



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

Vincent Divol, Théo Lacombe

## ► To cite this version:

Vincent Divol, Théo Lacombe. Understanding the Topology and the Geometry of the Space of Persistence Diagrams via Optimal Partial Transport. *Journal of Applied and Computational Topology*, Springer, 2020, 10.1007/s41468-020-00061-z . hal-01970466v3

**HAL Id: hal-01970466**

**<https://hal.inria.fr/hal-01970466v3>**

Submitted on 17 Jul 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Vincent Divol and Théo Lacombe

`firstname.lastname@inria.fr`  
Datashape, Inria Saclay

## Abstract

Despite the obvious similarities between the metrics used in topological data analysis and those of optimal transport, an optimal-transport based formalism to study persistence diagrams and similar topological descriptors has yet to come. In this article, by considering the space of persistence diagrams as a space of discrete measures, and by observing that its metrics can be expressed as optimal partial transport problems, we introduce a generalization of persistence diagrams, namely Radon measures supported on the upper half plane. 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. We explore 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, a geometric description of barycenters (Fréchet means) for any distribution of diagrams, and an exhaustive description of continuous linear representations of persistence diagrams. We also showcase the strength of this framework to study random persistence diagrams by providing several statistical results made meaningful thanks to this new formalism.

## 1 Introduction

### 1.1 Framework and motivations

Topological Data Analysis (TDA) is an emerging field in data analysis that has found applications in computer vision [47], material science [34, 41], shape analysis [14, 61], 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 the presence of a topological feature (connected component, loop,

hole, etc.) appearing at some scale in the *filtration* of an object  $\mathbb{X}$ . A complete description of the *persistent homology* machinery is not necessary for this work and the interested reader can refer to [26] for an introduction. The space of persistence diagrams, denoted by  $\mathcal{D}$  in the following, is usually equipped with partial matching extended metrics  $d_p$  (i.e. it can happen that  $d_p(\mu, \nu) = +\infty$  for some  $\mu, \nu \in \mathcal{D}$ ), sometimes called Wasserstein distances [26, Chapter VIII.2]: for  $p \in [1, +\infty)$  and  $a, b$  in  $\mathcal{D}$ , define

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

where  $d(\cdot, \cdot)$  denotes the  $q$ -norm on  $\mathbb{R}^2$  for some  $1 \leq q \leq \infty$ ,  $\Gamma(a, b)$  is the set of partial matchings between  $a$  and  $b$ , i.e. bijections between  $a \cup \partial\Omega$  and  $b \cup \partial\Omega$ , and  $\partial\Omega := \{(t, t), t \in \mathbb{R}\}$  is the boundary of  $\Omega$ , namely the diagonal (see Figure 1). When  $p \rightarrow \infty$ , we recover the so-called *bottleneck distance*:

$$d_\infty(a, b) := \inf_{\pi \in \Gamma(a, b)} \sup_{x \in a \cup \partial\Omega} d(x, \pi(x)). \quad (2)$$

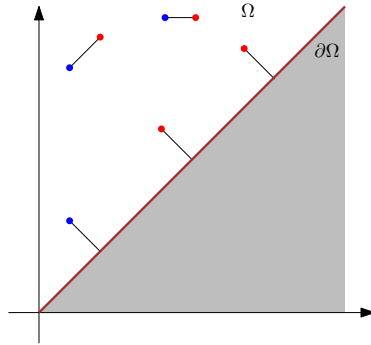


Figure 1: An example of optimal partial matching between two diagrams. The bottleneck distance between these two diagrams is the length of the longest edge in this matching, while their Wasserstein distance  $d_p$  is the  $p$ -th root of the sum of all edge lengths to the power  $p$ .

An equivalent viewpoint, developed in [16, Chapter 3], is to define 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  $\Omega$  with integer mass on each point of its support. This measure-based perspective suggests to consider more general Radon measures<sup>1</sup> supported on the upper half-plane  $\Omega$ . Besides this theoretical motivation, considering such measures allows us to address statistical and learning problems that appear in different applications of TDA:

(A1) **Continuity of representations.** When given a sample of persistence diagrams  $a_1, \dots, a_N$ , a common way to perform machine learning is to first map the diagrams into a vector space thanks to a *representation* (or

<sup>1</sup>A Radon measure supported on  $\Omega$  is a (Borel) measure that gives a finite mass to any compact subset  $K \subset \Omega$ . See Appendix A for a short reminder about measure theory.

feature map)  $\Phi : \mathcal{D} \rightarrow \mathcal{B}$ , where  $\mathcal{B}$  is a Banach space. In order to ensure the meaningfulness of the machine learning procedure, the stability of the representations with respect to the  $d_p$  distances is usually required. One of our contribution to the matter is to formulate an equivalence between  $d_p$ -convergence and convergence in terms of measures (Theorem 3.4). This result allows us to characterize a large class of continuous representations (Prop. 5.1) that includes some standard tools used in TDA such as the Betti curve [62], the persistence surface [1] and the persistence silhouette [18].

(A2) **Law of large numbers for diagrams generated by random point clouds.** A popular problem that generates random persistence diagrams is given by filtrations built on top of large random point clouds: if  $\mathbb{X}_n$  is a  $n$ -sample of i.i.d. points on, say, the cube  $[0, 1]^d$ , recent articles [32, 25] have investigated the asymptotic behavior of the persistence diagram  $a_n$  of the Čech (or Rips) filtration built on top of the rescaled point cloud  $n^{1/d}\mathbb{X}_n$ . In particular, it has been shown in [32] that the sequence of measures  $n^{-1}a_n$  converges vaguely to some limit measure  $\mu$  supported on  $\Omega$  (that is *not* a persistence diagram), and in [25] that the moments of  $n^{-1}a_n$  also converge to the moments of  $\mu$ . An interesting problem is to build a metric which generalizes  $d_p$  and for which the convergence of  $n^{-1}a_n$  to  $\mu$  holds.

(A3) **Stability of the expected diagrams.** Of particular interest in the literature are *linear* representations, that is of the form  $\Phi(a) := a(f)$ , the integral of a function  $f : \Omega \rightarrow \mathcal{B}$  against a persistence diagram  $a$  (seen as a measure). Given  $N$  i.i.d. diagrams  $a_1, \dots, a_N$  following some law  $P$ , and a linear representation  $\Phi$ , a natural object to consider is the sample mean  $N^{-1}(\Phi(a_1) + \dots + \Phi(a_N)) = \Phi(N^{-1}(a_1 + \dots + a_N))$ . By the law of large numbers, this quantity converges to  $\mathbb{E}_P[\mathbf{a}](f)$ , where  $\mathbb{E}_P[\mathbf{a}]$  is the expected persistence diagram of the process, introduced in [24]. Understanding how the object  $\mathbb{E}_P[\mathbf{a}]$  depends on the underlying process  $P$  generating  $\mathbf{a}$  invites one to define a notion of distance between  $\mathbb{E}_P[\mathbf{a}]$  and  $\mathbb{E}_{P'}[\mathbf{a}]$  for  $P, P'$  two distributions on the space of persistence diagrams, and relate this distance to a similarity measure between  $P$  and  $P'$ . In the same way that the expected value of an integer-valued random variable may not be an integer, the objects  $\mathbb{E}_P[\mathbf{a}]$  are not persistence diagrams in general, but Radon measures on  $\Omega$ . Therefore, extending the distances  $d_p$  to Radon measures in a consistent way will allow us to assess the closeness between those quantities.

**Remark 1.1.** *Note that, throughout this article, we consider persistence diagrams with possibly infinitely many points (although locally finite). This is motivated from a statistical perspective. Indeed, the space of finite persistence diagram is lacking completeness, as highlighted in [48, Definition 2], so that, for instance, the expectation of a probability distribution of diagrams with finite numbers of points may have an infinite mass. An alternative approach to recover a complete space is to study the space of persistence diagrams with total mass less than or equal to some fixed  $m \geq 0$ . However, this might be unsatisfactory as the number of points of a persistence diagram is known to be an unstable quantity with respect to perturbations of the input data. Note that infinite persistence diagrams may also help to model the topology of standard objects: for instance,*

the (random) persistence diagram built on the sub-level sets of a Brownian motion has infinitely many points (see [24, Section 6]). However, numerical applications generally involve finite sets of finite diagrams, which are studied in Section 3.2.

**Remark 1.2.** In general, persistence diagrams may contain points with coordinates of the form  $(t, +\infty)$ , called the essential points of the diagram. The distance between two persistence diagrams is then defined as the sum of two independent terms: the cost  $d_p$  that handle points with finite coordinates and the cost of a simple one-dimensional optimal matching between the first coordinates of the essential points (set to be  $+\infty$  if the cardinalities of the essential parts differ). We focus on persistence diagrams with only points with finite coordinates for the sake of simplicity, but all the results stated in this work may easily be adapted to the more general case including points with infinite coordinates.

## 1.2 Outline and main contributions

Examples (A2) and (A3) motivate the introduction of metrics on the space  $\mathcal{M}$  of Radon measures supported on  $\Omega$ , which generalize the distances  $d_p$  on  $\mathcal{D}$ : these are presented in Section 2. 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

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

where  $d(x, \partial\Omega) := \inf_{y \in \partial\Omega} d(x, y)$  is the distance from a point  $x \in \Omega$  to (its orthogonal projection onto) the diagonal  $\partial\Omega$ , and we define

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

We equip  $\mathcal{M}^p$  with metrics  $\text{OT}_p$  (see Definition 2.1), originally introduced in a work of Figalli and Gigli [28]. We show in Proposition 3.2 that  $\text{OT}_p$  and  $d_p$  coincide on  $\mathcal{D}^p := \mathcal{D} \cap \mathcal{M}^p$ , making  $\text{OT}_p$  a good candidate to address the questions raised in (A2) and (A3). To emphasize that we equip the space of Radon measures with a specific metric designed for our purpose, we will refer to elements of the metric space  $(\mathcal{M}^p, \text{OT}_p)$  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  $\mu$ , denoted by  $\mu_n \xrightarrow{v} \mu$ , if for any continuous compactly supported function  $f : \Omega \rightarrow \mathbb{R}$ ,  $\mu_n(f) \rightarrow \mu(f)$  (where the notation  $\mu(f) := \int f d\mu$  stands for the integration of the function  $f$  against  $\mu$ ). We prove the following equivalence between convergence for the metric  $\text{OT}_p$  and the vague convergence:

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

$$\text{OT}_p(\mu_n, \mu) \rightarrow 0 \Leftrightarrow \begin{cases} \mu_n \xrightarrow{v} \mu, \\ \text{Pers}_p(\mu_n) \rightarrow \text{Pers}_p(\mu). \end{cases} \quad (5)$$

This equivalence gives a positive answer to the issues raised by (A2), as detailed in Section 5. Note also that this characterization in particular holds for persistence diagrams in  $\mathcal{D}^p$ , and can thus be helpful to show the convergence

or the tightness of a sequence of diagrams. This theorem is analogous to the characterization of convergence of probability measures with respect to Wasserstein distances (see [64, Theorem 6.9]). A proof for Radon measures supported on a common bounded set can be found in [28, Proposition 2.7]. Our contribution consists in extending this result to non-bounded sets, in particular to the upper half plane  $\Omega$ .

Section 3.2 is dedicated to sets of measures with finite masses, appearing naturally in numerical applications. We show in particular that computing the  $\text{OT}_p$  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* mass (Prop. 3.7), a result having practical implications for the computation of  $\text{OT}_p$  distances between persistence measures (and diagrams).

Section 3.3 studies the case  $p = \infty$ , which is somewhat ill-behaved from a statistical analysis point of view (for instance, the space of persistence diagrams endowed with the bottleneck metric is not separable, as observed in [10, Theorem 5]), but is also of crucial interest in TDA as it is motivated by algebraic considerations [50] and satisfies stronger stability results [21] than its  $p < \infty$  counterparts [22]. In particular, we give in Propositions 3.11 and 3.13 a characterization of bottleneck convergence (in the vein of Theorem 3.4) for persistence diagrams satisfying some finiteness assumptions (namely, for each  $r > 0$ , the number of points with persistence greater than  $r$  must be finite).

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 [48, 60], where authors prove their existence for certain types of distributions [48, Theorem 28]. Using the framework of persistence measures, we show that this existence result is actually true for any distribution of persistence diagrams (and measures) with finite moment. Namely, we prove the following results:

**Theorem 4.3.** *Assume that  $1 < p < \infty$  and that  $d(\cdot, \cdot)$  denotes the  $q$ -norm for  $1 < q < \infty$ . For any probability distribution  $P$  supported on  $\mathcal{M}^p$  with finite  $p$ -th moment, the set of  $p$ -Fréchet means of  $P$  is a non-empty compact convex subset of  $\mathcal{M}^p$ .*

**Theorem 4.4.** *Assume that  $1 < p < \infty$  and that  $d(\cdot, \cdot)$  denotes the  $q$ -norm for  $1 < q < \infty$ . For any probability distribution  $P$  supported on  $\mathcal{D}^p$  with finite  $p$ -th moment, the set of  $p$ -Fréchet means of  $P$ , which is a subset of  $\mathcal{M}^p$ , contains an element in  $\mathcal{D}^p$ . Furthermore, if  $P$  is supported on a finite set of finite persistence diagrams, then the set of the  $p$ -Fréchet means of  $P$  is a convex set whose extreme points are in  $\mathcal{D}^p$ .*

Section 5 applies the formalism we developed to address the questions raised in (A1)—(A3). In Section 5.1, we prove a strong characterization of continuous linear representations of persistence measures (and diagrams), which answers to the issue raised by (A1) for the class of linear representations (see Figure 2).

**Proposition 5.1** *Let  $p \in [1, +\infty)$ ,  $d \geq 1$ , and  $f : \Omega \rightarrow \mathcal{B}$  for some Banach space  $\mathcal{B}$  (e.g.  $\mathbb{R}^d$ ). The representation  $\Phi : \mathcal{M}^p \rightarrow \mathcal{B}$  defined by  $\Phi(\mu) = \int_{\Omega} f(x) d\mu(x)$  is continuous with respect to  $\text{OT}_p$  if and only if  $f$  is of the form  $f(x) = g(x)d(x, \partial\Omega)^p$ , where  $g : \Omega \rightarrow \mathcal{B}$  is a continuous bounded map.*

This new result can be compared to the recent work [36, Theorem 13], which

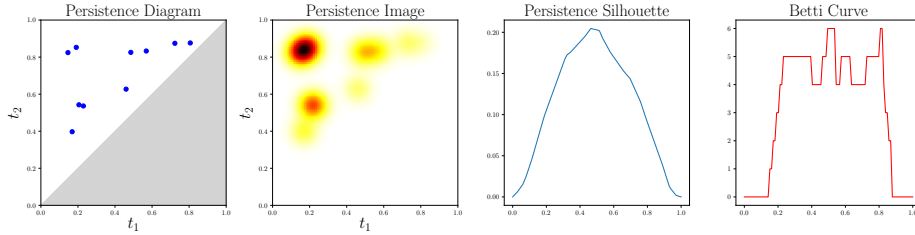


Figure 2: Some common linear representations of persistence diagrams. From left to right: A persistence diagram. Its persistence surface [1], which is a persistence measure. The corresponding persistence silhouette [18]. The corresponding Betti Curve [62]. See Section 5.1 for details.

gives a similar result in the case  $\mathcal{B} = \mathbb{R}$  on the space of finite persistence diagrams  $\mathcal{D}_f$ ), or the works [42, Proposition 8] and [25, Theorem 3], which show that linear representations can have more regularity (e.g. Lipschitz or Hölder) under additional assumptions.

Section 5.2 states a very concise law of large for persistence diagrams. Namely, building on the previous works [35, 25] along with Theorem 3.4, we prove the following:

**Proposition 5.3** *Let  $\mathbb{X}_n = \{X_1, \dots, X_n\}$  be a sample of  $n$  points on the  $d$ -dimensional cube  $[0, 1]^d$ , sampled from a density bounded from below and from above by positive constants, and let  $\mu_n = \frac{1}{n} \text{Dgm}(n^{1/d} \mathbb{X}_n)$ , where  $\text{Dgm}(n^{1/d} \mathbb{X}_n)$  is either the Rips or Čech complex built on the point cloud  $n^{1/d} \mathbb{X}_n$ . Then, there exists a measure  $\mu \in \mathcal{M}^p$  such that  $\text{OT}_p(\mu_n, \mu) \rightarrow 0$ .*

Finally, Section 5.3 considers the problem (A3), that is the stability of the expected persistence diagrams. In particular, we prove a stability result between an input point cloud  $\mathbb{X}_n$  in a random setting and its expected (Čech) diagrams  $\mathbb{E}(\text{Dgm}(\mathbb{X}_n))$ :

**Proposition 5.5** *Let  $\xi, \xi'$  be two probability measures supported on  $\mathbb{R}^d$ . Let  $\mathbb{X}_n$  (resp.  $\mathbb{X}'_n$ ) be a  $n$ -sample of law  $\xi$  (resp.  $\xi'$ ). Then, for any  $k > d$ , and any  $p \geq k + 1$ ,*

$$\text{OT}_p^p(\mathbb{E}[\text{Dgm}(\mathbb{X}_n)], \mathbb{E}[\text{Dgm}(\mathbb{X}'_n)]) \leq C_{k,d} \cdot n \cdot W_{p-k}^{p-k}(\xi, \xi') \quad (6)$$

where  $C_{k,d} := C \text{diam}(\mathbb{X})^{k-d} \frac{k}{k-d}$  for some constant  $C$  depending only on  $\mathbb{X}$ . In particular, letting  $p \rightarrow \infty$ , we obtain a bottleneck stability result:

$$\text{OT}_\infty(\mathbb{E}[\text{Dgm}(\mathbb{X}_n)], \mathbb{E}[\text{Dgm}(\mathbb{X}'_n)]) \leq W_\infty(\xi, \xi'). \quad (7)$$

## 2 Elements of optimal partial transport

In this section,  $(\mathcal{X}, d)$  denotes a Polish metric space.

## 2.1 Optimal transport between probability measures and Wasserstein distances

In its standard formulation, optimal transport is a widely developed theory providing tools to study and compare probability measures supported on  $\mathcal{X}$  [63, 64, 54], that is—up to a renormalization factor—non-negative measures of the same mass. Given two probability measures  $\mu, \nu$  supported on  $(\mathcal{X}, d)$ , the  $p$ -Wasserstein distance ( $p \geq 1$ ) induced by the metric  $d$  between  $\mu$  and  $\nu$  is defined as

$$W_{p,d}(\mu, \nu) := \left( \inf_{\pi \in \Pi(\mu, \nu)} \iint_{\mathcal{X} \times \mathcal{X}} d(x, y)^p d\pi(x, y) \right)^{\frac{1}{p}}, \quad (8)$$

where  $\Pi(\mu, \nu)$  denotes the set of *transport plans* between  $\mu$  and  $\nu$ , that is the set of measures on  $\mathcal{X} \times \mathcal{X}$  which have respective marginals  $\mu$ , and  $\nu$ . 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,  $\mu$  and  $\nu$  are required to have a finite  $p$ -th moment, that is there exists  $x_0 \in \mathcal{X}$  such that  $\int_{\mathcal{X}} d(x, x_0)^p d\mu(x)$  (resp.  $d\nu$ ) is finite. The set of such probability measures, endowed with the metric  $W_p$ , is referred to as  $\mathcal{W}^p(\mathcal{X})$ .

Wasserstein distances and  $d_p$  metrics defined in Eq. (1) share the key idea of defining a distance by minimizing a cost over some matchings. However, the set of transport plans  $\Pi(\mu, \nu)$  between two measures is non-empty if and only if the two measures have the same mass, while persistence diagrams with different masses can be compared, making a crucial difference between the  $W_p$  and  $d_p$  metrics.

## 2.2 Extension to Radon measures supported on a bounded space

Extending optimal transport to measures of different masses, generally referred to as optimal *partial* transport, has been addressed by different authors [27, 20, 40]. As it handles the case of measures with infinite masses, the work of Figalli and Gigli [28], is of particular interest for us. The authors propose to extend Wasserstein distances to Radon measures supported on a *bounded* open proper subset  $\mathcal{X}$  of  $\mathbb{R}^d$ , whose boundary is denoted by  $\partial\mathcal{X}$  (and  $\bar{\mathcal{X}} := \mathcal{X} \sqcup \partial\mathcal{X}$ ).

**Definition 2.1.** [28, Problem 1.1] *Let  $p \in [1, +\infty)$ . Let  $\mu, \nu$  be two Radon measures supported on  $\mathcal{X}$  satisfying*

$$\int_{\mathcal{X}} d(x, \partial\mathcal{X})^p d\mu(x) < +\infty, \quad \int_{\mathcal{X}} d(x, \partial\mathcal{X})^p d\nu(x) < +\infty.$$

*The set of admissible transport plans (or couplings)  $\text{Adm}(\mu, \nu)$  is defined as the set of Radon measures  $\pi$  on  $\bar{\mathcal{X}} \times \bar{\mathcal{X}}$  satisfying for all Borel sets  $A, B \subset \mathcal{X}$ ,*

$$\pi(A \times \bar{\mathcal{X}}) = \mu(A) \quad \text{and} \quad \pi(\bar{\mathcal{X}} \times B) = \nu(B).$$

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

$$C_p(\pi) := \iint_{\bar{\mathcal{X}} \times \bar{\mathcal{X}}} d(x, y)^p d\pi(x, y). \quad (9)$$



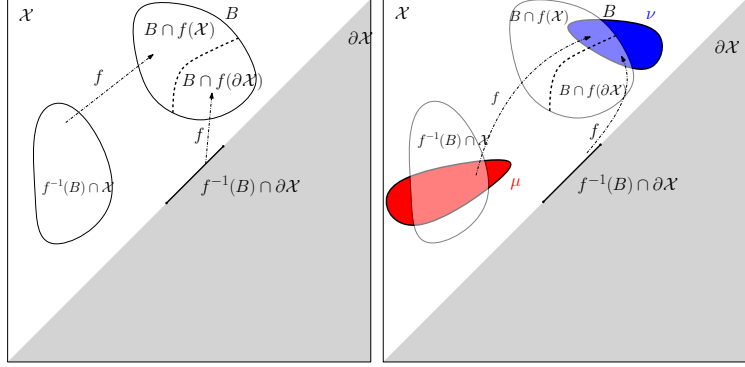


Figure 3: A transport map  $f$  must satisfy that the mass  $\nu(B)$  (light blue) is the sum of the mass  $\mu(f^{-1}(B) \cap \mathcal{X})$  given by  $\mu$  that is transported by  $f$  onto  $B$  (light red) and the mass  $\nu(B \cap f(\partial\mathcal{X}))$  coming from  $\partial\mathcal{X}$  and transported by  $f$  onto  $B$ .

The *Optimal Transport (with boundary) distance*  $\text{OT}_p(\mu, \nu)$  is then defined as

$$\text{OT}_p(\mu, \nu) := \left( \inf_{\pi \in \text{Adm}(\mu, \nu)} C_p(\pi) \right)^{1/p}. \quad (10)$$

Plans  $\pi \in \text{Adm}(\mu, \nu)$  realizing the infimum in (10) are called *optimal*. The set of optimal transport plans between  $\mu$  and  $\nu$  for the cost  $(x, y) \mapsto d(x, y)^p$  is denoted by  $\text{Opt}_p(\mu, \nu)$ .

We introduce the following definition, which shows how to build an element of  $\text{Adm}(\mu, \nu)$  given a map  $f : \overline{\mathcal{X}} \rightarrow \overline{\mathcal{X}}$  satisfying some balance condition (see Figure 3).

**Definition 2.2.** Let  $\mu, \nu \in \mathcal{M}$ . Consider  $f : \overline{\mathcal{X}} \rightarrow \overline{\mathcal{X}}$  a measurable function satisfying for all Borel set  $B \subset \mathcal{X}$

$$\mu(f^{-1}(B) \cap \mathcal{X}) + \nu(B \cap f(\partial\mathcal{X})) = \nu(B). \quad (11)$$

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

$$\pi(A \times B) = \mu(f^{-1}(B) \cap \mathcal{X} \cap A) + \nu(\mathcal{X} \cap B \cap f(A \cap \partial\mathcal{X})). \quad (12)$$

$\pi$  is called the *transport plan* induced by the transport map  $f$ .

One can easily check that we have indeed  $\pi(A \times \overline{\mathcal{X}}) = \mu(A)$  and  $\pi(\overline{\mathcal{X}} \times B) = \nu(B)$  for any Borel sets  $A, B \subset \mathcal{X}$ , so that  $\pi \in \text{Adm}(\mu, \nu)$  (see Figure 3).

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

$$E_{\mathcal{X}} := (\overline{\mathcal{X}} \times \overline{\mathcal{X}}) \setminus (\partial\mathcal{X} \times \partial\mathcal{X}). \quad (13)$$

### 3 Structure of the persistence measures and diagrams spaces

This section is dedicated to general properties of  $\mathcal{M}^p$ . Results in Section 3.1 are inspired from the ones of Figalli and Gigli in [28], which are stated for a bounded subset  $\mathcal{X}$  of  $\mathbb{R}^d$ . Our goal is to state properties over the space  $\Omega$ , which is of course *not bounded*. Adapting the results of [28] to our purpose is sometimes straightforward, in which case the proofs are delayed to Appendix B, and sometimes more involving, in which case the proofs are exposed in the main part of this article. Following Sections 3.2 and 3.3 are respectively dedicated to finite measures (involved in applications) and the case  $p = \infty$  (of major interest in topological data analysis).

**Remark 3.1.** *The results exposed in this section would remain true in a more general setting, namely for any locally compact Polish metric space  $\mathcal{X}$  that is partitioned into  $\mathcal{X} = A \sqcup B$ , where  $A$  is open and  $B$  is closed (here,  $A = \Omega$  and  $B = \partial\Omega$ ).*

#### 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. Consider the space  $\mathcal{M}^p$  defined in (4). First, we observe that the quantities introduced in Definition 2.1, in particular the metric  $\text{OT}_p$ , are still well-defined when  $\mathcal{X} = \Omega$  is not bounded.

**Proposition 3.1.** *Let  $\mu, \nu \in \mathcal{M}$ . The set of transport plans  $\text{Adm}(\mu, \nu)$  is sequentially compact for the vague topology on  $E_\Omega := \Omega \times \Omega \setminus \partial\Omega \times \partial\Omega$ . Moreover, if  $\mu, \nu \in \mathcal{M}^p$ , for this topology,*

- $\pi \in \text{Adm}(\mu, \nu) \mapsto C_p(\pi)$  is lower semi-continuous.
- $\text{Opt}_p(\mu, \nu)$  is a non-empty sequentially compact set.
- $\text{OT}_p$  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

$$\text{OT}_p(\mu, \nu) \leq \liminf_{n \rightarrow \infty} \text{OT}_p(\mu_n, \nu_n).$$

Moreover,  $\text{OT}_p$  is a metric on  $\mathcal{M}^p$ .

These properties are mentioned in [28, pages 4-5] in the bounded case, and corresponding proofs adapt straightforwardly to our framework. For the sake of completeness, we provide a detailed proof in Appendix B.

**Remark 3.2.** *If a (Borel) measure  $\mu$  satisfies  $\text{Pers}_p(\mu) < \infty$ , then for any Borel set  $A \subset \Omega$  satisfying  $d(A, \partial\Omega) := \inf_{x \in A} d(x, \partial\Omega) > 0$ , we have:*

$$\mu(A) d(A, \partial\Omega)^p \leq \int_A d(x, \partial\Omega)^p d\mu(x) \leq \int_\Omega d(x, \partial\Omega)^p d\mu(x) = \text{Pers}_p(\mu) < \infty, \quad (14)$$

so that  $\mu(A) < \infty$ . In particular,  $\mu$  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  $\mu$  to  $\Omega \setminus A_r$ . Then  $\text{OT}_p(\mu^{(r)}, \mu) \rightarrow 0$  when  $r \rightarrow 0$ . Similarly, if  $a \in \mathcal{D}^p$ , we have  $d_p(a^{(r)}, a) \rightarrow 0$ .*

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

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

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

The following proposition is central in our work: it shows that the metrics  $\text{OT}_p$  are extensions of the metrics  $d_p$ .

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

*Proof.* 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 [45, Proposition 1].

In the general case, let  $r > 0$ . Due to (14), the diagrams  $a^{(r)}$  and  $b^{(r)}$  defined in Lemma 3.1 have a finite mass (thus finite number of points). Therefore,  $d_p(a^{(r)}, b^{(r)}) = \text{OT}_p(a^{(r)}, b^{(r)})$ . By Lemma 3.1, the former converges to  $d_p(a, b)$  while the latter converges to  $\text{OT}_p(a, b)$ , giving the conclusion.  $\square$

As a consequence of this proposition, we will use  $\text{OT}_p$  to denote the distance between two elements of  $\mathcal{D}^p$  from now on.

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

As for Proposition 3.1, this proposition appears in [28, 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  $(\mathcal{M}^p, \text{OT}_p)$ .

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

$$\text{OT}_p(\mu_n, \mu) \rightarrow 0 \Leftrightarrow \begin{cases} \mu_n \xrightarrow{v} \mu, \\ \text{Pers}_p(\mu_n) \rightarrow \text{Pers}_p(\mu). \end{cases} \quad (15)$$

This result is analog to the characterization of convergence of probability measures in the Wasserstein space (see [64, Theorem 6.9]) and can be found in [28, Proposition 2.7] in the case where the ground space is bounded. While the proof of the direct implication can be easily adapted from [28] (it can be found in Appendix B), a new proof is needed for the converse implication.

*Proof of the converse implication.* For a given compact set  $K \subset \Omega$ , we denote its complementary set in  $\Omega$  by  $K^c$ , its interior set by  $\overset{\circ}{K}$ , and its boundary

by  $\partial K$ . Let  $\mu, \mu_1, \mu_2 \dots$  be elements of  $\mathcal{M}^p$  and assume that  $\mu_n \xrightarrow{v} \mu$  and  $\text{Pers}_p(\mu_n) \rightarrow \text{Pers}_p(\mu)$ . Since

$$\text{OT}_p(\mu_n, \mu) \leq \text{OT}_p(\mu_n, 0) + \text{OT}_p(\mu, 0) = \text{Pers}_p(\mu_n)^{1/p} + \text{Pers}_p(\mu)^{1/p},$$

the sequence  $(\text{OT}_p(\mu_n, \mu))_n$  is bounded. Thus, if we show that  $(\text{OT}_p(\mu_n, \mu))_n$  admits 0 as an unique accumulation point, then the convergence holds. Up to extracting a subsequence, we may assume that  $(\text{OT}_p(\mu_n, \mu))_n$  converges to some limit. Let  $(\pi_n)_n \in \text{Opt}(\mu_n, \mu)^{\mathbb{N}}$  be corresponding optimal transport plans. Let  $K$  be a compact subset of  $\Omega$ . Recall (Prop. A.1 in Appendix A) that relative compactness for the vague convergence of a sequence  $(\mu_n)_n$  is equivalent to  $\sup_n \{\mu_n(K)\} < \infty$  for every compact  $K \subset \Omega$ . Therefore, for any compact  $K \subset \Omega$ , and  $n \in \mathbb{N}$ ,

$$\pi_n((K \times \Omega) \cup (\Omega \times K)) \leq \mu_n(K) + \mu(K) \leq \sup_k \mu_k(K) + \mu(K) < \infty.$$

As any compact of  $E_\Omega = (\bar{\Omega} \times \bar{\Omega}) \setminus (\partial\Omega \times \partial\Omega)$  is included in some set of the form  $(K \times \Omega) \cup (\Omega \times K)$ , for  $K \subset \Omega$  any compact subset, using Proposition A.1 again, it follows that  $(\pi_n)_n$  is also relatively compact for the vague convergence.

Let thus  $\pi$  be the limit of any converging subsequence of  $(\pi_n)_n$ , whose indexes are still denoted by  $n$ . As  $\mu_n \xrightarrow{v} \mu$ ,  $\pi$  is necessarily in  $\text{Opt}_p(\mu, \mu)$  (see [28, Prop. 2.3]), i.e.  $\pi$  is supported on  $\{(x, x), x \in \Omega\}$ . The vague convergence of  $(\mu_n)_n$  and the convergence of  $(\text{Pers}_p(\mu_n))_n$  to  $\text{Pers}_p(\mu)$  imply that for a given compact set  $K \subset \Omega$ , we have

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \int_{K^c} d(x, \partial\Omega)^p d\mu_n(x) \\ &= \limsup_{n \rightarrow \infty} \left( \text{Pers}_p(\mu_n) - \int_K d(x, \partial\Omega)^p d\mu_n(x) \right) \\ &= \text{Pers}_p(\mu) - \liminf_n \int_{\tilde{K}} d(x, \partial\Omega)^p d\mu_n(x) - \liminf_n \int_{\partial K} d(x, \partial\Omega)^p d\mu_n(x) \\ &\leq \text{Pers}_p(\mu) - \int_{\tilde{K}} d(x, \partial\Omega)^p d\mu(x) \text{ by the Portmanteau theorem} \\ &= \int_{K^c} d(x, \partial\Omega)^p d\mu(x), \end{aligned}$$

where the Portmanteau theorem is recalled in Appendix A. As  $\text{Pers}_p(\mu)$  is finite, for  $\varepsilon > 0$ , there exists some compact set  $K \subset \Omega$  with

$$\limsup_n \int_{K^c} d(x, \partial\Omega)^p d\mu_n(x) < \varepsilon \quad \text{and} \quad \int_{K^c} d(x, \partial\Omega)^p d\mu(x) < \varepsilon. \quad (16)$$

Let  $s : \Omega \rightarrow \partial\Omega$  be the projection on  $\partial\Omega$  for the metric  $d$ . Such a projection is not unique for  $q = 1$  or for the more general Polish spaces  $\mathcal{X}$  of Remark 3.1, but we can always select a measurable projection  $s$  [15]. We 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):

$$\left\{ \begin{array}{ll} \tilde{\pi}_n = \pi_n & \text{on } K^2 \sqcup (K^c)^2, \\ \tilde{\pi}_n = 0 & \text{on } K \times K^c \sqcup K^c \times K, \\ \tilde{\pi}_n(A \times B) = \pi_n(A \times B) + \pi_n(A \times (s^{-1}(B) \cap K^c)) & \text{for } A \subset K, B \subset \partial\Omega, \\ \tilde{\pi}_n(A \times B) = \pi_n(A \times B) + \pi_n(A \times (s^{-1}(B) \cap K)) & \text{for } A \subset K^c, B \subset \partial\Omega, \\ \tilde{\pi}_n(A \times B) = \pi_n(A \times B) + \pi_n((s^{-1}(A) \cap K^c) \times B) & \text{for } A \subset \partial\Omega, B \subset K, \\ \tilde{\pi}_n(A \times B) = \pi_n(A \times B) + \pi_n((s^{-1}(A) \cap K) \times B) & \text{for } A \subset \partial\Omega, B \subset K^c. \end{array} \right. \quad (17)$$

Note that  $\tilde{\pi}_n \in \text{Adm}(\mu_n, \mu)$ : for instance, for  $A \subset K$  a Borel set,

$$\begin{aligned} \tilde{\pi}_n(A \times \bar{\Omega}) &= \tilde{\pi}_n(A \times K) + \tilde{\pi}_n(A \times K^c) + \tilde{\pi}_n(A \times \partial\Omega) \\ &= \pi_n(A \times K) + 0 + \pi_n(A \times \partial\Omega) + \pi_n(A \times (s^{-1}(\partial\Omega) \cap K^c)) \\ &= \pi_n(A \times \bar{\Omega}) = \mu_n(A), \end{aligned}$$

and it is shown likewise that the other constraints are satisfied. As  $\tilde{\pi}_n$  is suboptimal,  $\text{OT}_p^p(\mu_n, \mu) \leq \int_{\bar{\Omega}^2} d(x, y)^p d\tilde{\pi}_n(x, y)$ . The latter integral is equal to a sum of different terms, and we will show that each of them converges to 0. Assume without loss of generality that the compact set  $K$  belongs to an increasing sequence of compact sets whose union is  $\Omega$ , with  $\pi(\partial(K \times K)) = 0$  for all compacts of the sequence.

- We have  $\iint_{K^2} d(x, y)^p d\tilde{\pi}_n(x, y) = \iint_{K^2} d(x, y)^p d\pi_n(x, y)$ . The limsup of the integral is less than or equal to  $\iint_{K^2} d(x, y)^p d\pi(x, y)$  by the Portmanteau theorem (applied to the sequence  $(d(x, y)^p d\pi_n(x, y))_n$ ), and, recalling that  $\pi$  is supported on the diagonal of  $E_\Omega$ , this integral is equal to 0.
- For optimality reasons, any optimal transport plan must be supported on  $\{d(x, y)^p \leq d(x, \partial\Omega)^p + d(y, \partial\Omega)^p\}$  (this fact is detailed in [28, Prop. 2.3]). It follows that

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

Taking the limsup in  $n$ , and then letting  $K$  goes to  $\Omega$ , this quantity converges to 0 by (16).

- We have

$$\begin{aligned} &\iint_{K \times \partial\Omega} d(x, \partial\Omega)^p d\tilde{\pi}_n(x, y) \\ &= \iint_{K \times \partial\Omega} d(x, \partial\Omega)^p d\pi_n(x, y) + \iint_{K \times K^c} d(x, \partial\Omega)^p d\pi_n(x, y) \\ &= \iint_{K \times \bar{\Omega}} d(x, \partial\Omega)^p d\pi_n(x, y) - \iint_{K^2} d(x, \partial\Omega)^p d\pi_n(x, y) \\ &= \int_K d(x, \partial\Omega)^p d\mu_n(x) - \iint_{K^2} d(x, \partial\Omega)^p d\pi_n(x, y) \end{aligned}$$

By the Portmanteau theorem applied to the sequence  $(d(x, \partial\Omega)^p d\mu_n(x))_n$ , the limsup of the first term is less than or equal to  $\int_K d(x, \partial\Omega)^p d\mu(x)$ . Recall that we assume that  $\pi(\partial(K \times K)) = 0$ . By applying the second characterization of Portmanteau theorem (see Prop. A.4) on the second term to the sequence  $(d(x, y)^p d\pi_n(x, y))_n$ , and using that  $\pi$  is supported on the diagonal of  $E_\Omega$ , we obtain that the limsup of the second term is less than or equal to  $-\iint_{K^2} d(x, \partial\Omega)^p d\pi(x, y) = -\int_K d(x, \partial\Omega)^p d\mu(x)$ . Therefore, the limsup of the integral is equal to 0.

- The three remaining terms (corresponding to the three last lines of the definition (17)) are treated likewise this last case.

Finally, we have proven that  $(\text{OT}_p(\mu_n, \mu))_n$  is bounded and that for any converging subsequence  $(\mu_{n_k})_k$ ,  $\text{OT}_p(\mu_{n_k}, \mu)$  converges to 0. It follows that  $\text{OT}_p(\mu_n, \mu) \rightarrow 0$ .  $\square$

**Remark 3.3.** *The assumption  $\text{Pers}_p(\mu_n) \rightarrow \text{Pers}_p(\mu)$  is crucial to obtain  $\text{OT}_p$ -convergence assuming vague convergence. For example, the sequence defined by  $\mu_n := \delta_{(n, n+1)}$  converges vaguely to  $\mu = 0$  and  $(\text{Pers}_p(\mu_n))_n$  does converge (it is constant), while  $\text{OT}_p(\mu_n, 0) \not\rightarrow 0$ . This does not contradict Theorem 3.4 since  $\text{Pers}_p(\mu) = 0 \neq \lim_n \text{Pers}_p(\mu_n)$ .*

Theorem 3.4 implies some useful results. First, it entails that the topology of the metric  $\text{OT}_p$  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 topology).

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

We recover in particular that the space  $(\mathcal{D}^p, \text{OT}_p)$  is a Polish space (Proposition 3.3), a result already proved in [48, Theorems 7 and 12] with a different approach.

Secondly, we show that the vague convergence of  $\mu_n$  to  $\mu$  along with the convergence of  $\text{Pers}_p(\mu_n) \rightarrow \text{Pers}_p(\mu)$  is equivalent to the weak convergence of a weighted measure (see Appendix A for a definition of weak convergence, denoted by  $\xrightarrow{w}$  in the following). For  $\mu \in \mathcal{M}^p$ , let us introduce the Borel measure with finite mass  $\mu^{(p)}$  defined, for a Borel subset  $A \subset \Omega$ , as:

$$\mu^{(p)}(A) = \int_A d(x, \partial\Omega)^p d\mu(x). \quad (18)$$

**Corollary 3.2.** *For a sequence  $(\mu_n)_n$  and a persistence measure  $\mu \in \mathcal{M}^p$ , we have*

$$\text{OT}_p(\mu_n, \mu) \rightarrow 0 \text{ if and only if } \mu_n^{(p)} \xrightarrow{w} \mu^{(p)}.$$

*Proof.* Consider  $\mu, \mu_1, \mu_2, \dots \in \mathcal{M}^p$  and assume that  $\text{OT}_p(\mu_n, \mu) \rightarrow 0$ . By Theorem 3.4, this is equivalent to  $\mu_n \xrightarrow{v} \mu$  and  $\mu_n^{(p)}(\Omega) = \text{Pers}_p(\mu_n) \rightarrow \text{Pers}_p(\mu) = \mu^{(p)}(\Omega)$ . 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  $\mu_n^{(p)} \xrightarrow{v} \mu^{(p)}$ . Likewise, the map  $x \mapsto d(x, \partial\Omega)^{-p} f(x)$  is continuous and compactly supported, so that  $\mu_n^{(p)} \xrightarrow{v} \mu^{(p)}$  also implies  $\mu_n \xrightarrow{v} \mu$ . Hence,  $\mu_n \xrightarrow{v} \mu$  is equivalent to  $\mu_n^{(p)} \xrightarrow{v} \mu^{(p)}$ . By Proposition A.3, the vague convergence  $\mu_n^{(p)} \xrightarrow{v} \mu^{(p)}$  along with the convergence of the masses is equivalent to  $\mu_n^{(p)} \xrightarrow{w} \mu^{(p)}$ .  $\square$

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

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

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

**Remark 3.4.** *This characterization is equivalent to the one described in [48, Theorem 21] for persistence diagrams. The notions introduced by the authors of off-diagonally birth-death boundedness, and uniformness are rephrased using the notion of tightness, standard in measure theory.*

We end this section with a remark on the existence of transport maps, assuming that one of the two measures has a density with respect to the Lebesgue measure on  $\Omega$ . We denote by  $f_{\#}\mu$  the pushforward of a measure  $\mu$  by a map  $f$ , defined by  $f_{\#}\mu(A) = \mu(f^{-1}(A))$  for  $A$  a Borel set.

**Remark 3.5.** *Following [28, Corollary 2.5], one can prove that if  $\mu \in \mathcal{M}^2$  has a density with respect to the Lebesgue measure on  $\Omega$ , then for any measure  $\nu \in \mathcal{M}^2$ , there exists a unique optimal transport plan  $\pi$  between  $\mu$  and  $\nu$  for the  $\text{OT}_2$  metric. The restriction of this transport plan to  $\Omega \times \bar{\Omega}$  is equal to  $(\text{id}, T)_{\#}\mu$  where  $T : \Omega \rightarrow \bar{\Omega}$  is the gradient of some convex function, whereas the transport plan restricted to  $\partial\Omega \times \Omega$  is given by  $(s, \text{id})_{\#}(\nu - T_{\#}\mu)$ , where  $s : \Omega \rightarrow \partial\Omega$  is the projection on the diagonal. A proof of this fact in the context of persistence measures would require to introduce various notions that are out of the scope covered by this paper. We refer the interested reader to [28, Prop. 2.3] and [3, Theorem 6.2.4] for details.*

## 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 [44, 13], and the specific properties of  $\text{OT}_p$  in this setting are therefore of interest. Introduce for  $m \geq 0$  the subset  $\mathcal{M}_{\leq m}^p$  of  $\mathcal{M}^p$  defined as  $\mathcal{M}_{\leq m}^p := \{\mu \in \mathcal{M}^p, \mu(\Omega) \leq m\}$ , and the set  $\mathcal{M}_f^p$  of finite persistence measures,  $\mathcal{M}_f^p := \bigcup_{m \geq 0} \mathcal{M}_{\leq m}^p$ . Define similarly the set  $\mathcal{D}_{\leq m}$  (resp.  $\mathcal{D}_f$ ). Note that the assumption  $\text{Pers}_p(a) < \infty$  is always satisfied for a finite diagram  $a$  (which is not true for general Radon measures), so that the exponent  $p$  is not needed when defining  $\mathcal{D}_{\leq m}$  and  $\mathcal{D}_f$ .

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

*Proof.* This is a straightforward consequence of Lemma 3.1. Indeed, if  $\mu \in \mathcal{M}^p$  and  $r > 0$ , then (14) implies that  $\mu^{(r)}$  is of finite mass.  $\square$

Let  $\tilde{\Omega} = \Omega \sqcup \{\partial\Omega\}$  be the quotient of  $\bar{\Omega}$  by the closed subset  $\partial\Omega$ —i.e. we encode the diagonal by just one point (still denoted by  $\partial\Omega$ ). The distance  $d$  on  $\bar{\Omega}^2$  induces naturally a function  $\tilde{d}$  on  $\tilde{\Omega}^2$ , defined for  $x, y \in \Omega$  by  $\tilde{d}(x, y) = d(x, y)$ ,

$\tilde{d}(x, \partial\Omega) = \tilde{d}(\partial\Omega, x) = d(x, s(x))$  and  $\tilde{d}(\partial\Omega, \partial\Omega) = 0$ . However,  $\tilde{d}$  is not a distance since one can have  $\tilde{d}(x, y) > \tilde{d}(x, \partial\Omega) + \tilde{d}(y, \partial\Omega)$ . Define

$$\rho(x, y) := \min\{\tilde{d}(x, y), \tilde{d}(x, \partial\Omega) + \tilde{d}(y, \partial\Omega)\}. \quad (19)$$

It is straightforward to check that  $\rho$  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  $\rho$  for finite measures on  $\tilde{\Omega}$  which have the same masses, that is the infimum of  $\tilde{C}_p(\tilde{\pi}) := \iint_{\tilde{\Omega}^2} \rho(x, y)^p d\tilde{\pi}(x, y)$ , for  $\tilde{\pi}$  a transport plan with corresponding marginals (see Section 2.1). The following theorem states that the problem of computing the  $\text{OT}_p$  metric between two persistence measures with finite masses can be turned into the one of computing the Wasserstein distances between two measures supported on  $\tilde{\Omega}$  with the same mass. For the sake of simplicity, we assume in the following that  $d(x, y) = \|x - y\|_q$  with  $q > 1$ , ensuring that the quantity  $\arg \min_{y \in \partial\Omega} d(x, y)$  is reduced to the orthogonal projection  $s(x)$  of  $x$  onto the diagonal  $\partial\Omega$ . The following result could be seamlessly adapted to the case  $q = 1$ .

**Proposition 3.7.** *Let  $\mu, \nu \in \mathcal{M}_f^p$  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  $\text{OT}_p(\mu, \nu) = W_{p, \rho}(\tilde{\mu}, \tilde{\nu})$ .*

Before proving Proposition 3.7, we need the two following lemmas:

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

1. *Define  $T(\mu, \nu)$  the set of plans  $\pi \in \text{Adm}(\mu, \nu)$  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,  $\text{Opt}_p(\mu, \nu) \subset T(\mu, \nu)$ .*
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  $A, B \subset \Omega$ ,*

$$\begin{cases} \iota(\pi)(A \times B) = \pi(A \times B), \\ \iota(\pi)(A \times \{\partial\Omega\}) = \pi(A \times \partial\Omega), \\ \iota(\pi)(\{\partial\Omega\} \times B) = \pi(\partial\Omega \times B), \\ \iota(\pi)(\{\partial\Omega\} \times \{\partial\Omega\}) = r - \mu(\Omega) - \pi(\partial\Omega \times \Omega) \geq 0. \end{cases} \quad (20)$$

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

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

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

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

*Proof.*



1. Consider  $\pi \in \text{Adm}(\mu, \nu)$ , and define  $\pi'$  that coincides with  $\pi$  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  $C_p(\pi') \leq C_p(\pi)$ , with equality if and only if  $\pi \in T(\mu, \nu)$ , and thus  $\text{Opt}_p(\mu, \nu) \subset T(\mu, \nu)$ .

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  $A \subset \Omega$ ,

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

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  $\iota$  is well defined. Observe that

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

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

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

Similarly,  $\pi(\bar{\Omega} \times B) = \nu(B)$  for all  $B \subset \Omega$ . Therefore,  $\pi \in \text{Adm}(\mu, \nu)$ , and by construction, if a point  $x \in \Omega$  is transported on  $\partial\Omega$ , it is transported on  $s(x)$ , so that  $\pi \in T(\mu, \nu)$ . Observe that  $\mu(\Omega) + \pi(\partial\Omega \times \Omega) \leq \tilde{\pi}(\tilde{\Omega} \times \tilde{\Omega}) = r$ , so that  $\iota(\pi)$  is well defined. Also,  $\iota(\pi) = \tilde{\pi}$ , so that, according to point 2,  $C_p(\pi) = \iint_{\tilde{\Omega} \times \tilde{\Omega}} \tilde{d}(x, y)^p d\tilde{\pi}(x, y)$ .

□

We show that the inequality  $\text{OT}_p(\mu, \nu) \leq W_{p,\rho}(\tilde{\mu}, \tilde{\nu})$  holds as long as  $r \geq \max(\mu(\Omega), \nu(\Omega))$ .

**Lemma 3.3.** *Let  $\mu, \nu \in \mathcal{M}_f^p$  and  $r \geq \max(\mu(\Omega), \nu(\Omega))$ . Let  $\tilde{\mu} := \mu + (r - \mu(\Omega))\delta_{\partial\Omega}$ ,  $\tilde{\nu} := \nu + (r - \nu(\Omega))\delta_{\partial\Omega}$ . Then,  $\text{OT}_p(\mu, \nu) \leq W_{p,\rho}(\tilde{\mu}, \tilde{\nu})$ .*

*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  $(a + b)^p \geq a^p + b^p$  for positive  $a, b$ , we have

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

We conclude by taking the infimum on  $\tilde{\pi}$  that

$$W_{p,\rho}(\tilde{\mu}, \tilde{\nu}) \geq \inf_{\tilde{\pi}' \in \Pi(\tilde{\mu}, \tilde{\nu})} \iint_{\tilde{\Omega} \times \tilde{\Omega}} \tilde{d}(x, y)^p d\tilde{\pi}'(x, y).$$

Since  $\rho(x, y) \leq \tilde{d}(x, y)$ , it follows that

$$W_{p,\rho}^p(\tilde{\mu}, \tilde{\nu}) = \inf_{\tilde{\pi} \in \Pi(\tilde{\mu}, \tilde{\nu})} \iint_{\tilde{\Omega}^2} \tilde{d}(x, y)^p d\tilde{\pi}(x, y). \quad (21)$$

Since  $\tilde{d}$  is continuous, the infimum in the right hand side of (21) is reached [64, Theorem 4.1]. Consider thus  $\tilde{\pi} \in \Pi(\tilde{\mu}, \tilde{\nu})$  which realizes the infimum. We can write, using Lemma 3.2,

$$\begin{aligned} W_{p,\rho}^p(\tilde{\mu}, \tilde{\nu}) &= \iint_{\tilde{\Omega}^2} \tilde{d}(x, y)^p d\tilde{\pi}(x, y) = \iint_{\tilde{\Omega} \times \tilde{\Omega}} d(x, y)^p d\kappa(\tilde{\pi})(x, y) \\ &\geq \inf_{\pi \in T(\mu, \nu)} \iint_{\tilde{\Omega} \times \tilde{\Omega}} d(x, y)^p d\pi(x, y) = \text{OT}_p^p(\mu, \nu), \end{aligned}$$

which concludes the proof.  $\square$

*Proof of Proposition 3.7.* 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  $\rho(x, y) \leq \tilde{d}(x, y)$ , we have  $\tilde{C}_p(\tilde{\pi}) \leq \iint \tilde{d}(x, y)^p d\tilde{\pi}(x, y) = C_p(\pi)$  (Lemma 3.2). Taking infimum gives  $W_{p,\rho}(\tilde{\mu}, \tilde{\nu}) \leq \text{OT}_p(\mu, \nu)$ . The other inequality holds according to Lemma 3.3.  $\square$

**Remark 3.6.** *The starting idea of this theorem—informally, “adding the mass of one diagram to the other and vice-versa”—is known in TDA as a bipartite*

graph matching [26, Ch. VIII.4] and used in practical computations [39]. Here, Proposition 3.7 states that solving this bipartite graph matching problem can be formalized as computing a Wasserstein distance on the metric space  $(\tilde{\Omega}, \rho)$  and as such, makes sense (and remains true) for more general measures.

**Remark 3.7.** Proposition 3.7 is useful for numerical purposes since it allows us in applications, when dealing with a finite set of finite measures (in particular diagrams), to directly use the various tools developed in computational optimal transport [52] to compute Wasserstein distances. This alternative to the combinatorial algorithms considered in the literature [39, 60] is studied in details in [45]. This result is also helpful to prove the existence of  $p$ -Fréchet means of sets of persistence measures (see Section 4).

### 3.3 The $\text{OT}_\infty$ distance

In classical optimal transport, the  $\infty$ -Wasserstein distance is known to have a much more erratic behavior than its  $p < \infty$  counterparts [54, Section 5.5.1]. However, in the context of persistence diagrams, the  $d_\infty$  distance defined in Eq. (2) appears naturally as an interleaving distance between persistence modules and satisfies strong stability results: it is thus worthy 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  $\text{spt}(\mu)$  denote the support of a measure  $\mu$  and define  $\text{Pers}_\infty(\mu) := \sup\{d(x, \partial\Omega), x \in \text{spt}(\mu)\}$ . Let

$$\mathcal{M}^\infty := \{\mu \in \mathcal{M}, \text{Pers}_\infty(\mu) < \infty\} \quad \text{and} \quad \mathcal{D}^\infty := \mathcal{D} \cap \mathcal{M}^\infty. \quad (22)$$

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

$$\text{OT}_\infty(\mu, \nu) := \inf_{\pi \in \text{Adm}(\mu, \nu)} C_\infty(\pi). \quad (23)$$

The set of transport plans minimizing (23) is denoted by  $\text{Opt}_\infty(\mu, \nu)$ .

Recall that  $E_\Omega = (\bar{\Omega} \times \bar{\Omega}) \setminus (\partial\Omega \times \partial\Omega)$ .

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

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

Moreover,  $\text{OT}_\infty$  is a metric on  $\mathcal{M}^\infty$ .

The proofs of these results are found in Appendix B.

As in the case  $p < \infty$ ,  $\text{OT}_\infty$  and  $d_\infty$  coincide on  $\mathcal{D}^\infty$ .

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

*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 (possibly infinite) sets of indices. The marginal constraints imply that a plan  $\pi \in \text{Adm}(\mu, \nu)$  is supported on  $(\{x_i\}_i \cup \partial\Omega) \times (\{y_j\}_j \cup \partial\Omega)$ . If some of the mass  $\pi(\{x_i\}, \partial\Omega)$  (resp.  $\pi(\partial\Omega, \{y_j\})$ ) is sent on a point other than the projection of  $x_i$  (resp.  $y_j$ ) on the diagonal  $\partial\Omega$ , then the cost of such a plan can always be (strictly if  $q > 1$ ) reduced. Introduce the matrix  $C$  indexed on  $(-J \cup I) \times (-I \cup J)$  defined by

$$\begin{cases} C_{i,j} = d(x_i, y_j) & \text{for } i, j > 0, \\ C_{i,j} = d(\partial\Omega, y_j) & \text{for } i < 0, j > 0, \\ C_{i,j} = d(x_i, \partial\Omega) & \text{for } i > 0, j < 0, \\ C_{i,j} = 0 & \text{for } i, j < 0. \end{cases} \quad (24)$$

In this context, an element of  $\text{Opt}(a, b)$  can be written a matrix  $P$  indexed on  $(-J \cup I) \times (-I \cup J)$ , and marginal constraints state that  $P$  must belong to the set of doubly stochastic matrices  $\mathcal{S}$ . Therefore,  $\text{OT}_\infty(a, b) = \inf_{P \in \mathcal{S}} \sup\{C_{i,j}, (i, j) \in \text{spt}(P)\}$ , where  $\mathcal{S}$  is the set of doubly stochastic matrices indexed on  $(-J \cup I) \times (-I \cup J)$ , and  $\text{spt}(P)$  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  $\{i_1, \dots, i_k\} \subset -J \cup I$ , we have

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

Thus, the cardinality of  $\{j, \exists k' \text{ such that } (i_{k'}, j) \in \text{spt}(P)\}$  must be larger than  $k$ . Said differently, the marginal constraints impose that any set of  $k$  points in  $a$  must be matched to *at least*  $k$  points in  $b$  (points are counted with eventual repetitions here). Under such conditions, the Hall's marriage theorem (see [33, p. 51]) guarantees the existence of a permutation matrix  $P'$  with  $\text{spt}(P') \subset \text{spt}(P)$ . As a consequence,

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

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  $\text{OT}_\infty(a, b) = d_\infty(a, b)$ .  $\square$

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

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

$$\mu_n(K) = \pi_n(K \times \bar{\Omega}) = \pi_n(K \times K_\varepsilon) \leq \mu_{n_0}(K_\varepsilon). \quad (25)$$

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 as in the proof of the completeness for  $p < \infty$  (see Appendix B) show that  $(\mu_n)_n$  converges for the  $\text{OT}_\infty$  metric.  $\square$

**Remark 3.8.** *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  $I, I'$ ,  $\text{OT}_\infty(a_I, a_{I'}) = \frac{\sqrt{2}}{2}$ . This result is similar to [10, 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  $\text{OT}_\infty(\mu_n, \mu) \rightarrow 0$ , then  $(\mu_n)_n$  converges vaguely to  $\mu$  and  $\text{Pers}_\infty(\mu_n)$  converges to  $\text{Pers}_\infty(\mu)$ .*

*Proof.* First, the convergence of  $\text{Pers}_\infty(\mu_n)$  towards  $\text{Pers}_\infty(\mu)$  is a consequence of the reverse triangle inequality:

$$|\text{Pers}_\infty(\mu_n) - \text{Pers}_\infty(\mu)| = |\text{OT}_\infty(\mu_n, 0) - \text{OT}_\infty(\mu, 0)| \leq \text{OT}_\infty(\mu_n, \mu),$$

which converges to 0 as  $n$  goes to  $\infty$ .

We now prove the vague convergence. 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  $f_\varepsilon$ , 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 (25). Let  $\pi_n \in \text{Opt}_\infty(\mu_n, \mu)$ . 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,

$$\begin{aligned} |\mu_n(f_\varepsilon) - \mu(f_\varepsilon)| &= \left| \iint_{\overline{\Omega}^2} (f_\varepsilon(x) - f_\varepsilon(y)) d\pi_n(x, y) \right| \\ &\leq \iint_{\overline{\Omega}^2} |f_\varepsilon(x) - f_\varepsilon(y)| d\pi_n(x, y) \\ &\leq L \iint_{(K \times \overline{\Omega}) \cup (\overline{\Omega} \times K)} d(x, y) d\pi_n(x, y) \text{ as } f_\varepsilon \text{ is } L\text{-Lipschitz continuous} \\ &\leq LC_\infty(\pi_n)(\pi_n(K \times \overline{\Omega}) + \pi_n(\overline{\Omega} \times K)) \\ &\leq L\text{OT}_\infty(\mu_n, \mu) \left( \sup_k \mu_k(K) + \mu(K) \right). \end{aligned}$$

This last quantity converge to 0 as  $n$  goes to  $\infty$  for fixed  $\varepsilon$ . Therefore, taking the limsup in  $n$  and then letting  $\varepsilon$  go to 0, we obtain that  $\mu_n(f) \rightarrow \mu(f)$ .  $\square$

**Remark 3.9.** *As for the case  $1 \leq p < \infty$ , Proposition 3.11 implies that  $\text{OT}_\infty$  metrizes the vague convergence, and thus using Propositions 3.9 and A.5, we have that  $(\mathcal{D}^\infty, d_\infty)$  is closed in  $(\mathcal{M}^\infty, \text{OT}_\infty)$  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 4). To recover a space with a structure more similar to  $\mathcal{D}^p$ , it is useful to look at a smaller set. Introduce

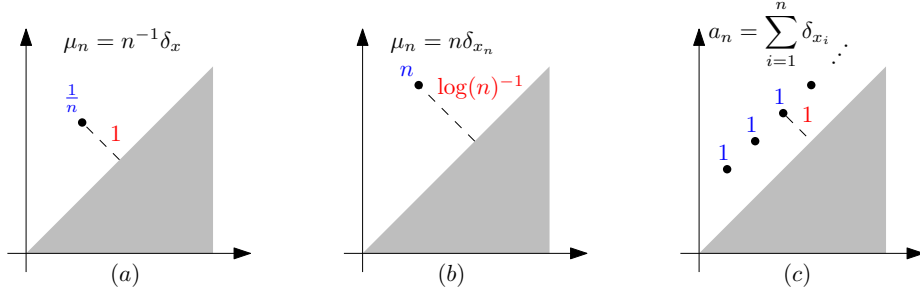


Figure 4: Illustration of differences between  $\text{OT}_p$ ,  $\text{OT}_\infty$ , and vague convergences. Blue color represents the mass on a point while red color designates distances. (a) A case where  $\text{OT}_p(\mu_n, 0) \rightarrow 0$  for any  $p < \infty$  while  $\text{OT}_\infty(\mu_n, 0) = 1$ . (b) A case where  $\text{OT}_\infty(\mu_n, 0) \rightarrow 0$  while for all  $p < \infty$ ,  $\text{OT}_p(\mu_n, \mu) \rightarrow \infty$ . (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  $\text{Pers}_\infty(a_n) = \text{Pers}_\infty(a)$ , but  $(a_n)$  does not converge to  $a$  for  $\text{OT}_\infty$ .

$\mathcal{D}_0^\infty$  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  $\text{OT}_\infty$  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  $\text{OT}_\infty(a, a_n) \leq \frac{1}{n} \rightarrow 0$ ,  $\mathcal{D}_0^\infty$  is contained in the closure of  $\mathcal{D}_f$ .

Conversely, consider a diagram  $a \in \mathcal{D}^\infty \setminus \mathcal{D}_0^\infty$ . 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  $\text{OT}_\infty(a', a) \geq r$ , so that  $a$  is not the limit for the  $\text{OT}_\infty$  metric of any sequence in  $\mathcal{D}_f$ .  $\square$

**Remark 3.10.** *The space  $\mathcal{D}_0^\infty$  is exactly the set introduced in [5, Theorem 3.5] as the completion of  $\mathcal{D}_f$  for the bottleneck metric  $d_\infty$ . 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  $\{x \in \Omega, d(x, \partial\Omega) > r\}$  (as in Lemma 3.1). The following characterization of convergence holds in  $\mathcal{D}_0^\infty$ .

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

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

*Proof.* Let us prove first the direct implication. Proposition 3.11 states that the convergence with respect to  $\text{OT}_\infty$  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.  $\bar{B}(x, \varepsilon)$ ) the open (resp. closed) ball of radius  $\varepsilon$  centered at  $x$ . By the Portmanteau theorem, for  $\varepsilon$  small enough,

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

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  $K' := K \setminus \bigcup_{i=1}^K B(x_i, \varepsilon)$  gives

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

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

$$\limsup_{n \rightarrow \infty} \text{OT}_\infty(a_n, a) \leq \limsup_{n \rightarrow \infty} \text{OT}_\infty(a_n, a^{(r)}) + r \leq 2r + \varepsilon.$$

Letting  $\varepsilon \rightarrow 0$  then  $r \rightarrow 0$ , the bottleneck convergence holds.  $\square$

## 4 $p$ -Fréchet means for distributions supported on $\mathcal{M}^p$

In this section, we state the existence of  $p$ -Fréchet means for probability distributions supported on  $\mathcal{M}^p$ . We start with the finite case (i.e. averaging finitely many persistence measures) and then extend the result to any probability distribution with finite  $p$ -th moment. We then study the specific case of distribution supported on  $\mathcal{D}^p$  (i.e. averaging persistence diagrams), and show that in the finite setting, the set of  $p$ -Fréchet means is a convex set whose extreme points are in  $\mathcal{D}^d$  (i.e. are actual persistence diagrams).

**Remark 4.1.** *In this section, we will assume that  $1 < p < \infty$  and  $1 < q < \infty$  (recall  $d(\cdot, \cdot) = \|\cdot - \cdot\|_q$ ). These assumptions will ensure that (i) the projection of  $x \in \Omega$  onto  $\partial\Omega$  is uniquely defined and (ii) the  $p$ -Fréchet mean of  $k$  points  $x_1 \dots x_k$  in  $\bar{\Omega}$ , i.e. minimizer of  $x \mapsto \sum_{i=1}^k \|x - x_i\|_q^p$ , is also uniquely defined; two facts used in our proofs.*

Recall that  $(\mathcal{M}^p, \text{OT}_p)$  is a Polish space, and let  $W_{p, \text{OT}_p}$  denote the Wasserstein distance (see Section 2.1) between probability measures supported on  $(\mathcal{M}^p, \text{OT}_p)$ . We denote by  $\mathcal{W}^p(\mathcal{M}^p)$  the space of probability measures  $P$  supported on  $\mathcal{M}^p$ , equipped with the  $W_{p, \text{OT}_p}$  metric, which are at a finite distance

from  $\delta_0$ —the Dirac mass supported on the empty diagram—i.e.

$$W_{p, \text{OT}_p}^p(P, \delta_0) = \int_{\nu \in \mathcal{M}^p} \text{OT}_p^p(\nu, 0) dP(\nu) = \int_{\nu \in \mathcal{M}^p} \text{Pers}_p(\nu) dP(\nu) < \infty.$$

**Definition 4.1.** Consider  $P \in \mathcal{W}^p(\mathcal{M}^p)$ . A measure  $\mu^* \in \mathcal{M}^p$  is a  $p$ -Fréchet mean of  $P$  if it minimizes  $\mathcal{E} : \mu \in \mathcal{M}^p \mapsto \int_{\nu \in \mathcal{M}^p} \text{OT}_p^p(\mu, \nu) dP(\nu)$ .

#### 4.1 $p$ -Fréchet means in the finite case

Let  $P$  be of the form  $\sum_{i=1}^N \lambda_i \delta_{\mu_i}$  with  $N \in \mathbb{N}$ ,  $\mu_i$  a persistence measure of finite mass  $m_i$ , and  $(\lambda_i)_i$  non-negative weights that sum to 1. Define  $m_{\text{tot}} := \sum_{i=1}^N m_i$ . To prove the existence of  $p$ -Fréchet means for such a  $P$ , we show that, in this case,  $p$ -Fréchet means correspond to  $p$ -Fréchet means for the Wasserstein distance of some distribution on  $\mathcal{M}_{m_{\text{tot}}}^p(\tilde{\Omega})$ , the sets of measures on  $\tilde{\Omega}$  that all have the same mass  $m_{\text{tot}}$  (see Section 3.2), a problem well studied in the literature [2, 11, 12].

We start with a lemma which affirms that if a measure  $\mu$  has too much mass (larger than  $m_{\text{tot}}$ ), then it cannot be a  $p$ -Fréchet mean of  $\mu_1 \dots \mu_N$ .

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

*Proof.* The idea of the proof is to show that if a measure  $\mu$  has some mass that is mapped to the diagonal in each transport plan between  $\mu$  and  $\mu_i$ , then we can build a measure  $\mu'$  by “removing” this mass, and then observe that such a measure  $\mu'$  has a smaller energy.

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

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

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

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

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

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



Also,

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

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

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

□

Recall that  $W_{p,\rho}$  denotes the Wasserstein distance between measures with same mass supported on the metric space  $(\tilde{\Omega}, \rho)$  (see Sections 2.1 and 3.2).

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

$$\begin{aligned}
\mathcal{E} : \mu \in \mathcal{M}_{\leq m_{\text{tot}}}^p &\mapsto \sum_{i=1}^N \lambda_i \text{OT}_p^p(\mu, \mu_i) \text{ and} \\
\mathcal{F} : \tilde{\mu} \in \mathcal{M}_{m_{\text{tot}}}^p(\tilde{\Omega}) &\mapsto \sum_{i=1}^N \lambda_i W_{p,\rho}^p(\tilde{\mu}, \Psi(\mu_i)),
\end{aligned}$$

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

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

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

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  $\pi_i \in \text{Opt}_p(\mu_i, \mu)$ . 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

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

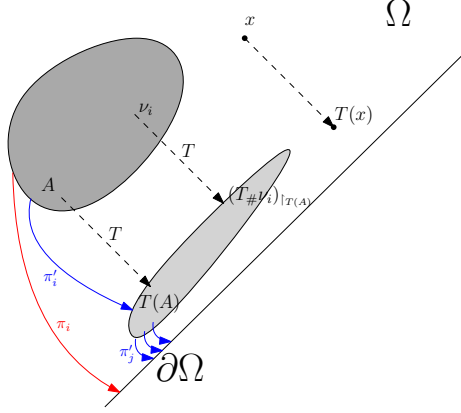


Figure 5: Global picture of the proof. The main idea is to observe that the cost induced by  $\pi_i$  (red) is strictly greater than the sum of costs induced by the  $\pi'_i$ 's (blue), which leads to a strictly better energy.

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 push-forward of  $\nu$  by the application  $T$ . Consider the transport plan  $\pi'_i$  deduced from  $\pi_i$  where  $\nu$  is transported onto  $T_{\#}\nu$  instead of being transported to  $\partial\Omega$  (see Figure 5). More precisely,  $\pi'_i$  is the measure on  $\Omega \times \Omega$  defined by, for Borel sets  $A, B \subset \Omega$ :

$$\begin{aligned}\pi'_i(A \times B) &= \pi_i(A \times B) + \nu(A \cap T^{-1}(B)), \\ \pi'_i(A \times \partial\Omega) &= 0, \quad \pi'_i(\partial\Omega \times B) = \pi_i(\partial\Omega \times B).\end{aligned}$$

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

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

and

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

Using  $\pi'_i$  instead of  $\pi_i$  changes the transport cost by the quantity

$$\int_{\Omega} [d(x, T(x))^p - d(x, \partial\Omega)^p] d\nu(x) < 0.$$

In a similar way, we define for  $j \neq i$  the plan  $\pi'_j \in \text{Adm}(\mu_j, \mu')$  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

$$\sum_{j \neq i} \lambda_j \int_{\Omega} d(T(x), \partial\Omega)^p d\nu(x).$$

One can observe that

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

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 less than  $\mathcal{E}(\mu)$ , and thus  $\mathcal{E}(\mu') < \mathcal{E}(\mu)$ . Finally, we have

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

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  $\Psi$  is a bijection. Also, if  $\mu$  is a minimizer of  $\mathcal{E}$  (should it exist), 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  $\mathcal{E}$ .  $\square$

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

## 4.2 Existence and consistency of $p$ -Fréchet means

We now extend the results of the previous section to the  $p$ -Fréchet means of general probability measures supported on  $\mathcal{M}^p$ . First, we show a *consistency* result, in the vein of [46, Theorem 3].

**Proposition 4.2.** *Let  $P_n, P$  be probability measures in  $\mathcal{W}^p(\mathcal{M}^p)$ . Assume that each  $P_n$  has a  $p$ -Fréchet mean  $\mu_n$  and that  $W_{p, \text{OT}_p}(P_n, P) \rightarrow 0$ . Then, the sequence  $(\mu_n)_n$  is relatively compact in  $(\mathcal{M}^p, \text{OT}_p)$ , and any limit of a converging subsequence is a  $p$ -Fréchet mean of  $P$ .*

*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 (14),

$$\begin{aligned} \mu_n(K)^{\frac{1}{p}} &\leq \frac{1}{d(K, \partial\Omega)} \text{OT}_p(\mu_n, 0) = \frac{1}{d(K, \partial\Omega)} W_{p, \text{OT}_p}(\delta_{\mu_n}, \delta_0) \\ &\leq \frac{1}{d(K, \partial\Omega)} (W_{p, \text{OT}_p}(\delta_{\mu_n}, P_n) + W_{p, \text{OT}_p}(P_n, \delta_0)) \end{aligned}$$

Since  $\mu_n$  is a  $p$ -Fréchet mean of  $P_n$ , it minimizes  $\{W_{p, \text{OT}_p}(\delta_{\nu}, P_n), \nu \in \mathcal{M}^p\}$ , and in particular  $W_{p, \text{OT}_p}(\delta_{\mu_n}, P_n) \leq W_{p, \text{OT}_p}(\delta_0, P_n)$ . Furthermore, as by assumption  $W_{p, \text{OT}_p}(P_n, P) \rightarrow 0$ , we have that  $\sup_n W_{p, \text{OT}_p}(P_n, \delta_0) < \infty$ . 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.

To conclude the proof, we use the following two lemmas, whose proofs are found in Appendix C.

**Lemma 4.2.** *There exists a subsequence  $(\mu_{n_k})_k$  of  $(\mu_n)_n$  which vaguely converges towards  $\mu$  a  $p$ -Fréchet mean of  $P$  and there exists  $\nu \in \mathcal{M}^p$  such that  $\text{OT}_p(\mu_{n_k}, \nu) \rightarrow \text{OT}_p(\mu, \nu)$  as  $k \rightarrow \infty$ .*

**Lemma 4.3.** *Let  $\mu, \mu_1, \mu_2, \dots \in \mathcal{M}^p$ . Then,  $\text{OT}_p(\mu_n, \mu) \rightarrow 0$  if and only if (i)  $\mu_n \xrightarrow{v} \mu$  and (ii) there exists a persistence measure  $\nu \in \mathcal{M}^p$  such that  $\text{OT}_p(\mu_n, \nu) \rightarrow \text{OT}_p(\mu, \nu)$ .*

Let  $\mu'_k = \mu_{n_k}$  be any subsequence of  $\mu_n$ . We want to show that there exists a subsequence of  $\mu'_k$  which converges with respect to the  $\text{OT}_p$  metric towards some  $p$ -Fréchet mean of  $P$ . By Lemma 4.2 applied to the sequence  $(\mu'_k)_k$ , there exists a subsequence  $\mu'_{k_l}$  which converges vaguely to some  $p$ -Fréchet mean  $\mu$  of  $P$ , and some  $\nu$  with  $\text{OT}_p(\mu'_{k_l}, \nu) \rightarrow \text{OT}_p(\mu, \nu)$  as  $l \rightarrow \infty$ . By Lemma 4.3, this implies that  $\mu'_{k_l}$  converges to  $\mu$  with respect to the  $\text{OT}_p$  metric, showing the conclusion.  $\square$

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

**Theorem 4.3.** *For any probability distribution  $P$  supported on  $\mathcal{M}^p$  with finite  $p$ -th moment, the set of  $p$ -Fréchet means of  $P$  is a non-empty compact convex set of  $\mathcal{M}^p$ .*

*Proof.* We first prove the non-emptiness. Let  $P = \sum_{i=1}^N \lambda_i \mu_i$  be a probability measure on  $\mathcal{M}^p$  with finite support  $\mu_1, \dots, \mu_N$ . According to Proposition 3.6, there exists sequences  $(\mu_i^{(n)})_n$  in  $\mathcal{M}_f^p$  with  $\text{OT}_p(\mu_i^{(n)}, \mu_i) \rightarrow 0$ . As a consequence of the result of Section 4.1, the probability measures  $P^{(n)} := \sum_i \lambda_i \delta_{\mu_i^{(n)}}$  admit  $p$ -Fréchet means. Furthermore,  $W_{p, \text{OT}_p}^p(P^{(n)}, P) \leq \sum_i \lambda_i \text{OT}_p^p(\mu_i^{(n)}, \mu_i)$  so that this quantity converges to 0 as  $n \rightarrow \infty$ . It follows from Proposition 4.2 that  $P$  admits a  $p$ -Fréchet mean.

If  $P$  has infinite support, following [46], it can be approximated (in  $W_{p, \text{OT}_p}$ ) by a empirical probability measure  $P_n = \frac{1}{n} \sum_{i=1}^n \delta_{\mu_i}$  where the  $\mu_i$  are i.i.d. from  $P$ . We know that  $P_n$  admits a  $p$ -Fréchet mean since its support is finite, and thus, applying Proposition 4.2 once again, we obtain that  $P$  admits a  $p$ -Fréchet mean.

Finally, the compactness of the set of  $p$ -Fréchet means follows from Proposition 4.2 applied with  $P_n = P$ : if  $(\mu_n)_n$  is a sequence of  $p$ -Fréchet means, then the sequence is relatively compact in  $(\mathcal{M}^p, \text{OT}_p)$ , and any converging subsequence is also a  $p$ -Fréchet mean of  $P$ . Also, the convexity of the set of  $p$ -Fréchet means follows from the convexity of  $\text{OT}_p^p$  (see Lemma D.1 in Appendix D below): if  $\mu_1, \mu_2$  are two  $p$ -Fréchet means with energy  $\mathcal{E}(\mu_1) = \mathcal{E}(\mu_2) = E_0$  and  $0 \leq \lambda \leq 1$ , then

$$\begin{aligned} \mathcal{E}(\lambda\mu_1 + (1-\lambda)\mu_2) &= \int_{\nu \in \mathcal{M}^p} \text{OT}_p^p(\lambda\mu_1 + (1-\lambda)\mu_2, \nu) dP(\nu) \\ &\leq \int_{\nu \in \mathcal{M}^p} (\lambda \text{OT}_p^p(\mu_1, \nu) + (1-\lambda) \text{OT}_p^p(\mu_2, \nu)) dP(\nu) \\ &= \lambda \mathcal{E}(\mu_1) + (1-\lambda) \mathcal{E}(\mu_2) = E_0, \end{aligned}$$

so that  $\lambda\mu_1 + (1-\lambda)\mu_2$  is also a  $p$ -Fréchet mean.  $\square$

### 4.3 $p$ -Fréchet means in $\mathcal{D}^p$

We now prove the existence of  $p$ -Fréchet means for distributions of persistence diagram (i.e. probability distributions supported on  $\mathcal{D}^p$ ), extending the results of [48], in which authors prove their existence for specific probability distributions (namely distributions with compact support or specific rates of decay). Theorem 4.4 below asserts two different things: that  $\operatorname{argmin}\{\mathcal{E}(a), a \in \mathcal{D}^p\}$  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  $p$ -Fréchet means in  $\mathcal{M}^p$ , we start with the finite case. The following lemma actually gives a geometric description of the set of  $p$ -Fréchet means obtained when averaging a finite number of finite diagrams.

**Lemma 4.4.** *Consider  $a_1, \dots, a_N \in \mathcal{D}_f$ , weights  $(\lambda_i)_i$  that sum to 1, and let  $P := \sum_{i=1}^N \lambda_i \delta_{a_i}$ . Then, the set of minimizers of  $\mu \mapsto \sum_{i=1}^N \lambda_i \operatorname{OT}_p^p(\mu, a_i)$  is a non empty convex subset of  $\mathcal{M}_f^p$  whose extreme points belong to  $\mathcal{D}_f$ . In particular,  $P$  admits a  $p$ -Fréchet mean in  $\mathcal{D}_f$ .*

The proof of this lemma is delayed to Appendix C. Note that, as a straightforward consequence, if  $P$  has a unique minimizer in  $\mathcal{D}_f$  (which is generically true [59]), then so it does in  $\mathcal{M}_f^p$ .

**Theorem 4.4.** *For any probability distribution  $P$  supported on  $\mathcal{D}^p$  with finite  $p$ -th moment, the set of  $p$ -Fréchet means of  $P$  contains an element of  $\mathcal{D}^p$ . Furthermore, if  $P$  is supported on a finite set of finite persistence diagrams, then the set of the  $p$ -Fréchet means of  $P$  is a convex set whose extreme points are in  $\mathcal{D}^p$ .*

*Proof.* The second assertion of the theorem is stated in Lemma 4.4. To prove the existence of a  $p$ -Fréchet mean which is a persistence diagram, we argue as in the proof of Theorem 4.3, using additionally the fact that  $\mathcal{D}^p$  is closed in  $\mathcal{M}^p$  (Proposition A.5).  $\square$

## 5 Applications

### 5.1 Characterization of continuous linear representations

As mentioned in the introduction, a *linear representation* of persistence measures (in particular persistence diagrams) is a mapping  $\Phi : \mathcal{M}^p \rightarrow \mathcal{B}$  for some Banach space  $\mathcal{B}$  of the form  $\mu \mapsto \mu(f)$ , where  $f : \Omega \rightarrow \mathcal{B}$  is some chosen function. Doing so, one can turn a sample of diagrams (or measures) into a sample of vectors, making the use of machine learning tools easier. Of course, a minimal expectation is that  $\Phi$  should be continuous. In practice, building a linear representations (see below for a list of examples) generally follows the same pattern: first consider a “nice” function  $g$ , e.g. a gaussian distribution, then introduce a weight with respect to the distance to the diagonal  $d(\cdot, \partial\Omega)^p$ , and prove that  $\mu \mapsto \mu(g(\cdot)d(\cdot, \partial\Omega)^p)$  has some regularity properties (continuity, stability, etc.). Applying Theorem 3.4, we show that this approach always gives a continuous linear representation, and that it is the only way to do so.

For  $\mathcal{B}$  a Banach space (typically  $\mathbb{R}^d$ ), define the class of functions:

$$\mathcal{C}_{b,p}^0 = \left\{ f : \Omega \rightarrow \mathcal{B}, f \text{ continuous and } x \mapsto \frac{f(x)}{d(x, \partial\Omega)^p} \text{ bounded} \right\} \quad (26)$$

**Proposition 5.1.** *Let  $\mathcal{B}$  be a Banach space and  $f : \Omega \rightarrow \mathcal{B}$  a function. The linear representation  $\Phi : \mathcal{M}^p \rightarrow \mathcal{B}$  defined by  $\Phi : \mu \mapsto \mu(f) = \int_{\Omega} f(x) d\mu(x)$  is continuous with respect to  $\text{OT}_p$  if and only if  $f \in C_{b,p}^0$ .*

*Proof.* Consider first the case  $\mathcal{B} = \mathbb{R}$ . Let  $f \in C_{b,p}^0$  and  $\mu, \mu_1, \mu_2 \dots \in \mathcal{M}^p$  be such that  $\text{OT}_p(\mu_n, \mu) \rightarrow 0$ . Recall the definition (18) of  $\mu^{(p)}$ . Using Corollary 3.2, having  $\text{OT}_p(\mu_n, \mu) \rightarrow 0$  means that  $\mu_n^{(p)} \xrightarrow{w} \mu^{(p)}$ , and thus that

$$\int_{\Omega} \frac{f(x)}{d(x, \partial\Omega)^p} d\mu_n^{(p)}(x) \rightarrow \int_{\Omega} \frac{f(x)}{d(x, \partial\Omega)^p} d\mu^{(p)}(x),$$

that is

$$\Phi(\mu_n) = \int_{\Omega} f(x) d\mu_n(x) \rightarrow \int_{\Omega} f(x) d\mu(x) = \Phi(\mu),$$

i.e.  $\Phi$  is continuous with respect to  $\text{OT}_p$ .

Now, let  $\mathcal{B}$  be any Banach space. [49, Theorem 2] states that if a sequence of measures  $(\mu_n)_n$  weakly converges to  $\mu$ , then  $\mu_n(f) \rightarrow \mu(f)$  for any continuous bounded function  $g : \Omega \rightarrow \mathcal{B}$ . Applying this result to the sequence  $(\mu_n^{(p)})$  with  $g = f/d(\cdot, \partial\Omega)^p$  yields the desired result.

Conversely, let  $f : \Omega \rightarrow \mathcal{B}$ . Assume first that  $f$  is not continuous in some  $x \in \Omega$ . There exist a sequence  $(x_n)_n \in \Omega^{\mathbb{N}}$  such that  $x_n \rightarrow x$  but  $f(x_n) \not\rightarrow f(x)$ . Let  $\mu_n = \delta_{x_n}$  and  $\mu = \delta_x$ . We have  $\text{OT}_p(\mu_n, \mu) \rightarrow 0$ , but  $\mu_n(f) = f(x_n) \not\rightarrow f(x) = \mu(f)$ , so that the linear representation  $\mu \mapsto \mu(f)$  cannot be continuous. Then, assume that  $f$  is continuous but that  $x \mapsto \frac{f(x)}{d(x, \partial\Omega)^p}$  is not bounded. Let thus  $(x_n)_n \in \Omega^{\mathbb{N}}$  be a sequence such that  $\left\| \frac{f(x_n)}{d(x_n, \partial\Omega)^p} \right\| \rightarrow +\infty$ . Define the measure  $\mu_n := \frac{1}{\|f(x_n)\|} \delta_{x_n}$ . Observe that  $\text{OT}_p(\mu_n, 0) = \frac{d(x_n, \partial\Omega)^p}{\|f(x_n)\|} \rightarrow 0$  by hypothesis. However,  $\|\mu_n(f)\| = 1$  for all  $n$ , allowing us to conclude once again that  $\mu \mapsto \mu(f)$  cannot be continuous.  $\square$

Let us give some examples of such linear representations (which are thus continuous) commonly used in applications of TDA. Note that the following definitions do not rely on the fact that the input must be a persistence diagram and actually make sense for any persistence measure in  $\mathcal{M}^p$ . See Figure 2 for an illustration, in which computations are done with  $p = 1$  (and  $p' = 1$  for the weighted Betti curve).

- *Persistence surface and its variations.* Let  $K : \mathbb{R}^2 \rightarrow \mathbb{R}$  be a nonnegative Lipschitz continuous bounded function (e.g.  $K(x, y) = \exp\left(-\frac{\|x-y\|^2}{2}\right)$ ) and define  $f : x \in \Omega \mapsto d(x, \partial\Omega)^p \times K(x, \cdot)$ , so that  $f(x) : \mathbb{R}^2 \rightarrow \mathbb{R}$  is a real-valued function. The corresponding representation  $\Phi$  takes its values in  $(C_b(\mathbb{R}^2), \|\cdot\|_{\infty})$ , the (Banach) space of continuous bounded functions. This representation is called the persistence surface and has been introduced with slight variations in different works [1, 19, 43, 53].
- *Persistence silhouettes.* Let  $\Lambda(x, t) = \max\left(\frac{x_2 - x_1}{2} - \left|t - \frac{x_2 + x_1}{2}\right|, 0\right)$  for  $t \in \mathbb{R}$  and  $x \in \Omega$ . Then, defining  $f : x \in \Omega \mapsto d(x, \partial\Omega)^{p-1} \times \Lambda(x, \cdot)$ , one has that  $\|f(x)\|_{\infty}$  is proportional to  $d(x, \partial\Omega)^p$ , so that the corresponding representation is continuous for  $\text{OT}_p$ . This representation is called the

persistence silhouette, and was introduced in [18]. In particular, it consists in a weighted sum of the different functions of the persistence landscape [9]. The corresponding Banach space is  $(C_b(\mathbb{R}), \|\cdot\|_\infty)$ .

- *Weighted Betti curves.* For  $t \in \mathbb{R}$ , define  $B_t$  the rectangle  $(-\infty, t] \times [t, +\infty)$ . Let  $p, p' \geq 1$ , and define  $f : x \in \Omega \mapsto (t \mapsto d(x, \partial\Omega)^{p-1/p'} \mathbf{1}\{x \in B_t\})$ . Then  $f(x) \in L_{p'}(\mathbb{R})$  with  $\|f(x)\|_{p'}$  proportional to  $d(x, \partial\Omega)^p$ . The corresponding function  $\Phi$  is the weighted Betti curve, which takes its values in the Banach space  $(L_{p'}(\mathbb{R}), \|\cdot\|_{p'})$ . In particular, one obtains the continuity of the classical Betti curves from  $(\mathcal{M}^1, \text{OT}_1)$  to  $L_1(\mathbb{R})$ .

**Stability in the case  $p = 1$ .** Continuity is a basic expectation when embedding a set of diagrams (or measures) in some Banach space  $\mathcal{B}$ . One could however ask for more, e.g. some Lipschitz regularity: given a representation  $\Phi : \mathcal{M}^p \rightarrow \mathcal{B}$ , one may want to have  $\|\Phi(\mu) - \Phi(\nu)\| \leq C \cdot \text{OT}_p(\mu, \nu)$  for some constant  $C$ . This property is generally referred to as “stability” in the TDA community and is generally obtained with  $p = 1$ , see for example [1, Theorem 5], [13, Theorem 3.3 & 3.4], [58, §4], [53, Theorem 2], etc.

Here, we still consider the case of linear representations, and show that stability always holds with respect to the distance  $\text{OT}_1$ . Informally, this is explained by the fact that when  $p = 1$ , the cost function  $(x, y) \mapsto d(x, y)^p$  is actually a distance.

**Proposition 5.2.** *Define  $\mathcal{L}$  the set of Lipschitz continuous functions  $f : \bar{\Omega} \rightarrow \mathbb{R}$  with Lipschitz constant less than 1 and that satisfy  $f(\partial\Omega) = 0$ . Let  $T$  be any set, and consider a family  $(f_t)_{t \in T}$  with  $f_t \in \mathcal{L}$ . Then the linear representation  $\Phi : \mu \mapsto (\mu(f_t))_{t \in T}$  is 1-Lipschitz continuous in the following sense:*

$$\|\Phi(\mu) - \Phi(\nu)\|_\infty := \sup_{t \in T} |(\mu - \nu)(f_t)| \leq \text{OT}_1(\mu, \nu), \quad (27)$$

for any measures  $\mu, \nu \in \mathcal{M}^1$ .

*Proof.* Consider  $\mu, \nu \in \mathcal{M}^1$ , and  $\pi \in \text{Opt}(\mu, \nu)$  an optimal transport plan. Let  $t \in T$ . We have:

$$\begin{aligned} (\mu - \nu)(f_t) &= \int_{\Omega} f_t(x) d\mu(x) - \int_{\Omega} f_t(y) d\nu(y) = \iint_{\bar{\Omega} \times \bar{\Omega}} (f_t(x) - f_t(y)) d\pi(x, y) \\ &\leq \iint_{\bar{\Omega} \times \bar{\Omega}} d(x, y) d\pi(x, y) = \text{OT}_1(\mu, \nu), \end{aligned}$$

and thus,  $\|\Phi(\mu) - \Phi(\nu)\|_\infty \leq \text{OT}_1(\mu, \nu)$ .  $\square$

In particular, if  $f : \bar{\Omega} \rightarrow \mathcal{B}$ , where  $\mathcal{B}$  is some Banach space, is 1-Lipschitz with  $f(\partial\Omega) = 0$ , then one can let  $T = \mathcal{B}_1^*$  (the unit ball of the dual of  $\mathcal{B}$ ) and  $f_t(x) := t(f(x))$  for  $t \in T$ . If  $\Phi(\mu) = \mu(f)$ , we then obtain that  $\|\Phi(\mu) - \Phi(\nu)\| \leq \text{OT}_1(\mu, \nu)$ , i.e. that  $\Phi : (\mathcal{M}^1, \text{OT}_1) \rightarrow (\mathcal{B}, \|\cdot\|)$  is 1-Lipschitz.

**Remark 5.1.** *One actually has a converse of such an inequality, i.e. it can be shown that*

$$\text{OT}_1(\mu, \nu) = \max_{f \in \mathcal{L}} (\mu - \nu)(f), \quad (28)$$

This equation is an adapted version of the very well-known Kantorovich-Rubinstein formula, which is itself a particular version in the case  $p = 1$  of the duality formula in optimal transport, see for example [64, Theorem 5.10] and [54, Theorem 1.39]. A proof of Eq. (28) would require to introduce several optimal transport notions. The interested reader can consult Proposition 2.3 in [28] for details.

## 5.2 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 [6, 25, 35, 32, 57] used techniques originating from this field to understand the persistent homology of such point processes. Let  $\mathbb{X}_n := \{X_1, \dots, X_n\}$  be a  $n$ -sample of a distribution having some density on the cube  $[0, 1]^d$ , bounded from below and above by positive constants. Extending the work of Hiraoka, Shirai and Trinh [35], Divol and Polonik [25] show laws of large numbers for the persistence diagrams  $\text{Dgm}(\mathbb{X}_n)$  of  $\mathbb{X}_n$ , built with either the Čech or Rips filtration. More precisely, [25, Theorem 5] states that there exists a Radon measure  $\mu$  on  $\Omega$  such that, almost surely, the sequence of measures  $\mu_n := n^{-1} \text{Dgm}(n^{1/d} \mathbb{X}_n)$  converges vaguely to  $\mu$  and [25, Theorem 6] implies that  $\text{Pers}_p(\mu_n)$  converges to  $\text{Pers}_p(\mu) < \infty$  for all  $p \geq 1$ . Those two facts (vague convergence and convergence of the total persistence), along with Theorem 3.4, gives the following result:

**Proposition 5.3.**  $\text{OT}_p(\mu_n, \mu) \xrightarrow[n \rightarrow \infty]{} 0$  almost surely.

### Numerical illustration of the convergence in the one-dimensional case.

In dimension  $d \geq 2$ , there is no known closed-form expression for the limit  $\mu$ . However, in the case  $d = 1$ , authors in [25, Remark 2 (b)] show that if the  $X_i$  are  $n$  i.i.d realizations of a random variable  $X$  admitting a density  $\kappa$  supported on  $[0, 1]$ , bounded from below and above by positive constants, then (the ordinate of)  $\mu$ —which is supported on  $\{0\} \times (0, +\infty)$  as we consider the Rips filtration in homology dimension 0—admits  $\varphi(u) := \mathbb{E}_{X \sim \kappa}[\exp(-u\kappa(X))]$  as density. In particular, if  $X$  is uniform, then (the ordinate of)  $\mu$  admits  $u \mapsto e^{-u}$  as density.

It allows us to realize a simple numerical experiment: we sample  $n$  points  $(X_1 \dots X_n)$  uniformly on  $[0, 1]$  and then compute the corresponding persistence diagram using the Rips filtration, whose points are denoted by  $(0, t_1), \dots, (0, t_{n-1})$  (note that we removed the point  $(0, +\infty)$ ). We can now introduce the one-dimensional measure  $\mu_n = \frac{1}{n} \sum_{k=1}^{n-1} \delta_{t_k}$  and compute  $\text{OT}_p(\mu_n, \mu)$  in closed form using Proposition 3.7 and the fact that computing the distance  $W_p$  in dimension  $d = 1$  is particularly easy (see [54, Chapter 2]). See Figure 6 for an illustration.

## 5.3 Stability of the expected persistence diagrams

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



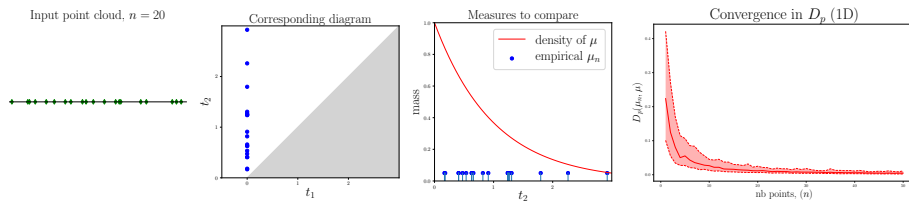


Figure 6: Numerical illustration of  $D_p$  convergence in the one dimensional case. From left to right: an example of input point cloud with  $n = 20$  points ; the corresponding persistence diagram ; the (ordinate of) the two measures we compare:  $\mu_n$  is the empirical one. Most right: median  $\text{OT}_p(\mu_n, \mu)$  for  $n = 2 \dots 50$ , over 100 runs along with 90% and 10% percentiles. Computations made with  $p = 2$ .

Bochner integral) and is detailed in Appendix D. It however satisfies the natural following characterization—that is sufficient to understand this section:

$$\forall K \subset \Omega \text{ compact}, \mathbb{E}[\boldsymbol{\mu}](K) = \mathbb{E}[\boldsymbol{\mu}(K)]. \quad (29)$$

The behavior of such measures is studied in [24], 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  $P$  and  $P'$  supported on  $\mathcal{M}^p$ , their linear expectations are close for the metric  $\text{OT}_p$ .

**Proposition 5.4.** *Let  $P, P' \in \mathcal{W}^p(\mathcal{M}^p)$ . Let  $\boldsymbol{\mu} \sim P$  and  $\boldsymbol{\mu}' \sim P'$ . Then, we have  $\text{OT}_p(\mathbb{E}[\boldsymbol{\mu}], \mathbb{E}[\boldsymbol{\mu}']) \leq W_{p, \text{OT}_p}(P, P')$ .*

The proof is postponed to Appendix D.

Using stability results on the distances  $d_p (= \text{OT}_p)$  between persistence diagrams [22], 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  $\text{Dgm}(\mathbb{Y})$  the persistence diagram of  $\mathbb{Y}$  built with the Čech filtration.

**Proposition 5.5.** *Let  $\xi, \xi'$  be two probability measures on  $\mathbb{R}^d$ . Let  $\mathbb{X}_n$  (resp.  $\mathbb{X}'_n$ ) be a  $n$ -sample of law  $\xi$  (resp.  $\xi'$ ). Then, for any  $k > d$ , and any  $p \geq k + 1$ ,*

$$\text{OT}_p^p(\mathbb{E}[\text{Dgm}(\mathbb{X}_n)], \mathbb{E}[\text{Dgm}(\mathbb{X}'_n)]) \leq C_{k,d} \cdot n \cdot W_{p-k}^{p-k}(\xi, \xi') \quad (30)$$

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

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

$$\text{OT}_\infty(\mathbb{E}[\text{Dgm}(\mathbb{X}_n)], \mathbb{E}[\text{Dgm}(\mathbb{X}'_n)]) \leq W_\infty(\xi, \xi'). \quad (31)$$

*Proof.* Let  $P$  be the law of  $\text{Dgm}(\mathbb{X}_n)$ ,  $P'$  be the law of  $\text{Dgm}(\mathbb{X}'_n)$  and let  $\gamma$  be any coupling between  $\mathbb{X}_n$  a  $n$ -sample of law  $\xi$ , and  $\mathbb{X}'_n$  a  $n$ -sample of law  $\xi'$ . Then, the law of  $(\text{Dgm}(\mathbb{X}_n), \text{Dgm}(\mathbb{X}'_n))$  is a coupling between  $P$  and  $P'$ . Thus, Proposition 5.4 yields

$$\text{OT}_p^p(\mathbb{E}[\text{Dgm}(\mathbb{X}_n)], \mathbb{E}[\text{Dgm}(\mathbb{X}'_n)]) \leq \mathbb{E}_\gamma[\text{OT}_p^p(\text{Dgm}(\mathbb{X}_n), \text{Dgm}(\mathbb{X}'_n))].$$

It is stated in [22, Wasserstein Stability Theorem] that

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

where  $C_{k,d} := C \text{diam}(\mathbb{X})^{k-d} \frac{k}{k-d}$  for some constant  $C$  depending only on  $\mathbb{X}$ , and  $H$  is the Hausdorff distance between sets. By taking the infimum on transport plans  $\gamma$ , we obtain

$$\text{OT}_p^p(\mathbb{E}[\text{Dgm}(\mathbb{X}_n)], \mathbb{E}[\text{Dgm}(\mathbb{X}'_n)]) \leq C_{k,d} W_{p-k,H}^{p-k}(\xi^{\otimes n}, (\xi')^{\otimes n}),$$

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 [17] states that

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

concluding the proof.  $\square$

Note that this proposition illustrates the usefulness of introducing new distances  $\text{OT}_p$ : considering the proximity between linear expectations requires to extend the metrics  $d_p$  to Radon measures.

## 6 Conclusion and further work.

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 of persistence diagrams and measures with respect to optimal transport metrics in terms of convergence for measures. We then prove existence and consistency of  $p$ -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: continuity of diagram representations, law of large number for persistence diagrams, stability of diagrams in a random settings.

We believe that closing the gap between optimal transport metrics and those of topological data analysis will help to develop new approaches in TDA and give a better understanding of the statistical behavior of topological descriptors. It allows one to address new problems, in particular those involving continuous counterpart of persistence diagrams, and invites one to consider various theoretical and computational tools developed in optimal transport theory that could be of major interest in topological data analysis: gradient flows [55] in the persistence diagrams space, entropic regularization [23] of metrics, use of diagram metrics in (deep) learning pipelines [31], along with the use of well developed libraries [29].

## References

- [1] Adams, H., Emerson, T., Kirby, M., Neville, R., Peterson, C., Shipman, P., Chepushtanova, S., Hanson, E., Motta, F., Ziegelmeier, L.: Persistence images: a stable vector representation of persistent homology. *Journal of Machine Learning Research* **18**(8), 1–35 (2017)
- [2] Agueh, M., Carlier, G.: Barycenters in the Wasserstein space. *SIAM Journal on Mathematical Analysis* **43**(2), 904–924 (2011)
- [3] Ambrosio, L., Gigli, N., Savaré, G.: Gradient flows: in metric spaces and in the space of probability measures. Springer Science & Business Media (2008)
- [4] Billingsley, P.: Convergence of probability measures. Wiley Series in Probability and Statistics. Wiley (2013)
- [5] Blumberg, A.J., Gal, I., Mandell, M.A., Pancia, M.: Robust statistics, hypothesis testing, and confidence intervals for persistent homology on metric measure spaces. *Foundations of Computational Mathematics* **14**(4), 745–789 (2014)
- [6] Bobrowski, O., Kahle, M., Skraba, P., et al.: Maximally persistent cycles in random geometric complexes. *The Annals of Applied Probability* **27**(4), 2032–2060 (2017)
- [7] Bochner, S.: Integration von funktionen, deren werte die elemente eines vektorraumes sind. *Fundamenta Mathematicae* **20**(1), 262–176 (1933)
- [8] Bogachev, V.: Measure theory. No. v. 1 in Measure Theory. Springer Berlin Heidelberg (2007)
- [9] Bubenik, P., Dłotko, P.: A persistence landscapes toolbox for topological statistics. *Journal of Symbolic Computation* **78**, 91–114 (2017)
- [10] Bubenik, P., Vergili, T.: Topological spaces of persistence modules and their properties. *Journal of Applied and Computational Topology* pp. 1–37 (2018)
- [11] Carlier, G., Ekeland, I.: Matching for teams. *Economic Theory* **42**(2), 397–418 (2010)
- [12] Carlier, G., Oberman, A., Oudet, E.: Numerical methods for matching for teams and Wasserstein barycenters. *ESAIM: Mathematical Modelling and Numerical Analysis* **49**(6), 1621–1642 (2015)
- [13] Carrière, M., Cuturi, M., Oudot, S.: Sliced Wasserstein kernel for persistence diagrams. In: 34th International Conference on Machine Learning (2017)
- [14] Carrière, M., Oudot, S.Y., Ovsjanikov, M.: Stable topological signatures for points on 3d shapes. *Computer Graphics Forum* **34**(5), 1–12 (2015). DOI 10.1111/cgf.12692
- [15] Cascales, B., Raja, M.: Measurable selectors for the metric projection. *Mathematische Nachrichten* **254**(1), 27–34 (2003)

- [16] Chazal, F., De Silva, V., Glisse, M., Oudot, S.: The structure and stability of persistence modules. Springer (2016)
- [17] Chazal, F., Fasy, B., Lecci, F., Michel, B., Rinaldo, A., Wasserman, L.: Subsampling methods for persistent homology. In: International Conference on Machine Learning, pp. 2143–2151 (2015)
- [18] Chazal, F., Fasy, B.T., Lecci, F., Rinaldo, A., Wasserman, L.A.: Stochastic convergence of persistence landscapes and silhouettes. *JoCG* **6**(2), 140–161 (2015). DOI 10.20382/jocg.v6i2a8
- [19] Chen, Y.C., Wang, D., Rinaldo, A., Wasserman, L.: Statistical analysis of persistence intensity functions. arXiv preprint arXiv:1510.02502 (2015)
- [20] Chizat, L., Peyré, G., Schmitzer, B., Vialard, F.X.: Unbalanced optimal transport: geometry and kantorovich formulation. arXiv preprint arXiv:1508.05216 (2015)
- [21] Cohen-Steiner, D., Edelsbrunner, H., Harer, J.: Stability of persistence diagrams. *Discrete & Computational Geometry* **37**(1), 103–120 (2007)
- [22] Cohen-Steiner, D., Edelsbrunner, H., Harer, J., Mileyko, Y.: Lipschitz functions have  $L_p$ -stable persistence. *Foundations of computational mathematics* **10**(2), 127–139 (2010)
- [23] Cuturi, M.: Sinkhorn distances: Lightspeed computation of optimal transport. In: *Advances in Neural Information Processing Systems*, pp. 2292–2300 (2013)
- [24] Divol, V., Chazal, F.: The density of expected persistence diagrams and its kernel based estimation. *JoCG* **10**(2), 127–153 (2019). DOI 10.20382/jocg.v10i2a7
- [25] Divol, V., Polonik, W.: On the choice of weight functions for linear representations of persistence diagrams. *Journal of Applied and Computational Topology* **3**(3), 249–283 (2019)
- [26] Edelsbrunner, H., Harer, J.: *Computational topology: an introduction*. American Mathematical Soc. (2010)
- [27] Figalli, A.: The optimal partial transport problem. *Archive for rational mechanics and analysis* **195**(2), 533–560 (2010)
- [28] Figalli, A., Gigli, N.: 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)
- [29] Flamary, R., Courty, N.: POT python optimal transport library (2017). URL <https://github.com/rflamary/POT>
- [30] Folland, G.: *Real analysis: modern techniques and their applications*. Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts. Wiley (2013)

- [31] Genevay, A., Peyre, G., Cuturi, M.: Learning generative models with sinkhorn divergences. In: International Conference on Artificial Intelligence and Statistics, pp. 1608–1617 (2018)
- [32] Goel, A., Trinh, K.D., Tsunoda, K.: Asymptotic behavior of Betti numbers of random geometric complexes. arXiv preprint arXiv:1805.05032 (2018)
- [33] Hall, M.: Combinatorial theory (2nd ed). John Wiley & Sons (1986)
- [34] Hiraoka, Y., Nakamura, T., Hirata, A., Escolar, E.G., Matsue, K., Nishiura, Y.: Hierarchical structures of amorphous solids characterized by persistent homology. Proceedings of the National Academy of Sciences (2016). DOI 10.1073/pnas.1520877113
- [35] Hiraoka, Y., Shirai, T., Trinh, K.D., et al.: Limit theorems for persistence diagrams. The Annals of Applied Probability **28**(5), 2740–2780 (2018)
- [36] Hofer, C.D., Kwitt, R., Niethammer, M.: Learning representations of persistence barcodes. Journal of Machine Learning Research **20**(126), 1–45 (2019)
- [37] Kallenberg, O.: Random measures. Elsevier Science & Technology Books (1983)
- [38] Kechris, A.: Classical descriptive set theory. Graduate Texts in Mathematics. Springer-Verlag (1995)
- [39] Kerber, M., Morozov, D., Nigmatov, A.: Geometry helps to compare persistence diagrams. Journal of Experimental Algorithmics (JEA) **22**(1), 1–4 (2017)
- [40] Kondratyev, S., Monsaingeon, L., Vorotnikov, D., et al.: A new optimal transport distance on the space of finite Radon measures. Advances in Differential Equations **21**(11/12), 1117–1164 (2016)
- [41] Kramar, M., Goulet, A., Kondic, L., Mischaikow, K.: Persistence of force networks in compressed granular media. Phys. Rev. E **87**, 042207 (2013). DOI 10.1103/PhysRevE.87.042207
- [42] Kusano, G., Fukumizu, K., Hiraoka, Y.: Kernel method for persistence diagrams via kernel embedding and weight factor. The Journal of Machine Learning Research **18**(1), 6947–6987 (2017)
- [43] Kusano, G., Hiraoka, Y., Fukumizu, K.: Persistence weighted gaussian kernel for topological data analysis. In: International Conference on Machine Learning, pp. 2004–2013 (2016)
- [44] Kwitt, R., Huber, S., Niethammer, M., Lin, W., Bauer, U.: Statistical topological data analysis - a kernel perspective. In: Advances in neural information processing systems, pp. 3070–3078 (2015)
- [45] Lacombe, T., Cuturi, M., Oudot, S.: Large scale computation of means and clusters for persistence diagrams using optimal transport. In: Advances in Neural Information Processing Systems (2018)

- [46] Le Gouic, T., Loubes, J.M.: Existence and consistency of Wasserstein barycenters. *Probability Theory and Related Fields* pp. 1–17 (2016)
- [47] Li, C., Ovsjanikov, M., Chazal, F.: Persistence-based structural recognition. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2014)
- [48] Mileyko, Y., Mukherjee, S., Harer, J.: Probability measures on the space of persistence diagrams. *Inverse Problems* **27**(12), 124007 (2011)
- [49] Nielsen, L.: Weak convergence and Banach space-valued functions: improving the stability theory of feynmans operational calculi. *Mathematical Physics, Analysis and Geometry* **14**(4), 279–294 (2011)
- [50] Oudot, S.Y.: *Persistence theory: from quiver representations to data analysis*, vol. 209. American Mathematical Society (2015)
- [51] Perlman, M.D.: Jensen’s inequality for a convex vector-valued function on an infinite-dimensional space. *Journal of Multivariate Analysis* **4**(1), 52–65 (1974)
- [52] Peyré, G., Cuturi, M.: *Computational optimal transport*. 2017-86 (2017)
- [53] Reininghaus, J., Huber, S., Bauer, U., Kwitt, R.: A stable multi-scale kernel for topological machine learning. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4741–4748 (2015)
- [54] Santambrogio, F.: *Optimal transport for applied mathematicians*. Birkäuser, NY (2015)
- [55] Santambrogio, F.: Euclidean, metric, and Wasserstein gradient flows: an overview. *Bulletin of Mathematical Sciences* **7**(1), 87–154 (2017)
- [56] Schrijver, A.: *Combinatorial optimization: polyhedra and efficiency*, vol. 24. Springer Science & Business Media (2003)
- [57] Schweinhart, B.: Weighted persistent homology sums of random Čech complexes. *arXiv preprint arXiv:1807.07054* (2018)
- [58] Som, A., Thopalli, K., Natesan Ramamurthy, K., Venkataraman, V., Shukla, A., Turaga, P.: Perturbation robust representations of topological persistence diagrams. In: *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 617–635 (2018)
- [59] Turner, K.: Means and medians of sets of persistence diagrams. *arXiv preprint arXiv:1307.8300* (2013)
- [60] Turner, K., Mileyko, Y., Mukherjee, S., Harer, J.: Fréchet means for distributions of persistence diagrams. *Discrete & Computational Geometry* **52**(1), 44–70 (2014)
- [61] Turner, K., Mukherjee, S., Boyer, D.M.: Persistent homology transform for modeling shapes and surfaces. *Information and Inference: A Journal of the IMA* **3**(4), 310–344 (2014). DOI 10.1093/imaiai/ia011

- [62] Umeda, Y.: Time series classification via topological data analysis. *Information and Media Technologies* **12**, 228–239 (2017)
- [63] Villani, C.: *Topics in optimal transportation*. 58. American Mathematical Soc. (2003)
- [64] Villani, C.: *Optimal transport: old and new*, vol. 338. Springer Science & Business Media (2008)

## A Elements of measure theory

In the following,  $\Omega$  denotes a locally compact Polish metric space (i.e. a Polish space equipped with a distinguished Polish metric).

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

Radon measures on a general space are also required to be regular (i.e. well approximated by above by open sets and by below by compact sets). However, on a locally compact Polish metric space (such as  $\Omega$ ), regularity is implied by the above definition (see [30, Section 7.1 and Theorem 7.8] for details).

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

See [8, Chapter 8] for more details on the weak topology on the set of finite Borel measures (which coincide with the set of Baire measures for  $\Omega$  a metrizable space). 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  $\mu(\Omega \setminus K) \leq \varepsilon$  for every  $\mu \in F$ .*

The following propositions are standard results. Corresponding proofs can be found for instance in [37, 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  $\Omega$ ,*

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

**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  $\mu, \mu_1, \mu_2, \dots$  be measures in  $\mathcal{M}_f(\Omega)$ . Then,  $\mu_n \xrightarrow{w} \mu$  if and only if  $\mu_n(\Omega) \rightarrow \mu(\Omega)$  and  $\mu_n \xrightarrow{v} \mu$ .*

**Proposition A.4** (The Portmanteau theorem). *Let  $\mu, \mu_1, \mu_2, \dots$  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  $F \subset \Omega$ ,

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

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



**Definition A.4.** *The set of point measures on  $\Omega$  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 [28] or which are close to standard proofs in optimal transport theory.

*Proofs of Proposition 3.1 and Proposition 3.8.*

- For  $\pi \in \text{Adm}(\mu, \nu)$  supported on  $E_\Omega$ , 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  $E_\Omega$  is included in a set of the form  $(K \times \overline{\Omega}) \cup (\overline{\Omega} \times K')$ , Proposition A.1 implies that  $\text{Adm}(\mu, \nu)$  is relatively compact for the vague convergence on  $E_\Omega$ . Also, if a sequence  $(\pi_n)_n$  in  $\text{Adm}(\mu, \nu)$  converges vaguely to some  $\pi \in \mathcal{M}(E_\Omega)$ , then the marginals of  $\pi$  are still  $\mu$  and  $\nu$ . Indeed, if  $f$  is a continuous function with compact support on  $\Omega$ , then

$$\int_{E_\Omega} f(x) d\pi(x, y) = \lim_n \int_{E_\Omega} f(x) d\pi_n(x, y) = \lim_n \int_\Omega f(x) d\mu_n(x) = \int_\Omega f(x) d\mu(x),$$

and we show likewise that the second marginal of  $\pi$  is  $\nu$ . Hence,  $\text{Adm}(\mu, \nu)$  is closed and relatively compact 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  $\pi'_n : A \mapsto \iint_A d(x, y)^p d\pi_n$ . The sequence  $(\pi'_n)_n$  still converges vaguely to  $\pi' : A \mapsto \iint_A d(x, y)^p d\pi$ . the Portmanteau theorem (Proposition A.4) applied with the open set  $E_\Omega$  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  $C_\infty$ . Let  $(\pi_n)_n$  be a sequence converging vaguely to  $\pi$  on  $E_\Omega$  and let  $r > \liminf_{n \rightarrow \infty} C_\infty(\pi_n)$ . The set  $U_r = \{(x, y) \in E_\Omega, d(x, y) > r\}$  is open. By the Portmanteau theorem (Proposition A.4), we have

$$0 = \liminf_{n \rightarrow \infty} \pi_n(U_r) \geq \pi(U_r).$$

Therefore,  $\text{spt}(\pi) \subset U_r^c$  and  $C_\infty(\pi) \leq r$ . As this holds for any  $r > \liminf_{n \rightarrow \infty} C_\infty(\pi_n)$ , we have  $\liminf_{n \rightarrow \infty} C_\infty(\pi_n) \geq C_\infty(\pi)$ .

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

1. Let  $(\pi_n)_n$  be a minimizing sequence of (10) or (23) in  $\text{Adm}(\mu, \nu)$ . As  $\text{Adm}(\mu, \nu)$  is sequentially compact, it has an adherence value  $\pi$ , and the lower semi-continuity implies that  $C_p(\pi) \leq \liminf_{n \rightarrow \infty} C_p(\pi_n) = \text{OT}_p^p(\mu, \nu)$ , so that  $\text{Opt}_p(\mu, \nu)$  is non-empty. Using once again the lower semi-continuity of  $C_p$ , if a sequence in  $\text{Opt}_p(\mu, \nu)$  converges to some limit, then the cost of the limit is less than or equal to (and thus equal to)  $\text{OT}_p^p(\mu, \nu)$ , i.e. the limit is in  $\text{Opt}_p(\mu, \nu)$ . The set  $\text{Opt}_p(\mu, \nu)$  being closed in the sequentially compact set  $\text{Adm}(\mu, \nu)$ , it is also sequentially compact.
2. Let  $\mu_n \xrightarrow{v} \mu$  and  $\nu_n \xrightarrow{v} \nu$ . One has  $\liminf_n \text{OT}_p(\mu_n, \nu_n) = \lim_k \text{OT}_p(\mu_{n_k}, \nu_{n_k})$  for some subsequence  $(n_k)_k$ . For ease of notation, we will still use the index  $n$  to denote this subsequence. If the limit is infinite, there is nothing to prove. Otherwise, consider  $\pi_n \in \text{Opt}_p(\mu_n, \nu_n)$ . For any compact sets  $K, K' \subset \Omega$ , one has  $\pi_n((K \times \bar{\Omega}) \cup (\bar{\Omega} \times K')) \leq \sup_n \mu_n(K) + \sup_n \nu_n(K') < \infty$ . Therefore, by Proposition A.1, there exists a subsequence  $(\pi_{n_k})_k$  which converges vaguely to some measure  $\pi \in \text{Adm}(\mu, \nu)$ . Note that the first (resp. second) marginal of  $\pi$  is equal to the limit  $\mu$  (resp.  $\nu$ ) of the first (resp. second) marginal of  $(\pi_{n_k})_k$ , so that  $\pi$  is in  $\text{Adm}(\mu, \nu)$ . Therefore,

$$\text{OT}_p^p(\mu, \nu) \leq C_p(\pi) \leq \liminf_{n \rightarrow \infty} C_p(\pi_n) = \liminf_{n \rightarrow \infty} \text{OT}_p^p(\mu_n, \nu_n).$$

- Finally, we prove that  $\text{OT}_p$  is a metric on  $\mathcal{M}^p$ . Let  $\mu, \nu, \lambda \in \mathcal{M}^p$ . The symmetry of  $\text{OT}_p$  is clear. If  $\text{OT}_p(\mu, \nu) = 0$ , then there exists  $\pi \in \text{Adm}(\mu, \nu)$  supported on  $\{(x, x), x \in \Omega\}$ . Therefore, for a Borel set  $A \subset \Omega$ ,  $\mu(A) = \pi(A \times \bar{\Omega}) = \pi(A \times A) = \pi(\bar{\Omega} \times A) = \nu(A)$ , and  $\mu = \nu$ . To prove the triangle inequality, we need a variant on the gluing lemma, stated in [28, Lemma 2.1]: for  $\pi_{12} \in \text{Opt}(\mu, \nu)$  and  $\pi_{23} \in \text{Opt}(\nu, \lambda)$  there exists a measure  $\gamma \in \mathcal{M}(\bar{\Omega}^3)$  such that the marginal corresponding to the first two entries (resp. two last entries), when restricted to  $E_\Omega$ , is equal to  $\pi_{12}$  (resp.  $\pi_{23}$ ), and induces a zero cost on  $\partial\Omega \times \partial\Omega$ . Therefore, by the triangle inequality and the Minkowski inequality,

$$\begin{aligned} \text{OT}_p(\mu, \lambda) &\leq \left( \int_{\bar{\Omega}^2} d(x, z)^p d\gamma(x, y, z) \right)^{1/p} \\ &\leq \left( \int_{\bar{\Omega}^2} d(x, y)^p d\gamma(x, y, z) \right)^{1/p} + \left( \int_{\bar{\Omega}^2} d(y, z)^p d\gamma(x, y, z) \right)^{1/p} \\ &= \left( \int_{\bar{\Omega}^2} d(x, y)^p d\pi_{12}(x, y) \right)^{1/p} + \left( \int_{\bar{\Omega}^2} d(y, z)^p d\pi_{23}(y, z) \right)^{1/p} \\ &= \text{OT}_p(\mu, \nu) + \text{OT}_p(\nu, \lambda). \end{aligned}$$

The proof is similar for  $p = \infty$ .

□

*Proof of Proposition 3.3.* We first show the separability. Consider for  $k > 0$  a partition of  $\Omega$  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  $\text{Pers}_p(\mu) < \infty$ , there exists a compact  $K \subset \Omega$  such that  $\text{Pers}_p(\mu) - \text{Pers}_p(\mu_0) < \varepsilon^p$ , where  $\mu_0$  is the restriction of  $\mu$  to  $K$ . By considering the transport plan between  $\mu$  and  $\mu_0$  induced by the identity map on  $K$  and the projection onto the diagonal on  $\bar{\Omega} \setminus K$ , it follows that  $\text{OT}_p^p(\mu, \mu_0) \leq \text{Pers}_p(\mu) - \text{Pers}_p(\mu_0) \leq \varepsilon^p$ .
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  $\mu_1 = \sum_{i \in I} \mu_0(C_i^k) \delta_{x_i^k}$ . One can create a transport map between  $\mu_0$  and  $\mu_1$  by mapping each square  $C_i^k$  to its center  $x_i^k$ , so that

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

3. Consider, for  $i \in I$ ,  $q_i$  a rational number satisfying  $q_i \leq \mu_0(C_i^k)$  and  $|\mu_0(C_i^k) - q_i| \leq \varepsilon^p / (\sum_{i \in I} d(x_i^k, \partial\Omega)^p)$ . Let  $\mu_2 = \sum_{i \in I} q_i \delta_{x_i^k}$ . Consider the transport plan between  $\mu_2$  and  $\mu_1$  that fully transports  $\mu_2$  onto  $\mu_1$ , and transport the remaining mass in  $\mu_1$  onto the diagonal. Then,

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

As  $\mu_2 \in F$  and  $\text{OT}_p(\mu, \mu_2) \leq 3\varepsilon$ , the separability is proven.

To prove that the space is complete, consider a Cauchy sequence  $(\mu_n)_n$ . As the sequence  $(\text{Pers}_p(\mu_n))_n = (\text{OT}_p^p(\mu_n, 0))_n$  is a Cauchy sequence, it is bounded. Therefore, for  $K \subset \Omega$  a compact set, (14) implies that  $\sup_n \mu_n(K) < \infty$ . Proposition A.1 implies that  $(\mu_n)_n$  is relatively compact for the vague topology on  $\Omega$ . Consider  $(\mu_{n_k})_k$  a subsequence converging vaguely on  $\Omega$  to some measure  $\mu$ . By the lower semi-continuity of  $\text{OT}_p$ ,

$$\text{Pers}_p(\mu) = \text{OT}_p^p(\mu, 0) \leq \liminf_{k \rightarrow \infty} \text{OT}_p^p(\mu_{n_k}, 0) < \infty,$$

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

$$\begin{aligned} \text{OT}_p(\mu_n, \mu) &\leq \liminf_{k \rightarrow \infty} \text{OT}_p(\mu_n, \mu_{n_k}) \\ \lim_{n \rightarrow \infty} \text{OT}_p(\mu_n, \mu) &\leq \lim_{n \rightarrow \infty} \liminf_{k \rightarrow \infty} \text{OT}_p(\mu_n, \mu_{n_k}) = 0, \end{aligned}$$

ensuring that  $\text{OT}_p(\mu_n, \mu) \rightarrow 0$ , that is the space is complete.  $\square$

*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  $(\text{OT}_p(\mu_n, \mu))_n$  converges to 0. The triangle inequality implies that  $\text{Pers}_p(\mu_n) = \text{OT}_p^p(\mu_n, 0)$  converges to  $\text{Pers}_p(\mu) = \text{OT}_p^p(\mu, 0)$ . Let  $f \in C_c(\Omega)$ , whose support is included in some compact set  $K$ . For any

$\varepsilon > 0$ , there exists a Lipschitz function  $f_\varepsilon$ , with Lipschitz constant  $L$  and whose support is included in  $K$ , with the  $\infty$ -norm  $\|f - f_\varepsilon\|_\infty$  less than or equal to  $\varepsilon$ . The convergence of  $\text{Pers}_p(\mu_n)$  and (14) imply that  $\sup_k \mu_k(K) < \infty$ . Let  $\pi_n \in \text{Opt}_p(\mu_n, \mu)$ , 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,

$$\begin{aligned} |\mu_n(f_\varepsilon) - \mu(f_\varepsilon)| &\leq \iint_{\bar{\Omega}^2} |f_\varepsilon(x) - f_\varepsilon(y)| d\pi_n(x, y) \quad \text{where } \pi_n \in \text{Opt}(\mu_n, \mu) \\ &\leq L \iint_{(K \times \bar{\Omega}) \cup (\bar{\Omega} \times K)} d(x, y) d\pi_n(x, y) \\ &\leq L \pi_n((K \times \bar{\Omega}) \cup (\bar{\Omega} \times K))^{1 - \frac{1}{p}} \left( \iint_{(K \times \bar{\Omega}) \cup (\bar{\Omega} \times K)} d(x, y)^p d\pi_n(x, y) \right)^{\frac{1}{p}} \\ &\quad \text{by Hölder's inequality.} \\ &\leq L \left( \sup_k \mu_k(K) + \mu(K) \right)^{1 - \frac{1}{p}} \text{OT}_p(\mu_n, \mu) \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

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

## C Proofs of the technical lemmas of Section 4

The following proof is already found in [46]. We reproduce it here for the sake of completeness.

*Proof of Lemma 4.2.* Recall that  $P_n$  is a sequence in  $\mathcal{W}^p(\mathcal{M}^p)$  such that each  $P_n$  has a  $p$ -Fréchet mean  $\mu_n$  and that  $W_{p, \text{OT}_p}(P_n, P) \rightarrow 0$  for some  $P \in \mathcal{W}^p(\mathcal{M}^p)$ . According to the beginning of the proof of Proposition 4.2, the sequence  $(\mu_n)_n$  is relatively compact for the vague convergence. Let  $\nu \in \mathcal{M}^p$  and let  $\mu$  be the vague limit of some subsequence, which, for ease of notations, will be denoted as the initial sequence. By Skorokhod's representation theorem [4, Theorem 6.7], as  $P_n$  converges weakly to  $P$ , there exists a probabilistic space on which are defined random variables  $\boldsymbol{\mu} \sim P$  and  $\boldsymbol{\mu}_n \sim P_n$  for  $n \geq 0$ , such that  $\boldsymbol{\mu}_n$  converges almost surely with respect to the  $\text{OT}_p$  metric towards  $\boldsymbol{\mu}$ . Using those random variables,

we have

$$\begin{aligned}
\mathcal{E}(\nu) &= \mathbb{E} \text{OT}_p^p(\nu, \boldsymbol{\mu}) = W_{p, \text{OT}_p}^p(\delta_\nu, P) \\
&= \lim_n W_{p, \text{OT}_p}^p(\delta_\nu, P_n) \text{ since } W_{p, \text{OT}_p}(P_n, P) \rightarrow 0 \\
&= \lim_n \mathbb{E} \text{OT}_p^p(\nu, \boldsymbol{\mu}_n) \\
&\geq \lim_n \mathbb{E} \text{OT}_p^p(\mu_n, \boldsymbol{\mu}_n) \text{ since } \mu_n \text{ is a barycenter of } P_n \\
&\geq \mathbb{E} \liminf_n \text{OT}_p^p(\mu_n, \boldsymbol{\mu}_n) \text{ by Fatou's lemma} \\
&\geq \mathbb{E} \text{OT}_p^p(\mu, \boldsymbol{\mu}) = \mathcal{E}(\mu) \text{ by lower semi-continuity of } \text{OT}_p \text{ (Prop. 3.1)}.
\end{aligned} \tag{32}$$

This implies that  $\mu$  is a barycenter of  $P$ . We are now going to show that, almost surely,  $\liminf_n \text{OT}_p(\mu_n, \boldsymbol{\mu}) = \text{OT}_p(\mu, \boldsymbol{\mu})$ . This concludes the proof by letting  $n_k$  be the subsequence attaining the liminf for some fixed realization of  $\boldsymbol{\mu}$ . By plugging in  $\nu = \mu$  in (32), all the inequalities become equalities, and in particular,

$$\lim_n W_{p, \text{OT}_p}^p(\delta_{\mu_n}, P_n) = \lim_n \mathbb{E} \text{OT}_p^p(\mu_n, \boldsymbol{\mu}_n) = \mathbb{E} \text{OT}_p^p(\mu, \boldsymbol{\mu}) = W_{p, \text{OT}_p}^p(\delta_\mu, P).$$

This yields

$$\begin{aligned}
0 &\leq W_{p, \text{OT}_p}(\delta_{\mu_n}, P) - W_{p, \text{OT}_p}(\delta_\mu, P) \\
&\leq W_{p, \text{OT}_p}(\delta_{\mu_n}, P_n) + W_{p, \text{OT}_p}(P_n, P) - W_{p, \text{OT}_p}(\delta_\mu, P) \rightarrow 0
\end{aligned}$$

as  $n$  goes to  $+\infty$ , i.e.  $\lim_n W_{p, \text{OT}_p}(\delta_{\mu_n}, P) = W_{p, \text{OT}_p}(\delta_\mu, P)$ . Therefore,

$$\begin{aligned}
\mathbb{E} \text{OT}_p^p(\mu, \boldsymbol{\mu}) &= W_{p, \text{OT}_p}^p(\delta_\mu, P) = \lim_n W_{p, \text{OT}_p}^p(\delta_{\mu_n}, P) = \lim_n \mathbb{E} \text{OT}_p^p(\mu_n, \boldsymbol{\mu}) \\
&\geq \mathbb{E} \liminf_n \text{OT}_p^p(\mu_n, \boldsymbol{\mu}) \text{ by Fatou's lemma} \\
&\geq \mathbb{E} \text{OT}_p^p(\mu, \boldsymbol{\mu}) \text{ by lower semi-continuity of } \text{OT}_p.
\end{aligned}$$

As  $\liminf_n \text{OT}_p^p(\mu_n, \boldsymbol{\mu}) \geq \text{OT}_p^p(\mu, \boldsymbol{\mu})$  and  $\mathbb{E} \liminf_n \text{OT}_p^p(\mu_n, \boldsymbol{\mu}) = \mathbb{E} \text{OT}_p^p(\mu, \boldsymbol{\mu})$ , we actually have  $\liminf_n \text{OT}_p^p(\mu_n, \boldsymbol{\mu}) = \text{OT}_p^p(\mu, \boldsymbol{\mu})$ , concluding the proof.  $\square$

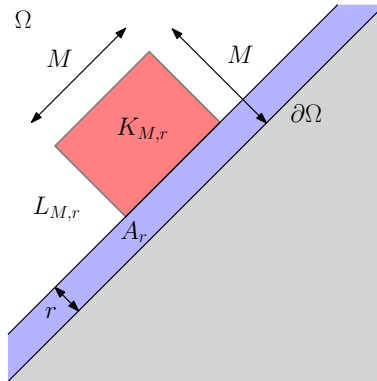


Figure 7: Partition of  $\Omega$  used in the proof of Lemma 4.3.

*Proof of Lemma 4.3.* 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  $\text{OT}_p(\mu_n, \nu) \rightarrow \text{OT}_p(\mu, \nu)$  for some  $\nu \in \mathcal{D}^p$ . The vague convergence of  $(\mu_n)_n$  implies that  $\mu^{(p)}$  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  $(\mu_n^{(p)})_n$  converges weakly to  $\mu^{(p)}$ , or equivalently by Proposition A.3 that  $\mu_n \xrightarrow{v} \mu$  and  $\text{Pers}_p(\mu_n) \rightarrow \text{Pers}_p(\mu)$ . 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 between  $\mu_n$  and  $\nu$ . The vague convergence of  $(\mu_n)_n$  implies that  $(\pi_n)_n$  is relatively compact with respect to the vague convergence on  $E_\Omega$ . Let  $\pi$  be a limit of any converging subsequence of  $(\pi_n)_n$ , which indexes are still denoted by  $n$ . One can prove that  $\pi \in \text{Opt}(\mu, \nu)$  (see [28, Prop. 2.3]). For  $r > 0$ , define  $A_r := \{x \in \Omega, d(x, \partial\Omega) \leq r\}$  and write  $\bar{A}_r$  for  $A_r \cup \partial\Omega$ . Consider  $\eta > 1$ . We can write

$$\begin{aligned}
\int_{A_r} d(x, \partial\Omega)^p d\mu_n(x) &= \iint_{A_r \times \bar{\Omega}} d(x, \partial\Omega)^p d\pi_n(x, y) \\
&= \iint_{A_r \times (\Omega \setminus A_{\eta r})} d(x, \partial\Omega)^p d\pi_n(x, y) + \iint_{\bar{A}_r \times \bar{A}_{\eta r}} d(x, \partial\Omega)^p d\pi_n(x, y) \\
&\stackrel{(*)}{\leq} \frac{1}{(\eta-1)^p} \iint_{A_r \times (\Omega \setminus A_{\eta r})} d(x, y)^p d\pi_n(x, y) + \iint_{\bar{A}_r \times \bar{A}_{\eta r}} d(x, \partial\Omega)^p d\pi_n(x, y) \\
&\leq \frac{1}{(\eta-1)^p} \text{OT}_p^p(\mu_n, \nu) + 2^{p-1} \left( \iint_{\bar{A}_r \times \bar{A}_{\eta r}} d(x, y)^p d\pi_n(x, y) + \iint_{\bar{A}_r \times \bar{A}_{\eta r}} d(y, \partial\Omega)^p d\pi_n(x, y) \right) \\
&\leq \frac{1}{(\eta-1)^p} \text{OT}_p^p(\mu_n, \nu) + 2^{p-1} \left( \text{OT}_p^p(\mu_n, \nu) - \iint_{E_\Omega \setminus (\bar{A}_r \times \bar{A}_{\eta r})} d(x, y)^p d\pi_n(x, y) + \int_{A_{\eta r}} d(y, \partial\Omega)^p d\nu(y) \right)
\end{aligned}$$

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

$$\begin{aligned}
\limsup_{n \rightarrow \infty} \int_{A_r} d(x, \partial\Omega)^p d\mu_n(x) &\leq \frac{1}{(\eta-1)^p} \text{OT}_p^p(\mu, \nu) + 2^{p-1} \left( \text{OT}_p^p(\mu, \nu) \right. \\
&\quad \left. - \iint_{E_\Omega \setminus (\bar{A}_r \times \bar{A}_{\eta r})} d(x, y)^p d\pi(x, y) + \int_{A_{\eta r}} d(y, \partial\Omega)^p d\nu(y) \right)
\end{aligned}$$

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

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

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

$$K_{M,r} := \{x \in \Omega \setminus A_r, d(x, \partial\Omega) < M, d(s(x), 0) < M\}$$

and  $L_{M,r}$  the closure of  $\Omega \setminus (A_r \cup K_{M,r})$  (see Figure 7). For  $r' > 0$ ,

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

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

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

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

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

Therefore,

$$\begin{aligned} \iint_{L_{M,r} \times K_{M/2,r'}} d(x, \partial\Omega)^p d\pi_n(x, y) &\leq 2^p \iint_{L_{M,r} \times K_{M/2,r'}} d(x, y)^p d\pi_n(x, y) \\ &\leq 2^p \iint_{L_{M,r} \times \bar{\Omega}} d(x, y)^p d\pi_n(x, y). \end{aligned}$$

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

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  $\int_{\Omega \setminus \overline{K_{M,r}}} d(x, \partial\Omega)^p d\mu_n = \mu_n^{(p)}(\Omega \setminus \overline{K_{M,r}})$  is uniformly small:  $(\mu_n^{(p)})_n$  is tight. As we have

$$\begin{aligned} \mu_n^{(p)}(\Omega) &= \text{Pers}_p(\mu_n) = \text{OT}_p^p(\mu_n, 0) \\ &\leq (\text{OT}_p(\mu_n, \nu) + \text{OT}_p(\nu, 0))^p \rightarrow (\text{OT}_p(\mu, \nu) + \text{OT}_p(\nu, 0))^p, \end{aligned}$$

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

*Proof of Lemma 4.4.* Let  $P = \sum_{i=1}^N \lambda_i \delta_{a_i}$  a probability distribution with  $a_i \in \mathcal{D}_f$  of mass  $m_i \in \mathbb{N}$ , and define  $m_{\text{tot}} = \sum_{i=1}^N m_i$ . By Proposition 4.1, every  $p$ -Fréchet mean  $a$  of  $P$  is in correspondence with a  $p$ -Fréchet mean for the Wasserstein distance  $\tilde{a}$  of  $\tilde{P} = \sum_{i=1}^N \lambda_i \delta_{\tilde{a}_i}$ , where  $\tilde{a}_i = a_i + (m_{\text{tot}} - m_i) \delta_{\partial\Omega}$ , with  $a$  being the restriction of  $\tilde{a}$  to  $\Omega$ .

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  $1 \leq i \leq N$ ,  $1 \leq j \leq m$ , with the  $x_{i,j}$ s non-necessarily distinct. Define

$$T : (x_1, \dots, x_N) \in \tilde{\Omega}^N \mapsto \arg \min \left\{ \sum_{i=1}^N \lambda_i \rho(x_i, y)^p, y \in \tilde{\Omega} \right\} \in \tilde{\Omega}. \quad (33)$$

Since we assume  $p > 1$ ,  $T$  is well-defined and is continuous (the minimizer is unique by strict convexity). Using the localization property stated in [12, Section 2.2], we know that the support of a  $p$ -Fréchet mean of  $\tilde{P}$  is included in the finite set

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

Let  $K = m^N$  and let  $z_1, \dots, z_K$  be an enumeration of the points of  $S$  (with potential repetitions). Denote by  $\text{Gr}(z_k)$  the  $N$  elements  $x_1, \dots, x_N$ , with  $x_i \in \text{spt}(\tilde{a}_i)$ , such that  $z_k = T(x_1, \dots, x_N)$ . It is explained in [12, §2.3], that finding a  $p$ -Fréchet mean of  $\tilde{P}$  is equivalent to finding a minimizer of the problem

$$\inf_{(\gamma_1, \dots, \gamma_N) \in \Pi} \sum_{i=1}^N \lambda_i \iint_{\tilde{\Omega}^2} \rho(x_i, y)^p d\gamma_i(x_i, y), \quad (34)$$

where  $\Pi$  is the set of plans  $(\gamma_i)_{i=1, \dots, N}$ , with  $\gamma_i$  having for first marginal  $\tilde{a}_i$ , and such that all  $\gamma_i$ s share the same (non-fixed) second marginal. Furthermore, we can assume without loss of generality that  $(\gamma_1 \dots \gamma_N)$  is supported on  $(\text{Gr}(z_k), z_k)_k$ , i.e. a point  $z_k$  in the  $p$ -Fréchet mean is necessary transported to its corresponding grouping  $\text{Gr}(z_k)$  by (optimal)  $\gamma_1, \dots, \gamma_N$  [12, §2.3]. For such a minimizer, the common second marginal is a  $p$ -Fréchet mean of  $\tilde{P}$ .

A potential minimizer of (34) is described by a vector  $\gamma = (\gamma_{i,j,k}) \in \mathbb{R}_+^{NmK}$  such that:

$$\begin{cases} \text{for } 1 \leq i \leq N, 1 \leq j \leq m, & \sum_{k=1}^K \gamma_{i,j,k} = 1 \text{ and} \\ \text{for } 2 \leq i \leq N, 1 \leq k \leq K, & \sum_{j=1}^m \gamma_{i,j,k} = \sum_{j=1}^m \gamma_{1,j,k}. \end{cases} \quad (35)$$

Let  $c \in \mathbb{R}^{NmK}$  be the vector defined by  $c_{i,j,k} = \mathbf{1}\{x_{i,j} \in \text{Gr}(z_k)\} \lambda_i \rho(x_{i,j}, z_k)^p$ . Then, the problem (34) is equivalent to

$$\underset{\gamma \in \mathbb{R}_+^{NmK}}{\text{minimize}} \gamma^T c \quad \text{under the constraints (35)}. \quad (36)$$



The set of  $p$ -Fréchet means of  $P$  are in bijection with the set of minimizers of this Linear Programming problem (see [56, §5.15]), which is given by a face of the polyhedron described by the equations (35). Hence, if we show that this polyhedron is integer (i.e. its vertices have integer values), then it would imply that the extreme points of the set of  $p$ -Fréchet means of  $P$  are point measures, concluding the proof. The constraints (35) 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 (35) if and only if  $A\gamma = b$ . A sufficient condition for the polyhedron  $\{Ax \leq b\}$  to be integer is to satisfy the following property (see [56, Section 5.17]): for all  $u \in \mathbb{Z}^{NmK}$ , the dual problem

$$\max\{y^T b, y \geq 0 \text{ and } y^T A = u\} \quad (37)$$

has either no solution (i.e. there is no  $y \geq 0$  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  $1 \leq i \leq N$ ,  $1 \leq j \leq m$  and  $y^1$  is indexed on  $2 \leq i \leq N$ ,  $1 \leq k \leq K$ . One can check that, for  $2 \leq i \leq N$ ,  $1 \leq j \leq m$ ,  $1 \leq k \leq K$ :

$$u_{1,j,k} = y_{1,j}^0 + \sum_{i'=2}^N y_{i',k}^1 \quad \text{and} \quad u_{i,j,k} = y_{i,j}^0 - y_{i,k}^1, \quad (38)$$

so that,

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

Therefore, the function  $y^T b$  is constant on the set  $P := \{y \geq 0, y^T A = u\}$ , 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  $y_{i,j}^0 = y_{i,k}^1 = \lambda_i$  for  $i \geq 2$  and  $y_{1,j}^0 = -\sum_{i=2}^N y_{i,k}^1 = -\sum_{i=2}^N \lambda_i$  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

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

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

## D Technical details regarding Section 5.3

Write  $\mathcal{M}_f$  for  $\mathcal{M}_f(\Omega)$  and define  $\mathcal{M}_\pm$  the space of finite signed measures on  $\Omega$ , 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 [7] is a generalization of the Lebesgue integral for functions taking their values in Banach space. We define the expected persistence measure of  $P \in \mathcal{W}^p(\mathcal{M}^p)$  as the Bochner integral of some pushforward of  $P$ . More precisely, recall the definition (18) of  $\mu^{(p)}$  and define

$$F : (\mathcal{M}^p, \text{OT}_p) \rightarrow (\mathcal{M}_\pm, |\cdot|) \\ \mu \mapsto \mu^{(p)}.$$

Note that  $F$  has an inverse  $G$  on  $\mathcal{M}_f$ , defined by  $G(\nu)(A) := \int_A d(x, \Omega)^{-p} d\nu(x)$  for  $A \subset \Omega$  a Borel set. Theorem 3.4 implies that  $G$  is a continuous function from  $(\mathcal{M}_f, |\cdot|)$  to  $(\mathcal{M}^p, \text{OT}_p)$ . In particular, as  $\mathcal{M}_f$  and  $\mathcal{M}^p$  are Polish spaces and  $G$  is injective, the map  $F$  is measurable (see [38, Theorem 15.1]). For  $P \in \mathcal{W}^p(\mathcal{M}^p(\Omega))$ , define for  $\boldsymbol{\mu} \sim P$ ,  $\mathbb{E}[\boldsymbol{\mu}]$  the *linear* expectation of  $P$  by

$$\mathbb{E}[\boldsymbol{\mu}] := G\left(\int \nu d(F_{\#}P)(\nu)\right) \in \mathcal{M}^p, \quad (39)$$

where the integral is the Bochner integral on the Banach space  $(\mathcal{M}_\pm, |\cdot|)$  and  $F_{\#}P$  is the pushforward of  $P$  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 5.4 consists in applying Jensen's inequality in an infinite-dimensional setting. We first show that the function  $\text{OT}_p^p$  is convex.

**Lemma D.1.** *For  $1 \leq p < \infty$ , the function  $\text{OT}_p^p : \mathcal{M}^p \times \mathcal{M}^p \rightarrow \mathbb{R}$  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

$$\text{OT}_p^p(t\mu_1 + (1-t)\mu_2, t\nu_1 + (1-t)\nu_2) \leq t\text{OT}_p^p(\mu_1, \nu_1) + (1-t)\text{OT}_p^p(\mu_2, \nu_2).$$

Let  $\pi_{11} \in \text{Opt}_p(\mu_1, \nu_1)$  and  $\pi_{22} \in \text{Opt}_p(\mu_2, \nu_2)$ . 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  $t\nu_1 + (1-t)\nu_2$ . The cost of this admissible plan is  $t\text{OT}_p^p(\mu_1, \nu_1) + (1-t)\text{OT}_p^p(\mu_2, \nu_2)$ , which is therefore larger than  $\text{OT}_p^p(t\mu_1 + (1-t)\mu_2, t\nu_1 + (1-t)\nu_2)$ .  $\square$

We then use the following result, which is a particular case of [51, Theorem 3.10].

**Proposition D.1.** *Let  $\mathcal{X}$  be a Hausdorff locally convex topological vector space and  $C \subset \mathcal{X}$  a closed convex set. Let  $Q$  be a probability measure on  $\mathcal{X}$  endowed with its borelian  $\sigma$ -algebra, which is supported on  $C$ . Assume that  $\int \|x\| dQ(x) < \infty$ . Let  $f : C \rightarrow [0, \infty)$  be a continuous convex function with  $\int f(x) dQ(x) < \infty$ . Then*

$$f\left(\int x dQ(x)\right) \leq \int f(x) dQ(x).$$

Let  $\mathcal{X} = \mathcal{M}_\pm \times \mathcal{M}_\pm$  which is a Banach space (endowed with the product norm), and thus in particular a Hausdorff locally convex topological vector space. Let  $C = \mathcal{M}_f \times \mathcal{M}_f$ , which is convex and closed (closedness follows immediately from the definition of the total variation  $|\cdot|$ ) and let  $f = \text{OT}_p^p \circ (G, G) : \mathcal{X} \rightarrow \mathbb{R}$ . The continuity of  $G$  implies that  $f$  is continuous and Lemma D.1 implies the convexity of  $f$ . Let  $P, P'$  be two probability measures in  $\mathcal{W}^p(\mathcal{M}^p)$  and  $\gamma$  be an optimal coupling between  $P$  and  $P'$ . We let  $Q$  be the image measure of  $\gamma$  by  $(F, F)$ , so that

$$\begin{aligned} \int_{x \in \mathcal{X}} \|x\| dQ(x) &= \int_{\mu, \mu' \in \mathcal{M}^p} \max(|\mu^{(p)}|, |(\mu')^{(p)}|) d\gamma(\mu, \mu') \\ &\leq \int_{\mu} \text{Pers}_p(\mu) dP(\mu) + \int_{\mu'} \text{Pers}_p(\mu') dP'(\mu') < \infty \end{aligned}$$

and that

$$\int_{x \in \mathcal{X}} f(x) dQ(x) = \int_{\mu, \mu' \in \mathcal{M}^p} \text{OT}_p^p(\mu, \mu') d\gamma(\mu, \mu') = W_{p, \text{OT}_p}^p(P, P') < \infty.$$

Also, we have

$$\begin{aligned} \int x dQ(x) &= \int_{\nu, \nu' \in \mathcal{M}^p} (\nu, \nu') d(F, F)_\# \gamma(\nu, \nu') \\ &= \left( \int_{\nu \in \mathcal{M}^p} \nu dF_\# P(\nu), \int_{\nu' \in \mathcal{M}^p} \nu' dF_\# P'(\nu') \right), \end{aligned}$$

so that by (39),  $f(\int x dQ(x)) = \text{OT}_p^p(\mathbb{E}[\boldsymbol{\mu}], \mathbb{E}[\boldsymbol{\mu}'])$ , where  $\boldsymbol{\mu} \sim P$  and  $\boldsymbol{\mu}' \sim P'$ . Proposition 5.4 yields the conclusion.