Large Scale computation of Means and Clusters for Persistence Diagrams using Optimal Transport

Théo Lacombe(1), Marco Cuturi(2), Steve Oudot(1)
(1)Inria Saclay, datashape. (2)CREST, ENSAE & Google Brain

Overview

Topological Data Analysis:

- Provides descriptors, called **persistence diagrams (PDs)**, of the topology of an object at all scales.
- Compares PDs with partial matching metrics.

Problem motivation:

- Hard to compute elementary statistics such as means.
- Current algorithm [1] to estimate PD barycenters is non-convex and intractable on large data.

Our contributions:

- Reformulate PD metrics as exact OT problems.
- Adapt the OT *entropic smoothing* [2] for PD metrics, in particular convolution on regular grids [3] allowing parallelization and GPU computations.
- Propose a convex formulation and scalable algorithm for PD barycenter estimation.

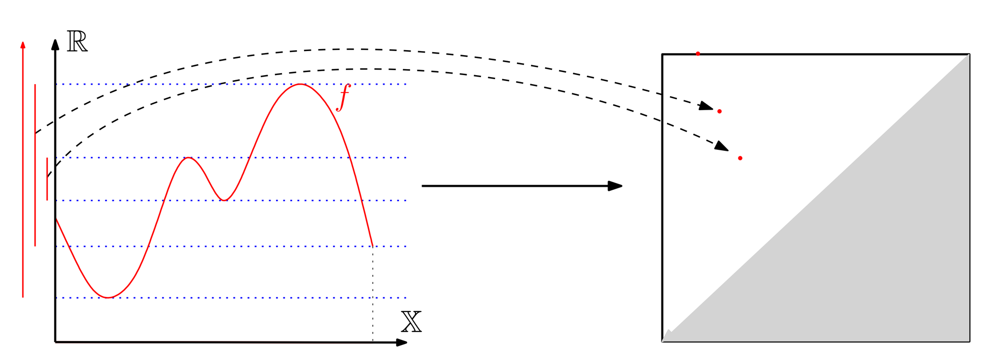


Figure 1: TDA sketch: filtration of a space $\mathbb{X}$ with a function f and corresponding PD accounting for the topology in the sublevel sets of f.

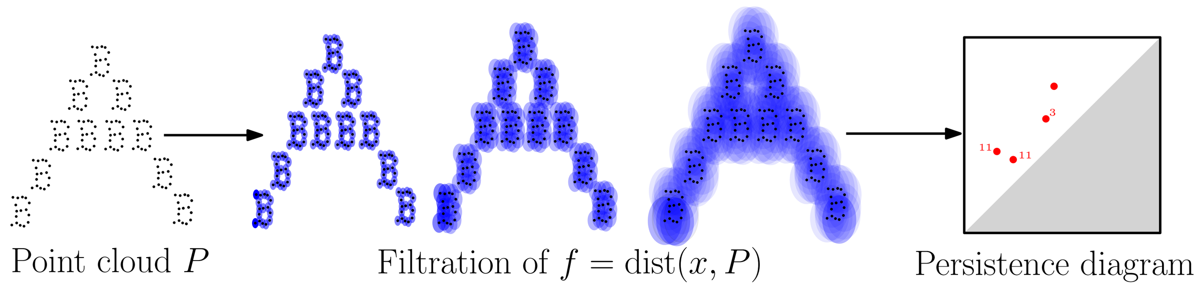


Figure 2: TDA sketch: filtration on a point cloud and corresponding PD.

I. Persistence diagrams and metrics

Persistence diagrams (PDs) are finite *point measures*, i.e. $\mu = \sum_{i=1}^{n} \delta_{x_i}$, with $x_i \in \{(t_1, t_2) \in \mathbb{R}^2, t_2 > t_1\}$. For $p \geq 1$,

$$d_p(\mu, \nu) := \left(\min_{\zeta \in \Gamma(\mu,\nu)} \sum_{(x,y) \in \zeta} \|x - y\|^p + \sum_{s \notin \zeta} \|s - \pi_\Delta(s)\|^p \right)^{\frac{1}{p}},$$

with $\Gamma(\mu, \nu)$: **partial** matchings between μ and ν, and $\pi_\Delta(s)$ the orthogonal projection of s onto the diagonal.

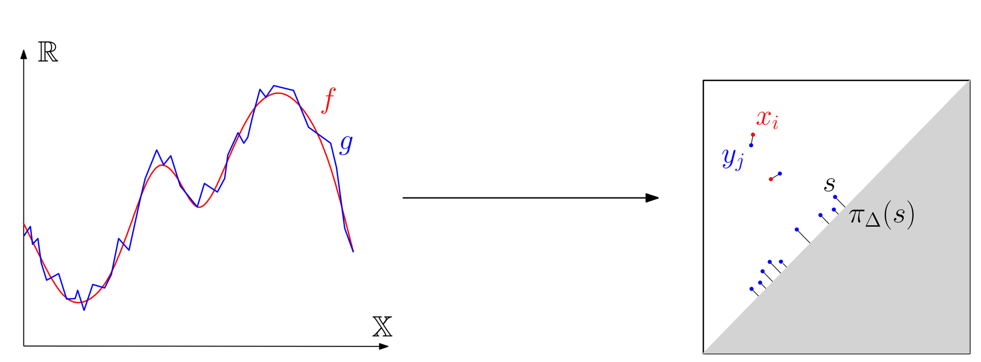
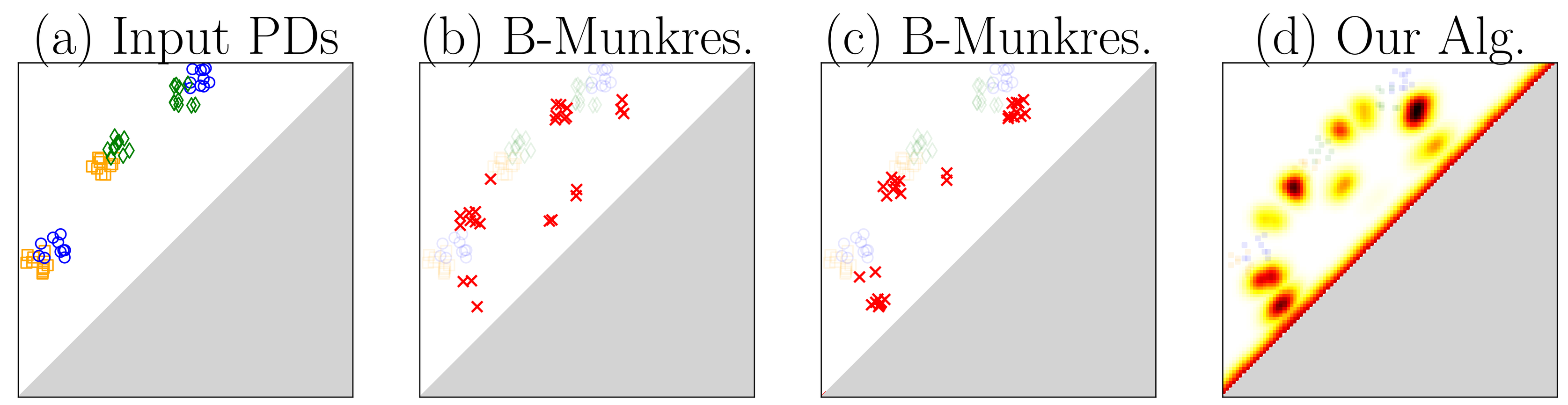


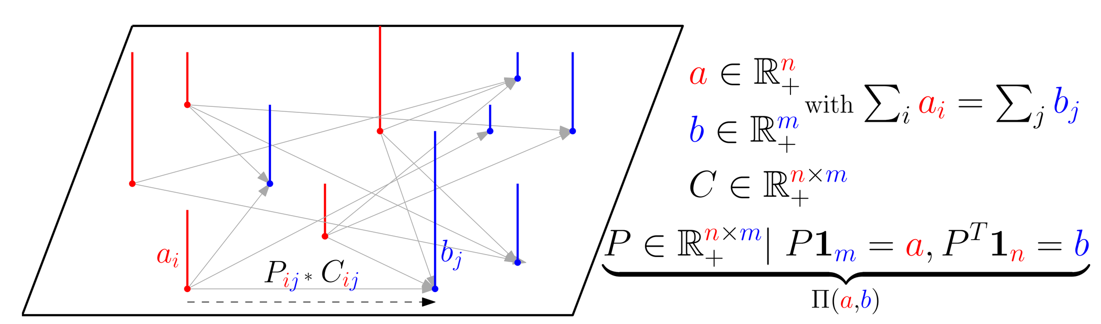
Figure 3: *(left)* Two functions $f, g : \mathbb{X} \to \mathbb{R}$. *(right)* Corresponding PDs and an optimal partial matching ζ (edges).



	Final energy
B-Munkres (b)	0.589
B-Munkres (c)	0.555
Our Alg. (d)	**0.542**

Figure 4: Illustration of our approach on a simple example. *(a)* 3 PDs for which we want to estimate a barycenter. *(b,c)* Outputs of B-Munkres algorithm [1] for two different initializations. Variability is due to non-convexity. *(d)* The output of our convex formulation. It performs better (lower energy).

II. Smoothed optimal transport (OT)



Smoothed OT problem ($\gamma > 0$):

$$\mathbf{L}_C^\gamma(a,b) := \min_{P \in \Pi(a,b)} \langle P, C \rangle - \gamma h(P)$$

where $h(P) := -\sum_{ij} P_{ij}(\log P_{ij} - 1)$.

Advantages:

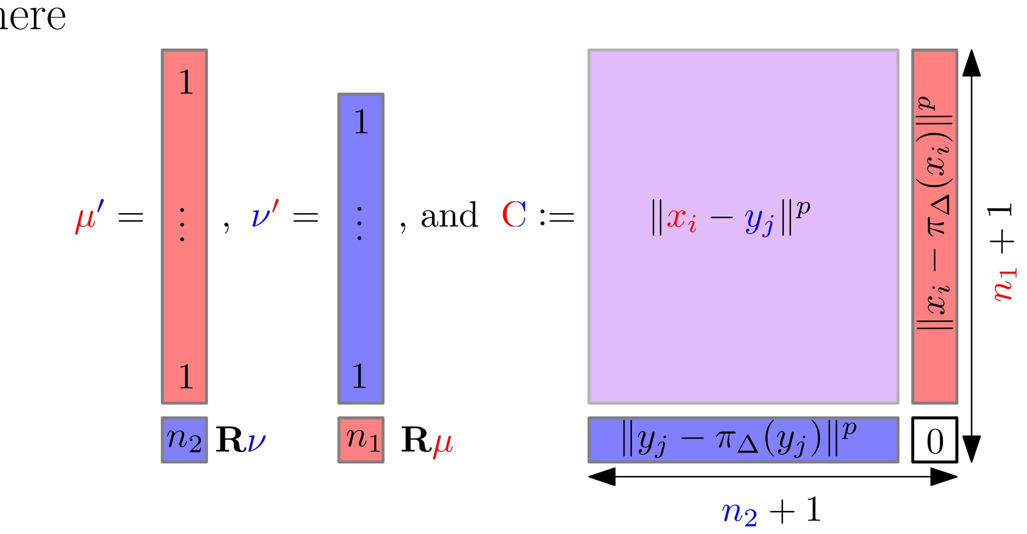
- Solved by iterating $(u, v) \mapsto \left(\frac{a}{Kv}, \frac{b}{K^T u}\right)$, with $K := e^{-\frac{C}{\gamma}}$.
- Converges to $\mathbf{L}_C(a,b) := \min\{\langle P, C\rangle ; P \in \Pi(a,b)\}$ when $\gamma \to 0$, with controllable error (upper and lower bounds).
- Numerically efficient to solve: GPU + Parallelism.
- Differentiable, with tractable gradient.

III. OT formulation of d_p

Consider $\mu = \sum_{i=1}^{n_1} \delta_{x_i}$ and $\nu = \sum_{j=1}^{n_2} \delta_{y_j}$. We have:

$$d_p(\mu, \nu) = (\mathbf{L}_C(\mu', \nu'))^{\frac{1}{p}},$$

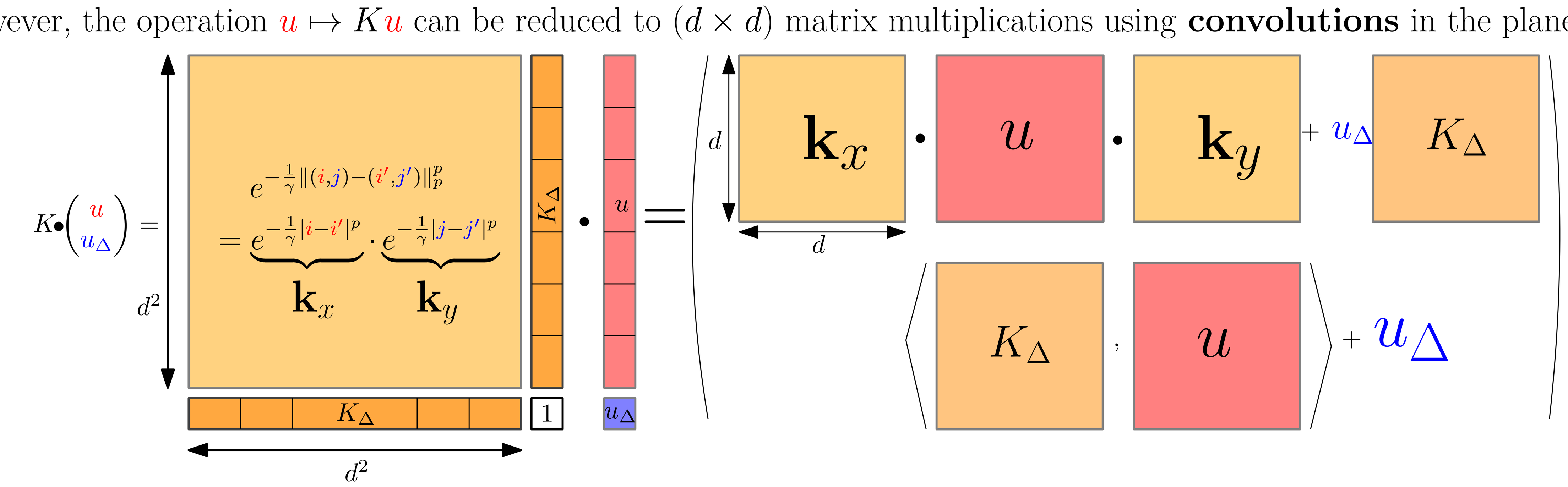
where



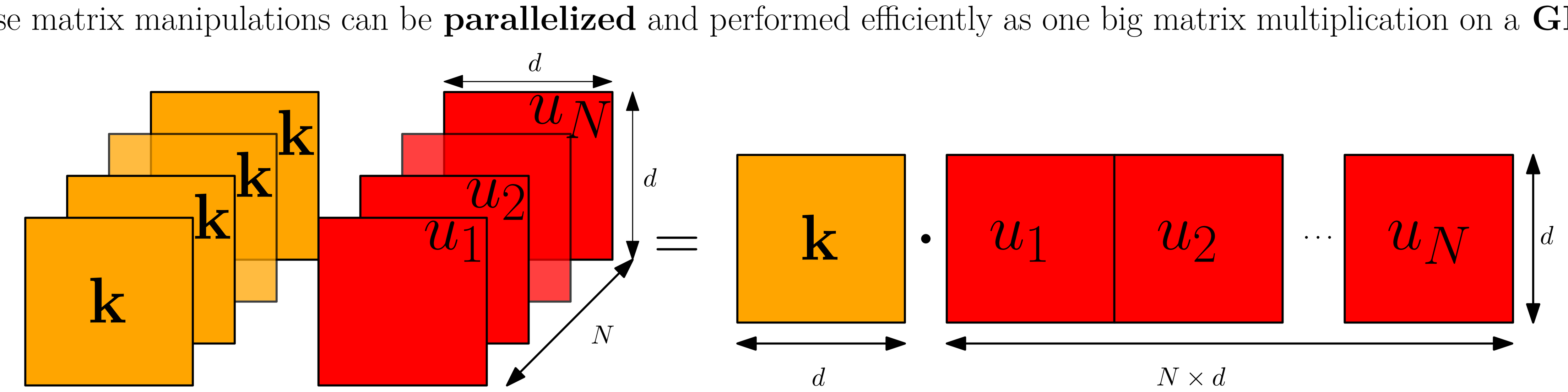
Idea: Approximate d_p with L_C^γ.

IV. Fast convolutions in the PD space

Discretize PDs on a $d \times d$ grid (+1 for the diagonal), $\Rightarrow$ $(d^2 + 1)$ **histograms**. C, K are $(d^2+1) \times (d^2+1)$ shaped. However, the operation $u \mapsto Ku$ can be reduced to $(d \times d)$ matrix multiplications using **convolutions** in the plane.



These matrix manipulations can be **parallelized** and performed efficiently as one big matrix multiplication on a **GPU**.



V. Smoothed barycenters for PDs

For $h_1 \ldots h_N$ histograms, a barycenter (Fréchet mean) is a minimizer of the energy:

$$\mathcal{E}^\gamma : x \mapsto \sum_{i=1}^{N} \mathbf{L}_C^\gamma(x + \mathbf{R}h_i, h_i + \mathbf{R}x),$$

which is **differentiable** with gradient

$$\nabla = \gamma \left(\sum_{i=1}^{N} \log(u_i^\gamma) + \mathbf{R}^T \log(v_i^\gamma) \right).$$

Advantages:

- Convex formulation: minimize with gradient descent. Gives better estimations in practice.
- GPU + Parallelism: drastically outperform previous algorithm (B-Munkres) developed in [1] on large scales.

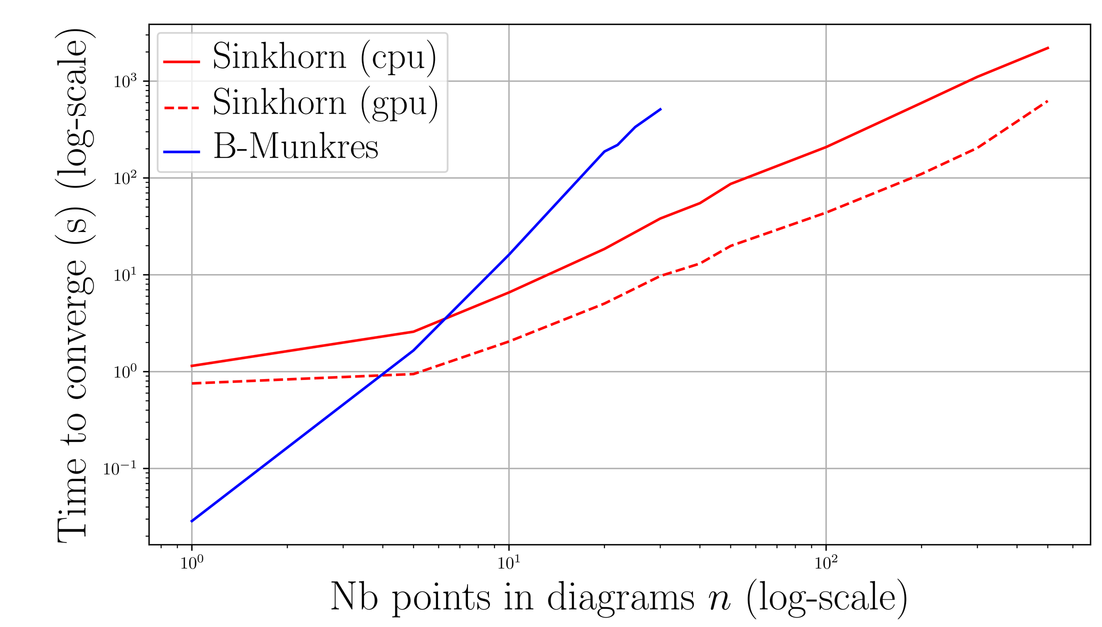


Figure 5: Running times of our algorithm (Sinkhorn, red) and algorithm described in [1] (B-Munkres, blue). Log-log scale.

Application: k-means clustering on thousands of PDs:

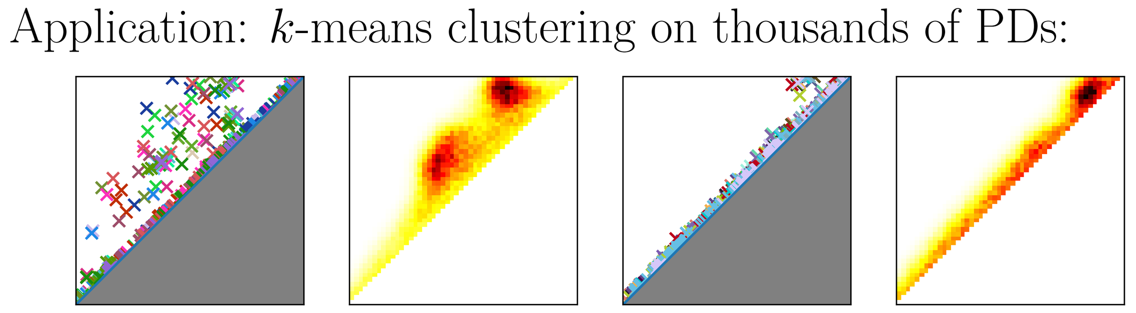


Figure 6: k-means on a real life dataset of 5000 persistence diagrams. Two identified clusters and their centroids.

References

[1] Katharine Turner et al.
Fréchet means for distributions of persistence diagrams.
Discrete & Computational Geometry, 52(1):44–70, 2014.

[2] Marco Cuturi.
Sinkhorn distances: Lightspeed computation of optimal transport.
In *Advances in Neural Information Processing Systems*, pages 2292–2300, 2013.

[3] Solomon et al.
Convolutional Wasserstein distances: Efficient optimal transportation on geometric domains.
In *ACM Transactions on Graphics (TOG)*, 2015.

Inria — inventeurs du monde numérique · ENSAE ParisTech · Google AI