Universidad Autónoma de Madrid

Mathematic Analysis Fundamentals

Optimal transport for Topological Data Analysis

End of Course Thesis.
2024-2025.

Author:
Gonzalo Ortega Carpintero

January 2025

Contents

1 Introduction

Transport maps were introduced in 1781 by Gaspard Monge to represent the idea of moving earth from one place into an other [5][1.1 Historical overview]. In this original formulation of the optimal transport problem, it was enough to consider $\mathbb{R}^3$ as the ambient space, using the Euclidean distance as the cost function of moving mass between two points.

In the 30's, Leonid Kantorovich reformulated the problem to describe the optimization process of supply and demand distributions of diverse problems. The mass could be divided between different origin and destinations, making it possible to interpret the problem as the way to measure the cost of transforming one probability distribution into an other. In this thesis, will introduce the p-Wasserstein distance as a metric on the probability measures with finite p-moment space. When $p = 1$, the distance will represent the metric introduced in the Kantorovich optimal transport problem, also used named Earth Mover's distance, used for machine learning algorithms and computer vision problems [7]. When $p = \infty$ it is named the bottleneck distance, and will be the main them of study of this thesis.

In topological data analysis (TDA), diagrams arise to represent the persistence of the homology groups of a data set through time. Those diagrams are named persistence diagrams, and those homology groups, persistence homology groups. We will introduce an analogous p-Wasserstein distance in the space of persistence diagrams and prove that there exists an isometric embedding from a separable metric space into the space of persistence diagrams with the Wasserstein distance.

The goal of this thesis is to give a brief introduction to optimal transport theory as starting point from where define the bottleneck distance in TDA. Note, however, that the Wasserstein distances we will define, first in probability spaces and then in persistence diagrams spaces are not equivalent. With probability measures, every point is *matched* to some other, while in persistence diagrams we will define *matchings* that might not relate all points between the two diagrams to compare, relating *unmatched* points to the diagonal of the plane.

Nevertheless, it is posible to define a general theory, thus far more abstract than needed for the purpose of this thesis. In this relative optimal transport theory, Wasserstein distances, in both, probability measure spaces and persistence diagram spaces are particular examples of a more general definition. Readers seeking further details may refer to [4].

2 Optimal transport

The main result of optimal transport theory is the solution of Kantorovich's problem for general costs: the existence of an optimal transport plan. In 2.1 introduce both,Monge's and Kantorovich's problems and explain its main differences. In 2.2 we define Wasserstein distance as it was originally conceived, as a way of computing a distance between two probability spaces.

2.1 Preliminaries

We fist define a way to compare probability measures from two different spaces. We will denote the set of probability measures over a space X by $\mathcal{P}(X)$, and the class of Borel-measurable sets by $\mathcal{B}(X)$.

Definition 2.1 (Push-forward measure)**.** Let $T : X \to Y$ be a Borel map, and $\mu \in \mathcal{P}(X)$. Let $A \in \mathcal{B}$. The *push-forward measure* $T_{\#}\mu \in \mathcal{P}(Y)$ is defined as

$$T_{\#}\mu(A) := \mu(T^{-1}(A)).$$

Now we can introduce transport maps, as functions witch transform one probability measure into an other.

Definition 2.2 (Transport map)**.** Given $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$, a *transport map from* μ *to* ν is a Borel map $T : X \to Y$ that satisfies $T_{\#}\mu = \nu$.

Definition 2.3 (Transport plan)**.** Let $\pi_X : (X \times Y) \to X$ and $\pi_Y : (X \times Y) \to Y$ such that for every $(x, y) \in (X, Y)$, $\pi_X(x, y) = x$ and $\pi_Y(x, y) = y$. A *transport plan between* μ *and* ν is a probability measure $\gamma \in \mathcal{P}(X \times Y)$ where

$$(\pi_X)_{\#}\gamma = \mu \text{ and } (\pi_Y)_{\#}\gamma = \nu.$$

The set of all transport plans between μ and ν is denoted $\Gamma(\mu, \nu)$.

While the set of transport maps between two given probability measures might be empty, transport plans are a more flexible generalization of them allowing to modulate one measure into the other. In probability theory, transport plans are named *couplings*, and $\Gamma(\mu, \nu)$ is the collection of all probability measures in $X \times Y$ with *marginals* μ and ν [6].

Given this definitions, we can introduce Monge and Kantorovich problems, $C_M(\mu, \nu)$ and $C_K(\mu, \nu)$ respectively, as follows.

Definition 2.4 (Transport problems)**.** Fix $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and consider a lower semicontinous map $c : X \times Y \to [0, \infty]$. Then

$$
\begin{aligned}
C_M(\mu, \nu) &:= \inf \left\{ \int_X c(x, T(x)) d\mu(x) : T_{\#}\mu = \nu \right\}, \\
C_K(\mu, \nu) &:= \inf \left\{ \int_{X \times Y} c(x, y) d\gamma(x, y) : \gamma \in \Gamma(\mu, \nu) \right\}.
\end{aligned}
$$

The function c in the above formulations is denoted *cost* function, and the integrals to minimice are the *transportation costs*. Monge's formulation consist on minimizing the transportation cost among all transport maps, while Kantorovich's formulation consists in minimizing the transport cost of all transport plans.

There are cases where Monge's problem has no solution, as sometimes there is no transport map T satisfying $T_{\#}\mu = \nu$. One of this cases arises when μ is a Dirac measure but ν is not. The key of Kantorovich problem is that it always has a transport plan that minimizes the problem. This will prove useful to verify that Wasserstein distance exists and it is a well defined metric.

Theorem 2.5 (Existance of an optimal coupling)**.** *Let $c : X \times Y \to [0, \infty]$ be lower semicontinous, and let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. Then there exists a coupling $\bar{\gamma} \in \Gamma(\mu, \nu)$ that verifies*

$$
C_K(\mu, \nu) = \int_{X \times Y} c(x, y) d\bar{\gamma}(x, y).
$$

The proof of this theorem implies some auxiliary results that outreach the purpose of this thesis. A complete proof of Theorem 2.5 can be found at [5][Theorem 2.3.2]. An alternative prove can also be found in [9][Theorem 4.1].

2.2 Wasserstein distance over probability spaces

In an analogous way as Lebesgue spaces is defined over complex measurable functions, that is, for functions over a measure space X with a positive measure μ such that for $0 < p < \infty$, the Lebesgue p-norm

$$
\|f\|_p = \left(\int_X |f|^p d\mu \right)^{\frac{1}{p}} < \infty,
$$

p-Wasserstein distance is defined over probability measure spaces where μ is a probability measure with finite p-moment. To give an example, in probability theory, the 1-moment represent the mean and the 2-moment, the variance.

Definition 2.6 (Probability measures with finite p-moment)**.** Let (X, d) be a locally compact and separable, metric space. Let $1 \leq p < \infty$. The *set of probability measures with finite p-moment* is defined As

$$\mathcal{P}_p(X) := \left\{ \sigma \in \mathcal{P}(X) : \int_X d(x, x_0)^p d\mu(x) < \infty \text{ for some } x_0 \in X \right\}.$$

Proposition 2.7. *The definition of $\mathcal{P}_p(X)$ is independent of the base point x_0*

Proof. Let $x_1 \in X$ be other arbitrary base point, and consider the function $f(s) = s^p$ with $s \in \mathbb{R}^+$, $p \in \mathbb{N}$. The function f is convex, that is, for every $a, b \in \mathbb{R}^+$ and for every $\lambda \in [0, 1]$, it satisfies

$$f(\lambda a + (1 - \lambda)b) \leq \lambda f(a) + (1 - \lambda) f(b).$$

Now, taking $\lambda = \frac{1}{2}$,

$$(a + b)^p \leq 2^{p-1}(a^p + b^p).$$

Thus, considering distances, and using the triangle inequality we have

$$d(x, x_1)^p \leq (d(x, x_0) + d(x_0, x_1))^p \leq 2^{p-1}(d(x, x_0)^p + d(x_0, x_1)^p).$$

Therefore, if a measure $\sigma \in \mathcal{P}(X)$ satisfies $\int_X d(x, x_0)^p d\mu(x) < \infty$, then it must also satisfy $\int_X d(x, x_1)^p d\mu(x) < \infty$. $\square$

Definition 2.8 (p-Wasserstein distance)**.** Let $0 < p < \infty$. Let $u, v \in \mathcal{P}_p(X)$ two probability measures with finite p-moment, the *p-Wasserstein distance* is defined as

$$W_p(u, v) := \left(\inf_{\gamma \in \Gamma(u,v)} \int_{X \times X} d(x, y)^p d\gamma(x, y) \right)^{\frac{1}{p}}.$$

For $p = \infty$ is posible to define the *∞-Wasserstein distance* making use of the usual L^∞ norm taken with respect to μ as

$$W_\infty(\mu, \nu) = \inf \|d(x, y)\|_\infty.$$

This particular case of the Wasserstein distance is named *bottleneck distance* and will be of first importance in its equivalent version for persistence diagrams that we will introduce in Section 3.

We will now prove that the p-Wasserstein distance is a metric in the space of Probability measures with finite p-moment. We limit to the case of $0 < p < \infty$, but for $p = \infty$ a similar procedure can me made. To check the triangle inequality, we will make use of the following auxiliary theorem of probability theory. Let $\mathbf{X}$ denote $X_1 \times \cdots \times X_n$ for some $n \in \mathbb{N}$.

Theorem 2.9 (Disintegration)**.** *Let $\mathbf{X}, X$ be Radon separable metric spaces, $\mu \in \mathcal{P}(\mathbf{X})$, let $\pi : \mathbb{X} \to X$ be a Borel map and let $\nu = \pi_{\#}\mu \in \mathcal{P}(X)$. Then there exists a ν-a.e. uniquely determined Borel family of probability measures $\{\mu_x\}_{x\in X} \subset \mathcal{P}(\mathbf{X})$ such that*

$$\mu_x(\mathbf{X} \setminus \pi^{-1}(x)) = 0 \textit{ for } \nu\textit{-a. } x \in X$$

and

$$\int_{\mathbf{X}} f(\mathbf{x})\, d\mu(\mathbf{x}) = \int_X \left(\int_{\pi^{-1}(x)} f(\mathbf{x})\, d\mu_x(\mathbf{x}) \right) d\nu(x).$$

The interested reader can find one proof of this theorem at [1][Theorem 5.3.1]. The disintegration theorem needs X to be a Radon space, that is, every finite Borel measure is a Radon measure. With a Radon measure being a measure that is finite on all compact sets, outer regular on all Borel sets, and inner regular on open sets.

It is also possible to prove the triangle inequality in a more elementary manner without the use of the disintegration theorem as seen in [3]. This ables to omit the assumption of the underlying space been Radon, but obtaining an obscurer proof in exchange. For the aim of this thesis, will be confortable with the requirements of the disintegration theorem and we present a proof following the steps made in [5][Theorem 3.1.5].

Proposition 2.10. *W_p is a distance on the space $\mathcal{P}_p(X)$.*

Proof. Let $\mu, \nu \in \mathcal{P}_p(X)$. Then, if $W_p(\nu, \mu) = 0$, Theorem 2.5 asures that there exists some $\bar{\gamma}$ such that

$$\int_{X\times X} d(x,y)^p d\bar{\gamma}(x,y) = 0.$$

This means $x = y$ $\bar{\gamma}$-.a.e. so $\bar{\gamma} = (\mathrm{Id} \times \mathrm{Id})_{\#}\mu$, making $\nu = (\pi_{\mathrm{Id}})_{\#}\bar{\gamma} = \mu$.

To prove the symmetry of W_p let $\gamma \in \Gamma(\nu, \mu)$ be the optimal coupling between ν and ν, and define $\tilde{\gamma} := S_{\#}\gamma$, with $S(x,y) = (y,x)$. Then $\tilde{\gamma} \in \Gamma(\nu, \mu)$ and using the symmetry of the distance, we have

$$W_p(\nu, \mu) \leq \int_{X\times X} d(x,y)^p d\tilde{\gamma} = \int_{X\times X} d(x,y)^p d\gamma = W_p(\mu, \nu).$$

Now repeating the same procedure but exchanging the order of μ and ν, we get $W_p(\mu, \nu) \leq W_p(\nu, \mu)$ so it has to be $W_p(\mu, \nu) = W_p(\nu, \mu)$.

To prove the triangle inequality, let $\mu_1, \mu_2, \mu_3 \in \mathcal{P}_p(X)$ and let $\gamma_{12} \in \Gamma(\mu_1, \mu_2)$ and $\gamma_{23} \in \Gamma(\mu_2, \mu_3)$ be optimal couplings. Let $\tilde{\gamma} \in \mathcal{P}(X \times X \times X)$. Using Theorem 2.9 we have

$$\begin{aligned}
&\int_{X\times X\times X} \phi(x_1, x_2) d\tilde{\gamma}(x_1, x_2, x_3) \\
&= \int_{X\times X} \left(\int_X \phi(x_1, x_2)\gamma_{12,x_2}(dx_1) d\gamma_{23}(x_3) \right) d\mu_2(x_2) \\
&= \int_{X\times X} \phi(x_1, x_2)\gamma_{12,x_2}(dx_1) \left(\int_X d\gamma_{23}(x_3) \right) d\mu_2(x_2) \\
&= \int_{X\times X} \phi(x_1, x_2)\gamma_{12}(x_1, x_2).
\end{aligned}$$

In the same way

$$\int_{X\times X\times X} \phi(x_2, x_3) d\tilde{\gamma}(x_1, x_2, x_3) = \int_{X\times X} \phi(x_2, x_3)\gamma_{23}(x_2, x_3).$$

Therefore, note that we also have

$$\int_{X\times X\times X} \psi(x_1) d\tilde{\gamma}(x_1, x_2, x_3) = \int_{X\times X} \psi(x_1) d\gamma_{12}(x_1, x_2) = \int_X \psi(x_1) d\gamma_1(x_1)$$

and

$$\int_{X\times X\times X} \psi(x_3) d\tilde{\gamma}(x_1, x_2, x_3) = \int_X \psi(x_3) d\gamma_3(x_3).$$

This two equations imply that

$$\int_{X\times X\times X} \phi(x_1 x_3) d\tilde{\gamma}(x_1, x_2, x_3) = \int_X \phi(x_1, x_3) d\tilde{\gamma}(x_1, x_2, x_3).$$

Then, if we set $\bar{\gamma}_{13} := \int_X \tilde{\gamma}(x_1, x_2, x_3)$, we have that $\bar{\gamma}_{13} \in \Gamma(\mu_1, \mu_3)$. Finally, using the triangle inequality of the Lebesgue space $L^p(X \times X \times X, \tilde{\gamma})$ we have

$$\begin{aligned}
W_p(\mu_1, \mu_3) &\le \left(\int_{X\times X} d(x_1, x_3)^p d\bar{\gamma}_{13}(x_1, x_3) \right)^{\frac{1}{p}} \\
&= \left(\int_{X\times X} d(x_1, x_3)^p d\tilde{\gamma}_{13}(x_1, x_2 x_3) \right)^{\frac{1}{p}} \\
&= ||d(x_1, x_3)||_{L^p(\tilde{\gamma})} \le ||d(x_1, x_2) + d(x_2, x_3)||_{L^p(\tilde{\gamma})} \\
&\le ||d(x_1, x_2)||_{L^p(\tilde{\gamma})} + ||d(x_2, x_3)||_{L^p(\tilde{\gamma})} \\
&\le ||d(x_1, x_2)||_{L^p(\gamma 12)} + ||d(x_2, x_3)||_{L^p(\gamma 23)} \\
&= W_p(\mu_1, \mu_2) + W_p(\mu_2, \mu_3).
\end{aligned}$$

$\square$

To conclude this section introducing Wasserstein distances over provability measures we will give a concrete example putting theory into practice.

Example 2.11. We fill fix $p = 2$ and $X = \mathbb{R}$ to compute the 2-Wasserstein distance $W_2(\mu, \nu)$ between two continuous probability measures on $\mathbb{R}$. Let $\mu(x) = \mathbb{I}_{[0,1]}(x)$ and $\nu(x) = 2 \cdot \mathbb{I}_{[0,1/2]}(x)$ where $\mathbb{I}_{[a,b]}$ is the indicator function

$$\mathbb{I}_{[a,b]}(x) = \begin{cases} 1 & \text{if } x \in [a,b], \\ 0 & \text{otherwise.} \end{cases}$$

That is, μ is a uniform distribution on $[0,1]$, and ν is a uniform distribution on $[0,1/2]$, scaled to integrate to 1. Their density functions are of $f_\mu(x) = 1$ for $x \in [0,1]$ and $f_\nu(x) = 2$ for $x \in [0,1/2]$. The W_2-distance between μ and ν is then given by

$$W_2^2(\mu, \nu) = \inf_{\gamma \in \Gamma(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} d(x,y)^2 \, d\gamma(x,y),$$

where $\Gamma(\mu, \nu)$ is the set of all couplings of μ and ν.

For measures on $\mathbb{R}$, the optimal coupling can be determined using the *monotone transport map* $T(x)$, where $T : \mathbb{R} \to \mathbb{R}$ is the map that pushes μ onto ν. The W_2-distance then simplifies to

$$W_2^2(\mu, \nu) = \int_{\mathbb{R}} |x - T(x)|^2 f_\mu(x) \, dx.$$

To compute $T(x)$, we observe that the transport map must verify

$$\int_{-\infty}^{T(x)} f_\nu(y) \, dy = \int_{-\infty}^{x} f_\mu(y) \, dy,$$

witch for our μ and ν becomes

$$\int_0^{T(x)} 2 \, dy = \int_0^x 1 \, dy.$$

Thus, the optimal transport map is $T(x) = x/2$, which maps the distribution of μ onto ν. Substituting into the formula for W_2^2

$$W_2^2(\mu, \nu) = \int_0^1 \left| x - \frac{x}{2} \right|^2 f_\mu(x) \, dx.$$

Since $f_\mu(x) = 1$ for $x \in [0,1]$, this simplifies to

$$W_2(\mu, \nu) = \left(\int_0^1 \left(\frac{x}{2} \right)^2 dx \right)^{\frac{1}{2}} = \left(\int_0^1 \frac{x^2}{4} \, dx \right)^{\frac{1}{2}} = \left(\frac{1}{4} \left[\frac{x^3}{3} \right]_0^1 \right)^{\frac{1}{2}} = \sqrt{\frac{1}{12}} = \frac{1}{2\sqrt{3}}.$$

3 Reformulations for TDA

In last section we have exposed the original optimal transport problem where the objetive was to measure distance between probability measures. We will now define a new Wasserstein distance, inspired in the original one, looking forward to measure the distance between persistence diagrams. In 3.1 we make a brief introduction to TDA concepts required to define our new Wasserstein distance. We also check that it is actually a distance between persistence diagrams. In 3.2 and we will conclude with the main result of this thesis: the existence of an isometric embedding from a separable metric space into the space of persistence diagrams.

3.1 Wasserstein distance over persistence diagrams

We will denote the strict upper triangular region of the Euclidean plane as $\mathbb{R}^2_< := \{(x, y) \in \mathbb{R}^2 : x < y\}$, and the diagonal of the plane as $\Delta := \{(x, y) \in \mathbb{R}^2 : x = y\}$.

Definition 3.1 (Persistence diagram)**.** Let I be a countable set. A *persistence diagram* is a function $D : I \to \mathbb{R}^2_<$.

Persistence diagrams are just a way of presenting the out coming from computing the persistence homology groups of a set of data. This output comes in the so called *barcodes*, which are multisets of intervals. As every interval is given with its *birth* and *death* parameters, it can as well be seen as a point in $\mathbb{R}^2_<$.

Definition 3.2 (Partial matching)**.** Let $D_1 : I_1 \to \mathbb{R}^2_<$ and $D_2 : I_2 \to \mathbb{R}^2_<$ be persistence diagrams. A *partial matching* between D_1 and D_2 is the triple (I'_1, I'_2, f) such that $f : I'_1 \to I'_2$ is a bijection with $I'_1 \subseteq I_1$ and $I'_2 \subseteq I_2$.

Example 3.3. To give some visual intuition of how persistence diagrams are matched between each other, in Figure 3.3 we represent two diagrams, D_1 and D_2 over $\mathbb{R}^2_<$ with the diagonal Δ and a posible partial matching between them.

Instead of probability measures, now we are actually dealing with countable sets of points in $\mathbb{R}$. We will make use of the l^p norm at countable spaces to measure the distance between matched pairs and the distance between unmatched pairs and the diagonal Δ. For a more detailed explanation of Lebesgue measures check [8][Definition 3.7]. This norm is named after Pafnuty Chebyshev.

Definition 3.4 (Chebyshev distance)**.** Let $a, b \in \mathbb{R}^2$ with $a = (a_x, a_y)$ and $b = (b_x, b_y)$. The *Chebyshev distance* is defined as

$$d_\infty(a, b) := ||a - b||_\infty := \max\{|a_x - b_x|, |a_y - b_y|\}.$$

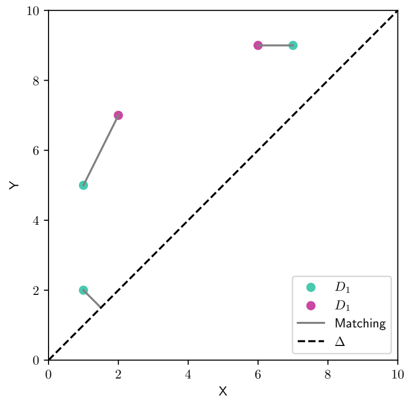

Figure 1: Partial matching over two persistence diagrams

To define our adapted Wasserstein distance we need to check how Chebyshev distance measures distances between points of $\mathbb{R}^2_<$ and Δ.

Proposition 3.5. *If $a = (a_x, a_y) \in \mathbb{R}^2_<$, then $d_\infty(a, \Delta) = \inf_{t \in \Delta} d_\infty(a, t) = \frac{a_y - a_x}{2}$.*

Proof. The t which minimizes the distance is the midpoint of a_x and a_y, that is $t = \left(\frac{a_x + a_y}{2}, \frac{a_x + a_y}{2}\right)$. Then,

$$\left| a_x - \frac{a_x + a_y}{2} \right| = \left| \frac{a_x - a_y}{2} \right| = \left| \frac{a_y - a_x}{2} \right| = \left| a_y - \frac{a_x + a_y}{2} \right|,$$

and as $a_y > a_x$ we have

$$d_\infty(a, t) = \left| \frac{a_y - a_x}{2} \right| = \frac{a_y - a_x}{2}.$$

□

We now verify that the upper triangular region of the Euclidean plane with the Chebyshev distance adapted to measure distances in Δ is a metric space.

Proposition 3.6. *d_∞ is a distance in $\mathbb{R}^2_<$ with the diagonal Δ.*

Proof. For points $a, b \in \mathbb{R}^2_< \subset \mathbb{R}^2$, d_∞ is a distance as usual Lebesgue norms are well defined. See [8][Chapter 3]. To verify that the metric requirements are fulfilled for $d_\infty(a, \Delta)$, it is enough to consider $t = \frac{a_y - a_x}{2}$ as in Proposition 3.5. □

Definition 3.7 (p-cost)**.** Let $D_1 : I_1 \to \mathbb{R}^2_<$ and $D_2 : I_2 \to \mathbb{R}^2_<$ be persistence diagrams. Let (I'_1, I'_2, f) be a partial matching between them. If $p < \infty$, the *p-cost of f* is defined as

$$\begin{aligned}\mathrm{cost}_p(f) := \Bigg(&\sum_{i \in I'_1} d_\infty(D_1(i), D_2(f(i)))^p \\ &+ \sum_{i \in I_1 \setminus I'_1} d_\infty(D_1(i), \Delta)^p \\ &+ \sum_{i \in I_2 \setminus I'_2} d_\infty(D_2(i), \Delta)^p \Bigg)^{\frac{1}{p}}.\end{aligned}$$

For $p = \infty$, the *∞-cost of f* is defined as

$$\begin{aligned}\mathrm{cost}_\infty(f) := \max \Bigg\{ &\sup_{i \in I'_1} d_\infty(D_1(i), D_2(f_i)), \\ &\sup_{i \in I_1 \setminus I'_1} d_\infty(D_1(i), \Delta), \\ &\sup_{i \in I_2 \setminus I'_2} d_\infty(D_2(i), \Delta) \Bigg\}.\end{aligned}$$

Definition 3.8 (p-Wasserstein distance)**.** Let D_1, D_2 be persistence diagrams. Let $1 \leq p \leq \infty$. Define

$$\tilde{\omega}_p(D_1, D_2) = \inf\{\mathrm{cost}_p(f) : f \text{ is a partial matching between } D_1 \text{ and } D_2\}.$$

Let $\emptyset$ denote the unique persistence diagram with empty indexing set. Let $(\mathrm{Dgm}_p, \omega_p)$ be the space of persistence diagrams D that satisfy $\tilde{\omega}_p(D, \emptyset) < \infty$ modulo the equivalence relation $D_1 \sim D_2$ if $\tilde{\omega}_p(D_1, D_2) = 0$. The metric ω_p is called the *p-Wasserstein distance.*

Definition 3.9 (Bottleneck distance)**.** In the conditions of Definition 3.8, if $p = \infty$, the metric ω_∞ is called the *bottleneck distance.*

Proposition 3.10. *There is only one matching between $D : I \to \mathbb{R}^2_<$ and $\emptyset$. Hence, if $p \leq \infty$,*

$$\tilde{\omega}_p(D, \emptyset) = \left(\sum_{i \in I} d_\infty(D(i), \Delta)^p \right)^{\frac{1}{p}},$$

and, if $p = \infty$,

$$\tilde{\omega}_\infty(D, \emptyset) = \sup_{i \in I} d_\infty(D_1(i), \Delta)$$

Proof. Let $I' \subseteq D$. If f is a partial matching between D and $\emptyset$, means that $f(I') = \emptyset$ is a bijection. That is only posible if $I' = \emptyset$ too. Therefore $I \setminus I' = I \setminus \emptyset = I$ and following Definition 3.7 we conclude our proof. □

Next proposition will prove that, in indeed, the space of persistence diagrams with the p-Wasserstein distance $(\mathrm{Dgm}_p, \omega_p)$ is a metric space. Its proof is usually omitted in literature, as it based on the simple fact that d_∞ is a distance. We will give, however, an step by step version here.

Proposition 3.11. *ω_p is a distance on the space* $(\mathrm{Dgm}_p, \omega_p)$.

Proof. Let $D_1, D_2, D_3 \in \mathrm{Dgm}_p$, with $1 \leq p \leq \infty$.

First of all, $\omega_p(D_1, D_2) \geq 0$ because $d_\infty \geq 0$. $\omega_p(D_1, D_2) = 0$ if and only if $\tilde{\omega}_p(D_1, D_2) = 0$. Thus, because of the equivalence relationship used to define ω_p, it has to be $D_1 \sim D_2$.

To check symmetry, note that every partial matching f is bijective, therefore f^{-1} is a partial matching. But, for all $i \in I'_1$, exists $j \in I'_2$ such that $f(i) = j$ and

$$d_\infty(D_1(i), D_2(f(i))) = d_\infty(D_2(f(i)), D_1(i)) = d_\infty(D_2(j), D_1(f^{-1}(j))).$$

Then, $\mathrm{cost}_p(f) = \mathrm{cost}_p(f^{-1})$ and we have

$$\begin{aligned}
\omega_p(D_1, D_2) &= \inf\{\mathrm{cost}_p(f) : f \text{ is a partial matching between } D_1 \text{ and } D_2\} \\
&= \inf\{\mathrm{cost}_p(f^{-1}) : f^{-1} \text{ is a partial matching between } D_2 \text{ and } D_1\} \\
&= \omega_p(D_2, D_1).
\end{aligned}$$

Finally, lets prove the triangle inequality. If $f : I'_1 \to I'_2$ is a partial matching between D_1 and D_2 and $g : I'_2 \to I'_3$ is a partial matching between D_2 and D_3, $g \circ f : I'_1 \to I'_3$ is a partial matching between D_1 and D_3 as both f and g are bijective. Computing the cost of the matchings for $p < \infty$, we notice that

$$\begin{aligned}
&\sum_{i \in I'_1} d_\infty(D_1(i), D_2(f(i))) + \sum_{i \in I_1 \setminus I'_1} d_\infty(D_1(i), \Delta) + \sum_{i \in I_2 \setminus I'_2} d_\infty(D_2(i), \Delta) \\
&+ \sum_{i \in I'_2} d_\infty(D_2(i), D_3(g(i))) + \sum_{i \in I_2 \setminus I'_2} d_\infty(D_2(i), \Delta) + \sum_{i \in I_3 \setminus I'_3} d_\infty(D_3(i), \Delta) \\
&\geq \sum_{i \in I'_1} d_\infty(D_1(i), D_3(g \circ f(i))) + \sum_{i \in I_1 \setminus I'_1} d_\infty(D_1(i), \Delta) + \sum_{i \in I_3 \setminus I'_3} d_\infty(D_3(i), \Delta)
\end{aligned}$$

as $d_\infty(D_1(i), D_2(f(i))) + d_\infty(D_2(f(i)), D_2(g(f(i)))) \geq d_\infty(D_1(i), D_3(g \circ f(i)))$ using the triangle inequality of d_∞. Therefore, for all partial matchings f and g as described, we have $\text{cost}_p(f) + \text{cost}_p(g) \geq \text{cost}_p(g \circ f)$. Using the same reasoning, por $p = \infty$ we also obtain $\text{cost}_\infty(f) + \text{cost}_\infty(g) \geq \text{cost}_\infty(g \circ f)$. Hence, we have verified that

$$\omega_p(D_1, D_2) + \omega_p(D_2, D_3) \geq \omega_p(D_1, D_3).$$

□

3.2 Metric spaces into persistence diagrams

Finally, we will present the main theorem of this thesis, witch asserts that any separable bounded metric space can be seen in the form of a persistence diagram. This interesting fact allows the bottleneck distance to be used in much more general cases than the ones offered by the output of persistence homology.

Definition 3.12 (Isometric embedding)**.** Let $(X, d_X), (Y, d_Y)$ be metric spaces. An *isometric embedding* $\eta : (X, d_X) \to (Y, d_Y)$ is a mapping that satisfies

$$d_X(x_1, x_2) = d_Y(\eta(x_1), \eta(x_2))$$

for all $x_1, x_2 \in X$.

Definition 3.13 (Ball in persistence diagrams)**.** Let $1 \leq p \leq \infty$. Let $D_0 \in \text{Dgm}_p$. The *ball* at the space of persistence diagrams is defined as $B_p(D_0, r) := \{D \in \text{Dgm}_p : w_p(D, D_0) < r\}$.

Theorem 3.14 (Isometric embedding of metric spaces into persistence diagrams)**.** *Let (X, d) be a separable, bounded metric space. Then there exists an isometric embedding to the space of persistence diagrams $\eta : (X, d) \to (\text{Dgm}_\infty, \omega_\infty)$ such that $\eta(X) \subseteq B(\emptyset, \frac{3c}{c}) \backslash B(\emptyset, c)$.*

Proof. We will follow the procedure followed in [2][Theorem 19]. As (X, d) is bounded, we can let $c > \sup\{d(x, y) : \ x, y \in X\}$. As (X, d) is separable, we can take $\{x_k\}_{k=1}^\infty$, a countable, dense subset of (X, d). Consider

$$\begin{aligned} \eta : (X, d) &\to (\text{Dgm}_\infty, \omega_\infty) \\ x &\mapsto \{(2c(k-1), 2ck + d(x, x_k))\}_{k=1}^\infty \end{aligned}$$

For any $x \in X$ and $k \in \mathbb{N}$,

$$\begin{aligned} d_\infty((2c(k-1), 2ck + d(x, x_k)), \Delta) &= \frac{2ck + d(x, x_k) - 2c(k-1)}{2} \\ &= c + \frac{d(x, x_k)}{2} \\ &< c + \frac{c}{2} = \frac{3c}{2}. \end{aligned}$$

Because of Proposition 3.10, for every $x \in X$, $\tilde{\omega}_\infty(\eta(x), \emptyset) < \infty$ and η is well defined. Note that

$$\omega_\infty(\eta(x), \emptyset) = \sup_{1 \leq k < \infty} d_\infty((2c(k-1), 2ck + d(x, x_k)), \Delta),$$

so $\eta(x) \in B(\emptyset, \frac{3c}{c}) \backslash B(\emptyset, c)$.

Let $\eta(x)$ and $\eta(y)$ two equivalence classes of $(\mathrm{Dgm}_\infty, \omega_\infty)$. Choose the representative diagrams $D_x : \mathbb{N} \to \mathbb{R}^2_<$ and $D_y : \mathbb{N} \to \mathbb{R}^2_<$ and consider the partial matching $(\mathbb{N}, \mathbb{N}, \mathrm{id}_\mathbb{N})$. With it, for every $k \in \mathbb{N}$, $(2c(k-1), 2ck + d(x, x_k))$ is matched with $(2c(k-1), 2ck + d(y, x_k))$. The Chebyshev distance between those points is

$$\begin{aligned} d_\infty(D_x(k), D_y(k)) &= \max \big\{ |2c(k-1) - 2c(k-1)|, \\ &\qquad |2ck + d(x, x_k) - b_y - (2ck + d(y, x_k))| \big\} \\ &= \max\{0, |d(x, x_k) - d(y, x_k)|\} \\ &= |d(x, x_k) - d(y, x_k)|. \end{aligned}$$

Hence, because of the triangle inequality, the cost of this partial matching is

$$\mathrm{cost}_\infty(\mathrm{id}_\mathbb{N}) = \sup_k |d(x, x_k - d(y, x_k)) \leq d(x, y).$$

Since $\{x_k\}_{k=1}^\infty$ is dense, for every $\epsilon > 0$, there exist a $k \in \mathbb{N}$ such that $d(x, x_k) \leq \epsilon$, so

$$\begin{aligned} |d(x, x_k) - d(y, x_k)| &\geq d(y, x_k) - d(x, x_k) \\ &= d(y, x_k) + d(x, x_k) - d(x, x_k) - d(x, x_k) \\ &\geq d(x, y) - 2d(x, x_k) \\ &> d(x, y) - 2\epsilon. \end{aligned}$$

Therefore, $\sup_k |d(x, x_k - d(y, x_k)) \geq d(x, y)$ and

$$\mathrm{cost}_\infty(\mathrm{id}_\mathbb{N}) = \sup_k |d(x, x_k - d(y, x_k)) = d(x, y).$$

Suppose $I, J \subseteq \mathbb{N}$ and (I, J, f) is a different partial matching between D_x and D_y. Then there exist a $k \in \mathbb{N}$ such that either $k \notin I$ or $k \in I$ and $f(k) = k \neq k$. If $k \notin I$, then

$$\text{cost}_\infty(f) \geq d_\infty((2c(k-1), 2ck + d(x, x_k)), \Delta) \geq c.$$

If $k \in I$ and $f(k) = k \neq k$, then

$$\text{cost}_\infty(f) \geq ||(2c(k-1), 2ck + d(x, x_k)) - (2c(k'-1), 2ck' + d(x, x_{k'}))||_\infty \geq 2\epsilon.$$

Hence, $\text{cost}_\infty(f) \geq c > d(x, y)$ and $d(x, y) = \omega_\infty(\eta(x), \eta(y))$, proving that η is an isometric embedding of a metric space into the space of persistence diagrams. $\square$

Lastly, we show in a toy example how the isometric embedding works for a discrete metric space of tree points.

Example 3.15. In Figure 3.15 we follow the proceeding used in the proof of Theorem 3.14. We have pictured a metric space with three points, x_1, x_2 and x_3, with the respective distances between them, and its image for the map ν with $c = 6$, were $x_i \mapsto \{12(k-1), 12k + d(x_i, x_k)\}_{k=1}^3$.

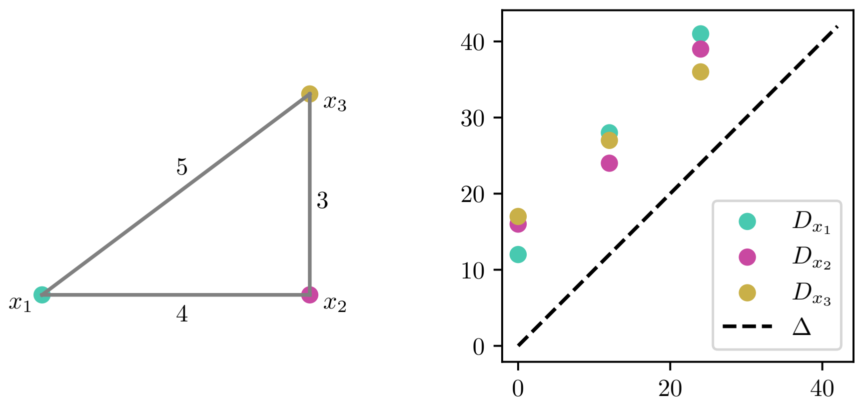

Figure 2: Isometric embeeding over a finite metric space.

References

[1] Ambrosio, L., Gili, N., and Savaré, G. *Gradient Flows in Metric Spaces and in the Space of Probability Measures.* Birkhäuser Verlag, 2005.

[2] Bubenik, P., and Wagner, A. Embeddings of persistence diagrams into Hilbert spaces. *Journal of Applied and Computational Topology* (2020).

[3] Clement, P., and Desch, W. An elementary proof of the triangle inequality for the Wasserstein metric. *Proceedings of the American Mathematical Society* (2007).

[4] Elchesen, A. Universality of the Wasserstein distances and relative optimal transport. *University of Florida* (2022).

[5] Figalli, A., and Glaudo, F. *An invitation to Optimal Transport, Wasserstein distances, and gradient flows.* EMS Press, 2020.

[6] Givens, C. R., and Shortt, R. M. A class of Wasserstein metrics for probability distributions. *Michigan Technological University* (1984).

[7] Rubner, Y., Tomasi, C., and Guibas, L. J. The Earth Mover's Distance as a metric for image retrieval. *International Journal of Computer Vision* (2000).

[8] Rudin, W. *Real and Complex Analysis.* McGraw-Hill, 1987.

[9] Villani, C. *Optimal transport, old and new.* Springer, 2008.