



HAL
open science

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. 2019. hal-01970466v2

HAL Id: hal-01970466

<https://inria.hal.science/hal-01970466v2>

Preprint submitted on 29 Apr 2019 (v2), last revised 17 Jul 2020 (v3)

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 Persistence Diagram Space via Optimal Partial Transport

Vincent Divol, 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 measure space, and by observing that its metrics can be expressed as solutions of 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 study the 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 transport metrics, the existence of Fréchet means for any distribution of diagrams, and an exhaustive description of continuous linear representations of persistence diagrams. We also showcase the usefulness 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 [43], material science [30, 37], shape analysis [12, 56], 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) 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 [25] for an introduction. The space of persistence diagrams, denoted by \mathcal{D} in the following, is usually equipped with partial matching metrics d_p , sometimes called Wasserstein distances [25, Chapter VIII.2]: for $p \in [1, +\infty)$ and a, b in \mathcal{D} , define

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

where $\|\cdot\|$ denotes the Euclidean distance, $\Gamma(a, b)$ 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 Ω , namely the diagonal (see Figure 1). When $p \rightarrow \infty$, we recover the so-called *bottleneck distance*:

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

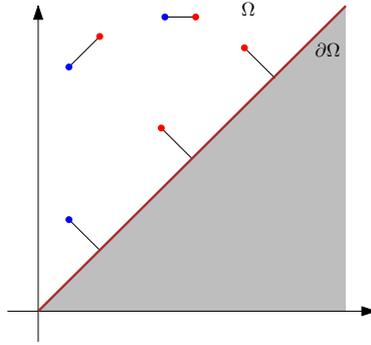


Figure 1: An example of optimal partial matching between two diagrams. The bottleneck distance between these two diagrams is the length of the longest edge in this matching, while their Wasserstein distance d_p is the p -th root of the sum of all edge lengths to the power p .

An equivalent viewpoint, developed in [13, 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 Ω with integer mass on each point of its support. This measure-based perspective suggests to consider more general Radon measures¹ supported on the upper half-plane Ω . 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 *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

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

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 3.1). This result allows to characterize a large class of continuous representations (Prop. 5.1) that includes some standard tools used in TDA such as the Betti curve [57], the persistence surface [1] and the persistence silhouette [16].

- (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 [28, 24] 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 [28] that the sequence of measures $n^{-1}a_n$ converges vaguely to some limit measure μ supported on Ω (that is *not* a persistence diagram), and in [24] that the moments of $n^{-1}a_n$ also converge to the moments of μ . An interesting problem is to build a metric which generalizes d_p and for which the convergence of $n^{-1}a_n$ to μ holds.
- (A3) **Stability of the expected diagrams.** Of particular interests 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 Φ , 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 [14]. Understanding how the object $\mathbb{E}_P[\mathbf{a}]$ depends on the underlying process P generating \mathbf{a} invites 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 Ω . Therefore, extending the distances d_p to Radon measures in a consistent way will allow to assess the closeness between those quantities.

Remark 1.1. *Note that, throughout this article, we consider persistence diagrams with possibly infinitely many points (but still locally finite). This is motivated from a statistical perspective. Indeed, the space of finite persistence diagram is lacking completeness, as highlighted in [44, Definition 2]. Furthermore, infinite persistence diagrams can be helpful to describe natural phenomena: for instance, the (random) persistence diagram built on the sub-level sets of a Brownian motion has infinitely many points (see [14, Section 6]). Finite persistence diagrams, useful for numerical applications, are studied in Section 3.2.*

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 Ω , 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 persistence² of $\mu \in \mathcal{M}$ as

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

where $d(x, \partial\Omega)$ 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 OT_p (see Definition 2.1), originally introduced in a work of Figalli and Gigli [27]. We show in Proposition 3.2 that OT_p and d_p coincide on $\mathcal{D}^p := \mathcal{D} \cap \mathcal{M}^p$, making 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 μ , denoted by $\mu_n \xrightarrow{v} \mu$, if for any continuous compactly supported function $f : \Omega \rightarrow \mathbb{R}$, $\mu_n(f) \rightarrow \mu(f)$. We prove the following equivalence between convergence for the metric OT_p and the vague convergence:

Theorem 3.1. *Let $1 \leq p < \infty$. Let μ, μ_1, μ_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 in the Wasserstein space (see [59, Theorem 6.9]). A proof for Radon measures supported on a common bounded set can be found in [27, Proposition 2.7]. Our contribution consists in extending this result to non-bounded sets, in particular to the upper half plane Ω .

Section 3.2 is dedicated to sets of measures with finite masses, appearing naturally in numerical applications. We show in particular that the problem of computing the 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. 3.6), a result having practical implications for the computation of OT_p distances between persistence measures (and diagrams).

²The total persistence of a diagram is usually defined with $d(x, \partial\Omega)$ being replaced by $t_2 - t_1$ where $x = (t_1, t_2) \in \Omega$. This would only lead to a multiplication by a factor $\sqrt{2}^p$.

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 [7, Theorem 4.20]), but is also of crucial interest in TDA as it is motivated by algebraic considerations (see [45]) and satisfies stronger stability results [19] than its $p < \infty$ counterparts (see [20]). In particular, we give in Propositions 3.10 and 3.12 a characterization of bottleneck convergence (in the vein of Theorem 3.1) 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 [44, 55], where authors prove their existence for certain types of distributions [44, Theorem 28]. Using the framework of persistence measures and existing works on Fréchet means in optimal transport, we show that this existence result is actually true for any distribution of persistence diagrams with finite moment. By doing so, we also prove a similar result for persistence measures. Namely, we prove the following:

Theorem 4.1. *For any probability distribution P supported on \mathcal{M}^p with finite p -th moment, the set of Fréchet means of P is not empty.*

Theorem 4.2. *If P is supported on \mathcal{D}^p and has a finite p -th moment, then P admits a Fréchet mean 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 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 works [38, Proposition 3.2] and [24, 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 [31, 24] along with Theorem 3.1, 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 ξ, ξ' be two probability measures on some compact d -*

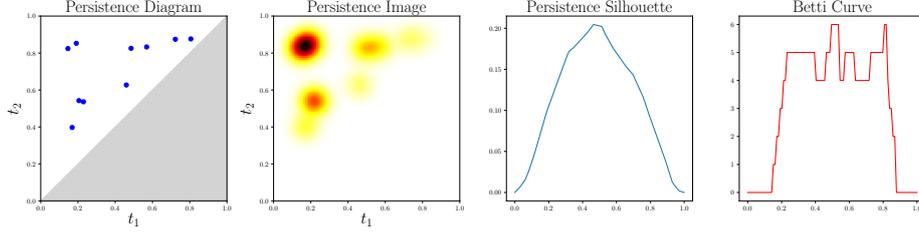


Figure 2: Some common linear representations of persistence diagrams. From left to right: A persistence diagram. Its persistence surface [1], which is a persistence measure. The corresponding persistence silhouette [16]. The corresponding Betti Curve [57]. See Section 5.1 for details. Computations are made using the `sklearn-tda` library: https://github.com/MathieuCarriere/sklearn_tda.

dimensional Riemannian manifold (\mathbb{X}, ρ) . Let \mathbb{X}_n (resp. \mathbb{X}'_n) be a n -sample of law ξ (resp. ξ'). Then, for any $k > d$, and any $p \geq k + 1$,

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

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:

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

1.3 Notations

We use the following notations throughout the article:

- $\Omega := \{(t_1, t_2) \in \mathbb{R}^2, t_2 > t_1\}$ denotes the upper half-plane and $\partial\Omega := \{(t, t), t \in \mathbb{R}\}$ denotes its boundary, namely the diagonal. $\bar{\Omega} := \Omega \cup \partial\Omega$. Ω is endowed with the Euclidean distance, denoted either by d or by $\|\cdot\|$ in the following. For $x \in \Omega$, $d(x, \partial\Omega) := \|x - s(x)\|$, where s denotes the orthogonal projection of x onto the diagonal $\partial\Omega$.
- E_Ω denotes the set $\bar{\Omega} \times \bar{\Omega} \setminus (\partial\Omega \times \partial\Omega)$.
- \mathcal{M} denotes the space of Radon measures supported on Ω . \mathcal{M}_f denotes the subspace of \mathcal{M} of measures μ satisfying $\mu(\Omega) < \infty$. $\mathcal{M}_{\leq m}$ denotes the subset of \mathcal{M} of measures satisfying $\mu(\Omega) \leq m$ (for $m > 0$), and \mathcal{M}_m is the subset of \mathcal{M} of measures satisfying $\mu(\Omega) = m$. For $1 \leq p < \infty$, \mathcal{M}^p denotes the subspace of \mathcal{M} of measures μ with finite p -persistence, that is $\text{Pers}_p(\mu) := \int_\Omega d(x, \partial\Omega)^p d\mu(x) < \infty$.
- $\mathcal{M}^\infty \subset \mathcal{M}$ denotes the subset of \mathcal{M} of measures μ such that the μ -essential supremum of map $x \mapsto d(x, \partial\Omega)$ is finite, that is $\sup\{d(x, \partial\Omega), x \in \text{spt}(\mu)\} < \infty$, where $\text{spt}(\mu)$ is the support of the measure μ .
- \mathcal{D} denotes the space of persistence diagrams, namely the subset of \mathcal{M} of measures of the form $\sum_{x \in X} n_x \delta_x$, where $X \subset \Omega$ is a locally finite set and $n_x \in \mathbb{N}$. We define $\mathcal{D}_f := \mathcal{D} \cap \mathcal{M}_f$, $\mathcal{D}^p := \mathcal{M}^p \cap \mathcal{D}$, $\mathcal{D}^\infty := \mathcal{M}^\infty \cap \mathcal{D}$, etc.

- $W_{p,d}$ (or W_p when there is no ambiguity) denotes the Wasserstein distance between probability measures supported on a Polish metric space (\mathcal{X}, d) . $\mathcal{W}^p(\mathcal{X})$ denotes the space of probability measures supported on \mathcal{X} with finite p -th moment, equipped with the metric $W_{p,d}$.

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} [58, 59, 49], that is—up to a renormalization factor—non-negative measures of the same mass. Intuitively, optimal transport can be described in the following informal way: given a pile of sand and a hole of the same volume, how to fill the hole such that we minimize the distance traveled while transporting the sand? This turns up to find an optimal way to match each (infinitesimal) element of the initial pile of sand to a final position in the hole. Formally, given two probability measures μ, ν supported on (\mathcal{X}, d) , the *Wasserstein- p* distance ($p \geq 1$) induced by the metric d between μ and ν 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 μ and ν , that is the set of measures on $\mathcal{X} \times \mathcal{X}$ which have respective marginals μ , and ν . When there is no ambiguity on the distance d used, we simply write W_p instead of $W_{p,d}$. In order to have W_p finite, μ and ν are required to have a finite p -th moment, that is there exists $x_0 \in \mathcal{X}$ 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 matching points. However, the constraint on masses appearing in optimal transport³—while persistence diagrams can have different (even infinite) masses—makes a crucial difference between the W_p and d_p metrics, and appeals for an extension of optimal transport to a wider class of measures.

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 [26, 18, 36]. As it handles the case of measures with infinite masses, the work of Figalli and Gigli [27], is of particular interest for us. 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 $\overline{\mathcal{X}} := \mathcal{X} \sqcup \partial\mathcal{X}$). Informally, they

³Note that if μ and ν do not have the same masses, the set $\Pi(\mu, \nu)$ is empty.

propose to use the boundary $\partial\mathcal{X}$ as an infinite reservoir of mass. To do that, they relax the marginal constraints, proposing to match an element of mass $\mu(dx)$ either to some $\nu(dy)$, with cost $d(x, y)^p$, or to its projection $s(x)$ onto the boundary $\partial\mathcal{X}$, with cost $d(x, s(x))^p$ (and similarly for ν). Formally, they introduce the following problem:

Definition 2.1. [27, Problem 1.1] Let $p \in [1, +\infty)$. Let μ, ν 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 π on $\overline{\mathcal{X}} \times \overline{\mathcal{X}}$ satisfying for all Borel sets $A, B \subset \mathcal{X}$,

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

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

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

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 μ and ν 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 $A, B \subset \overline{\mathcal{X}}$,

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

π 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 π satisfies $\pi(\partial\mathcal{X} \times \partial\mathcal{X}) = 0$, so that measures $\pi \in \text{Adm}(\mu, \nu)$ are supported on

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

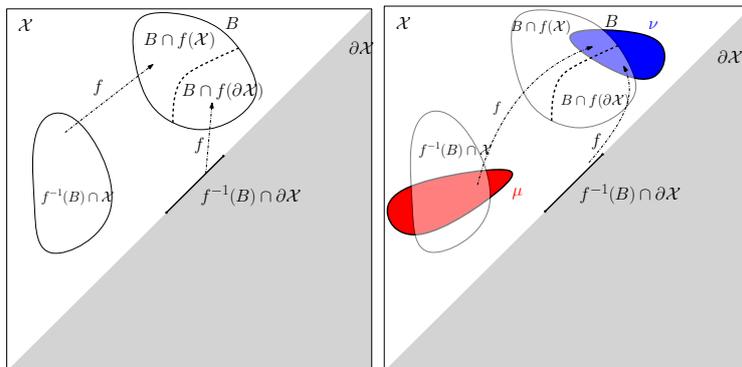


Figure 3: A transport map f must satisfy that the mass $\nu(B)$ (light blue) is the sum of the mass $\mu(f^{-1}(B) \cap \mathcal{X})$ given by μ that is transported by f onto B (light red) and the mass $\nu(B \cap f(\partial\mathcal{X}))$ coming from $\partial\mathcal{X}$ and transported by f onto B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

3 Structure of the persistence measures and diagrams spaces

This section is dedicated to general properties of \mathcal{M}^p . Most of the results in Section 3.1 are inspired from the ones of Figalli and Gigli in [27], which are stated for a bounded subset \mathcal{X} of \mathbb{R}^d . Our goal is to state properties of the space Ω , which is of course *not bounded*, and also to study the case $p = \infty$ (Section 3.3). Adapting the results of [27] 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.

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). Our first proposition states preliminary results on the problem stated in Definition 2.1.

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

These properties are mentioned in [27, pages 4-5] in the bounded case, and adapt straightforwardly to our framework. For the sake of completeness, we provide a detailed proof in Appendix B.

Remark 3.1. *If a (Borel) measure μ satisfies $\text{Pers}_p(\mu) < \infty$, 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, μ is automatically a Radon measure.

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

Lemma 3.1. *Let $\mu \in \mathcal{M}^p$. Fix $r > 0$, and let $A_r := \{x \in \Omega, d(x, \partial\Omega) \leq r\}$. Let $\mu^{(r)}$ be the restriction of μ to $\Omega \setminus A_r$. 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 π 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 μ 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 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 [41, 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 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 [27, 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 3.1. *Let μ, μ_1, μ_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 [59, Theorem 6.9]) and can be found in [27, Proposition 2.7] in the case where the ground space is bounded. While the proof of the direct implication can be easily adapted from [27] (it can be found in Appendix B), a new proof is needed for the converse implication.

Proof of the converse implication. Let $\mu, \mu_1, \mu_2 \dots$ 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 Ω . 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_Ω is included in some set of the form $(K \cup \Omega) \times (\Omega \cup K)$, for $K \subset \Omega$ compact, using Proposition A.1 again, it follows that $(\pi_n)_n$ is also relatively compact for the vague convergence.

Let thus π be the limit of any converging subsequence of $(\pi_n)_n$, which indexes are still denoted by n . As $\mu_n \xrightarrow{v} \mu$, π is necessarily in $\text{Opt}_p(\mu, \mu)$ (see [27, Prop. 2.3]), i.e. π 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 $K \subset \Omega$, whose complementary set in Ω is denoted by K^c , its interior set is denoted by $\overset{\circ}{K}$, and its boundary by ∂K we have

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

Recall that $\text{Pers}_p(\mu) < \infty$. Therefore, for $\varepsilon > 0$, there exists some compact set $K \subset \Omega$, with

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

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

onto the diagonal, while everything else is unchanged):

$$\begin{cases} \tilde{\pi}_n = \pi_n & \text{on } K^2 \sqcup (K^c)^2, \\ \tilde{\pi}_n = 0 & \text{on } K \times K^c \sqcup K^c \times K, \\ \tilde{\pi}_n(A \times B) = \pi_n(A \times (K^c \sqcup B)) & \text{for } A \subset K, B \subset \partial\Omega, \\ \tilde{\pi}_n(A \times B) = \pi_n(A \times (K \sqcup B)) & \text{for } A \subset K^c, B \subset \partial\Omega, \\ \tilde{\pi}_n(A \times B) = \pi_n((K^c \sqcup A) \times B) & \text{for } A \subset \partial\Omega, B \subset K, \\ \tilde{\pi}_n(A \times B) = \pi_n((K \sqcup A) \times B) & \text{for } A \subset \partial\Omega, B \subset K^c. \end{cases} \quad (17)$$

Note that $\tilde{\pi}_n \in \text{Adm}(\mu_n, \mu)$: for instance, for $A \subset K$ a Borel set, $\tilde{\pi}_n(A \times \bar{\Omega}) = \pi_n(A \times K) + \pi_n(A \times (K^c \sqcup \partial\Omega)) = \pi_n(A \times \bar{\Omega}) = \mu_n(A)$, 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 show that each of them converges to 0. Assume without loss of generality that the compact set K belongs to an increasing sequence $(K_m)_m$ of compact sets of Ω so that $\bigcup_m K_m = \Omega$ and $\pi(\partial(K_m \times K_m)) = 0$ for all m .

- We have $\iint_{K^2} d(x, y)^p d\tilde{\pi}_n(x, y) = \iint_{K^2} d(x, y)^p d\pi_n(x, y)$. The lim sup of the integral is smaller than $\iint_{K^2} d(x, y)^p d\pi(x, y)$ by Portmanteau theorem (applied to the sequence $(d(x, y)^p d\pi_n(x, y))_n$), and, recalling that π is supported on the diagonal of E_Ω , 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 [27, 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 lim sup in n , and then letting K goes to Ω , this quantity converges to 0 by (16).

- We have

$$\iint_{K \times \partial\Omega} d(x, \partial\Omega)^p d\tilde{\pi}_n(x, y) = \int_K d(x, \partial\Omega)^p d\mu_n(x) - \iint_{K^2} d(x, \partial\Omega)^p d\pi_n(x, y).$$

By Portmanteau theorem applied to the sequence $(d(x, \partial\Omega)^p d\mu_n(x))_n$, the lim sup of the first term is smaller than $\int_K d(x, \partial\Omega)^p d\mu(x)$. Applying once again Portmanteau theorem on the second term to the sequence $(d(x, y)^p d\pi_n(x, y))_n$, and using that π is supported on the diagonal of E_Ω , the limsup of the second term is smaller than $-\iint_{K^2} d(x, \partial\Omega)^p d\pi(x, y) = -\int_K d(x, \partial\Omega)^p d\mu(x)$ (recall that $\pi(\partial(K \times K)) = 0$). Therefore, the lim sup 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 3.2. *The assumption $\text{Pers}_p(\mu_n) \rightarrow \text{Pers}_p(\mu)$ is crucial to obtain 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 (as it is constant equal to $\frac{\sqrt{2}}{2}$), while $\text{OT}_p(\mu_n, 0) \not\rightarrow 0$. This does not contradict Theorem 3.1 since $\text{Pers}_p(\mu) = 0 \neq \frac{\sqrt{2}}{2} = \lim_n \text{Pers}_p(\mu_n)$.*

Theorem 3.1 implies some useful results. First, it turns out that the topology of the metric 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 convergence).

Corollary 3.1. *\mathcal{D}^p is closed in \mathcal{M}^p for the metric 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 [44, Theorems 7 and 12] with a different approach.

Secondly, we show that the vague convergence of μ_n to μ 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 measure with finite mass μ^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 3.1, 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, $\mu_n \xrightarrow{v} \mu$ is equivalent to $\mu_n^p \xrightarrow{v} \mu^p$. By Proposition A.3, the vague convergence along with the convergence of the masses is also 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 3.4. *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 OT_p is equivalent to the relative compactness of the set $\{\mu^p, \mu \in F\}$ for the weak convergence. Recall that all μ^p have a finite mass, as $\mu^p(\Omega) = \text{Pers}_p(\mu) < \infty$. Therefore, one can use Prokhorov's theorem (Proposition A.2) to conclude. \square

Remark 3.3. *This characterization is equivalent to the one described in [44, 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.*

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 [40, 11], and the specific properties of 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 3.5. \mathcal{M}_f^p (resp. \mathcal{D}_f) is dense in \mathcal{M}^p (resp. \mathcal{D}^p) for the metric OT_p .

Proof. This is a straightforward consequence of Lemma 3.1. \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 ρ is a distance on $\tilde{\Omega}$ and that $(\tilde{\Omega}, \rho)$ is a Polish space. One can then define the Wasserstein distance $W_{p, \rho}$ with respect to ρ for finite measures on $\tilde{\Omega}$ which have the same masses, that is the infimum of $\tilde{C}_p(\tilde{\pi}) := \iint_{\tilde{\Omega}^2} \rho(x, y)^p d\tilde{\pi}(x, y)$, for $\tilde{\pi}$ a transport plan with corresponding marginals (see Section 2.1). The following theorem states that the problem of computing the 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.

Proposition 3.6. 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 $\text{OT}_p(\mu, \nu) = W_{p, \rho}(\tilde{\mu}, \tilde{\nu})$.

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

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

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

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

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

Proof.

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

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

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

□

We show that $\text{OT}_p(\mu, \nu) \leq W_{p,\rho}(\tilde{\mu}, \tilde{\nu})$ 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 $\tilde{\mu} := \mu + (r - \mu(\Omega))\delta_{\partial\Omega}$, $\tilde{\nu} := \nu + (r - |\nu|)\delta_{\partial\Omega}$. 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

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

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

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

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

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

Since d is continuous, the infimum in the right hand side of (21) is reached [59, Theorem 4.1]. Consider thus $\tilde{\pi} \in \Pi(\tilde{\mu}, \tilde{\nu})$ which realizes the infimum. We can write, using Lemma 3.2,

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

which concludes the proof. □

We can now prove the theorem. Let $\pi \in T(\mu, \nu)$. As $\mu(\Omega) + \pi(\partial\Omega \times \Omega) \leq \mu(\Omega) + \nu(\Omega) \leq r$, one can define $\tilde{\pi} = \iota(\pi)$. Since $\rho(x, y) \leq d(x, y)$, we have $\tilde{C}_p(\tilde{\pi}) \leq \iint d(x, y)^p d\tilde{\pi}(x, y) = C_p(\pi)$ (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 3.4. *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 [25, Ch. VIII.4] and used in practical computations [34]. Here, Proposition 3.6 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 3.5. *Proposition 3.6 is useful for numerical purpose since it allows 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 [46] to compute Wasserstein distances. This alternative to the combinatorial algorithms considered in the literature [34, 55] is studied in details in [41]. This result is also helpful to prove the existence of Fréchet means of sets of persistence measures (see Section 4).*

3.3 The OT_∞ distance

In classical optimal transport, the ∞ -Wasserstein distance is known to have a much more erratic behavior than its $p < \infty$ counterparts [49, Section 5.5.1]. However, in the framework of persistence diagrams, the d_∞ distance defined in Eq. (2) appears naturally as an interleaving distance between persistence modules and satisfies strong stability results: it is thus worth of interest. It also happens that, when restricted to diagrams having some specific finiteness properties, most irregular behaviors are suppressed and a convenient characterization of convergence exists.

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

Proposition 3.7. *Let $\mu, \nu \in \mathcal{M}^\infty$. For the vague topology on E_Ω ,*

- *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.*
- *OT_∞ is lower semi-continuous.*

Moreover, OT_∞ is a distance on \mathcal{M}^∞ .

The proofs of these results are found in Appendix B (as it is mixed with the proof of Proposition 3.1 for $p < \infty$). As in the case $p < \infty$, we prove that OT_∞ and d_∞ coincide on \mathcal{D}^∞ .

Proposition 3.8. 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 (potentially infinite) sets of indices. The marginals constraints imply that a plan $\pi \in \text{Adm}(\mu, \nu)$ is supported on $(\{x_i\}_i \cup \partial\Omega) \times (\{y_j\}_j \cup \partial\Omega)$, and the cost of such a plan can always be reduced if some of the mass $\pi(\{x_i\}, \partial\Omega)$ (resp. $\pi(\partial\Omega, \{y_j\})$) is sent on another point than the projection of x_i (resp. y_j) on the diagonal $\partial\Omega$. Introduce the matrix C indexed on $(-J \cup I) \times (-I \cup J)$ defined by

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

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 marginals 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 [29, 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 3.9. The space $(\mathcal{M}^\infty, \text{OT}_\infty)$ is complete.

Proof. Let $(\mu_n)_n$ be a Cauchy sequence for OT_∞ . 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 for the vague convergence. Up to extracting a subsequence, let μ be a measure such that $\mu_n \xrightarrow{v} \mu$.

By the lower semi-continuity of OT_∞ ,

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

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

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

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

Remark 3.6. *Contrary to the case $p < \infty$, the space \mathcal{D}^∞ (and therefore \mathcal{M}^∞) 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 [7, Theorem 4.20].*

We now show that the direct implication in Theorem 3.1 still holds in the case $p = \infty$.

Proposition 3.10. *Let μ, μ_1, μ_2, \dots be measures in \mathcal{M}^∞ . If $\text{OT}_\infty(\mu_n, \mu) \rightarrow 0$, then $(\mu_n)_n$ converges vaguely to μ .*

Proof. Let $f \in C_c(\Omega)$, whose support is included in some compact set K . For any $\varepsilon > 0$, there exists a L -Lipschitz function f_ε , whose support is included in K , with $\|f - f_\varepsilon\|_\infty \leq \varepsilon$. Observe that $\sup_k \mu_k(K) < \infty$ using the same arguments than for (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,

$$\begin{aligned} |\mu_n(f_\varepsilon) - \mu(f_\varepsilon)| &\leq \iint_{\bar{\Omega}^2} |f_\varepsilon(x) - f_\varepsilon(y)| d\pi_n(x, y) \\ &\leq L \iint_{(K \times \bar{\Omega}) \cup (\bar{\Omega} \times K)} d(x, y) d\pi_n(x, y) \\ &\leq LC_\infty(\pi_n)(\pi_n(K \times \bar{\Omega}) + \pi_n(\bar{\Omega} \times K)) \\ &\leq L \text{OT}_\infty(\mu_n, \mu) \left(\sup_k \mu_k(K) + \mu(K) \right) \rightarrow 0. \end{aligned}$$

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

Remark 3.7. *As for the case $1 \leq p < \infty$, Proposition 3.10 implies that OT_∞ metricizes the vague convergence, and thus using Proposition 3.8, we have that $(\mathcal{D}^\infty, d_\infty)$ is closed in $(\mathcal{M}^\infty, \text{OT}_\infty)$ and is—in particular—complete.*

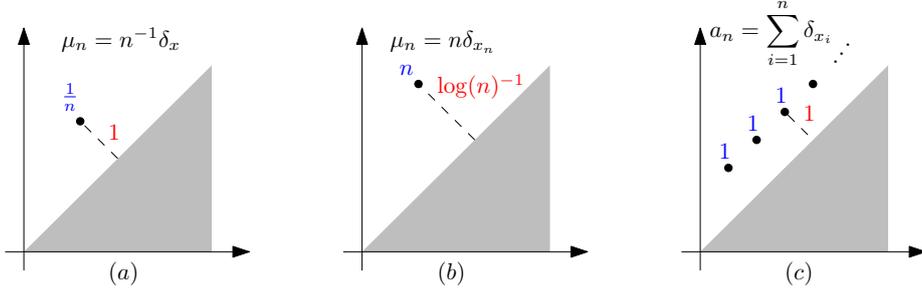


Figure 4: Illustration of differences between OT_p , OT_∞ , and vague convergences. Blue color represents the mass on a point while red color designates distances. (a) A case where $\text{OT}_p(\mu_n, 0) \rightarrow 0$ for any $p < \infty$ while $\text{OT}_\infty(\mu_n, 0) = 1$. (b) A case where $\text{OT}_\infty(\mu_n, 0) \rightarrow 0$ 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 OT_∞ .

Contrary to the $p < \infty$ case, a converse of Proposition 3.10 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 \mathcal{D}_0^∞ 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 3.11. *The closure of \mathcal{D}_f for the distance OT_∞ is \mathcal{D}_0^∞ .*

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^∞ 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 OT_∞ metric of any sequence in \mathcal{D}_f . \square

Remark 3.8. *The space \mathcal{D}_0^∞ is exactly the set introduced in [4, Theorem 3.5] as the completion of \mathcal{D}_f for the bottleneck metric d_∞ . Here, we recover that \mathcal{D}_0^∞ is complete as a closed subset of the complete space \mathcal{D}^∞ .*

For $r > 0$ and $a \in \mathcal{D}$, let $a^{(r)}$ denote the restriction of a to $\{x \in \Omega, d(x, \partial\Omega) > r\}$ (as in Lemma 3.1). The following characterization of convergence holds in \mathcal{D}_0^∞ .

Proposition 3.12. *Let a, a_1, a_2, \dots be persistence diagrams in \mathcal{D}_0^∞ . 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 3.10 states that the convergence with respect to OT_∞ implies the vague convergence. Fix $r > 0$.

By definition, $a^{(r)}$ is made of a finite number of points, all included in some open bounded set $U \subset \Omega$. As $a_n^{(r)}(U^c)$ is a sequence of integers, the bottleneck convergence implies that for n large enough, $a_n^{(r)}(U^c)$ is equal to 0. Thus, $(a_n^{(r)})_n$ is tight.

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

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

so that, for n large enough, there are exactly m_k points of a_n in $B(x_k, \varepsilon)$ (since the sequence $(a_n(B(x_k, \varepsilon)))_n$ 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 smaller than $r + \varepsilon$. 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 Fréchet means for distributions supported on \mathcal{M}^p

In this section, we state the existence of 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 persistence diagrams, showing that the Fréchet mean of a probability distribution on \mathcal{D}^p —that exists in \mathcal{M}^p —also exists in \mathcal{D}^p . We suppose that $p > 1$ (the case $p = 1$ would correspond to a notion of *median*).

Recall that $(\mathcal{M}^p, \text{OT}_p)$ is Polish, and let W_p denote the Wasserstein distance between probability measures supported on $(\mathcal{M}^p, \text{OT}_p)$ (see Section 2.1). 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 metric, which are at a finite distance from δ_0 —the dirac mass supported on the empty diagram—i.e. $W_p(P, \delta_0) = \int_{\nu \in \mathcal{M}^p} \text{OT}_p^p(\nu, 0) dP(\nu) = \int_{\nu \in \mathcal{M}^p} \text{Pers}_p(\nu) dP(\nu) < \infty$.

Definition 4.1. Consider $P \in \mathcal{W}^p(\mathcal{M}^p)$. A measure $\mu^* \in \mathcal{M}^p$ is a 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 Fréchet means in the finite case

Let P be of the form $\sum_{i=1}^N \lambda_i \delta_{\mu_i}$ with $N \in \mathbb{N}$, μ_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 Fréchet means for such a P , we show that, in this case, Fréchet means correspond to Fréchet mean 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_{tot} (see Section 3.2), a problem well studied in the literature [2, 8, 10].

We start with a lemma which affirms that if a measure μ has too much mass (larger than m_{tot}), then it cannot be a 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 μ has some mass that is mapped to the diagonal in each transport plan between μ and μ_i , then we can build a measure μ' by “removing” this mass, and then observe that such a measure μ' has a smaller energy.

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

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

and, for $i = 1, \dots, N$, a measure π'_i , equal to π_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*

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

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

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

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

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

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

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

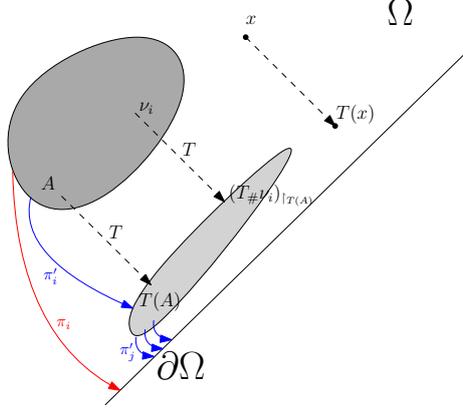


Figure 5: Global picture of the proof. The main idea is to observe that the cost induced by π_i (red) is strictly larger than the sum of costs induced by the π'_i 's (blue), which leads to a strictly better energy.

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

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

and

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

Using π'_i instead of π_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 smaller 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 Ψ is a bijection. Also, if μ 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. 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 $\mathcal{M}_{\leq m_{\text{tot}}}^p$, and thus a Fréchet mean of P according to Lemma 4.1. We therefore have proved the existence of Fréchet means in the finite case.

4.2 Existence and consistency of Fréchet means

To extend the result of the previous section to Fréchet means of infinitely many measures, we adapt the work of Le Gouic and Loubes [42] to probability distributions supported on \mathcal{M}^p . Our Proposition 4.2 corresponds to the Theorem 3 in [42] (consistency of barycenters), and our Theorem 4.1 corresponds to the Theorem 2 in [42] (existence of barycenters).

Proposition 4.2. *Let P_n, P be probability measures in $\mathcal{W}^p(\mathcal{M}^p)$. Assume that each P_n has a Fréchet mean μ_n and that $W_p(P_n, P) \rightarrow 0$. 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 Fréchet mean of P .*

We follow the same “sketch of proof” of Theorem 3 in [42], namely:

- Proving relative compactness of the sequence $(\mu_n)_n$ (for the vague convergence).
- Proving that any accumulation point is a Fréchet mean of P .
- Observing that the convergence actually holds for the OT_p metric.

Proof. In order to prove relative compactness of $(\mu_n)_n$, we use the characterization stated in Proposition A.1. Consider a compact set $K \subset \Omega$. We have, because of (14),

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

Since μ_n is a Fréchet mean of P_n , it minimizes $\{W_p(\delta_\nu, P_n), \nu \in \mathcal{M}^p\}$, and in particular $W_p(\delta_{\mu_n}, P_n) \leq W_p(\delta_0, P_n)$. Furthermore, since we assume that $W_p(P_n, P) \rightarrow 0$, we have in particular that $\sup_n W_p(P_n, \delta_0) < \infty$. 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.

At that point, the exact same computations as the **proof of Theorem 3** in [42] show the existence of a subsequence $\mu_{n_k} \xrightarrow{v} \mu$ where μ is a Fréchet mean of P and of some $\nu \in \mathcal{M}^p$ such that $\text{OT}_p(\mu_{n_k}, \nu) \rightarrow \text{OT}_p(\mu, \nu)$. In order to get the desired conclusion, we use the following lemma:

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

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

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

Theorem 4.1. *Let $P \in \mathcal{W}^p(\mathcal{M}^p)$. The set of Fréchet means of P is not empty.*

Proof. Let $P = \sum_{i=1}^N \lambda_i \mu_i$ be a probability measure on \mathcal{M}^p with finite support μ_1, \dots, μ_N (but no assumption on finiteness of masses). According to Proposition 3.5, 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 Fréchet means. Furthermore, $W_p^p(P^{(n)}, P) \leq \sum_i \lambda_i \text{OT}_p^p(\mu_i^{(n)}, \mu_i)$ so that this quantity converges to 0 as $n \rightarrow \infty$. It follows from Proposition 4.2 that P admits a Fréchet mean.

If P has infinite support, following [42], it can be approximated (in W_p) by an empirical probability measure $P_n = \frac{1}{n} \sum_{i=1}^n \delta_{\mu_i}$ where the μ_i are i.i.d. from P . We know that P_n admits a Fréchet mean since its support is finite, and thus, applying Proposition 4.2 once again, we obtain that P admits a Fréchet mean. \square

4.3 Fréchet means in \mathcal{D}^p

We now prove the existence of Fréchet means in \mathcal{D}^p for any distribution of persistence diagram (i.e. probability distribution supported on \mathcal{D}^p), extending the results of [44], in which authors prove their existence for specific probability distributions (namely distributions with compact support or specific rates of decay). The following theorem asserts two different things: that $\arg \min\{\mathcal{E}(a), a \in \mathcal{D}^p\}$ is non empty, and that $\min\{\mathcal{E}(a), a \in \mathcal{D}^p\} = \min\{\mathcal{E}(\mu), \mu \in \mathcal{M}^p\}$, i.e a persistence measure cannot perform strictly better than an optimal persistence diagram when averaging diagrams. As for Fréchet means in \mathcal{M}^p , we start with the finite case.

Lemma 4.3. *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, P admits a Fréchet mean in \mathcal{D}_f .*

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

Theorem 4.2. *Consider a distribution P supported on \mathcal{D}^p with finite p -th moment. There exists a Fréchet mean of P which is in \mathcal{D}^p .*

Proof. The proof is the same as the one of Theorem 4.1, using additionally the fact that \mathcal{D}^p is closed in \mathcal{M}^p (Proposition A.5) to conclude that P has a Fréchet mean in \mathcal{D}^p . \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 an application $\Phi : \mathcal{M}^p \rightarrow \mathcal{B}$ 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 Φ 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 then prove that $\mu \mapsto \mu(g(\cdot)d(\cdot, \partial\Omega)^p)$ has some regularity properties (continuity, stability, etc.). Applying Theorem 3.1, we show that this approach always gives a continuous linear representation, and that it is the only way to do.

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 OT_p if and only if $f \in \mathcal{C}_{b,p}^0$.*

Proof. Let $f \in \mathcal{C}_{b,p}^0$ and $\mu, \mu_1, \mu_2 \dots \in \mathcal{M}^p$ be such that $\text{OT}_p(\mu_n, \mu) \rightarrow 0$. Using Corollary 3.2, it 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. Φ is continuous with respect to OT_p .

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 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 $q = 1$ 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 Φ 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, 17, 39, 48].
- *Persistence silhouettes.* Let $\Lambda(x, t) = \max\left(\frac{x_1-x_2}{2} - \left|t - \frac{x_1+x_2}{2}\right|, 0\right)$ 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 OT_p . This representation is called the persistence silhouette, and was introduced in [16]. 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 $p, q \geq 1$, and define $f : x \in \Omega \mapsto (t \mapsto d(x, \partial\Omega)^{p-1/q} \mathbf{1}\{x \in B_t\})$. Then $f(x) \in L_q(\mathbb{R})$ with $\|f(x)\|_q$ proportional to $d(x, \partial\Omega)^p$. The corresponding function Φ is the weighted Betti curve, which takes its values in the Banach space $L_q(\mathbb{R})$ (equipped with the norm $\|\cdot\|_q$). 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. Heuristically, such regularity is generally obtained when working with $p = 1$, see for example [1, Theorem 2], [11, Theorem 3.3 & 3.4], [53, §4], [48, Theorem 2].

Here, we still consider the case of linear representations, and show that stability always holds with respect to the distance OT_1 . Informally, this is explained by the fact that when $p = 1$, the cost function $(x, y) \mapsto d(x, y)^p$ is an actual distance.

Proposition 5.2. *Define \mathcal{L} the set of Lipschitz continuous functions $f : \overline{\Omega} \rightarrow \mathbb{R}$ with Lipschitz constant less than 1 and that satisfy $f(\partial\Omega) = 0$. Let $T \subset \mathbb{R}$, 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 $\pi \in \text{Opt}_1(\mu, \nu)$ an optimal transport plan. Let $t \in T$. We have, using that f_t vanishes on $\partial\Omega$:

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

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

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 [59, Theorem 5.10] and [49, 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 [27] 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 [5, 24, 31, 28, 51] 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 Duy, Hiraoka and Shirai [31], Divol and Polonik [24] 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, it is shown in [31, Theorem 1.5] that the rescaled diagram

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

which is a persistence measure, almost surely converges vaguely to some non-degenerate Radon measure μ , and Divol and Polonik [24, Theorem 1] show that μ has moments of all orders and that for all $p \geq 1$, the total persistence $\text{Pers}_p(\mu_n)$ also almost surely converges to $\text{Pers}_p(\mu)$. The formalism developed here, and more specifically Theorem 3.1, 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 μ . However, in the case $d = 1$, authors in [24] show that if the X_i are n i.i.d realizations of a random variable X admitting a density κ supported on $[0, 1]$, bounded from below and above by positive constants, then (the ordinate of) μ —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) μ 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 close form using Proposition 3.6 and the fact that computing the distance W_p in dimension $d = 1$ is particularly easy (see [49, Chapter 2]). See Figure 6 for an illustration.

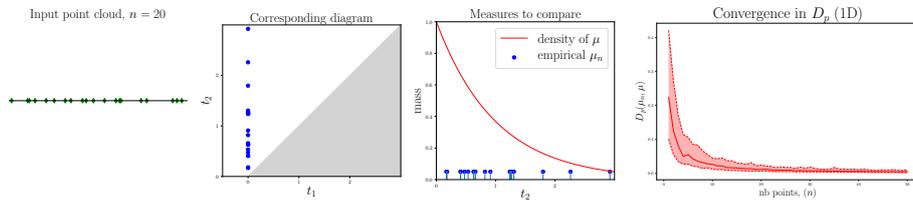


Figure 6: Numerical illustration of D_p convergence in the one dimensional case. From left to right: an example of input point cloud with $n = 20$ points ; the corresponding persistence diagram ; the (ordinate of) the two measures we compare: μ_n is the empirical one. Most right: median $\text{OT}_p(\mu_n,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

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 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 [14], 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 OT_p .

Proposition 5.4. *Let $P, P' \in \mathcal{W}^p(\mathcal{M}^p)$. Then, for any coupling $\pi \in \Pi(P, P')$ between P and P' , we have $\text{OT}_p^p(\mathbb{E}[\boldsymbol{\mu}], \mathbb{E}[\boldsymbol{\mu}']) \leq \mathbb{E}_\pi[\text{OT}_p^p(\boldsymbol{\mu}, \boldsymbol{\mu}')$.*

The proof is postponed to Appendix D.

Using stability results on the distances $d_p (= \text{OT}_p)$ between persistence diagrams [20], 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 ξ, ξ' be two probability measures on some compact d -dimensional Riemannian manifold (\mathbb{X}, ρ) . Let \mathbb{X}_n (resp. \mathbb{X}'_n) be a n -sample of law ξ (resp. ξ'). Then, for any $k > d$, and any $p \geq k + 1$,*

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

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 π be any coupling between \mathbb{X}_n a n -sample of law ξ , and \mathbb{X}'_n a n -sample of law ξ' . According to Proposition 5.4,

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

It is stated in [20, 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 π , we obtain

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

where $W_{H,p}$ is the p -Wasserstein distance between probability distributions on compact sets of the manifold \mathbb{X} , endowed with the Hausdorff distance. Lemma 15 of [15] states that

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

concluding the proof. \square

Note that this proposition illustrates the usefulness of introducing new distances 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 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 the formalism presented in this article will help to develop new approaches in TDA and give a better understanding of the statistical behavior of topological descriptors. The new notions and techniques introduced here raise some new questions:

- **Kernel between measures.** Section 5.1 considers the stability of linear representations of diagrams in a Banach space. Aside linear representations, a common approach to turn persistence diagrams into a machine-learning

friendly framework is to use a *kernel* to perform the embedding [11, 39, 39]. Kernels between measures are well studied from both a theoretical and practical point of view [22, 23, 3, 54] and it is likely that some results and methods could be transposed to persistence diagrams and measures. The *sliced-Wasserstein kernel* [47, 35] adapted for persistence diagrams in [11] is such an example.

- **Approximating persistence diagrams with persistence measures.**

In order to perform statistical analysis, it can be natural to consider optimization problems over \mathcal{D} . For instance, given a sample of diagrams $a_1 \dots a_N$, a standard statistical descriptor would be its Fréchet means (or barycenters), defined as the minimizers of $\mathcal{E} : a \mapsto \sum_{i=1}^N d_p(a, a_i)$. However, optimizing in \mathcal{D} is challenging and it can be useful to relax the problem, optimizing on the larger set of Radon measures instead. As an example, authors in [55] provide an algorithm to minimize \mathcal{E} on \mathcal{D} . Their optimization scheme is however non-convex, and does not scale numerically as the numbers of diagram and of points per diagrams increase. To improve on this, following [21, 52], authors in [41] observe that when extended to a larger space of measures, the functional \mathcal{E} becomes convex. Thus, they propose to minimize, introducing a smoothing parameter $\gamma > 0$ (the smaller γ , the better the approximation), an approximation \mathcal{E}^γ of \mathcal{E} over \mathcal{M} instead, that is also convex. The output μ^γ is no longer a persistence diagram, but a Radon measure. A natural question is then to quantify how close μ^γ is from a given persistence diagram, in particular as $\gamma \rightarrow 0$ (in the vein of [9, Prop. 2.18]). More generally, measuring the distance between a persistence diagram and a more general (e.g. continuous) distribution is a natural problem and the formalism developed in this article gives tools to address it.

- **An algebraic construction of persistence measures.**

In this article, persistence measures arise from the statistical or numerical analysis of a set of persistence diagrams (averaging sample of diagrams, relaxing diagrams, etc.). For instance, using the formalism of Section 5.3, we can “define” the persistence measure of a random point process to be the expected persistence diagram. One should however recall that persistence diagrams are representations of *persistence modules* where the presence of a point (t_1, t_2) in the persistence diagrams corresponds to the presence of an interval module $\mathbb{I}_{(t_1, t_2)}$ in the (decomposition of the) corresponding persistence module \mathbb{M} , that is $\mathbb{M} = \bigoplus_{(t_1, t_2) \in A} \mathbb{I}_{(t_1, t_2)}$ where A is a locally finite subset of Ω . As such, one can wonder if it would be possible to define an algebraic construction that would generalize the decomposition theorem of persistence modules and such that its natural representation would be a persistence measure.

References

- [1] Henry Adams, Tegan Emerson, Michael Kirby, Rachel Neville, Chris Peterson, Patrick Shipman, Sofya Chepushtanova, Eric Hanson, Francis Motta, and Lori Ziegelmeier. Persistence images: a stable vector representation of persistent homology. *Journal of Machine Learning Research*, 18(8):1–35, 2017.
- [2] Martial Agueh and Guillaume Carlier. Barycenters in the wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011.
- [3] Francis R Bach. Graph kernels between point clouds. In *Proceedings of the 25th international conference on Machine learning*, pages 25–32. ACM, 2008.
- [4] Andrew J Blumberg, Itamar Gal, Michael A Mandell, and Matthew Pancia. Robust statistics, hypothesis testing, and confidence intervals for persistent homology on metric measure spaces. *Foundations of Computational Mathematics*, 14(4):745–789, 2014.
- [5] Omer Bobrowski, Matthew Kahle, Primoz Skraba, et al. Maximally persistent cycles in random geometric complexes. *The Annals of Applied Probability*, 27(4):2032–2060, 2017.
- [6] Salomon Bochner. Integration von funktionen, deren werte die elemente eines vektorraumes sind. *Fundamenta Mathematicae*, 20(1):262–176, 1933.
- [7] Peter Bubenik and Tane Vergili. Topological spaces of persistence modules and their properties. *Journal of Applied and Computational Topology*, pages 1–37, 2018.
- [8] G. Carlier and I. Ekeland. Matching for teams. *Economic Theory*, 42(2):397–418, 2010.
- [9] Guillaume Carlier, Vincent Duval, Gabriel Peyré, and Bernhard Schmitzer. Convergence of entropic schemes for optimal transport and gradient flows. *SIAM Journal on Mathematical Analysis*, 49(2):1385–1418, 2017.
- [10] Guillaume Carlier, Adam Oberman, and Edouard Oudet. Numerical methods for matching for teams and wasserstein barycenters. *ESAIM: Mathematical Modelling and Numerical Analysis*, 49(6):1621–1642, 2015.
- [11] Mathieu Carrière, Marco Cuturi, and Steve Oudot. Sliced wasserstein kernel for persistence diagrams. In *34th International Conference on Machine Learning*, 2017.
- [12] Mathieu Carrière, Steve Y. Oudot, and Maks Ovsjanikov. Stable topological signatures for points on 3d shapes. *Computer Graphics Forum*, 34(5):1–12, 2015.
- [13] Frédéric Chazal, Vin De Silva, Marc Glisse, and Steve Oudot. *The structure and stability of persistence modules*. Springer, 2016.

- [14] Frédéric Chazal and Vincent Divol. The Density of Expected Persistence Diagrams and its Kernel Based Estimation. In Bettina Speckmann and Csaba D. Tóth, editors, *34th International Symposium on Computational Geometry (SoCG 2018)*, volume 99 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 26:1–26:15, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
- [15] Frédéric Chazal, Brittany Fasy, Fabrizio Lecci, Bertrand Michel, Alessandro Rinaldo, and Larry Wasserman. Subsampling methods for persistent homology. In *International Conference on Machine Learning*, pages 2143–2151, 2015.
- [16] Frédéric Chazal, Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, and Larry Wasserman. Stochastic convergence of persistence landscapes and silhouettes. In *Proceedings of the thirtieth annual symposium on Computational geometry*, page 474. ACM, 2014.
- [17] Yen-Chi Chen, Daren Wang, Alessandro Rinaldo, and Larry Wasserman. Statistical analysis of persistence intensity functions. *arXiv preprint arXiv:1510.02502*, 2015.
- [18] Lenaïc Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. Unbalanced optimal transport: geometry and kantorovich formulation. *arXiv preprint arXiv:1508.05216*, 2015.
- [19] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete & Computational Geometry*, 37(1):103–120, 2007.
- [20] David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Yuriy Mileyko. Lipschitz functions have l_p -stable persistence. *Foundations of computational mathematics*, 10(2):127–139, 2010.
- [21] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, pages 2292–2300, 2013.
- [22] Marco Cuturi, Kenji Fukumizu, and Jean-Philippe Vert. Semigroup kernels on measures. *Journal of Machine Learning Research*, 6(Jul):1169–1198, 2005.
- [23] Marco Cuturi and Jean-Philippe Vert. Semigroup kernels on finite sets. In *Advances in neural information processing systems*, pages 329–336, 2005.
- [24] Vincent Divol and Wolfgang Polonik. On the choice of weight functions for linear representations of persistence diagrams. *arXiv preprint arXiv:1807.03678*, 2018.
- [25] Herbert Edelsbrunner and John Harer. *Computational topology: an introduction*. American Mathematical Soc., 2010.
- [26] Alessio Figalli. The optimal partial transport problem. *Archive for rational mechanics and analysis*, 195(2):533–560, 2010.

- [27] Alessio Figalli and Nicola Gigli. A new transportation distance between non-negative measures, with applications to gradients flows with dirichlet boundary conditions. *Journal de mathématiques pures et appliquées*, 94(2):107–130, 2010.
- [28] Akshay Goel, Khanh Duy Trinh, and Kenkichi Tsunoda. Asymptotic behavior of betti numbers of random geometric complexes. *arXiv preprint arXiv:1805.05032*, 2018.
- [29] Marshall Hall. *Combinatorial theory (2nd ed)*. John Wiley & Sons, 1986.
- [30] Yasuaki Hiraoka, Takenobu Nakamura, Akihiko Hirata, Emerson G. Escolar, Kaname Matsue, and Yasumasa Nishiura. Hierarchical structures of amorphous solids characterized by persistent homology. *Proceedings of the National Academy of Sciences*, 2016.
- [31] Yasuaki Hiraoka, Tomoyuki Shirai, Khanh Duy Trinh, et al. Limit theorems for persistence diagrams. *The Annals of Applied Probability*, 28(5):2740–2780, 2018.
- [32] Olav Kallenberg. *Random measures, theory and applications*. Springer, 2017.
- [33] A.S. Kechris. *Classical Descriptive Set Theory*. Graduate Texts in Mathematics. Springer-Verlag, 1995.
- [34] Michael Kerber, Dmitriy Morozov, and Arnur Nigmatov. Geometry helps to compare persistence diagrams. *Journal of Experimental Algorithmics (JEA)*, 22(1):1–4, 2017.
- [35] Soheil Kolouri, Yang Zou, and Gustavo K Rohde. Sliced wasserstein kernels for probability distributions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5258–5267, 2016.
- [36] Stanislav Kondratyev, Léonard Monsaingeon, Dmitry Vorotnikov, et al. A new optimal transport distance on the space of finite radon measures. *Advances in Differential Equations*, 21(11/12):1117–1164, 2016.
- [37] M. Kramar, A. Goulet, L. Kondic, and K. Mischaikow. Persistence of force networks in compressed granular media. *Phys. Rev. E*, 87:042207, Apr 2013.
- [38] Genki Kusano, Kenji Fukumizu, and Yasuaki Hiraoka. Kernel method for persistence diagrams via kernel embedding and weight factor. *The Journal of Machine Learning Research*, 18(1):6947–6987, 2017.
- [39] Genki Kusano, Yasuaki Hiraoka, and Kenji Fukumizu. Persistence weighted gaussian kernel for topological data analysis. In *International Conference on Machine Learning*, pages 2004–2013, 2016.
- [40] Roland Kwitt, Stefan Huber, Marc Niethammer, Weili Lin, and Ulrich Bauer. Statistical topological data analysis—a kernel perspective. In *Advances in neural information processing systems*, pages 3070–3078, 2015.

- [41] Théo Lacombe, Marco Cuturi, and Steve Oudot. Large scale computation of means and clusters for persistence diagrams using optimal transport. In *Advances in Neural Information Processing Systems*, 2018.
- [42] Thibaut Le Gouic and Jean-Michel Loubes. Existence and consistency of wasserstein barycenters. *Probability Theory and Related Fields*, pages 1–17, 2016.
- [43] Chunyuan Li, Maks Ovsjanikov, and Frederic Chazal. Persistence-based structural recognition. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.
- [44] Yuriy Mileyko, Sayan Mukherjee, and John Harer. Probability measures on the space of persistence diagrams. *Inverse Problems*, 27(12):124007, 2011.
- [45] Steve Y Oudot. *Persistence theory: from quiver representations to data analysis*, volume 209. American Mathematical Society, 2015.
- [46] Gabriel Peyré and Marco Cuturi. *Computational Optimal Transport*. Number 2017-86. December 2017.
- [47] Julien Rabin, Gabriel Peyré, Julie Delon, and Marc Bernot. Wasserstein barycenter and its application to texture mixing. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 435–446. Springer, 2011.
- [48] Jan Reininghaus, Stefan Huber, Ulrich Bauer, and Roland Kwitt. A stable multi-scale kernel for topological machine learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4741–4748, 2015.
- [49] Filippo Santambrogio. *Optimal transport for applied mathematicians*. Birkhäuser, NY, 2015.
- [50] Alexander Schrijver. *Combinatorial optimization: polyhedra and efficiency*, volume 24. Springer Science & Business Media, 2003.
- [51] Benjamin Schweinhart. Weighted persistent homology sums of random čech complexes. *arXiv preprint arXiv:1807.07054*, 2018.
- [52] Justin Solomon, Fernando De Goes, Gabriel Peyré, Marco Cuturi, Adrian Butscher, Andy Nguyen, Tao Du, and Leonidas Guibas. Convolutional wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Transactions on Graphics (TOG)*, 34(4):66, 2015.
- [53] Anirudh Som, Kowshik Thopalli, Karthikeyan Natesan Ramamurthy, Vinay Venkataraman, Ankita Shukla, and Pavan Turaga. Perturbation robust representations of topological persistence diagrams. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 617–635, 2018.
- [54] Bharath Sriperumbudur, Kenji Fukumizu, and Gert Lanckriet. On the relation between universality, characteristic kernels and rkhs embedding of measures. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 773–780, 2010.

- [55] Katharine Turner, Yuriy Mileyko, Sayan Mukherjee, and John Harer. Fréchet means for distributions of persistence diagrams. *Discrete & Computational Geometry*, 52(1):44–70, 2014.
- [56] Katharine Turner, Sayan Mukherjee, and Doug M. Boyer. Persistent homology transform for modeling shapes and surfaces. *Information and Inference: A Journal of the IMA*, 3(4):310–344, 2014.
- [57] Yuhei Umeda. Time series classification via topological data analysis. *Information and Media Technologies*, 12:228–239, 2017.
- [58] Cédric Villani. *Topics in optimal transportation*. Number 58. American Mathematical Soc., 2003.
- [59] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.
- [60] August M Zapala et al. Jensens inequality for conditional expectations in banach spaces. *Real Analysis Exchange*, 26(2):541–552, 2000.

A Elements of measure theory

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

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

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

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

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

The following propositions are standard results. Corresponding proofs can be found for instance in [32, Section 15.7].

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

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

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

Proposition A.3. *Let μ, μ_1, μ_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 (Portmanteau theorem). *Let μ, μ_1, μ_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 Ω is the subset $\mathcal{D}(\Omega) \subset \mathcal{M}(\Omega)$ of Radon measures with discrete support and integer mass on each point, that is of the form*

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

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

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

B Delayed proofs of Section 3

For the sake of completeness, we present in this section proofs which either require very few adaptations from corresponding proofs in [27] or which are close to standard proofs in optimal transport theory.

Proofs of Proposition 3.1 and Proposition 3.7.

- For $\pi \in \text{Adm}(\mu, \nu)$ supported on E_Ω , and for any compact sets $K, K' \subset \Omega$, one has $\pi((K \times \overline{\Omega}) \cup (\overline{\Omega} \times K')) \leq \mu(K) + \nu(K') < \infty$. As any compact subset of E_Ω is included in a set of the form $(K \times \overline{\Omega}) \cup (\overline{\Omega} \times K')$, Proposition A.1 implies that $\text{Adm}(\mu, \nu)$ is relatively compact for the vague convergence on E_Ω . 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 π are still μ and ν , implying that $\text{Adm}(\mu, \nu)$ is closed in $\mathcal{M}(E_\Omega)$: it is therefore sequentially compact.
- To prove the second point of Proposition 3.1, consider π, π_1, π_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 $\pi' : A \mapsto \iint_A d(x, y)^p d\pi$. Portmanteau theorem (Proposition A.4) applied with the open set E_Ω to the measures $\pi'_n \xrightarrow{v} \pi'$ implies that $C_p(\pi) = \pi'(E_\Omega) \leq \liminf_n \pi'_n(E_\Omega) = \liminf_n C_p(\pi_n)$, i.e. C_p is lower semi-continuous.
- We now prove the lower semi-continuity of C_∞ . Let $(\pi_n)_n$ be a sequence converging vaguely to π on E_Ω and let $r > \liminf_{n \rightarrow \infty} C_\infty(\pi_n)$. The set $U_r = \{(x, y) \in E_\Omega, d(x, y) > r\}$ is open. By Portmanteau theorem (Proposition A.4), we have $0 = \liminf_{n \rightarrow \infty} \pi_n(U_r) \geq \pi(U_r)$. 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_Ω and that 2. OT_p is lower semi-continuous.
 1. Let $(\pi_n)_n$ be a minimizing sequence of (9) in $\text{Adm}(\mu, \nu)$. As $\text{Adm}(\mu, \nu)$ is sequentially compact, it has an adherence value π , and the lower semi-continuity implies that $C_p(\pi) \leq \liminf_{n \rightarrow \infty} C_p(\pi_n) = \text{OT}_p^p(\mu, \nu)$, 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 smaller than (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 $\nu_n \xrightarrow{v} \nu$. Up to taking a subsequence, one may assume that $\text{OT}_p(\mu_n, \nu_n)$ converges (eventually to infinity). Consider $\pi_n \in \text{Opt}_p(\mu_n, \nu_n)$. For any compact sets $K, K' \subset \Omega$, one has $\pi_n((K \times \overline{\Omega}) \cup (\overline{\Omega} \times K')) \leq \sup_n \mu_n(K) + \sup_n \nu_n(K') < \infty$. Therefore, by Proposition A.1, there exists a subsequence $(\pi_{n_k})_k$ which converges vaguely to some $\pi \in \text{Adm}(\mu, \nu)$. As the marginals of π_{n_k} converges to μ and ν , $\pi \in \text{Adm}(\mu, \nu)$. 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 OT_p is a metric on \mathcal{M}^p . Let $\mu, \nu, \lambda \in \mathcal{M}^p$. The symmetry of 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 [27, Lemma 2.1]: for $\pi_{12} \in \text{Opt}_p(\mu, \nu)$ and $\pi_{23} \in \text{Opt}_p(\nu, \lambda)$ 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_Ω , is equal to π_{12} (resp. π_{23}), and induces a zero cost on $\partial\Omega \times \partial\Omega$. Therefore,

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

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

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

1. Since $\text{Pers}_p(\mu) < \infty$, there exists a compact $K \subset \Omega$ such that $\text{Pers}_p(\mu) - \text{Pers}_p(\mu_0) < \varepsilon^p$, where μ_0 is the restriction of μ to K . By considering the transport plan between μ and μ_0 induced by the identity map on K and the projection onto the diagonal on $\bar{\Omega} \setminus K$, 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 μ_0 and μ_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 μ_2 and μ_1 that fully transports μ_2 onto μ_1 , and transport the remaining mass in μ_1 onto the diagonal. Then,

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

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 $(\text{Pers}_p(\mu_n))_n = (\text{OT}_p^p(\mu_n, 0))_n$ 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 convergence on Ω : consider $(\mu_{n_k})_k$ a subsequence converging vaguely on Ω to some measure μ . By the lower semi-continuity of 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,

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

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

Proof of the direct implication of Theorem 3.1. Let μ, μ_1, μ_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_ε , with Lipschitz constant L and whose support is included in K , with the ∞ -norm $\|f - f_\varepsilon\|_\infty$ smaller than ε . The convergence of $\text{Pers}_p(\mu_n)$ 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,

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

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

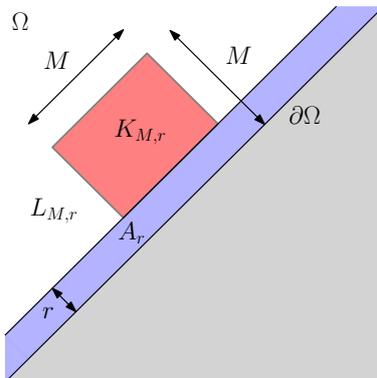


Figure 7: Partition of Ω used in the proof of Lemma 4.2.

C Proofs of the technical lemmas of Section 4

Proof of Lemma 4.2. For the direct implication, take $\nu = 0$ and apply Theorem 3.1.

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 μ^p is the only possible accumulation point for weak convergence of the sequence $(\mu_n^p)_n$. Therefore, it is sufficient to show that the sequence $(\mu_n^p)_n$ is relatively compact for weak convergence (i.e. tight and bounded in total variation, see Proposition A.2). Indeed, this would mean that (μ_n^p) converges weakly to μ^p , or equivalently by Proposition A.3 that $\mu_n \xrightarrow{v} \mu$ and $\text{Pers}_p(\mu_n) \rightarrow \text{Pers}_p(\mu)$. The conclusion is then obtained thanks to Theorem 3.1.

Thus, let $(\mu_n)_n$ be any subsequence and $(\pi_n)_n$ be corresponding optimal transport plans between μ_n and ν . The vague convergence of $(\mu_n)_n$ implies that $(\pi_n)_n$ is relatively compact with respect to the vague convergence on E_Ω . Let π be a limit of any converging subsequence of $(\pi_n)_n$, which indexes are still denoted by n . One can prove that $\pi \in \text{Opt}_p(\mu, \nu)$ (see [27, 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

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

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 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 η 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

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

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

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

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

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

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,

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

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

Finally, we have shown, that by taking r small enough and M large enough, one can find a compact set $\overline{K_{M,r}}$ such that $\int_{\Omega \setminus \overline{K_{M,r}}} d(x, \partial\Omega)^p d\mu_n = \mu_n^p(\Omega \setminus \overline{K_{M,r}})$ 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 4.3. 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 $m_{\text{tot}} = \sum_{i=1}^N m_i$. To prove the theorem, it suffices to show that $\tilde{P} = \sum_{i=1}^N \lambda_i \delta_{\tilde{a}_i}$, where $\tilde{a}_i = a_i + (m_{\text{tot}} - m_i) \delta_{\partial\Omega}$ has a Fréchet mean for the Wasserstein distance \tilde{a} which is a finite point measure supported on $\tilde{\Omega}$ (of mass m_{tot}). Indeed, taking the restriction of \tilde{a} to Ω will then give a Fréchet mean of P according to Proposition 4.1.

Let thus fix $m \in \mathbb{N}$, and let $\tilde{a}_1, \dots, \tilde{a}_N$ be point measures of mass m in $\tilde{\Omega}$. Write $\tilde{a}_i = \sum_{j=1}^m \delta_{x_{i,j}}$, so that $x_{i,j} \in \tilde{\Omega}$ for $1 \leq i \leq N$, $1 \leq j \leq m$, with the $x_{i,j}$ s non-necessarily distinct. Define

$$T : (x_1, \dots, x_N) \in \tilde{\Omega}^N \mapsto \arg \min \left\{ \sum_{i=1}^N \lambda_i \rho(x_i, y)^p, y \in \tilde{\Omega} \right\} \in \tilde{\Omega}. \quad (32)$$

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 [10, Section 2.2], we know that the support of a Fréchet mean 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 eventual 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 [10, §2.3], that finding a Fréchet mean of \tilde{P} is equivalent to finding a minimizer of the problem

$$\inf_{(\gamma_1, \dots, \gamma_N) \in \Pi} \sum_{i=1}^N \lambda_i \iint_{\tilde{\Omega}^2} \rho(x_i, y)^p d\gamma_i(x_i, y), \quad (33)$$

where Π is the set of plans $(\gamma_i)_{i=1, \dots, N}$, with γ_i having for first marginal \tilde{a}_i , and such that all γ_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 Fréchet mean is necessary transported to its corresponding grouping $\text{Gr}(z_k)$ by (optimal) $\gamma_1, \dots, \gamma_N$ [10, §2.3]. For such a minimizer, the common second marginal is a Fréchet mean of \tilde{P} . Therefore, if we show that there exists a minimizer of (33) which is a point measure, then the lemma is proven.

A potential minimizer of (33) is described by a vector $\gamma = (\gamma_{i,j,k}) \in \mathbb{R}_+^{NmK}$ such that:

$$\begin{cases} \text{for } 1 \leq i \leq N, 1 \leq j \leq m, & \sum_{k=1}^K \gamma_{i,j,k} = 1 \text{ and} \\ \text{for } 1 \leq i \leq N, 1 \leq k \leq K, & \sum_{j=1}^m \gamma_{i,j,k} = \sum_{j=1}^m \gamma_{i,j,k}. \end{cases} \quad (34)$$

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 (33) is equivalent to

$$\underset{\gamma \in \mathbb{R}_+^{NmK}}{\text{minimize}} \gamma^T c \quad \text{under the constraints (34)}.$$

This Linear Programming problem (see [50, §5.15]) has an integer solution if the polyhedron described by the equations (34) is integer (i.e. its vertices have integer values). The constraints (34) 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 (34) 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 [50, Section 5.17]): for all $u \in \mathbb{Z}^{NmK}$, the dual problem

$$\max\{y^T b, y \geq 0 \text{ and } y^T A = u\} \quad (35)$$

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 $2 \leq i \leq N$, $1 \leq j \leq m$, $1 \leq k \leq K$:

$$u_{1,j,k} = y_{1,j}^0 + \sum_{i'=2}^N y_{i',k}^1 \quad \text{and} \quad u_{i,j,k} = y_{i,j}^0 - y_{i,k}^1, \quad (36)$$

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

Define \mathcal{M}_\pm the space of signed measures on Ω , i.e. a measure $\mu \in \mathcal{M}_\pm$ is written $\mu_+ - \mu_-$ for two finite measures $\mu_+, \mu_- \in \mathcal{M}_f$. The total variation distance $|\cdot|$

is a norm on \mathcal{M}_\pm , and $(\mathcal{M}_\pm, |\cdot|)$ is a Banach space. The Bochner integral [6] 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, 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 $G(\nu)(f) := \int_\Omega \frac{f(x)}{d(x, \Omega)^p} d\nu(x)$. Theorem 3.1 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, F is measurable (see [33, 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

$$\mathbb{E}[\boldsymbol{\mu}] := G \left(\int \nu d(F_\# P)(\nu) \right) \in \mathcal{M}^p, \quad (37)$$

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 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 general result:

Proposition D.1 (see [60]). *Let $(V, \|\cdot\|)$ be a separable Banach space and \mathcal{X} a closed convex subset of V with non-empty interior. Let P be a probability measure on \mathcal{X} endowed with its Borelian σ -algebra. Assume that $\int \|x\| dP(x) < \infty$. Let $f : \mathcal{X} \rightarrow \mathbb{R}$ be a continuous convex function so that $\int f(x) dP(x) < \infty$. Then,*

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

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