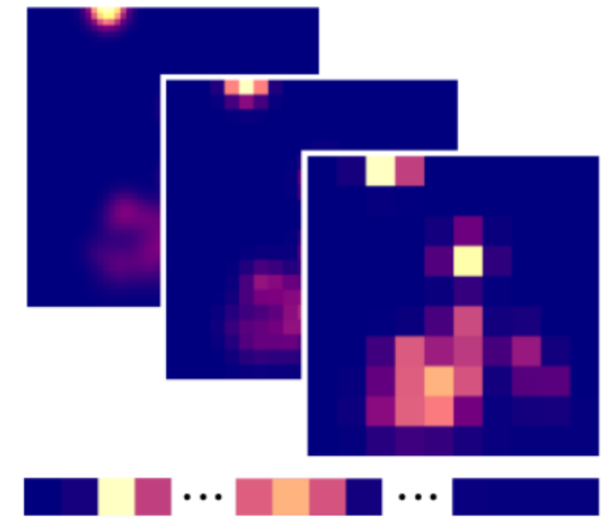
Representations of persistence diagrams

State of the Art: define ϕ via:

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

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



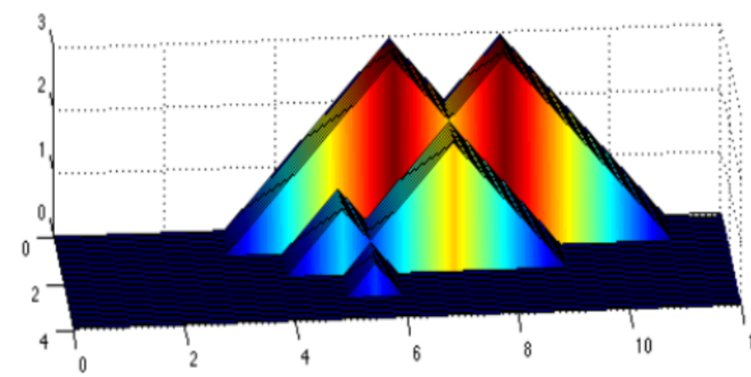


- finite metric spaces [*Stable topological signatures for points on 3D shapes*, C., Oudot, Ovsjanikov, SGP, 2015]

- polynomial roots or evaluations [*Tropical coordinates on the space of persistence barcodes*, Kalisnik, FoCM, 2018]

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

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



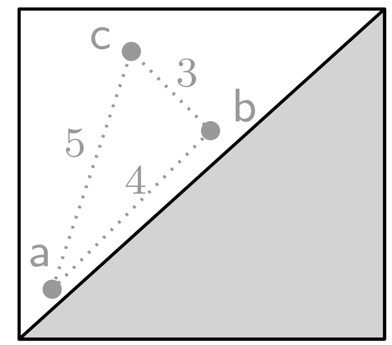
Representations of persistence diagrams

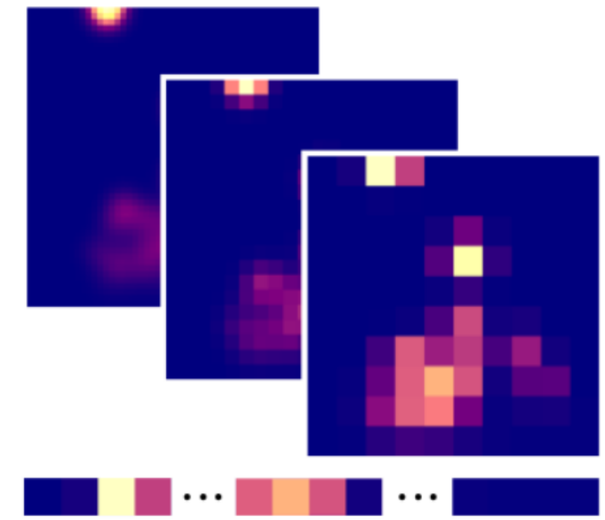
State of the Art: define ϕ via:

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

$$\begin{array}{c} \\ a \\ b \\ c \end{array} \begin{array}{ccc} a & b & c \\ \hline \end{array}$$

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



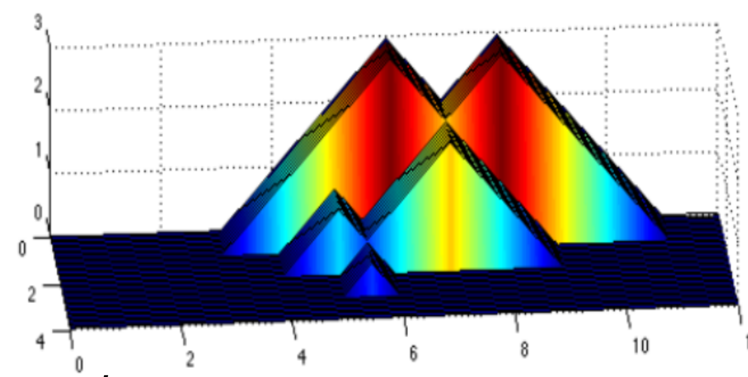


- finite metric spaces [*Stable topological signatures for points on 3D shapes*, C., Oudot, Ovsjanikov, SGP, 2015]

- polynomial roots or evaluations [*Tropical coordinates on the space of persistence barcodes*, Kalisnik, FoCM, 2018]

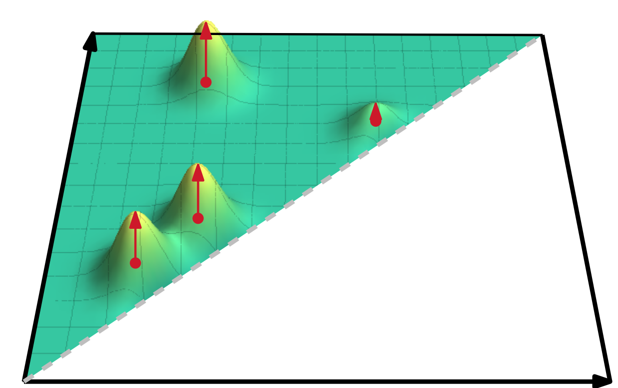
$$\{p_1, \ldots, p_n\} \mapsto (P_1(p_1, \ldots, p_n), \ldots, P_r(p_1, \ldots, p_n), \ldots)$$

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



- discrete measures:

$\rightarrow$ Fisher information [*Persistence Fisher kernel: a Riemannian manifold kernel for persistence diagrams*, Le, Yamada, NeurIPS, 2018]

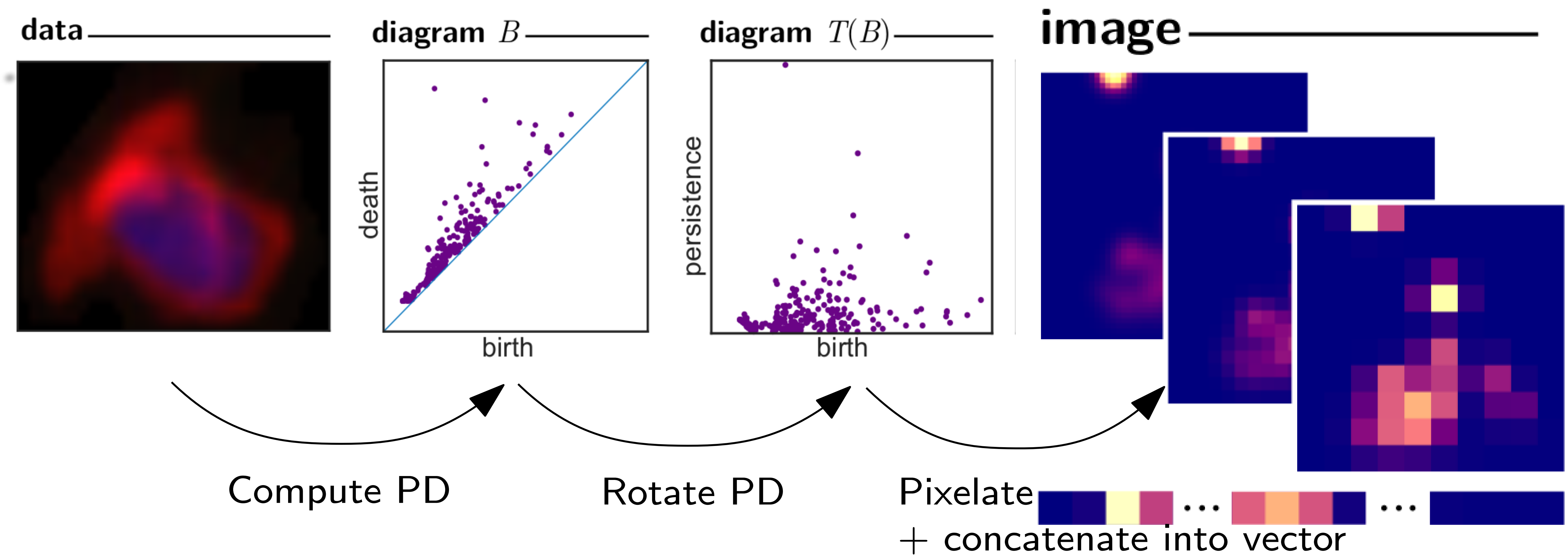


$\rightarrow$ convolution with weighted kernel [*Persistence weighted Gaussian kernel for topological data analysis*, Kusano, Hiraoka, Fukumizu, ICML, 2016]

$\rightarrow$ heat diffusion [*A stable multi-scale kernel for topological machine learning*, Reininghaus et al., CVPR, 2015]

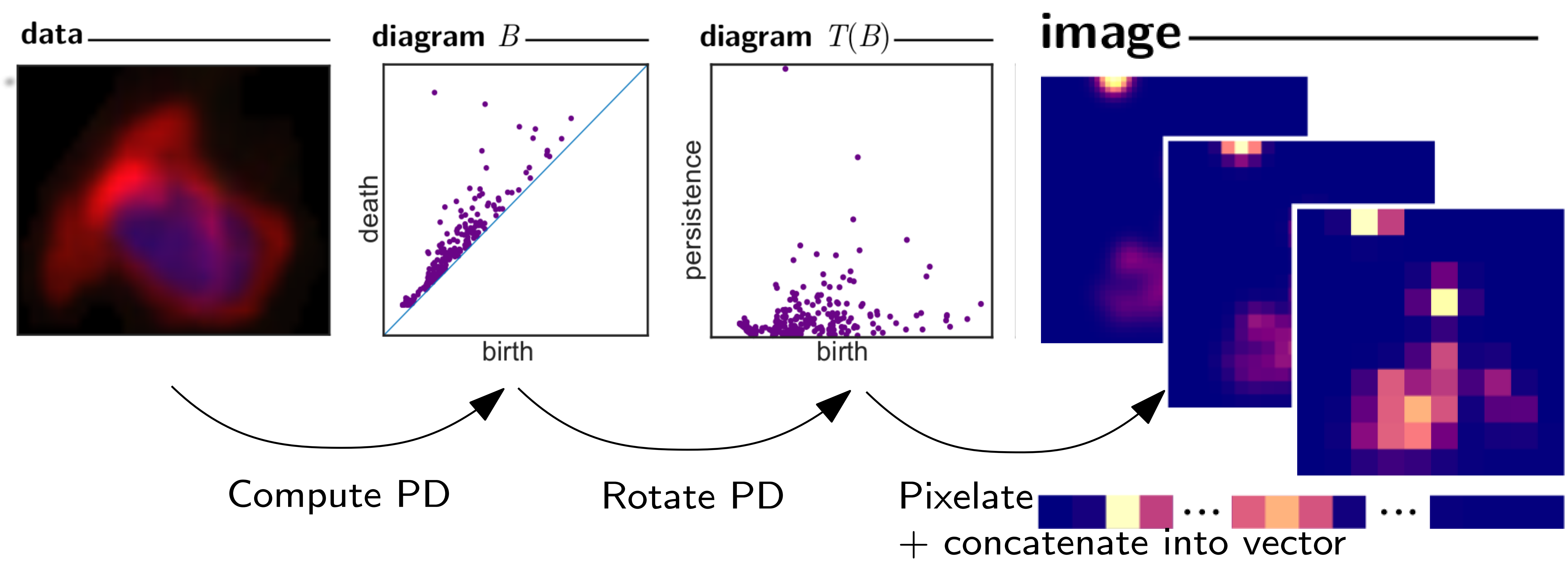
Persistence image

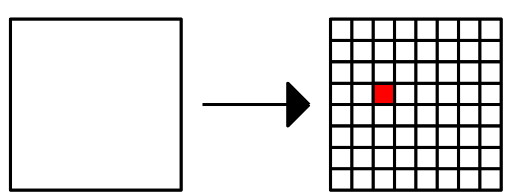
[*Persistence Images: A Stable Vector Representation of Persistent Homology*, Adams et al., JMLR, 2017]



Persistence image

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



Discretize plane into a grid: 

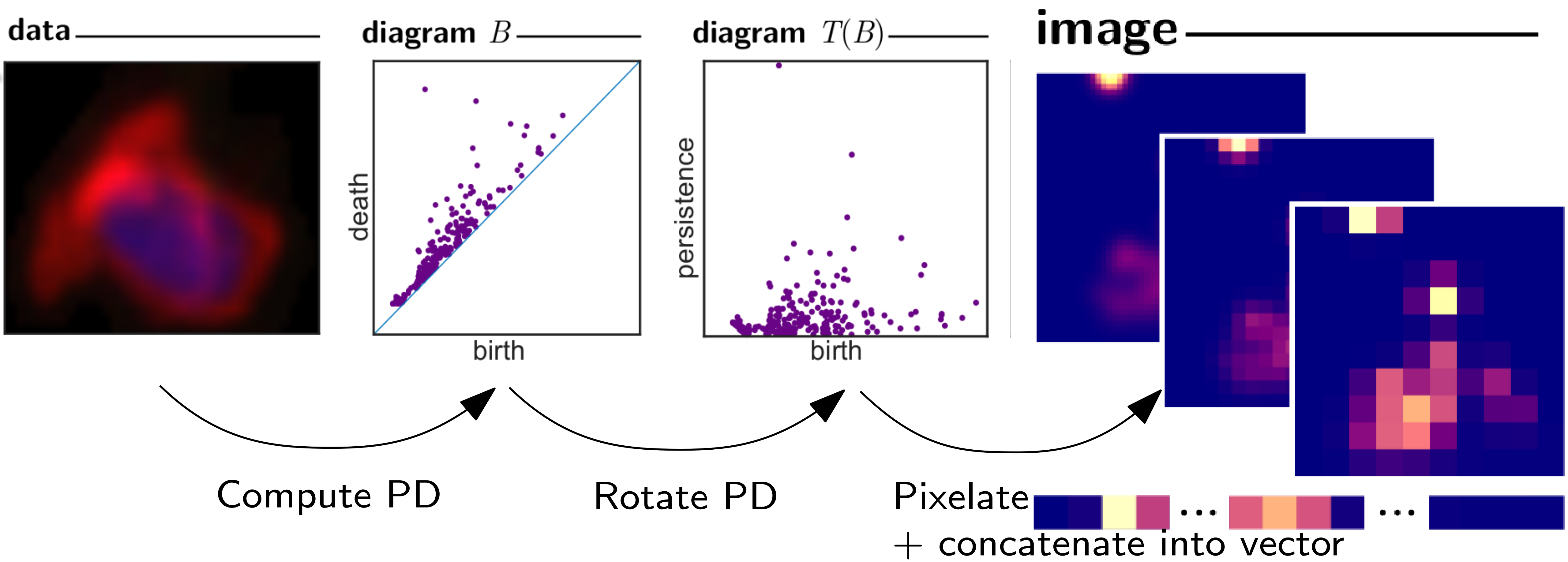
For each grid pixel P, compute $I(P) = \sum_{p \in D} \int\int_P w(p) \cdot \phi_p$.

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

Ex: $\phi_p = \mathcal{N}(p, \sigma)$.

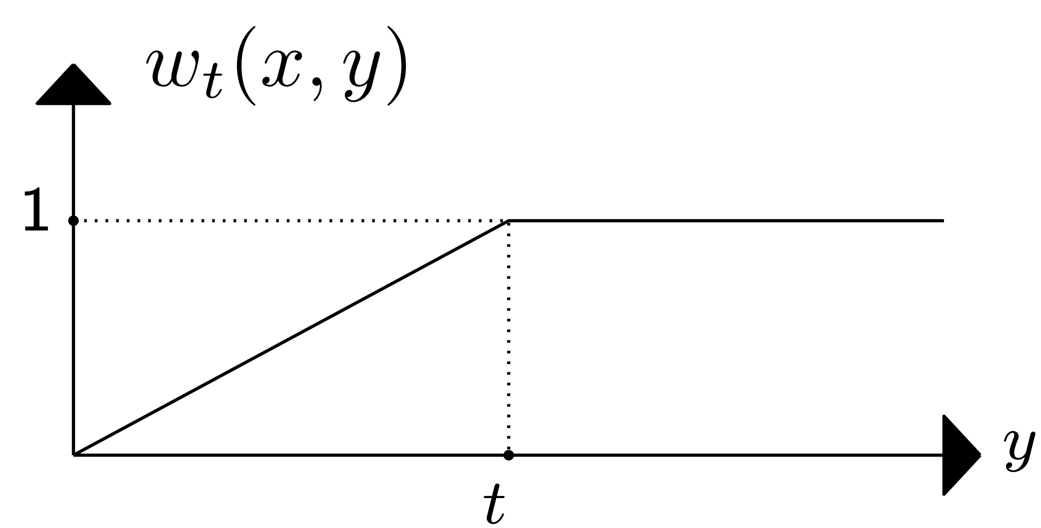
Persistence image

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



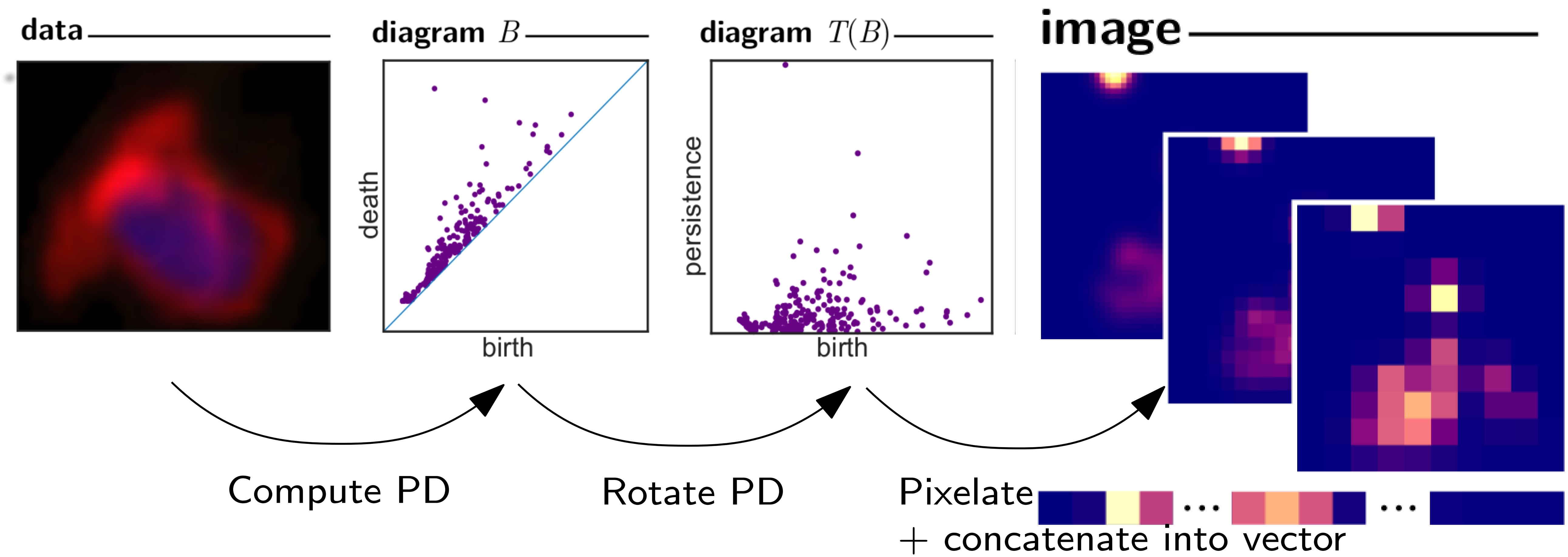
Weight functions that preserve stability must satisfy $w(p) \to 0$ when $d(p, \Delta) \to 0$.

[*Understanding the topology and the geometry of the persistence diagram space via optimal partial transport*, Divol, Lacombe, JACT, 2020]



Persistence image

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

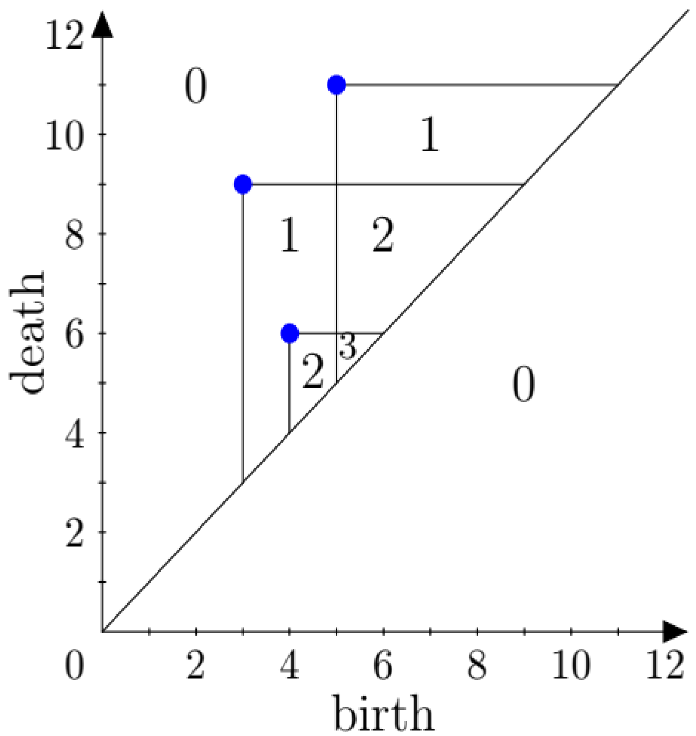


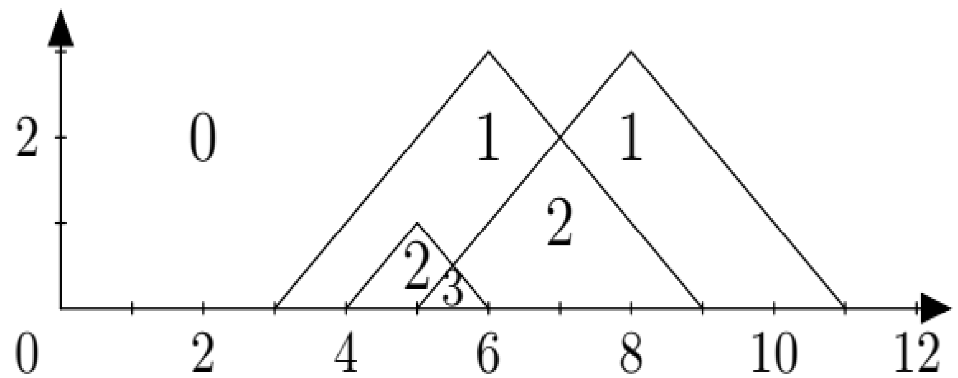
Prop: The following inequalities hold:

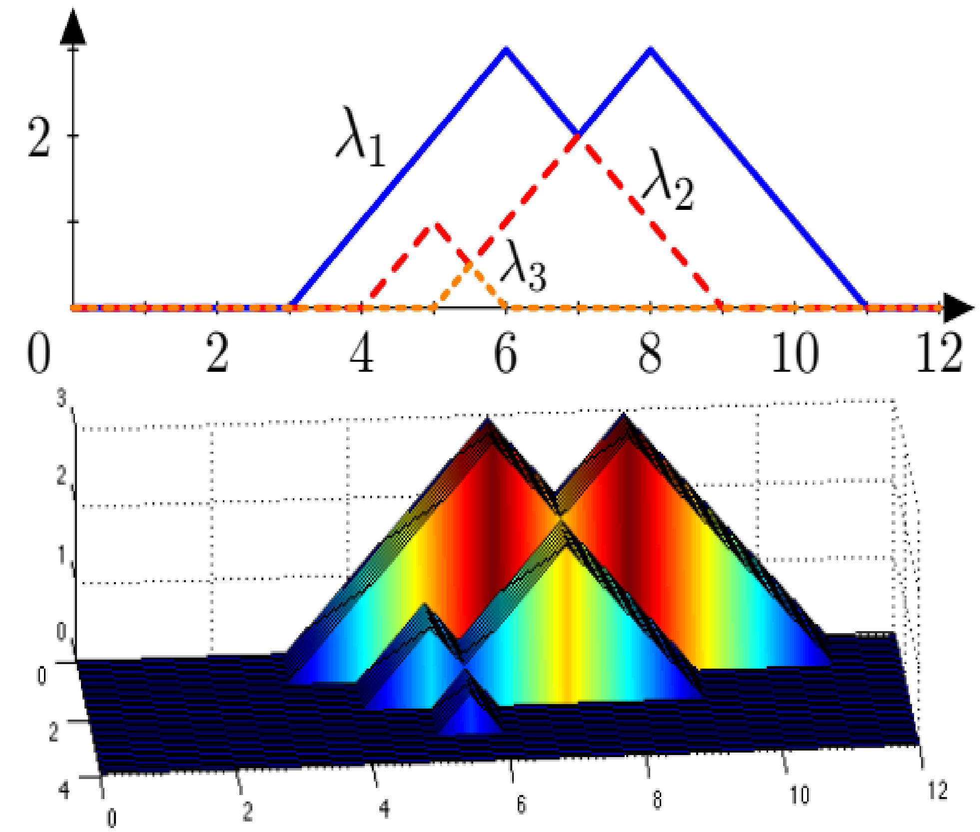
- $\|\mathrm{PI}(D) - \mathrm{PI}(D')\|_\infty \leq C(w, \phi_p)\,\mathrm{d}_1(D, D')$.
- $\|\mathrm{PI}(D) - \mathrm{PI}(D')\|_2 \leq \sqrt{d}C(w, \phi_p)\,\mathrm{d}_1(D, D')$.

Persistence landscape

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





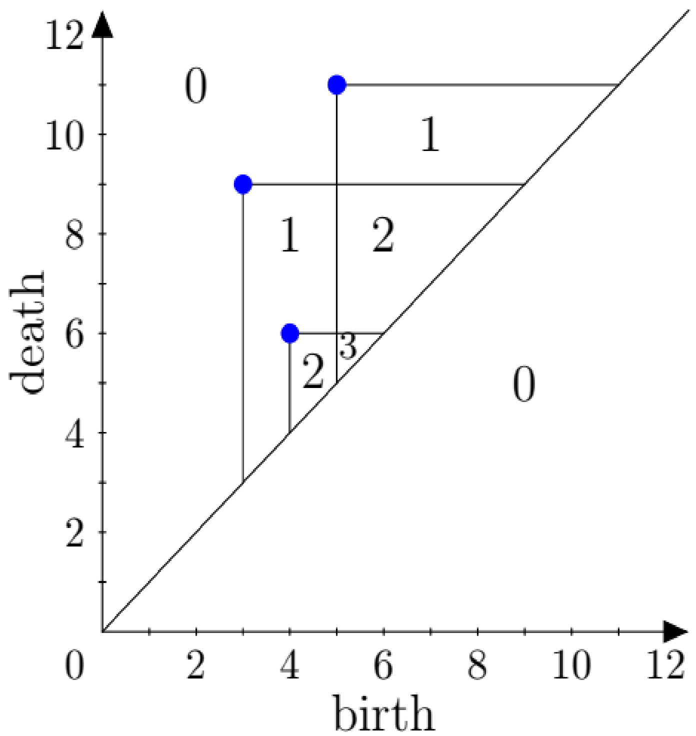


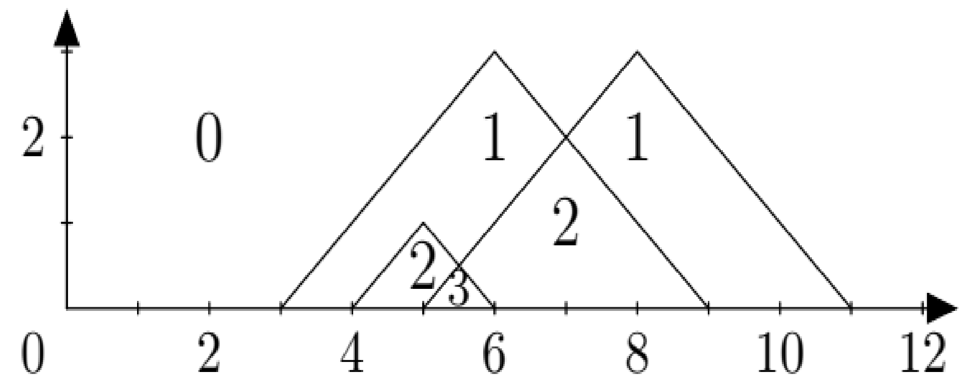
Rotate PD
Compute rank function

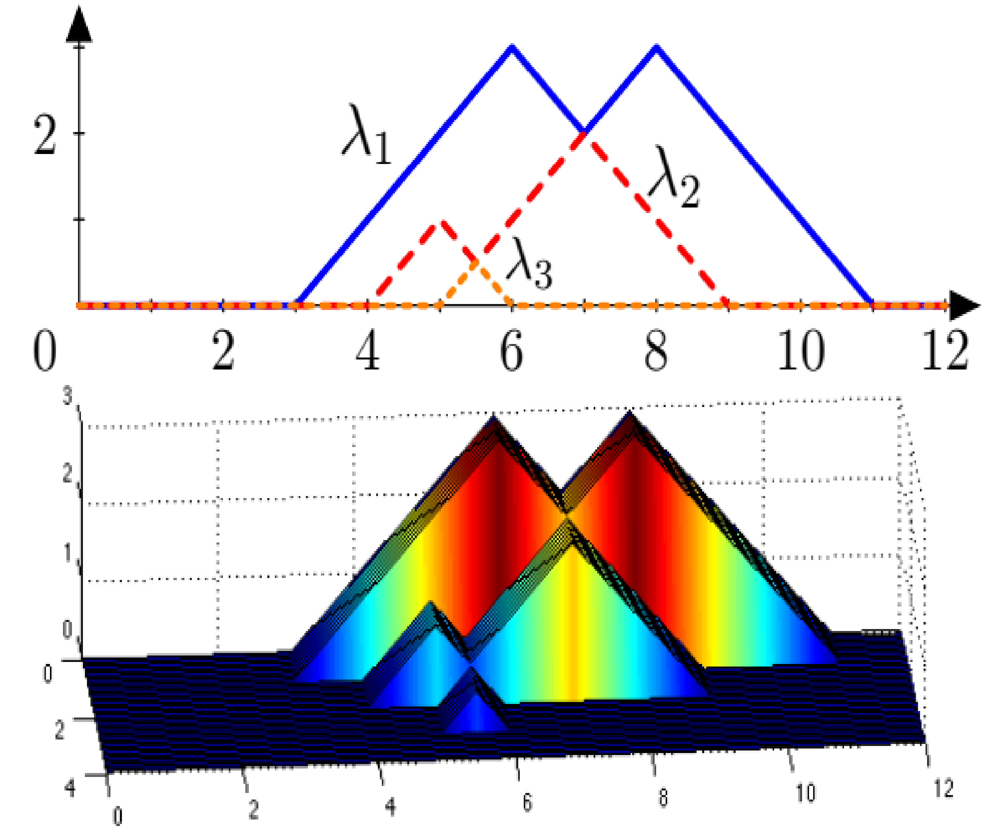
Use boundaries of rank function

Persistence landscape

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







Rotate PD
Compute rank function

Use boundaries of rank function

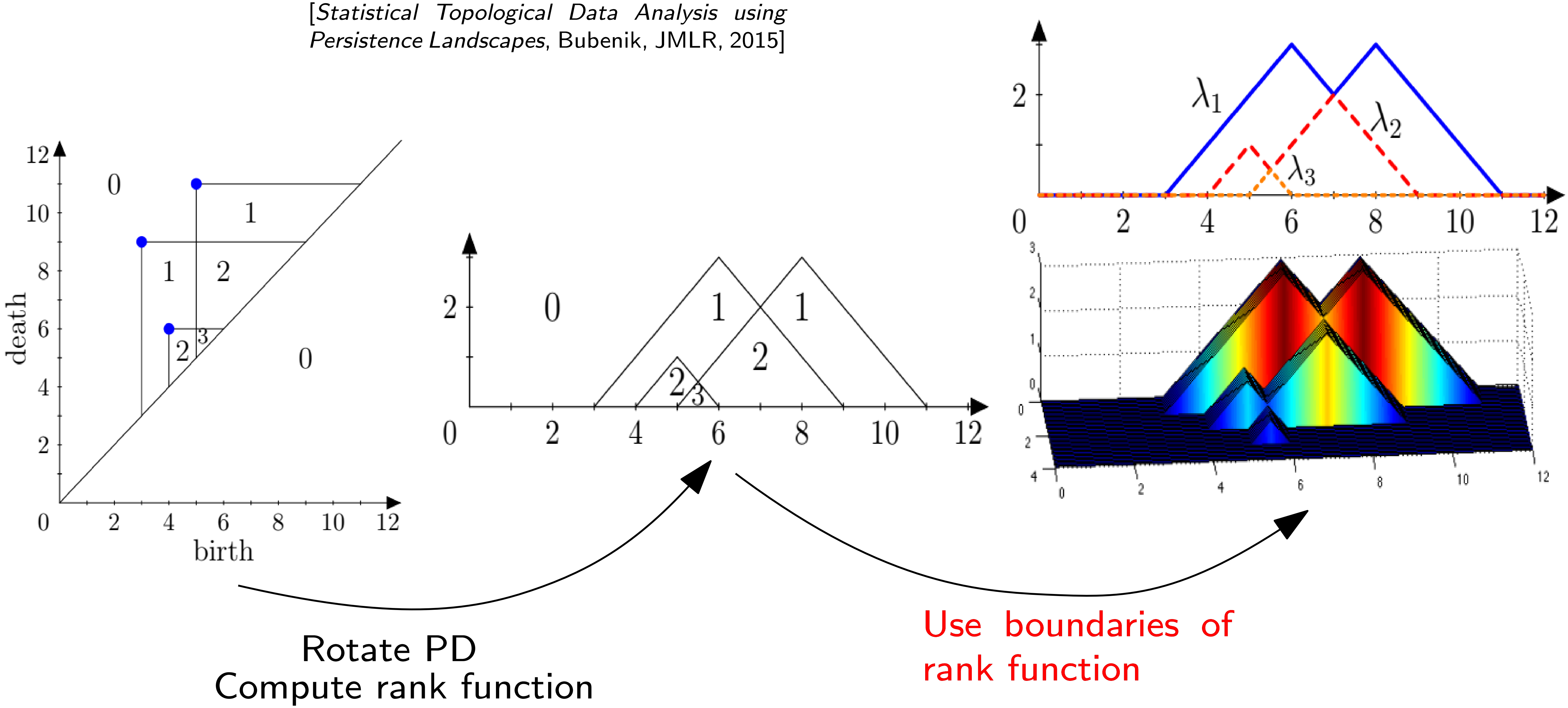
$$x \leq y \Longrightarrow f^{-1}(-\infty, x) \subseteq f^{-1}(-\infty, y)$$

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

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

Persistence landscape

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



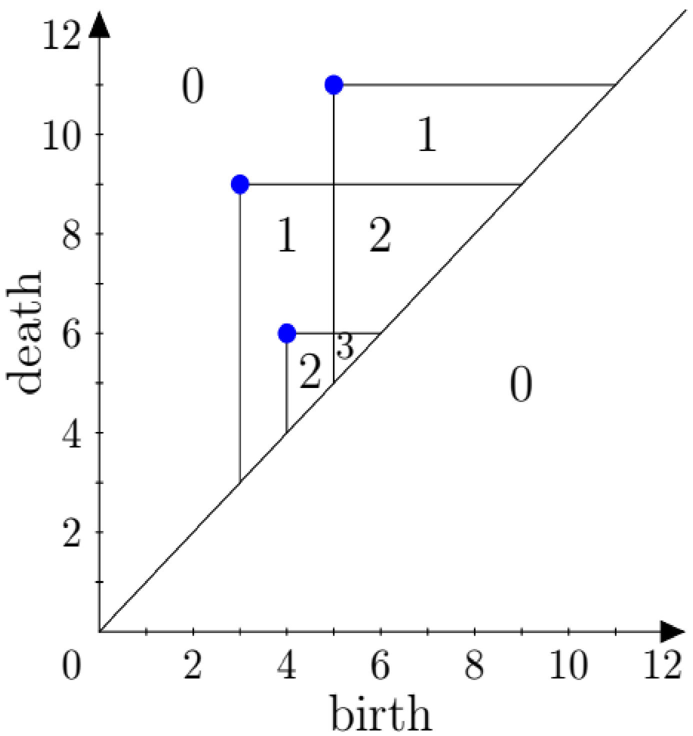
Boundaries of rank function: $\lambda_i(t) = \sup\{s \geq 0 : \lambda(t-s, t+s) \geq i\}$

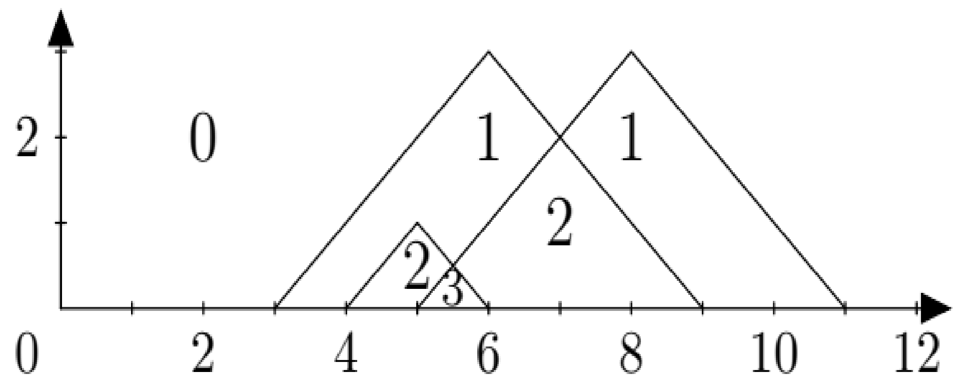
Landscape $\Lambda : \mathbb{R}^2 \to \mathbb{R}$ is defined as: $\Lambda(i,t) = \lambda_{\lfloor i \rfloor}(t)$

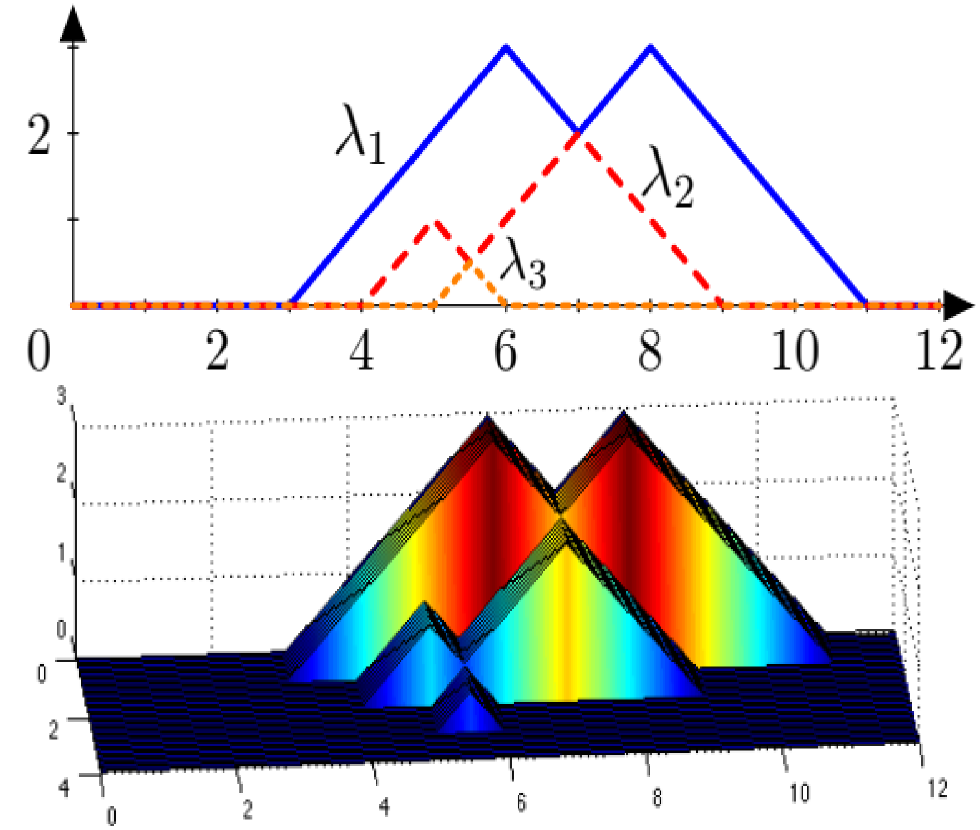
They can equivalently be defined as: $\Lambda(i,t) = i$-th $\max\{\lambda_j(t)\}$

Persistence landscape

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







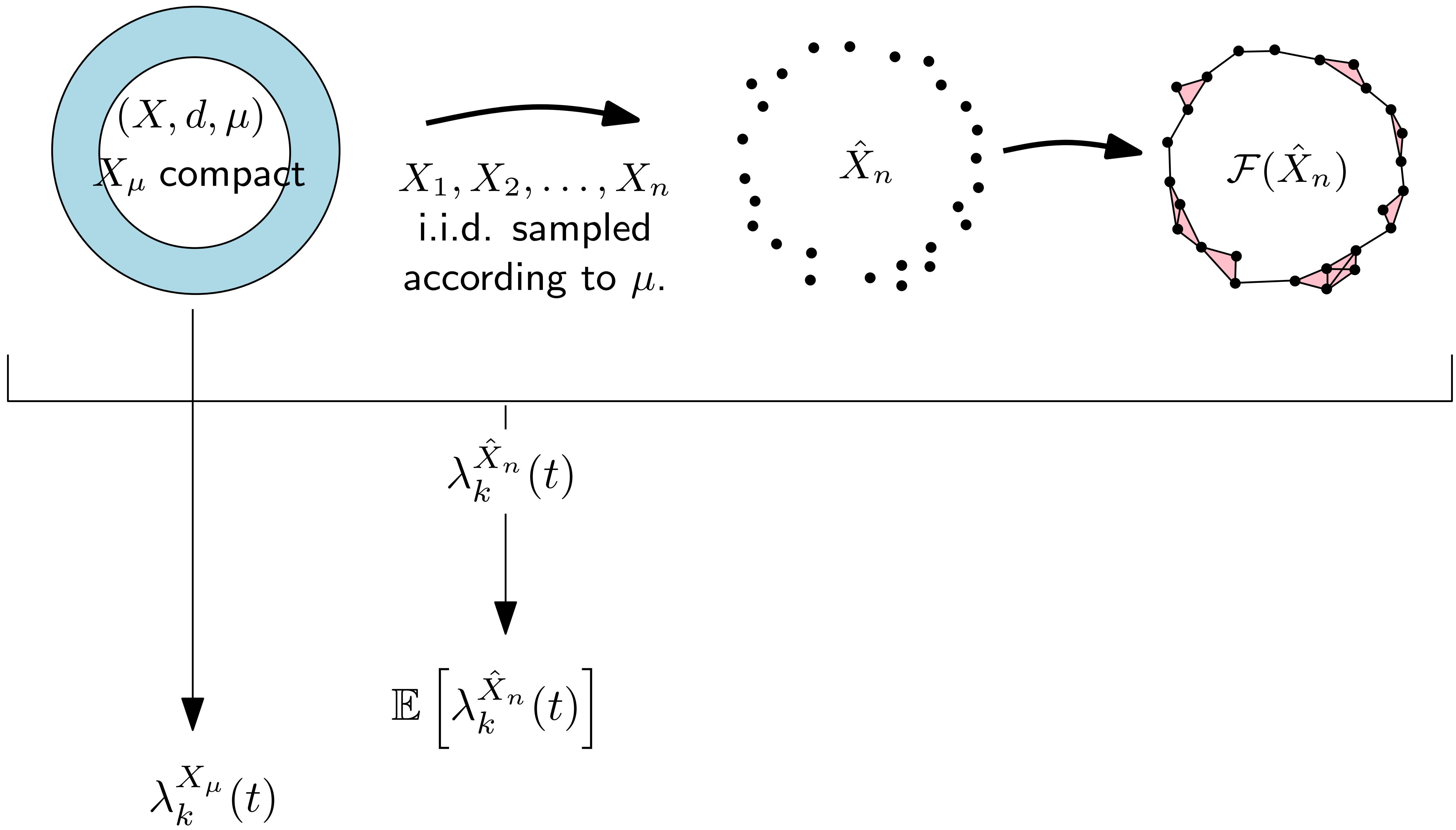
Rotate PD
Compute rank function

Use boundaries of rank function

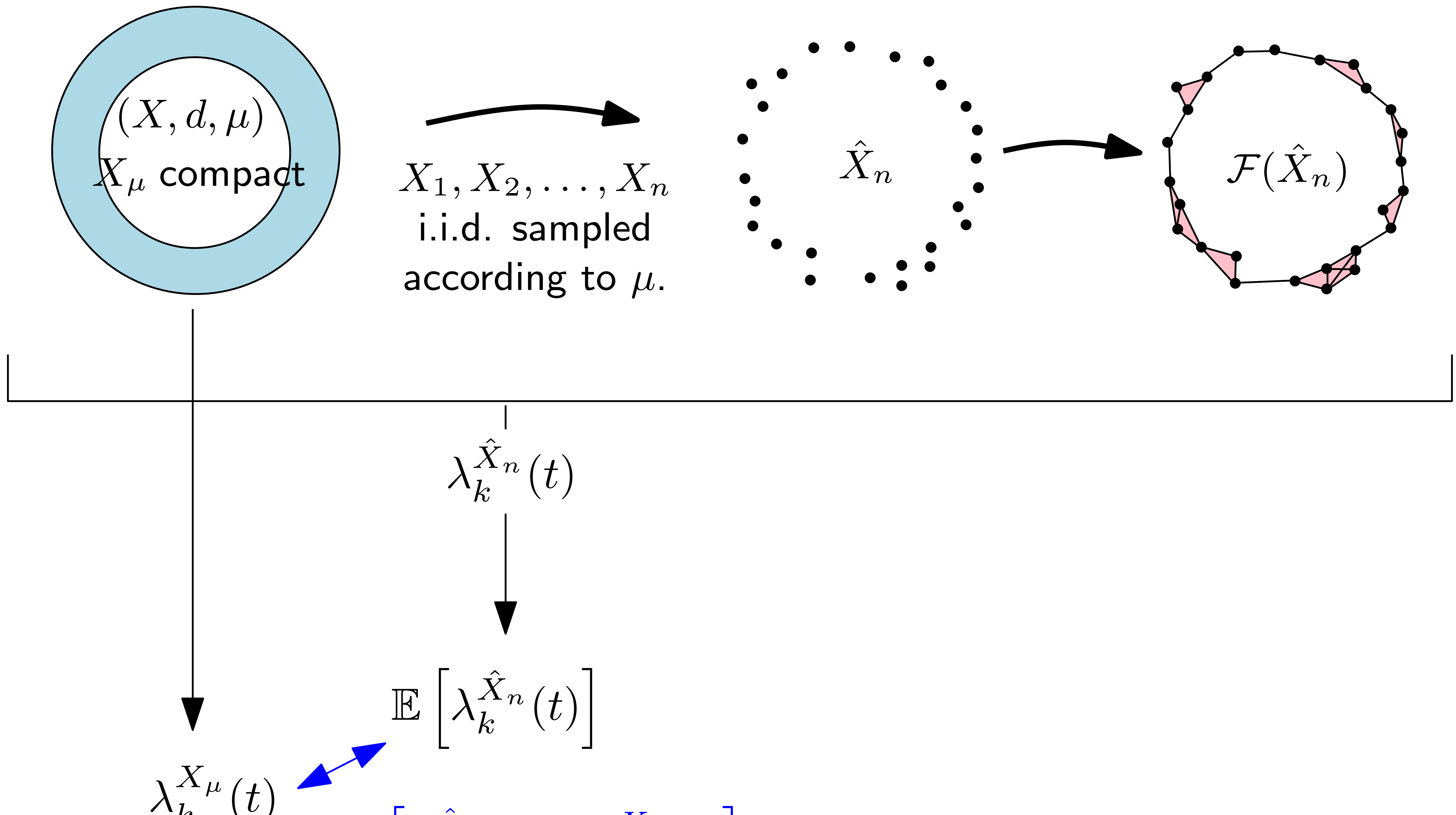
Prop: The following inequalities hold:

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

Statistics with landscapes

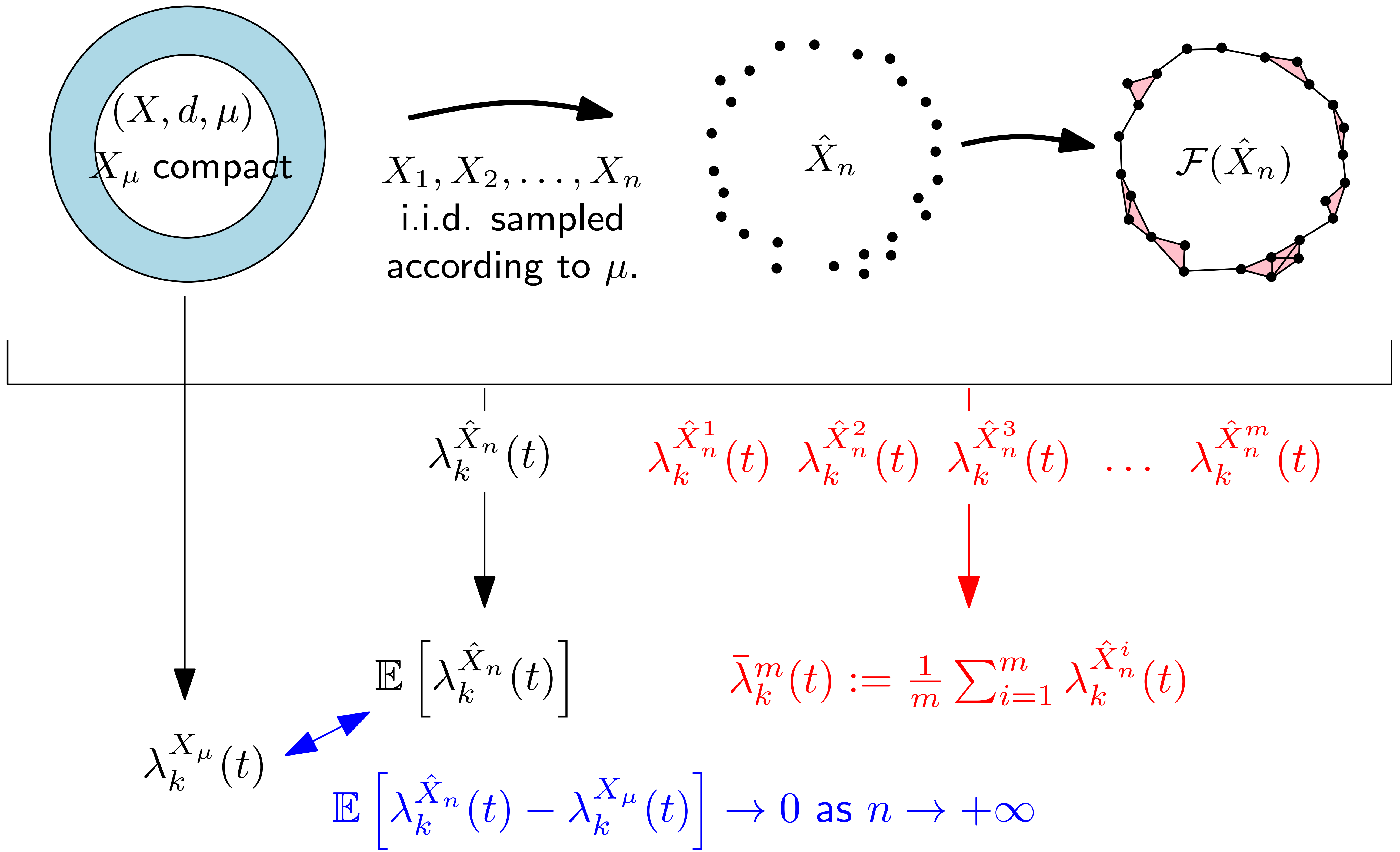


Statistics with landscapes

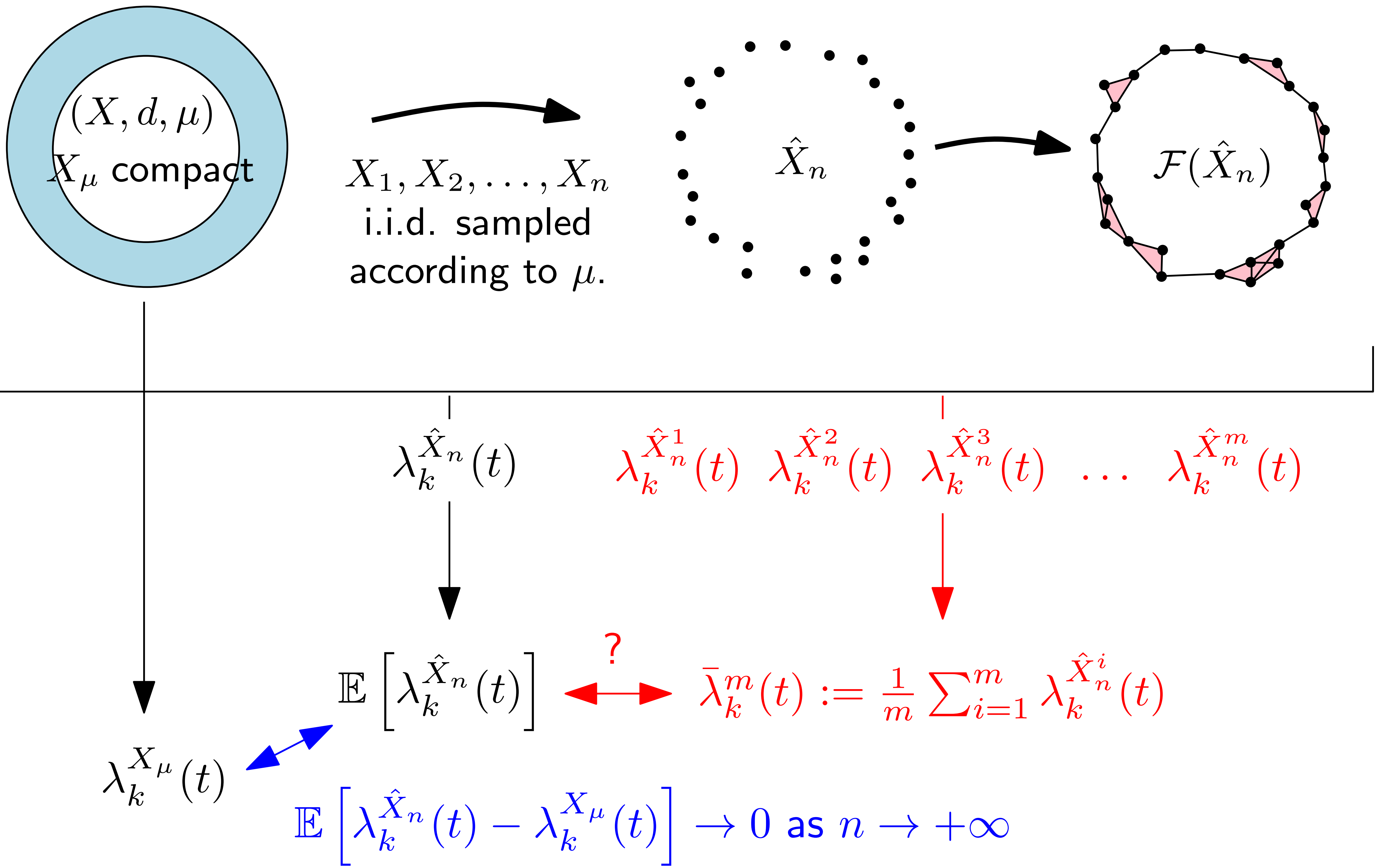


$\mathbb{E}\left[\lambda_k^{\hat{X}_n}(t) - \lambda_k^{X_\mu}(t)\right] \to 0$ as $n \to +\infty$

Statistics with landscapes



Statistics with landscapes



Bootstrapping landscapes

[*Stochastic convergence of persistence landscapes and silhouettes*, Chazal et al., JoCG, 2015]

Thm: Suppose that $\mathrm{var}(\bar{\lambda}_k^m(t)) > c > 0$ in an interval $[t_*, t^*] \subset [0,T]$, for some constant c. Then, given a confidence level $1-\alpha$, one has:

$$\mathbb{P}\left(\left|\mathbb{E}\left[\lambda_k^{\hat{X}_n}(t)\right] - \bar{\lambda}_k^m(t)\right| \leq \frac{Z_{B,\alpha}}{\sqrt{m}}\ \forall t \in [t_*, t^*]\right) \geq 1 - \alpha - O\left(\frac{(\log m)^{7/8}}{m^{1/8}}\right),$$

where $Z_{B,\alpha}$ is a quantile of a multiplier bootstrap distribution.

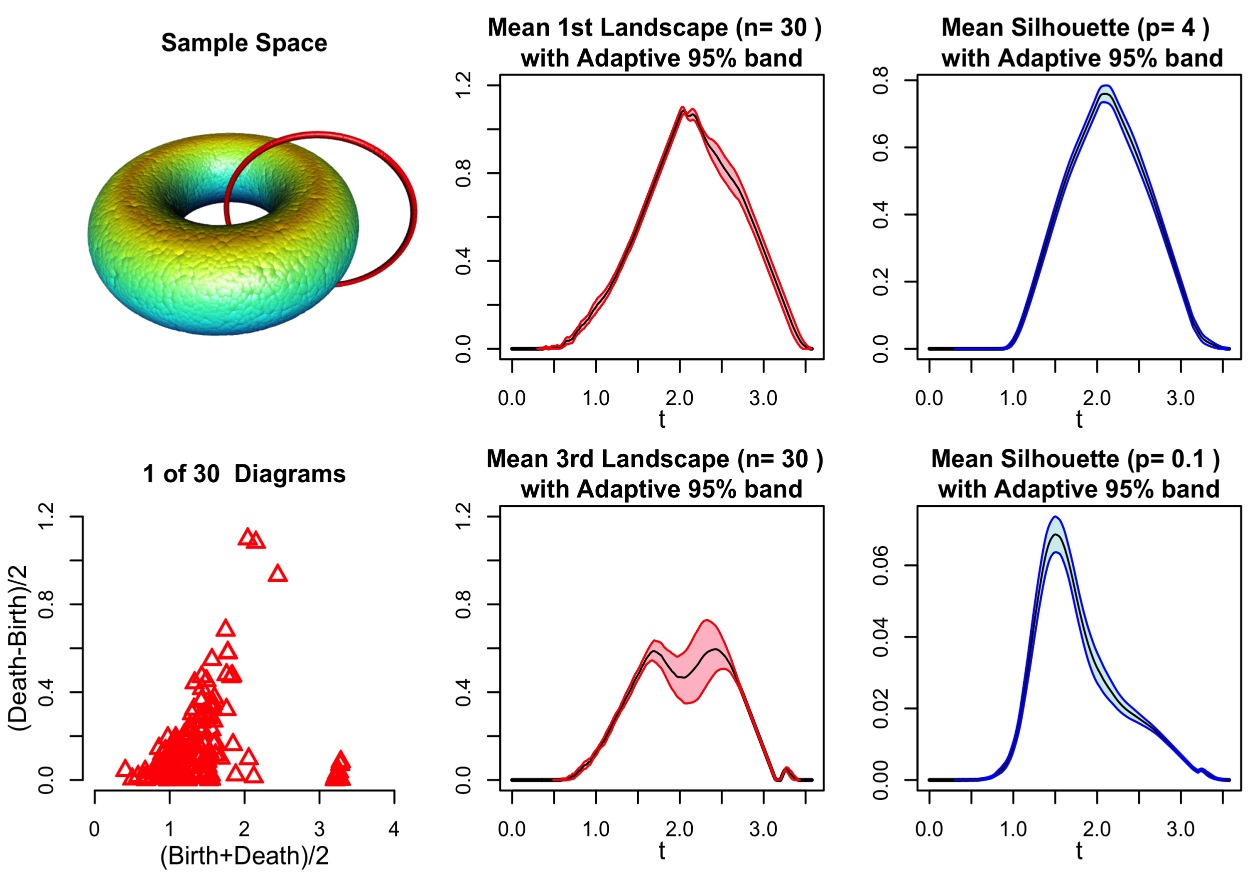
Bootstrapping landscapes

[*Stochastic convergence of persistence landscapes and silhouettes*, Chazal et al., JoCG, 2015]

Thm: Suppose that $\mathrm{var}(\bar{\lambda}_k^m(t)) > c > 0$ in an interval $[t_*, t^*] \subset [0, T]$, for some constant c. Then, given a confidence level $1-\alpha$, one has:

$$\mathbb{P}\left(\left|\mathbb{E}\left[\lambda_k^{\hat{X}_n}(t)\right] - \bar{\lambda}_k^m(t)\right| \leq \frac{Z_{B,\alpha}}{\sqrt{m}}\ \forall t \in [t_*, t^*]\right) \geq 1 - \alpha - O\left(\frac{(\log m)^{7/8}}{m^{1/8}}\right),$$

where $Z_{B,\alpha}$ is a quantile of a multiplier bootstrap distribution.



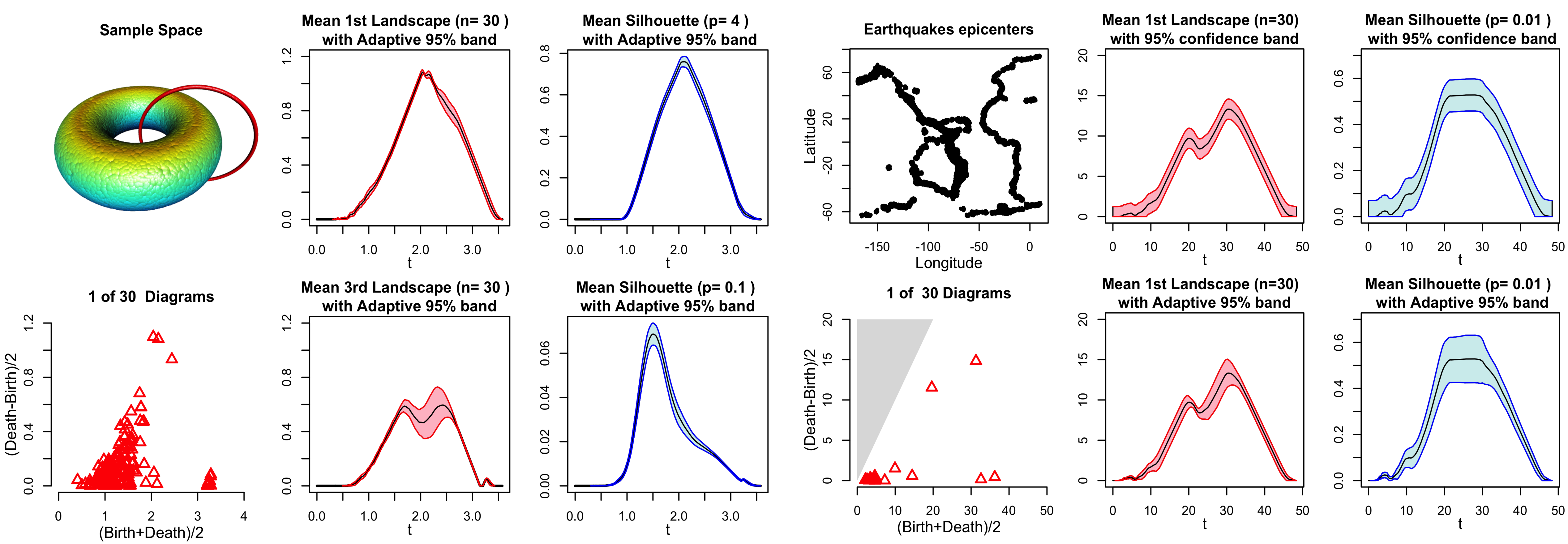
Bootstrapping landscapes

[*Stochastic convergence of persistence landscapes and silhouettes*, Chazal et al., JoCG, 2015]

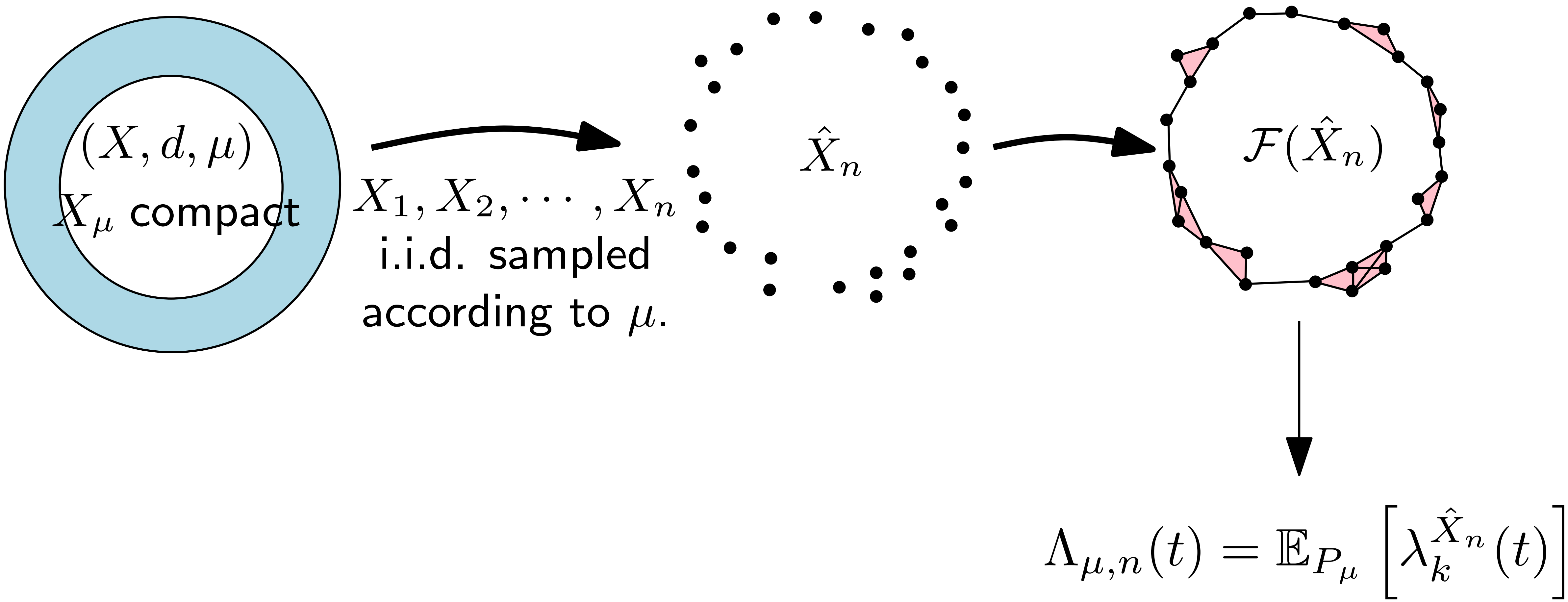
Thm: Suppose that $\mathrm{var}(\bar{\lambda}_k^m(t)) > c > 0$ in an interval $[t_*, t^*] \subset [0, T]$, for some constant c. Then, given a confidence level $1-\alpha$, one has:

$$\mathbb{P}\left(\left|\mathbb{E}\left[\lambda_k^{\hat{X}_n}(t)\right] - \bar{\lambda}_k^m(t)\right| \leq \frac{Z_{B,\alpha}}{\sqrt{m}}\ \forall t \in [t_*, t^*]\right) \geq 1 - \alpha - O\left(\frac{(\log m)^{7/8}}{m^{1/8}}\right),$$

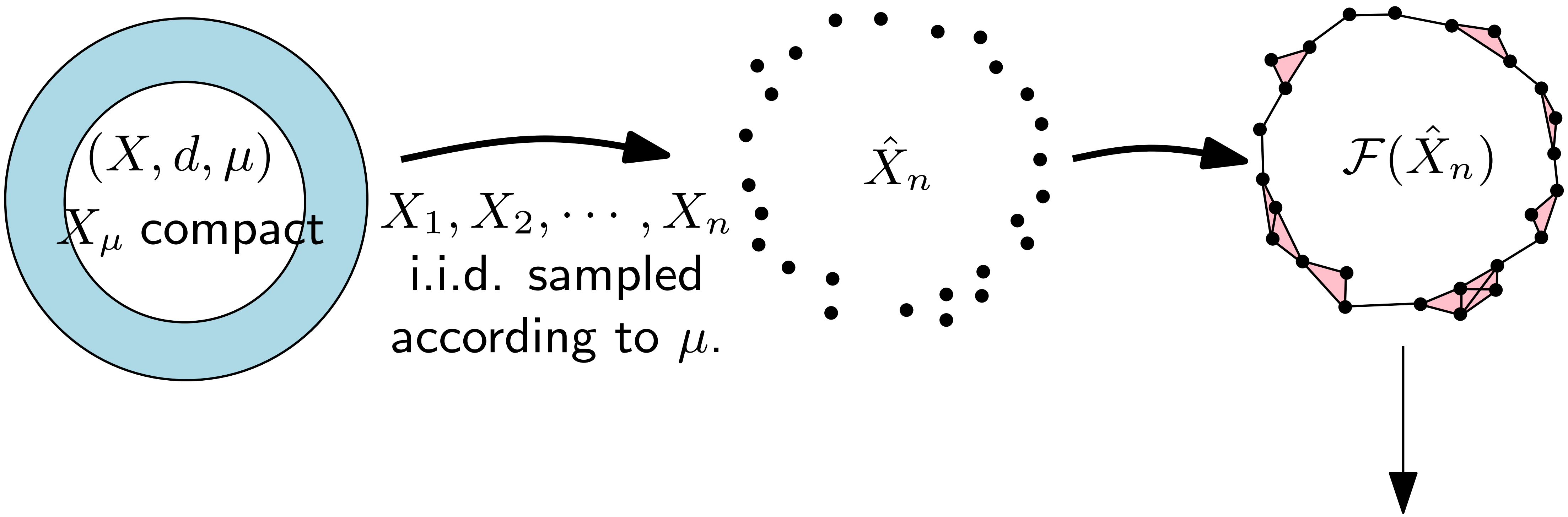
where $Z_{B,\alpha}$ is a quantile of a multiplier bootstrap distribution.



Stability of the mean landscape



Stability of the mean landscape

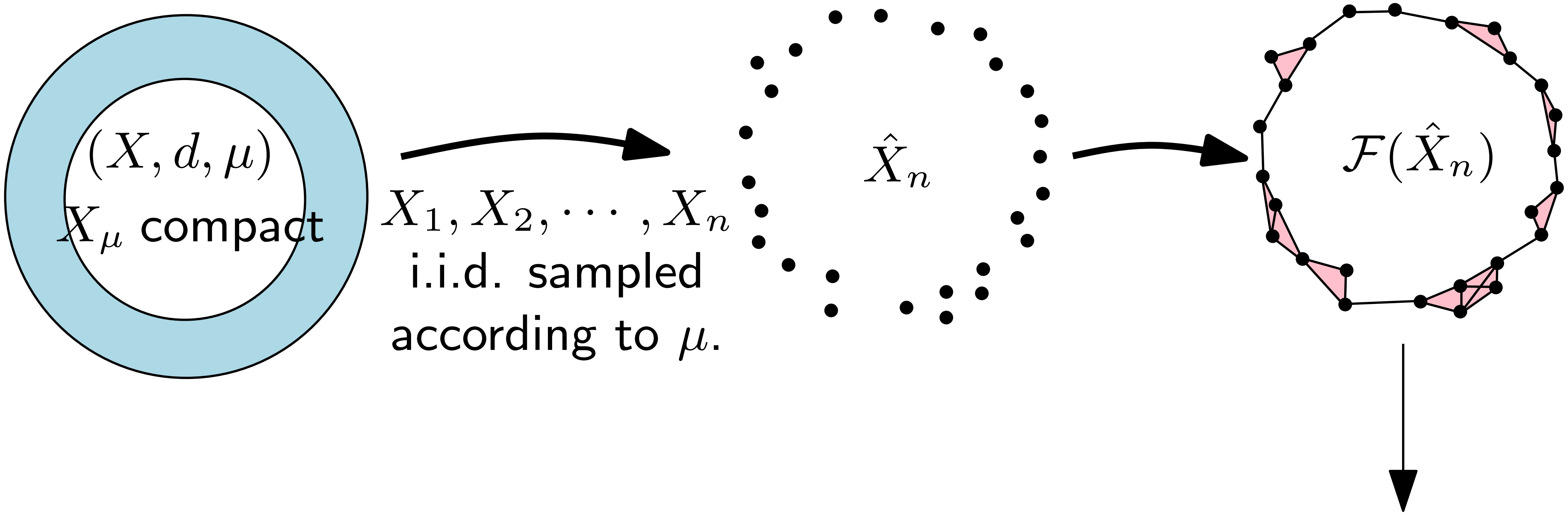


$$\Lambda_{\mu,n}(t) = \mathbb{E}_{P_\mu}\left[\lambda_k^{\hat{X}_n}(t)\right]$$

How bad is the dependence in μ?

Stability of the mean landscape

[*Subsampling methods for persistent homology*, Chazal et al., ICML, 2015]



$$\Lambda_{\mu,n}(t) = \mathbb{E}_{P_\mu}\left[\lambda_k^{\hat{X}_n}(t)\right]$$

How bad is the dependence in μ?

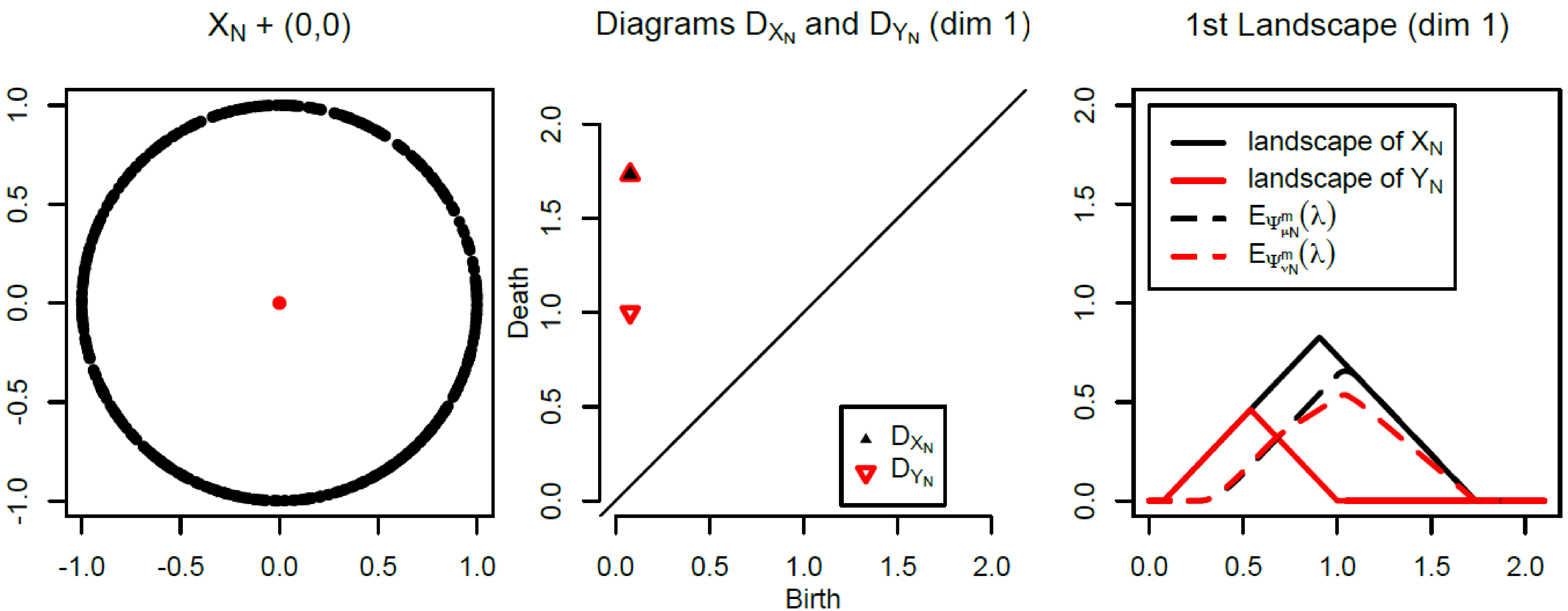
Thm: Let (X, d) be a metric space and let μ, ν be probability measures on X with compact supports. Then one has:

$$\|\Lambda_{\mu,n} - \Lambda_{\nu,n}\|_\infty \leq n^{\frac{1}{p}} W_p(\mu, \nu),$$

where W_p denotes the Wasserstein distance with cost function $d(\cdot, \cdot)^p$.

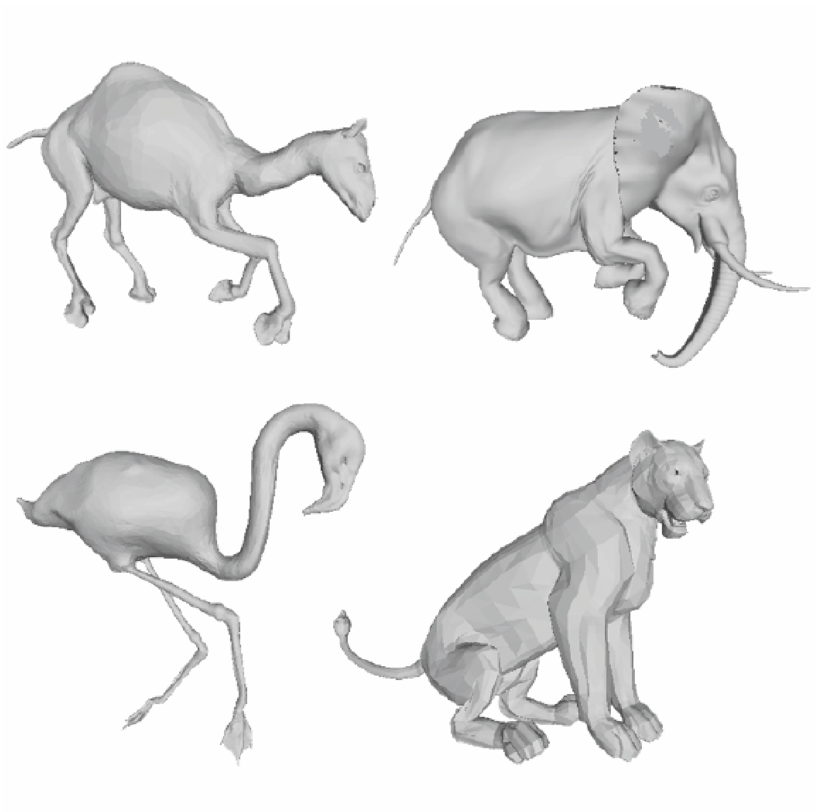
Numerical illustrations: confidence for landscapes

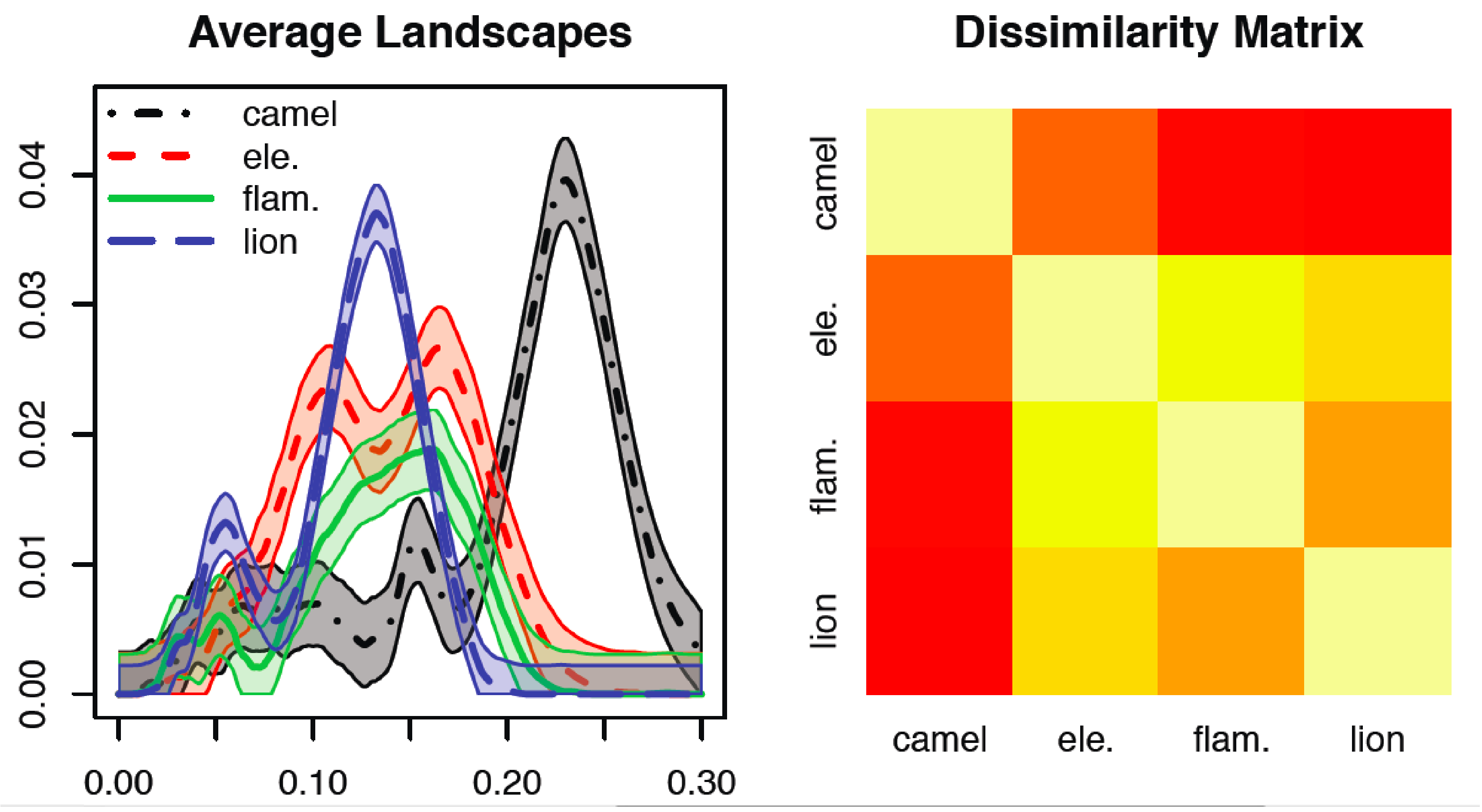
Example: Circle with one outlier.



Numerical illustrations: confidence for landscapes

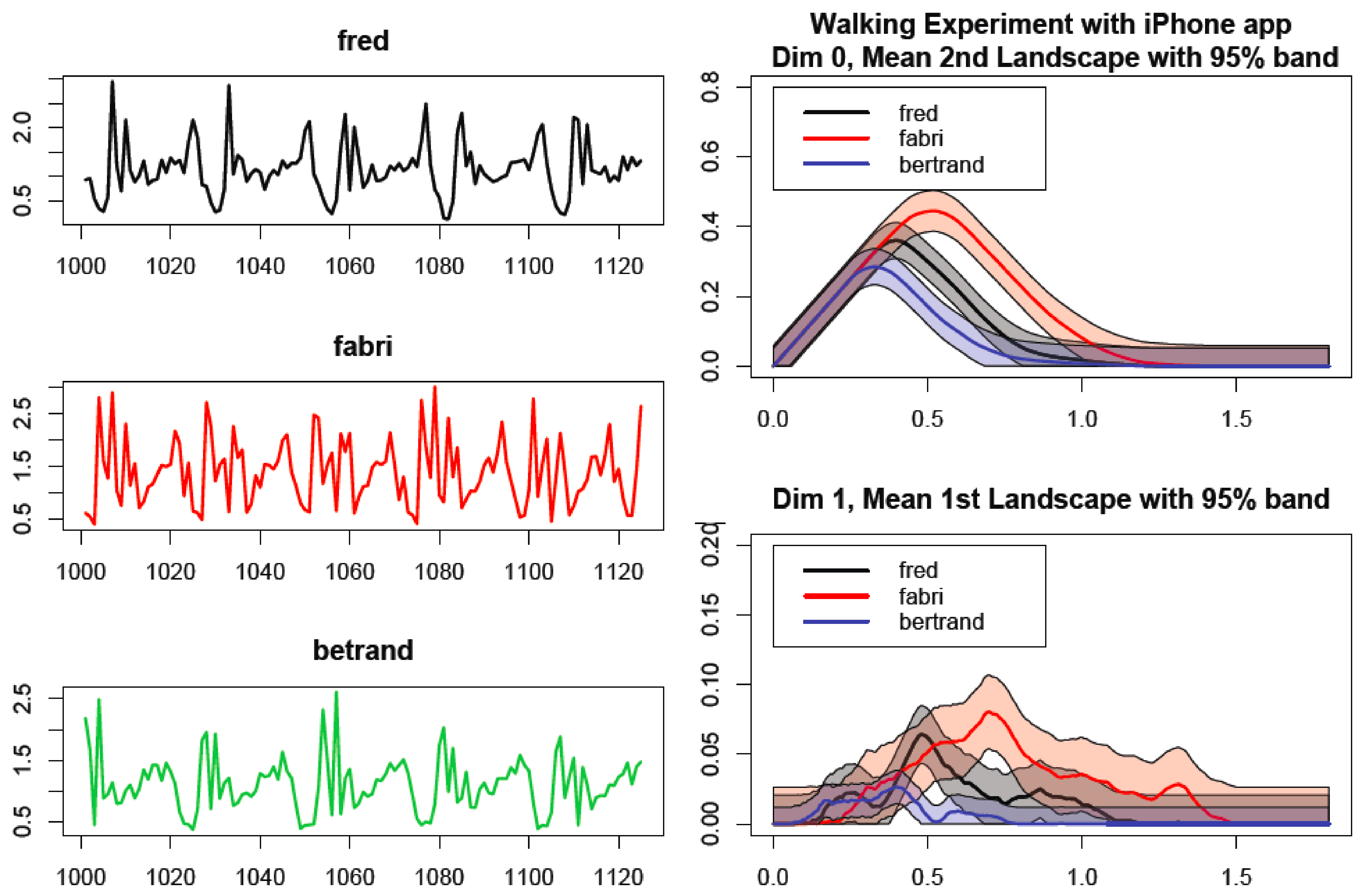
Example: 3D shapes





Numerical illustrations: confidence for landscapes

Example: Accelerometer data from smartphone.



Topological Machine Learning (I): Statistics and Representations

1. Topological Inference
2. Persistence Representations
3. **Learning Representations**

The space of persistence diagrams

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

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

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

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

The space of persistence diagrams

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

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

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

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

Prop: $\mathcal{H}$ 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.

(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 $L, N \to +\infty$

The space of persistence diagrams

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

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

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

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

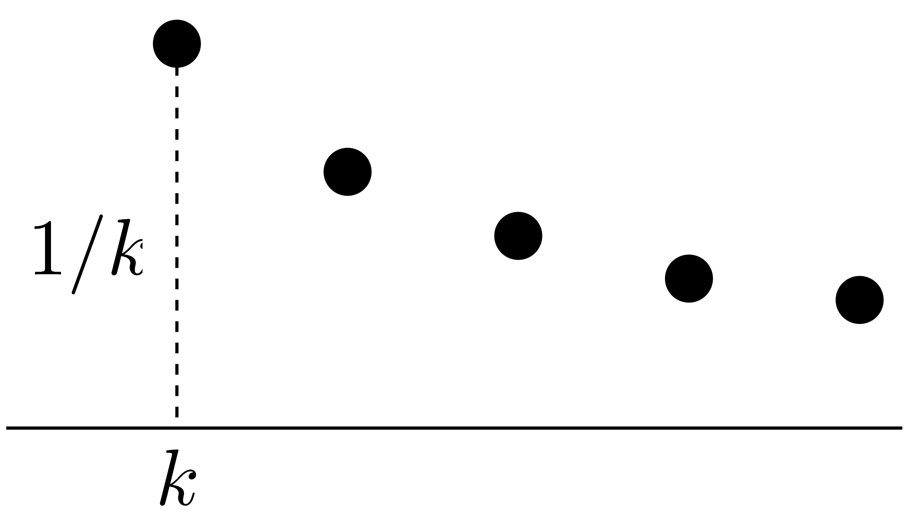
Proof:

(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, C., SoCG, 2019]

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

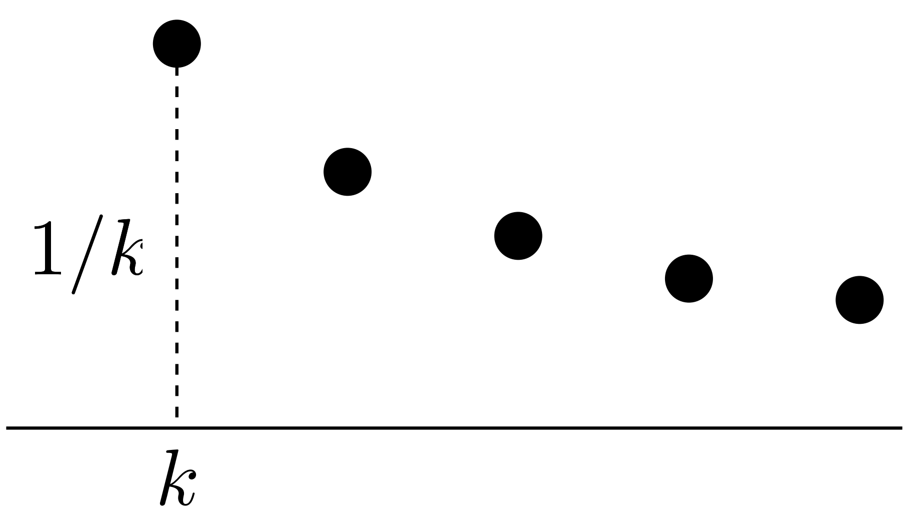
Def: Two metrics d, d' are *equivalent* if

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

Proof:

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

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



The space of persistence diagrams

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

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

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

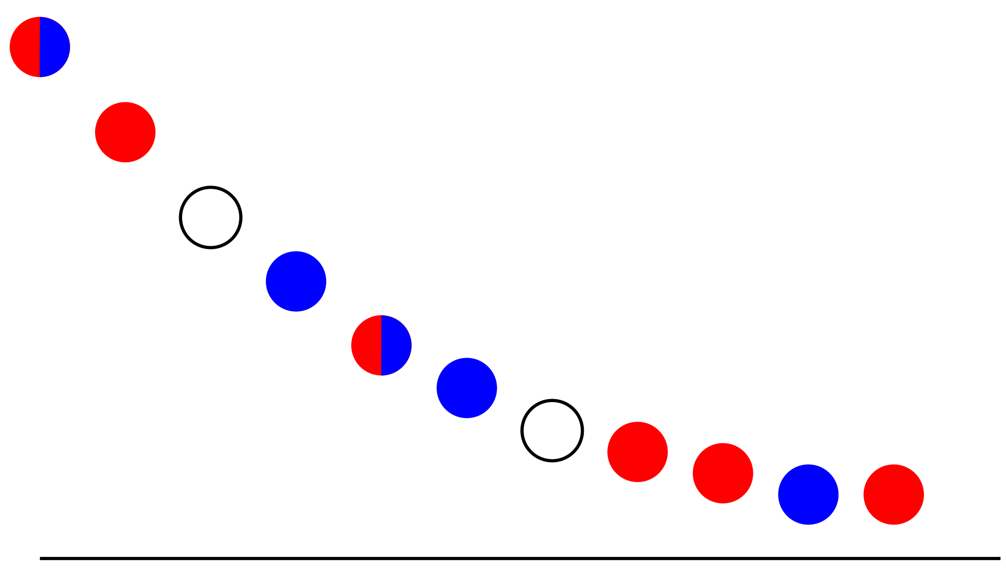
$$\exists 0 < A, B < +\infty \text{ s.t. } A\,\mathrm{d}(\cdot,\cdot) \leq \mathrm{d}'(\cdot,\cdot) \leq B\,\mathrm{d}(\cdot,\cdot)$$

Proof:

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

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

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



The space of persistence diagrams

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

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

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

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

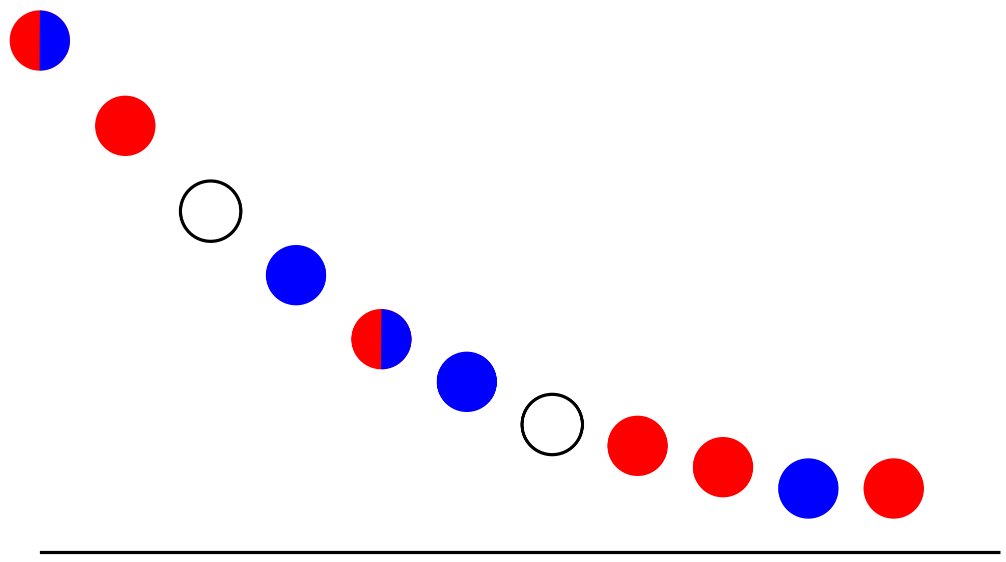
Proof:

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

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



The space of persistence diagrams

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

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

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

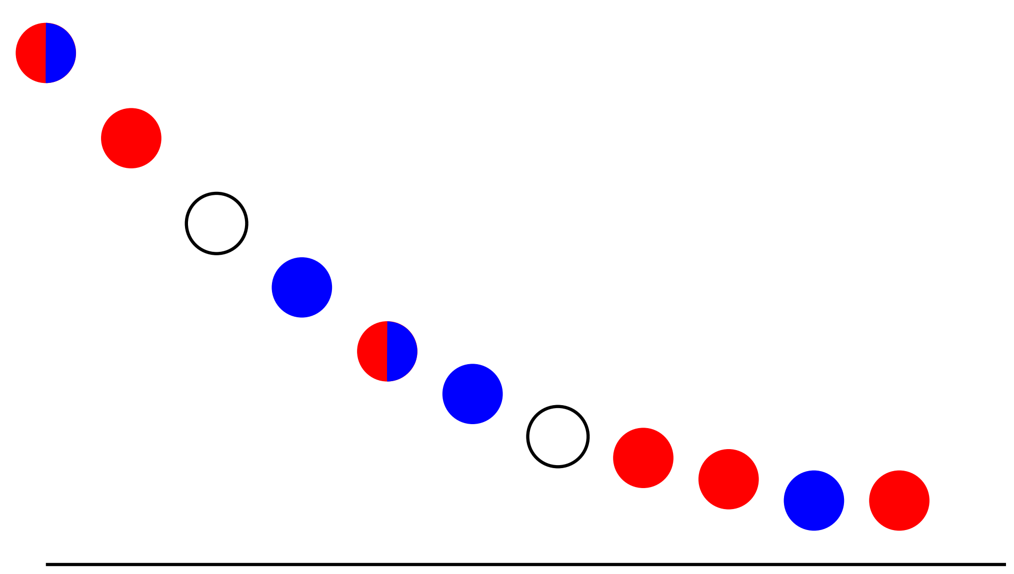
$$\exists 0 < A, B < +\infty \text{ s.t. } A\,\mathrm{d}(\cdot,\cdot) \leq \mathrm{d}'(\cdot,\cdot) \leq B\,\mathrm{d}(\cdot,\cdot)$$

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$(S') \geq$ card$(S/\sim)$

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

The space of persistence diagrams

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

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

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

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

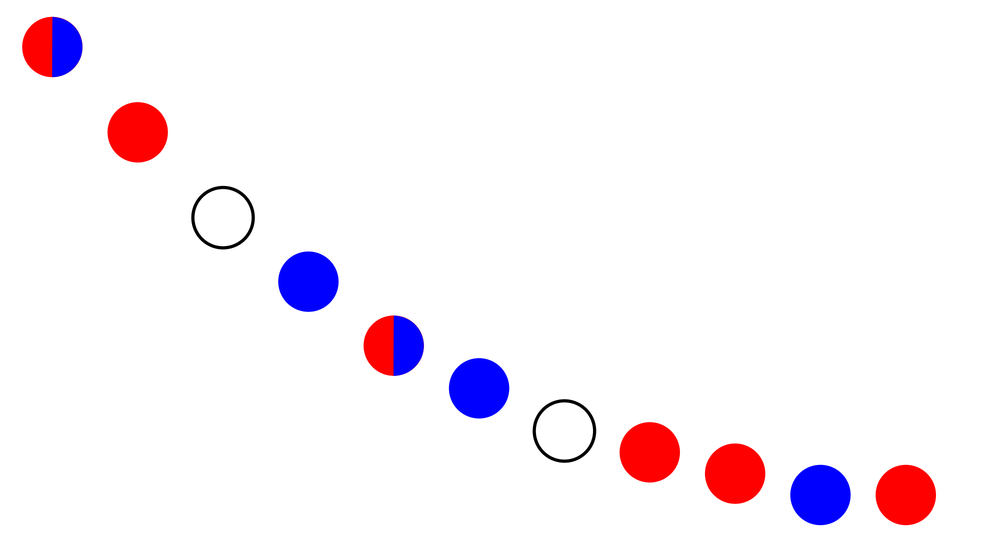
Proof:

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

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



$\Rightarrow \text{card}(S') \geq \boxed{\text{card}(S/\sim)}$ 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, C., SoCG, 2019]

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

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

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

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

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

The space of persistence diagrams

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

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

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

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

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

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, C., SoCG, 2019]

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

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

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

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:

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

The space of persistence diagrams

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

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

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

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

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

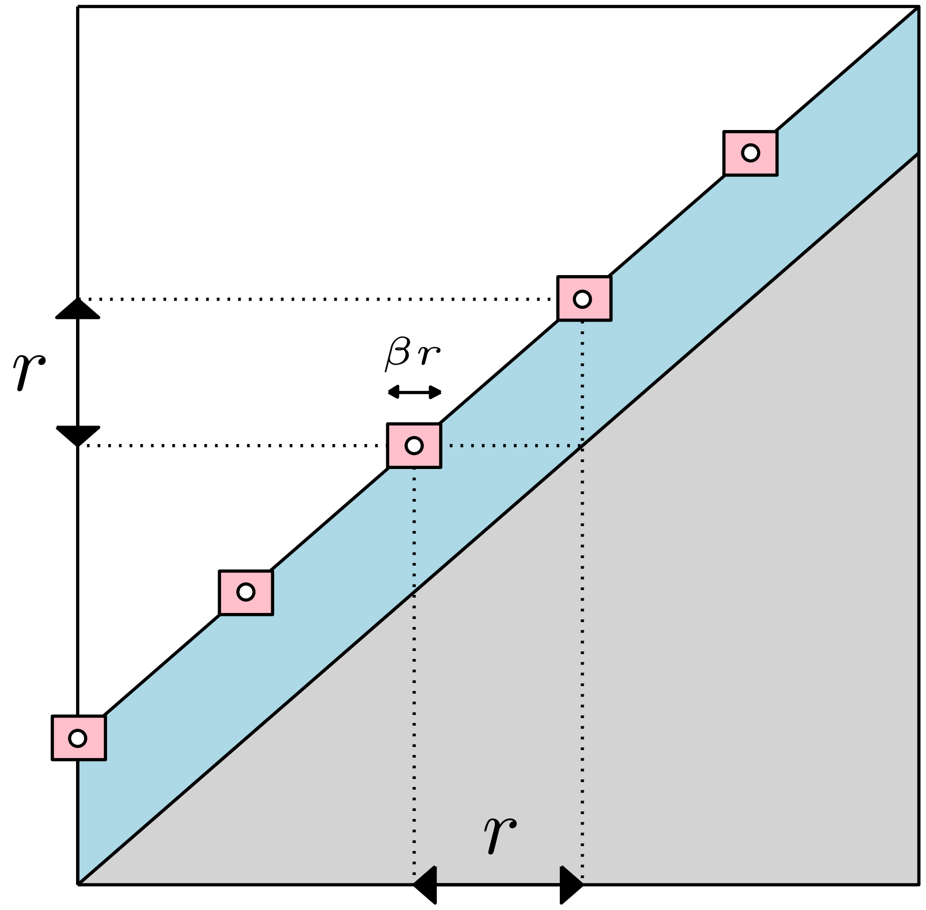
[*On the Metric Distortion of Embedding Persistence Diagrams into separable Hilbert Spaces*, Bauer, C., SoCG, 2019]

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

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

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

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

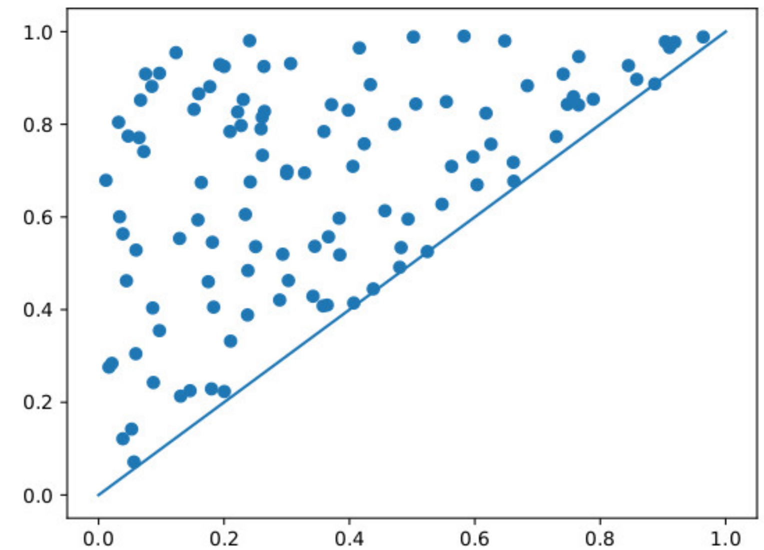
$\dim_A(\mathcal{D}, d_p) = +\infty$ whereas $\dim_A(\mathbb{R}^d) = d$

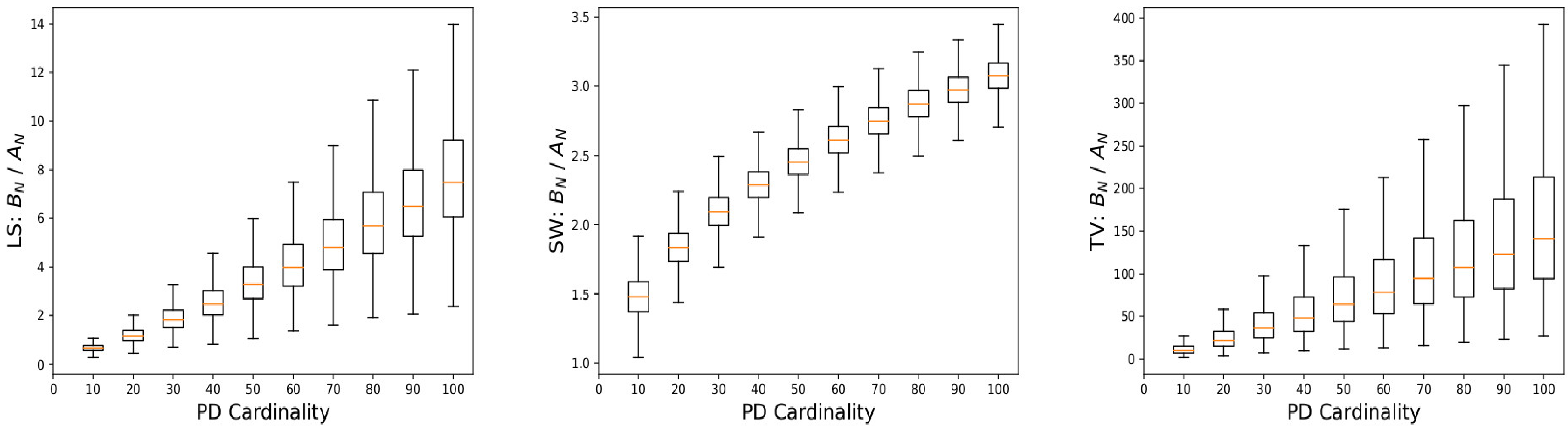
The space of persistence diagrams

[*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



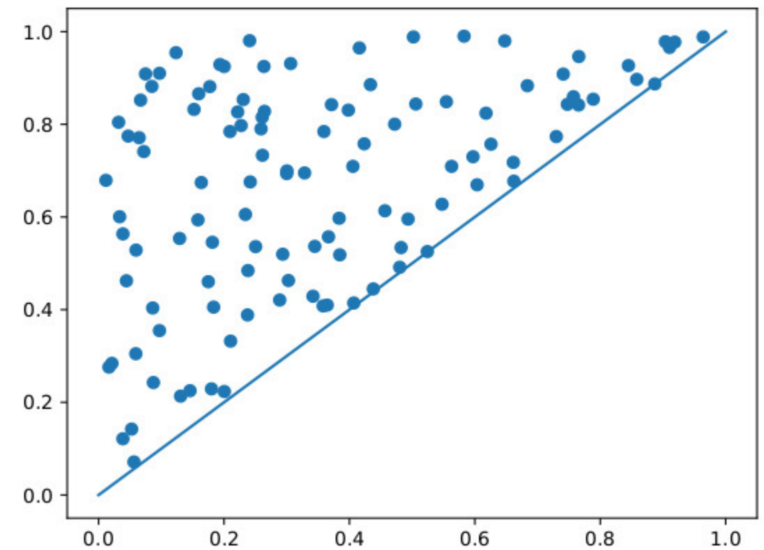


The space of persistence diagrams

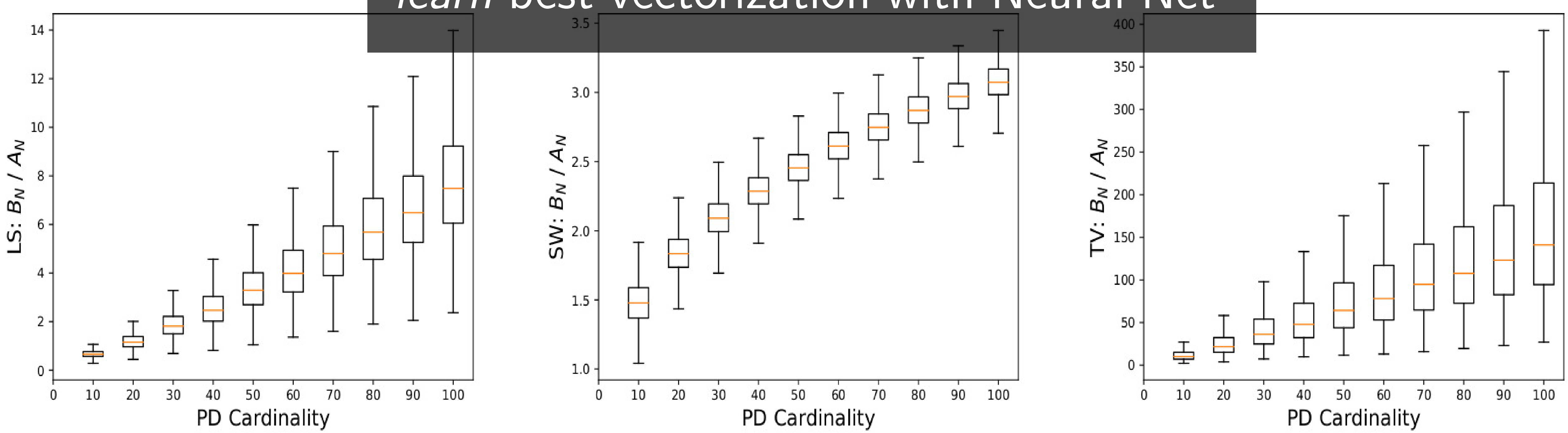
[*On the Metric Distortion of Embedding Persistence Diagrams into separable Hilbert Spaces*, Bauer, Carrière, SoCG, 2019]

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 $x \in \mathbb{R}^d$

The Deep Set architecture

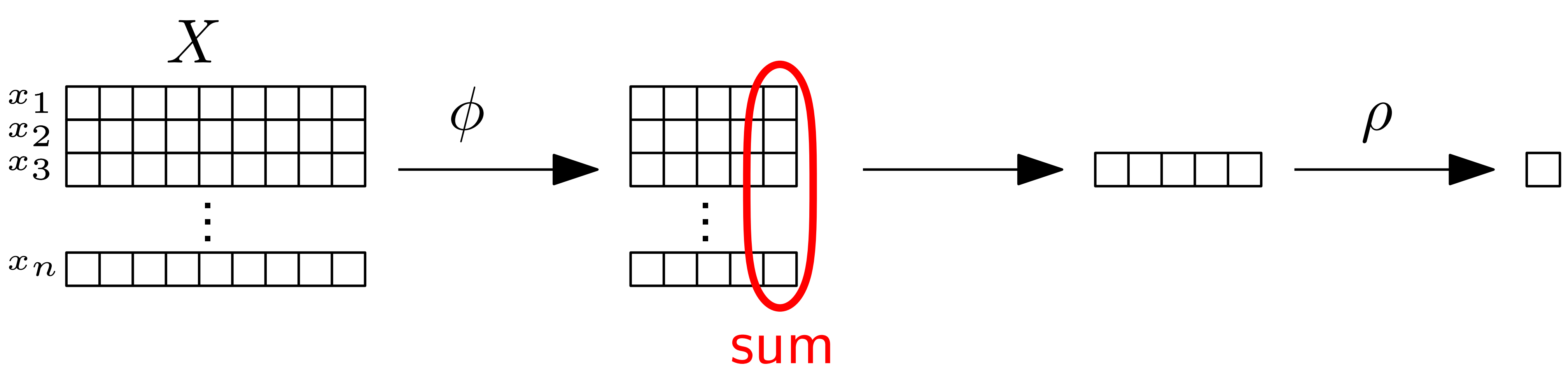
[*Deep Sets*, Zaheer, Kottur, Ravanbakhsh, Poczos, Salakhutdinov, Smola, NeurIPS, 2017]

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

Permutation invariant layers generalize several TDA approaches

Application to PDs

Permutation invariant layers generalize several TDA approaches

$\longrightarrow$ persistence images

Application to PDs

Permutation invariant layers generalize several TDA approaches

$\longrightarrow$ persistence images $\quad\longrightarrow$ landscapes

Application to PDs

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]

Application to PDs

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

Application to PDs

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

Application to PDs

[*PersLay: A Neural Network Layer for Persistence Diagrams and New Graph Topological Signatures*, C., 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

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

Application to PDs

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

Permutation invariant layers generalize several TDA approaches

$\rightarrow$ persistence images $\quad\rightarrow$ landscapes $\quad\rightarrow$ 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

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

Permutation-invariant operation

Application to PDs

[*PersLay: A Neural Network Layer for Persistence Diagrams and New Graph Topological Signatures*, C., 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

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

Permutation-invariant operation

Weight function

Application to PDs

[*PersLay: A Neural Network Layer for Persistence Diagrams and New Graph Topological Signatures*, C., 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

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

Permutation-invariant operation

Weight function

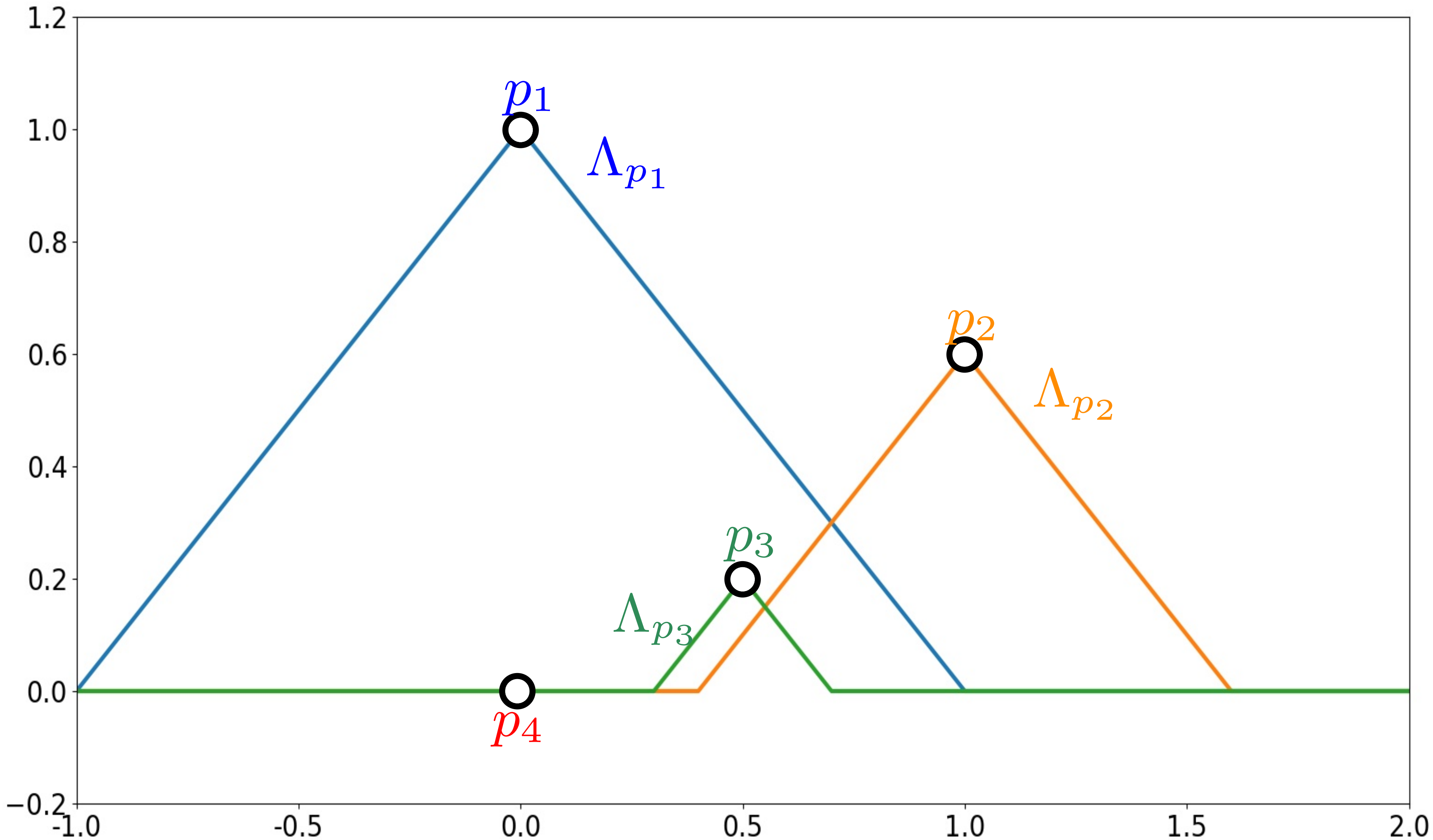
Point transformation

Application to PDs

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

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

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



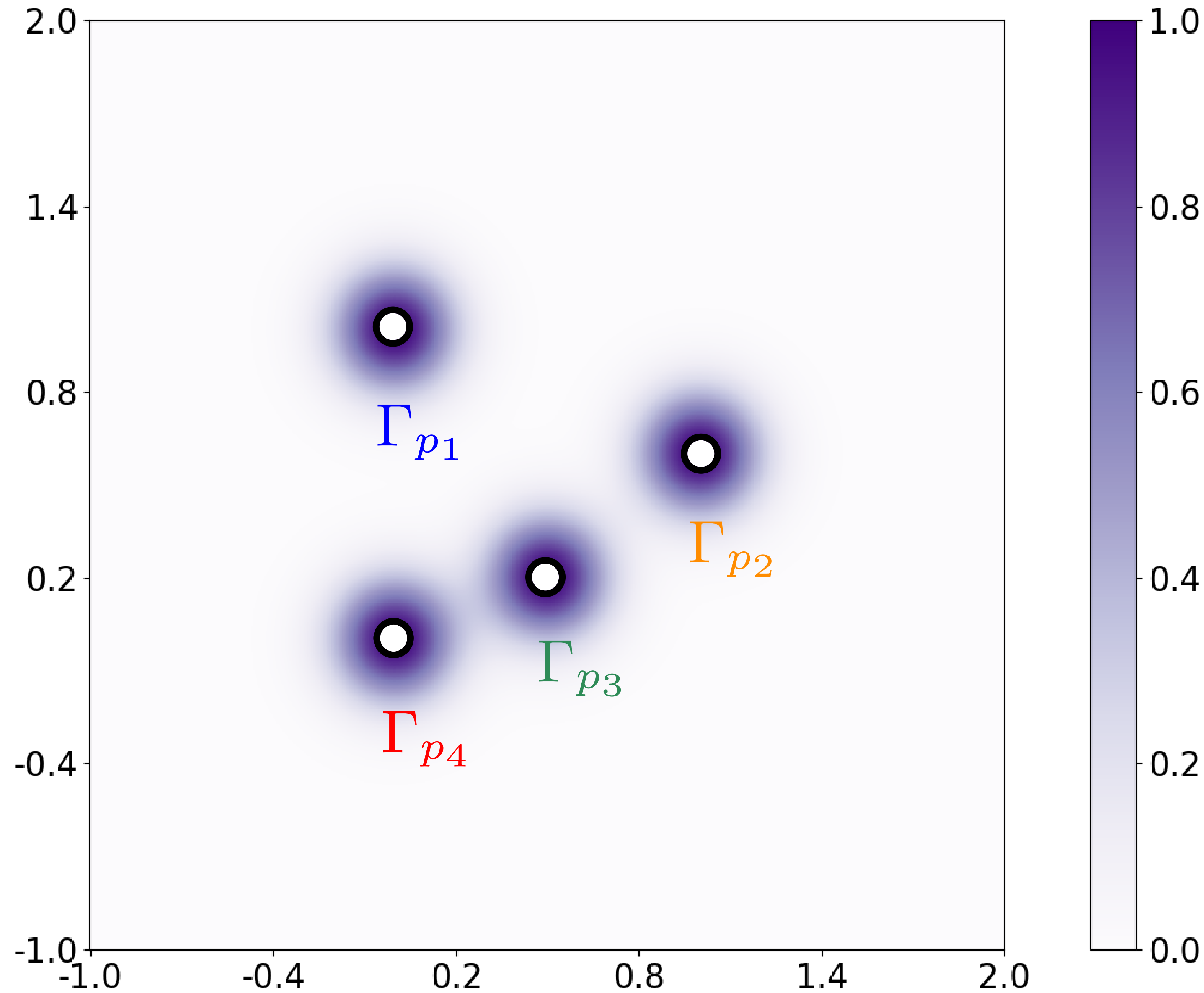
Application to PDs

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

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

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

op = sum



Application to PDs

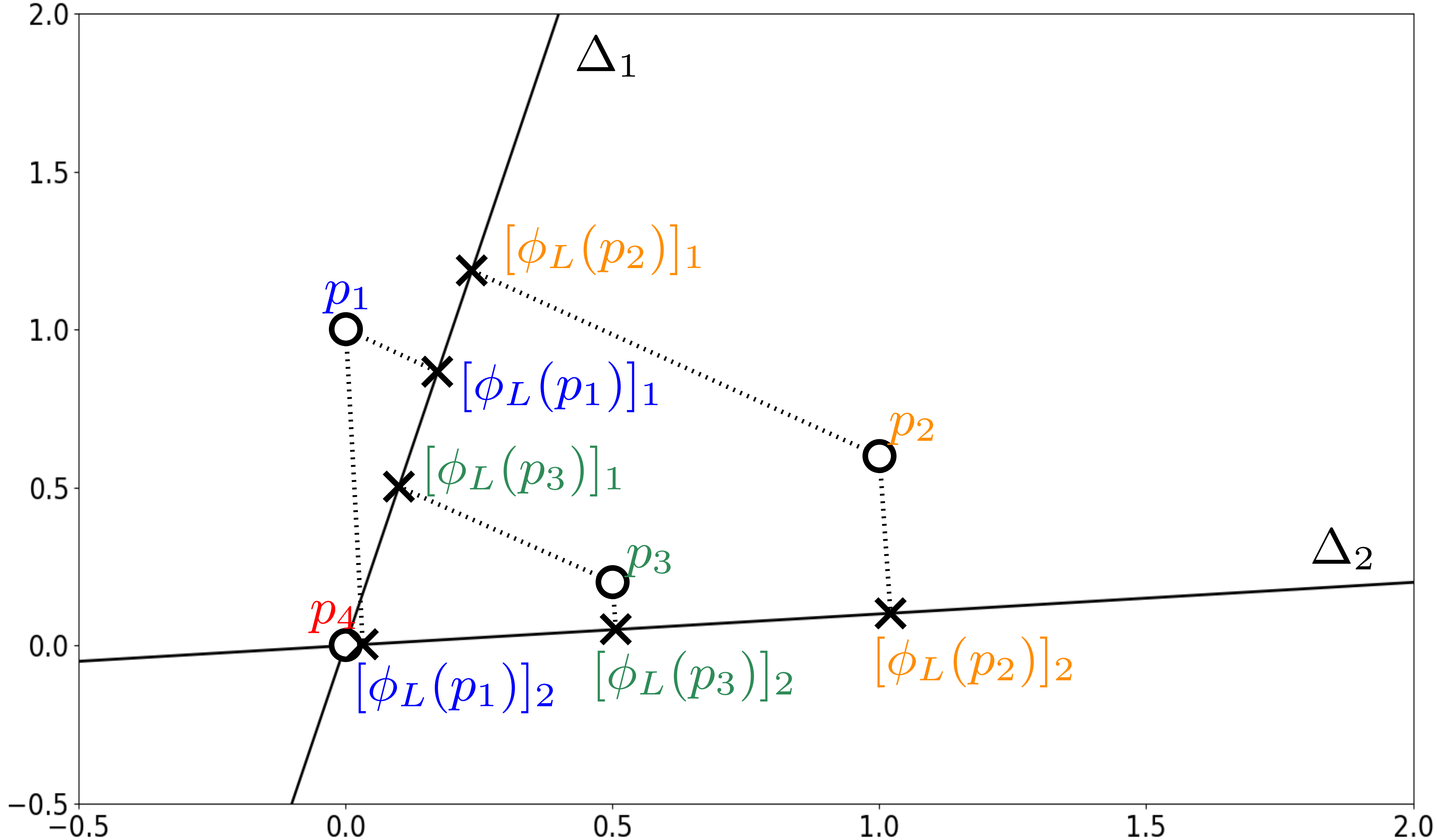
[*PersLay: A Neural Network Layer for Persistence Diagrams and New Graph Topological Signatures*, C., Chazal, Ike, Lacombe, Royer, Umeda, AISTATS, 2019]

Parameters $\Delta_1, \cdots, \Delta_q \in [-\frac{\pi}{2}, \frac{\pi}{2}]$
$b_{\Delta_1}, \cdots, b_{\Delta_q} \in \mathbb{R}$

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

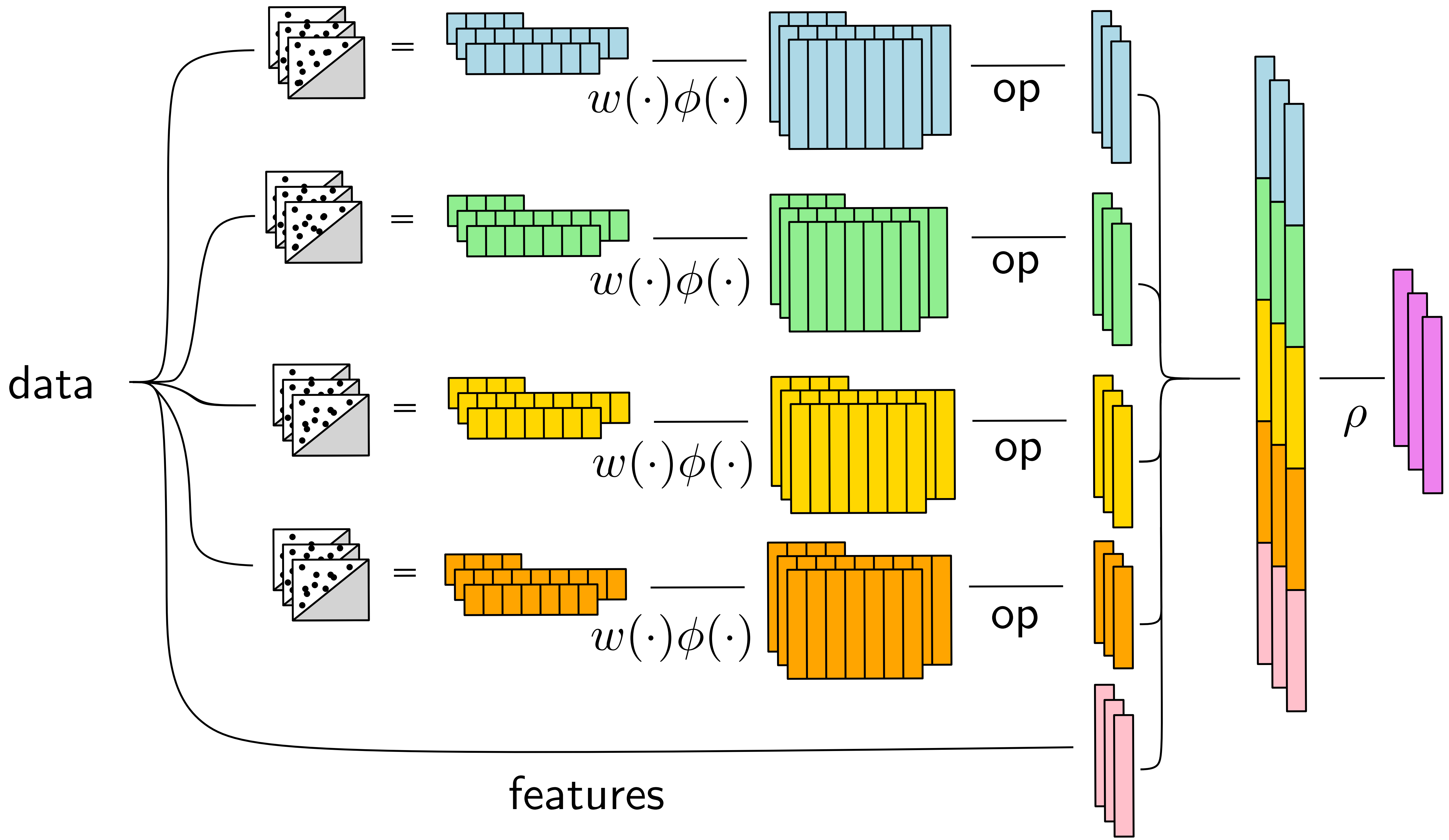
$w(p) = 1$

op $=$ top-k



Application to PDs

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



Application to graph classification

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

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

D its degree matrix

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

Application to graph classification

[*PersLay: A Neural Network Layer for Persistence Diagrams and New Graph Topological Signatures*, C., 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 $0 \leq \lambda_1 \leq \dots \leq \lambda_n \leq 2$

Application to graph classification

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

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

D its degree matrix

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

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

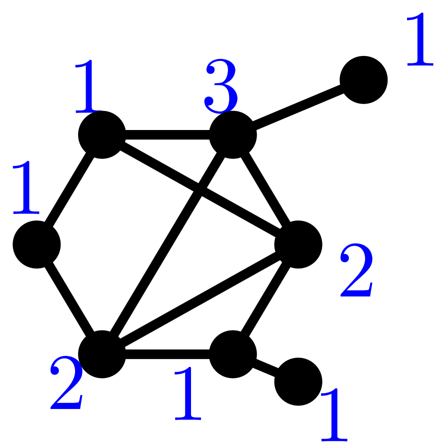
with eigenvalues $0 \leq \lambda_1 \leq \cdots \leq \lambda_n \leq 2$

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*, C., Chazal, Ike, Lacombe, Royer, Umeda, AISTATS, 2019]

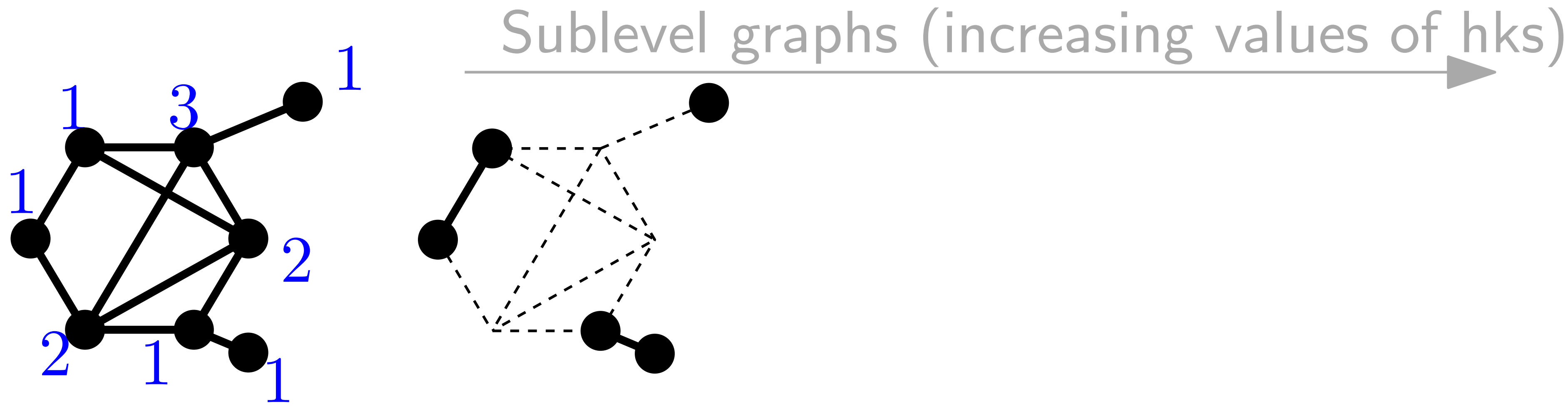


Def: Let $t \geq 0$, and define the *Heat Kernel Signature* of param t:

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

Application to graph classification

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

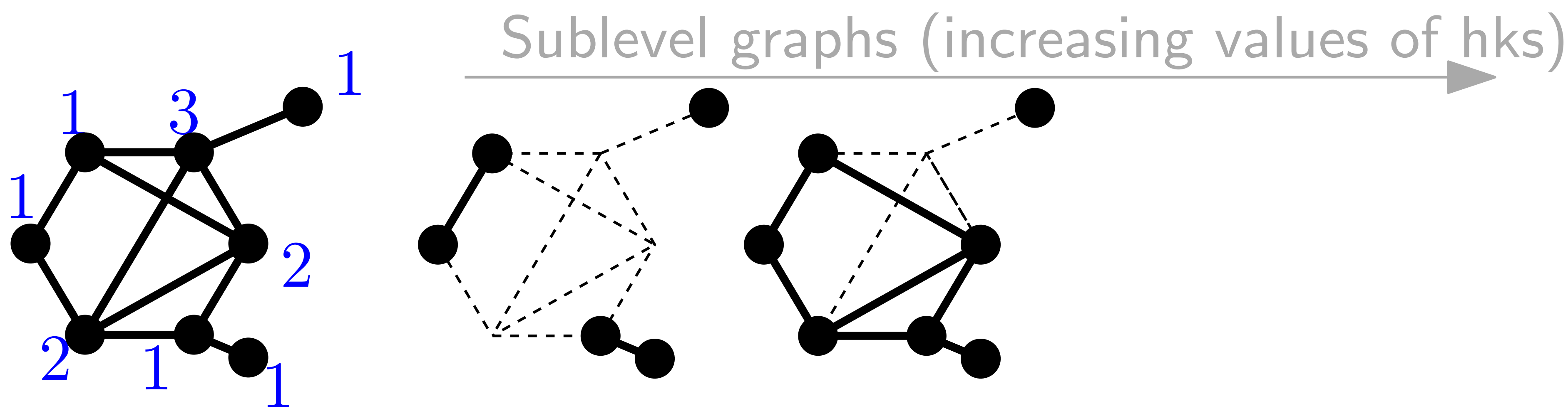


Def: Let $t \geq 0$, and define the *Heat Kernel Signature* of param t:

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

Application to graph classification

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

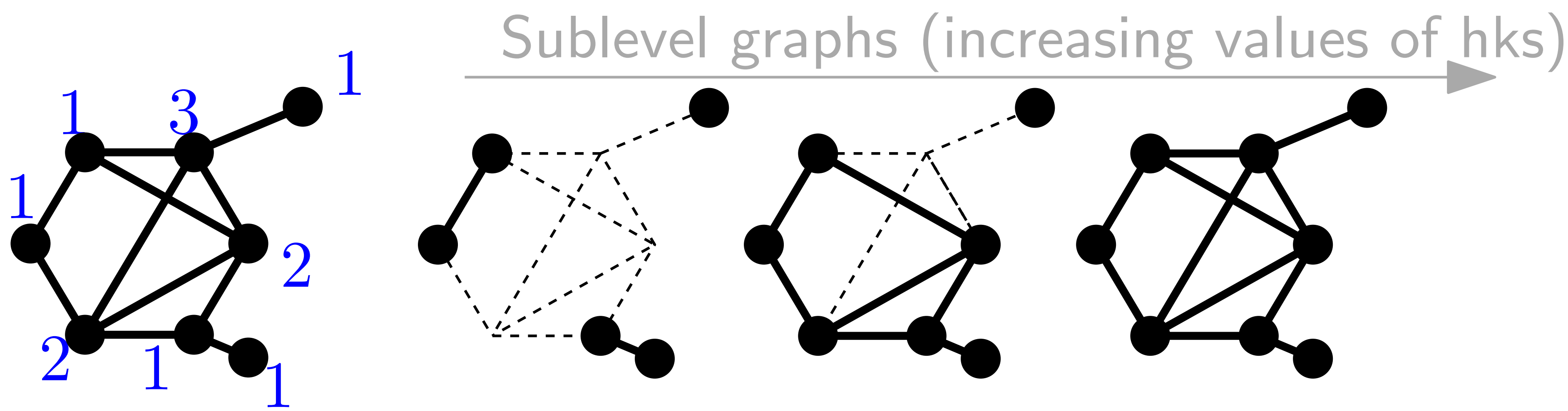


Def: Let $t \geq 0$, and define the *Heat Kernel Signature* of param t:

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

Application to graph classification

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

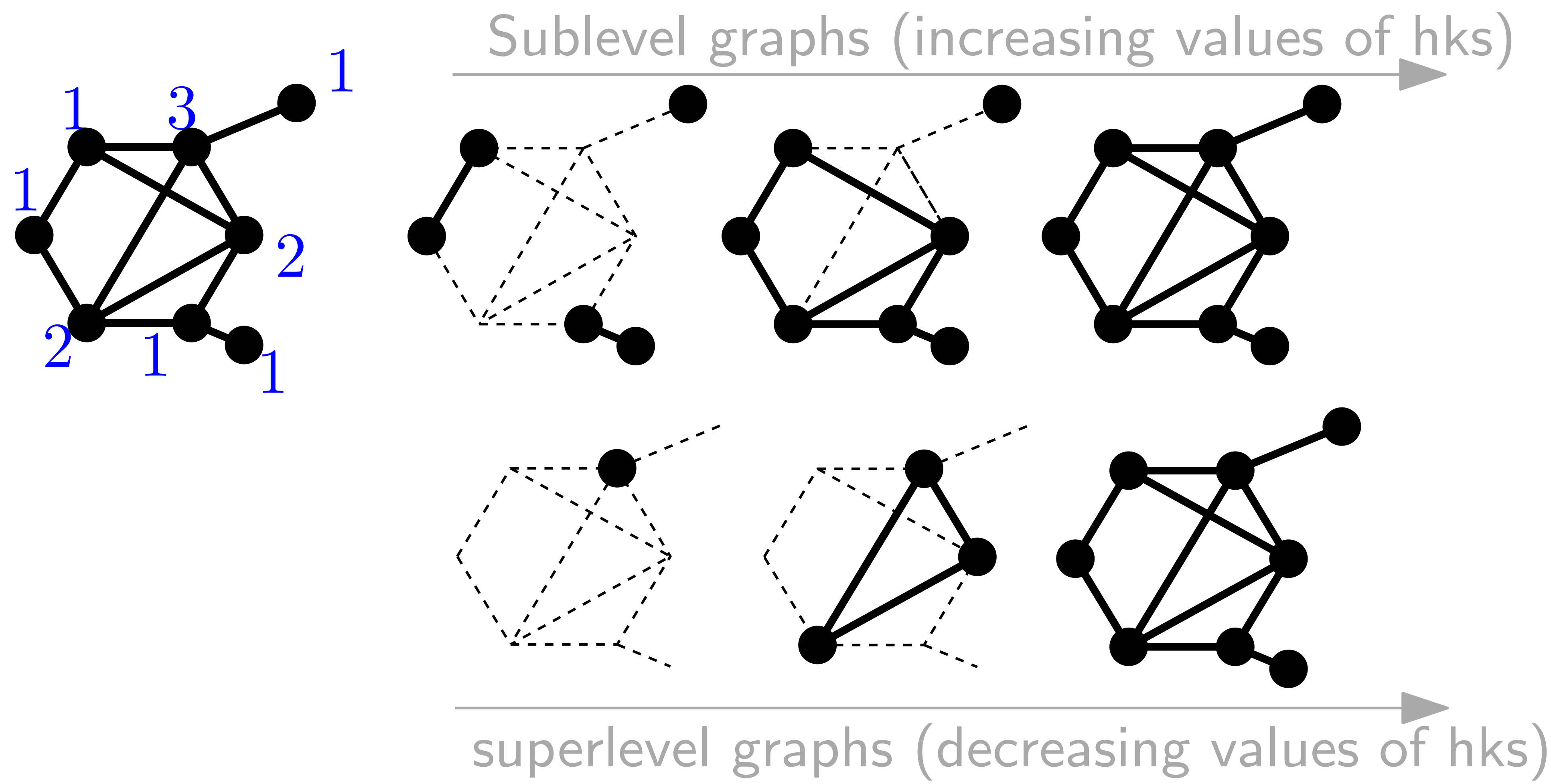


Def: Let $t \geq 0$, and define the *Heat Kernel Signature* of param t:

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

Application to graph classification

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



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

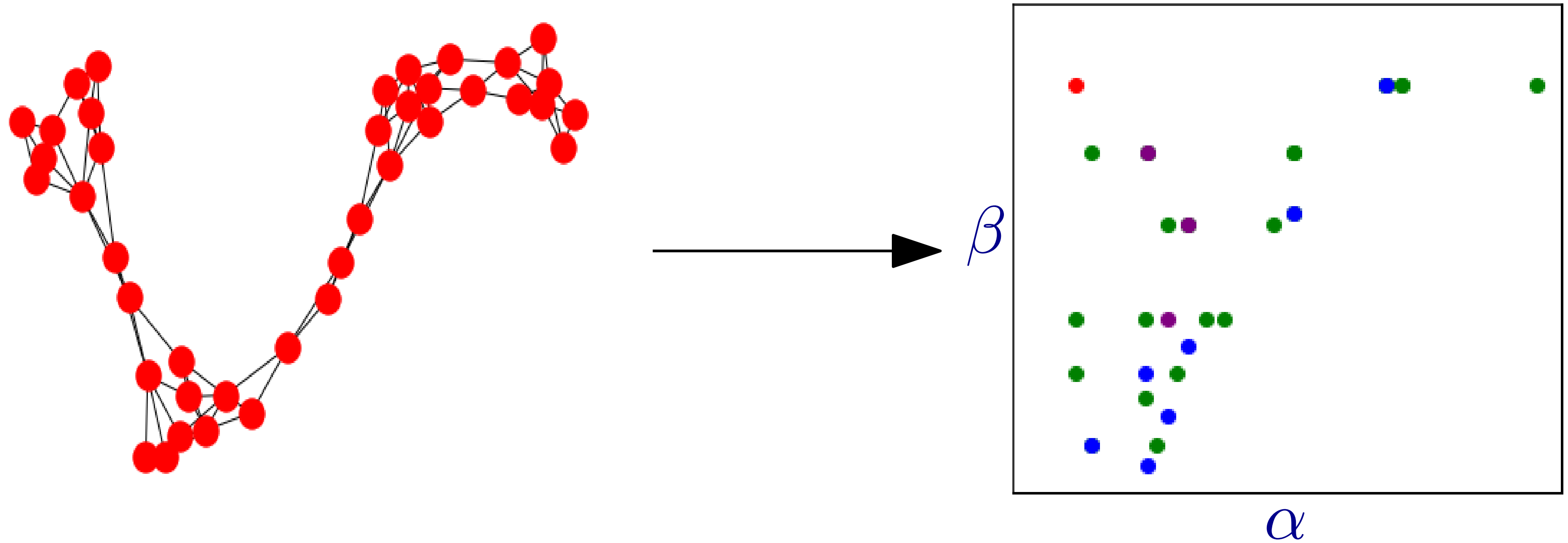
$$\mathrm{hks}_{G,t} : v \mapsto \sum_{k=1}^{n} \exp(-\lambda_k t)\phi_k(v)^2$$

Application to graph classification

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

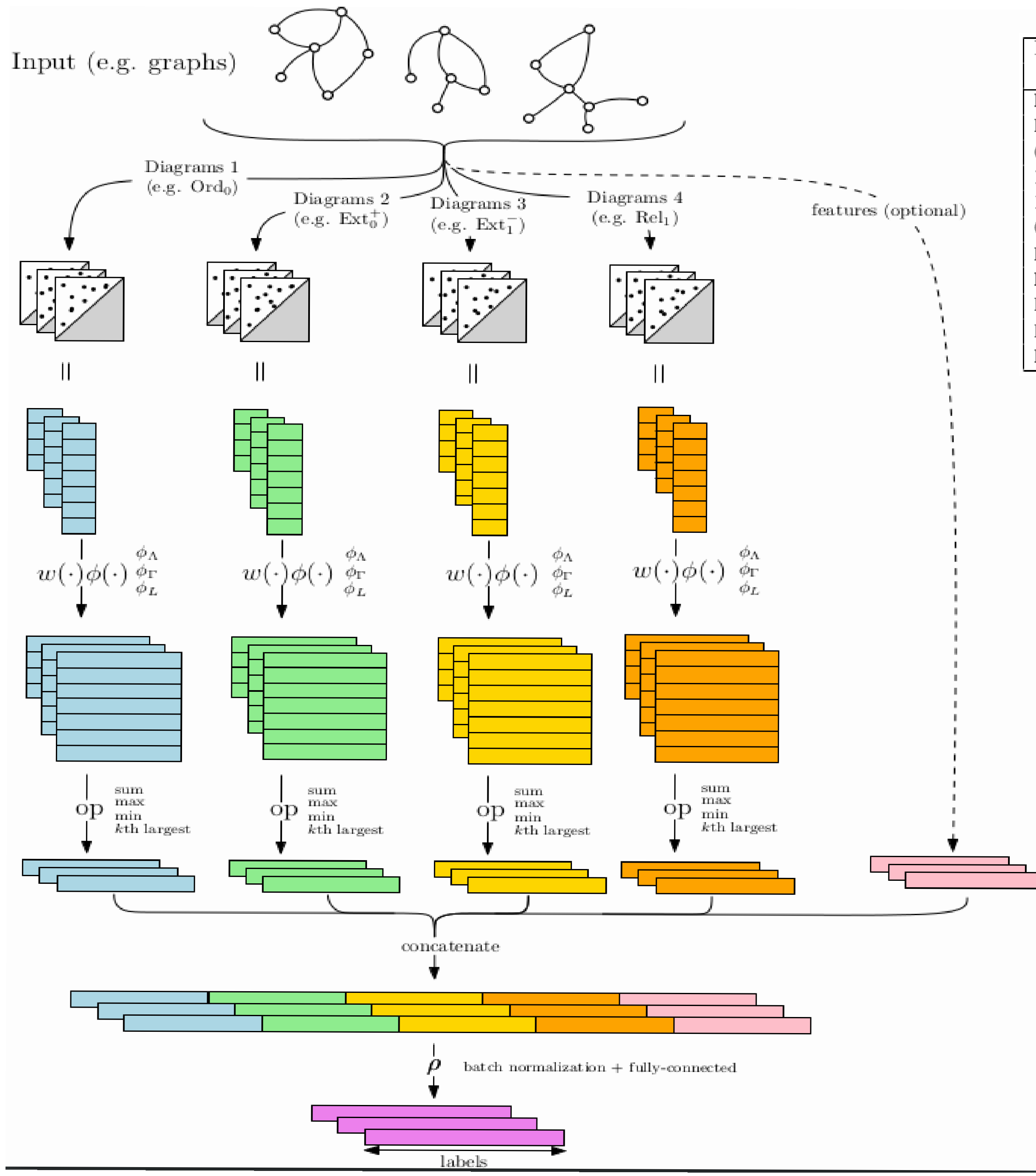
Graph from the `PROTEINS` dataset

Corresponding persistence diagram



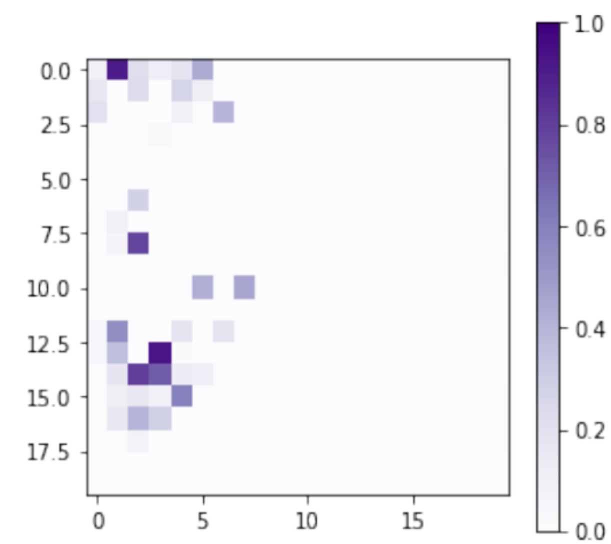
Application to graph classification

[*PersLay: A Neural Network Layer for Persistence Diagrams and New Graph Topological Signatures*, C., 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)

Summary

In this class, I introduced the basics of persistence representations.

We have seen that we can derive confidence regions on persistence diagrams using the stability theorem and (a, b)-standard measures.

We have seen that we turn persistence diagrams into vectors using persistence images and persistence landscapes.

We have seen how to automatically learn representations using PersLay and how to use it for graph classification.

In the next class, we will study how to guide models with persistence diagrams, with examples in clustering and regularization.

One kernel to rule them all...

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

Sliced Wasserstein Kernel

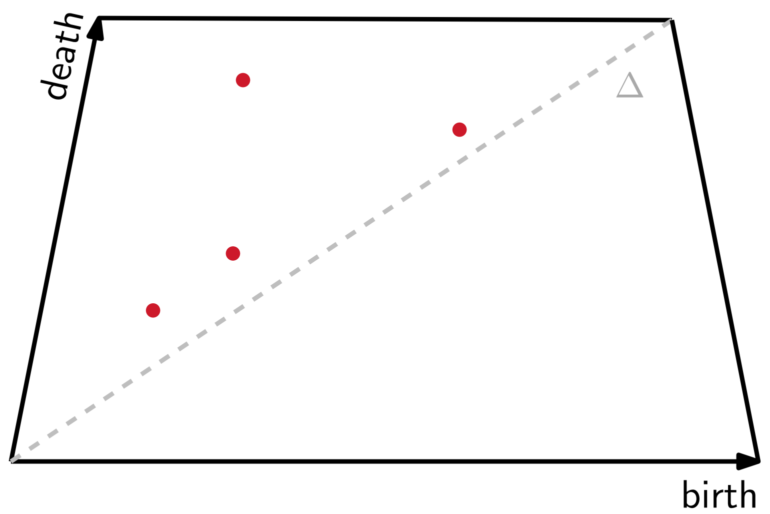
Provably stable

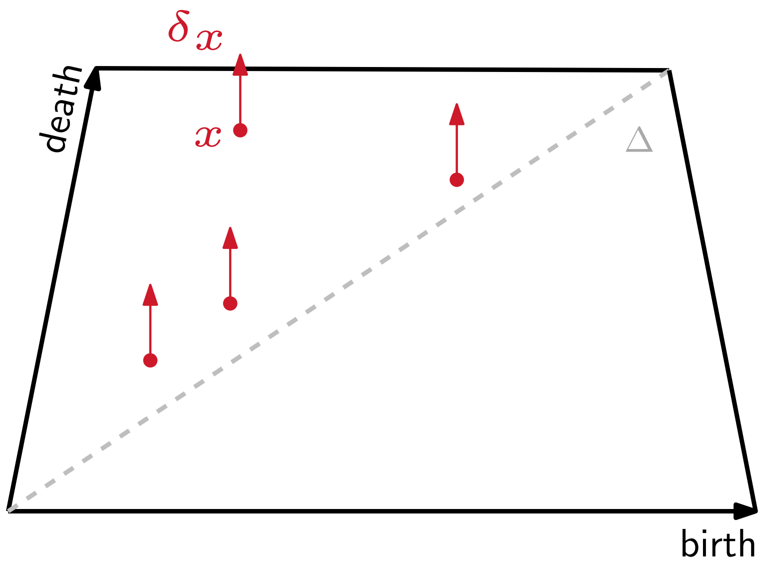
Provably discriminative

Mimicks the Gaussian kernel

View diagrams as discrete measures w/o density functions

Persistence diagrams as discrete measures

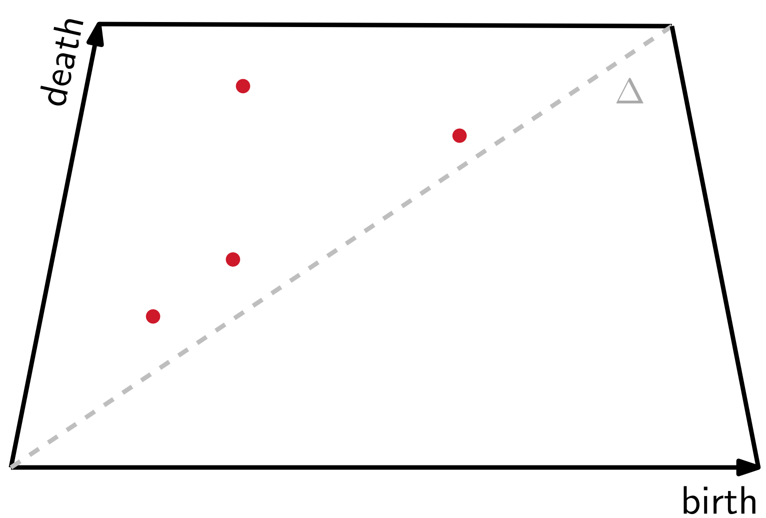


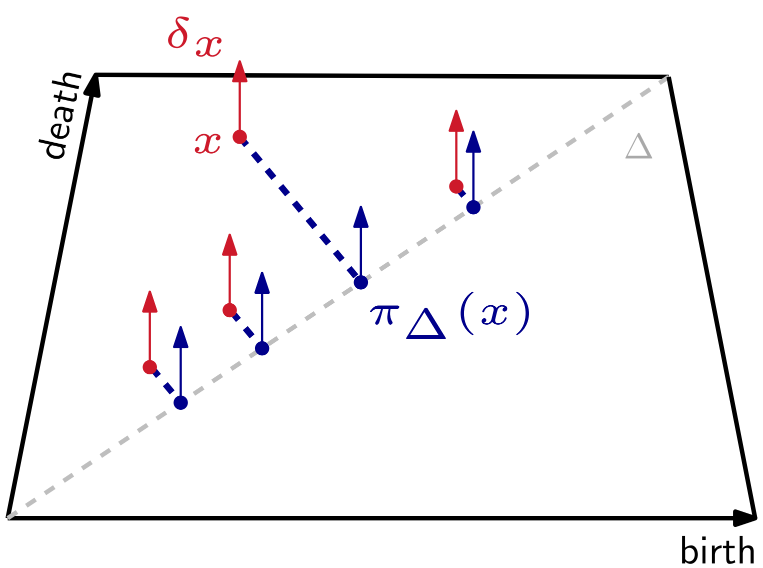


$$\mu_D := \sum_{x\in D} \delta_x$$

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





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

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

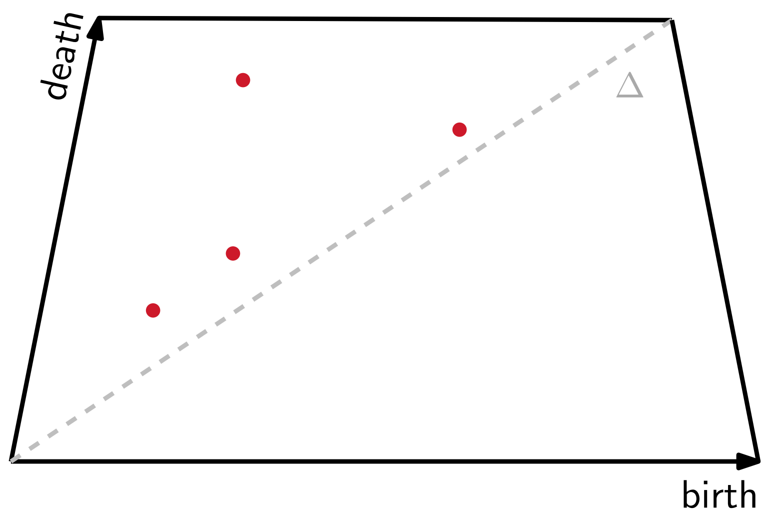
$\rightarrow$ given D, D', let

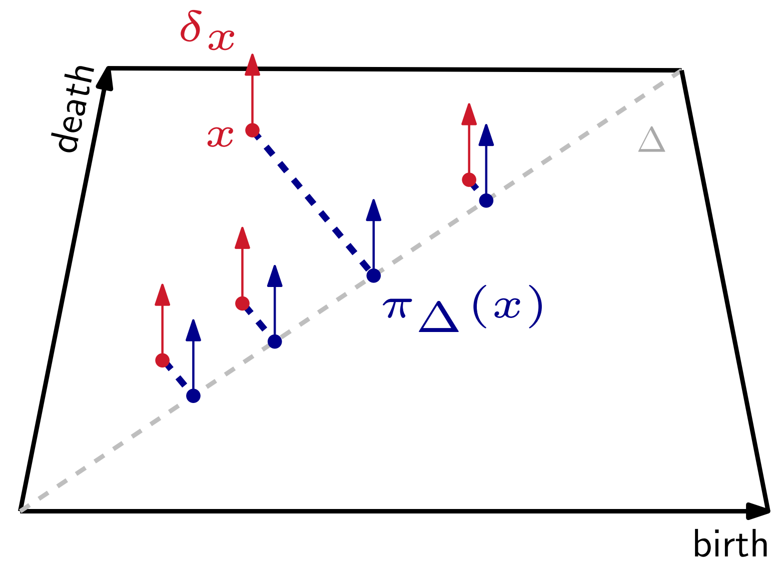
$$\bar{\mu}_D := \sum_{x \in D} \delta_x + \sum_{y \in D'} \delta_{\pi_\Delta(y)}$$

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

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





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

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

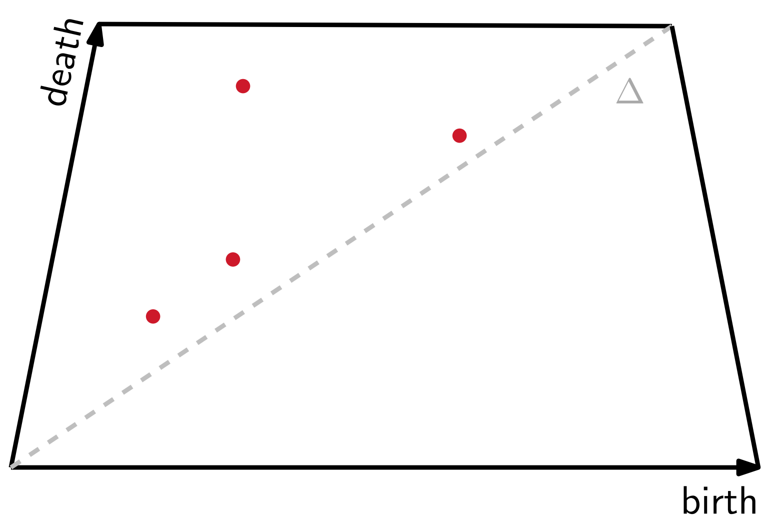
$\rightarrow$ given D, D', let

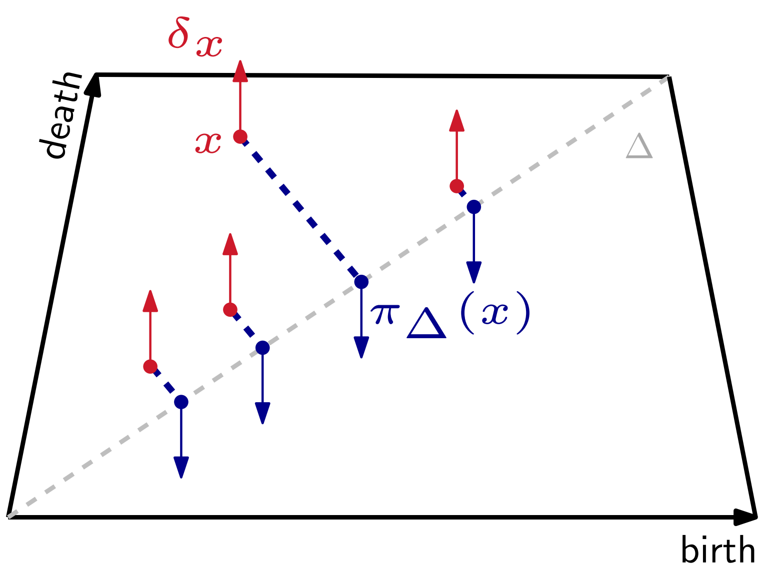
$$\bar{\mu}_D := \sum_{x\in D} \delta_x + \sum_{y\in D'} \delta_{\pi_\Delta(y)}$$

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

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





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

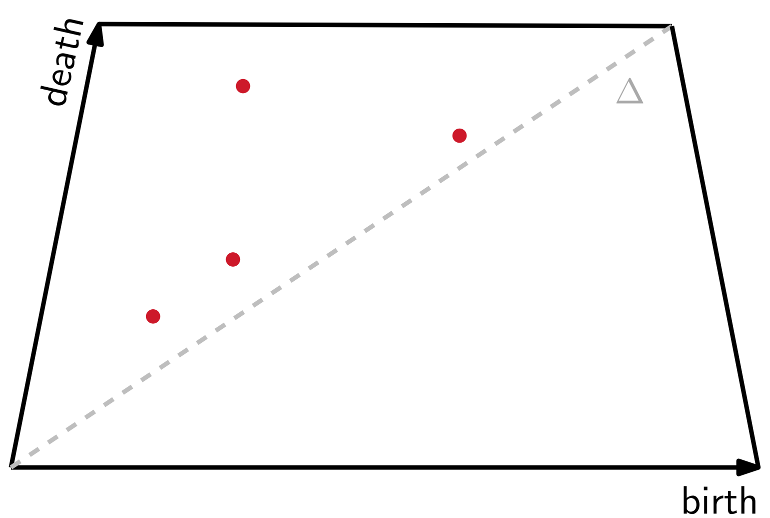
Pb: $\mathrm{d}_p(D, D') \not\propto W_p(\mu_D, \mu_{D'})$ (W_p does not even make sense here)

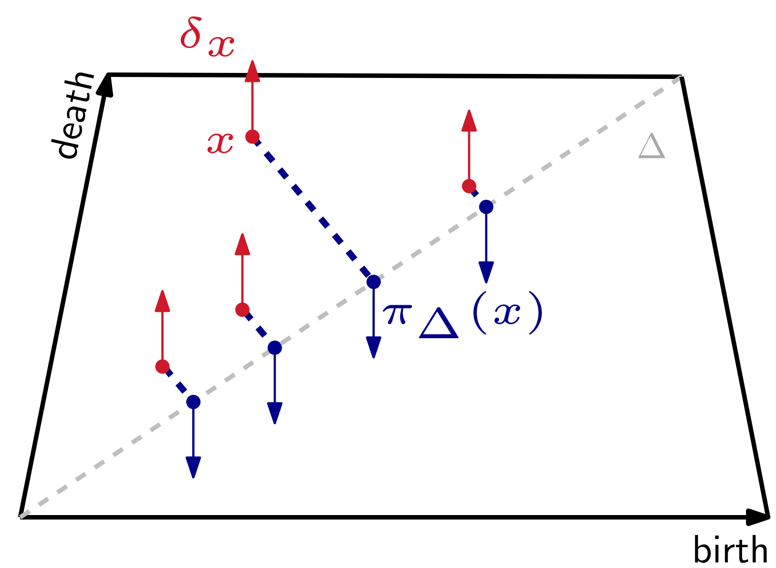
Solution: transfer mass negatively to μ_D:

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

$\rightarrow$ signed discrete measure of total mass zero

Persistence diagrams as discrete measures





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

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 to μ_D:

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

$\rightarrow$ signed discrete measure of total mass zero

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

Persistence diagrams as discrete measures

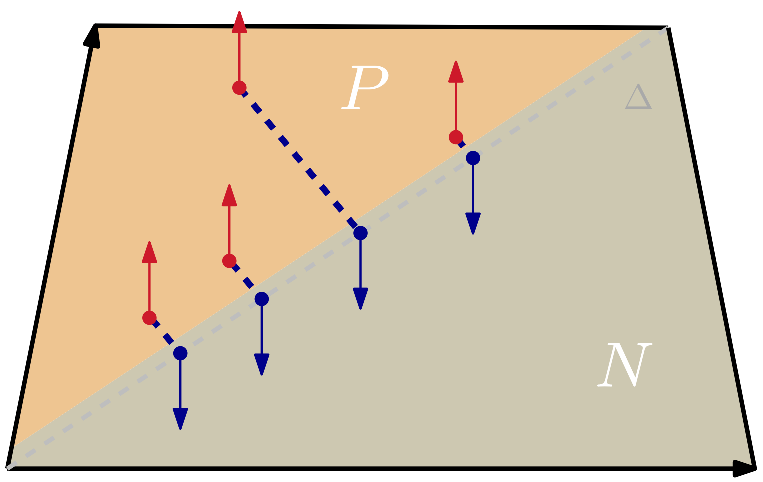
Hahn decomp. thm: For any $\mu \in \mathcal{M}_0(X, \Sigma)$ there exist measurable sets P, N such that:

(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: $\forall \mu, \nu \in \mathcal{M}_0(X)$, $\quad W_1(\mu^+ + \nu^-, \nu^+ + \mu^-) = \|\mu - \nu\|_K$

Persistence diagrams as discrete measures

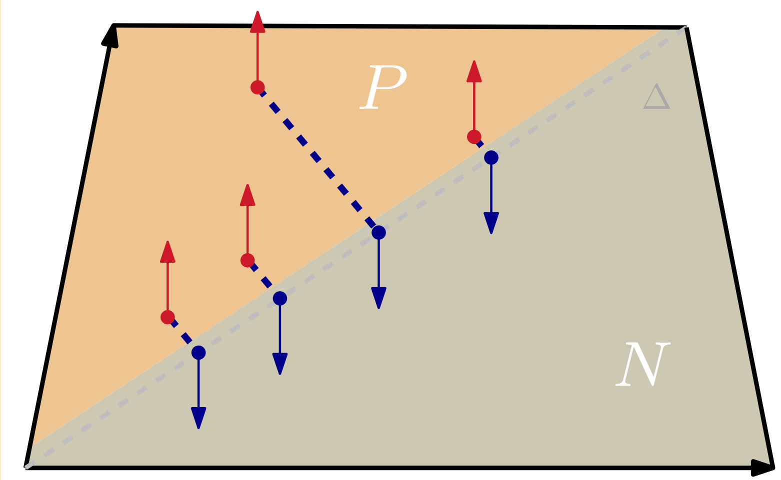
Hahn decomp. thm: For any $\mu \in \mathcal{M}_0(X, \Sigma)$ there exist measurable sets P, N such that:

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

for persistence diagrams:

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

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

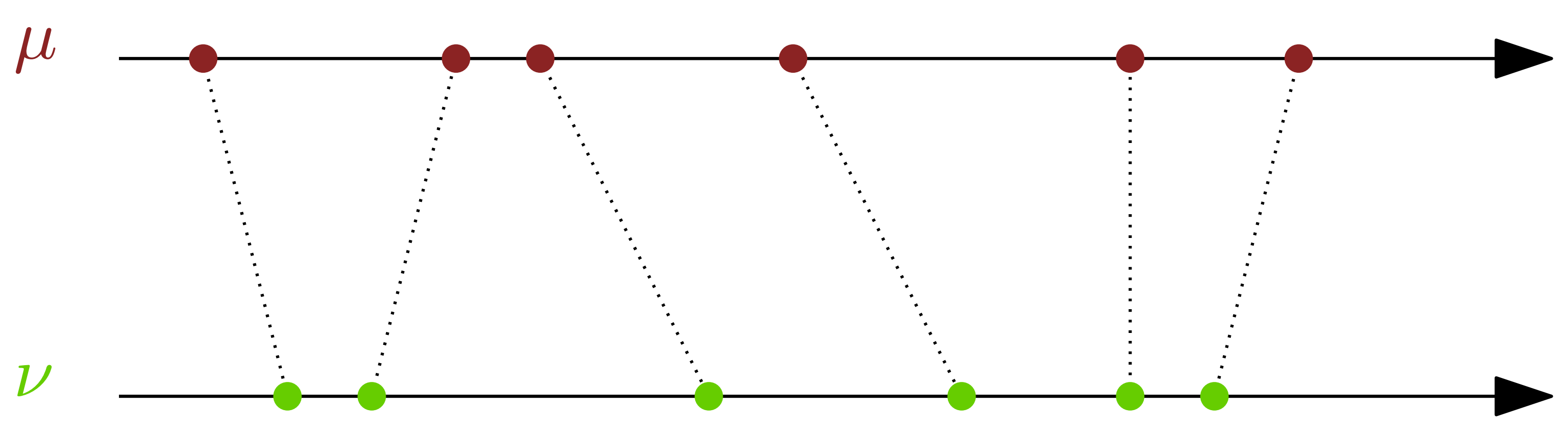
[*Sliced Wasserstein Kernel for persistence diagrams*, Carrière, Cuturi, Oudot, ICML, 2017]

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



Sliced Wasserstein metric

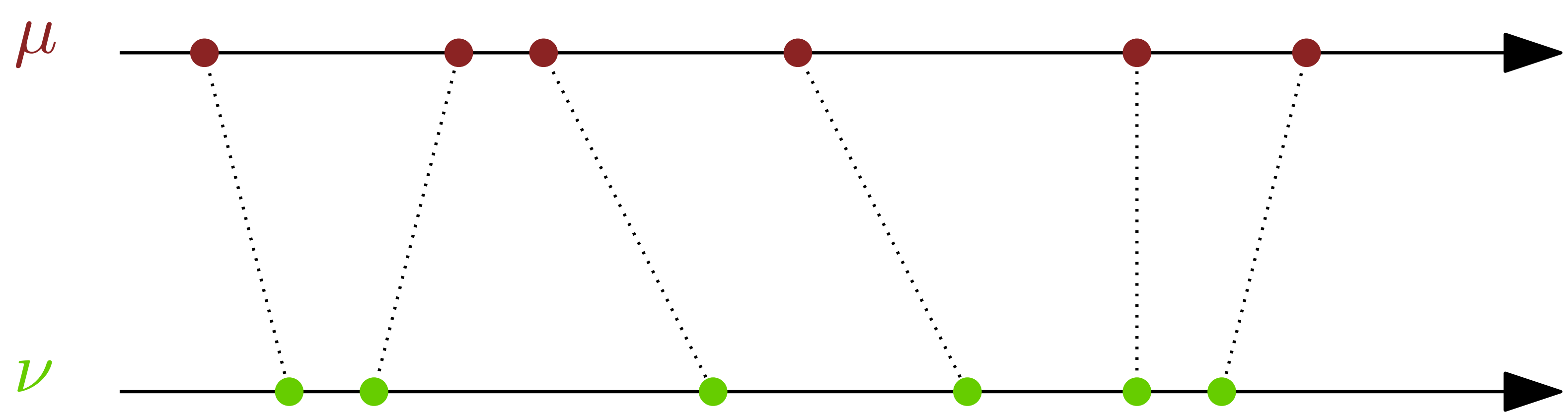
[*Sliced Wasserstein Kernel for persistence diagrams*, Carrière, Cuturi, Oudot, ICML, 2017]

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)

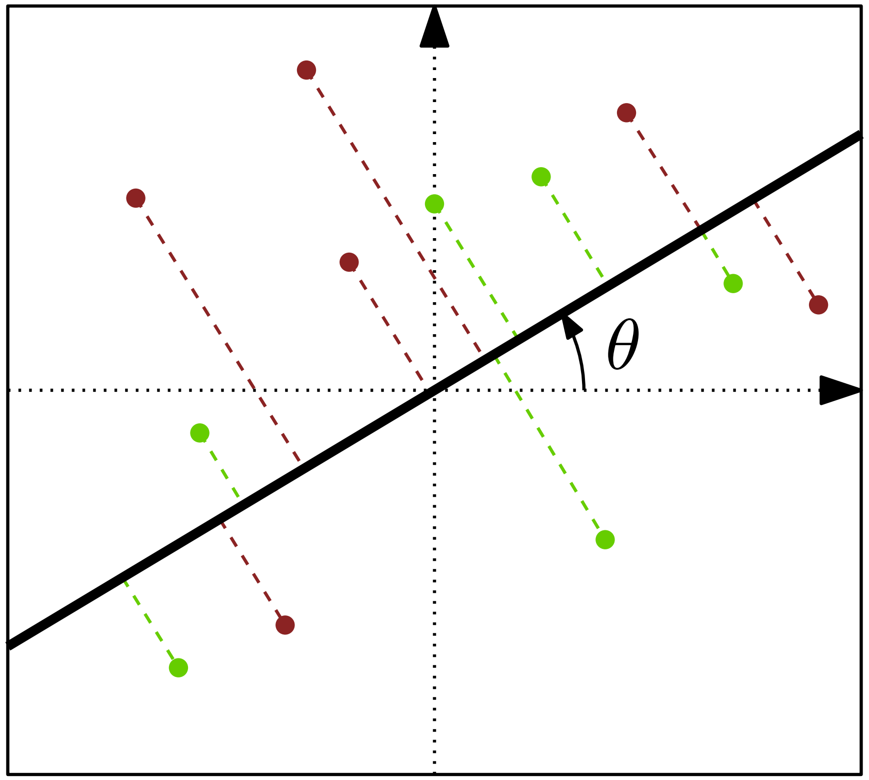
Sliced Wasserstein metric

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



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.

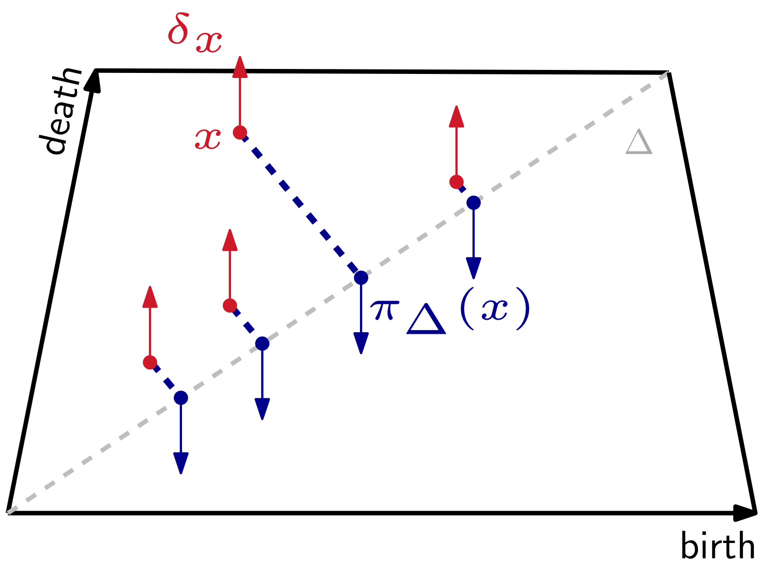
Sliced Wasserstein kernel

[*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.



$\rightarrow$ application to persistence diagrams:

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

$$\mapsto \tilde{\mu}_D := \mu_D - \pi_\Delta \# \mu_D$$

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

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

Sliced Wasserstein kernel

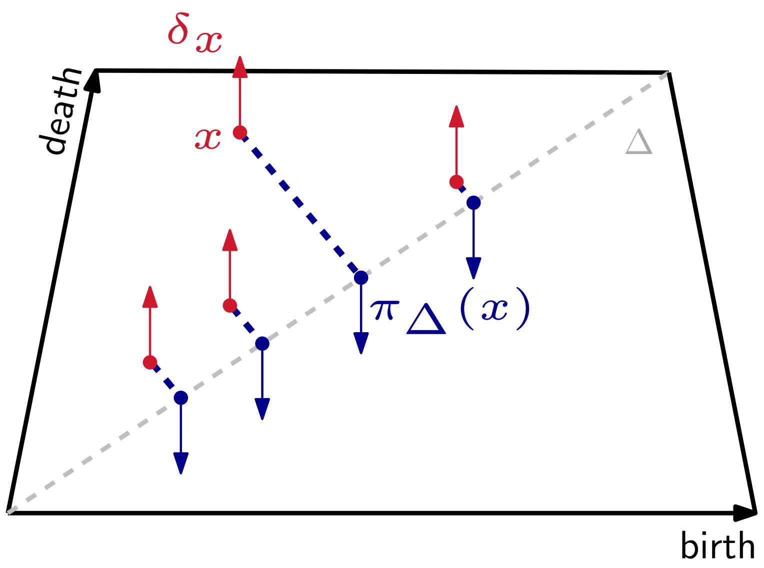
[*Sliced Wasserstein Kernel for persistence diagrams*, Carrière, Cuturi, Oudot, ICML, 2017]

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.

$\rightarrow$ application to persistence diagrams:



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

$$\mapsto \tilde{\mu}_D := \mu_D - \pi_\Delta \# \mu_D$$

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

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

- positive semidefinite
- simple and fast to compute

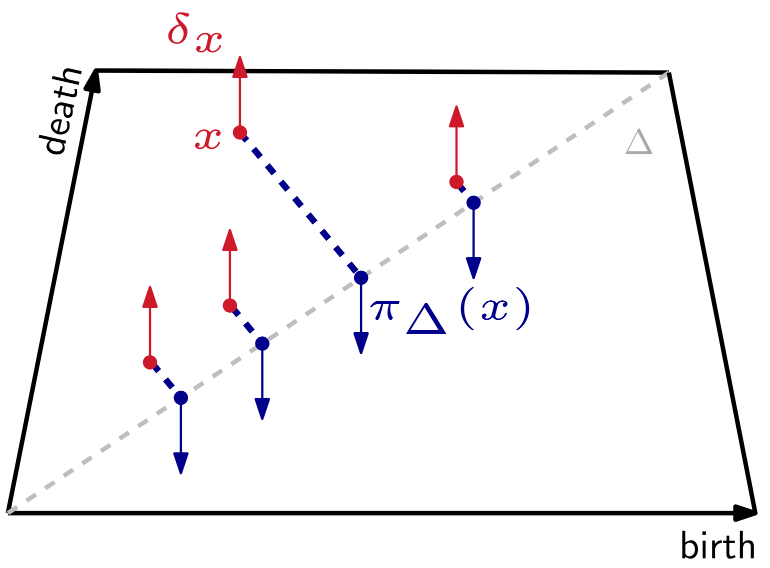
Sliced Wasserstein kernel

[*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 $D, D' \in \mathcal{D}_N$,

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

$\rightarrow$ application to persistence diagrams:



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

$$\mapsto \tilde{\mu}_D := \mu_D - \pi_\Delta \# \mu_D$$

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

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

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

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

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