Ph.D. defense

Statistical tools for topological descriptors using optimal transport

September 8th, 2020

Théo Lacombe

Under the supervision of Steve Oudot and Marco Cuturi.

École polytechnique

DataShape - Inria Saclay

theo.lacombe@inria.fr

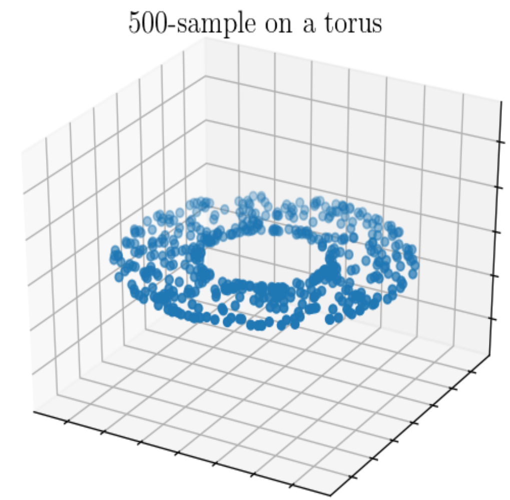
TDA: from shapes to persistence diagrams

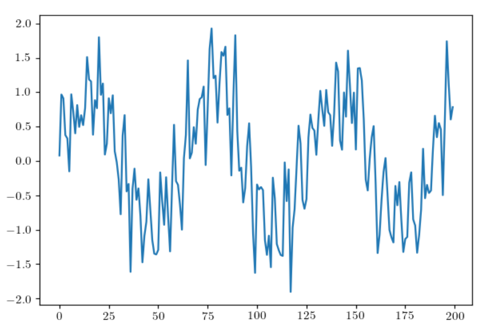
A (very) concise summary:

Input: some complex object

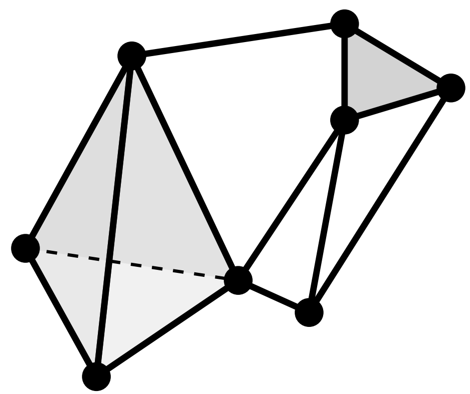
↓

Topological information

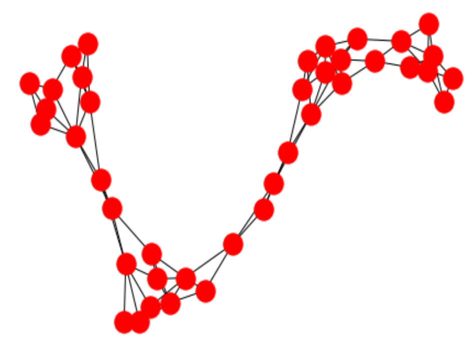




time series



Simplicial complex



Graph

TDA: from shapes to persistence diagrams

A (very) concise summary:

Input: some complex object

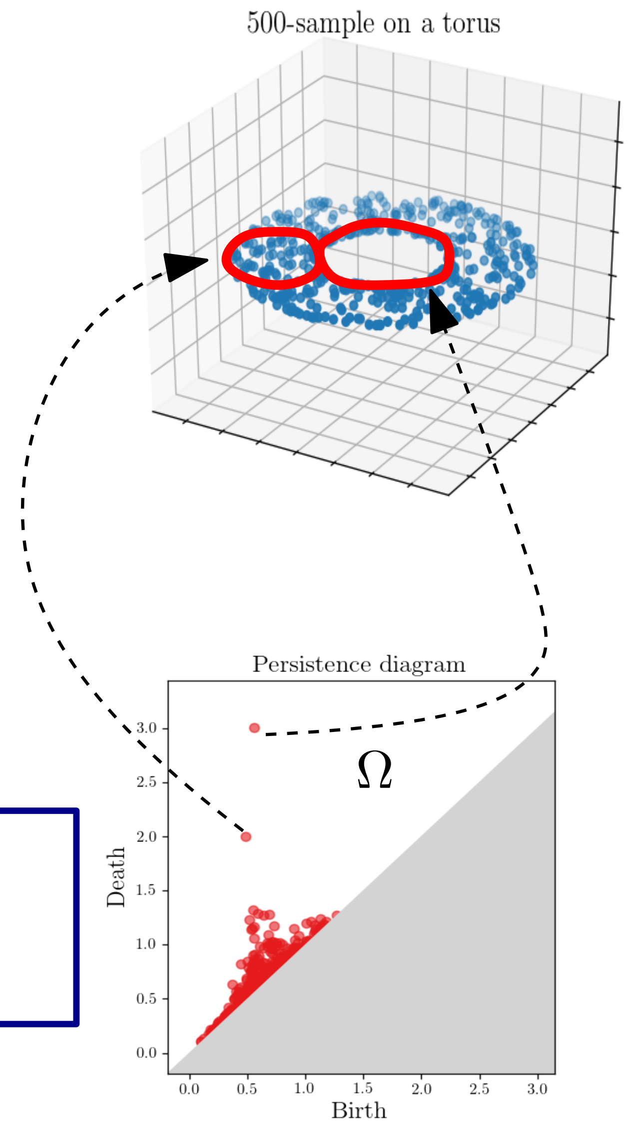
Persistent homology theory [EH10, Oud15]

Output: persistence diagram

- Multiset of points

$\{x_1, x_2, \ldots, x_i, \ldots\}$

$$x_i \in \Omega := \{(t_1, t_2) \in \mathbb{R}^2,\ t_2 > t_1\}.$$



[EH10]: *An Introduction to Computational Topology*, Edelsbrunner and Harer, 2010.
[Oud15]: *Persistence theory: from quiver representation to data analysis*, Oudot, 2015.

TDA: from shapes to persistence diagrams

A (very) concise summary:

Input: some complex object

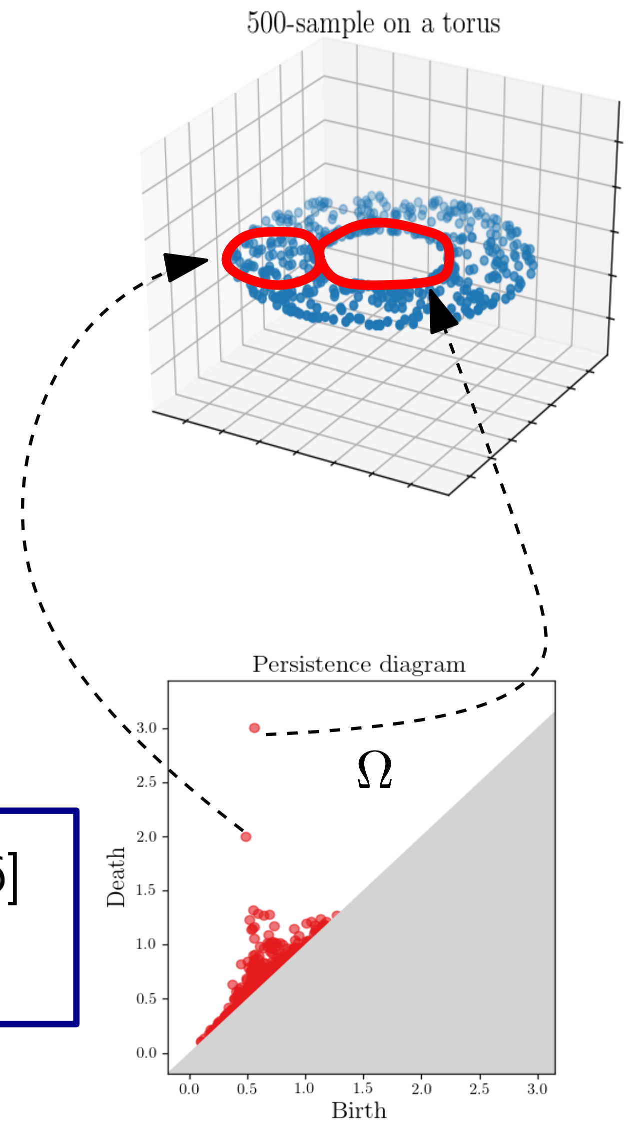
Persistent homology theory [EH10, Oud15]

Output: persistence diagram

- Multiset of points $\leftrightarrow$ (Radon) Point measures [CSGO16]

$\{x_1, x_2, \ldots, x_i, \ldots\} \leftrightarrow \sum_i \delta_{x_i}$

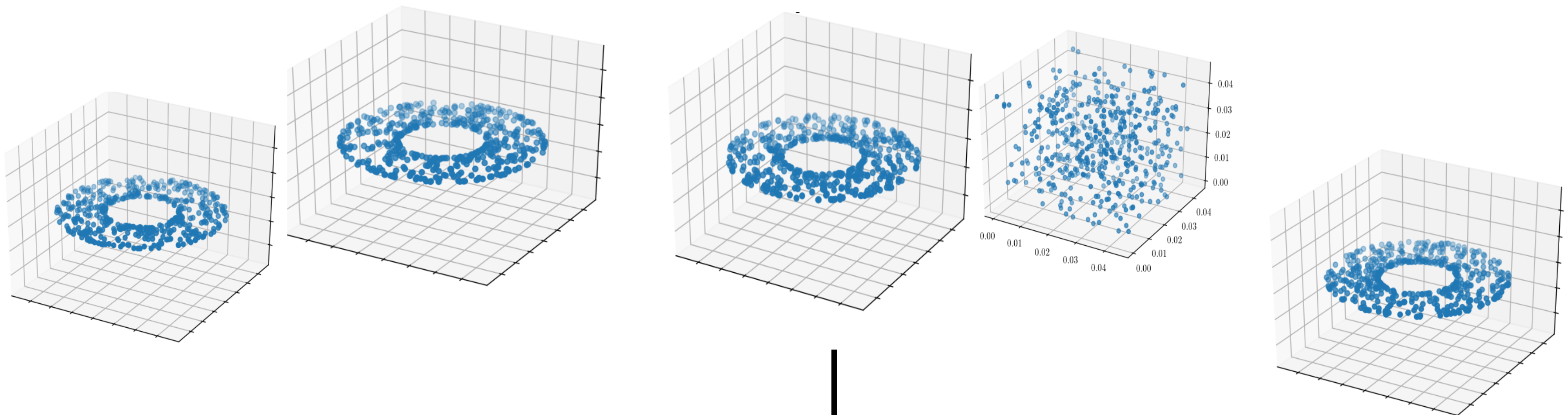
$$x_i \in \Omega := \{(t_1, t_2) \in \mathbb{R}^2,\ t_2 > t_1\}.$$



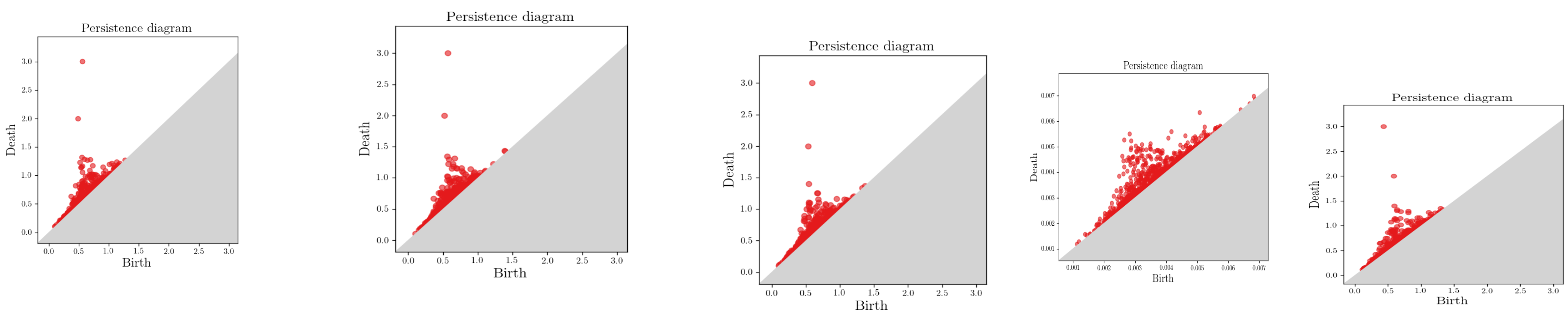
[EH10]: *An Introduction to Computational Topology*, Edelsbrunner and Harer, 2010.
[Oud15]: *Persistence theory: from quiver representation to data analysis*, Oudot, 2015.
[CSGO16]: *The structure and Stability of Persistence Modules*, Chazal, Da Silva, Glisse, Oudot, 2016.

TDA: from shapes to persistence diagrams

- Collection of complex objects



- Collection of persistence diagrams

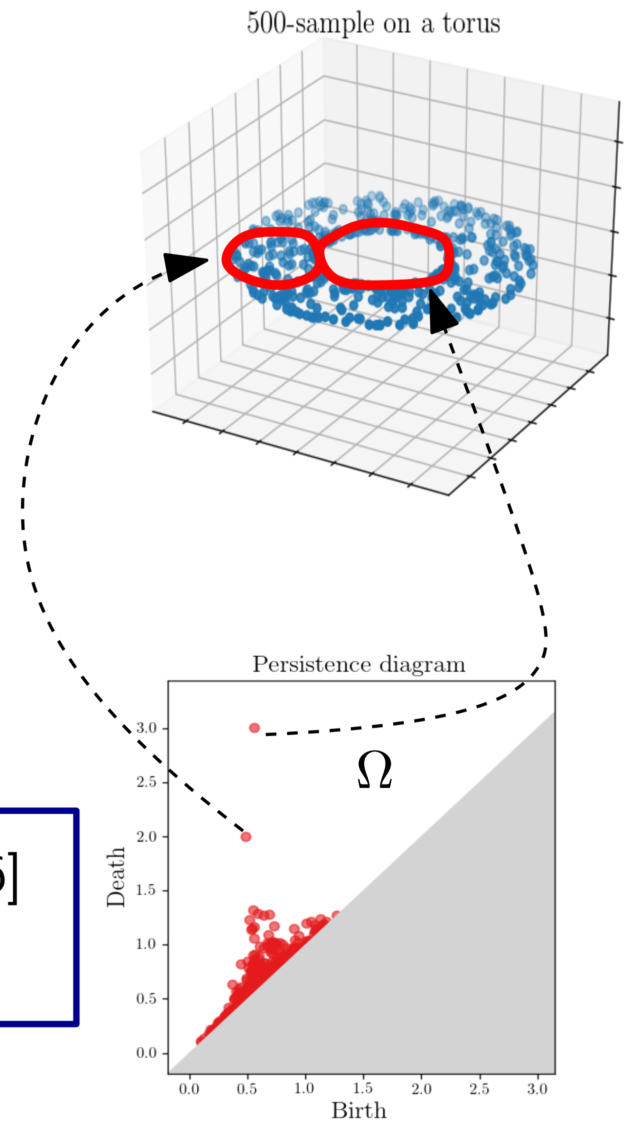


$\Rightarrow$ Understanding and making use of the metric and statistical properties of the space of persistence diagrams.

TDA: from shapes to persistence diagrams

Disclaimer:

- We do not consider points with $+\infty$ coordinates.



Output: persistence diagram

- Multiset of points $\leftrightarrow$ (Radon) Point measures [CSGO16]

$\{x_1, x_2, \ldots, x_i, \ldots\} \leftrightarrow \sum_i \delta_{x_i}$

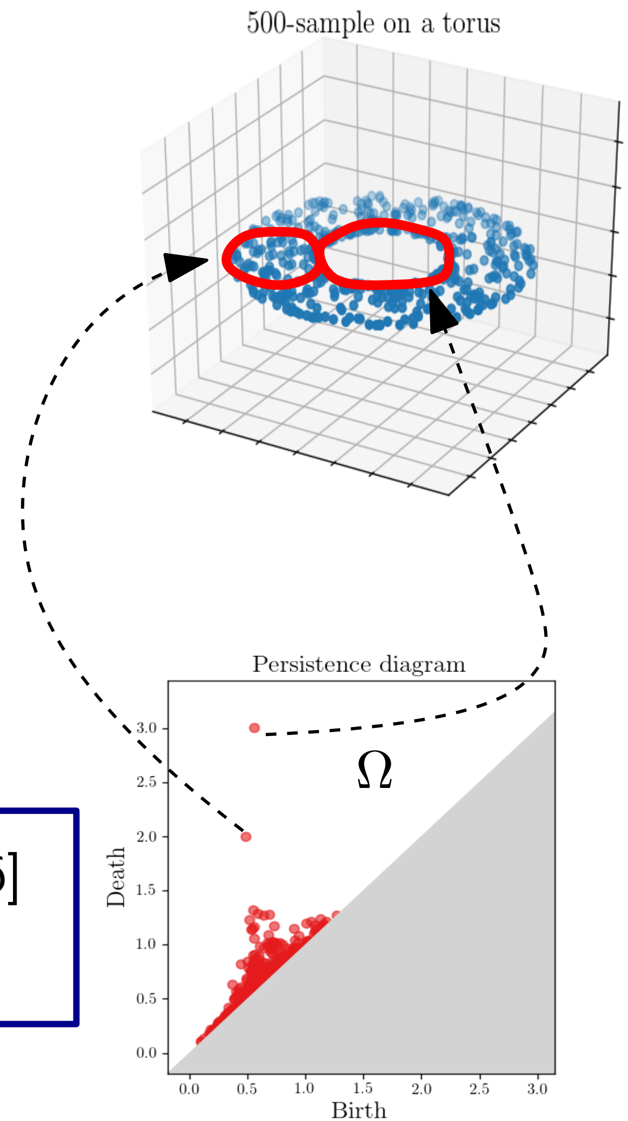
$x_i \in \Omega := \{(t_1, t_2) \in \mathbb{R}^2,\ t_2 > t_1\}$.

[EH10]: *An Introduction to Computational Topology*, Edelsbrunner and Harer, 2010.
[Oud15]: *Persistence theory: from quiver representation to data analysis*, Oudot, 2015.
[CSGO16]: *The structure and Stability of Persistence Modules*, Chazal, Da Silva, Glisse, Oudot, 2016.

TDA: from shapes to persistence diagrams

Disclaimer:

- We do not consider points with $+\infty$ coordinates.
- We allow for PDs with infinitely many points
  (but locally finite).
  $\rightarrow$ Retrieve a complete space [MMH11]



Output: persistence diagram

- Multiset of points $\leftrightarrow$ (Radon) Point measures [CSGO16]

$\{x_1, x_2, \ldots, x_i, \ldots\} \leftrightarrow \sum_i \delta_{x_i}$

$$x_i \in \Omega := \{(t_1, t_2) \in \mathbb{R}^2,\ t_2 > t_1\}.$$

[EH10]: *An Introduction to Computational Topology*, Edelsbrunner and Harer, 2010.
[Oud15]: *Persistence theory: from quiver representation to data analysis*, Oudot, 2015.
[CSGO16]: *The structure and Stability of Persistence Modules*, Chazal, Da Silva, Glisse, Oudot, 2016.
[MMH11]: *Probability measures on the space of persistence diagrams*, Mileyko, Mukherjee, Harer, 2011.

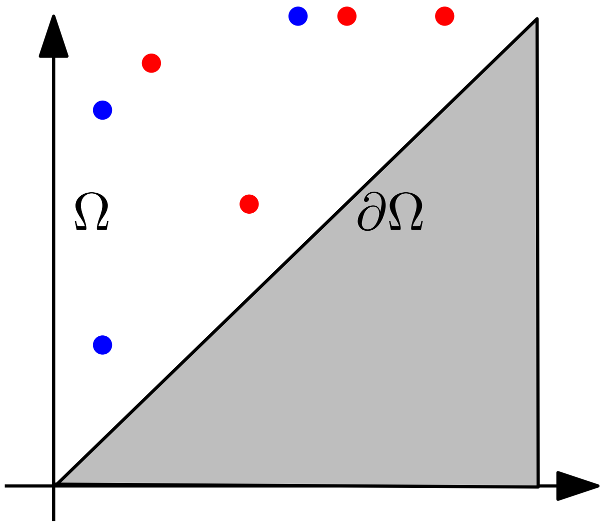
The space of persistence diagrams $(\mathcal{D}^p, d_p)$

Definition: standard metrics between diagrams [EH10,Ch.VIII]

- For a, b two PDs (viewed as multisets), define the p-th diagram distance:

$$d_p(a,b) = \left(\inf_{\zeta} \sum_{(x \in M)} \|x - \zeta(x)\|^p + \sum_{s \in (a \setminus M) \cup (b \setminus \zeta(M))} \|s - \mathrm{proj}_{\partial\Omega}(s)\|^p \right)^{1/p}$$

where $\zeta : M \subset a \xrightarrow{\mathrm{bij}} N \subset b$ (partial matching), and $\mathrm{proj}_{\partial\Omega} : \Omega \to \partial\Omega$ is the projection on $\partial\Omega$. When $p = \infty$, d_∞ is called the bottleneck distance.



[EH10]: *An Introduction to Computational Topology*, Edelsbrunner and Harer, 2010.

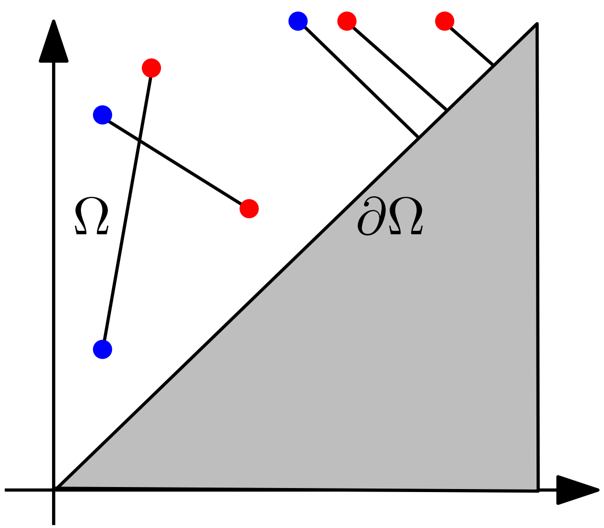
The space of persistence diagrams $(\mathcal{D}^p, d_p)$

Definition: standard metrics between diagrams [EH10,Ch.VIII]

- For a, b two PDs (viewed as multisets), define the p-th diagram distance:

$$d_p(a,b) = \left(\inf_{\zeta} \sum_{(x \in M)} \|x - \zeta(x)\|^p + \sum_{s \in (a \setminus M) \cup (b \setminus \zeta(M))} \|s - \mathrm{proj}_{\partial\Omega}(s)\|^p \right)^{1/p}$$

where $\zeta : M \subset a \xrightarrow{\text{bij}} N \subset b$ (partial matching), and $\mathrm{proj}_{\partial\Omega} : \Omega \to \partial\Omega$ is the projection on $\partial\Omega$. When $p = \infty$, d_∞ is called the bottleneck distance.



[EH10]: *An Introduction to Computational Topology*, Edelsbrunner and Harer, 2010.

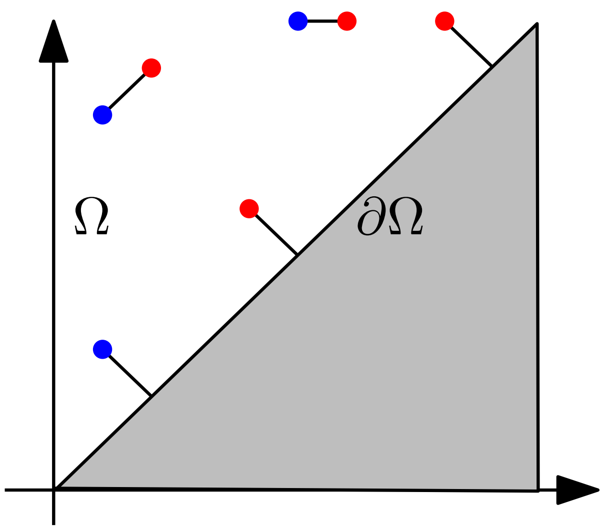
The space of persistence diagrams $(\mathcal{D}^p, d_p)$

Definition: standard metrics between diagrams [EH10,Ch.VIII]

- For a, b two PDs (viewed as multisets), define the p-th diagram distance:

$$d_p(a,b) = \left(\inf_{\zeta} \sum_{(x \in M)} \|x - \zeta(x)\|^p + \sum_{s \in (a \setminus M) \cup (b \setminus \zeta(M))} \|s - \mathrm{proj}_{\partial\Omega}(s)\|^p \right)^{1/p}$$

where $\zeta : M \subset a \xrightarrow{\text{bij}} N \subset b$ (partial matching), and $\mathrm{proj}_{\partial\Omega} : \Omega \to \partial\Omega$ is the projection on $\partial\Omega$. When $p = \infty$, d_∞ is called the bottleneck distance.



[EH10]: *An Introduction to Computational Topology*, Edelsbrunner and Harer, 2010.

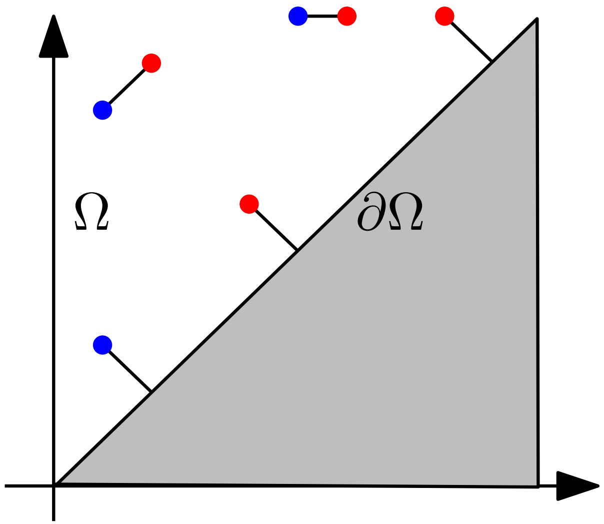
The space of persistence diagrams $(\mathcal{D}^p, d_p)$

Definition: standard metrics between diagrams [EH10,Ch.VIII]

- For a, b two PDs (viewed as multisets), define the p-th diagram distance:

$$d_p(a,b) = \left(\inf_{\zeta} \sum_{(x \in M)} \|x - \zeta(x)\|^p + \sum_{s \in (a \setminus M) \cup (b \setminus \zeta(M))} \|s - \mathrm{proj}_{\partial\Omega}(s)\|^p \right)^{1/p}$$

where $\zeta : M \subset a \xrightarrow{\text{bij}} N \subset b$ (partial matching), and $\mathrm{proj}_{\partial\Omega} : \Omega \to \partial\Omega$ is the projection on $\partial\Omega$. When $p = \infty$, d_∞ is called the bottleneck distance.



Definition [MMH11]

The space of PDs $(\mathcal{D}^p, d_p)$ is the space of loc. finite point measures $\mu = \sum_i \delta_{x_i}$ on Ω with finite total persistence:

$$\mathrm{Pers}_p(\mu) := \sum_i \|x_i - \mathrm{proj}_{\partial\Omega}(x_i)\|^p < \infty$$

[EH10]: *An Introduction to Computational Topology*, Edelsbrunner and Harer, 2010.
[MMH11]: *Probability measures on the space of persistence diagrams*, Mileyko, Mukherjee, Harer, 2011.

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

A concise history of diagram metrics

- (2005) Introduction of the bottleneck distance d_∞ and proof of a stability theorem [CSEH05].
- (2007) Introduction of d_p, with $p < \infty$ [CSE+07], called therein Wasserstein distances between persistence diagrams and proof of their stability $\to$ analogy with optimal transport literature.

[CSEH05]: *Stability of persistence diagrams*, Cohen-Steiner, Edelsbrunner, Harer, 2009.
[CSE+07]: *Lipschitz Functions Have L_p-Stable Persistence*, Cohen-Steiner, Edelsbrunner, Harer, Mileyko, 2007.

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

A concise history of diagram metrics

- (2005) Introduction of the bottleneck distance d_∞ and proof of a stability theorem [CSEH05].
- (2007) Introduction of d_p, with $p < \infty$ [CSE+07], called therein Wasserstein distances between persistence diagrams and proof of their stability $\rightarrow$ analogy with optimal transport literature.
- (2009, 2011) Isometry theorem between d_∞ and the *interleaving distance*, a metric between algebraic objects [CSGO09, Les11] $\rightarrow$ algebraic motivation.

[CCS+09]: *Proximity of Persistence Modules and their Diagrams*, Chazal, Cohen-Steiner, Glisse, Guibas, Oudot, 2009.
[Les11]: *The theory of the interleaving distance on multidimensional persistence modules*, Lesnick, 2011.

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

A concise history of diagram metrics

- (2005) Introduction of the bottleneck distance d_∞ and proof of a stability theorem [CSEH05].
- (2007) Introduction of d_p, with $p < \infty$ [CSE+07], called therein Wasserstein distances between persistence diagrams and proof of their stability $\rightarrow$ analogy with optimal transport literature.
- (2009, 2011) Isometry theorem between d_∞ and the *interleaving distance*, a metric between algebraic objects [CSGO09, Les11] $\rightarrow$ algebraic motivation.
- (2011) Seminal study of the space of persistence diagrams $(\mathcal{D}^p, d_p)$ [MMH11].
- (2014) Turner et al. study Fréchet means in $\mathcal{D}^2$ and propose an algorithm to estimate them [TMMH14] $\rightarrow$ statistical object **in** $\mathcal{D}^2$ (not a vector space).

[MMH11]: *Probability measures on the space of persistence diagrams*, Mileyko, Mukherjee, Harer, 2011.
[TMMH14]: *Fréchet means for distributions of persistence diagrams*, Turner, Mileyko, Mukherjee, Harer, 2014.

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

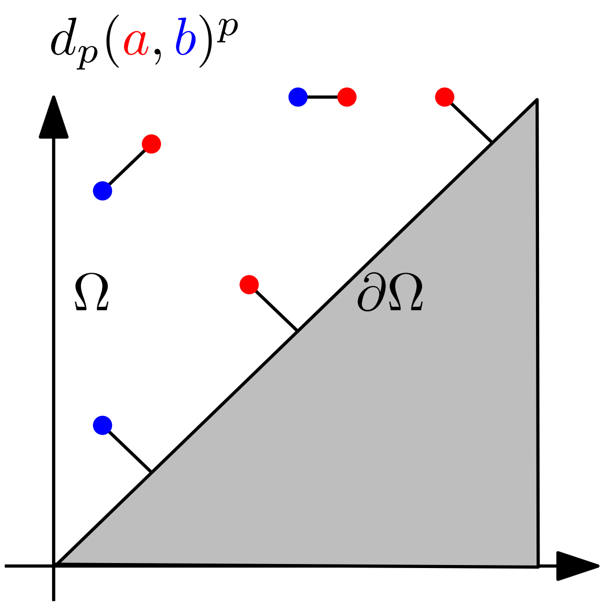
A concise history of diagram metrics

- (2005) Introduction of the bottleneck distance d_∞ and proof of a stability theorem [CSEH05].
- (2007) Introduction of d_p, with $p < \infty$ [CSE+07], called therein Wasserstein distances between persistence diagrams and proof of their stability $\to$ analogy with optimal transport literature.
- (2009, 2011) Isometry theorem between d_∞ and the *interleaving distance*, a metric between algebraic objects [CSGO09, Les11] $\to$ algebraic motivation.
- (2011) Seminal study of the space of persistence diagrams $(\mathcal{D}^p, d_p)$ [MMH11].
- (2014) Turner et al. study Fréchet means in $\mathcal{D}^2$ and propose an algorithm to estimate them [TMMH14] $\to$ statistics with PDs.
- (2017) Sliced-Wasserstein kernels for PDs [CCO17] $\to$ Adapting tools from OT to TDA $\Rightarrow$ More than just an analogy?

[CCO17]: *Sliced-Wasserstein kernel for persistence diagrams*, Carrière, Cuturi, Oudot, 2017.

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

The metrics d_p are **not** standard Wasserstein distances, although sharing key ideas.



$\mathcal{W}^p(\mathcal{X})$: space of **probability** measures with finite p-th moment on a Polish metric space $\mathcal{X}$ equipped with the Wasserstein distance:

$$W_p(\mu,\nu)^p := \inf_P \iint_{\mathbb{R}^d \times \mathbb{R}^d} C(x,y)\mathrm{d}P(x,y)$$

where:

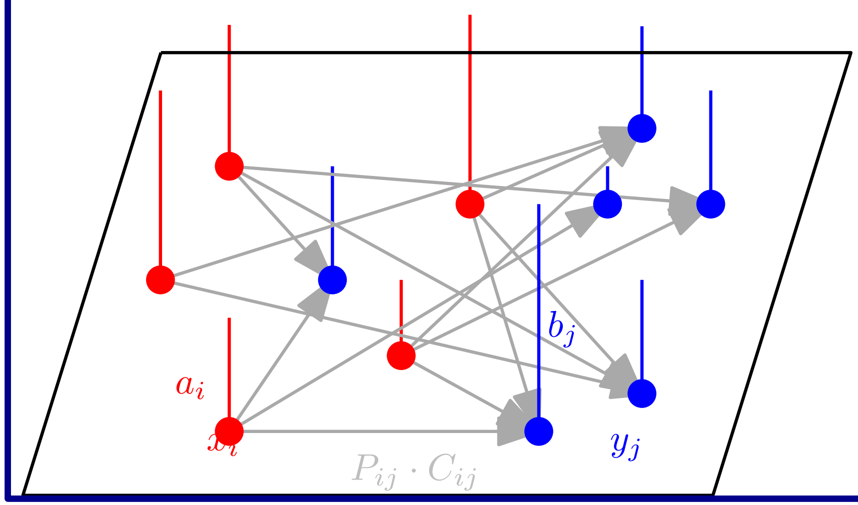
$$C(x,y) = d(x,y)^p$$

subject to (for all $A, B \subset \mathcal{X}$)

marginal constraints

$$P(A \times \mathbb{R}^d) = \mu(A)$$

$$P(\mathbb{R}^d \times B) = \nu(B)$$

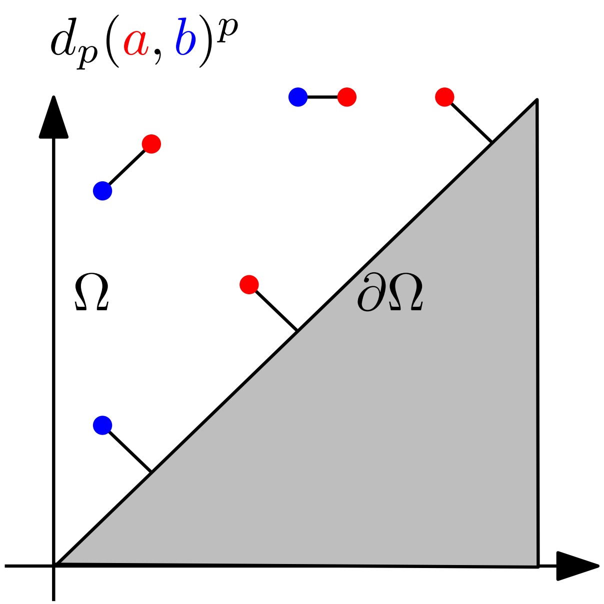


[EH10]: *An Introduction to Computational Topology*, Edelsbrunner and Harer, 2010.
[CSEH07]: *Stability of Persistence Diagrams*, Cohen-Steiner, Edelsbrunner, Harer, 2007.
[CCSG+09]: *Gromov-Hausdorff stable signatures for shapes using persistence*, Chazal, Cohen-Stein, Guibas, Mémoli, Oudot, 2009.

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

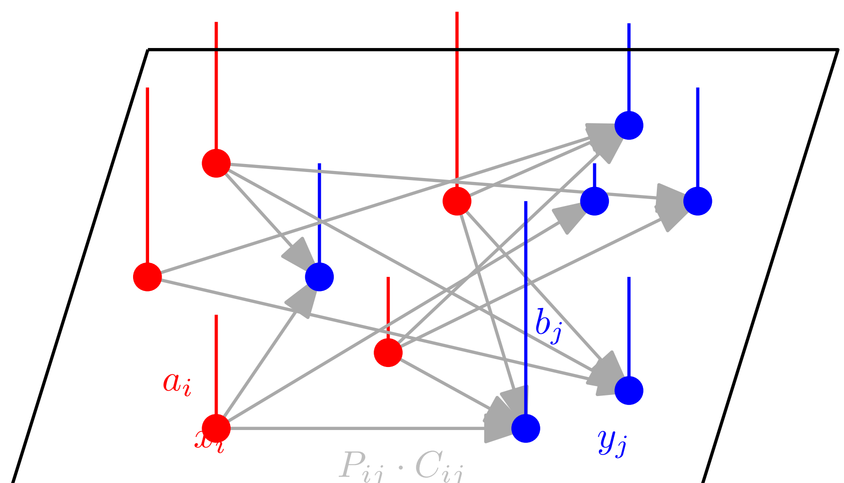
The metrics d_p are **not** standard Wasserstein distances, although sharing key ideas.

Persistence Diagrams $(\mathcal{D}^p, d_p)$	Optimal Transport $(\mathcal{W}^p, W_p)$
• Discrete point measures	• Discrete and/or **continuous** support.
• Measures with different total masses	• Measures with same total masses



$\mathcal{W}^p(\mathcal{X})$: space of **probability** measures with finite p-th moment on a Polish metric space $\mathcal{X}$ equipped with the Wasserstein distance:

$$W_p(\mu,\nu)^p := \inf_P \iint_{\mathbb{R}^d\times\mathbb{R}^d} C(x,y)\mathrm{d}P(x,y)$$



where:

$$C(x,y) = d(x,y)^p$$

subject to (for all $A, B \subset \mathcal{X}$)

marginal constraints

$$P(A\times\mathbb{R}^d) = \mu(A)$$

$$P(\mathbb{R}^d\times B) = \nu(B)$$

[EH10]: *An Introduction to Computational Topology*, Edelsbrunner and Harer, 2010.
[CSEH07]: *Stability of Persistence Diagrams*, Cohen-Steiner, Edelsbrunner, Harer, 2007.
[CCSG+09]: *Gromov-Hausdorff stable signatures for shapes using persistence*, Chazal, Cohen-Stein, Guibas, Mémoli, Oudot, 2009.

Filling the gap: what can TDA gain from OT?

Why is it important?

• OT: well-developed theoretical field [V08, S15] while computational OT has known impressive progress in recent years [FC17, PC19].
$\rightarrow$ A connection between PD-metrics and OT would allow us to adapt the various tools developed in OT to manipulate PDs and would be highly beneficial to TDA.

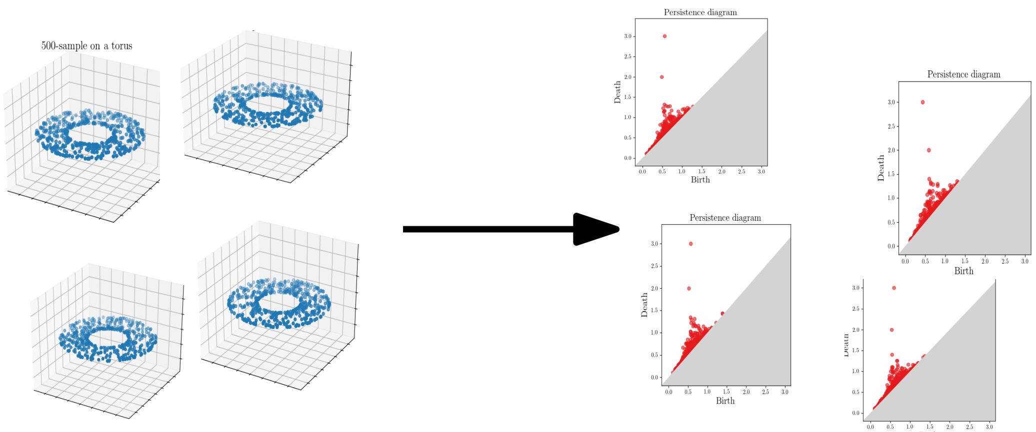
[V08]: *Optimal Transport, old and new*, Villani, 2008.
[S15]: *Optimal Transport for Applied Mathematicians*, Santambrogio, 2015.
[PC19]: *Computational Optimal Transport: with applications to Data Science*, Peyré and Cuturi, 2019.
[FC17]: *POT: Python Optimal Transport library*, Flamary and Courty, 2017.

Filling the gap: what can TDA gain from OT?

Why is it important?

• OT: well-developed theoretical field [V08, S15] while computational OT has known impressive progress in recent years [FC17, PC19].
$\rightarrow$ A connection between PD-metrics and OT would allow us to adapt the various tools developed in OT to manipulate PDs and would be highly beneficial to TDA.

$\rightarrow$ Would enable natural generalizations of persistence diagrams that appear in applications of TDA.
$\rightarrow$ Needs first to extend the metrics d_p to measures with continuous support.



$$\frac{1}{K}\sum_{i=1}^{K}\mu^{(i)}$$

K point clouds i.i.d. $\qquad$ K PDs $\mu^{(1)} \dots \mu^{(K)}$

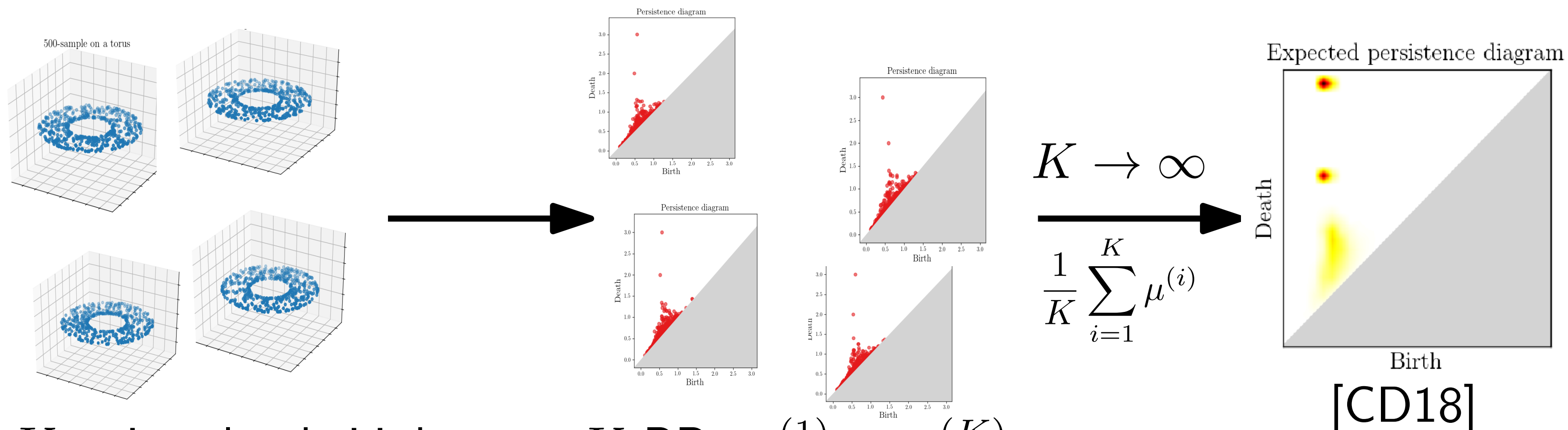
[V08]: *Optimal Transport, old and new*, Villani, 2008.
[S15]: *Optimal Transport for Applied Mathematicians*, Santambrogio, 2015.
[PC19]: *Computational Optimal Transport: with applications to Data Science*, Peyré and Cuturi, 2019.
[FC17]: *POT: Python Optimal Transport library*, Flamary and Courty, 2017.

Filling the gap: what can TDA gain from OT?

Why is it important?

• OT: well-developed theoretical field [V08, S15] while computational OT has known impressive progress in recent years [FC17, PC19].
$\rightarrow$ A connection between PD-metrics and OT would allow us to adapt the various tools developed in OT to manipulate PDs and would be highly beneficial to TDA.

$\rightarrow$ Would enable natural generalizations of persistence diagrams that appear in applications of TDA.
$\rightarrow$ Needs first to extend the metrics d_p to measures with continuous support.



Stability? Convergence rates?

K point clouds i.i.d. $\quad$ K PDs $\mu^{(1)} \ldots \mu^{(K)}$

[V08]: *Optimal Transport, old and new*, Villani, 2008.
[S15]: *Optimal Transport for Applied Mathematicians*, Santambrogio, 2015.
[PC19]: *Computational Optimal Transport: with applications to Data Science*, Peyré and Cuturi, 2019.
[FC17]: *POT: Python Optimal Transport library*, Flamary and Courty, 2017.
[CD18]: *The density of expected persistence diagrams and its kernel based estimation*, Chazal and Divol, 2018.

Filling the gap: what can TDA gain from OT?

My main contributions:

Part I.

The bridge: Optimal Partial Transport with boundary and new results for persistence diagrams [D**L**19].

→ Characterization of convergence, continuity of vectorizations, manipulation of generalized PDs...

Part II.

Computational benefits: estimation of Fréchet means for persistence diagrams using modern OT tools [**L**CO18]

→ Entropic regularization for metrics d_p, efficient estimation of Fréchet means...

[DL19]: *Understanding the Topology and the Geometry of the Space of Persistence Diagrams via Optimal Partial Transport*, Divol, L, 2019.
[LCO18]: *Large scale computation of means and clusters for persistence diagrams using optimal transport*, L, Cuturi, Oudot, 2010.

1. The bridge: Optimal Partial Transport with boundary

Part I.

The bridge: Optimal Partial Transport with boundary and new results for persistence diagrams [D**L**19].

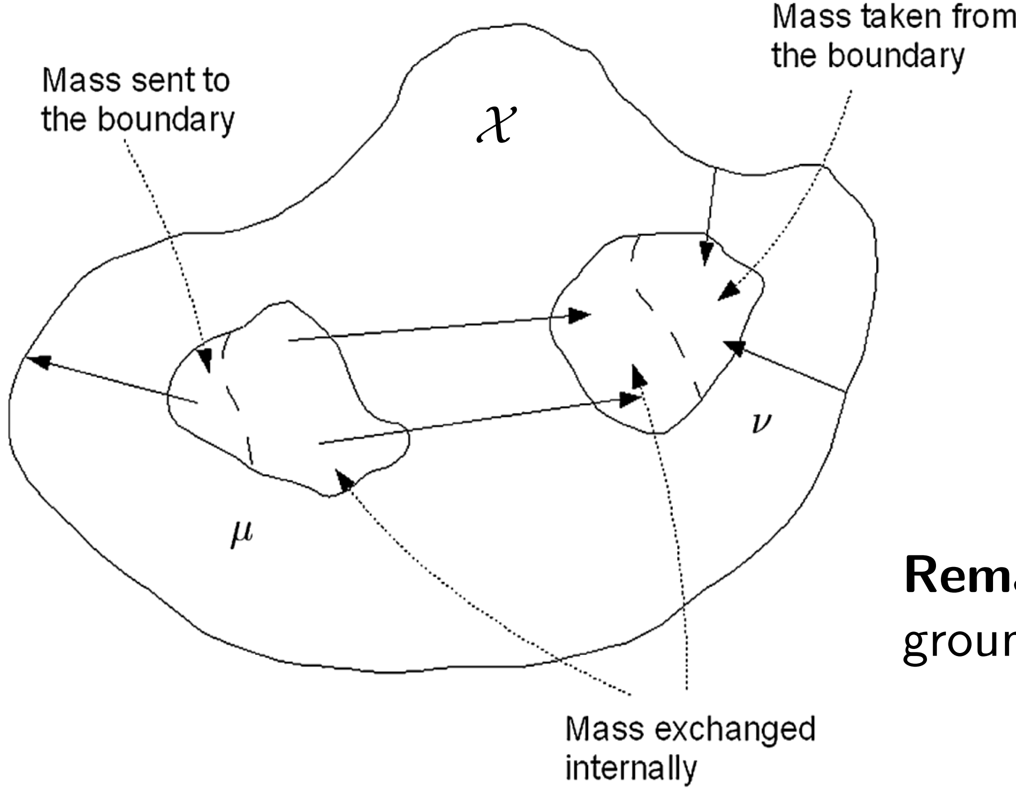
Part II.

Computational benefits: estimation of Fréchet means for persistence diagrams using modern OT tools [**L**CO18]

1. The bridge: Optimal Partial Transport with boundary

A.Figalli and N.Gigli [FG10]:

A. Figalli, N. Gigli / J. Math. Pures Appl. 94 (2010) 107–130



Remark: Assumption: ground space $\mathcal{X}$ compact.

> This means that we can use $\partial\mathcal{X}$ as an infinite reserve: we can ‘take’ as mass as we wish from the boundary, or ‘give’ it back some of the mass, provided we pay the transportation cost, see Fig. 1. This is why this distance is well defined for measures which do not have the same mass.

[FG10]: *A new transportation distance between non-negative measures, with applications to gradients flows with Dirichlet boundary conditions*, Figalli, Gigli, 2010.

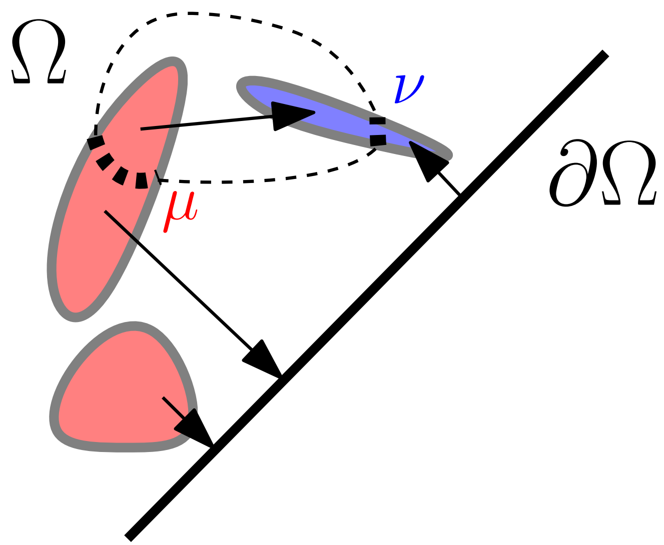
1. The bridge: Optimal Partial Transport with boundary

Core idea: Consider sub-marginal constraints!

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

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

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



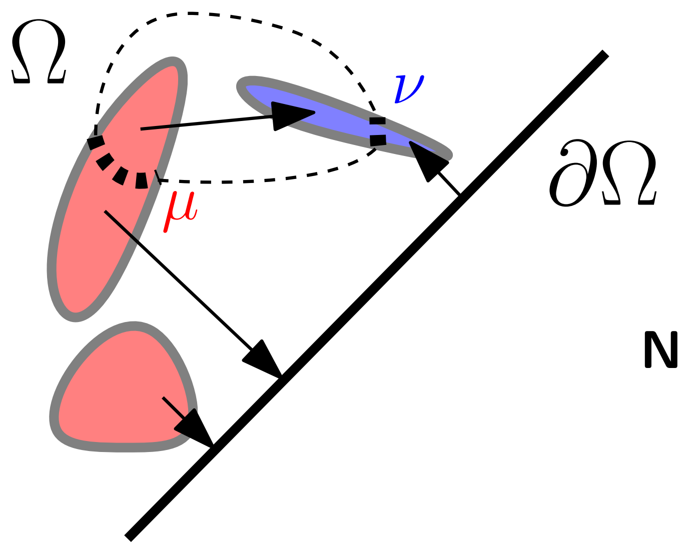
1. The bridge: Optimal Partial Transport with boundary

Core idea: Consider sub-marginal constraints!

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

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

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



and then simply define

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

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

Note: restrict to $\mathcal{M}^p := \{\mu : \int_\Omega d(x,\partial\Omega)^p \mathrm{d}\mu(x) < +\infty\}$

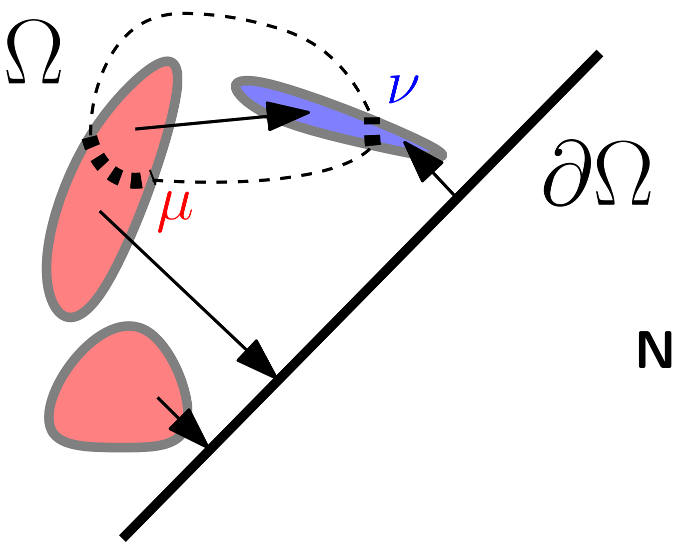
1. The bridge: Optimal Partial Transport with boundary

Core idea: Consider sub-marginal constraints!

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

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

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



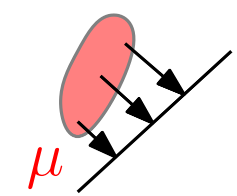
and then simply define

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

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

Note: restrict to $\mathcal{M}^p := \{\mu : \int_\Omega d(x,\partial\Omega)^p \mathrm{d}\mu(x) < +\infty\}$

$\mathrm{Pers}_p(\mu)$



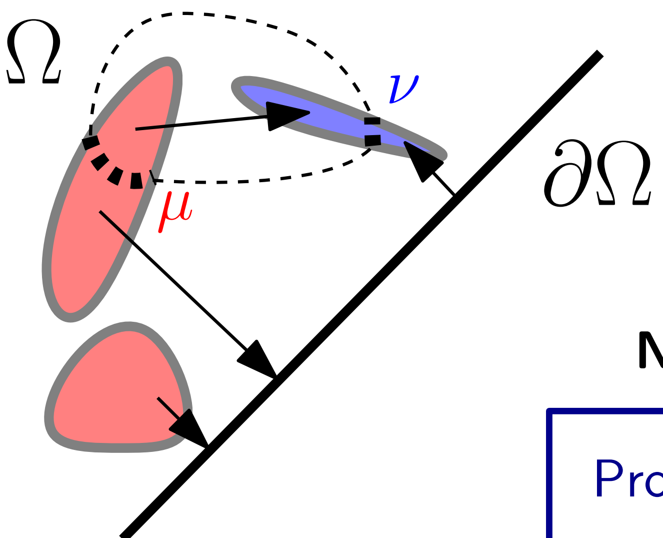
1. The bridge: Optimal Partial Transport with boundary

Core idea: Consider sub-marginal constraints!

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

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

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



and then simply define

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

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

Note: restrict to $\mathcal{M}^p := \{\mu : \int_\Omega d(x,\partial\Omega)^p \mathrm{d}\mu(x) < +\infty\}$

Proposition [D**L**19]: Let $1 \leq p \leq \infty$.

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

$\rightarrow$ The metric OT_p is a natural extension of the metric d_p, and we can consider elements of the larger space $(\mathcal{M}^p, \mathrm{OT}_p)$, called persistence measures.

1. The bridge: Optimal Partial Transport with boundary

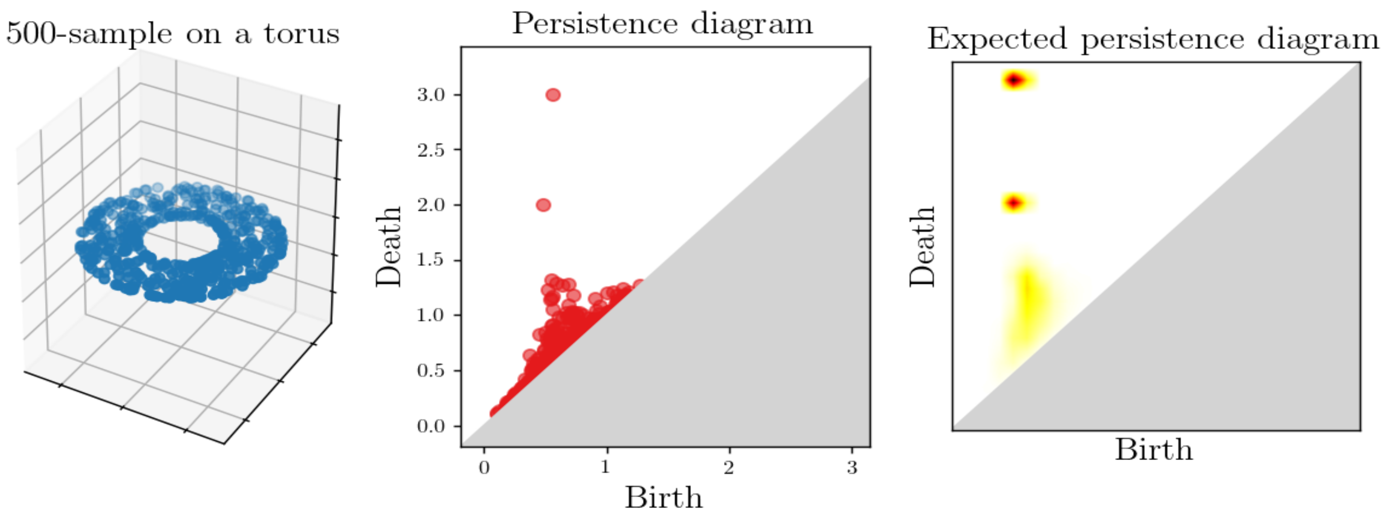
Definition: expected persistence diagrams [CD18]

Let P be a probability distribution supported on $(\mathcal{D}^p, \mathrm{OT}_p)$ satisfying $\mathbb{E}_{\mu\sim P}[\mathrm{Pers}_p(\mu)] < +\infty$.

Define for $A \subset \Omega$ Borel,

$$E[P](A) = \mathbb{E}_{\mu\sim P}[\mu(A)],$$

Remark: The EPD $E[P]$ is $\mathcal{M}^p$ but **not** in $\mathcal{D}^p$ in general.



1. The bridge: Optimal Partial Transport with boundary

Definition: expected persistence diagrams [CD18]

Let P be a probability distribution supported on $(\mathcal{D}^p, \mathrm{OT}_p)$ satisfying $\mathbb{E}_{\mu\sim P}[\mathrm{Pers}_p(\mu)] < +\infty$.

Define for $A \subset \Omega$ Borel,

$$E[P](A) = \mathbb{E}_{\mu\sim P}[\mu(A)],$$

Remark: The EPD $E[P]$ is $\mathcal{M}^p$ but **not** in $\mathcal{D}^p$ in general.

Proposition [DL**19]:**

Let P_1, P_2 be two probability distributions on $(\mathcal{D}^p, \mathrm{OT}_p)$, then

$$\mathrm{OT}_p(E[P_1], E[P_2]) \leq W_{p,\mathrm{OT}_p}(P_1, P_2).$$

- Similar distributions in $(\mathcal{D}^p, \mathrm{OT}_p)$ have similar EPDs.
- This result requires to introduce the metric OT_p to make sense.

1. The bridge: Optimal Partial Transport with boundary

Theorem [DL19]: Let $(\mu_n)_n, \mu \in \mathcal{M}^p$. One has

$$\mathrm{OT}_p(\mu_n, \mu) \to 0 \Leftrightarrow \begin{cases} \mu_n \to \mu \text{ vaguely} \\ \mathrm{Pers}_p(\mu_n) \to \mathrm{Pers}_p(\mu) \end{cases}.$$

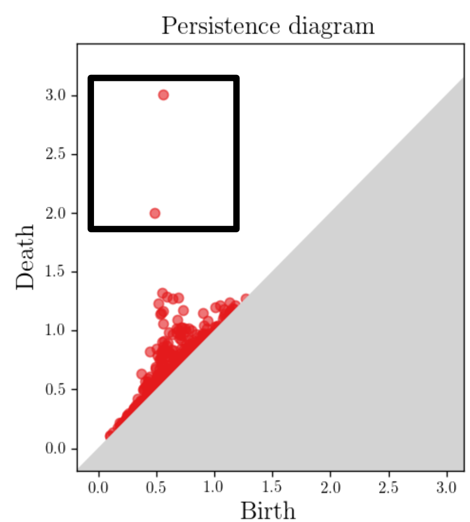
Note: In particular, $\mathcal{D}^p$ is closed in $\mathcal{M}^p$ for the metric OT_p (as it is closed for the vague convergence).

Note: Similar result in [FG10] when the groundspace is compact. Adaptation to Ω (non-compact) requires care.

Recall (*vague* convergence):

For all $f : \Omega \to \mathbb{R}$ continuous, compactly supported:

$$\mu_n(f) = \int f(x) \mathrm{d}\mu_n(x) \to \int f(x) \mathrm{d}\mu(x) = \mu(f)$$



1. The bridge: Optimal Partial Transport with boundary

Theorem [DL19]: Let $(\mu_n)_n, \mu \in \mathcal{M}^p$. One has

$$\mathrm{OT}_p(\mu_n, \mu) \to 0 \Leftrightarrow \begin{cases} \mu_n \to \mu \text{ vaguely} \\ \mathrm{Pers}_p(\mu_n) \to \mathrm{Pers}_p(\mu) \end{cases}.$$

Application: Understanding convergence helps to study continuity.
$\to$ E.g. obtaining a characterization of continuous *linear vectorizations*, routinely used in practice in TDA, i.e. maps of the form

$$\mathcal{M}^p \ni \mu \mapsto \mu(f) = \int f(x) \mathrm{d}\mu(x) \in \mathcal{B}$$

for $f : \Omega \to \mathcal{B}$, where $\mathcal{B}$ is a Banach space (in practice, $\mathcal{B} = \mathbb{R}^d$).

2. Computational benefits: Fréchet means in $(\mathcal{M}^p, \mathrm{OT}_p)$.

Part I.

The bridge: Optimal Partial Transport with boundary and new results for persistence diagrams [D**L**19].

Part II.

Computational benefits: estimation of Fréchet means for persistence diagrams using modern OT tools [**L**CO18]

2. Computational benefits: Fréchet means in $(\mathcal{M}^p, \mathrm{OT}_p)$.

Definition (Fréchet means, aka barycenters):

Consider $b_1, \dots, b_N$ a set of diagrams. Fréchet means are minimizers (if they exist) of

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

2. Computational benefits: Fréchet means in $(\mathcal{M}^p, \mathrm{OT}_p)$.

Definition (Fréchet means, aka barycenters):

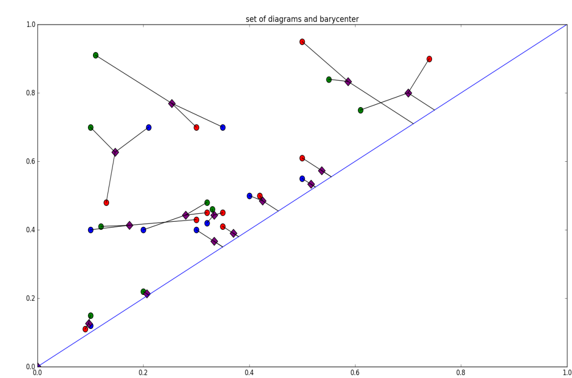
Consider $b_1, \ldots, b_N$ a set of diagrams. Fréchet means are minimizers (if they exist) of

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

First results [TMMH14]:

- $\mathcal{E} : \mathcal{D}^2 \to \mathbb{R}$ is not convex. It admits global (and local) minimizers

- Local minimizers can be computed ($\sim$[CD14]) by iteratively computing N optimal partial matchings between a and the b_is (expensive).



[TMMH14]: *Fréchet means for distributions of persistence diagrams*, Turner, Mileyko, Mukherjee, Harer, 2014.
[CD14]: *Fast computation of Wasserstein barycenters*, Cuturi, Doucet, 2014.

2. Computational benefits: Fréchet means in $(\mathcal{M}^p, \mathrm{OT}_p)$.

Definition (Fréchet means, aka barycenters):

Consider $b_1, \dots, b_N$ a set of diagrams. Fréchet means are minimizers (if they exist) of

$$a \mapsto \mathcal{E}(a) = \frac{1}{N} \sum_{i=1}^{N} \cancel{d_2(a, b_i)}^2, \quad \cancel{a \text{ persistence diagram}}.$$

$$\mathrm{OT}_2(a, b_i)^2 \quad \text{any element of } \mathcal{M}^2$$

2. Computational benefits: Fréchet means in $(\mathcal{M}^p, \mathrm{OT}_p)$.

Definition (Fréchet means, aka barycenters):

Consider $b_1, \dots, b_N$ a set of diagrams. Fréchet means are minimizers (if they exist) of

$$a \mapsto \mathcal{E}(a) = \frac{1}{N} \sum_{i=1}^{N} \cancel{d_2(a, b_i)}^2, \quad \cancel{a \text{ persistence diagram}}.$$

$$\mathrm{OT}_2(a, b_i)^2 \quad \text{any element of } \mathcal{M}^2$$

Proposition [DL19]:

- $\mathcal{E} : \mathcal{M}^2 \to \mathbb{R}$ is now convex, admits global minimizers.
- Some of them are actual diagrams.

$\to$ Extending to $\mathcal{M}^p$ can be beneficial when considering optimization problems in $\mathcal{D}^p$.

2. Computational benefits: Fréchet means in $(\mathcal{M}^p, \mathrm{OT}_p)$.

Definition (Fréchet means, aka barycenters):

Consider $b_1, \dots, b_N$ a set of diagrams. Fréchet means are minimizers (if they exist) of

$$a \mapsto \mathcal{E}(a) = \frac{1}{N} \sum_{i=1}^{N} \underbrace{\cancel{d_2(a, b_i)^2}}_{\mathrm{OT}_2(a, b_i)^2}, \quad a \text{ \underbrace{\cancel{\text{persistence diagram}}}_{\text{any element of } \mathcal{M}^2}}.$$

Proposition [DL19]:

- $\mathcal{E} : \mathcal{M}^2 \to \mathbb{R}$ is now convex, admits global minimizers.
- Some of them are actual diagrams.

$\to$ Extending to $\mathcal{M}^p$ can be beneficial when considering optimization problems in $\mathcal{D}^p$.

Proposition [LCO18]

- These can be approximated efficiently (Sinkhorn algorithm).

2. Computational benefits: Fréchet means in $(\mathcal{M}^p, \mathrm{OT}_p)$.

Problem: Can't optimize in $\mathcal{M}^2$ in practice $\rightarrow$ need finite-dim parameters.

• First idea (Lagrangian): fix k, and consider $\sum_{i=1}^k m_i \delta_{x_i}$, where $m_i \in \mathbb{R}_+, x_i \in \Omega$ are parameters to be optimized.

$\Rightarrow$ Goes back to a non-convex problem (∼[CD14]). ✗

[CD14]: *Fast computation of Wasserstein barycenters*, Cuturi, Doucet, 2014.

2. Computational benefits: Fréchet means in $(\mathcal{M}^p, \mathrm{OT}_p)$.

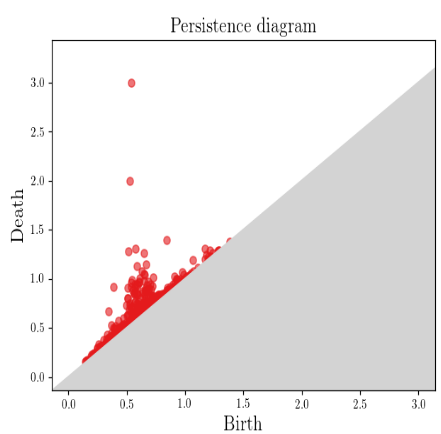
Problem: Can't optimize in $\mathcal{M}^2$ in practice $\rightarrow$ need finite-dim parameters.

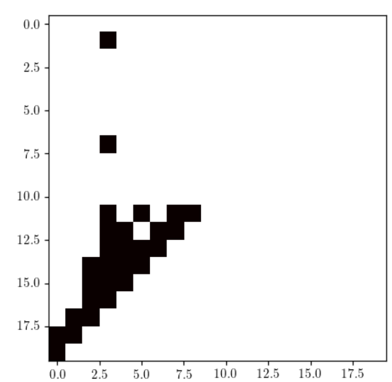
- First idea (Lagrangian): fix k, and consider $\sum_{i=1}^{k} m_i \delta_{x_i}$, where $m_i \in \mathbb{R}_+, x_i \in \Omega$ are parameters to be optimized.

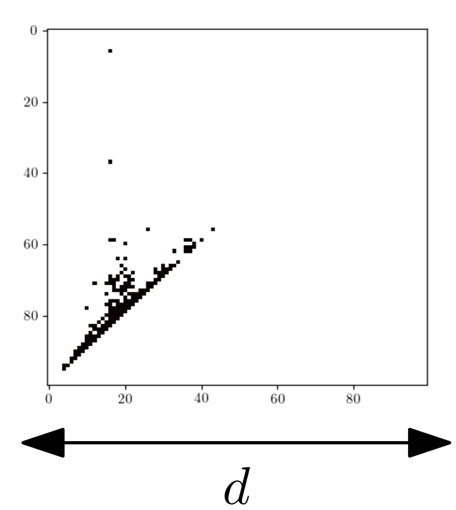
  $\Rightarrow$ Goes back to a non-convex problem ($\sim$[CD14]). ✗

- Idea 2 (Eulerian):

  Discretize the ground space, $\mu, \nu \rightarrow A, B \in \mathbb{R}^{d \times d}$ (2D-histograms) and optimize masses in bins.







$\Rightarrow$ Lead to a convex problem. ✓

[CD14]: *Fast computation of Wasserstein barycenters*, Cuturi, Doucet, 2014.

2. Computational benefits: Fréchet means in $(\mathcal{M}^p, \mathrm{OT}_p)$.

- Discretize the ground space, $\mu, \nu \to A, B \in \mathbb{R}^{d\times d}$ (2D-histograms)

  Use (standard) computational OT tools on histograms with A, B.

  Problem: $|A| \neq |B|$, and we want to take $\partial\Omega$ into account.

2. Computational benefits: Fréchet means in $(\mathcal{M}^p, \mathrm{OT}_p)$.

- Discretize the ground space, $\mu, \nu \to A, B \in \mathbb{R}^{d\times d}$ (2D-histograms)

  Use (standard) computational OT tools on histograms with A, B.

  Problem: $|A| \neq |B|$, and we want to take $\partial\Omega$ into account.

  Solution:

$$\mathrm{OT}_2^2(A, B) = \min_P \sum_{i,j,i',j'} P_{i,j,i',j'} \mathsf{C}_{i,j,i',j'}$$

  where $P \in \mathbb{R}^{(d\times d+1)\times(d\times d+1)}$ has marginals $(A, |B|)$ and $(B, |A|)$,

  and $$\mathsf{C} := \begin{array}{cc} \begin{pmatrix} d((i,j),(i',j'))^p & d((i,j),\partial\Omega)^p \\ d((i',j'),\partial\Omega)^p & 0 \end{pmatrix} & \updownarrow d^2 \\ \underleftrightarrow{\;d^2\;} & \end{array}$$

Remark: C does not depend on A, B.

2. Computational benefits: Fréchet means in $(\mathcal{M}^p, \mathrm{OT}_p)$.

- Discretize the ground space, $\mu, \nu \to A, B \in \mathbb{R}^{d \times d}$ (2D-histograms)

  Problem: Solving standard OT is not tractable on large histograms.

2. Computational benefits: Fréchet means in $(\mathcal{M}^p, \mathrm{OT}_p)$.

- Discretize the ground space, $\mu, \nu \to A, B \in \mathbb{R}^{d\times d}$ (2D-histograms)

  Problem: Solving standard OT is not tractable on large histograms.

Solution:

- Entropic regularization [Cut13]:

$$\mathrm{OT}_p^p(A, B) \simeq \min_P \sum_{i,j,i',j'} P_{i,j,i',j'} \mathsf{C}_{i,j,i',j'} + \varepsilon h(P) =: S_\varepsilon(A, B),$$

where P has marginals $(A, |B|)$ and $(B, |A|)$; and $h(P) := -\langle P, \log(P) - 1\rangle$.

[Cut13]: *Sinkhorn distances: Lightspeed computation of optimal transport*, Cuturi, 2013.

2. Computational benefits: Fréchet means in $(\mathcal{M}^p, \mathrm{OT}_p)$.

- Discretize the ground space, $\mu, \nu \to A, B \in \mathbb{R}^{d\times d}$ (2D-histograms)

  Problem: Solving standard OT is not tractable on large histograms.

Solution:

- Entropic regularization [Cut13]:

$$\mathrm{OT}_p^p(A, B) \simeq \min_P \sum_{i,j,i',j'} P_{i,j,i',j'} \mathsf{C}_{i,j,i',j'} + \varepsilon h(P) =: S_\varepsilon(A, B),$$

where P has marginals $(A, |B|)$ and $(B, |A|)$; and $h(P) := -\langle P, \log(P) - 1\rangle$.

- $S_\varepsilon(A, B)$ is estimated by iterating, starting from any $u_0, v_0 \in \mathbb{R}^{d^2+1}$:

$$u_{t+1} \leftarrow \frac{(A, |B|)}{\mathsf{K}\cdot v_t}, \qquad v_{t+1} \leftarrow \frac{(B, |A|)}{\mathsf{K}\cdot u_{t+1}}, \qquad \text{with } \mathsf{K} := e^{-\mathsf{C}/\varepsilon}.$$

[Cut13]: *Sinkhorn distances: Lightspeed computation of optimal transport*, Cuturi, 2013.

2. Computational benefits: Fréchet means in $(\mathcal{M}^p, \mathrm{OT}_p)$.

- Discretize the ground space, $\mu, \nu \to A, B \in \mathbb{R}^{d \times d}$ (2D-histograms)

  Problem: Solving standard OT is not tractable on large histograms.

Solution:

- Entropic regularization [Cut13]:

$$\mathrm{OT}_p^p(A, B) \simeq \min_P \sum_{i,j,i',j'} P_{i,j,i',j'} \mathsf{C}_{i,j,i',j'} + \varepsilon h(P) =: S_\varepsilon(A, B),$$

where P has marginals $(A, |B|)$ and $(B, |A|)$; and $h(P) := -\langle P, \log(P) - 1\rangle$.

- $S_\varepsilon(A, B)$ is estimated by iterating, starting from any $u_0, v_0 \in \mathbb{R}^{d^2+1}$:

$$u_{t+1} \leftarrow \frac{(A, |B|)}{\mathsf{K} \cdot v_t}, \qquad v_{t+1} \leftarrow \frac{(B, |A|)}{\mathsf{K} \cdot u_{t+1}}, \qquad \text{with } \mathsf{K} := e^{-\mathsf{C}/\varepsilon}.$$

- Has access to an approx gradient of $A \mapsto S_\varepsilon(A, B)$ from (u_T, v_T).

  In particular, of $A \mapsto \sum_{i=1}^N S_\varepsilon(A, B_i)$, which is convex $\Rightarrow$ Gradient descent.

[Cut13]: *Sinkhorn distances: Lightspeed computation of optimal transport*, Cuturi, 2013.

2. Computational benefits: Fréchet means in $(\mathcal{M}^p, \mathrm{OT}_p)$.

$$u_{t+1} \leftarrow \frac{A}{\mathsf{K} \cdot v_t}, \quad v_{t+1} \leftarrow \frac{B}{\mathsf{K} \cdot u_{t+1}}, \quad \text{with } \mathsf{K} := e^{-\mathsf{C}/\varepsilon}.$$

(Last) Problem: $P, \mathsf{C}, \mathsf{K}$ is of size $(d^2+1) \times (d^2+1)$, too large in practice.

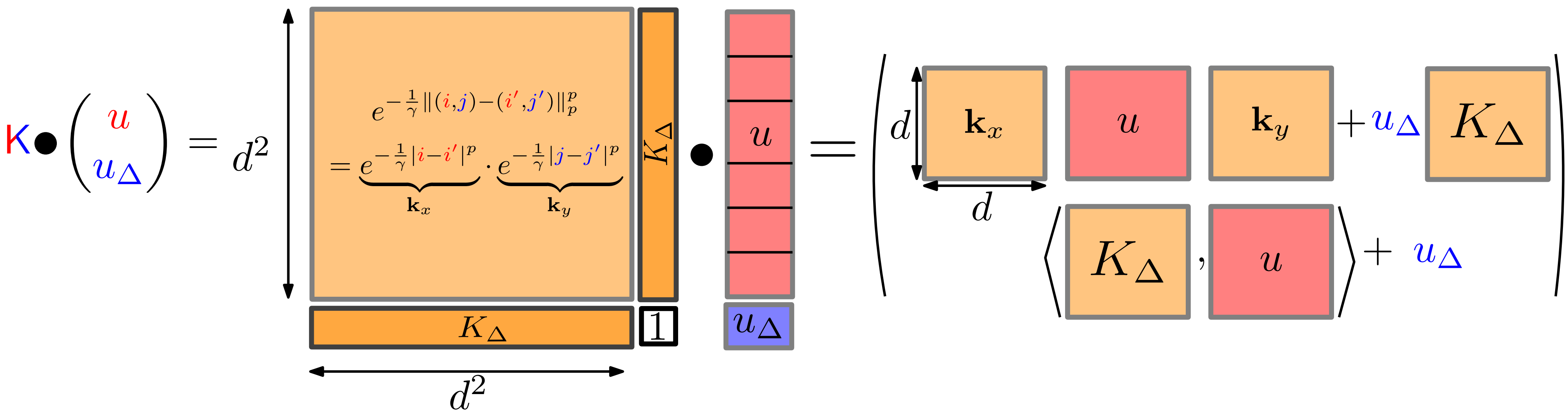
2. Computational benefits: Fréchet means in $(\mathcal{M}^p, \mathrm{OT}_p)$.

$$u_{t+1} \leftarrow \frac{A}{\mathsf{K} \cdot v_t}, \quad v_{t+1} \leftarrow \frac{B}{\mathsf{K} \cdot u_{t+1}}, \quad \text{with } \mathsf{K} := e^{-\mathsf{C}/\varepsilon}.$$

(Last) Problem: $P, \mathsf{C}, \mathsf{K}$ is of size $(d^2+1) \times (d^2+1)$, too large in practice.

Solution:

- Reduce to $(d \times d)$-matrix multiplications using 2D-convolutions [SGP+15] (never instantiate $\mathsf{K}, \mathsf{C}, P \ldots$) + can be parallelized and run on GPUs.

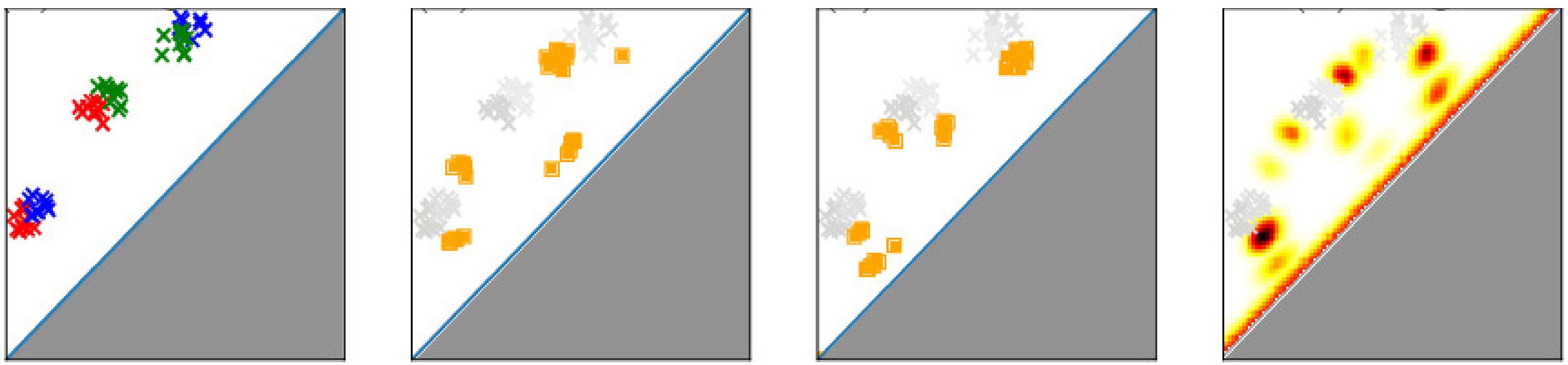


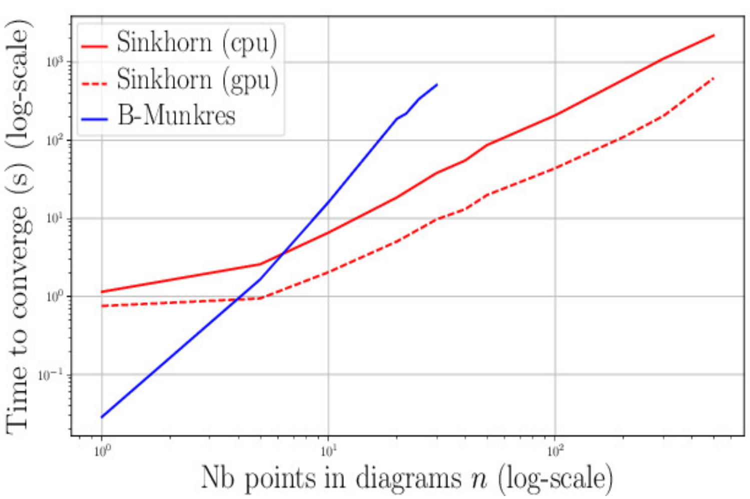
[Cut13]: *Sinkhorn distances: Lightspeed computation of optimal transport*, Cuturi, 2013.

[SGP+15]: *Convolutional wasserstein distances: Efficient optimal transportation on geometric domains*, Solomon, DeGoes, Peyré, Cuturi, Butscher, Nguyen, Du, Guibas, 2015.

2. Computational benefits: Fréchet means in $(\mathcal{M}^p, \mathrm{OT}_p)$.

Input (3 PDs) — Outputs of [THMM14] for different init. — Our output [**L**CO18]





	[TMMH14]	[**L**CO18]
Parameter-free	✓	✗ $(\varepsilon, T, d\ldots)$
CV in finitely many steps	✓	✗
Convexity	✗	✓
Scalability	✗	✓(//, GPU)
General input	✗ (only in $\mathcal{D}^p$)	✓ in $\mathcal{M}^p$

[TMMH14]: *Fréchet means for distributions of persistence diagrams*, Turner, Mileyko, Mukherjee, Harer, 2014.

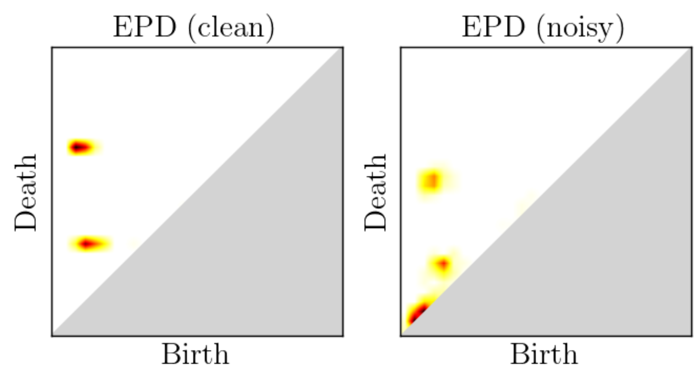
Conclusion

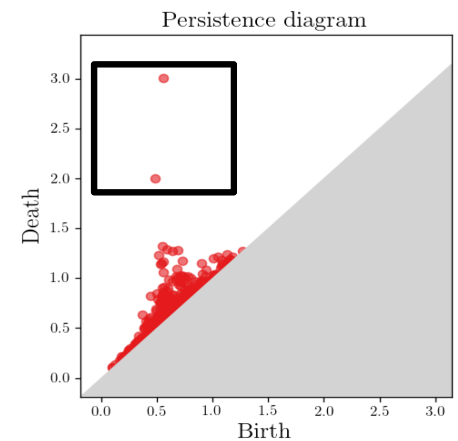
Most important take home messages:

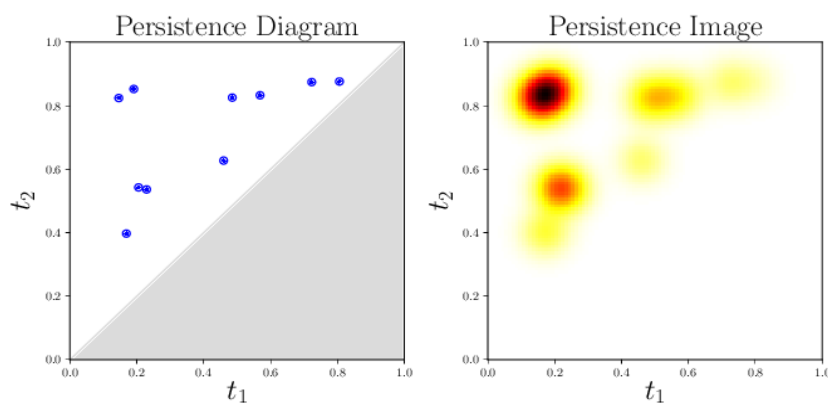
- PD metrics can be formulated as a true OT problem.
- Establishing this formal connection allows us to

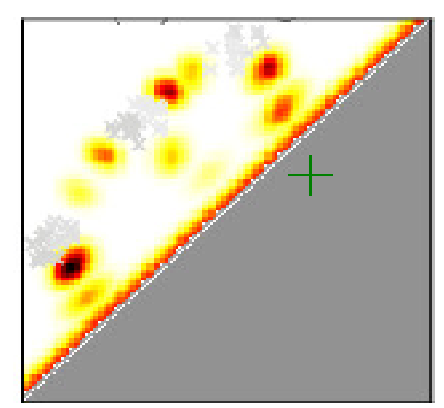
  $\rightarrow$ adapt theoretical ($\Rightarrow$ new results) and computational ($\Rightarrow$ new algorithms) tools from OT-literature to the context of PDs.

  $\rightarrow$ consider more general measures that appear in statistical applications of TDA









Conclusion

Most important take home messages:

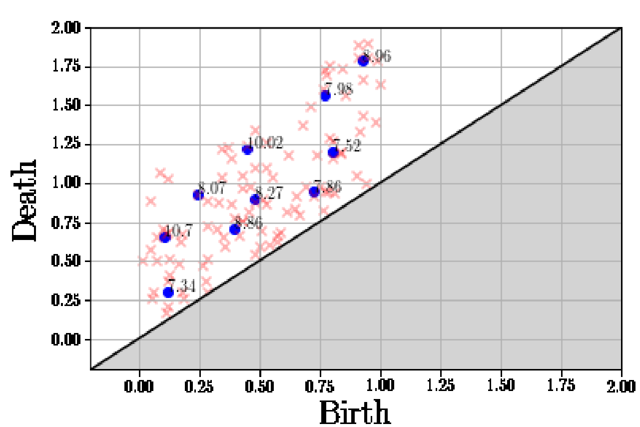
- PD metrics can be formulated as a true OT problem.
- Establishing this formal connection allows us to

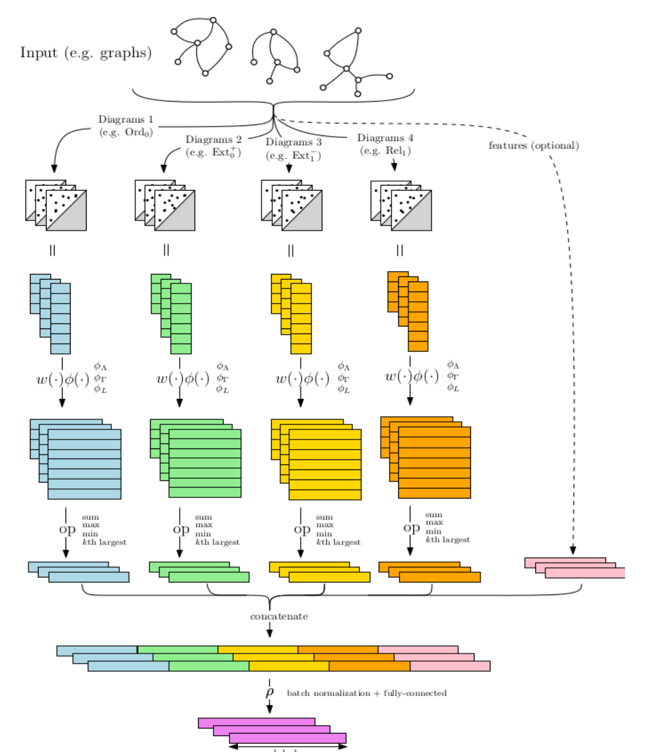
  $\to$ adapt theoretical ($\Rightarrow$ new results) and computational ($\Rightarrow$ new algorithms) tools from OT-literature to the context of PDs.

  $\to$ consider more general measures that appear in statistical applications of TDA

Not presented today:

- Duality
- Learning representations of PDs using PersLay [CCI**L**RU20]
- Quantization of PDs.





Conclusion

Most important take home messages:

- PD metrics can be formulated as a true OT problem.
- Establishing this formal connection allows us to

  $\rightarrow$ adapt theoretical ($\Rightarrow$ new results) and computational ($\Rightarrow$ new algorithms) tools from OT-literature to the context of PDs.

  $\rightarrow$ consider more general measures that appear in statistical applications of TDA

Not presented today:

- Duality
- Learning representations of PDs using PersLay [CCI**L**RU20]
- Quantization of PDs.

Perspectives:

- More tools from OT in TDA $\rightarrow$ Theoretical properties of (entropic?) regularizations for OT_p? $\rightarrow$ Gradient flows for PDs ?
- **Persistence measures:** Using optimal transport *maps* between PMs? Algebraic meaning of OT_∞?
- New (first?) use-case in OT with boundary, in particular with $p = \infty$.

Conclusion

Most important take home messages:

- PD metrics can be formulated as a true OT problem.
- Establishing this formal connection allows us to

  $\rightarrow$ adapt theoretical ($\Rightarrow$ new results) and computational ($\Rightarrow$ new algorithms) tools from OT-literature to the context of PDs.

  $\rightarrow$ consider more general measures that appear in statistical applications of TDA

Not presented today:

- Duality
- Learning representations of PDs using PersLay [CCI**L**RU20]
- Quantization of PDs.

Perspectives:

- More tools from OT in TDA $\rightarrow$ Theoretical properties of (entropic?) regularizations for OT_p? $\rightarrow$ Gradient flows for PDs ?
- **Persistence measures:** Using optimal transport *maps* between PMs? Algebraic meaning of OT_∞?
- New (first?) use-case in OT with boundary, in particular with $p = \infty$.

Thank you for your attention!