A Kernel on Persistence Diagrams for Machine Learning

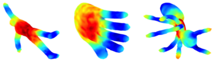
Jan Reininghaus[1] Stefan Huber[1] Roland Kwitt[2] Ulrich Bauer[1]

[1]Institute of Science and Technology Austria

[2]FB Computer Science
Universität Salzburg, Austria

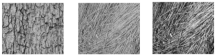
Toposys Annual Meeting
IST Austria — September 9, 2014

Pipeline in topological machine learning

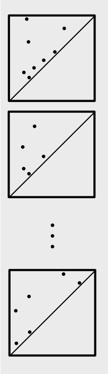


Task: shape retrieval

Task: object recognition

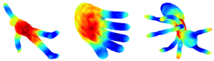


Task: texture recognition



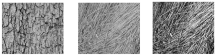
Topological data analysis

Pipeline in topological machine learning

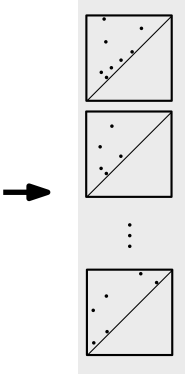


Task: shape retrieval

Task: object recognition



Task: texture recognition



SVM

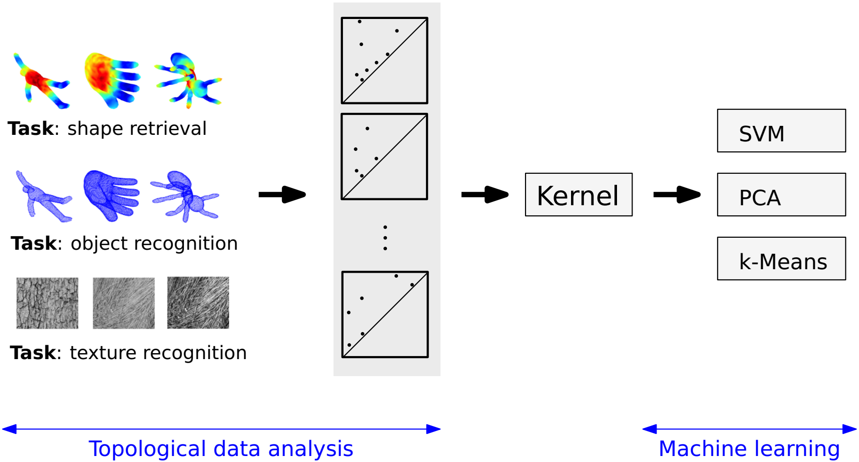
PCA

k-Means

Topological data analysis

Machine learning

Pipeline in topological machine learning



We want a **robust, multi-scale kernel** on the set of persistence diagrams!

Background on kernels

Definition

Given a non-empty set X, a (positive definite) kernel is a symmetric, positive-definite function $k\colon X \times X \to \mathbb{R}$, i.e.

$$k(x, x') = k(x', x) \qquad \forall x, x' \in X \tag{1}$$

$$[k(x_i, x_j)]_{i,j=1,1}^{n,n} \text{ is pos. semi-def.} \qquad \forall x_1, \ldots, x_n \in X, n > 0. \tag{2}$$

Example: An inner product on any vector space V is a kernel.

Background on kernels

Theorem

For any kernel k on a set X there exists a Hilbert space $\mathcal{H}$ and a mapping $\varphi\colon X \to \mathcal{H}$ with

$$k(x, x') = (\varphi(x), \varphi(x'))_{\mathcal{H}} \qquad \forall x, x' \in X. \tag{3}$$

Background on kernels: Examples

Given a Vector space V with inner product $(\cdot, \cdot)$.

- $k(x, x') = (x, x')$ is a kernel.
- $k(x, x') = (x, x')^d$, with $d \geq 0$, is a kernel.
- $k(x, x') = e^{-\lambda \|x - x'\|^2}$ is a kernel for all $\lambda > 0$.
  Radial basis function kernel (RBF-kernel).

Let $k_1, \ldots, k_m$ be kernels on a set X.

- $\sum_{k=1}^{m} \lambda_i k_i$, with $\lambda_i \geq 0$, is a kernel.

Background on kernels: Optimal assignment kernel

Goal: A kernel on the set of finite point sets in $\mathbb{R}^2$.

Definition

Let X, Y be two finite point sets in $\mathbb{R}^2$. The optimal assignment kernel is given as

$$k_A(X,Y) = \begin{cases} \max_{\pi:\, X \hookrightarrow Y} \sum_{x \in X} k(x, \pi(x)) & \text{for } |X| \leq |Y| \\ \max_{\pi:\, Y \hookrightarrow X} \sum_{y \in Y} k(\pi(y), y) & \text{otherwise} \end{cases}, \tag{4}$$

with $k(\cdot,\cdot)$ being a positive definite kernel for points in $\mathbb{R}^2$.

The optimal assignment kernel with $k(x,y) = e^{-\lambda \|x-y\|^2}$ and $\lambda > 0$ is **not** positive definite, i.e., not a kernel.

Hilbert-space embedding

Goal: A kernel on persistence diagrams, i.e., on finite multi-sets of points in $\mathbb{R}^2$.

Hilbert-space embedding

~~**Goal:** A kernel on persistence diagrams, i.e., on finite multi-sets of points in $\mathbb{R}^2$.~~

Goal: An embedding $\Psi\colon \mathcal{D} \to \mathcal{H}$ from the set $\mathcal{D}$ of persistence diagrams into a Hilbert space $\mathcal{H}$.

- The inner product of $\mathcal{H}$ will serve as kernel:

$$k(F, G) = (\Psi(F), \Psi(G))_{\mathcal{H}}$$

  with $F, G \in \mathcal{D}$.
- Or we use the RBF-kernel on $\mathcal{H}$, or a linear-combination of kernels, or ...

Embedding a multi-set of points

Given a persistence diagram D:

- Replace each point $p \in D$ with a Dirac delta δ_p centered at p:

$$\mathcal{D} \to H^{-2}(\mathbb{R}^2) \colon D \mapsto \sum_{p \in D} \delta_p$$

- However, the induced metric in $H^{-2}(\mathbb{R}^2)$ is not stable.
- We apply heat diffusion with $\sum_{p \in D} \delta_p$ as initial condition and Dirichlet boundary conditions at the diagonal.
  - Solutions are in $L_2(\mathbb{R}^2)$.
  - **Claim:** Solutions are stable.

Heat diffusion: Definition

Let $\Omega = \{(x, y) \in \mathbb{R}^2 \colon y \geq x\}$.

Heat-diffusion PDE

For a given diagram D we consider the solution $u \colon \Omega \times \mathbb{R}_{\geq 0} \to \mathbb{R}$ of

$$\partial_{xx} u + \partial_{yy} u = \partial_t u \quad \text{in } \Omega \times \mathbb{R}_{>0}, \tag{5}$$

$$u = 0 \quad \text{on } \partial\Omega \times \mathbb{R}_{\geq 0}, \tag{6}$$

$$u = \sum_{p \in D} \delta_p \quad \text{on } \Omega \times \{0\}. \tag{7}$$

- At scale $\sigma > 0$, the embedding of diagrams is given as:

$$\Psi_\sigma(D) = u(\cdot, \cdot, \sigma) \tag{8}$$

- $\Psi_\sigma(D)$ is in $L_2(\Omega)$, that is, Ψ_σ is an L_2-embedding.

Heat diffusion: The solution

Heat-diffusion PDE

For a given diagram D we consider the solution $u\colon \Omega \times \mathbb{R}_{\geq 0} \to \mathbb{R}$ of

$$\begin{aligned} \partial_{xx} u + \partial_{yy} u &= \partial_t u && \text{in } \Omega \times \mathbb{R}_{>0}, \\ u &= 0 && \text{on } \partial\Omega \times \mathbb{R}_{\geq 0}, \\ u &= \sum_{p \in D} \delta_p && \text{on } \Omega \times \{0\}. \end{aligned}$$

Heat diffusion: The solution

Heat-diffusion PDE

For a given diagram D we consider the solution $u \colon \mathbb{R}^2 \times \mathbb{R}_{\geq 0} \to \mathbb{R}$ of

$$\partial_{xx} u + \partial_{yy} u = \partial_t u \qquad \text{in } \mathbb{R}^2 \times \mathbb{R}_{>0}, \tag{9}$$

$$u = \sum_{p \in D} \delta_p - \delta_{\overline{p}} \qquad \text{on } \mathbb{R}^2 \times \{0\}. \tag{10}$$

Extend Ω to $\mathbb{R}^2$:

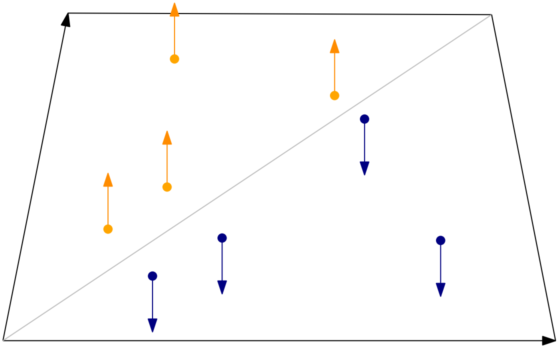
- Mirror and negate each δ_p at diagonal.
  - For $p = (a, b)$, we set $\overline{p} = (b, a)$.
- Restricting its solution to Ω solves the original problem.
- Solution is given by convolution with Gaussian kernel.

$$u(x, y, t) = \frac{1}{4\pi t} \sum_{p \in D} e^{-\frac{\|(x,y)-p\|^2}{4t}} - e^{-\frac{\|(x,y)-\overline{p}\|^2}{4t}} \tag{11}$$

Heat diffusion: The solution

The L_2-embedding at scale σ results in:

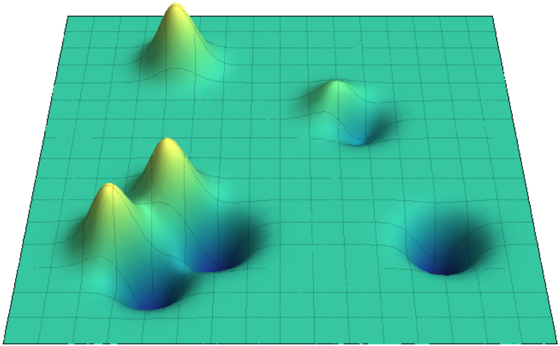
$$\Psi_\sigma(D)\colon \Omega \to \mathbb{R}\colon (x, y) \mapsto \frac{1}{4\pi\sigma} \sum_{p \in D} e^{-\frac{\|(x,y)-p\|^2}{4\sigma}} - e^{-\frac{\|(x,y)-\overline{p}\|^2}{4\sigma}} \qquad (12)$$



Heat diffusion: The solution

The L_2-embedding at scale σ results in:

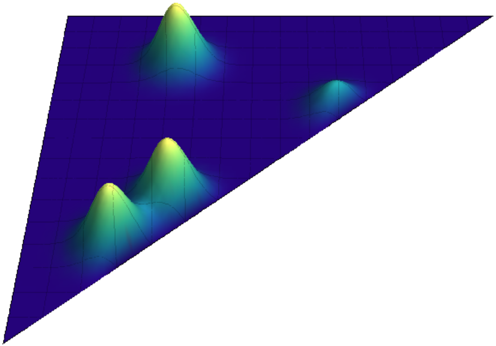
$$\Psi_\sigma(D)\colon \Omega \to \mathbb{R}\colon (x,y) \mapsto \frac{1}{4\pi\sigma} \sum_{p \in D} e^{-\frac{\|(x,y)-p\|^2}{4\sigma}} - e^{-\frac{\|(x,y)-\overline{p}\|^2}{4\sigma}} \qquad (12)$$



Heat diffusion: The solution

The L_2-embedding at scale σ results in:

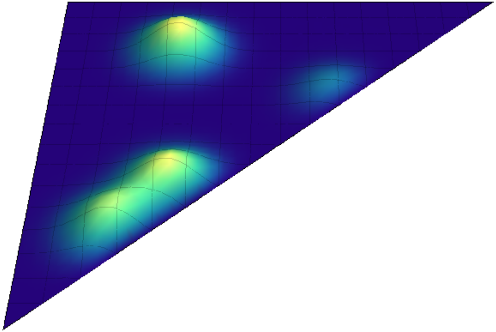
$$\Psi_\sigma(D)\colon \Omega \to \mathbb{R}\colon (x,y) \mapsto \frac{1}{4\pi\sigma} \sum_{p \in D} e^{-\frac{\|(x,y)-p\|^2}{4\sigma}} - e^{-\frac{\|(x,y)-\overline{p}\|^2}{4\sigma}} \tag{12}$$



Heat diffusion: The solution

The L_2-embedding at scale σ results in:

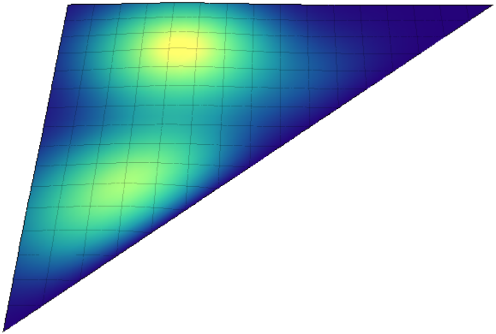
$$\Psi_\sigma(D)\colon \Omega \to \mathbb{R}\colon (x,y) \mapsto \frac{1}{4\pi\sigma} \sum_{p \in D} e^{-\frac{\|(x,y)-p\|^2}{4\sigma}} - e^{-\frac{\|(x,y)-\overline{p}\|^2}{4\sigma}} \qquad (12)$$



Heat diffusion: The solution

The L_2-embedding at scale σ results in:

$$\Psi_\sigma(D)\colon \Omega \to \mathbb{R}\colon (x,y) \mapsto \frac{1}{4\pi\sigma} \sum_{p \in D} e^{-\frac{\|(x,y)-p\|^2}{4\sigma}} - e^{-\frac{\|(x,y)-\overline{p}\|^2}{4\sigma}} \qquad (12)$$



The inner product

Given two diagrams F and G.

- We compute the inner product explicitly:

$$\begin{aligned}(\Psi_\sigma(F), \Psi_\sigma(G)) &= \int_\Omega \Psi_\sigma(F)\Psi_\sigma(G) \\ &= \frac{1}{8\pi\sigma} \sum_{\substack{p\in F \\ q\in G}} e^{-\frac{\|p-q\|^2}{8\sigma}} - e^{-\frac{\|p-\overline{q}\|^2}{8\sigma}} \end{aligned} \tag{13}$$

- $O(|F||G|)$ time, no approximation.

Robustness

Goal: We want to bound $\|\Psi_\sigma(F) - \Psi_\sigma(G)\|_{L_2}$ by $W_q(F, G)$.

Observation

A linear embedding $\Psi\colon \mathcal{D} \to V$, with a normed vector space V, cannot be stable w.r.t W_q for $q \in \{2, \ldots, \infty\}$.

- By *linear* we mean $\Psi(F + G) = \Psi(F) + \Psi(G)$ and $\Psi(\emptyset) = 0$.
- Set $G = \emptyset$ and $F_n = \{\underbrace{u, \ldots, u}_{n \text{ times}}\}$.
- We have

$$W_q(F_n, G) = \min_{\eta\colon F_n \leftrightarrow G} \sqrt[q]{\sum_{(u,v)\in\eta} \|u - v\|_\infty^q} = \sqrt[q]{n} \cdot W_q(F_1, G).$$

- But $\|\Psi(F_n) - \Psi(G)\|_V = n \cdot \|\Psi(F_1)\|_V$.

Robustness

Theorem

For any two persistence diagrams F and G, and any $\sigma > 0$ we have

$$\|\Psi_\sigma(F) - \Psi_\sigma(G)\|_{L_2(\Omega)} \leq \frac{8}{\sigma\sqrt{\pi}} W_1(F, G). \quad (14)$$

Multi-scale and noise:

- Noise on the input data causes $W_1(F, G)$ to increase.
- We can counteract by increasing σ.

Robustness: Proof I

Proof.

- Let $\eta^* = \arg\min_{\eta:\ F \leftrightarrow G} \sum_{(u,v)\in\eta} \|u - v\|_\infty$.
  - Augmenting D with diagonal points leaves $\Psi_\sigma(D)$ unchanged.
- We denote by $N_u(x) = \frac{1}{4\pi\sigma} e^{-\frac{\|x-u\|^2}{4\sigma}}$.

$$
\begin{aligned}
\|\Psi_\sigma(F) - \Psi_\sigma(G)\|_{L_2(\Omega)} &= \left\| \sum_{(u,v)\in\eta^*} (N_u - N_{\overline{u}}) - (N_v - N_{\overline{v}}) \right\|_{L_2(\Omega)} \\
&\leq 2 \sum_{(u,v)\in\eta^*} \|N_u - N_v\|_{L_2(\Omega)}
\end{aligned}
$$

Robustness: Proof II

Using $\|N_u - N_v\|_{L_2(\Omega)} = \frac{1}{\sqrt{4\pi\sigma}} \cdot \sqrt{1 - e^{-\frac{\|u-v\|_2^2}{8\sigma}}}$ we have

$$\|\Psi_\sigma(F) - \Psi_\sigma(G)\|_{L_2(\Omega)} \leq \frac{1}{\sqrt{\pi\sigma}} \sum_{(u,v)\in\eta^*} \sqrt{1 - e^{-\frac{\|u-v\|_2^2}{8\sigma}}}.$$

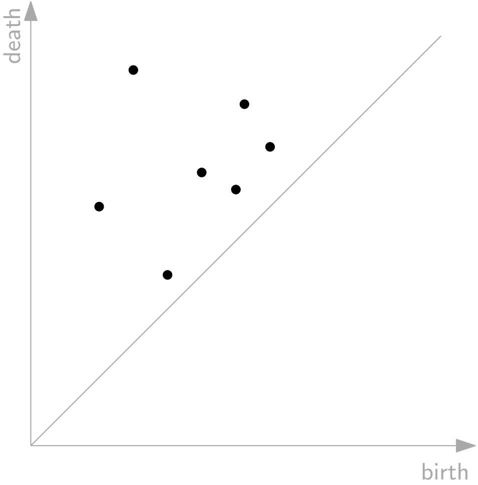
By the convexity of $x \mapsto e^x$, we have $e^{-\lambda x} \geq 1 - \lambda x$, and hence

$$\begin{aligned}
\|\Psi_\sigma(F) - \Psi_\sigma(G)\|_{L_2(\Omega)} &\leq \frac{1}{\sqrt{\pi\sigma}} \cdot \sum_{(u,v)\in\eta^*} \sqrt{\frac{1}{8\sigma}\|u-v\|_2^2} \\
&\leq \frac{1}{\sigma\sqrt{8\pi}} \cdot \sum_{(u,v)\in\eta^*} \|u-v\|_\infty \\
&\leq \frac{1}{\sigma\sqrt{8\pi}} W_1(F, G)
\end{aligned}$$

□

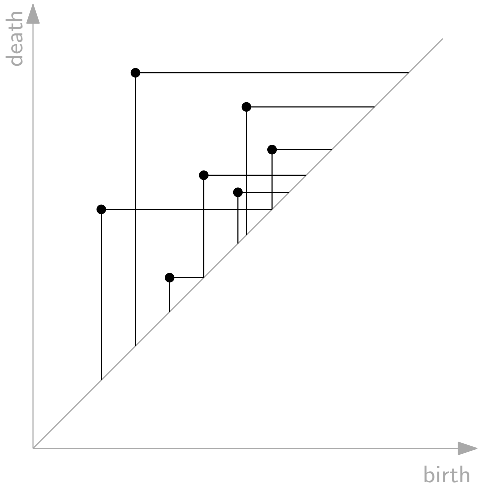
Persistence landscapes: Definition

- Introduced by Bubenik [Bubenik, 2013].



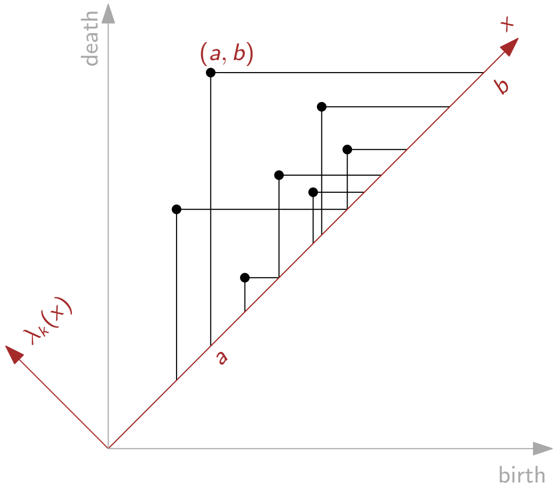
Persistence landscapes: Definition

- Introduced by Bubenik [Bubenik, 2013].



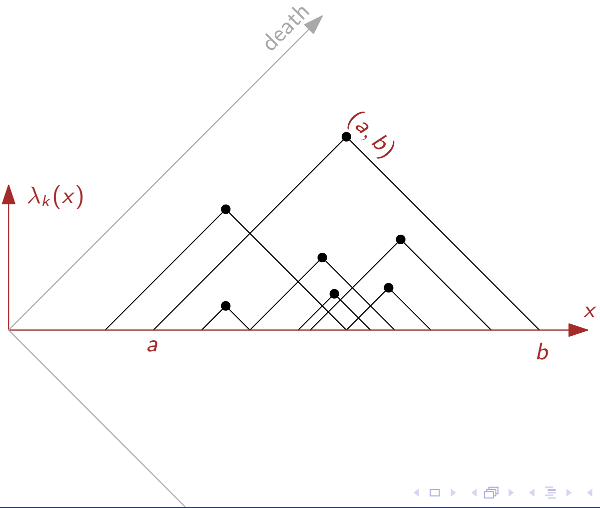
Persistence landscapes: Definition

- Introduced by Bubenik [Bubenik, 2013].



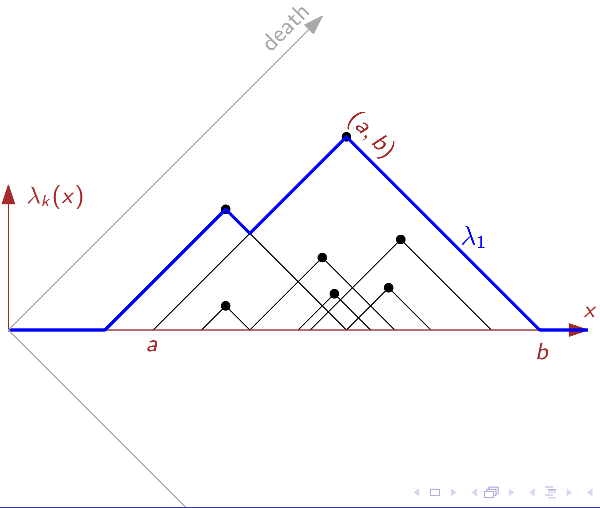
Persistence landscapes: Definition

- Introduced by Bubenik [Bubenik, 2013].



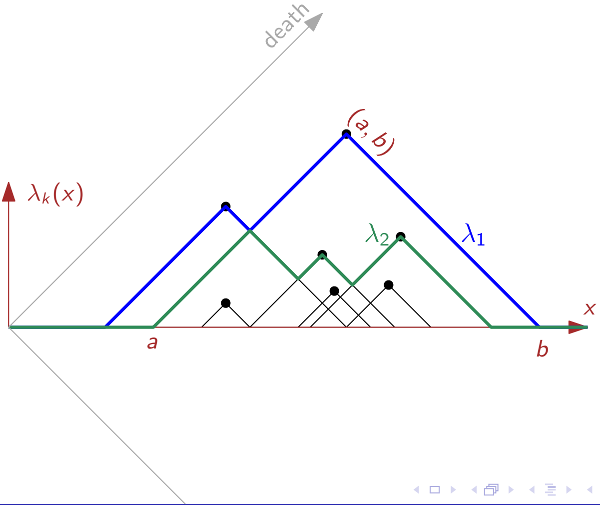
Persistence landscapes: Definition

- Introduced by Bubenik [Bubenik, 2013].



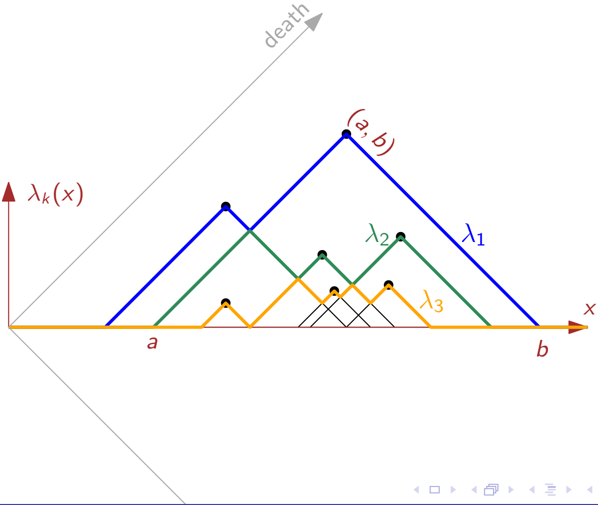
Persistence landscapes: Definition

- Introduced by Bubenik [Bubenik, 2013].



Persistence landscapes: Definition

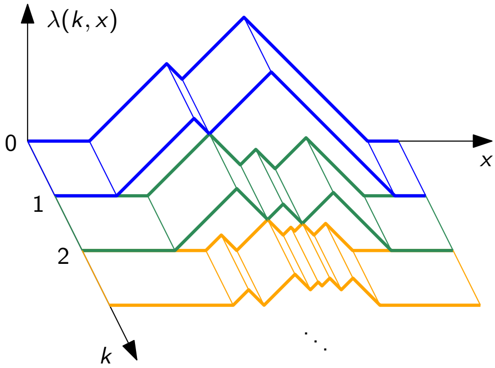
- Introduced by Bubenik [Bubenik, 2013].



Persistence landscapes: Definition

The persistence landscape of a persistence diagram can be interpreted as a function in $L_p(\mathbb{R}^2)$.

- Let $\Psi^L \colon \mathcal{D} \to L_p(\mathbb{R}^2) \colon D \mapsto \lambda(\cdot, \cdot)$ denote this embedding.



Persistence landscapes: Stability

Lemma (Bottleneck stability)

For two persistence diagrams F and G it holds that

$$\|\Psi^L(F) - \Psi^L(G)\|_{L_\infty(\mathbb{R}^2)} \leq W_\infty(F, G) \tag{15}$$

Lemma (Weighted Wasserstein stability)

For two persistence diagrams F and G it holds that

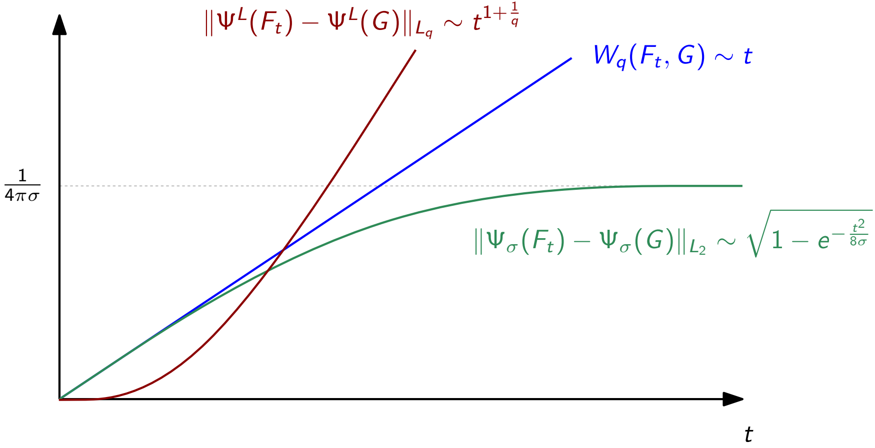
$$\|\Psi^L(F) - \Psi^L(G)\|_{L_q(\mathbb{R}^2)} \leq$$

$$\min_{\eta\colon F \leftrightarrow G} \sqrt[q]{\sum_{(u,v)\in\eta} \operatorname{pers}(u)\|u - v\|_\infty^q + \frac{2}{q+1}\|u - v\|_\infty^{q+1}} \tag{16}$$

- $\Psi^L\colon \mathcal{D} \to L_q(\mathbb{R}^2)$ is not a linear embedding.

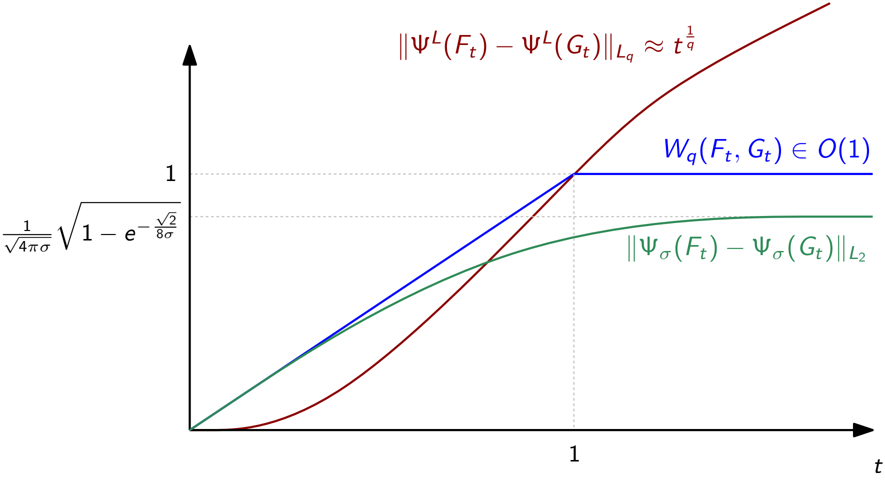
Stability comparison: Experiment 1

Let $G = \emptyset$ and let $F_t = \{(-t, t)\}$, with $t \in \mathbb{R}_{\geq 0}$.



Stability comparison: Experiment 2

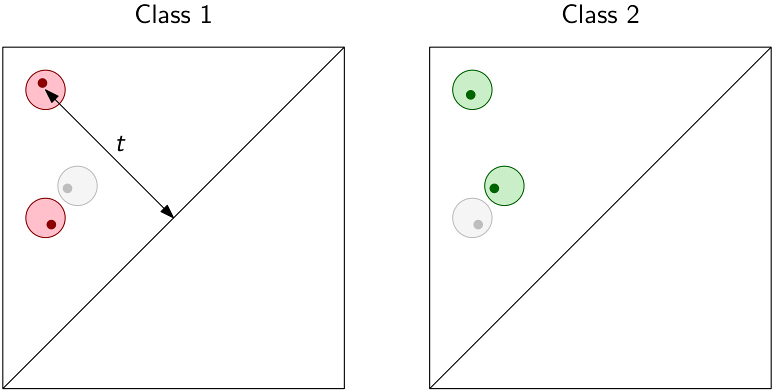
Let $F_t = \{(-t, t)\}$ and $G_t = \{(-t+1, t+1)\}$, with $t \in \mathbb{R}_{\geq 0}$.



Stability comparison: Experiment 3

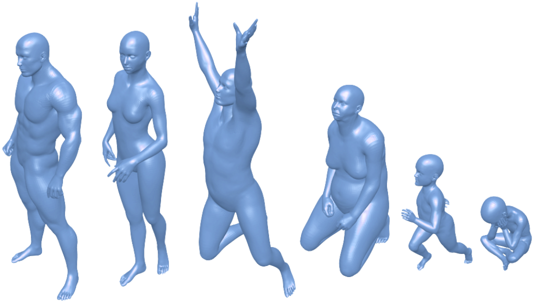
Persistence landscapes emphasize high-persistence features.

- Noise at high-persistence points can hide significant mid-range features.
- It becomes hard to distinguish between Class 1 and Class 2.



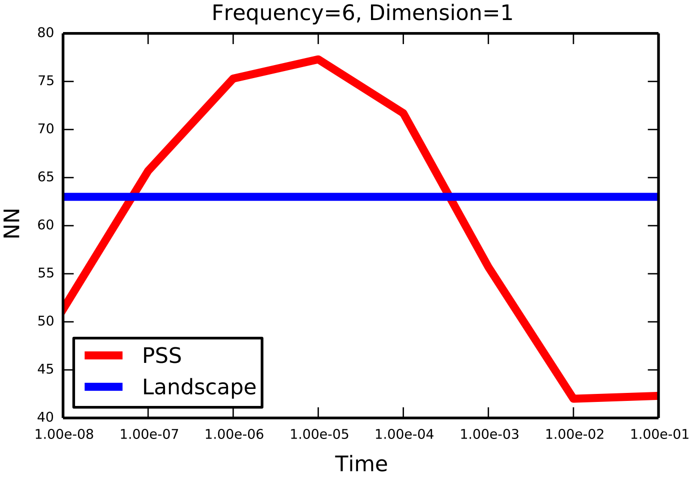
Experiments: PSS- vs. Landscape-kernel

- SHREC 2014 dataset of shapes:
  - 300 meshes, 20 classes (poses).



- Morse-filtration of heat-signature scalar field on each mesh.
  - An additional scale parameter ("frequency").
- 2 Tasks:
  - Retrieval: Get the nearest neighbor in Hilbert space.
  - Classification: Train a SVM with the kernel.

Experiments: Retrieval



Experiments: Classification

Freq.	**PSS [%]**	**Landscapes [%]**	Diff.
1	95.67 ± 2.53	70.67 ± 3.65	+25.00
2	99.33 ± 0.91	90.67 ± 3.03	+8.67
3	96.67 ± 2.64	76.00 ± 3.84	+20.67
4	97.67 ± 1.49	85.33 ± 6.39	+12.33
5	97.33 ± 1.90	86.67 ± 2.04	+10.67
6	93.67 ± 3.21	71.67 ± 6.56	+22.00
7	89.33 ± 2.79	73.33 ± 7.17	+16.00
8	88.67 ± 5.58	81.33 ± 4.31	+7.33
9	90.67 ± 5.35	74.33 ± 7.32	+16.33
10	92.00 ± 3.21	59.67 ± 5.19	+32.33
11	92.00 ± 2.74	70.00 ± 6.56	+22.00
12	86.67 ± 2.64	61.67 ± 8.08	+25.00
13	86.67 ± 3.33	70.67 ± 6.93	+16.00
14	89.67 ± 2.98	58.67 ± 7.49	+31.00
15	87.33 ± 3.46	55.67 ± 8.04	+31.67
16	82.33 ± 4.18	55.67 ± 6.08	+26.67
17	86.00 ± 4.50	54.67 ± 5.70	+31.33
18	85.67 ± 5.08	52.33 ± 4.94	+33.33
19	79.00 ± 5.60	48.00 ± 5.19	+31.00
20	72.33 ± 2.53	43.00 ± 3.21	+29.33

Table : Comparison of the PSS- *vs.* Landscape-kernel using 5-fold cross-validation on the SHREC2014 (Synthetic) dataset for dimension 1.

Bibliography I

Bubenik, P. (2013).
Statistical topological data analysis using persistence landscapes.
arXiv, available at http://arxiv.org/abs/1207.6437.