Understanding the Topology and the Geometry of the Persistence Diagram Space via Optimal Partial Transport

Vincent Divol and Théo Lacombe

firstname.lastname@inria.fr
Datashape, Inria Saclay

Abstract

We consider a generalization of persistence diagrams, namely Radon measures supported on the upper half plane for which we define natural extensions of Wasserstein and bottleneck distances between persistence diagrams. Such measures naturally appear in topological data analysis when considering continuous representations of persistence diagrams (e.g. persistence surfaces) but also as limits for laws of large numbers on persistence diagrams or as expectations of probability distributions on the persistence diagrams space. Introducing a formalism originating from the theory of optimal partial transport, we build a convenient framework to prove topological properties of this new space, which will also hold for the closed subspace of persistence diagrams. New results include a characterization of convergence with respect to Wasserstein metrics, and the existence of barycenters (Fréchet means) for any distribution of diagrams. We also showcase the strength of this framework by providing several statistical results made meaningful thanks to this new formalism.

1 Introduction

Topological Data Analysis (TDA) is an emerging field in data analysis that has found applications in computer vision [30], material science [21, 25], shape analysis [10, 37], to name a few. The aim of TDA is to provide interpretable descriptors of the underlying topology of a given object. One of the most used (and theoretically studied) descriptors in TDA is the *persistence diagram*. This descriptor consists in a locally finite multiset of points in the upper half plane $\Omega := \{(t_1, t_2) \in \mathbb{R}^2, t_2 > t_1\}$, each point in the diagram corresponding informally to a topological feature (connected component, loop, hole) of a given object. The space of persistence diagrams $\mathcal{D}$ is usually equipped with partial matching metrics d_p, sometimes called Wasserstein distances [17, Chapter VIII.2]: for $p \in [1, +\infty)$ and a, b in $\mathcal{D}$, define

$$d_p(a,b) := \left(\inf_{\pi \in \Gamma(a,b)} \sum_{x \in a \cup \partial\Omega} \|x - \pi(x)\|^p \right)^{\frac{1}{p}}, \tag{1}$$

where $\Gamma(a,b)$ is the set of partial matchings between a and b, i.e. bijections between $a \cup \partial\Omega$ and $b \cup \partial\Omega$, $\partial\Omega := \{(t,t),\ t \in \mathbb{R}\}$ being the boundary of Ω, namely the diagonal (see Figure 1). When $p \to \infty$, we recover the so-called *bottleneck distance*:

$$d_\infty(a,b) := \inf_{\pi \in \Gamma(a,b)} \max_{x \in a \cup \partial\Omega} \|x - \pi(x)\|. \tag{2}$$

An equivalent viewpoint, developed in [11, Chapter 3], is to consider a persistence diagram as a measure of the form $a = \sum_{x \in X} n_x \delta_x$, where $X \subset \Omega$ is locally finite and $n_x \in \mathbb{N}$ for all $x \in X$, so that a is a locally finite measure supported on Ω with integer mass on each point of its support. Considering persistence diagrams with such a perspective suggests to study more general Radon measures[1] supported on the upper half-plane Ω. Such objects appear naturally in several applications, e.g. when taking representations of persistence diagrams such as persistent surfaces [1], studying laws of large numbers for persistence diagrams [15, 16, 19], linear expectations of random persistence diagrams [12], or when estimating barycenters of persistence diagrams [28].

In Section 3, we study the space $\mathcal{M}$ of Radon measures supported on Ω. For finite $p \geq 1$ (the case $p = \infty$ is studied in Section 3.3), we define the persistence of $\mu \in \mathcal{M}$ as

$$\mathrm{Pers}_p(\mu) := \int_\Omega d(x, \partial\Omega)^p \mathrm{d}\mu(x), \tag{3}$$

[1] A short reminder about Radon measure theory is provided in Appendix A.

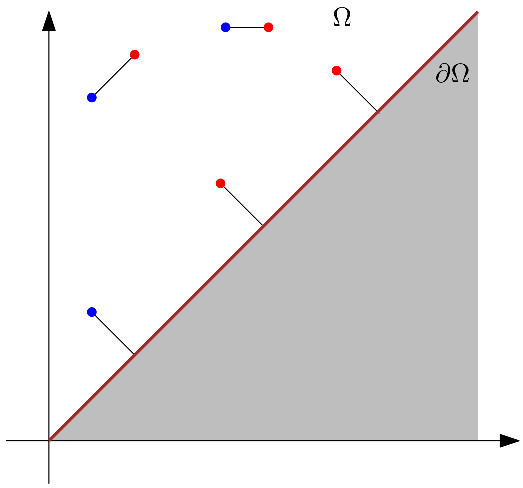

Figure 1: An example of optimal partial matching between two diagrams. The bottleneck distance between these two diagrams is—by definition—the length of the longest edge in this matching, while their Wasserstein distance d_p is the p-th root of the sum of all edge lengths to the power p.

where $d(x, \partial\Omega)$ is the distance from a point $x \in \Omega$ to the diagonal $\partial\Omega$, and we define

$$\mathcal{M}^p := \{\mu \in \mathcal{M},\ \mathrm{Pers}_p(\mu) < \infty\}. \tag{4}$$

Let $\mathcal{D}^p := \mathcal{D} \cap \mathcal{M}^p$ be the space of persistence diagrams with finite persistence p. We equip $\mathcal{M}^p$ with metrics D_p (see Definition 2.1), originally introduced in a work of Figalli and Gigli [18]. These metrics appear as extensions of metrics d_p, as they both coincide on $\mathcal{D}^p$ (Proposition 3.2). Elements of the metric space $(\mathcal{M}^p, D_p)$ are referred to as *persistence measures* in the following. As $\mathcal{D}^p$ is closed in $\mathcal{M}^p$ (Corollary 3.1), most properties of $\mathcal{M}^p$ hold for $\mathcal{D}^p$ too (e.g. being Polish, Proposition 3.3).

A sequence of Radon measures $(\mu_n)_n$ is said to *converge vaguely* to a measure μ, denoted by $\mu_n \xrightarrow{v} \mu$, if for any continuous compactly supported function $f : \Omega \to \mathbb{R}$, $\mu_n(f) \to \mu(f)$. We prove the following equivalence between convergence for the metric D_p and the vague convergence:

Theorem 3.4. *Let $1 \leq p < \infty$. Let $\mu, \mu_1, \mu_2, \ldots$ be measures in $\mathcal{M}^p$. Then,*

$$D_p(\mu_n, \mu) \to 0 \Leftrightarrow \begin{cases} \mu_n \xrightarrow{v} \mu, \\ \mathrm{Pers}_p(\mu_n) \to \mathrm{Pers}_p(\mu). \end{cases} \tag{5}$$

This characterization in particular holds for persistence diagrams in $\mathcal{D}^p$. This result is analog to the characterization of convergence of probability measures in the Wasserstein space (see [39, Theorem 6.9]). A proof for Radon measures supported on a bounded set can be found in [18, Proposition 2.7]. Our work consists in extending this result to non-bounded sets, in particular to the upper half plane Ω. As a corollary of this theorem, we obtain the following general result on the continuity of linear representations of persistence measures (and diagrams).

Corollary 3.2. *Let $p \in [1, +\infty)$, $d \geq 1$, and $f : \Omega \to \mathbb{R}^d$ be a continuous bounded function. The feature map $\Phi : \mathcal{M}^p \to \mathbb{R}^d$ defined by $\Phi(\mu) = \int_\Omega d(x, \partial\Omega)^p f(x) \mathrm{d}\mu(x)$ is continuous with respect to D_p.*

This new result can be compared to the recent works [26, Proposition 3.2] and [15, Theorem 3], which show that linear feature maps can have more regularity (e.g. Lipschitz or Hlder) under additional assumptions. We also show in Section 3.2, that the problem of computing the D_p metric between two measures of finite mass can be turned into the known problem of computing a Wasserstein distance (see Section 2) between two measures with the same masses, a result having practical implications for the computation of distances d_p between diagrams.

Section 4 studies Fréchet means (i.e. barycenters, see Definition 4.1) for probability distributions of persistence measures. In the specific case of persistence diagrams, the study of Fréchet means was initiated in [31, 36], where authors prove the existence of barycenters for certain types of distributions [31, Theorem 28]. We show that this existence result is actually true for any distributions of persistence measures, and adapt them to persistence diagrams. Namely, we prove the following results:

Theorem 4.3. *For any probability distribution $\mathbb{P}$ supported on $\mathcal{M}^p$ with finite p-th moment, the set of barycenters of $\mathbb{P}$ is not empty.*

Theorem 4.4. *If $\mathbb{P}$ is supported on $\mathcal{D}^p$ and has a finite p-th moment, then $\mathbb{P}$ admits a barycenter in $\mathcal{D}^p$.*

Theorem 4.3 follows the work of [29, Theorem 2] (itself following the seminal paper of Agueh and Carlier [2]), where authors prove the existence of Fréchet mean of probability measures endowed with the Wasserstein metric. We adapt this result to the space of Radon measures endowed with the D_p metrics.

Section 5 presents two applications of Theorem 3.4: Proposition 5.1 gives a law of large numbers in terms of the metric D_p, while Proposition 5.3 states a stability result between (Čech) diagrams and input point cloud in a random setting.

2 Elements of optimal transport

In this section, (Ω, d) denotes a Polish space. Optimal transport is a widely developed theory providing tools to study and compare probability measures supported on Ω [38, 39, 33].

2.1 Wasserstein distances

Given two probability measures μ, ν supported on (Ω, d), the *Wasserstein-p* distance ($p \geq 1$) induced by the metric d between these probability measures is defined as

$$W_{p,d}(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu,\nu)} \iint_{\Omega \times \Omega} d(x,y)^p \mathrm{d}\pi(x,y) \right)^{\frac{1}{p}}, \tag{6}$$

where $\Pi(\mu, \nu)$ denotes the set of *transport plans* between μ and ν, that is the set of measures on $\Omega \times \Omega$ which have respective marginals $\pi^{(1)} = \mu$, $\pi^{(2)} = \nu$. When there is no ambiguity on the distance d used, we simply write W_p instead of $W_{p,d}$. In order to have W_p finite, μ and ν are required to have a finite p-th moment, that is there exists $x_0 \in \Omega$ such that $\int_\Omega d(x, x_0)^p \mathrm{d}\mu(x)$ (resp. $\mathrm{d}\nu$) is finite. The set of such probability measures, endowed with the metric W_p, is referred to as $\mathcal{W}^p(\Omega)$.

2.2 Extension to Radon measures [18]

Classic optimal transport only deals with probability measures, that is—up to a renormalization factor—positive measures with the same mass. In [18], Figalli and Gigli propose to extend Wasserstein distances to Radon measures supported on an open proper subset Ω of $\mathbb{R}^d$, whose boundary is denoted by $\partial\Omega$ (and $\overline{\Omega} := \Omega \sqcup \partial\Omega$), considering the following optimal *partial* transport problem:

Definition 2.1. *[18, Problem 1.1] Let $p \in [1, +\infty)$. Let μ, ν be two Radon measures supported on Ω satisfying $\int_\Omega d(x, \partial\Omega)^p \mathrm{d}\mu(x) < +\infty$ (resp. $\int_\Omega d(x, \partial\Omega)^p \mathrm{d}\nu(x) < +\infty$). The set of admissible transport plans (or couplings) $\mathrm{Adm}(\mu, \nu)$ is defined as the set of Radon measures π on $\overline{\Omega} \times \overline{\Omega}$ satisfying:*

$$\text{for all Borel sets } A, B \subset \Omega, \quad \pi(A, \overline{\Omega}) = \mu(A) \quad \text{and} \quad \pi(\overline{\Omega}, B) = \nu(B).$$

The cost of $\pi \in \mathrm{Adm}(\mu, \nu)$ is defined as

$$C_p(\pi) := \iint_{\overline{\Omega} \times \overline{\Omega}} d(x,y)^p \mathrm{d}\pi(x,y). \tag{7}$$

The distance $D_p(\mu, \nu)$ is then defined as

$$D_p(\mu, \nu) := \left(\inf_{\pi \in \mathrm{Adm}(\mu,\nu)} C_p(\pi) \right)^{1/p}. \tag{8}$$

Plans $\pi \in \mathrm{Adm}(\mu, \nu)$ realizing the infimum in (8) are called optimal. *The set of optimal transport plans is denoted by $\mathrm{Opt}_p(\mu, \nu)$.*

The following definition shows how to build an element of $\mathrm{Adm}(\mu, \nu)$ given a map $f : \overline{\Omega} \to \overline{\Omega}$ satisfying some balance condition.

Definition 2.2. *Let $\mu, \nu \in \mathcal{M}$. Consider $f : \overline{\Omega} \to \overline{\Omega}$ satisfying for all Borel set $B \subset \Omega$*

$$\mu(f^{-1}(B) \cap \Omega) + \nu(B \cap f(\partial\Omega)) = \nu(B). \tag{9}$$

Define for all Borel sets $A, B \subset \overline{\Omega}$,

$$\pi(A \times B) = \mu(f^{-1}(B) \cap \Omega \cap A) + \nu(\Omega \cap B \cap f(A \cap \partial\Omega)). \tag{10}$$

π is called the transport plan induced *by the* transport map *f.*

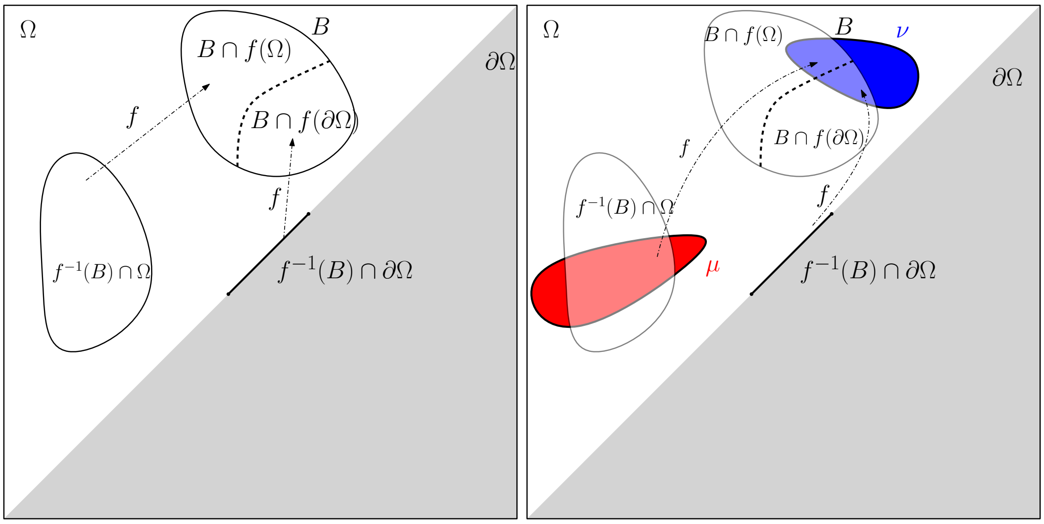

Figure 2: A transport map f must satisfy that the mass $\nu(B)$ (light blue) is the sum of the mass $\mu(f^{-1}(B) \cap \Omega)$ given by μ that is transported by f onto B (light red) and the mass $\nu(B \cap f(\partial\Omega))$ coming from $\partial\Omega$ and transported by f onto B.

One can easily check that we have indeed $\pi(A \times \overline{\Omega}) = \mu(A)$ and $\pi(\overline{\Omega} \times B) = \nu(B)$ for any Borel sets $A, B \subset \Omega$, so that $\pi \in \mathrm{Adm}(\mu, \nu)$ (see Figure 2).

Remark 2.1. *Since we have no constraints on $\pi(\partial\Omega \times \partial\Omega)$, one may always assume that a plan π satisfies $\pi(\partial\Omega \times \partial\Omega) = 0$, so that measures $\pi \in \mathrm{Adm}(\mu, \nu)$ are supported on*

$$E_\Omega := (\overline{\Omega} \times \overline{\Omega}) \backslash (\partial\Omega \times \partial\Omega). \tag{11}$$

Under the assumption that Ω is **bounded**, and assuming $p = 2$ (but the authors mention that their proofs work for any finite $p \geq 1$), it is proved in [18] that:

- [18, Theorem 2.2]: D_p is a indeed a distance over Radon measures supported on Ω.
- [18, Proposition 2.7]: Given $\mu, \mu_1, \mu_2 \dots$ Radon measures, we have

$$D_p(\mu_n, \mu) \to 0 \Leftrightarrow \begin{cases} \mu_n \xrightarrow{v} \mu, \\ \int_\Omega d(x, \partial\Omega)^p \mathrm{d}\mu_n(x) \to \int_\Omega d(x, \partial\Omega)^p \mathrm{d}\mu(x). \end{cases}$$

3 Structure of the persistence measures and diagrams spaces

For the remainder, we fix $\Omega := \{x = (t_1, t_2) \in \mathbb{R}^2,\ t_2 > t_1\}$ endowed with the euclidean metric $d : (x, y) \mapsto \|x - y\|_2$.

3.1 General properties of $\mathcal{M}^p$

It is assumed for now that $1 \leq p < \infty$. The case $p = \infty$ is studied in Section 3.3. The first proposition states preliminary results on the problem stated in Definition 2.1.

Proposition 3.1. *Let $\mu, \nu \in \mathcal{M}$. The set of transport plans $\mathrm{Adm}(\mu, \nu)$ is sequentially compact for the vague topology on E_Ω. Moreover, if $\mu, \nu \in \mathcal{M}^p$, for this topology,*

- *$\pi \in \mathrm{Adm}(\mu, \nu) \mapsto C_p(\pi)$ is lower semi-continuous.*
- *$\mathrm{Opt}_p(\mu, \nu)$ is a non-empty sequentially compact set.*
- *D_p is lower semi-continuous, in the sense that for sequences $(\mu_n)_n, (\nu_n)_n$ in $\mathcal{M}^p$ satisfying $\mu_n \xrightarrow{v} \mu$ and $\nu_n \xrightarrow{v} \nu$, we have $D_p(\mu, \nu) \leq \liminf_{n \to \infty} D_p(\mu_n, \nu_n)$.*

Moreover, D_p is a distance on $\mathcal{M}^p$.

These properties are mentioned in [18, pages 4-5] in the bounded case, and adapt straightforwardly to our framework. For the sake of completeness, we provide a detailed proof in Appendix B.

Remark 3.1. *If a (Borel) measure μ satisfies* $\mathrm{Pers}_p(\mu) < \infty$*, then for any Borel set* $A \subset \Omega$ *satisfying* $d(A, \partial\Omega) := \inf_{x\in A} d(x, \partial\Omega) > 0$*, we have:*

$$\mu(A) d(A, \partial\Omega)^p \leq \int_A d(x, \partial\Omega)^p \mathrm{d}\mu(x) \leq \int_\Omega d(x, \partial\Omega)^p \mathrm{d}\mu(x) = \mathrm{Pers}_p(\mu) < \infty, \tag{12}$$

so that $\mu(A) < \infty$*. In particular,* μ *is automatically a Radon measure.*

The following lemma gives a simple way to approximate a persistence measure (resp. diagram) with ones of finite masses.

Lemma 3.1. *Let* $\mu \in \mathcal{M}^p$*. Fix* $r > 0$*, and let* $A_r := \{x \in \Omega, d(x, \partial\Omega) \leq r\}$*. Let* $\mu^{(r)}$ *be the restriction of* μ *to* $\Omega\backslash A_r$*. Then* $D_p(\mu^{(r)}, \mu) \to 0$ *when* $r \to 0$*. Similarly, if* $a \in \mathcal{D}^p$*, we have* $d_p(a^{(r)}, a) \to 0$*.*

Proof. Let $\pi \in \mathrm{Adm}(\mu, \mu^{(r)})$ be the transport plan induced by the identity map on $\overline{\Omega}\backslash A_r$, and the projection onto $\partial\Omega$ on A_r. As π is sub-optimal, one has:

$$D_p^p(\mu, \mu^{(r)}) \leq C_p(\pi) = \int_{A_r} d(x, \partial\Omega)^p \mathrm{d}\mu(x) = \mathrm{Pers}_p(\mu) - \mathrm{Pers}_p(\mu^{(r)}).$$

Thus, by the monotone convergence theorem applied to μ with the functions $f_r : x \mapsto d(x, \partial\Omega)^p \cdot \mathbf{1}_{\Omega\backslash A_r}(x)$, $D_p(\mu, \mu^{(r)}) \to 0$ as $r \to 0$. Similar arguments show that $d_p(a^{(r)}, a) \to 0$ as $r \to 0$. $\square$

The following proposition is central in our work: it shows that the metrics D_p are extensions of the metrics d_p.

Proposition 3.2. *For* $a, b \in \mathcal{D}^p$*,* $D_p(a, b) = d_p(a, b)$*.*

Proof of 3.2. Let $a, b \in \mathcal{D}^p$ be two persistence diagrams. The case where a, b have a finite number of points is already treated in [28, Proposition 1].

In the general case, let $r > 0$. Due to (12), the diagrams $a^{(r)}$ and $b^{(r)}$ defined in Lemma 3.1 have a finite mass (thus finite number of points). Therefore, $d_p(a^{(r)}, b^{(r)}) = D_p(a^{(r)}, b^{(r)})$. By Lemma 3.1, the left hand side converges to $d_p(a, b)$ while the right hand side converges to $D_p(a, b)$, giving the conclusion. $\square$

Proposition 3.3. *The space* $(\mathcal{M}^p, D_p)$ *is a Polish space.*

As for Proposition 3.1, this proposition appears in [18, Proposition 2.7] in the bounded case, and its proof is straightforwardly adapted to our framework. For the sake of completeness, we provide a detailed proof in Appendix B.

We now state one of our main result: a characterization of convergence in $(\mathcal{M}^p, D_p)$.

Theorem 3.4. *Let* $\mu, \mu_1, \mu_2, \ldots$ *be measures in* $\mathcal{M}^p$*. Then,*

$$D_p(\mu_n, \mu) \to 0 \Leftrightarrow \begin{cases} \mu_n \xrightarrow{v} \mu, \\ \mathrm{Pers}_p(\mu_n) \to \mathrm{Pers}_p(\mu). \end{cases} \tag{13}$$

This result is analog to the characterization of convergence of probability measures in the Wasserstein space (see [39, Theorem 6.9]) and can be found in [18, Proposition 2.7] in the case where the ground space is bounded. While the proof of the direct implication can be adapted (it can be found in Appendix B), a new proof is needed for the converse implication.

Proof of the converse implication. Let $\mu, \mu_1, \mu_2 \ldots$ be elements of $\mathcal{M}^p$ and assume that $\mu_n \xrightarrow{v} \mu$ and $\mathrm{Pers}_p(\mu_n) \to \mathrm{Pers}_p(\mu)$. Since $D_p(\mu_n, \mu) \leq D_p(\mu_n, 0) + D_p(\mu, 0)$, the sequence $(D_p(\mu_n, \mu))_n$ is bounded. Thus, if we show that $(D_p(\mu_n, \mu))_n$ admits 0 as an unique accumulation point, then the convergence holds. Up to extracting a subsequence, we may assume that $(D_p(\mu_n, \mu))_n$ converges to some limit. Let $(\pi_n)_n \in \mathrm{Opt}(\mu_n, \mu)^{\mathbb{N}}$ be corresponding optimal transport plans. The vague convergence of $(\mu_n)_n$, together with Proposition A.1, imply that $(\pi_n)_n$ is relatively compact with respect to the vague convergence on E_Ω. Let π be the limit of any converging subsequence of $(\pi_n)_n$, which indexes are still denoted by n. As $\mu_n \xrightarrow{v} \mu$, standard arguments of optimal transport theory assert that π is necessarily in $\mathrm{Opt}_p(\mu, \mu)$ (see [18, Proposition 2.3]), i.e. π is supported on $\{(x, x), x \in \Omega\}$. The vague convergence of $(\mu_n)_n$ and the convergence of $(\mathrm{Pers}_p(\mu_n))_n$ to $\mathrm{Pers}_p(\mu)$ imply that for a given compact set $K \subset \Omega$, whose complementary set in Ω is denoted by K^c and interior set is denoted by $\mathring{K}$, we have

$$\begin{aligned}
\limsup_{n\to\infty} \int_{K^c} d(x, \partial\Omega)^p \mathrm{d}\mu_n(x) &= \limsup_{n\to\infty} \left(\mathrm{Pers}_p(\mu_n) - \int_K d(x, \partial\Omega)^p \mathrm{d}\mu_n(x) \right) \\
&\leq \mathrm{Pers}_p(\mu) - \int_{\mathring{K}} d(x, \partial\Omega)^p \mathrm{d}\mu(x) \text{ by Portmanteau theorem (see Appendix A)} \\
&= \int_{\overline{K^c}} d(x, \partial\Omega)^p \mathrm{d}\mu(x).
\end{aligned}$$

Therefore, for $\varepsilon > 0$, there exists some compact set $K \subset \Omega$, with

$$\limsup_n \int_{K^c} d(x, \partial\Omega)^p \mathrm{d}\mu_n(x) < \varepsilon \quad \text{and} \quad \int_{K^c} d(x, \partial\Omega)^p \mathrm{d}\mu(x) < \varepsilon. \tag{14}$$

For some compact set $K \subset \Omega$, consider the following transport plan $\tilde{\pi}_n$ (consider informally that what went from K to K^c and from K^c to K is now transported onto the diagonal, while everything else is unchanged):

$$\begin{cases} \tilde{\pi}_n = \pi_n & \text{on } K^2 \sqcup (K^c)^2, \\ \tilde{\pi}_n = 0 & \text{on } K \times K^c \sqcup K^c \times K, \\ \tilde{\pi}_n(A \times B) = \pi_n(A \times (K^c \sqcup B)) & \text{for } A \subset K,\ B \subset \partial\Omega, \\ \tilde{\pi}_n(A \times B) = \pi_n(A \times (K \sqcup B)) & \text{for } A \subset K^c,\ B \subset \partial\Omega, \\ \tilde{\pi}_n(A \times B) = \pi_n((K^c \sqcup A) \times B) & \text{for } A \subset \partial\Omega,\ B \subset K, \\ \tilde{\pi}_n(A \times B) = \pi_n((K \sqcup A) \times B) & \text{for } A \subset \partial\Omega,\ B \subset K^c. \end{cases} \tag{15}$$

Note that $\tilde{\pi}_n \in \mathrm{Adm}(\mu_n, \mu)$: for instance, for $A \subset K$ a Borel set, $\tilde{\pi}_n(A \times \overline{\Omega}) = \pi_n(A \times K) + \pi_n(A \times (K^c \sqcup \partial\Omega)) = \pi_n(A \times \overline{\Omega}) = \mu_n(A)$, and it is shown likewise that the other constraints are satisfied. As $\tilde{\pi}_n$ is suboptimal, $D_p^p(\mu_n, \mu) \leq \int_{\overline{\Omega}^2} d(x, y)^p \mathrm{d}\tilde{\pi}_n(x, y)$. The latter integral is equal to a sum of different terms, and we show that each of them converges to 0. Assume that the compact set K belongs to an increasing sequence $(K_m)_m$ of compact sets of Ω so that $\bigcup_m K_m = \Omega$ and $\pi(\partial(K_m \times K_m)) = 0$ for all m.

- We have $\iint_{K^2} d(x, y)^p \mathrm{d}\tilde{\pi}_n(x, y) = \iint_{K^2} d(x, y)^p \mathrm{d}\pi_n(x, y)$. The lim sup of the integral is smaller than $\iint_{K^2} d(x, y)^p \mathrm{d}\pi(x, y)$ by Portmanteau theorem (applied to the sequence $(d(x, y)^p \mathrm{d}\pi_n(x, y))_n$), and, recalling that π is supported on the diagonal of E_Ω, this integral is equal to 0.
- Figalli and Gigli [18, Proposition 2.3] show that an optimal transport plan, such as π_n, must be supported on $\{d(x, y)^p \leq d(x, \partial\Omega)^p + d(y, \partial\Omega)^p\}$. It follows that

$$\begin{aligned} \iint_{(K^c)^2} d(x, y)^p \mathrm{d}\tilde{\pi}_n(x, y) &= \iint_{(K^c)^2} d(x, y)^p \mathrm{d}\pi_n(x, y) \\ &\leq \int_{K^c} d(x, \partial\Omega)^p \mathrm{d}\mu_n(x) + \int_{K^c} d(y, \partial\Omega)^p \mathrm{d}\mu(y). \end{aligned}$$

  Taking the lim sup in n, and then letting K goes to Ω, this quantity converges to 0 by (14).
- We have $\iint_{K \times \partial\Omega} d(x, \partial\Omega)^p \mathrm{d}\tilde{\pi}_n(x, y) = \int_K d(x, \partial\Omega)^p \mathrm{d}\mu_n(x) - \iint_{K^2} d(x, \partial\Omega)^p \mathrm{d}\pi_n(x, y)$. By Portmanteau theorem applied to the sequence $(d(x, \partial\Omega)^p \mathrm{d}\mu_n(x))_n$), the lim sup of the first term is smaller than $\int_K d(x, \partial\Omega)^p \mathrm{d}\mu(x)$. Applying once again Portmanteau theorem on the second term (applied to the sequence $(d(x, y)^p \mathrm{d}\pi_n(x, y))_n$), and using that π is supported on the diagonal of E_Ω, the limsup of the second term is smaller than $-\iint_{K^2} d(x, \partial\Omega)^p \mathrm{d}\pi(x, y) = -\int_K d(x, \partial\Omega)^p \mathrm{d}\mu(x)$ (recall that $\pi(\partial(K \times K)) = 0$). Therefore, the limsup of the integral is equal to 0.
- The three remaining terms (corresponding to the three last lines of the definition (15)) are treated likewise this last case.

Finally, we have proven that $(D_p(\mu_n, \mu))_n$ is bounded and that for any converging subsequence, $\lim_k D_p(\mu_{n_k}, \mu) = 0$. It follows that $D_p(\mu_n, \mu) \to 0$. $\square$

Remark 3.2. *The assumption* $\mathrm{Pers}_p(\mu_n) \to \mathrm{Pers}_p(\mu)$ *is crucial to obtain* D_p*-convergence assuming vague convergence. For example, the sequence defined by* $\mu_n := \delta_{(n,n+1)}$ *converges vaguely to* $\mu = 0$ *and* $(\mathrm{Pers}_p(\mu_n))_n$ *does converge (as it is constant equal to* $\frac{\sqrt{2}}{2}$*), while* $D_p(\mu_n, 0) \nrightarrow 0$*. This does not contradict Theorem 3.4 since* $\mathrm{Pers}_p(\mu) = 0 \neq \frac{\sqrt{2}}{2} = \lim_n \mathrm{Pers}_p(\mu_n)$*.*

The direct implication of Theorem 3.4 implies in particular that the topology of the metric D_p is stronger than the vague topology. As a consequence, the following corollary holds, using Proposition A.5 ($\mathcal{D}^p$ is closed in $\mathcal{M}^p$ for the vague convergence).

Corollary 3.1. $\mathcal{D}^p$ *is closed in* $\mathcal{M}^p$ *for the metric* D_p*.*

We recover in particular that the space $(\mathcal{D}^p, d_p)$ is a Polish space (Proposition 3.3), a result already proved in [31, Theorems 7 and 12] with a different approach.

For a persistence measure $\mu \in \mathcal{M}^p$, let μ^p be the measure (with finite mass) defined by

$$\text{for } A \subset \Omega,\ \mu^p(A) := \int_A d(x, \partial\Omega)^p \mathrm{d}\mu(x). \tag{16}$$

Corollary 3.2. *Let $\mu, \mu_1, \mu_2, \dots$ be measures in $\mathcal{M}^p$. Then, $D_p(\mu_n, \mu) \to 0$ if and only if $(\mu_n^p)_n$ converges weakly*[2] *to μ^p.*

In particular, if $f : \Omega \to \mathbb{R}^d$ is a continuous bounded function, then the feature map $\Phi : \mathcal{M}^p \to \mathbb{R}^d$ defined by $\Phi : \mu \mapsto \mu^p(f) = \int_\Omega d(x, \partial\Omega)^p f(x) \mathrm{d}\mu(x)$ is continuous with respect to D_p.

Proof. Consider $\mu, \mu_1, \mu_2, \dots \in \mathcal{M}^p$ and assume that $D_p(\mu_n, \mu) \to 0$. By Theorem 3.4, $\mu_n \xrightarrow{v} \mu$ and $\mu_n^p(\Omega) \to \mu^p(\Omega)$. Since for any continuous function f compactly supported, the map $x \mapsto d(x, \partial\Omega)^p f(x)$ is also continuous and compactly supported, $\mu_n \xrightarrow{v} \mu$ implies in particular that $\mu_n^p \xrightarrow{v} \mu^p$. By Proposition A.3, the vague convergence along with the convergence of the masses imply $\mu_n^p \xrightarrow{w} \mu^p$. ☐

We end this section with a characterization of relatively compact sets in $(\mathcal{M}^p, D_p)$.

Proposition 3.5. *A set F is relatively compact in $(\mathcal{M}^p, D_p)$ if and only if the set $\{\mu^p, \mu \in F\}$ is tight and $\sup_{\mu \in F} \mathrm{Pers}_p(\mu) < \infty$.*

Proof. From Corollary 3.2, the relative compactness of a set $F \subset \mathcal{M}^p$ for the metric D_p is equivalent to the relative compactness of the set $\{\mu^p,\ \mu \in F\}$ for the weak convergence. Recall that all μ^p have a finite mass, as $\mu^p(\Omega) = \mathrm{Pers}_p(\mu) < \infty$. Therefore, one can use Prokhorov's theorem (Proposition A.2) to conclude. ☐

Remark 3.3. *This characterization is equivalent to the one described in [31, Theorem 21] for persistence diagrams. The notions introduces by the authors of off-diagonally birth-death boundedness, and uniformness are rephrased using the notion of tightness, standard in measure theory.*

3.2 Persistence measures in the finite setting

In practice, many statistical results regarding persistence diagrams are stated for sets of diagrams with uniformly bounded number of points [27, 9], and the specific properties of D_p in this setting are therefore of interest. Introduce for $m \geq 0$ the subset $\mathcal{M}^p_{\leq m}$ of $\mathcal{M}^p$ defined as $\mathcal{M}^p_{\leq m} := \{\mu \in \mathcal{M}^p,\ \mu(\Omega) \leq m\}$, and the set $\mathcal{M}^p_f$ of finite persistence measures, $\mathcal{M}^p_f := \bigcup_{m \geq 0} \mathcal{M}^p_{\leq m}$. Define similarly the set $\mathcal{D}_{\leq m}$ (resp. $\mathcal{D}_f$).

Proposition 3.6. *$\mathcal{M}^p_f$ (resp $\mathcal{D}_f$) is dense in $\mathcal{M}^p$ (resp $\mathcal{D}^p$) for the metric D_p.*

Proof. This is a straightforward consequence of Lemma 3.1. ☐

Let $\tilde{\Omega} = \Omega \sqcup \{\partial\Omega\}$ be the quotient of $\overline{\Omega}$ by $\partial\Omega$—i.e. we encode the diagonal by just one point. The distance d on $\overline{\Omega}^2$ induces naturally a function (still denoted by d) on $\tilde{\Omega}^2$. However, d is not a distance since one can have $d(x, y) > d(x, \partial\Omega) + d(y, \partial\Omega)$. Define

$$\rho(x, y) := \min\{d(x, y), d(x, \partial\Omega) + d(y, \partial\Omega)\}. \tag{17}$$

It is straightforward to check that ρ is a distance on $\tilde{\Omega}$ and that $(\tilde{\Omega}, \rho)$ is a Polish space. One can then define the Wasserstein distance $W_{p,\rho}$ with respect to ρ for finite measures on $\tilde{\Omega}$ which have the same masses, that is the infimum of $\tilde{C}_p(\tilde{\pi}) := \iint_{\tilde{\Omega}^2} \rho(x, y)^p \mathrm{d}\tilde{\pi}(x, y)$, for $\tilde{\pi}$ a transport plan with corresponding marginals (see Section 2.1). The following theorem states that computing the D_p metric between two persistence measures with finite masses can be turn into computing the Wasserstein distances between two measures supported on $\tilde{\Omega}$ with the same (finite) masses.

Theorem 3.7. *Let $\mu, \nu \in \mathcal{M}^p_f$ and $r \geq \mu(\Omega) + \nu(\Omega)$. Define $\tilde{\mu} = \mu + (r - \mu(\Omega))\delta_{\partial\Omega}$ and $\tilde{\nu} = \nu + (r - \nu(\Omega))\delta_{\partial\Omega}$. Then $D_p(\mu, \nu) = W_{p,\rho}(\tilde{\mu}, \tilde{\nu})$.*

Proof. We first introduce a lemma that explicits correspondences between $\mathrm{Adm}(\mu, \nu)$ and $\Pi(\tilde{\mu}, \tilde{\nu})$.

Lemma 3.2. *Let $\mu, \nu \in \mathcal{M}^p_f$ and $r \geq \max(\mu(\Omega), \nu(\Omega))$. Let $\tilde{\mu} := \mu + (r - \mu(\Omega))\delta_{\partial\Omega}$, $\tilde{\nu} := \nu + (r - |\nu|)\delta_{\partial\Omega}$ and $s : \Omega \to \partial\Omega$ be the orthogonal projection on the diagonal.*

1. *Define $T(\mu, \nu)$ the set of plans $\pi \in \mathrm{Adm}(\mu, \nu)$ satisfying $\pi(\{(x, y) \in \Omega \times \partial\Omega,\ y \neq s(x)\}) = \pi(\{(x, y) \in \partial\Omega \times \Omega,\ x \neq s(y)\}) = 0$ along with $\pi(\partial\Omega \times \partial\Omega) = 0$. Then, $\mathrm{Opt}_p(\mu, \nu) \subset T(\mu, \nu)$.*
2. *Let $\pi \in T(\mu, \nu)$ be such that $\mu(\Omega) + \pi(\partial\Omega \times \Omega) \leq r$. Define $\iota(\pi) \in \Pi(\tilde{\mu}, \tilde{\nu})$ by, for Borel sets $A, B \subset \Omega$,*

$$\begin{cases} \iota(\pi)(A \times B) = \pi(A \times B), \\ \iota(\pi)(A \times \{\partial\Omega\}) = \pi(A \times \partial\Omega), \\ \iota(\pi)(\{\partial\Omega\} \times B) = \pi(\partial\Omega \times B), \\ \iota(\pi)(\{\partial\Omega\} \times \{\partial\Omega\}) = r - \mu(\Omega) - \pi(\partial\Omega \times \Omega) \geq 0. \end{cases} \tag{18}$$

Then, $C_p(\pi) = \iint_{\tilde{\Omega} \times \tilde{\Omega}} d(x, y)^p \mathrm{d}\iota(\pi)(x, y)$.

[2] The weak convergence $\xrightarrow{w}$ is defined in Appendix A.

3. *Let $\tilde\pi \in \Pi(\tilde\mu,\tilde\nu)$. Define $\kappa(\tilde\pi) \in T(\mu,\nu)$ by,*

$$\begin{cases} \kappa(\tilde\pi)(A\times B) = \tilde\pi(A\times B) & \text{for Borel sets } A,B\subset\Omega,\\ \kappa(\tilde\pi)(A\times B) = \tilde\pi((A\cap s^{-1}(B))\times\{\partial\Omega\}) & \text{for Borel sets } A\subset\Omega, B\subset\partial\Omega,\\ \kappa(\tilde\pi)(A\times B) = \tilde\pi(\{\partial\Omega\}\times(B\cap s^{-1}(A))) & \text{for Borel sets } A\subset\partial\Omega, B\subset\Omega,\\ \kappa(\tilde\pi)(\partial\Omega,\partial\Omega) = 0. \end{cases}$$

Then, $\iint_{\tilde\Omega\times\tilde\Omega} d(x,y)^p \mathrm{d}\tilde\pi(x,y) = C_p(\kappa(\tilde\pi))$.

Proof.

1. Consider $\pi \in \mathrm{Adm}(\mu,\nu)$, and define π' that coincides with π on $\Omega\times\Omega$, and is such that we enforce mass transported on the diagonal to be transported on its orthogonal projection: more precisely, for all Borel set $A\subset\Omega$, $B\subset\partial\Omega$, $\pi'(A\times B) = \pi((s^{-1}(B)\cap A)\times B)$ and $\pi'(B\times A) = \pi(B\times(s^{-1}(B)\cap A))$. Note that $\pi' \in T(\mu,\nu)$. Since $s(x)$ is the unique minimizer of $y\mapsto d(x,y)^p$, it follows that $C_p(\pi')\le C_p(\pi)$, with equality if and only if $\pi\in T(\mu,\nu)$, and thus $\mathrm{Opt}_p(\mu,\nu)\subset T(\mu,\nu)$.

2. Write $\tilde\pi = \iota(\pi)$. The mass $\tilde\pi(\{\partial\Omega\}\times\{\partial\Omega\})$ is nonnegative by definition. One has for all Borel sets $A\subset\Omega$,

$$\begin{aligned}\tilde\pi(A\times\tilde\Omega) &= \tilde\pi(A\times\Omega) + \tilde\pi(A\times\{\partial\Omega\})\\ &= \pi(A\times\Omega)+\pi(A\times\partial\Omega) = \pi(A\times\overline{\Omega}) = \mu(A) = \tilde\mu(A).\end{aligned}$$

Similarly, $\tilde\pi(\tilde\Omega\times B) = \tilde\nu(B)$ for all $B\subset\Omega$. Observe also that

$$\tilde\pi(\{\partial\Omega\}\times\tilde\Omega) = \tilde\pi(\{\partial\Omega\}\times\{\partial\Omega\}) + \tilde\pi(\{\partial\Omega\}\times\Omega) = r-\mu(\Omega) = \tilde\mu(\{\partial\Omega\}).$$

Similarly, $\tilde\pi(\tilde\Omega\times\{\partial\Omega\}) = \tilde\nu(\{\partial\Omega\})$. It gives that $\iota(\pi)\in\Pi(\tilde\mu,\tilde\nu)$, so that ι is well defined. Observe that

$$\begin{aligned}\iint_{\tilde\Omega\times\tilde\Omega} d(x,y)^p\mathrm{d}\tilde\pi(x,y) &= \iint_{\Omega\times\Omega} d(x,y)^p\mathrm{d}\pi(x,y)\\ &\quad+\int_\Omega d(x,\partial\Omega)^p\mathrm{d}\pi(x,\partial\Omega) + \int_\Omega d(\partial\Omega,y)^p\mathrm{d}\pi(\partial\Omega,y)+0\\ &= C_p(\pi) \text{ as } \pi\in T(\mu,\nu).\end{aligned}$$

3. Write $\pi = \kappa(\tilde\pi)$. For $A\subset\Omega$ a Borel set,

$$\begin{aligned}\pi(A\times\overline{\Omega}) &= \pi(A\times\Omega)+\pi(A\times\partial\Omega)\\ &= \tilde\pi(A\times\Omega)+\tilde\pi(A\times\{\partial\Omega\}) = \tilde\pi(A\times\tilde\Omega) = \mu(A).\end{aligned}$$

Similarly, $\pi(\overline{\Omega}\times B) = \nu(B)$ for all $B\subset\Omega$. Therefore, $\pi\in\mathrm{Adm}(\mu,\nu)$, and by construction, if a point $x\in\Omega$ is transported on $\partial\Omega$, it cannot be transported elsewhere than on $s(x)$, so that $\pi\in T(\mu,\nu)$. Observe that $\mu(\Omega)+\pi(\partial\Omega\times\Omega)\le\tilde\pi(\tilde\Omega\times\tilde\Omega) = r$, so that $\iota(\pi)$ is well defined. Also, $\iota(\pi)=\tilde\pi$, so that, according to point 2, $C_p(\pi) = \iint_{\tilde\Omega\times\tilde\Omega} d(x,y)^p\mathrm{d}\tilde\pi(x,y)$.

□

We show that the inequality $D_p(\mu,\nu)\le W_{p,\rho}(\tilde\mu,\tilde\nu)$ holds as long as $r\ge\max(\mu(\Omega),\nu(\Omega))$.

Lemma 3.3. *Let $\mu,\nu\in\mathcal{M}_f^p$ and $r\ge\max(\mu(\Omega),\nu(\Omega))$. Let $\tilde\mu := \mu+(r-\mu(\Omega))\delta_{\partial\Omega}$, $\tilde\nu := \nu+(r-|\nu|)\delta_{\partial\Omega}$. Then, $D_p(\mu,\nu)\le W_{p,\rho}(\tilde\mu,\tilde\nu)$.*

Proof. Let $\tilde\pi\in\Pi(\tilde\mu,\tilde\nu)$. Define the set $H := \{(x,y)\in\tilde\Omega^2,\ \rho(x,y) = d(x,y)\}$, and let H^c be its complementary set in $\tilde\Omega^2$, i.e. the set where $\rho(x,y) = d(x,\partial\Omega)+d(\partial\Omega,y)$. Define $\tilde\pi'\in\mathcal{M}(\tilde\Omega^2)$ by, for Borel sets $A,B\subset\Omega$:

$$\begin{cases}\tilde\pi'(A\times B) = \tilde\pi((A\times B)\cap H)\\ \tilde\pi'(A\times\{\partial\Omega\}) = \tilde\pi((A\times\tilde\Omega)\cap H^c)+\tilde\pi(A\times\{\partial\Omega\})\\ \tilde\pi'(\{\partial\Omega\}\times B) = \tilde\pi((\tilde\Omega\times B)\cap H^c)+\tilde\pi(\{\partial\Omega\}\times B).\end{cases}$$

We easily check that $\tilde\pi'\in\Pi(\tilde\mu,\tilde\nu)$. Also, using $(a+b)^p\ge a^p+b^p$ for positive a,b, we have

$$\begin{aligned}\iint_{\tilde\Omega\times\tilde\Omega}\rho(x,y)^p\mathrm{d}\tilde\pi(x,y) &= \iint_H d(x,y)^p\mathrm{d}\tilde\pi(x,y) + \iint_{H^c}(d(x,\partial\Omega)+d(\partial\Omega,y))^p\mathrm{d}\tilde\pi(x,y)\\ &\ge \iint_H d(x,y)^p\mathrm{d}\tilde\pi'(x,y) + \iint_{H^c}\left(d(x,\partial\Omega)^p+d(y,\partial\Omega)^p\right)\mathrm{d}\tilde\pi(x,y)\\ &= \iint_{\tilde\Omega\times\tilde\Omega} d(x,y)^p\mathrm{d}\tilde\pi'(x,y)\\ &\ge \inf_{\tilde\pi'\in\Pi(\tilde\mu,\tilde\nu)}\iint_{\tilde\Omega\times\tilde\Omega} d(x,y)^p\mathrm{d}\tilde\pi'(x,y).\end{aligned}$$

We conclude by taking the infimum on $\tilde{\pi}$ that $W_{p,\rho}(\tilde{\mu}, \tilde{\nu}) \geq \inf_{\tilde{\pi}' \in \Pi(\tilde{\mu}, \tilde{\nu})} \iint_{\tilde{\Omega} \times \tilde{\Omega}} d(x,y)^p \mathrm{d}\tilde{\pi}'(x,y)$. Since $\rho(x,y) \leq d(x,y)$, it follows that

$$W^p_{p,\rho}(\tilde{\mu}, \tilde{\nu}) = \inf_{\tilde{\pi} \in \Pi(\tilde{\mu}, \tilde{\nu})} \iint_{\tilde{\Omega}^2} d(x,y)^p \mathrm{d}\tilde{\pi}(x,y). \tag{19}$$

Since d is continuous, the infimum in the right hand side of (19) is reached [39, Theorem 4.1]. Consider thus $\tilde{\pi} \in \Pi(\tilde{\mu}, \tilde{\nu})$ which realizes the infimum. We can write, using Lemma 3.2,

$$\begin{aligned} W^p_{p,\rho}(\tilde{\mu}, \tilde{\nu}) &= \iint_{\tilde{\Omega}^2} d(x,y)^p \mathrm{d}\tilde{\pi}(x,y) = \iint_{\overline{\Omega} \times \overline{\Omega}} d(x,y)^p \mathrm{d}\kappa(\tilde{\pi})(x,y) \\ &\geq \inf_{\pi \in T(\mu,\nu)} \iint_{\overline{\Omega} \times \overline{\Omega}} d(x,y)^p \mathrm{d}\pi(x,y) = D^p_p(\mu, \nu), \end{aligned}$$

which concludes the proof. □

We can now prove the theorem. Let $\pi \in T(\mu, \nu)$. As $\mu(\Omega) + \pi(\partial\Omega \times \Omega) \leq \mu(\Omega) + \nu(\Omega) \leq r$, one can define $\tilde{\pi} = \iota(\pi)$. Since $\rho(x,y) \leq d(x,y)$, we have $\tilde{C}_p(\tilde{\pi}) \leq \iint d(x,y)^p \mathrm{d}\tilde{\pi}(x,y) = C_p(\pi)$ (Lemma 3.2). Taking infimum gives $W_{p,\rho}(\tilde{\mu}, \tilde{\nu}) \leq D_p(\mu, \nu)$. The other inequality holds according to Lemma 3.3.

□

Theorem 3.7 is useful for numerical purpose since it allows in applications, when dealing with a finite set of diagrams, to directly use the various tools developed in computational optimal transport [32] to compute Wasserstein distances. This alternative to the combinatorial algorithms considered in the literature [24, 36] is studied in details in [28]. This result is also useful to prove the existence of barycenters of sets of persistence measures in the finite setting (see Section 4).

3.3 The D_∞ distance

In classical optimal transport, the ∞-Wasserstein distance is known to have a much more erratic behavior than its $p < \infty$ counterparts [33, Section 5.5.1]. However, in the framework of persistence diagrams, the d_∞ distance (2) appears naturally as an interleaving distance between persistence modules and satisfies strong stability results: it is thus worth of interest. It also happens that, when restricted to diagrams having some specific finiteness properties, most irregular behaviors are suppressed and a convenient characterization of convergence exists.

Definition 3.1. *Let* $\mathrm{spt}(\mu)$ *denotes the support of a measure* μ *and* $\mathrm{Pers}_\infty(\mu) := \sup\{d(x, \partial\Omega),\ x \in \mathrm{spt}(\mu)\}$. *Let*

$$\mathcal{M}^\infty := \{\mu \in \mathcal{M},\ \mathrm{Pers}_\infty(\mu) < \infty\} \quad \text{and} \quad \mathcal{D}^\infty := \mathcal{D} \cap \mathcal{M}^\infty. \tag{20}$$

For $\mu, \nu \in \mathcal{M}^\infty$ *and* $\pi \in \mathrm{Adm}(\mu, \nu)$*, let* $C_\infty(\pi) := \sup\{d(x,y),\ (x,y) \in \mathrm{spt}(\pi)\}$ *and let*

$$D_\infty(\mu, \nu) := \inf_{\pi \in \mathrm{Adm}(\mu,\nu)} C_\infty(\pi). \tag{21}$$

The set of transport plans minimizing (21) *is denoted by* $\mathrm{Opt}_\infty(\mu, \nu)$.

Proposition 3.8. *Let* $\mu, \nu \in \mathcal{M}^\infty$. *For the vague topology on* E_Ω,

- *the map* $\pi \in \mathrm{Adm}(\mu, \nu) \mapsto C_\infty(\pi)$ *is lower semi-continuous.*
- *The set* $\mathrm{Opt}_\infty(\mu, \nu)$ *is a non-empty sequentially compact set.*
- D_∞ *is lower semi-continuous.*

Moreover, D_∞ *is a distance on* $\mathcal{M}^\infty$.

The proofs of these results are found in Appendix B. As in the case $p < \infty$, we prove that D_∞ is an extension of d_∞.

Proposition 3.9. *For* $a, b \in \mathcal{D}^\infty$, $D_\infty(a,b) = d_\infty(a,b)$.

Proof. Consider two diagrams $a, b \in \mathcal{D}^\infty$, written as $a = \sum_{i \in I} \delta_{x_i}$ and $b = \sum_{j \in J} \delta_{y_j}$, where $I, J \subset \mathbb{N}^*$ are (potentially infinite) sets of indices. The marginal constraints imply that a plan $\pi \in \mathrm{Adm}(\mu, \nu)$ is supported on $(\{x_i\}_i \cup \partial\Omega) \times (\{y_j\}_j \cup \partial\Omega)$, and the cost of such a plan can always be reduced if some of the mass $\pi(\{x_i\}, \partial\Omega)$

(resp. $\pi(\partial\Omega, \{y_j\})$) is sent on another point than the projection of x_i (resp. y_j) on the diagonal $\partial\Omega$. Introduce the matrix C indexed on $(-J \cup I) \times (-I \cup J)$ defined by

$$\begin{cases} C_{i,j} = d(x_i, y_j) & \text{for } i, j > 0, \\ C_{i,j} = d(\partial\Omega, y_j) & \text{for } i < 0, j > 0, \\ C_{i,j} = d(x_i, \partial\Omega) & \text{for } i > 0, j < 0, \\ C_{i,j} = 0 & \text{for } i, j < 0. \end{cases} \tag{22}$$

Then, $D_\infty(a, b) = \inf_{P \in \mathcal{S}} \sup\{C_{i,j},\ (i,j) \in \mathrm{spt}(P)\}$, where $\mathcal{S}$ is the set of doubly stochastic matrices indexed on $(-J \cup I) \times (-I \cup J)$, and $\mathrm{spt}(P)$ denotes the support of P, that is the set $\{(i,j),\ P_{i,j} > 0\}$.

Let $P \in \mathcal{S}$. For any $k \in \mathbb{N}$, and any set of distinct indices $\{i_1, \dots, i_k\} \subset I$, we have

$$k = \sum_{k'=1}^{k} \underbrace{\sum_{j \in J} P_{i_{k'},j}}_{=1} = \sum_{j \in J} \underbrace{\sum_{k'=1}^{k} P_{i_{k'},j}}_{\leq 1}.$$

Thus, the cardinality of $\{j,\ \exists k' \text{ such that } (i_{k'}, j) \in \mathrm{spt}(P)\}$ must be larger than k. Under such conditions, the Hall's marriage theorem (see [20, p. 51]) guarantees the existence of a permutation matrix P' with $\mathrm{spt}(P') \subset \mathrm{spt}(P)$. As a consequence,

$$\begin{aligned} \sup\{C_{i,j},\ (i,j) \in \mathrm{spt}(P)\} &\geq \sup\{C_{i,j},\ (i,j) \in \mathrm{spt}(P')\} \\ &\geq \inf_{P' \in \mathcal{S}'} \sup\{C_{i,j},\ (i,j) \in \mathrm{spt}(P')\} = d_\infty(a, b), \end{aligned}$$

where $\mathcal{S}'$ denotes the set of permutations matrix indexed on $(-J \cup I) \times (-I \cup J)$. Taking the infimum on $P \in \mathcal{S}$ on the left-hand side and using that $\mathcal{S}' \subset \mathcal{S}$ finally gives that $D_\infty(a, b) = d_\infty(a, b)$. □

Proposition 3.10. *The space $(\mathcal{M}^\infty, D_\infty)$ is complete.*

A proof can be found in Appendix B. Note that contrary to the case $p < \infty$, the space $\mathcal{D}^\infty$ (and therefore $\mathcal{M}^\infty$) is not separable. Indeed, for $I \subset \mathbb{N}$, define the diagram $a_I := \sum_{i \in I} \delta_{(i,i+1)} \in \mathcal{D}^\infty$. The family $\{a_I,\ I \subset \mathbb{N}\}$ is uncountable, and for two distinct I, I', $D_\infty(a_I, a_{I'}) = \frac{\sqrt{2}}{2}$. This result is similar to [6, Theorem 4.20].

We now show that the direct implication in Theorem 3.4 still holds in the case $p = \infty$.

Proposition 3.11. *Let $\mu, \mu_1, \mu_2, \dots$ be measures in $\mathcal{M}^\infty$. If $D_\infty(\mu_n, \mu) \to 0$, then $(\mu_n)_n$ converges vaguely to μ.*

The proof is delayed to Appendix B.

Remark 3.4. *As for the case $1 \leq p < \infty$, Proposition 3.11 implies that D_∞ metricizes the vague convergence, and thus using Proposition 3.9, we have that $(\mathcal{D}^\infty, d_\infty)$ is closed in $(\mathcal{M}^\infty, D_\infty)$ and is—in particular—complete.*

Contrary to the $p < \infty$ case, a converse of Proposition 3.11 does not hold, even on the subspace of persistence diagrams (see Figure 3). To recover a space with a structure more similar to $\mathcal{D}^p$, it is useful to look at a smaller set. Introduce $\mathcal{D}_0^\infty$ the set of persistence diagrams such that for all $r > 0$, there is a finite number of points of the diagram of persistence larger than r and recall that $\mathcal{D}_f$ denotes the set of persistence diagrams with finite number of points.

Proposition 3.12. *The closure of $\mathcal{D}_f$ for the distance D_∞ is $\mathcal{D}_0^\infty$.*

Proof. Consider $a \in \mathcal{D}_0^\infty$. By definition, for all $n \in \mathbb{N}$, a has a finite number of points with persistence larger than $\frac{1}{n}$, so that the restriction a_n of a to points with persistence larger than $\frac{1}{n}$ belongs to $\mathcal{D}_f$. As $D_\infty(a, a_n) \leq \frac{1}{n} \to 0$, $\mathcal{D}_0^\infty$ is contained in the closure of $\mathcal{D}_f$.

Conversely, consider a diagram $a \in \mathcal{D}^\infty \backslash \mathcal{D}_0^\infty$, that is there is a constant $r > 0$ such that a has infinitely many points with persistence larger than r. For any finite diagram $a' \in \mathcal{D}_f$, we have $D_\infty(a', a) \geq r$, so that a is not the limit for the D_∞ metric of any sequence in $\mathcal{D}_f$. □

Remark 3.5. *The space $\mathcal{D}_0^\infty$ is exactly the set introduced in [3, Theorem 3.5] as the completion of $\mathcal{D}_f$ for the bottleneck metric d_∞. Here, we recover that $\mathcal{D}_0^\infty$ is complete as a closed subset of the complete space $\mathcal{D}^\infty$.*

Define for $r > 0$ and $a \in \mathcal{D}$, $a^{(r)}$ the persistence diagram restricted to $\{x \in \Omega,\ d(x, \partial\Omega) > r\}$. The following characterization of convergence holds in $\mathcal{D}_0^\infty$.

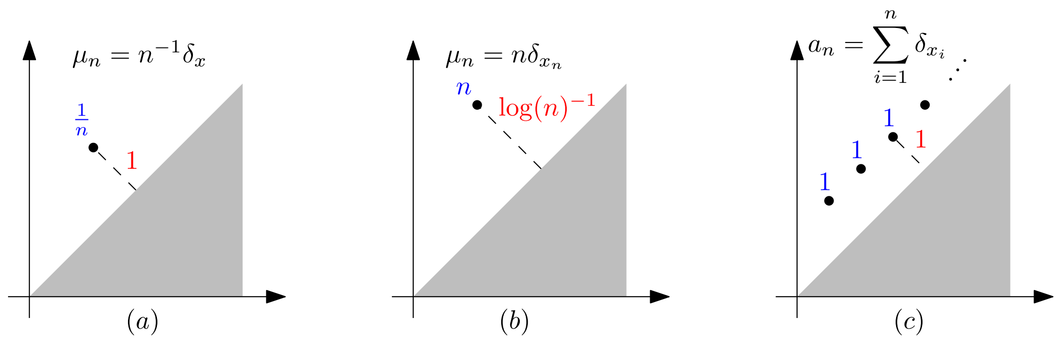

Figure 3: Illustration of differences between D_p, D_∞, and vague convergences. Blue color represents the mass on a point while red color designates distances. (a) A case where $D_p(\mu_n, 0) \to 0$ for any $p < \infty$ while $D_\infty(\mu_n, 0) = 1$. (b) A case where $D_\infty(\mu_n, 0) \to 0$ while for all $p < \infty$, $D_p(\mu_n, \mu) \to \infty$. (c) A sequence of persistence diagrams $a_n \in \mathcal{D}^\infty$, where $(a_n)_n$ converges vaguely to $a = \sum_i \delta_{x_i}$ and $\mathrm{Pers}_\infty(a_n) = \mathrm{Pers}_\infty(a)$, but (a_n) does not converge to a for D_∞.

Proposition 3.13. *Let $a, a_1, a_2, \dots$ be persistence diagrams in $\mathcal{D}_0^\infty$. Then,*

$$D_\infty(a_n, a) \to 0 \Leftrightarrow \begin{cases} a_n \xrightarrow{v} a, \\ (a_n^{(r)})_n \text{ is tight for all positive } r. \end{cases}$$

Proof of Proposition 3.13. Let us prove first the direct implication. Proposition 3.11 states that the convergence with respect to D_∞ implies the vague convergence. Fix $r > 0$. By definition, $a^{(r)}$ is made of a finite number of points, all included in some open bounded set $U \subset \Omega$. As $a_n^{(r)}(U^c)$ is a sequence of integers, the bottleneck convergence implies that for n large enough, $a_n^{(r)}(U^c)$ is equal to 0. Thus, $(a_n^{(r)})_n$ is tight.

Let us prove the converse. Consider $a \in \mathcal{D}_0^\infty$ and a sequence $(a_n)_n$ that converges vaguely to a, with $(a_n^{(r)})$ tight for all $r > 0$. Fix $r > 0$ and let $x_1, \dots, x_K$ be an enumeration of the points in $a^{(r)}$, the point x_k being present with multiplicity $m_k \in \mathbb{N}$. Denote by $B(x, \varepsilon)$ (resp. $\overline{B}(x, \varepsilon)$) the open (resp. closed) ball of radius ε centered at x. By Portmanteau theorem, for ε small enough,

$$\begin{cases} \liminf_{n\to\infty} a_n(B(x_k, \varepsilon)) \geq a(B(x_k, \varepsilon)) = m_k \\ \limsup_{n\to\infty} a_n(\overline{B}(x_k, \varepsilon)) \leq a(\overline{B}(x_k, \varepsilon)) = m_k, \end{cases}$$

so that, for n large enough, there are exactly m_k points of a_n in $B(x_k, \varepsilon)$ (since $(a_n(B(x_k, \varepsilon)))_n$ is a converging sequence of integers). The tightness of $(a_n^{(r)})_n$ implies the existence of some compact $K \subset \Omega$ such that for n large enough, $a_n^{(r)}(K^c) = 0$ (as the measures take their values in $\mathbb{N}$). Applying Portmanteau's theorem to the closed set $K' := K \backslash \{B(x_i, \varepsilon),\ i = 1 \cdots K\}$ gives

$$\limsup_{n\to\infty} a_n^{(r)}(K') \leq a^{(r)}(K') = 0.$$

This implies that for n large enough, there are no other points in a_n with persistence larger than r and thus $D_\infty(a^{(r)}, a_n)$ is smaller than $r + \varepsilon$. Finally,

$$\limsup_{n\to\infty} D_\infty(a_n, a) \leq \limsup_{n\to\infty} D_\infty(a_n, a^{(r)}) + r \leq 2r + \varepsilon.$$

Letting $\varepsilon \to 0$ then $r \to 0$, the bottleneck convergence holds. □

4 Barycenters for distributions supported on $\mathcal{M}^p$

In this section, we state the existence of barycenters for probability distributions supported on $\mathcal{M}^p$, starting with the finite case and extending to any probability distribution with finite p-th moment. We then study the specific case of persistence diagrams, showing that the barycenter of a probability distribution on $\mathcal{D}^p$ exists *in* $\mathcal{D}^p$. We suppose that $p > 1$ (the case $p = 1$ would correspond to a notion of *median*).

Recall that $(\mathcal{M}^p, D_p)$ is Polish, and let W_p denote the Wasserstein distance between probability measures supported on $(\mathcal{M}^p, D_p)$ (see Section 2). We denote by $\mathcal{W}^p(\mathcal{M}^p)$ the space of probability measures $\mathbb{P}$ supported on $\mathcal{M}^p$, endowed with the W_p metric, which are at a finite distance from δ_0—the dirac mass supported on the empty diagram—i.e. $\int_{\nu\in\mathcal{M}^p} D_p^p(\nu, 0) \mathrm{d}\mathbb{P}(\nu) = W_p^p(\mathbb{P}, \delta_0) < \infty$.

Definition 4.1. *Consider $\mathbb{P} \in \mathcal{W}^p(\mathcal{M}^p)$. A measure $\mu^* \in \mathcal{M}^p$ is a* barycenter *of $\mathbb{P}$ if it minimizes $\mathcal{E} : \mu \in \mathcal{M}^p \mapsto \int_{\nu\in\mathcal{M}^p} D_p^p(\mu, \nu) \mathrm{d}\mathbb{P}(\mu)$.*

4.1 Barycenters in the finite case

Let $\mathbb{P}$ be of the form $\sum_{i=1}^N \lambda_i \delta_{\mu_i}$ with $N \in \mathbb{N}$, μ_i a persistence measure of finite mass m_i, and $(\lambda_i)_i$ are non-negative weights that sum to 1. Define $m_{\text{tot}} := \sum_{i=1}^N m_i$. To prove the existence of barycenters for such a $\mathbb{P}$, we show that, in this case, barycenters correspond to barycenters of some distribution on $\mathcal{M}^p_{m_{\text{tot}}}(\tilde{\Omega})$, the sets of measures on $\tilde{\Omega}$ that all have the same mass m_{tot} (see Section 3.2), a problem well studied in the literature [2, 7, 8].

Lemma 4.1. *We have* $\inf\{\mathcal{E}(\mu),\ \mu \in \mathcal{M}^p\} = \inf\{\mathcal{E}(\mu), \mu \in \mathcal{M}^p_{\leq m_{\text{tot}}}\}$.

Proof. The idea of the proof is to show that if a measure μ has some mass that is mapped to the diagonal in each transport plan between μ and μ_i, then we can build a measure μ' by "removing" this mass, and then observe that such a measure μ' has a smaller energy.

Let thus $\mu \in \mathcal{M}^p$. Let $\pi_i \in \mathrm{Opt}_p(\mu_i, \mu)$ for $i = 1, \dots, N$. The measure $A \subset \Omega \mapsto \pi_i(\partial\Omega \times A)$ is absolutely continuous with respect to μ. It therefore has a density f_i with respect to μ. Define for $A \subset \Omega$ a Borel set,

$$\mu'(A) := \mu(A) - \int_A \min_j f_j(x) \mathrm{d}\mu(x),$$

and, for $i = 1, \dots, N$, a measure π'_i, equal to π_i on $\Omega \times \overline{\Omega}$ and which satisfies for $A \subset \Omega$ a Borel set,

$$\pi'_i(\partial\Omega \times A) = \pi'_i(s(A) \times A) := \pi_i(\partial\Omega \times A) - \int_A \min_j f_j(x) \mathrm{d}\mu(x),$$

where s is the orthogonal projection on $\partial\Omega$. As $\pi_i(\partial\Omega \times A) = \int_A f_i(x)\mathrm{d}\mu(x)$, $\pi'_i(A)$ is nonnegative, and as $\pi_i(\partial\Omega \times A) \leq \mu(A)$, $\mu'(A)$ is nonnegative. To prove that $\pi'_i \in \mathrm{Adm}(\mu_i, \mu')$, it is enough to check that for $A \subset \Omega$, $\pi'_i(\overline{\Omega} \times A) = \mu'(A)$:

$$\begin{aligned}
\pi'_i(\overline{\Omega} \times A) &= \pi_i(\Omega \times A) + \pi_i(\partial\Omega \times A) - \int_A \min_j f_j(x) \mathrm{d}\mu(x) \\
&= \mu(A) - \int_A \min_j f_j(x) \mathrm{d}\mu(x) = \mu'(A).
\end{aligned}$$

Also,

$$\begin{aligned}
\mu'(\Omega) &= \int_\Omega (1 - \min_j f_j) \mathrm{d}\mu(x) \leq \sum_{j=1}^N \int_\Omega (1 - f_j) \mathrm{d}\mu(x) \\
&= \sum_{j=1}^N (\mu(\Omega) - \pi_j(\partial\Omega \times \Omega)) = \sum_{j=1}^N (\pi_j(\overline{\Omega} \times \Omega) - \pi_j(\partial\Omega \times \Omega)) \\
&= \sum_{j=1}^N \pi_j(\Omega \times \Omega) \leq \sum_{j=1}^N \pi_j(\Omega \times \overline{\Omega}) = \sum_{j=1}^N m_j = m_{\text{tot}}.
\end{aligned}$$

and thus $\mu'(\Omega) \leq m_{\text{tot}}$. To conclude, observe that

$$\begin{aligned}
\mathcal{E}(\mu') \leq \sum_{i=1}^N \lambda_i C_p(\pi'_i) &= \sum_{i=1}^N \lambda_i \left(\iint_{\Omega \times \overline{\Omega}} d(x,y)^p \mathrm{d}\pi_i(x,y) \right. \\
&\quad \left. + \iint_{\partial\Omega \times \Omega} d(x,y)^p \mathrm{d}\pi_i(x,y) - \int_\Omega d(x, \partial\Omega)^p \min_j f_j(x) \mathrm{d}\mu(x) \right) \\
&\leq \sum_{i=1}^N \lambda_i C_p(\pi) = \mathcal{E}(\mu).
\end{aligned}$$

□

Proposition 4.1. *Let* $\Psi : \nu \in \mathcal{M}^p_{\leq m_{\text{tot}}} \mapsto \tilde{\nu} \in \mathcal{M}^p_{m_{\text{tot}}}(\tilde{\Omega})$, *where* $\tilde{\nu} := \nu + (m_{\text{tot}} - \nu(\Omega))\delta_{\partial\Omega}$. *The functionals*

$$\mathcal{E} : \mu \in \mathcal{M}^p_{\leq m_{\text{tot}}} \mapsto \sum_{i=1}^N \lambda_i D_p^p(\mu, \mu_i) \ \textit{and}$$

$$\mathcal{F} : \mu \in \mathcal{M}^p_{m_{\text{tot}}}(\tilde{\Omega}) \mapsto \sum_{i=1}^N \lambda_i W^p_{p,\rho}(\tilde{\mu}, \tilde{\mu_i}),$$

have the same infimum values and $\arg\min \mathcal{E} = \Psi^{-1}(\arg\min \mathcal{F})$.

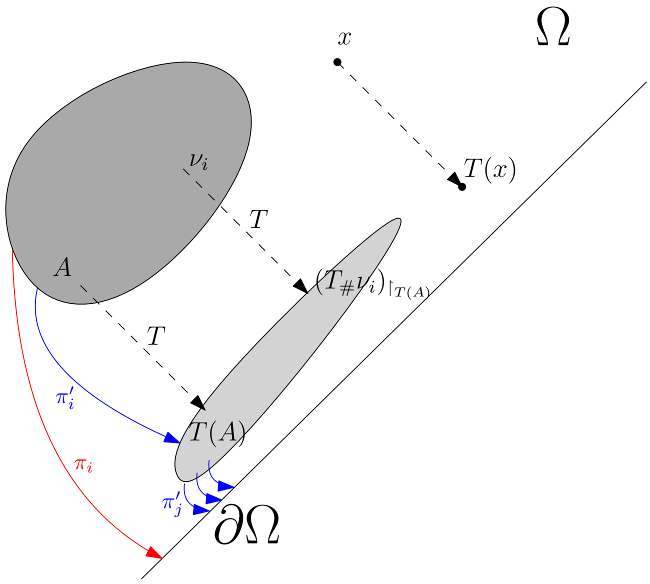

Figure 4: Global picture of the proof. The main idea is to observe that the cost induced by π_i (red) is strictly larger than the sum of costs induces by the π_i's (blue), which leads to a strictly better energy.

Proof. Let G be the set of $\mu \in \mathcal{M}^p$ such that, for all i, there exists $\pi_i \in \mathrm{Opt}_p(\mu_i, \mu)$ with $\pi_i(\Omega, \partial\Omega) = 0$. By point 2 of Lemma 3.2, for $\mu \in G$ and $\pi_i \in \mathrm{Opt}_p(\mu_i, \mu)$ with $\pi_i(\Omega, \partial\Omega) = 0$, $\iota(\pi_i)$ is well defined and satisfies

$$D_p^p(\mu_i, \mu) = C_p(\pi_i) = \iint_{\tilde{\Omega}\times\tilde{\Omega}} d(x,y)^p \mathrm{d}\iota(\pi_i)(x,y) \geq \tilde{C}_p(\iota(\pi_i)) \geq W_{p,\rho}^p(\tilde{\mu_i}, \tilde{\mu}),$$

so that $\mathcal{F}(\Psi(\mu)) \leq \mathcal{E}(\mu)$. As, by Lemma 3.3, $\mathcal{E} \leq \mathcal{F} \circ \Psi$, we therefore have $\mathcal{E}(\mu) = \mathcal{F}(\Psi(\mu))$ for $\mu \in G$.

We now show that if $\mu \notin G$, then there exists $\mu' \in \mathcal{M}^p$ with $\mathcal{E}(\mu') < \mathcal{E}(\mu)$. Let $\mu \notin G$ and $\pi_i \in \mathrm{Opt}_p(\mu_i, \mu)$. Assume that for some i, we have $\pi_i(\Omega, \partial\Omega) > 0$, and introduce $\nu \in \mathcal{M}^p$ defined as $\nu(A) = \pi_i(A, \partial\Omega)$ for $A \subset \Omega$. Define

$$T : \Omega \ni x \mapsto \operatorname*{arg\,min}_{y\in\Omega} \left\{ \lambda_i d(x,y)^p + \sum_{j\neq i} \lambda_j d(y, \partial\Omega)^p \right\} \in \Omega.$$

Note that since $p > 1$, this map is well defined (the minimizer is unique due to strict convexity) and continuous thus measurable. Consider the measure $\mu' = \mu + (T_{\#}\nu)$, where $T_{\#}\nu$ is the psuh-forward of ν by the application T. Consider the transport plan π_i' deduced from π_i where ν is transported onto $T_{\#}\nu$ instead of being transported to $\partial\Omega$ (see Figure 4). More precisely, π_i' is the measure on $\overline{\Omega} \times \overline{\Omega}$ defined by, for Borel sets $A, B \subset \Omega$:

$$\pi_i'(A, B) = \pi_i(A, B) + \nu(A \cap T^{-1}(B)), \quad \pi_i'(A, \partial\Omega) = 0, \quad \pi_i'(\partial\Omega, B) = \pi_i(\partial\Omega, B).$$

We have $\pi_i' \in \mathrm{Adm}(\mu_i, \mu')$. Indeed, for Borel sets $A, B \subset \Omega$:

$$\pi_i'(A, \overline{\Omega}) = \pi_i'(A, \Omega) = \pi_i(A, \Omega) + \nu(A) = \pi_i(A \times \overline{\Omega}) = \mu_i(A),$$

and

$$\begin{aligned}\pi_i'(\overline{\Omega}, B) &= \pi_i'(\Omega, B) + \pi_i'(\partial\Omega, B) = \pi_i(\Omega, B) + \nu(T^{-1}(B)) + \pi_i(\partial\Omega, B) \\ &= \mu(B) + T_{\#}\nu(B) = \mu'(B).\end{aligned}$$

Using π_i' instead of π_i lowers the transport cost by a quantity

$$\int_\Omega [d(x, T(x))^p - d(x, \partial\Omega)^p] \mathrm{d}\nu(x).$$

In a similar way, we define for $j \neq i$ the plan $\pi_j' \in \mathrm{Adm}(\mu_j, \mu')$ by transporting the mass induced by the newly added $(T_{\#}\nu)$ to the diagonal $\partial\Omega$. Using these modified transport plans increases the total cost by $\sum_{j\neq i} \lambda_j \int_\Omega d(T(x), \partial\Omega)^p \mathrm{d}\nu(x)$.

One can observe that

$$\int_\Omega \left[\lambda_i \left(d(x, T(x))^p - d(x, \partial\Omega)^p \right) + \sum_{j\neq i} \lambda_j d(T(x), \partial\Omega)^p \right] \mathrm{d}\nu(x) < 0$$

due to the definition of T and $\nu(\Omega) > 0$.

Therefore, the total transport cost induced by the $(\pi'_i)_{i=1\dots N}$ is strictly smaller than $\mathcal{E}(\mu)$, and thus $\mathcal{E}(\mu') < \mathcal{E}(\mu)$. Finally, we have

$$\inf_{\mu \in \mathcal{M}^p_{\leq m_{tot}}} \mathcal{E}(\mu) = \inf_{\mu \in G} \mathcal{E}(\mu) = \inf_{\mu \in G} \mathcal{F}(\Psi(\mu)) \geq \inf_{\mu \in \mathcal{M}^p_{\leq m_{tot}}} \mathcal{F}(\Psi(\mu)) \geq \inf_{\mu \in \mathcal{M}^p_{\leq m_{tot}}} \mathcal{E}(\mu),$$

where the last inequality comes from $\mathcal{F} \circ \Psi \geq \mathcal{E}$ (Lemma 3.3). Therefore, $\inf \mathcal{E} = \inf \mathcal{F} \circ \Psi$, which is equal to $\inf \mathcal{F}$, as Ψ is a bijection. Also, if μ is a minimizer of $\mathcal{E}$ (should it exists), then $\mu \in G$ and $\mathcal{E}(\mu) = \mathcal{F}(\Psi(\mu))$. Therefore, as the infimum are equal, $\Psi(\mu)$ is a minimizer of $\mathcal{F}$. Reciprocally, if $\tilde{\mu}$ is a minimizer of $\mathcal{F}$, then, by Lemma 3.3, $\mathcal{F}(\tilde{\mu}) \geq \mathcal{E}(\Psi^{-1}(\tilde{\mu}))$, and, as the infimum are equal, $\Psi^{-1}(\tilde{\mu})$ is a minimizer of $\mathcal{E}$. □

The existence of minimizers $\tilde{\mu}$ of $\mathcal{F}$, that is barycenters of $\tilde{\mathbb{P}} := \sum_{i=1}^N \lambda_i \delta_{\tilde{\mu}_i}$, is well-known (see [2, Theorem 8]). Proposition 4.1 asserts that $\Psi^{-1}(\tilde{\mu})$ is a minimizer of $\mathcal{E}$ on $\mathcal{M}^p_{\leq m_{\text{tot}}}$, and thus a barycenter of $\mathbb{P}$ according to Lemma 4.1. We therefore have proved the existence of barycenters in the finite case.

4.2 Existence and consistency of barycenters

To extend the result of the previous section to barycenters of infinitely many measures, we adapt the work of Le Gouic and Loubes [29] to probability distributions supported on $\mathcal{M}^p$. Our Proposition 4.2 corresponds to the Theorem 3 in [29] (consistency of barycenters), and our Theorem 4.3 corresponds to the Theorem 2 in [29] (existence of barycenters).

Proposition 4.2. *Let $\mathbb{P}_n, \mathbb{P}$ be probability measures in $\mathcal{W}^p(\mathcal{M}^p)$. Assume that each $\mathbb{P}_n$ has a barycenter μ_n and that $W_p(\mathbb{P}_n, \mathbb{P}) \to 0$. Then, the sequence $(\mu_n)_n$ is relatively compact in $(\mathcal{M}^p, D_p)$, and any limit of a converging subsequence is a barycenter of $\mathbb{P}$.*

We follow the same "sketch of proof" as Theorem 3 in [29], namely:

- Proving relative compactness of the sequence $(\mu_n)_n$ (for the vague convergence).
- Proving that any accumulation point is a barycenter of $\mathbb{P}$.
- Observing that the convergence actually holds for the D_p metric.

Proof. In order to prove relative compactness of $(\mu_n)_n$, we use the characterization stated in Proposition A.1. Consider a compact set $K \subset \Omega$. We have, because of (12),

$$\begin{aligned}\mu_n(K)^{\frac{1}{p}} &\leq \frac{1}{d(K, \partial\Omega)} D_p(\mu_n, 0) = \frac{1}{d(K, \partial\Omega)} W_p(\delta_{\mu_n}, \delta_0) \\ &\leq \frac{1}{d(K, \partial\Omega)} \left(W_p(\delta_{\mu_n}, \mathbb{P}_n) + W_p(\mathbb{P}_n, \delta_0)\right)\end{aligned}$$

Since μ_n is a barycenter of $\mathbb{P}_n$, it minimizes $\{W_p(\delta_\nu, \mathbb{P}_n),\ \nu \in \mathcal{M}^p\}$, and in particular $W_p(\delta_{\mu_n}, \mathbb{P}_n) \leq W_p(\delta_0, \mathbb{P}_n)$. Furthermore, since we assume that $W_p(\mathbb{P}_n, \mathbb{P}) \to 0$, we have in particular that $\sup_n W_p(\mathbb{P}_n,\ \delta_0) < \infty$. As a consequence $\sup_n \mu_n(K) < \infty$, and Proposition A.1 allows us to conclude that the sequence $(\mu_n)_n$ is relatively compact for the vague convergence.

At that point, the exact same computations as the **proof of Theorem 3** in [29] show the existence of a subsequence $\mu_{n_k} \xrightarrow{v} \mu$ where μ is a barycenter of $\mathbb{P}$ and of some $\nu \in \mathcal{M}^p$ such that $D_p(\mu_{n_k}, \nu) \to D_p(\mu, \nu)$. In order to get the desired conclusion, we use the following lemma:

Lemma 4.2. *Let $\mu, \mu_1, \mu_2, \dots \in \mathcal{M}^p$. Then, $D_p(\mu_n, \mu) \to 0$ if and only if $\mu_n \xrightarrow{v} \mu$ and there exists a persistence measure $\nu \in \mathcal{M}^p$ such that $D_p(\mu_n, \nu) \to D_p(\mu, \nu)$.*

The proof of this lemma is technical and is delayed to Appendix C. □

As the finite case is solved, generalization follows easily using Proposition 4.2.

Theorem 4.3. *Let $\mathbb{P} \in \mathcal{W}^p(\mathcal{M}^p)$. The set of barycenters of $\mathbb{P}$ is not empty.*

Proof. Let $\mathbb{P} = \sum_{i=1}^N \lambda_i \mu_i$ be a probability measure on $\mathcal{M}^p$ with finite support $\mu_1, \dots, \mu_N$ (but no assumption on finiteness of masses). According to Proposition 3.6, there exists sequences $(\mu_i^{(n)})_n$ in $\mathcal{M}^p_f$ with $D_p(\mu_i^{(n)}, \mu_i) \to 0$. As a consequence of the result of Section 4.1, the probability measures $\mathbb{P}^{(n)} := \sum_i \lambda_i \delta_{\mu_i^{(n)}}$ admit barycenters. Furthermore, $W_p^p(\mathbb{P}^{(n)}, \mathbb{P}) \leq \sum_i \lambda_i D_p^p(\mu_i^{(n)}, \mu_i) \to 0$ as $n \to \infty$. It follows from Proposition 4.2 that $\mathbb{P}$ admits a barycenter.

If $\mathbb{P}$ has infinite support, following [29], it can be approximated (in W_p) by a empirical probability measure $\mathbb{P}_n = \frac{1}{n} \sum_{i=1}^n \delta_{\boldsymbol{\mu_i}}$ where the $\boldsymbol{\mu_i}$ are i.i.d. from $\mathbb{P}$. We know that $\mathbb{P}_n$ admits a barycenter since its support is finite, and thus, applying Proposition 4.2 once again, we obtain that $\mathbb{P}$ admits a barycenter. □

4.3 Barycenters in $\mathcal{D}^p$

We now prove the existence of barycenters in $\mathcal{D}^p$ for any distribution supported on $\mathcal{D}^p$, extending the results of [36], which prove their existence for specific probability distributions (namely distributions with compact support or specific rates of decay). The following theorem asserts two different things: that $\arg\min\{\mathcal{E}(a),\ a \in \mathcal{D}^p\}$ is non empty, and that $\min\{\mathcal{E}(a),\ a \in \mathcal{D}^p\} = \min\{\mathcal{E}(\mu),\ \mu \in \mathcal{M}^p\}$, i.e a persistence measure cannot perform strictly better than an optimal persistence diagram when averaging diagrams. As for barycenters in $\mathcal{M}^p$, we start with the finite case.

Lemma 4.3. *Consider $a_1, \ldots, a_N \in \mathcal{D}_f$, weights $(\lambda_i)_i$ that sum to 1, and let $\mathbb{P} := \sum_{i=1}^N \lambda_i \delta_{a_i}$. Then, $\mathbb{P}$ admits a barycenter in $\mathcal{D}_f$.*

The proof of this lemma is delayed to Appendix C. Note that we actually prove a stronger result, namely that if $\mathbb{P} = \sum_{i=1}^N \lambda_i \delta_{c_i}$, with the c_i point measures (see Appendix A) with the same mass in some sufficiently regular metric space (X, d) (see the proof for the technical details), then there exists a barycenter of $\mathbb{P}$ for the W_p distance which is a point measure.

Theorem 4.4. *Consider a distribution $\mathbb{P}$ supported on $\mathcal{D}^p$ with finite p-th moment. There exists a barycenter of $\mathbb{P}$ which is in $\mathcal{D}^p$.*

Proof. The proof is the same as the one of Theorem 4.3, using additionally the fact that $\mathcal{D}^p$ is closed in $\mathcal{M}^p$ (Proposition A.5) to conclude that $\mathbb{P}$ has a barycenter in $\mathcal{D}^p$. $\square$

5 Applications

5.1 Convergence of random persistence measures in the thermodynamic regime

Geometric probability is the study of geometric quantities arising naturally from point processes in $\mathbb{R}^d$. Recently, several works [4, 15, 16, 19, 35] used techniques originating from this field to understand the persistent homology of such point processes. Let $\mathbb{X}_n := \{X_1, \ldots, X_n\}$ be a n-sample of a distribution having some density bounded by below on the cube $[0, 1]^d$. Extending the work of Duy, Hiraoka and Shirai [16], Divol and Polonik [15] show laws of large numbers for the persistence diagrams $\mathrm{Dgm}(\mathbb{X}_n)$ of $\mathbb{X}_n$, built with either the Čech or Rips filtration. More precisely, it is shown in [16, Theorem 1.5] that the rescaled diagram

$$\mu_n := \frac{\mathrm{Dgm}(n^{1/d}\mathbb{X}_n)}{n},$$

which is a persistence measure, almost surely converges vaguely to some non-degenerate Radon measure μ, and Divol and Polonik [15, Theorem 1] show that μ has moments of all orders and that for all $p \geq 1$, the total persistence $\mathrm{Pers}_p(\mu_n)$ also almost surely converges to $\mathrm{Pers}_p(\mu)$. The formalism developed here, and more specifically Theorem 3.4, gives the following result:

Proposition 5.1. $D_p(\mu_n, \mu) \xrightarrow[n\to\infty]{} 0$ *almost surely.*

5.2 Stability of the expected persistence diagrams

Given an i.i.d. sample of N persistence diagrams $a_1 \ldots a_N$, a natural persistence measure to consider is their linear sample mean $\overline{a} := \frac{1}{N}\sum_{1\leq i\leq N} a_i$. More generally, given $\mathbb{P} \in \mathcal{W}^p(\mathcal{M}^p)$, and $\boldsymbol{\mu} \sim \mathbb{P}$, one may want to define $\mathbb{E}[\boldsymbol{\mu}]$ the linear expectation of $\mathbb{P}$ in the same vein. A well-suited definition of the linear expectation requires technical care (basically, turning the finite sum into a Bochner integral) and is detailed in Appendix D. It however satisfies the natural following characterization:

$$\forall K \subset \Omega \text{ compact},\ \mathbb{E}[\boldsymbol{\mu}](K) = \mathbb{E}[\boldsymbol{\mu}(K)]. \tag{23}$$

The behavior of such measures is studied in [12], which shows that they have densities with respect to the Lebesgue measure in a wide variety of settings. A natural question is the stability of the linear expectations of random diagrams with respect to the underlying phenomenon generating them. The following proposition gives a positive answer to this problem, showing that given two close probability distributions $\mathbb{P}$ and $\mathbb{P}'$ supported on $\mathcal{M}^p$, their linear expectations are close for the metric D_p.

Proposition 5.2. *Let $\mathbb{P}, \mathbb{P}' \in \mathcal{W}^p(\mathcal{M}^p)$. Then, for any coupling $\pi \in \Pi(\mathbb{P}, \mathbb{P}')$ between $\mathbb{P}$ and $\mathbb{P}'$, we have* $D_p^p(\mathbb{E}[\boldsymbol{\mu}], \mathbb{E}[\boldsymbol{\mu}']) \leq \mathbb{E}_\pi[D_p^p(\boldsymbol{\mu}, \boldsymbol{\mu}')]$.

The proof is postponed to Appendix D.

Using stability results on the distances d_p [14], one is able to obtain a more precise control between the expectations in some situations. For $\mathbb{Y}$ a sample in some metric space, denote by $\mathrm{Dgm}(\mathbb{Y})$ the persistence diagram of $\mathbb{Y}$ built with the Čech filtration.

Proposition 5.3. *Let ξ, ξ' be two probability measures on some some d-dimensional Riemannian manifold $(\mathbb{X}, \rho)$. Let $\mathbb{X}_n$ (resp. $\mathbb{X}'_n$) be a n-sample of law ξ (resp. ξ'). Then, for any $k > d$, and any $p \geq k+1$,*

$$D_p^p(\mathbb{E}[\mathrm{Dgm}(\mathbb{X}_n)], \mathbb{E}[\mathrm{Dgm}(\mathbb{X}'_n)]) \leq C_{k,d} \cdot n \cdot W_{p-k}^{p-k}(\xi, \xi') \tag{24}$$

where $C_{k,d} := C\mathrm{diam}(\mathbb{X})^{k-d}\frac{k}{k-d}$ for some constant C depending only on $\mathbb{X}$.

In particular, letting $p \to \infty$, we obtain a bottleneck stability result:

$$D_\infty(\mathbb{E}[\mathrm{Dgm}(\mathbb{X}_n)], \mathbb{E}[\mathrm{Dgm}(\mathbb{X}'_n)]) \leq W_\infty(\xi, \xi'). \tag{25}$$

Proof. Let π be any coupling between $\mathbb{X}_n$ a n-sample of law ξ, and $\mathbb{X}'_n$ a n-sample of law ξ'. According to Proposition 5.2, $D_p^p(\mathbb{E}[\mathrm{Dgm}(\mathbb{X}_n)], \mathbb{E}[\mathrm{Dgm}(\mathbb{X}'_n)]) \leq \mathbb{E}_\pi[D_p^p(\mathrm{Dgm}(\mathbb{X}_n), \mathrm{Dgm}(\mathbb{X}'_n))]$. It is stated in [14, Wasserstein Stability Theorem] that

$$D_p^p(\mathrm{Dgm}(\mathbb{X}_n), \mathrm{Dgm}(\mathbb{X}'_n)) \leq C_{k,d} H(\mathbb{X}_n, \mathbb{X}'_n)^{p-k},$$

where $C_{k,d} := C\mathrm{diam}(\mathbb{X})^{k-d}\frac{k}{k-d}$ for some constant C depending only on $\mathbb{X}$, and H is the Hausdorff distance between sets. By taking the infimum on transport plans π, we obtain $D_p^p(\mathbb{E}[\mathrm{Dgm}(\mathbb{X}_n)], \mathbb{E}[\mathrm{Dgm}(\mathbb{X}'_n)]) \leq C_{k,d} W_{H,p-k}^{p-k}(\xi^{\otimes n}, (\xi')^{\otimes n})$, where $W_{H,p}$ is the p-Wasserstein distance between probability distributions on compact sets of the manifold $\mathbb{X}$, endowed with the Hausdorff distance. Lemma 15 of [13] states that

$$W_{H,p-k}^{p-k}(\xi^{\otimes n}, (\xi')^{\otimes n}) \leq n \cdot W_{p-k}^{p-k}(\xi, \xi'),$$

concluding the proof. ☐

Note that this proposition illustrates the usefulness of introducing new distances D_p: the question of the proximity between linear expectations requires to extend the metrics d_p to Radon measures.

6 Conclusion

In this article, we introduce the space of persistence measures, a generalization of persistence diagrams which naturally appears in different applications (e.g. when studying persistence diagrams coming from a random process). We provide an analysis of this space that also holds for the subspace of persistence diagrams. In particular, we observe that many notions used for the statistical analysis of persistence diagrams can be expressed naturally using this formalism based on optimal partial transport. We give characterizations of convergence with respect to optimal transport metrics in terms of convergence for measures. We then prove existence and consistency of Fréchet means for any probability distribution of persistence diagrams and measures, extending previous work in the TDA community. We illustrate the interest of introducing the persistence measures space and its metrics in statistical applications of TDA.

Aknowledgement This work was partially supported by the advanced Grant of the European Research Council GUDHI (Geometric Understanding in Higher Dimensions). TL is supported by the AMX grant, cole polytechnique. Authors thank Frédéric Chazal for fruitful discussions.

References

[1] Henry Adams, Tegan Emerson, Michael Kirby, Rachel Neville, Chris Peterson, Patrick Shipman, Sofya Chepushtanova, Eric Hanson, Francis Motta, and Lori Ziegelmeier. Persistence images: a stable vector representation of persistent homology. *Journal of Machine Learning Research*, 18(8):1–35, 2017.

[2] Martial Agueh and Guillaume Carlier. Barycenters in the wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011.

[3] Andrew J Blumberg, Itamar Gal, Michael A Mandell, and Matthew Pancia. Robust statistics, hypothesis testing, and confidence intervals for persistent homology on metric measure spaces. *Foundations of Computational Mathematics*, 14(4):745–789, 2014.

[4] Omer Bobrowski, Matthew Kahle, Primoz Skraba, et al. Maximally persistent cycles in random geometric complexes. *The Annals of Applied Probability*, 27(4):2032–2060, 2017.

[5] Salomon Bochner. Integration von funktionen, deren werte die elemente eines vektorraumes sind. *Fundamenta Mathematicae*, 20(1):262–176, 1933.

[6] Peter Bubenik and Tane Vergili. Topological spaces of persistence modules and their properties. *arXiv preprint arXiv:1802.08117*, 2018.

[7] G. Carlier and I. Ekeland. Matching for teams. *Economic Theory*, 42(2):397–418, 2010.

[8] Guillaume Carlier, Adam Oberman, and Edouard Oudet. Numerical methods for matching for teams and wasserstein barycenters. *ESAIM: Mathematical Modelling and Numerical Analysis*, 49(6):1621–1642, 2015.

[9] Mathieu Carrière, Marco Cuturi, and Steve Oudot. Sliced wasserstein kernel for persistence diagrams. In *34th International Conference on Machine Learning*, 2017.

[10] Mathieu Carrire, Steve Y. Oudot, and Maks Ovsjanikov. Stable topological signatures for points on 3d shapes. *Computer Graphics Forum*, 34(5):1–12, 2015.

[11] Frédéric Chazal, Vin De Silva, Marc Glisse, and Steve Oudot. *The structure and stability of persistence modules*. Springer, 2016.

[12] Frédéric Chazal and Vincent Divol. The Density of Expected Persistence Diagrams and its Kernel Based Estimation. In Bettina Speckmann and Csaba D. Tóth, editors, *34th International Symposium on Computational Geometry (SoCG 2018)*, volume 99 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 26:1–26:15, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[13] Frédéric Chazal, Brittany Fasy, Fabrizio Lecci, Bertrand Michel, Alessandro Rinaldo, and Larry Wasserman. Subsampling methods for persistent homology. In *International Conference on Machine Learning*, pages 2143–2151, 2015.

[14] David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Yuriy Mileyko. Lipschitz functions have l p-stable persistence. *Foundations of computational mathematics*, 10(2):127–139, 2010.

[15] Vincent Divol and Wolfgang Polonik. On the choice of weight functions for linear representations of persistence diagrams. *arXiv preprint arXiv:1807.03678*, 2018.

[16] Trinh Khanh Duy, Yasuaki Hiraoka, and Tomoyuki Shirai. Limit theorems for persistence diagrams. *arXiv preprint arXiv:1612.08371*, 2016.

[17] Herbert Edelsbrunner and John Harer. *Computational topology: an introduction*. American Mathematical Soc., 2010.

[18] Alessio Figalli and Nicola Gigli. A new transportation distance between non-negative measures, with applications to gradients flows with dirichlet boundary conditions. *Journal de mathématiques pures et appliquées*, 94(2):107–130, 2010.

[19] Akshay Goel, Khanh Duy Trinh, and Kenkichi Tsunoda. Asymptotic behavior of betti numbers of random geometric complexes. *arXiv preprint arXiv:1805.05032*, 2018.

[20] Marshall Hall. *Combinatorial theory (2nd ed)*. John Wiley & Sons, 1986.

[21] Yasuaki Hiraoka, Takenobu Nakamura, Akihiko Hirata, Emerson G. Escolar, Kaname Matsue, and Yasumasa Nishiura. Hierarchical structures of amorphous solids characterized by persistent homology. *Proceedings of the National Academy of Sciences*, 2016.

[22] Olav Kallenberg. *Random measures, theory and applications.* Springer, 2017.

[23] A.S. Kechris. *Classical Descriptive Set Theory.* Graduate Texts in Mathematics. Springer-Verlag, 1995.

[24] Michael Kerber, Dmitriy Morozov, and Arnur Nigmetov. Geometry helps to compare persistence diagrams. *Journal of Experimental Algorithmics (JEA)*, 22(1):1–4, 2017.

[25] M. Kramar, A. Goullet, L. Kondic, and K. Mischaikow. Persistence of force networks in compressed granular media. *Phys. Rev. E*, 87:042207, Apr 2013.

[26] Genki Kusano, Kenji Fukumizu, and Yasuaki Hiraoka. Kernel method for persistence diagrams via kernel embedding and weight factor. *arXiv preprint arXiv:1706.03472*, 2017.

[27] Roland Kwitt, Stefan Huber, Marc Niethammer, Weili Lin, and Ulrich Bauer. Statistical topological data analysis-a kernel perspective. In *Advances in neural information processing systems*, pages 3070–3078, 2015.

[28] Théo Lacombe, Marco Cuturi, and Steve Oudot. Large scale computation of means and clusters for persistence diagrams using optimal transport. In *Advances in Neural Information Processing Systems*, 2018.

[29] Thibaut Le Gouic and Jean-Michel Loubes. Existence and consistency of wasserstein barycenters. *Probability Theory and Related Fields*, pages 1–17, 2016.

[30] Chunyuan Li, Maks Ovsjanikov, and Frederic Chazal. Persistence-based structural recognition. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[31] Yuriy Mileyko, Sayan Mukherjee, and John Harer. Probability measures on the space of persistence diagrams. *Inverse Problems*, 27(12):124007, 2011.

[32] Gabriel Peyr and Marco Cuturi. *Computational Optimal Transport.* Number 2017-86. December 2017.

[33] Filippo Santambrogio. Optimal transport for applied mathematicians. *Birkäuser, NY*, 2015.

[34] Alexander Schrijver. *Combinatorial optimization: polyhedra and efficiency*, volume 24. Springer Science & Business Media, 2003.

[35] Benjamin Schweinhart. Weighted persistent homology sums of random cech complexes. *arXiv preprint arXiv:1807.07054*, 2018.

[36] Katharine Turner, Yuriy Mileyko, Sayan Mukherjee, and John Harer. Fréchet means for distributions of persistence diagrams. *Discrete & Computational Geometry*, 52(1):44–70, 2014.

[37] Katharine Turner, Sayan Mukherjee, and Doug M. Boyer. Persistent homology transform for modeling shapes and surfaces. *Information and Inference: A Journal of the IMA*, 3(4):310–344, 2014.

[38] Cédric Villani. *Topics in optimal transportation.* Number 58. American Mathematical Soc., 2003.

[39] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.

[40] August M Zapała et al. Jensens inequality for conditional expectations in banach spaces. *Real Analysis Exchange*, 26(2):541–552, 2000.

A Elements of measure theory

In the following, Ω denotes a locally compact Polish space.

Definition A.1. *The space $\mathcal{M}(\Omega)$ of Radon measures supported on Ω is the space of Borel measures which give finite mass to every compact set of Ω. The vague topology on $\mathcal{M}(\Omega)$ is the coarsest topology such that the applications $\mu \mapsto \mu(f)$ are continuous for every $f \in C_c(\Omega)$, the space of continuous functions with compact support in Ω.*

Remark A.1. *We will extensively use the fact that to define a Borel measure on $\Omega \times \Omega$, it is sufficient to define it on the sets $A \times B$, for $A, B \subset \Omega$ Borel sets, since these sets generate the Borel sigma-algebra on $\Omega \times \Omega$.*

Definition A.2. *Denote by $\mathcal{M}_f(\Omega)$ the space of finite Borel measures on Ω. The weak topology on $\mathcal{M}_f(\Omega)$ is the coarsest topology such that the applications $\mu \mapsto \mu(f)$ are continuous for every $f \in C_b(\Omega)$, the space of continuous bounded functions in Ω.*

We denote by $\xrightarrow{v}$ the vague convergence and $\xrightarrow{w}$ the weak convergence.

Definition A.3. *A set $F \subset \mathcal{M}(\Omega)$ is said to be tight if, for every $\varepsilon > 0$, there exists a compact set K with $\mu(\Omega \backslash K) \leq \varepsilon$ for every $\mu \in F$.*

The following propositions are standard results. Corresponding proofs can be found for instance in [22, Section 15.7].

Proposition A.1. *A set $F \subset \mathcal{M}(\Omega)$ is relatively compact for the vague topology if and only if for every compact set K included in Ω,*

$$\sup\{\mu(K),\ \mu \in F\} < \infty.$$

Proposition A.2 (Prokhorov's theorem)**.** *A set $F \subset \mathcal{M}_f(\Omega)$ is relatively compact for the weak topology if and only if F is tight and $\sup_{\mu \in F} \mu(\Omega) < \infty$.*

Proposition A.3. *Let $\mu, \mu_1, \mu_2, \ldots$ be measures in $\mathcal{M}_f(\Omega)$. Then, $\mu_n \xrightarrow{w} \mu$ if and only if $\mu_n(\Omega) \to \mu(\Omega)$ and $\mu_n \xrightarrow{v} \mu$.*

Proposition A.4 (Portmanteau theorem)**.** *Let $\mu, \mu_1, \mu_2, \ldots$ be measures in $\mathcal{M}(\Omega)$. Then, $\mu_n \xrightarrow{v} \mu$ if and only if one of the following propositions holds:*

- *for all open sets $U \subset \Omega$ and all* bounded *closed sets $F \subset \Omega$,*

$$\limsup_{n\to\infty} \mu_n(F) \leq \mu(F) \text{ and } \liminf_{n\to\infty} \mu_n(U) \geq \mu(U).$$

- *for all* bounded *Borel sets A with $\mu(\partial A) = 0$, $\lim_{n\to\infty} \mu_n(A) = \mu(A)$.*

Definition A.4. *The set of* point measures *on Ω is the subset $\mathcal{D}(\Omega) \subset \mathcal{M}(\Omega)$ of Radon measures with discrete support and integer mass on each point, that is of the form*

$$\sum_{x \in X} n_x \delta_x$$

where $n_x \in \mathbb{N}$ and $X \subset \Omega$ is some locally finite set.

Proposition A.5. *The set $\mathcal{D}(\Omega)$ is closed in $\mathcal{M}(\Omega)$ for the vague topology.*

B Delayed proofs of Section 3

For the sake of completeness, we present in this section proofs which either require very few adaptations from corresponding proofs in [18] or which are close to standard proofs in optimal transport theory.

Proofs of Proposition 3.1 and Proposition 3.8.

- For $\pi \in \mathrm{Adm}(\mu, \nu)$ supported on E_Ω, and for any compact sets $K,\ K' \subset \Omega$, one has $\pi((K \times \overline{\Omega}) \cup (\overline{\Omega} \times K')) \leq \mu(K) + \nu(K') < \infty$. As any compact subset of E_Ω is included in a set of the form $(K \times \overline{\Omega}) \cup (\overline{\Omega} \times K')$, Proposition A.1 implies that $\mathrm{Adm}(\mu, \nu)$ is relatively compact for the vague convergence on E_Ω. Also, if a sequence $(\pi_n)_n$ in $\mathrm{Adm}(\mu, \nu)$ converges vaguely to some $\pi \in \mathcal{M}(E_\Omega)$, then the marginals of π are still μ and ν, implying that $\mathrm{Adm}(\mu, \nu)$ is closed in $\mathcal{M}(E_\Omega)$: it is therefore sequentially compact.

- To prove the second point of Proposition 3.1, consider $\pi, \pi_1, \pi_2, \dots$ such that $\pi_n \xrightarrow{v} \pi$, and introduce $\pi'_n : A \mapsto \iint_A d(x,y)^p \mathrm{d}\pi_n$. The sequence $(\pi'_n)_n$ still converges vaguely to $\pi' : A \mapsto \iint_A d(x,y)^p \mathrm{d}\pi$. Portmanteau theorem (Proposition A.4) applied with the open set E_Ω to the measures $\pi'_n \xrightarrow{v} \pi'$ implies that $C_p(\pi) = \pi'(E_\Omega) \leq \liminf_n \pi'_n(E_\Omega) = \liminf_n C_p(\pi_n)$, i.e. C_p is lower semi-continuous.

- We now prove the lower semi-continuity of C_∞. Let $(\pi_n)_n$ be a sequence converging vaguely to π on E_Ω and let $r > \liminf_{n\to\infty} C_\infty(\pi_n)$. The set $U_r = \{(x,y) \in E_\Omega,\ d(x,y) > r\}$ is open. By Portmanteau theorem (Proposition A.4), we have $0 = \liminf_{n\to\infty} \pi_n(U_r) \geq \pi(U_r)$. Therefore, $\mathrm{spt}(\pi) \subset U_r^c$ and $C_\infty(\pi) \leq r$. As this holds for any $r > \liminf_{n\to\infty} C_\infty(\pi_n)$, we have $\liminf_{n\to\infty} C_\infty(\pi_n) \geq C_\infty(\pi)$.

- We show that for any $1 \leq p \leq \infty$, the lower semi-continuity of C_p and the sequential compactness of $\mathrm{Adm}(\mu,\nu)$ imply that 1. $\mathrm{Opt}_p(\mu,\nu)$ is a non-empty compact set for the vague topology on E_Ω and that 2. D_p is lower semi-continuous.

  1. Let $(\pi_n)_n$ be a minimizing sequence of (7) in $\mathrm{Adm}(\mu,\nu)$. As $\mathrm{Adm}(\mu,\nu)$ is sequentially compact, it has an adherence value π, and the lower semi-continuity implies that $C_p(\pi) \leq \liminf_{n\to\infty} C_p(\pi_n) = D_p^p(\mu,\nu)$, so that $\mathrm{Opt}_p(\mu,\nu)$ is non-empty. Using once again the lower semi-continuity of C_p, if a sequence in $\mathrm{Opt}_p(\mu,\nu)$ converges to some limit, then the cost of the limit is smaller than (and thus equal to) $D_p^p(\mu,\nu)$, i.e. the limit is in $\mathrm{Opt}_p(\mu,\nu)$. The set $\mathrm{Opt}_p(\mu,\nu)$ being closed in the sequentially compact set $\mathrm{Adm}(\mu,\nu)$, it is also sequentially compact.
  2. Let $\mu_n \xrightarrow{v} \mu$ and $\nu_n \xrightarrow{v} \nu$. Up to taking a subsequence, on may assume that $D_p(\mu_n,\nu_n)$ converges (eventually to infinity). Consider $\pi_n \in \mathrm{Opt}_p(\mu_n,\nu_n)$. For any compact sets $K,\ K' \subset \Omega$, one has $\pi_n((K \times \overline{\Omega}) \cup (\overline{\Omega} \times K')) \leq \sup_n \mu_n(K) + \sup_n \nu_n(K') < \infty$. Therefore, by Proposition A.1, there exists a subsequence $(\pi_{n_k})_k$ which converges vaguely to some $\pi \in \mathrm{Adm}(\mu,\nu)$. As the marginals of π_{n_k} converges to μ and ν, $\pi \in \mathrm{Adm}(\mu,\nu)$. Therefore,
  $$D_p^p(\mu,\nu) \leq C_p(\pi) \leq \liminf_{n\to\infty} C_p(\pi_n) = \liminf_{n\to\infty} D_p^p(\mu_n,\nu_n).$$

- Finally, we prove that D_p is a metric on $\mathcal{M}^p$. Let $\mu,\nu,\lambda \in \mathcal{M}^p$. The symmetry of D_p is clear. If $D_p(\mu,\nu) = 0$, then there exists $\pi \in \mathrm{Adm}(\mu,\nu)$ supported on $\{(x,x),\ x \in \Omega\}$. Therefore, for a Borel set $A \subset \Omega$, $\mu(A) = \pi(A \times \overline{\Omega}) = \pi(A \times A) = \pi(\overline{\Omega} \times A) = \nu(A)$, and $\mu = \nu$. To prove the triangle inequality, we need a variant on the gluing lemma, stated in [18, Lemma 2.1]: for $\pi_{12} \in \mathrm{Opt}(\mu,\nu)$ and $\pi_{23} \in \mathrm{Opt}(\nu,\lambda)$ there exists a measure $\gamma \in \mathcal{M}(\overline{\Omega}^3)$ such that the marginal corresponding to the first two entries (resp. two last entries), when restricted to E_Ω, is equal to π_{12} (resp. π_{23}), and induces a zero cost on $\partial\Omega \times \partial\Omega$. Therefore,
  $$\begin{aligned} D_p(\mu,\lambda) &= \left(\int_{\overline{\Omega}^2} d(x,z)^p \mathrm{d}\gamma(x,y,z)\right)^{1/p} \\ &\leq \left(\int_{\overline{\Omega}^2} d(x,y)^p \mathrm{d}\gamma(x,y,z)\right)^{1/p} + \left(\int_{\overline{\Omega}^2} d(y,z)^p \mathrm{d}\gamma(x,y,z)\right)^{1/p} \\ &= \left(\int_{\overline{\Omega}^2} d(x,y)^p \mathrm{d}\pi_{12}(x,y)\right)^{1/p} + \left(\int_{\overline{\Omega}^2} d(y,z)^p \mathrm{d}\pi_{23}(y,z)\right)^{1/p} \\ &= D_p(\mu,\nu) + D_p(\nu,\lambda). \end{aligned}$$
  The proof is similar for $p = \infty$.

□

Proof of Proposition 3.3. We first show the separability. Consider for $k > 0$ a partition of Ω into squares (C_i^k) of side length 2^{-k}, centered at points x_i^k. Let F be the set of all measures of the form $\sum_{i\in I} q_i \delta_{x_i^k}$ for q_i positive rationals, $k > 0$ and I a finite subset of $\mathbb{N}$. Our goal is to show that the countable set F is dense in $\mathcal{M}^p$. Fix $\varepsilon > 0$, and $\mu \in \mathcal{M}^p$. The proof is in three steps.

1. Since $\mathrm{Pers}_p(\mu) < \infty$, there exists a compact $K \subset \Omega$ such that $\mathrm{Pers}_p(\mu) - \mathrm{Pers}_p(\mu_0) < \varepsilon^p$, where μ_0 is the restriction of μ to K. By considering the transport plan between μ and μ_0 induced by the identity map on K and the projection onto the diagonal on $\overline{\Omega}\backslash K$, it follows that $D_p^p(\mu,\mu_0) \leq \mathrm{Pers}_p(\mu) - \mathrm{Pers}_p(\mu_0) \leq \varepsilon^p$.

2. Consider k such that $2^{-k} \leq \varepsilon/(\sqrt{2}\mu(K)^{1/p})$ and denote by I the indices corresponding to squares C_i^k intersecting K. Let $\mu_1 = \sum_{i\in I}^{\infty} \mu_0(C_i^k)\delta_{x_i^k}$. One can create a transport map between μ_0 and μ_1 by mapping each square C_i^k to its center x_i^k, so that

$$D_p(\mu_0,\mu_1) \leq \left(\sum_i \mu_0(C_i^k)(\sqrt{2}\cdot 2^{-k})^p\right)^{1/p} \leq \mu(K)^{1/p}\sqrt{2}\cdot 2^{-k} \leq \varepsilon.$$

3. Consider, for $i \in I$, q_i a rational number satisfying $q_i \leq \mu_0(C_i)$ and $|\mu_0(C_i^k) - q_i| \leq \varepsilon^p / \left(\sum_{i\in I} d(x_i^k, \partial\Omega)^p\right)$. Let $\mu_2 = \sum_{i\in I} q_i \delta_{x_i^k}$. Consider the transport plan between μ_2 and μ_1 that fully transports μ_2 onto μ_1, and transport the remaining mass in μ_1 onto the diagonal. Then,

$$D_p(\mu_1,\mu_2) \leq \left(\sum_{i\in I} |\mu_0(C_i^k) - q_i| d(x_i^k,\partial\Omega)^p\right)^{1/p} \leq \varepsilon.$$

As $\mu_2 \in F$ and $D_p(\mu,\mu_2) \leq 3\varepsilon$, the separability is proven.

To prove that the space is complete, consider a Cauchy sequence $(\mu_n)_n$. As $(\mathrm{Pers}_p(\mu_n))_n = (D_p^p(\mu_n,0))_n$ is a Cauchy sequence, it is bounded. Therefore, for $K \subset \Omega$ a compact set, (12) implies that $\sup_n \mu_n(K) < \infty$. Proposition A.1 implies that $(\mu_n)_n$ is relatively compact for the vague convergence on Ω: consider $(\mu_{n_k})_k$ a subsequence converging vaguely on Ω to some measure μ. By the lower semi-continuity of D_p,

$$\mathrm{Pers}_p(\mu) = D_p^p(\mu,0) \leq \liminf_{k\to\infty} D_p^p(\mu_{n_k},0) < \infty,$$

so that $\mu \in \mathcal{M}^p$. Using once again the lower semi-continuity,

$$\begin{aligned} D_p(\mu_n,\mu) &\leq \liminf_{k\to\infty} D_p(\mu_n,\mu_{n_k}) \\ \lim_{n\to\infty} D_p(\mu_n,\mu) &\leq \lim_{n\to\infty}\liminf_{k\to\infty} D_p(\mu_n,\mu_{n_k}) = 0, \end{aligned}$$

ensuring that $D_p(\mu_n,\mu) \to 0$, that is the space is complete. □

Proof of the direct implication of Theorem 3.4. Let $\mu,\mu_1,\mu_2,\dots$ be elements of $\mathcal{M}^p$ and assume that the sequence $(D_p(\mu_n,\mu))_n$ converges to 0. The triangle inequality implies that $\mathrm{Pers}_p(\mu_n) = D_p^p(\mu_n,0)$ converges to $\mathrm{Pers}_p(\mu) = D_p^p(\mu,0)$. Let $f \in C_c(\Omega)$, whose support is included in some compact set K. For any $\varepsilon > 0$, there exists a Lipschitz function f_ε, with Lipschitz constant L and whose support is included in K, with the ∞-norm $\|f - f_\varepsilon\|_\infty$ smaller than ε. The convergence of $\mathrm{Pers}_p(\mu_n)$ and (12) imply that $\sup_k \mu_k(K) < \infty$. Let $\pi_n \in \mathrm{Opt}_p(\mu_n,\mu)$, we have

$$\begin{aligned} |\mu_n(f) - \mu(f)| &\leq |\mu_n(f - f_\varepsilon)| + |\mu(f - f_\varepsilon)| + |\mu_n(f_\varepsilon) - \mu(f_\varepsilon)| \\ &\leq (\mu_n(K) + \mu(K))\varepsilon + |\mu_n(f_\varepsilon) - \mu(f_\varepsilon)| \\ &\leq (\sup_k \mu_k(K) + \mu(K))\varepsilon + |\mu_n(f_\varepsilon) - \mu(f_\varepsilon)|. \end{aligned}$$

Also,

$$\begin{aligned} |\mu_n(f_\varepsilon) - \mu(f_\varepsilon)| &\leq \iint_{\overline{\Omega}^2} |f_\varepsilon(x) - f_\varepsilon(y)| \mathrm{d}\pi_n(x,y) \quad \text{where } \pi_n \in \mathrm{Opt}(\mu_n,\mu) \\ &\leq L \iint_{(K\times\overline{\Omega})\cup(\overline{\Omega}\times K)} d(x,y)\mathrm{d}\pi_n(x,y) \\ &\leq L\pi_n((K\times\overline{\Omega})\cup(\overline{\Omega}\times K))^{1-\frac{1}{p}} \left(\iint_{(K\times\overline{\Omega})\cup(\overline{\Omega}\times K)} d(x,y)^p \mathrm{d}\pi_n(x,y)\right)^{\frac{1}{p}} \\ &\qquad \text{by Hlder's inequality.} \\ &\leq L\left(\sup_k \mu_k(K) + \mu(K)\right)^{1-\frac{1}{p}} D_p(\mu_n,\mu) \to 0. \end{aligned}$$

Therefore, taking the limsup in n and then letting ε goes to 0, we obtain that $\mu_n(f) \to \mu(f)$. □

Proof of Proposition 3.10. Let $(\mu_n)_n$ be a Cauchy sequence in D_∞. Fix a compact $K \subset \Omega$, and pick $\varepsilon = d(K, \partial\Omega)/2$. There exists n_0 such that for $n > n_0$, $D_\infty(\mu_n, \mu_{n_0}) < \varepsilon$. Let $K_\varepsilon := \{x \in \Omega,\ d(x,K) \leq \varepsilon\}$. By considering $\pi_n \in \mathrm{Opt}_\infty(\mu_n, \mu_{n_0})$, and since $D_\infty(\mu_n, \mu_{n_0}) < \varepsilon$, we have that

$$\mu_n(K) = \pi_n(K \times \overline{\Omega}) = \pi_n(K \times K_\varepsilon) \leq \mu_{n_0}(K_\varepsilon). \tag{26}$$

Therefore, $(\mu_n(K))_n$ is uniformly bounded, and Proposition A.1 implies that $(\mu_n)_n$ is relatively compact. Finally, the exact same computations than in the proof of the completeness for $p < \infty$ show that $(\mu_n)_n$ converges for the D_∞ metric. □

Proof of Proposition 3.11. Let $f \in C_c(\Omega)$, whose support is included in some compact set K. For any $\varepsilon > 0$, there exists a L-Lipschitz function f_ε, whose support is included in K, with $\|f - f_\varepsilon\|_\infty \leq \varepsilon$. Observe that $\sup_k \mu_k(K) < \infty$ using the same arguments than for (26). Let $\pi_n \in \mathrm{Opt}_\infty(\mu_n, \mu)$. We have

$$\begin{aligned}
|\mu_n(f) - \mu(f)| &\leq |\mu_n(f - f_\varepsilon)| + |\mu(f - f_\varepsilon)| + |\mu_n(f_\varepsilon) - \mu(f_\varepsilon)| \\
&\leq (\mu_n(K) + \mu(K))\varepsilon + |\mu_n(f_\varepsilon) - \mu(f_\varepsilon)| \\
&\leq (\sup_k \mu_k(K) + \mu(K))\varepsilon + |\mu_n(f_\varepsilon) - \mu(f_\varepsilon)|.
\end{aligned}$$

Also,

$$\begin{aligned}
|\mu_n(f_\varepsilon) - \mu(f_\varepsilon)| &\leq \iint_{\overline{\Omega}^2} |f_\varepsilon(x) - f_\varepsilon(y)| \mathrm{d}\pi_n(x,y) \\
&\leq L \iint_{(K \times \overline{\Omega}) \cup (\overline{\Omega} \times K)} d(x,y) \mathrm{d}\pi_n(x,y) \\
&\leq L C_\infty(\pi_n)(\pi_n(K \times \overline{\Omega}) + \pi_n(\overline{\Omega} \times K)) \\
&\leq L D_\infty(\mu_n, \mu) \left(\sup_k \mu_k(K) + \mu(K) \right) \to 0.
\end{aligned}$$

Therefore, taking the limsup in n and then letting ε goes to 0, we obtain that $\mu_n(f) \to \mu(f)$. □

C Proofs of the technical lemmas of Section 4

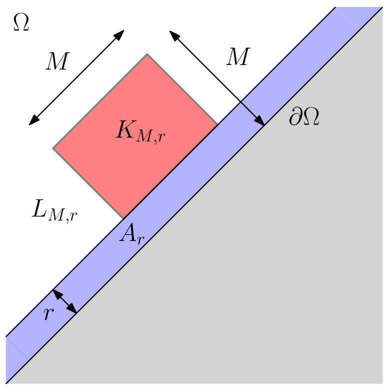

Figure 5: Partition of Ω used in the proof of Lemma 4.2.

Proof of Lemma 4.2. For the direct implication, take $\nu = 0$ and apply Theorem 3.4.

Let us prove the converse implication. Assume that $\mu_n \xrightarrow{v} \mu$ and $D_p(\mu_n, \nu) \to D_p(\mu, \nu)$ for some $\nu \in \mathcal{D}^p$. The vague convergence of $(\mu_n)_n$ implies that μ^p is the only possible accumulation point for weak convergence of the sequence $(\mu_n^p)_n$. Therefore, it is sufficient to show that the sequence $(\mu_n^p)_n$ is relatively compact for weak convergence (i.e. tight and bounded in total variation, see Proposition A.2). Indeed, this would mean that (μ_n^p) converges weakly to μ^p, or equivalently by Proposition A.3 that $\mu_n \xrightarrow{v} \mu$ and $\mathrm{Pers}_p(\mu_n) \to \mathrm{Pers}_p(\mu)$. The conclusion is then obtained thanks to Theorem 3.4.

Thus, let $(\mu_n)_n$ be any subsequence and $(\pi_n)_n$ be corresponding optimal transport plans. The vague convergence of $(\mu_n)_n$ implies that $(\pi_n)_n$ is relatively compact with respect to the vague convergence on E_Ω. Let π be a limit of any converging subsequence of $(\pi_n)_n$, which indexes are still denoted by n. By [18, Proposition

2.3], $\pi \in \mathrm{Opt}(\mu,\nu)$. For $r>0$, define $A_r := \{x \in \Omega,\ d(x,\partial\Omega) \le r\}$ and write $\overline{A}_r$ for $A_r \cup \partial\Omega$. Consider $\eta > 1$. We can write

$$\begin{aligned}
&\int_{A_r} d(x,\partial\Omega)^p \mathrm{d}\mu_n(x) = \iint\limits_{A_r\times\overline{\Omega}} d(x,\partial\Omega)^p \mathrm{d}\pi_n(x,y)\\
&= \iint\limits_{A_r\times A_{\eta r}^c} d(x,\partial\Omega)^p \mathrm{d}\pi_n(x,y) + \iint\limits_{\overline{A}_r\times\overline{A}_{\eta r}} d(x,\partial\Omega)^p \mathrm{d}\pi_n(x,y)\\
&\overset{(*)}{\le} \frac{1}{(\eta-1)^p} \iint\limits_{A_r\times A_{\eta r}^c} d(x,y)^p \mathrm{d}\pi_n(x,y) + \iint\limits_{\overline{A}_r\times\overline{A}_{\eta r}} d(x,\partial\Omega)^p \mathrm{d}\pi_n(x,y)\\
&\le \frac{1}{(\eta-1)^p} D_p^p(\mu_n,\nu) + 2^{p-1}\left(\iint\limits_{\overline{A}_r\times\overline{A}_{\eta r}} d(x,y)^p \mathrm{d}\pi_n(x,y) + \iint\limits_{\overline{A}_r\times\overline{A}_{\eta r}} d(y,\partial\Omega)^p \mathrm{d}\pi_n(x,y)\right)\\
&\le \frac{1}{(\eta-1)^p} D_p^p(\mu_n,\nu) + 2^{p-1}\left(D_p^p(\mu_n,\nu) - \iint\limits_{E_\Omega\backslash(\overline{A}_r\times\overline{A}_{\eta r})} d(x,y)^p \mathrm{d}\pi_n(x,y) + \int_{A_{\eta r}} d(y,\partial\Omega)^p \mathrm{d}\nu(y)\right)
\end{aligned}$$

where $(*)$ holds because $d(x,y) \ge (\eta-1)r \ge (\eta-1)d(x,\partial\Omega)$ for $(x,y) \in A_r \times A_{\eta r}^c$. Therefore,

$$\begin{aligned}
\limsup_{n\to\infty} \int_{A_r} d(x,\partial\Omega)^p \mathrm{d}\mu_n(x) \le& \frac{1}{(\eta-1)^p} D_p^p(\mu,\nu) + 2^{p-1}\Bigg(D_p^p(\mu,\nu)\\
&- \iint\limits_{E_\Omega\backslash(\overline{A}_r\times\overline{A}_{\eta r})} d(x,y)^p \mathrm{d}\pi(x,y) + \int_{A_{\eta r}} d(y,\partial\Omega)^p \mathrm{d}\nu(y)\Bigg)
\end{aligned}$$

Note that at the last line, we used Portmanteau theorem (see Proposition A.4) on the sequence of measures $(d(x,y)^p \mathrm{d}\pi_n(x,y))_n$ for the open set $E_\Omega\backslash(\overline{A}_r \times \overline{A}_{\eta r})$. Letting r goes to 0, then η goes to infinity, one obtains

$$\lim_{r\to 0} \limsup_{n\to\infty} \int_{A_r} d(x,\partial\Omega)^p \mathrm{d}\mu_n(x) = 0.$$

The second part consists in showing that there can not be mass escaping "at infinity" in the subsequence $(\mu_n^p)_n$. Fix $r, M > 0$. For $x \in \Omega$, denote $s(x)$ the projection of x on $\partial\Omega$. Pose

$$K_{M,r} := \{x \in \overline{A}_r^c,\ d(x,\partial\Omega) < M, \|s(x)\| < M\}$$

and $L_{M,r}$ the closure of $\overline{A_r}^c \backslash K_{M,r}$ (see Figure 5). For $r' > 0$,

$$\begin{aligned}
&\int_{L_{M,r}} d(x,\partial\Omega)^p \mathrm{d}\mu_n(x) = \iint\limits_{L_{M,r}\times\overline{\Omega}} d(x,\partial\Omega)^p \mathrm{d}\pi_n(x,y)\\
&= \iint\limits_{L_{M,r}\times(L_{M/2,r'}\cup\overline{A}_{r'})} d(x,\partial\Omega)^p \mathrm{d}\pi_n(x,y) + \iint\limits_{L_{M,r}\times K_{M/2,r'}} d(x,\partial\Omega)^p \mathrm{d}\pi_n(x,y)\\
&\le 2^{p-1} \iint\limits_{L_{M,r}\times(L_{M/2,r'}\cup\overline{A}_{r'})} d(x,y)^p \mathrm{d}\pi_n(x,y)\\
&+ 2^{p-1} \iint\limits_{L_{M,r}\times(L_{M/2,r'}\cup\overline{A}_{r'})} d(\partial\Omega,y)^p \mathrm{d}\pi_n(x,y)\\
&+ \iint\limits_{L_{M,r}\times K_{M/2,r'}} d(x,\partial\Omega)^p \mathrm{d}\pi_n(x,y).
\end{aligned}$$

We treat the three parts of the sum separately. As before, taking the $\limsup$ in n and letting M goes to ∞, the first part of the sum converges to 0 (apply Portmanteau theorem on the open set $E_\Omega\backslash(L_{M,r}\times(L_{M/2,r'}\cup\overline{A}_{r'}))$. The second part is smaller than

$$2^{p-1} \int_{L_{M/2,r'}\cup A_{r'}} d(y,\partial\Omega)^p \mathrm{d}\nu(y),$$

which converges to 0 as $M \to \infty$ and $r' \to 0$. For the third part, notice that if $(x,y) \in L_{M,r} \times K_{M/2,r'}$, then

$$d(x,\partial\Omega) \le d(x,s(y)) \le d(x,y) + d(y,s(y)) \le d(x,y) + \frac{M}{2} \le 2d(x,y).$$

Therefore,

$$\begin{aligned}\iint\limits_{L_{M,r}\times K_{M/2,r'}} d(x,\partial\Omega)^p \mathrm{d}\pi_n(x,y) &\le 2^p \iint\limits_{L_{M,r}\times K_{M/2,r'}} d(x,y)^p \mathrm{d}\pi_n(x,y) \\ &\le 2^p \iint\limits_{L_{M,r}\times\overline{\Omega}} d(x,y)^p \mathrm{d}\pi_n(x,y).\end{aligned}$$

As before, it is shown that $\limsup_n \iint_{L_{M,r}\times\overline{\Omega}} d(x,y)^p \mathrm{d}\pi_n(x,y)$ converges to 0 when M goes to infinity by applying Portmanteau theorem on the open set $E_\Omega \backslash (L_{M,r} \times \overline{\Omega})$.

Finally, we have shown, that by taking r small enough and M large enough, one can find a compact set $\overline{K_{M,r}}$ such that $\int_{\Omega\backslash\overline{K_{M,r}}} d(x,\partial\Omega)^p \mathrm{d}\mu_n = \mu_n^p(\Omega\backslash\overline{K_{M,r}})$ is uniformly small: $(\mu_n^p)_n$ is tight. As we have

$$\mu_n^p(\Omega) = \mathrm{Pers}_p(\mu_n) = D_p^p(\mu_n, 0) \le (D_p(\mu_n,\nu) + D_p(\nu,0))^p \to (D_p(\mu,\nu) + D_p(\nu,0))^p,$$

it is also bounded in total variation. Hence, $(\mu_n^p)_n$ is relatively compact for the weak convergence: this concludes the proof. □

Proof of Lemma 4.3. Given $\mathbb{P} = \sum_{i=1}^N \lambda_i \delta_{a_i}$ a probability distribution with $a_i \in \mathcal{D}_f$, it suffices to show that $\tilde{\mathbb{P}} = \sum_{i=1}^N \lambda_i \delta_{\tilde{a_i}}$, where $\tilde{a_i} = a_i + (m_{\mathrm{tot}} - m_i)\delta_{\partial\Omega}$—where $m_i = a_i(\Omega)$—has a barycenter $\tilde{a}$ which is a finite point measure supported on $\tilde{\Omega}$ (of mass m_{tot}). Indeed, taking the restriction of $\tilde{a}$ to Ω will then give a barycenter of $\mathbb{P}$ according to Proposition 4.1.

Let thus fix $m \in \mathbb{N}$, and let $\tilde{a_1}, \dots, \tilde{a_N}$ be point measures of mass m in $\tilde{\Omega}$. Write $\tilde{a_i} = \sum_{j=1}^m \delta_{x_{i,j}}$, so that $x_{i,j} \in \tilde{\Omega}$ for $1 \le i \le N,\ 1 \le j \le m$, with the $x_{i,j}$s non-necessarily distinct. Define

$$T : (x_1, \dots, x_N) \in \tilde{\Omega}^N \mapsto \arg\min \left\{ \sum_{i=1}^N \lambda_i \rho(x_i, y)^p,\ y \in \tilde{\Omega} \right\} \in \tilde{\Omega}.$$

Since we assume $p > 1$, T is well-defined and is continuous (the minimizer is unique by strict convexity). Note that the proof holds if we replace the $\tilde{a_i}$s by any point measures with the same mass in some separable metric space (X, d) for which the application T is measurable. Using the localization property stated in [8, Section 2.2], we know that the support of a barycenter is included in the finite set

$$S := \{T(x_{1,j_1}, \dots, x_{N,j_N}),\ 1 \le j_1, \dots, j_N \le m\}.$$

Let $K = m^N$ and let $z_1, \dots, z_K$ be an enumeration of the points of S. Denote by $\mathrm{Gr}(z_k)$ the N elements $x_1, \dots, x_N$, with $x_i \in \mathrm{spt}(\tilde{a_i})$, such that $z_k = T(x_1, \dots, x_N)$. It is explained in [8, Section 2.3], that finding a barycenter of $\tilde{\mathbb{P}}$ is equivalent to finding a minimizer of the problem

$$\inf_{(\gamma_1,\dots,\gamma_N)\in\Pi} \sum_{i=1}^N \lambda_i \iint_{\tilde{\Omega}^2} \rho(x_i, y)^p \mathrm{d}\gamma_i(x_i, y), \tag{27}$$

where Π is the set of plans $(\gamma_i)_{i=1,\dots,N}$, with γ_i having for first marginal $\tilde{a_i}$, and such that all γ_is share the same second marginal. For such a minimizer, the common second marginal is a barycenter of $\tilde{\mathbb{P}}$. Therefore, if we prove that there exists a minimizer of (27) which is a point measure, then the lemma is proven.

A potential minimizer of (27) is described by a vector $\gamma = (\gamma_{i,j,k}) \in \mathbb{R}_+^{NmK}$ such that:

$$\begin{cases} \text{for } 1 \le i \le N,\ 1 \le j \le m, & \sum_{k=1}^K \gamma_{i,j,k} = 1 \text{ and} \\ \text{for } 2 \le i \le N,\ 1 \le k \le K, & \sum_{j=1}^m \gamma_{1,j,k} = \sum_{j=1}^m \gamma_{i,j,k}. \end{cases} \tag{28}$$

Let $c \in \mathbb{R}^{NmK}$ be the vector defined by $c_{i,j,k} = \mathbf{1}\{x_{i,j} \in \mathrm{Gr}(z_k)\}\lambda_i \rho(x_{i,j}, z_k)^p$. Then, the problem (27) is equivalent to

$$\underset{\gamma \in R_+^{NmK}}{\mathrm{minimize}}\, \gamma^T c \quad \text{under the constraints (28).}$$

This Linear Programming problem (see [34, Section 5.15]) has an integer solution if the polyhedron described by the equations (28) is integer (i.e. its vertices have integer values). The constraints (28) are described by a matrix A of size $(Nm + (N-1)K) \times NmK$ and a vector $b = [\mathbf{1}_{Nm}, \mathbf{0}_{(N-1)K}]$, such that $\gamma \in \mathbb{R}^{NmK}$ satisfies (28) if and only if $A\gamma = b$. A sufficient condition for the polyhedron $\{Ax \le b\}$ to be integer is to satisfy the following property (see [34, Section 5.17]): for all $u \in \mathbb{Z}^{NmK}$, the dual problem

$$\max\{y^T b,\ y \ge 0 \text{ and } y^T A = u\} \tag{29}$$

has either no solution (i.e. there is no $y \ge 0$ satisfying $y^T A = u$), or it has an integer optimal solution y.

For y satisfying $y^T A = u$, write $y = [y^0, y^1]$ with $y^0 \in \mathbb{R}^{Nm}$ and $y^1 \in \mathbb{R}^{(N-1)K}$, so that y^0 is indexed on $1 \le i \le N,\ 1 \le j \le m$ and y^1 is indexed on $2 \le i \le N,\ 1 \le k \le K$. One can check that, for $2 \le i \le N,\ 1 \le j \le m,\ 1 \le k \le K$:

$$u_{1,j,k} = y^0_{1,j} + \sum_{i'=2}^{N} y^1_{i',k} \quad \text{and} \quad u_{i,j,k} = y^0_{i,j} - y^1_{i,k}, \tag{30}$$

so that,

$$\begin{aligned} y^T b &= \sum_{i=1}^{N}\sum_{j=1}^{m} y^0_{i,j} = \sum_{j=1}^{m} y^0_{1,j} + \sum_{i=2}^{N}\sum_{j=1}^{m} y^0_{i,j} \\ &= \sum_{j=1}^{m}(u_{1,j,k} - \sum_{i=2}^{N} y^1_{i,k}) + \sum_{i=2}^{N}\sum_{j=1}^{m}(u_{i,j,k} + y^1_{i,k}) \\ &= \sum_{i=1}^{N}\sum_{j=1}^{m} u_{i,j,k}. \end{aligned}$$

Therefore, the function $y^T b$ is constant on the set $P := \{y \ge 0,\ y^T A = u\}$, and any point of the set is an argmax. We need to check that if the set P is non-empty, then it contains a vector with integer coordinates: this would conclude the proof. A solution of the homogeneous equation $y^T A = 0$ satisfies $y^0_{i,j} = y^1_{i,k} = \lambda_i$ for $i \ge 2$ and $y^0_{1,j} = -\sum_{i=2}^{N} y^1_{i,k} = -\sum_{i=2}^{N} \lambda_i$ and reciprocally, any choice of $\lambda_i \in \mathbb{R}$ gives rise to a solution of the homogeneous equation. For a given u, one can verify that the set of solutions of $y^T A = u$ is given, for $\lambda_i \in \mathbb{R}$, by

$$\begin{cases} y^0_{1,j} = \sum_{i=1}^{N} u_{i,j,k} - \sum_{i=2}^{N} \lambda_i \\ y^0_{i,j} = \lambda_i \text{ for } i \ge 2, \\ y^1_{i,k} = -u_{i,j,k} + \lambda_i \text{ for } i \ge 2. \end{cases}$$

Such a solution exists if and only if for all j, $U_j := \sum_{i=1}^{N} u_{i,j,k}$ does not depend on k and for $i \ge 2$, $U_{i,k} := u_{i,j,k}$ does not depend on j. For such a vector u, P corresponds to the $\lambda_i \ge 0$ with $\lambda_i \ge \max_k U_{i,k}$ and $U_j \ge \sum_{i=1}^{N} \lambda_i$. If this set is non empty, it contains as least the point corresponding to $\lambda_i = \max_k U_{i,k}$, which is an integer: this point is integer valued, concluding the proof. □

D Technical details regarding Section 5.2

Define $\mathcal{M}_\pm$ the space of signed measures on Ω, i.e. a measure $\mu \in \mathcal{M}_\pm$ is written $\mu_+ - \mu_-$ for two finite measures $\mu_+, \mu_- \in \mathcal{M}_f$. The total variation distance $|\cdot|$ is a norm on $\mathcal{M}_\pm$, and $(\mathcal{M}_\pm, |\cdot|)$ is a Banach space. The Bochner integral [5] is a generalization of the Lebesgue integral for functions taking their values in Banach space. We define the expected persistence measure of $\mathbb{P} \in \mathcal{W}^p(\mathcal{M}^p)$ as the Bochner integral of some pushforward of $\mathbb{P}$. More precisely, define

$$\begin{aligned} F : (\mathcal{M}^p, D_p) &\to (\mathcal{M}_\pm, |\cdot|) \\ \mu &\mapsto \mu^p. \end{aligned}$$

Note that F has an inverse G on $\mathcal{M}_f$, defined by $G(\nu)(f) := \int_\Omega \frac{f(x)}{d(x,\Omega)^p} \mathrm{d}\nu(x)$. Theorem 3.4 implies that G is a continuous function from $(\mathcal{M}_f, |\cdot|)$ to $(\mathcal{M}^p, D_p)$. In particular, as $\mathcal{M}_f$ and $\mathcal{M}^p$ are Polish spaces, F is measurable (see [23, Theorem 15.1]). For $\mathbb{P} \in \mathcal{W}^p(\mathcal{M}^p(\Omega))$, define for $\boldsymbol{\mu} \sim \mathbb{P}$, $\mathbb{E}[\boldsymbol{\mu}]$ the *linear* expectation of $\mathbb{P}$ by

$$\mathbb{E}[\boldsymbol{\mu}] := G\left(\int \nu \mathrm{d}(F_\# \mathbb{P})(\nu)\right) \in \mathcal{M}^p, \tag{31}$$

where the integral is the Bochner integral on the Banach space $(\mathcal{M}_\pm, |\cdot|)$ and $F_\#\mathbb{P}$ is the pushforward of $\mathbb{P}$ by F. It is straightforward to check that $\mathbb{E}[\boldsymbol{\mu}]$ defined in that way satisfies the relation

$$\forall K \subset \Omega \text{ compact, } \mathbb{E}[\boldsymbol{\mu}](K) = \mathbb{E}[\boldsymbol{\mu}(K)].$$

The proof of Proposition 5.2 consists in applying Jensen's inequality in an infinite-dimensional setting. We first show that the function D_p^p is convex.

Lemma D.1. *For $1 \leq p < \infty$, the function $D_p^p : \mathcal{M}^p \times \mathcal{M}^p \to \mathbb{R}$ is convex.*

Proof. Fix $\mu_1, \mu_2, \nu_1, \nu_2 \in \mathcal{M}^p$ and $t \in [0,1]$. Our goal is to show that

$$D_p^p(t\mu_1 + (1-t)\mu_2, t\nu_1 + (1-t)\nu_2) \leq tD_p^p(\mu_1, \nu_1) + (1-t)D_p^p(\mu_2, \nu_2).$$

Let $\pi_{11} \in \mathrm{Opt}_p(\mu_1, \nu_1)$ and $\pi_{22} \in \mathrm{Opt}_p(\mu_2, \nu_2)$. It is straightforward to check that $\pi := t\pi_{11} + (1-t)\pi_{22}$ is an admissible plan between $t\mu_1 + (1-t)\mu_2$ and $t\nu_2 + (1-t)\nu_2$. The cost of this admissible plan is $tD_p^p(\mu_1, \nu_1) + (1-t)D_p^p(\mu_2, \nu_2)$, which is therefore larger than $D_p^p(t\mu_1 + (1-t)\mu_2, t\nu_1 + (1-t)\nu_2)$. $\square$

We then use the following general result:

Proposition D.1 (see [40])**.** *Let $(V, \|\cdot\|)$ be a separable Banach space and $\mathcal{X}$ a closed convex subset of V with non-empty interior. Let $\mathbb{P}$ be a probability measure on $\mathcal{X}$ endowed with its Borelian σ-algebra. Assume that $\int \|x\| \mathrm{d}\mathbb{P}(x) < \infty$. Let $f : \mathcal{X} \to \mathbb{R}$ be a continuous convex function so that $\int f(x)\mathrm{d}\mathbb{P}(x) < \infty$. Then,*

$$f\left(\int x \mathrm{d}\mathbb{P}(x)\right) \leq \int f(x)\mathrm{d}\mathbb{P}(x).$$

To conclude, take $V = \mathcal{M}_\pm \times \mathcal{M}_\pm$, $\mathcal{X} = \mathcal{M}_f \times \mathcal{M}_f$ and $f = D_p^p \circ (G, G) : \mathcal{X} \to \mathbb{R}$. The continuity of G implies that f is continuous and Lemma D.1 implies the convexity of f. Apply Proposition D.1 with π to conclude.