Optimal Partial Transport for Topological Data Analysis

Saint-Flour
10 juillet 2019

Théo Lacombe

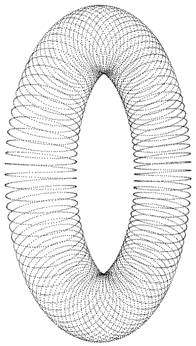
DataShape - Inria Saclay

theo.lacombe@inria.fr

The TDA pipeline: persistent homology

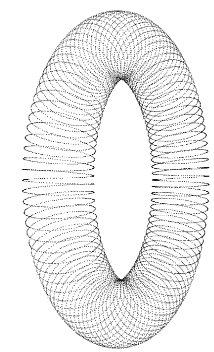
A (very) concise summary:

Input: some (complex) object



The TDA pipeline: persistent homology

A (very) concise summary:



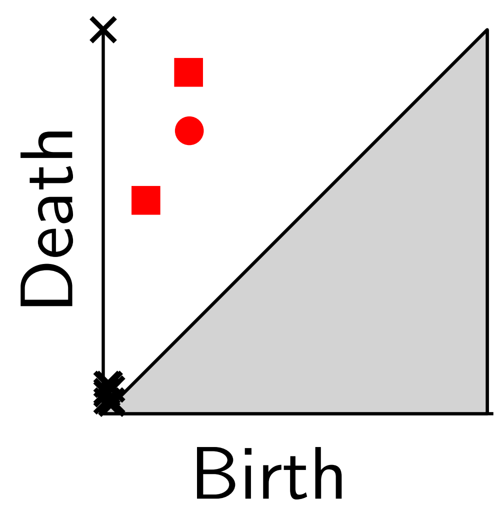
Input: some (complex) object

$\downarrow$ filtration f

Sequence of increasing topo. spaces

$\downarrow$ persistent homology theory

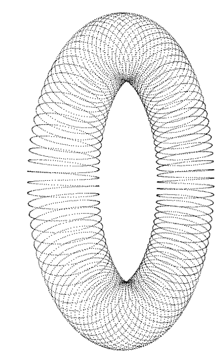
Persistence diagram



- Point cloud on the plane
  $\{{\color{red}x_1 \ldots x_i \ldots}\}$

The TDA pipeline: persistent homology

A (very) concise summary:



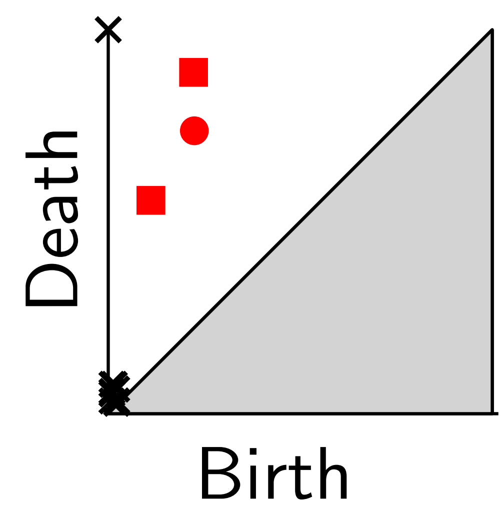
Input: some (complex) object

$\downarrow$ filtration f

Sequence of increasing topo. spaces

$\downarrow$ persistent homology theory

Persistence diagram



- Radon measures (locally finite Borel measures)
  $\{x_1 \ldots x_i \ldots\} \leftrightarrow \sum_i \delta_{x_i}$

The TDA pipeline: persistent homology

A (very) concise summary:

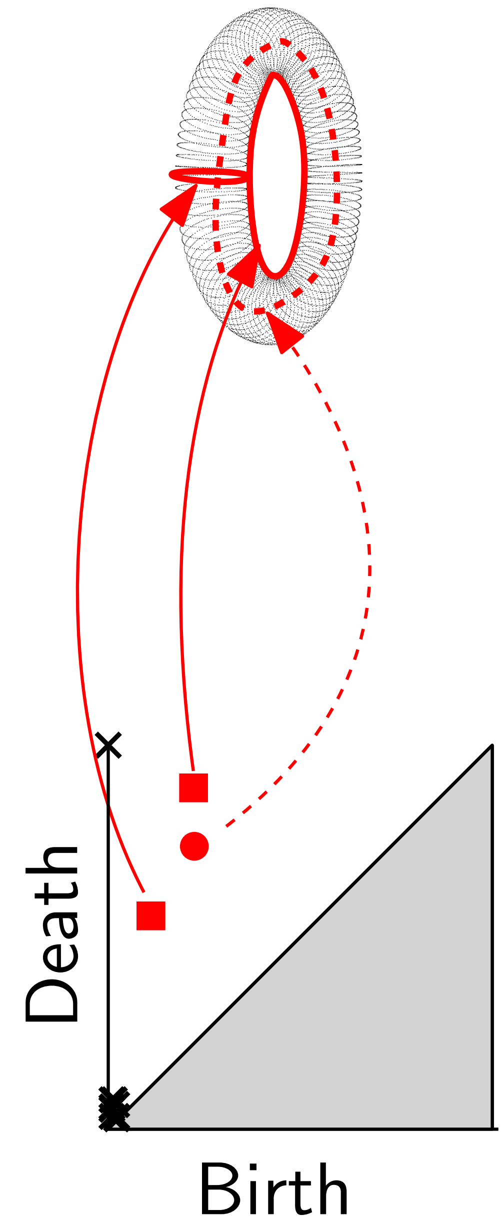
Input: some (complex) object

↓ filtration f

Sequence of increasing topo. spaces

↓ persistent homology theory

Persistence diagram



- Radon measures (locally finite Borel measures)
  $\{x_1 \ldots x_i \ldots\} \leftrightarrow \sum_i \delta_{x_i}$

The space of persistence diagrams $(\mathcal{D}^p, d_p)$

Persistence diagrams are comparable

- Using matching-like metrics

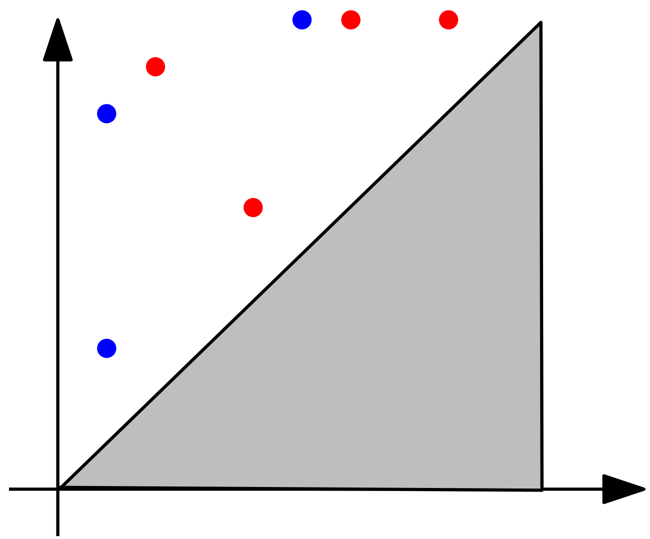
$$d_p(a,b) = \left(\inf_{\zeta} \sum_{(x,y)\in\zeta} \|x-y\|^p + \sum_{u\notin\zeta} \|u - s(u)\|^p \right)^{1/p}$$

partial matching

matched points

unmatched points

proj on diag



The space of persistence diagrams $(\mathcal{D}^p, d_p)$

Persistence diagrams are comparable

- Using matching-like metrics

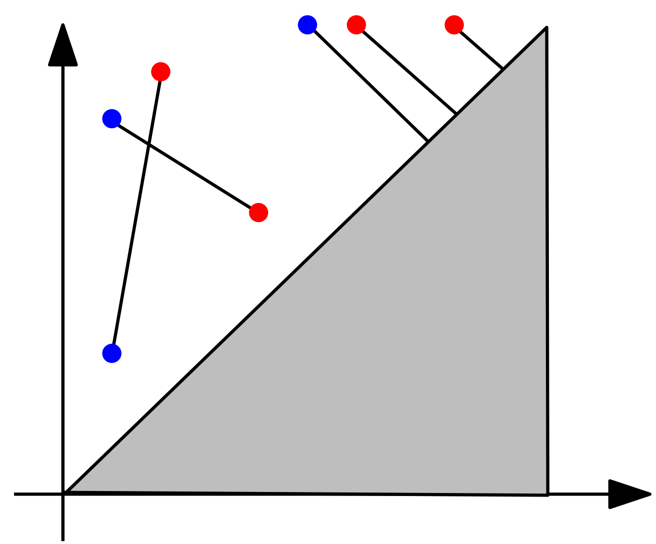
$$d_p(a,b) = \left(\inf_{\zeta} \sum_{(x,y)\in\zeta} \|x - y\|^p + \sum_{u \notin \zeta} \|u - s(u)\|^p \right)^{1/p}$$

partial matching

matched points

unmatched points

proj on diag



The space of persistence diagrams $(\mathcal{D}^p, d_p)$

Persistence diagrams are comparable

- Using matching-like metrics

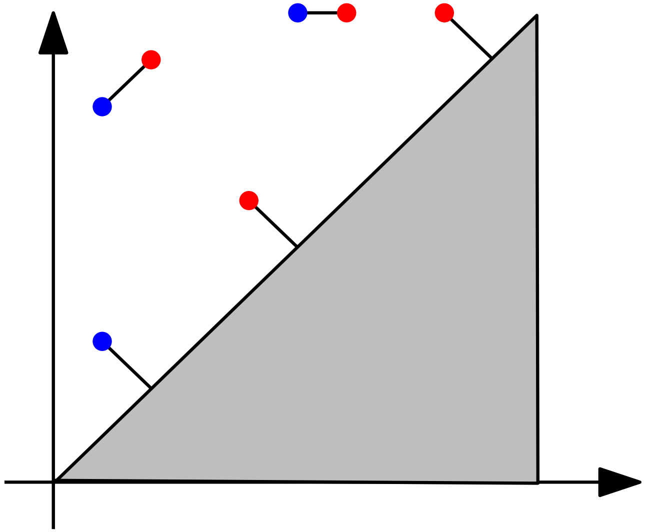
$$d_p(a,b) = \left(\inf_{\zeta} \sum_{(x,y)\in\zeta} \|x-y\|^p + \sum_{u\notin\zeta} \|u-s(u)\|^p \right)^{1/p}$$

partial matching

matched points

unmatched points

proj on diag



The space of persistence diagrams $(\mathcal{D}^p, d_p)$

Persistence diagrams are comparable

- Using matching-like metrics

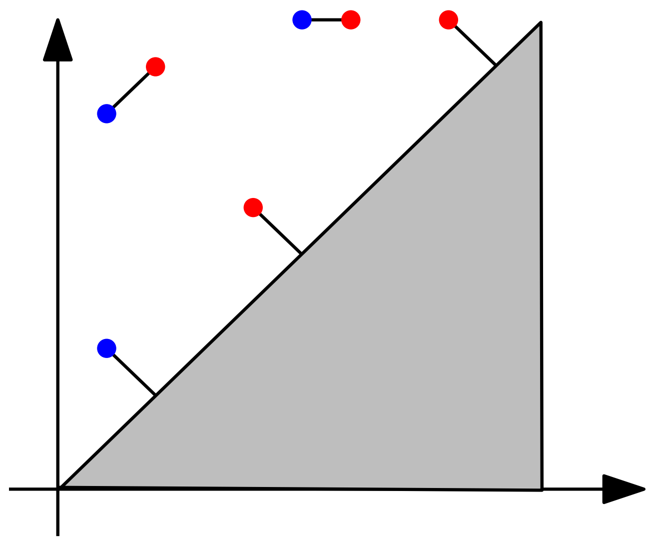
$$d_p(a, b) = \left(\inf_{\zeta} \sum_{(x,y)\in\zeta} \|x - y\|^p + \sum_{u \notin \zeta} \|u - s(u)\|^p \right)^{1/p}$$

partial matching

unmatched points

matched points

proj on diag



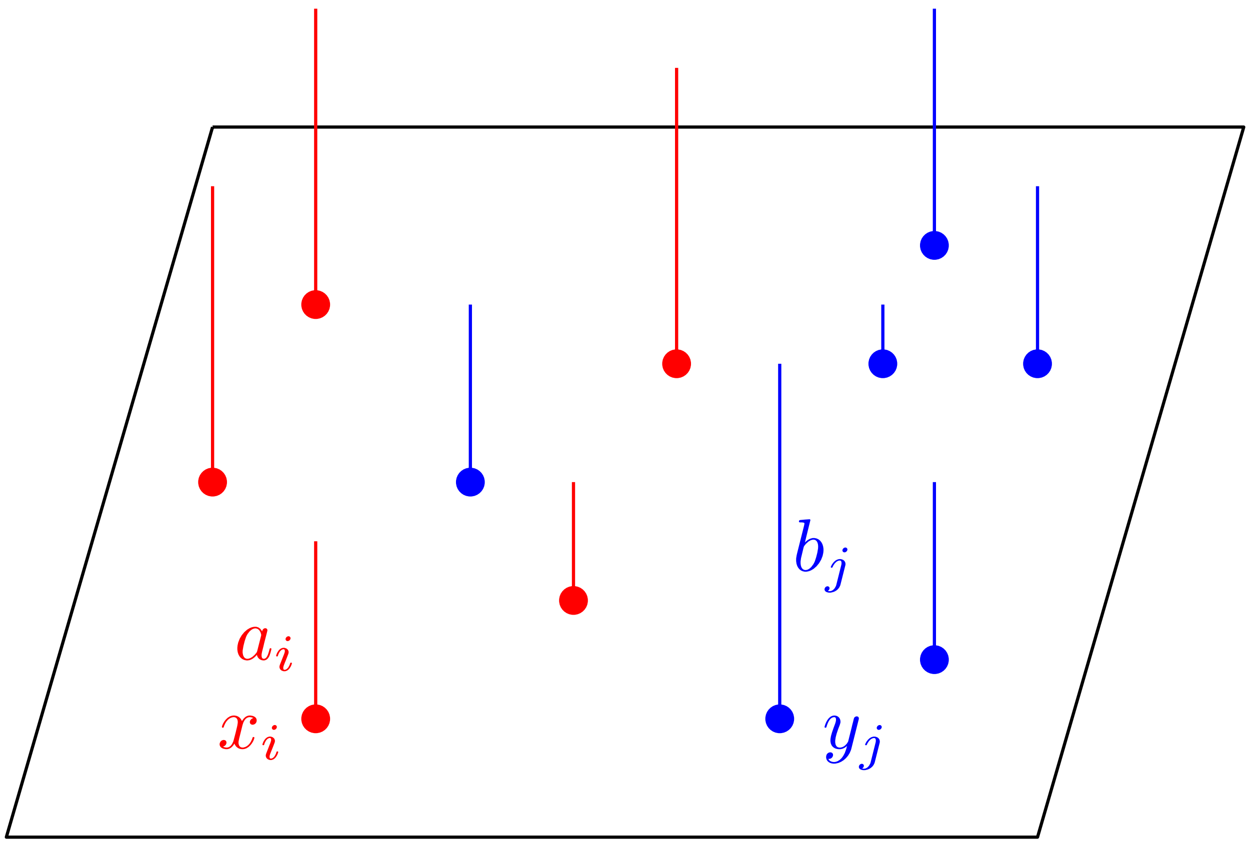
- Theoretically motivated
- Stable wrt input data

Assume also
$d_p(a, \emptyset) < \infty$

Optimal Transport - Generalities

Discrete formulation : μ and ν two probability measures

$$\mu = \sum_i a_i \delta_{x_i} \qquad \nu = \sum_j b_j \delta_{y_j}$$

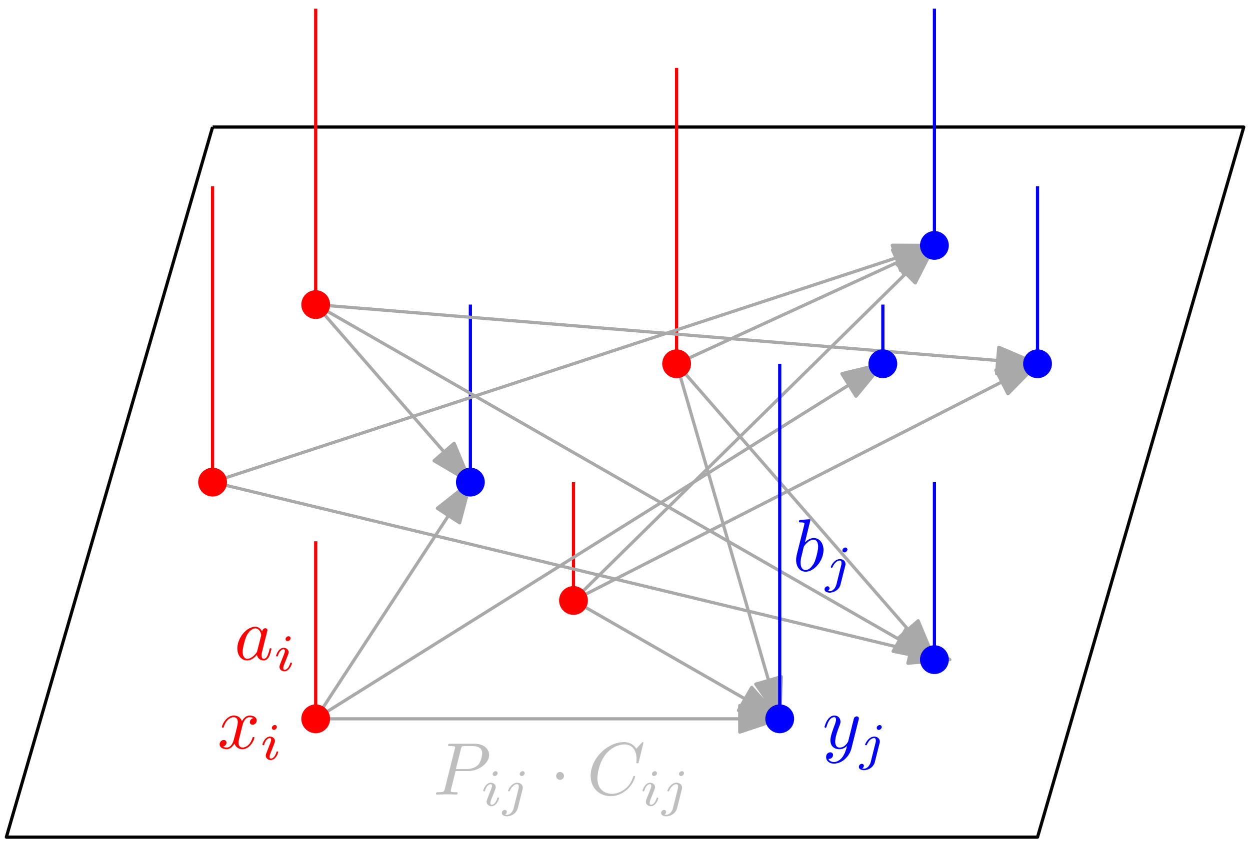


Optimal Transport - Generalities

Discrete formulation : μ and ν two probability measures

$$\mu = \sum_i a_i \delta_{x_i} \qquad \nu = \sum_j b_j \delta_{y_j}$$

(P_{ij}) transport plan between μ and ν



$$\sum_j P_{ij} = a_i$$

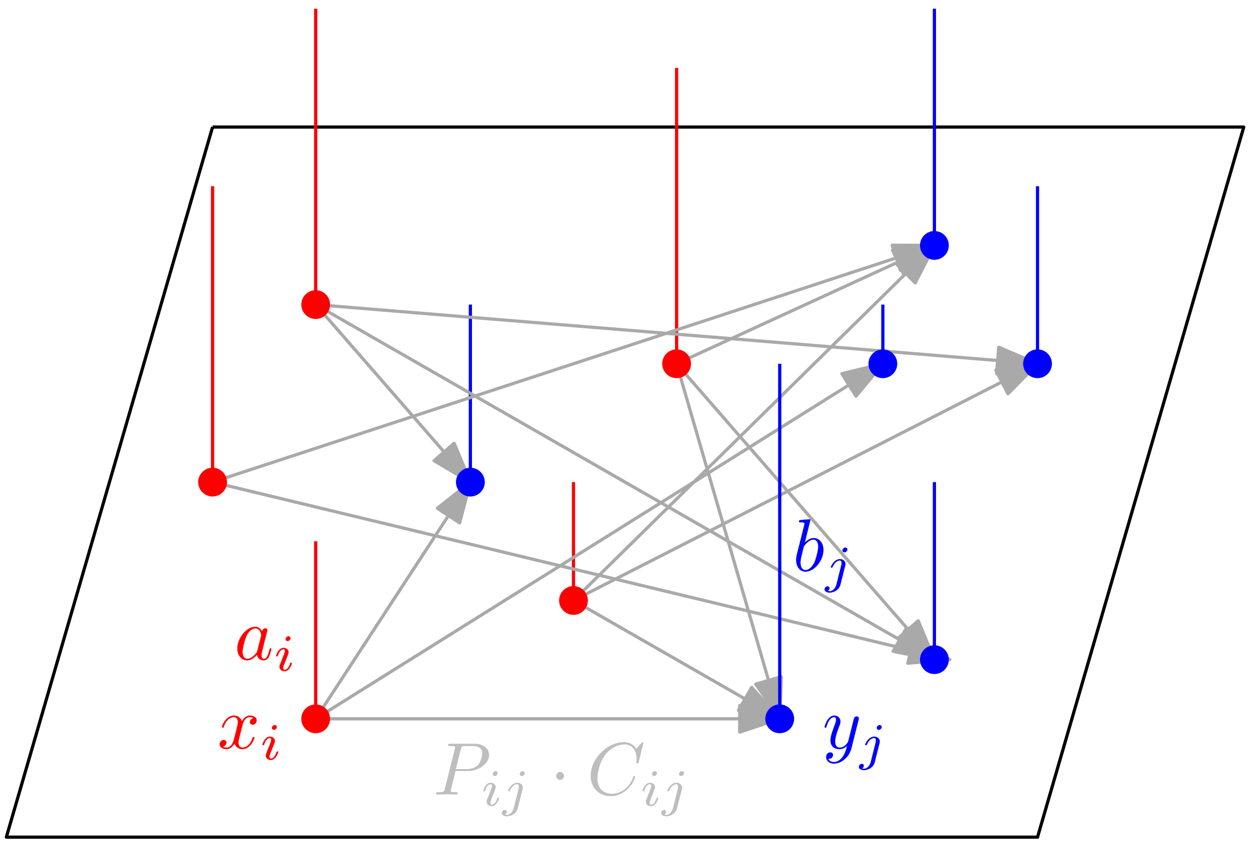
$$\sum_i P_{ij} = b_j$$

Optimal Transport - Generalities

Discrete formulation : μ and ν two probability measures

$$\mu = \sum_i a_i \delta_{x_i} \qquad \nu = \sum_j b_j \delta_{y_j}$$

$$\langle P, C \rangle = \sum_{ij} P_{ij} C_{ij}$$



where

$$C_{ij} = d(x_i, y_j)^p$$

subject to

$$\sum_j P_{ij} = a_i$$

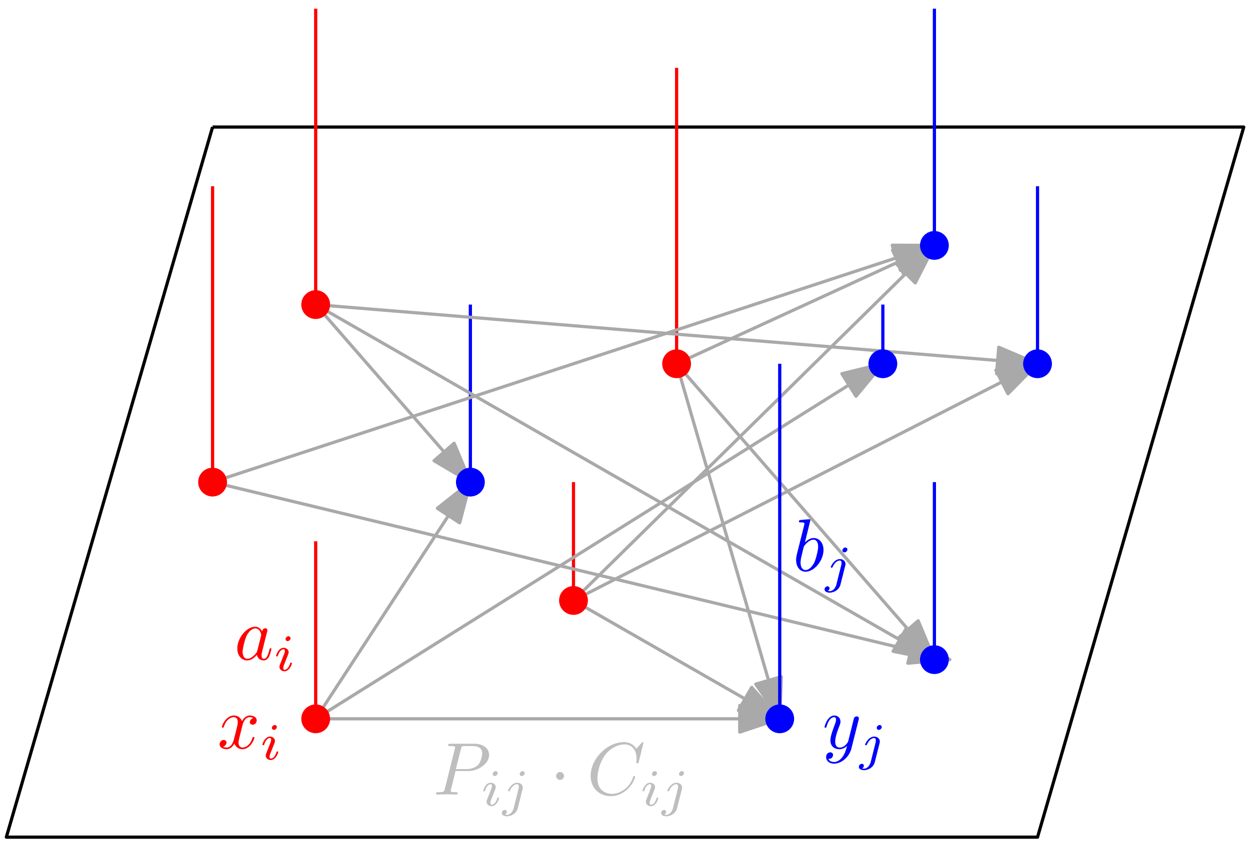
$$\sum_i P_{ij} = b_j$$

Optimal Transport - Generalities

Discrete formulation : μ and ν two probability measures

$$\mu = \sum_i a_i \delta_{x_i} \qquad \nu = \sum_j b_j \delta_{y_j}$$

$$W_p(\mu,\nu)^p = \inf_P \langle P, C \rangle = \inf_P \sum_{ij} P_{ij} C_{ij}$$



where
$$C_{ij} = d(x_i, y_j)^p$$

subject to

$$\sum_j P_{ij} = a_i$$
$$\sum_i P_{ij} = b_j$$

Optimal Transport - Generalities

General formulation

Consider μ, ν two probability measures on a Polish metric space $(\mathcal{X}, d)$

$\pi \in \mathcal{M}(\mathcal{X} \times \mathcal{X})$ is a *transport plan* between μ and ν if

$$\pi(A, \mathcal{X}) = \mu(A) \text{ and } \pi(\mathcal{X}, B) = \nu(B)$$

The cost of π is $C_p(\pi) := \iint_{\mathcal{X}\times\mathcal{X}} d(x, y)^p \mathrm{d}\pi(x, y)$

and the Wasserstein-p distance between μ and ν is

$$W_p(\mu, \nu) = \left(\inf_\pi C_p(\pi)\right)^{\frac{1}{p}}$$

Optimal Transport - Generalities

Properties:

- W_p is a distance over $\{\mu \in \mathcal{P}(\mathcal{X}) : \overbrace{\int_{\mathcal{X}} d(x, x_0)^p \mathrm{d}\mu(x)}^{W_p(\mu, \delta_{x_0})^p} < \infty\}$

- It metricizes the weak convergence and the p-th moment convergence.

$$W_p(\mu_n, \mu) \to 0 \Leftrightarrow \begin{cases} \mu_n \to \mu \text{ weakly} \\ W_p(\mu_n, \delta_{x_0}) \to W_p(\mu, \delta_{x_0}) \end{cases}$$

Optimal Transport - Generalities

Properties:

- W_p is a distance over $\{\mu \in \mathcal{P}(\mathcal{X}) : \overbrace{\int_{\mathcal{X}} d(x, x_0)^p \mathrm{d}\mu(x)}^{W_p(\mu,\delta_{x_0})^p} < \infty\}$

- It metrizes the weak convergence and the p-th moment convergence.

$$W_p(\mu_n, \mu) \to 0 \Leftrightarrow \begin{cases} \mu_n \to \mu \text{ weakly} \\ W_p(\mu_n, \delta_{x_0}) \to W_p(\mu, \delta_{x_0}) \end{cases}$$

Reminder:

- $\mu_n \to \mu$ *weakly* means:

for all f continuous, bounded, $\int_{\mathcal{X}} f(x)\mathrm{d}\mu_n(x) = \mu_n(f) \to \mu(f)$

- $\mu_n \to \mu$ *vaguely* means:

for all f continuous, compactly supported, $\mu_n(f) \to \mu(f)$

Optimal Transport - Generalities

Properties:

- W_p is a distance over $\{\mu \in \mathcal{P}(\mathcal{X}) : \overbrace{\int_{\mathcal{X}} d(x, x_0)^p \mathrm{d}\mu(x)}^{W_p(\mu,\delta_{x_0})^p} < \infty\}$

- It metricizes the weak convergence and the p-th moment convergence.

$$W_p(\mu_n, \mu) \to 0 \Leftrightarrow \begin{cases} \mu_n \to \mu \text{ weakly} \\ W_p(\mu_n, \delta_{x_0}) \to W_p(\mu, \delta_{x_0}) \end{cases}$$

- And many other nice properties:
  - Know about barycenters (Fréchet means).
  - Know the geodesics.
  - Many numerical tools (algorithms, libraries)...

Bridging TDA and OT

Persistence Diagrams $(\mathcal{D}^p, d_p)$	Optimal Transport $(\mathcal{W}^p, W_p)$

Bridging TDA and OT

Persistence Diagrams $(\mathcal{D}^p, d_p)$

- Discrete support (+integer mass)

Optimal Transport $(\mathcal{W}^p, W_p)$

- General support

Bridging TDA and OT

Persistence Diagrams $(\mathcal{D}^p, d_p)$

- Discrete support ($+$integer mass)
- Measures with different (potentially infinite) masses

Optimal Transport $(\mathcal{W}^p, W_p)$

- General support
- Measures same masses

Bridging TDA and OT

Persistence Diagrams $(\mathcal{D}^p, d_p)$

- Discrete support ($+$integer mass)
- Measures with different (potentially infinite) masses
- *Partial* matching distances (can match to the diagonal)

Optimal Transport $(\mathcal{W}^p, W_p)$

- General support
- Measures same masses
- Exact *transportation* distances

Bridging TDA and OT

Persistence Diagrams $(\mathcal{D}^p, d_p)$

- Discrete support (+integer mass)
- Measures with different (potentially infinite) masses
- *Partial* matching distances (can match to the diagonal)

Optimal Transport $(\mathcal{W}^p, W_p)$

- General support
- Measures same masses
- Exact *transportation* distances
- Well-studied theoretically
- Efficients algorithms/libraries

Optimal Partial Transport

Various ways to extend OT to measures with different masses

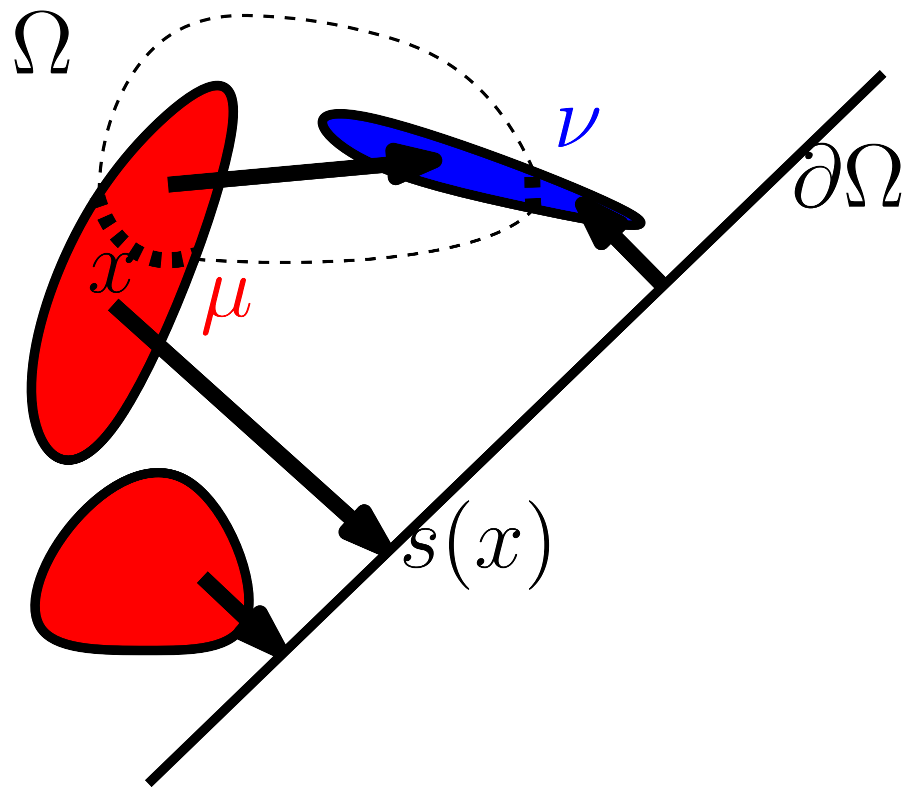
$\longrightarrow$ see [Chizat, 2017]

Optimal Partial Transport

Various ways to extend OT to measures with different masses

$\rightarrow$ see [Chizat, 2017]

A.Figalli and N.Gigli (2010)



Introduced to study heat equation with Dirichlet boundary conditions

Optimal Partial Transport

Core idea: Just consider sub-marginal constraints!

Let $\partial\Omega$ be the boundary of Ω, and $\overline{\Omega} = \Omega \cup \partial\Omega$

Given μ, ν two Radon measures on Ω, consider admissible transport plans

$$\pi \in \mathcal{M}(\overline{\Omega} \times \overline{\Omega}) \qquad \text{such that} \qquad \begin{aligned} \pi(A \times \overline{\Omega}) &= \mu(A) \qquad A \subset \Omega \\ \pi(\overline{\Omega} \times B) &= \nu(B) \qquad B \subset \Omega \end{aligned}$$

And then just define

$$C_p(\pi) = \iint_{\overline{\Omega} \times \overline{\Omega}} d(x,y)^p \mathrm{d}\pi(x,y)$$

$$\mathrm{OT}_p(\mu,\nu) = \left(\inf_{\pi \in \mathrm{Adm}(\mu,\nu)} C_p(\pi)\right)^{1/p}$$

Rem: measures must satisfy $\int_\Omega d(x, \partial\Omega)^p \mathrm{d}\mu(x) < +\infty$

Studying $(\mathcal{D}^p, d_p)$ using Optimal Partial Transport

Equivalence of formalisms [Divol, L, 2019]

If a, b are persistence diagrams, then $\mathrm{OT}_p(a, b) = d_p(a, b)$

Studying $(\mathcal{D}^p, d_p)$ using Optimal Partial Transport

Equivalence of formalisms [Divol, L, 2019]

If a, b are persistence diagrams, then $\mathrm{OT}_p(a, b) = d_p(a, b)$

Elements for the proof:

- Assume a and b are finite (masses n_1, n_2), let $n = n_1 + n_2$.

$\mathfrak{S}_n$: permutation matrices, $\mathcal{B}_n$: bi-stochastic matrices.

- We must show that $\inf_{P \in \mathcal{B}_n} \langle P, C \rangle = \inf_{A \in \mathfrak{S}_n} \langle A, C \rangle$

Studying $(\mathcal{D}^p, d_p)$ using Optimal Partial Transport

Equivalence of formalisms [Divol, L, 2019]
If a, b are persistence diagrams, then $\mathrm{OT}_p(a,b) = d_p(a,b)$

Elements for the proof:

- Assume a and b are finite (masses n_1, n_2), let $n = n_1 + n_2$.

$\mathfrak{S}_n$: permutation matrices, $\mathcal{B}_n$: bi-stochastic matrices.

- We must show that $\inf_{P \in \mathcal{B}_n} \langle P, C \rangle = \inf_{A \in \mathfrak{S}_n} \langle A, C \rangle$

$\leq$ (obvious)

Studying $(\mathcal{D}^p, d_p)$ using Optimal Partial Transport

Equivalence of formalisms [Divol, L, 2019]

If a, b are persistence diagrams, then $\mathrm{OT}_p(a,b) = d_p(a,b)$

Elements for the proof:

- Assume a and b are finite (masses n_1, n_2), let $n = n_1 + n_2$.

$\mathfrak{S}_n$: permutation matrices, $\mathcal{B}_n$: bi-stochastic matrices.

- We must show that $\inf_{P \in \mathcal{B}_n} \langle P, C \rangle = \inf_{A \in \mathfrak{S}_n} \langle A, C \rangle$

$\leq$ (obvious)

— [Birkhoff] The convex hull of $\mathfrak{S}_n$ is exactly $\mathcal{B}_n$

Studying $(\mathcal{D}^p, d_p)$ using Optimal Partial Transport

Equivalence of formalisms [Divol, L, 2019]

If a, b are persistence diagrams, then $\mathrm{OT}_p(a, b) = d_p(a, b)$

Elements for the proof:

- Assume a and b are finite (masses n_1, n_2), let $n = n_1 + n_2$.

$\mathfrak{S}_n$: permutation matrices, $\mathcal{B}_n$: bi-stochastic matrices.

- We must show that $\inf_{P \in \mathcal{B}_n} \langle P, C \rangle = \inf_{A \in \mathfrak{S}_n} \langle A, C \rangle$

  $\leq$ (obvious)

  — [Birkhoff] The convex hull of $\mathfrak{S}_n$ is exactly $\mathcal{B}_n$

  — For $P^* \in \mathcal{B}_n$ optimal, write $P^* = \sum_{i=1}^{n!} \lambda_i A_i$, with $\sum_i \lambda_i = 1$

Studying $(\mathcal{D}^p, d_p)$ using Optimal Partial Transport

Equivalence of formalisms [Divol, L, 2019]

If a, b are persistence diagrams, then $\mathrm{OT}_p(a,b) = d_p(a,b)$

Elements for the proof:

- Assume a and b are finite (masses n_1, n_2), let $n = n_1 + n_2$.

$\mathfrak{S}_n$: permutation matrices, $\mathcal{B}_n$: bi-stochastic matrices.

- We must show that $\inf_{P \in \mathcal{B}_n} \langle P, C \rangle = \inf_{A \in \mathfrak{S}_n} \langle A, C \rangle$

$\leq$ (obvious)

— [Birkhoff] The convex hull of $\mathfrak{S}_n$ is exactly $\mathcal{B}_n$

— For $P^* \in \mathcal{B}_n$ optimal, write $P^* = \sum_{i=1}^{n!} \lambda_i A_i$, with $\sum_i \lambda_i = 1$

— $\mathrm{OT}_p(a,b) = \langle P^*, C \rangle = \langle \sum_i \lambda_i A_i, C \rangle = \sum_i \lambda_i \langle A_i, C \rangle \geq d_p(a,b)$

Studying $(\mathcal{D}^p, d_p)$ using Optimal Partial Transport

Equivalence of formalisms [Divol, L, 2019]

If a, b are persistence diagrams, then $\mathrm{OT}_p(a,b) = d_p(a,b)$

Elements for the proof:

- Assume a and b are finite (masses n_1, n_2), let $n = n_1 + n_2$.

$\mathfrak{S}_n$: permutation matrices, $\mathcal{B}_n$: bi-stochastic matrices.

- We must show that $\inf_{P \in \mathcal{B}_n} \langle P, C \rangle = \inf_{A \in \mathfrak{S}_n} \langle A, C \rangle$

  $\leq$ (obvious)

  — [Birkhoff] The convex hull of $\mathfrak{S}_n$ is exactly $\mathcal{B}_n$

  — For $P^* \in \mathcal{B}_n$ optimal, write $P^* = \sum_{i=1}^{n!} \lambda_i A_i$, with $\sum_i \lambda_i = 1$

  — $\mathrm{OT}_p(a,b) = \langle P^*, C \rangle = \langle \sum_i \lambda_i A_i, C \rangle = \sum_i \lambda_i \langle A_i, C \rangle \geq d_p(a,b)$

- Conclude using some density argument.

Studying $(\mathcal{D}^p, d_p)$ using Optimal Partial Transport

Equivalence of formalisms [Divol, L, 2019]
If a, b are persistence diagrams, then $\mathrm{OT}_p(a,b) = d_p(a,b)$

Elements for the proof:

- Assume a and b are finite (masses n_1, n_2), let $n = n_1 + n_2$.

$\mathfrak{S}_n$: permutation matrices, $\mathcal{B}_n$: bi-stochastic matrices.

- We must show that $\inf_{P \in \mathcal{B}_n} \langle P, C \rangle = \inf_{A \in \mathfrak{S}_n} \langle A, C \rangle$

  $\leq$ (obvious)

  — [Birkhoff] The convex hull of $\mathfrak{S}_n$ is exactly $\mathcal{B}_n$
  — For $P^* \in \mathcal{B}_n$ optimal, write $P^* = \sum_{i=1}^{n!} \lambda_i A_i$, with $\sum_i \lambda_i = 1$
  — $\mathrm{OT}_p(a,b) = \langle P^*, C \rangle = \langle \sum_i \lambda_i A_i, C \rangle = \sum_i \lambda_i \langle A_i, C \rangle \geq d_p(a,b)$

- Conclude using some density argument.

Question: Is there a direct proof in infinite dimension?

Studying $(\mathcal{D}^p, d_p)$ using Optimal Partial Transport

Equivalence of formalisms [Divol, L, 2019]

If a, b are persistence diagrams, then $\mathrm{OT}_p(a, b) = d_p(a, b)$

Convergence in $(\mathcal{D}^p, d_p)$

- We have $d_p(a_n, a) \to 0 \Leftrightarrow \begin{cases} a_n \to a \text{ vaguely} \\ d_p(a_n, \emptyset) \to d_p(a, \emptyset) \end{cases}$

Barycenters of persistence diagrams

$\mathcal{D}^p$ is a Polish space [Mukherjee et al. 2011]

Barycenters of persistence diagrams

$\mathcal{D}^p$ is a Polish space [Mukherjee et al. 2011]

Fréchet means (aka barycenters):

Consider $b_1 \dots b_N$ a set of diagrams

Estimating their Fréchet mean consists in computing

$$\operatorname{argmin} \left\{ \mathcal{E}(a) = \frac{1}{N} \sum_{i=1}^{N} d_2(a, b_i)^2, \ a \text{ persistence diagram} \right\}.$$

Barycenters of persistence diagrams

$\mathcal{D}^p$ is a Polish space [Mukherjee et al. 2011]

Fréchet means (aka barycenters):

Consider $b_1 \dots b_N$ a set of diagrams

Estimating their Fréchet mean consists in computing

$$\operatorname{argmin} \left\{ \mathcal{E}(a) = \frac{1}{N} \sum_{i=1}^{N} d_2(a, b_i)^2, \ a \text{ persistence diagram} \right\}.$$

First results [Turner et al. 2013]:

- $\mathcal{E}$ is not convex. It admits global (and local) minimizers
- Local minimizers can be computed (expensive)

Barycenters of persistence diagrams

$\mathcal{D}^p$ is a Polish space [Mukherjee et al. 2011]

Fréchet means (aka barycenters):

Consider $b_1 \dots b_N$ a set of diagrams

Estimating their Fréchet mean consists in computing

$$\operatorname{argmin} \left\{ \mathcal{E}(a) = \frac{1}{N} \sum_{i=1}^{N} \begin{array}{l} \cancel{d_2(a, b_i)^2, \ a \text{ persistence diagram}} \\ \mathrm{OT}_2(a, b)^2, a \text{ any Radon measure} \end{array} \right\}.$$

Barycenters of persistence diagrams

$\mathcal{D}^p$ is a Polish space [Mukherjee et al. 2011]

Fréchet means (aka barycenters):

Consider $b_1 \dots b_N$ a set of diagrams

Estimating their Fréchet mean consists in computing

$$\operatorname{argmin} \left\{ \mathcal{E}(a) = \frac{1}{N} \sum_{i=1}^{N} \begin{matrix} \cancel{d_2(a, b_i)^2}, \ \cancel{a \text{ persistence diagram}} \\ \mathrm{OT}_2(a, b)^2, a \text{ any Radon measure} \end{matrix} \right\}.$$

Properties [Divol, L, 2019]

- $\mathcal{E}$ is now convex, admits global minimizers.
- Some of them are actual diagrams.

Barycenters of persistence diagrams

$\mathcal{D}^p$ is a Polish space [Mukherjee et al. 2011]

Fréchet means (aka barycenters):

Consider $b_1 \dots b_N$ a set of diagrams

Estimating their Fréchet mean consists in computing

$$\operatorname{argmin} \left\{ \mathcal{E}(a) = \frac{1}{N} \sum_{i=1}^{N} \begin{matrix} \cancel{d_2(a, b_i)^2}, \ \cancel{a \text{ persistence diagram}} \\ \mathrm{OT}_2(a, b)^2, a \text{ any Radon measure} \end{matrix} \right\}.$$

Properties [Divol, L, 2019]

- $\mathcal{E}$ is now convex, admits global minimizers.
- Some of them are actual diagrams.

Numerical considerations [L, Cuturi, Oudot, 2018]

- These can be approximated efficiently (Sinkhorn algorithm).

Conclusion

Take home messages:

- Optimal transport can be extended (in various ways) to handle measures with different masses
- Figalli&Gigli's formulation (2010) of Optimal Partial Transport is a prowerful tool to study TDA objects.

  (and Radon measures in general)