Università degli Studi di Padova
Dipartimento di Matematica "Tullio Levi-Civita"
Corso di Laurea Magistrale in Matematica

The use of Persistent Homology in Data Science for Topological Data Analysis

Relatore
Prof. Marco Andrea Garuti

Laureanda
Clara Isabel López González
1218532

21 Luglio 2021

Anno Accademico 2020–2021

Preface

Don't disturb my circles!

Archimedes

Topological data analysis (TDA) is a young field that has been rapidly growing over the last years and which blends algebraic topology, statistics and computer science. It arises motivated by the fact that the topology of a space gives useful information about it. In particular, when working with scientific data, information about their internal structure may provide important properties about the phenomena that it represents or its behaviour. And sometimes this information is topological. The goal of topological data analysis is to provide well-founded mathematical, statistical and algorithmic methods to infer the topological structure underlying the data.

The aims of this thesis are two. On the one hand, to introduce this field and its mathematical tools, focusing on persistent homology and its representation by persistence diagrams. On the other hand, the study of the latter using techniques from optimal transport, proposing a vectorization based on transport maps. The thesis is divided into 6 chapters as follows.

In chapter 1 we introduce topological data analysis and its relation with simplicial complexes, which allow to turn data into geometric objects. The homotopy equivalence between these complexes and the internal structure of the data is studied through the Nerve (1.2.1) and Reconstruction theorems (1.3.8). Here we will follow basically [1], [2] and [3].

In chapter 2 we present the theory of persistent homology, the main tool of TDA. To this aim we introduce persistence modules, a mathematical object designed to carry topological information about a data set at many different scales simultaneously. This information can be compactly represented by persistence diagrams. We will focus on the case of decomposable persistence modules.

In chapter 3 the definition of persistence diagram is extended to non-decomposable modules, by taking a measure-theoretical point of view. The main goal is to introduce persitence diagrams for the general case of q-tame modules. The main reference for these last two chapters is [4].

In chapter 4 we focus on the study of persistence diagrams and their stability properties. Notice that, since scientific data almost always depends on a probability distribution, it is important to know the statistical behaviour of these methods. To this end, distances between persistence diagrams are introduced. In particular, the bottleneck distance and the p-th or Wasserstein distance. For this, we refer again to [1] and [4]. Moreover, persistence diagrams can be seen as point measures. This

let us present the metric space of persistence diagrams with the optimal transport distance. We follow [5] and [6].

In chapter 5 we introduce a series of vectorizations for persistence diagrams, which are necessary in order to perform linear tasks with them. In particular, linear representations are studied [6]. These are easy to handle vectorizations for which, using results of chapter 4, we can deduce continuity characterizations. Moreover, they include usual TDA vectorizations as persistence surfaces, silhouettes and Betti curves. Persistence landscapes [1] are also introduced. We finish by proposing a new vectorization based on transport maps.

Finally, so as to exemplify the theory, computations of the notions introduced are made. They are included, together with their representations, in a Python Jupyter Notebook. We summarize it in chapter 6. The data set used was taken from GUDHI TDA tutorial.

Furthermore, two appendixes are added for the sake of completeness. The first is a reminder of the basics of simplicial complexes and simplicial homology. The second one is a summary of the principal results on measure theory that are needed through the text.

Acknowledgements

First and foremost, I would like to thank my advisor from the University of Bordeaux, Jérémie Bigot, for accepting to supervise my study in this topic, for introducing me to optimal transport theory and for his continuous guidance and support throughout this work.

I would also like to express my gratitude to the ALGANT master program, for providing me with an opportunity to spend a year in Padova and another in Bordeaux. Specially to coordinators in each university, for all the help offered.

Thanks are also due to my friends in Padova and Bordeaux who made these two years a truly remarkable experience, despite their peculiarity. In particular to Bianca: we constantly helped each other, both in life and university. I am thankful for having overcome last year's difficulties together.

I also want to express my gratitude towards all my hometown friends. Specially to Álvaro, who has always been by my side, in good and bad times, and who has shown me the beauty of mathematics.

Last but not least, I want to thank my parents, Maribel and Benito, and my family, for their constant support and encouragement through my life.

Contents

Chapter 1

Introduction to topological data analysis

The recent field of **topological data analysis** (TDA) has its roots in various works in applied algebraic topology and computational geometry. The most important ones are the works due to Edelsbrunnet et al. (2002) [7] and Zomorodian and Carlsson (2005) [8]. TDA exploits the fact that topological information of a space provides important properties, characteristics and invariants. Therefore, it may be used to study scientific data sets in order to obtain information about their global structure.

In this chapter we provide an introduction to the mathematical methods used in TDA to infer the topological information encoded in data. We will follow basically [1], [2] and [3]. It should be pointed out that computational aspects are crucial in TDA, and mentioned throughout the text. Indeed, robust and efficient algorithms are implemented and available, for instance, in the GUDHI library.

1.1 Introduction

Let $\mathbb{X}$ be a data set, i.e, a finite set of points with a notion of distance (or similarity) between them. It will be considered as a metric space on its own or a sample of a metric space, with the induced metric. For example, $\mathbb{X}$ may be a noisy sampling of some geometric structure, spatial coordinates obtained from some phenomena or discrete orbits of some dynamical system. Notice that we are not asking for the data to be embedded into an euclidean space, just a notion of distance.

Data sets are studied in order to obtain as much information as possible about their underlying structure. In particular, topological information might provide significant features. Nevertheless, point clouds are totally disconnected spaces with extremely simple topology. We need to assemble these discrete points into a global structure where we can study relevant topological features, and that still gives us information about the underlying shape of the data. The most obvious way to transform points $\mathbb{X}$ in a metric space into a global object is to use them as the vertices of a graph, whose edges are determined by proximity. Such a graph, while showing connectivity, ignores plenty of higher order features as cycles, voids, etc. Hence, simplicial complexes are built instead (see appendix A). As a result, we end up with non-topologically trivial spaces that hopefully retain the topological properties of

the original space where the data comes from.

Given a data set $\mathbb{X}$ on a metric space (M, ρ), there are many ways of connecting n-tuples of points in order to build simplexes of dimension n and form a simplicial complex. Note that, in practice, we will only deal with finite abstract simplicial complexes. The most used in TDA are the followings. Let $\alpha > 0$ be a real number.

Definition 1.1.1. The **Vietoris-Rips complex** $\mathrm{Rips}_\alpha(\mathbb{X})$ is the collection of simplexes $\{x_0, \dots, x_k\}$ where $x_i \in \mathbb{X}$ are such that $\rho_{\mathbb{X}}(x_i, x_j) < \alpha$ for all (i, j). That is, $\{x_0, \dots, x_k\}$ is a simplex if the open balls $B_\rho(x_i, \alpha/2)$, for $i = 0, \dots, k$, have nonempty pairwise intersection.

Definition 1.1.2. The **Čech complex** $\check{\mathrm{C}}\mathrm{ech}_\alpha(\mathbb{X})$ is the set of simplexes $\{x_0, \dots, x_k\}$ where $x_i \in \mathbb{X}$ are such that $B_\rho(x_0, \alpha) \cap \cdots \cap B_\rho(x_k, \alpha) \neq \emptyset$.

The Čech complex is just a particular case of the nerve simplicial complex (example A.1.8) for the open cover $\{B_\rho(x_i, \alpha)\}_{x_i \in \mathbb{X}}$.

Notice that, even if $\mathbb{X} \subset \mathbb{R}^m$, both complexes can have higher dimension. They are related by

$$\mathrm{Rips}_\alpha(\mathbb{X}) \subseteq \check{\mathrm{C}}\mathrm{ech}_\alpha(\mathbb{X}) \subseteq \mathrm{Rips}_{2\alpha}(\mathbb{X}). \tag{1.1}$$

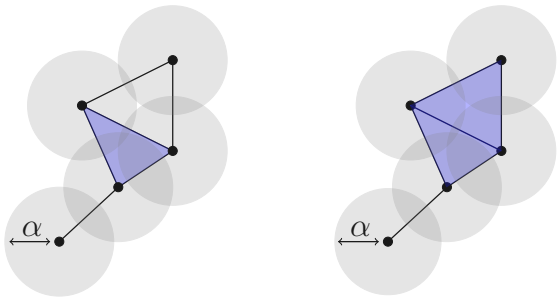


Figure 1.1: Čech complex $\check{\mathrm{C}}\mathrm{ech}_\alpha(\mathbb{X})$ (left) and Rips complex $\mathrm{Rips}_{2\alpha}(\mathbb{X})$(right) .

Remark 1.1.3. Despite having in general more simpleces, the Rips complex Rips_α is computationally cheaper than the corresponding Čech complex $\check{\mathrm{C}}\mathrm{ech}_{\alpha/2}$. Indeed, in order to compute the Čech complex we must remember either the entire complex and its boundary operator (i.e. the number of simpleces of each kind and how to glue them through the boundary operator) or the distances between vertices. Meanwhile, the Rips complex is a flag complex: it is maximal among all simplicial complexes with the given 1-skeleton. Consequently, it is completely determined by the 1-skeleton and can be stored as a graph and reconstituted. Since this text is not focused on the computational aspects, for forms of storage and reconstruction of simplicial complexes we refer to [9], with the simplex tree structure, and to [10] with the tidy set, especially designed for flag complexes. The drawback of Rips complexes is that homotopy equivalence results are not as immediate as those for Čech complexes, as it will be shown. ♢

1.1.1 Mapper algorithm

Another possibility when constructing the simplicial complex associated to a data set $\mathbb{X}$ is to consider subsets of points as vertices, instead of each point being a vertex of the complex. The use of covers and their corresponding nerve complexes is perfect for this. A well chosen cover is able to highlight certain areas of interest, and the associated nerve complex provides a compact description of the relationships between these areas through their intersection patterns. This is useful when data sets are huge and we want to obtain a low dimensional representation easy to understand.

In order to chose a cover according to our needs we can proceed as follows. In the general framework, given a topological space X select a continuous function $f : X \to Z$, where Z is called **parameter space**, and $\mathcal{U} = \{U_i\}_{i\in I}$ an open cover of $f(X)$. Consider the open cover of X given by the inverse images of these open subsets $\{f^{-1}(U_i)\}_{i\in I}$, which is called the **pull back cover** induced by $(f,\mathcal{U})$. Then, if a wise choice of Z, f and $\mathcal{U}$ has been made, the **refined pull back cover**, which is the collection of connected components of the open sets $f^{-1}(U_i)$ for $i \in I$, provides a nerve complex which summarizes the required information.

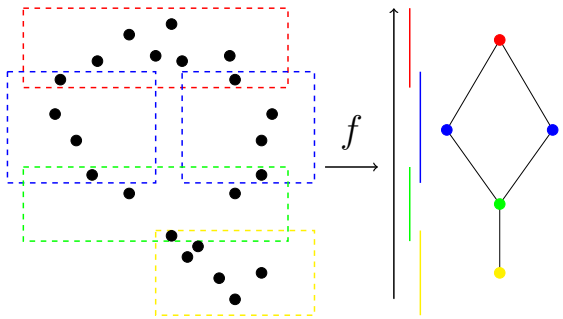


Figure 1.2: The mapper algorithm on a point cloud.

To adapt this to the case in which $\mathbb{X}$ is a point cloud, the **Mapper algorithm** has been developed in [11]. Let $f : \mathbb{X} \to \mathbb{R}^m$ for $m \geq 1$, a function called **filter**. In addition, consider a cover $\mathcal{U}$ of $f(\mathbb{X})$ and compute the pull back cover. The partitioning of the opens of this cover into their connected components has to be substituted for some clustering of the inverse images. Then, we compute the nerve complex generated by these clusters. For a well chosen cover the nerve will be a graph, which has nodes corresponding to clusters in the data, but of course this method could provide higher dimensional complexes. All in all, this algorithm receives as input a data set $\mathbb{X}$, a function $f : \mathbb{X} \to \mathbb{R}^m$ and a cover $\mathcal{U}$ and outputs a simplicial complex. Observe that it does not need the data to be embedded into an Euclidean space.

Some examples of functions f used in the literature are density estimates (may help to understand the structure of high density areas) and distance functions to a given point.

On the other hand, if $f : \mathbb{X} \to \mathbb{R}$, a standard cover choice is a finite collection of intervals $\{I_j\}_{j\in J}$. It can be parametrize by the length and the percentage overlap between successive intervals. For each of them, the inverse image is computed $\mathbb{X}_j :=$

$f^{-1}(I_j) = \{x \in \mathbb{X} \ : \ f(x) \in I_j\}$, providing a cover of $\mathbb{X}$. Then, for each set $\mathbb{X}_j$, we find clusters $\{\mathbb{X}_{jk}\}$ and compute the nerve complex of these clusters. That is, each cluster is a vertex and an edge exists between them if $\mathbb{X}_{jk} \cap \mathbb{X}_{lt} \neq \emptyset$. This method can be generalized for $f : \mathbb{X} \to \mathbb{R}^m$ with $m > 1$.

Finally, notice that when working with data sets a third choice is needed, the clustering method.

Remark 1.1.4. Small changes in the cover can dramatically change the output of the Mapper algorithm, which makes this method inestable. This could be solved by studying some range of the parameters of the cover and selecting the ones that turn out to provide the better output from the user perspective. ♢

1.2 Nerve theorem

All in all, the discrete set of points $\mathbb{X}$ has been transformed into a higher dimensional topological space from which we can obtained topological properties. It would be ideal if both are related, so that the topological features of the simplicial complex lead us to properties of the data set and its underlying structure. The following general statement holds, whose proof can be found in section 4G of [12].

Theorem 1.2.1 (Nerve Theorem)**.** *Let $\mathcal{U} = \{U_i\}_{i \in I}$ be an open cover of a paracompact space X such that the intersection of any finite subcollection of the U_i's is either empty or contractible. Then, X and the nerve $N(\mathcal{U})$ are homotopy equivalent.*

Obviously if $\mathcal{U}$ consists of convex open subsets, the nerve theorem follows. That is the case of open balls in $(\mathbb{R}^m, ||\cdot||_q)$, with $1 \leq q \leq \infty$. Hence, if $\mathbb{X} \subset \mathbb{R}^m$ for some $m > 0$, $\check{\text{C}}\text{ech}_\alpha(\mathbb{X})$ is homotopy equivalent to the union of balls $\bigcup_{x \in \mathbb{X}} B(x, \alpha)$. We expect that, for some radius α, this union reflects the underlying structure of the data set. However, choosing the right radius is not an easy task. If α is too small no new information is obtained, whereas if it is too large all the balls will overlap. As a consequence, the best approach is not to chose just a value for α but some range of values and study how the cover and the simplicial complex evolve.

With the same idea, for a well chosen cover, the mapper algorithm also preserves the topological structure of the original space.

The following extension of the nerve theorem will be needed.

Theorem 1.2.2 (Persistent Nerve Theorem)**.** *Let $X \subseteq X'$ be two paracompact spaces and let $\mathcal{U} = \{U_i\}_{i \in I}$ and $\mathcal{U}' = \{U'_j\}_{j \in J}$ be two open covers, of X and X' respectively. Suppose that these covers satisfy that the intersection of any finite subcollection of open subsets is either empty or contractible. Assume moreover that they are based on finite parameter sets $I \subseteq J$ such that $U_i \subseteq U'_i$ for all $i \in I$.*

Then, the homotopy equivalences $N(\mathcal{U}) \to X$ and $N(\mathcal{U}') \to X'$ provided by the nerve theorem commute with the canonical inclusions $X \hookrightarrow X'$ and $N(\mathcal{U}) \hookrightarrow N(\mathcal{U}')$ at homology level.

The proof is similar to the one of the nerve theorem, noticing that the canonical inclusions commute with the homotopy equivalences defined there. Anyway, it can be found in Lemma 3.4 of [13].

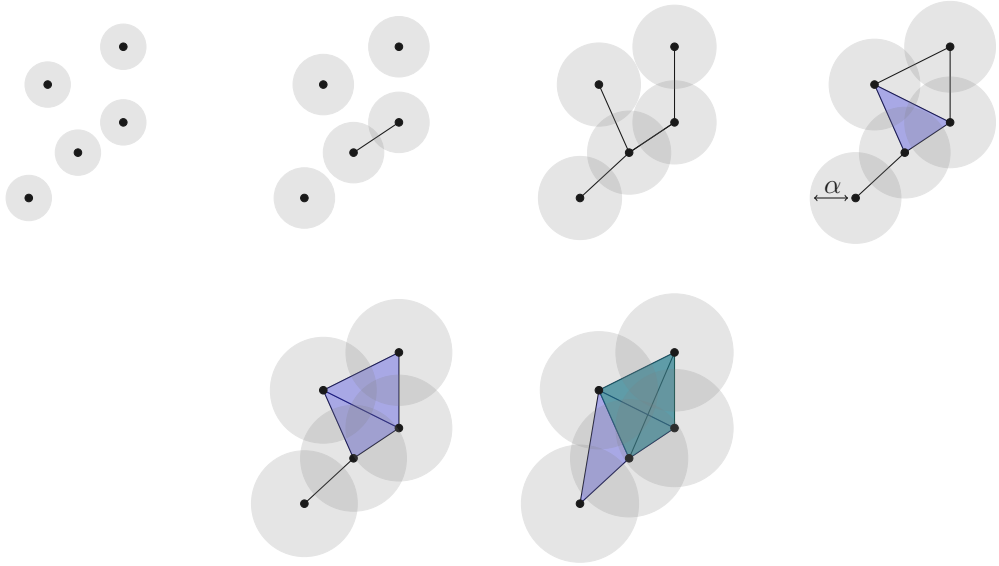


Figure 1.3: Sequence of Čech complexes with increasing radius α.

1.3 Reconstruction theorem

Assume that the data set $\mathbb{X}$ is a finite discrete subset of $\mathbb{R}^m$, with the euclidean distance d, sampled i.i.d according to a probability measure μ with compact support $M \subset \mathbb{R}^m$. If $\mathbb{X}$ and M are close we expect to be able to infer topological features of M from $\mathbb{X}$. This notion of closeness between compact sets is formalized as follows. Let $(\mathcal{A}, \rho)$ be a metric space and, for every $K \subset \mathcal{A}$, consider the distance function $\rho_K : \mathcal{A} \to \mathbb{R}_+$ defined by $\rho_K(x) = \inf_{y \in K} \rho(x, y)$.

Definition 1.3.1. The **Hausdorff distance** between two compact subsets K, K' of a metric space $(\mathcal{A}, \rho)$ is defined as

$$\begin{aligned} d_H(K, K') &= \max\{ \sup_{x' \in K'} \rho_K(x'),\ \sup_{x \in K} \rho_{K'}(x)\} \\ &= \sup_{x \in \mathcal{A}} |\rho_K(x) - \rho_{K'}(x)| = ||\rho_K - \rho_{K'}||_\infty. \end{aligned}$$

From the previous section, the cover of $\mathbb{X}$ by open balls centered on the points of the data set is homotopy equivalent to the Čech complex. We are going to see that, under some regularity assumptions on M, one can relate the topology of this union of balls to the one of M. Therefore, by studying the Čech complex of $\mathbb{X}$ we may obtain topological properties of M.

Definition 1.3.2. Let $K \subset \mathcal{A}$ be a compact subset of a metric space $(\mathcal{A}, \rho)$ and $r > 0$ a real number. The union of closed balls of radius r centered on K is called the r-**offset** of K and denoted by

$$K^r = \bigcup_{x \in K} \bar{B}_\rho(x, r).$$

Notice that K^r is just the r-**sublevel set** of the distance function. That is,

$$K^r = \{x \in \mathcal{A} \ : \ \rho_K(x) \leq r\} = \rho_K^{-1}([0, r]).$$

Let $\{B(x, \alpha)\}_{x \in \mathbb{X}}$ be a cover of the finite set $\mathbb{X}$ by open balls for the euclidean distance d. In general, it is possible to find $0 < r < \alpha$ close enough to α such that the two subsets $\cup_{x \in \mathbb{X}} B(x, \alpha)$ and $\cup_{x \in \mathbb{X}} \bar{B}(x, r)$ are homotopy equivalent and, moreover, define the same Čech complex. But the latter union is just $\mathbb{X}^r$. As a result, the Čech complex of $\mathbb{X}$ is homotopy equivalent to $\mathbb{X}^r$. Hence, if we relate the r-offsets of M with those of $\mathbb{X}$, just by studying the Čech complex of the data set we could obtain topological information about M.

All in all, it seems useful to understand the topological behaviour of the r-sublevel sets of distance functions of compact subsets. In fact, the statements are true in a more general framework.

Definition 1.3.3. A continuous function $\phi : \mathbb{R}^m \to \mathbb{R}_+$ is **distance-like** if it is proper, i.e. the preimage of any compact set is a compact set, and ϕ^2 is semiconcave, i.e. $x \mapsto ||x||^2 - \phi^2(x)$ is convex.

For any compact set $K \subset \mathbb{R}^m$ the distance function d_K is distance-like.

Remark 1.3.4. Due to semiconcavity distance-like functions have a well defined (but not continuous) notion of gradient $\nabla\phi : \mathbb{R}^m \to \mathbb{R}^m$. This is essential to introduce the next notions. However, the study of semiconcave and distance-like functions is outside the scope of this text. Hence, we refer to [14] and [15] for more information. ♢

Definition 1.3.5. Let ϕ be a distance-like function and let $\phi^r = \phi^{-1}([0, r])$ be the r-sublevel set of ϕ.

- A point $x \in \mathbb{R}^m$ is α-**critical** if $||\nabla_x \phi|| \leq \alpha$. The corresponding value $\phi(x)$ is said to be an α-critical value. A 0-critical point (value) is simply called a critical point (value).
- The **weak feature size** of ϕ at r, denoted by $\text{wfs}_\phi(r)$, is the maximum $r' > 0$ such that ϕ does not have any critical value between r and $r + r'$.
- For any $0 < \alpha < 1$ the α-**reach** of ϕ, denoted by $\text{reach}_\alpha(\phi)$, is the maximum r such that $\phi^{-1}((0, r])$ does not contain any α-critical point. The $\text{reach}_\alpha(\phi)$ is always a lower bound for the weak feature size with $r = 0$.

A relevant result involving distance-like functions is the ensuing isotopy lemma, which is proven in [16], proposition 1.8.

Definition 1.3.6. Let $X, Y \subseteq \mathbb{R}^m$. We say that they are **ambient isotopic** if there exists a continuous map $H : [0, 1] \times \mathbb{R}^m \to \mathbb{R}^m$ such that $H_t = H(t, \cdot)$ is a homeomorphism for every $t \in [0, 1]$, $H_0 = \text{id}_{\mathbb{R}^m}$ and $H_1(X) = Y$.

Lemma 1.3.7 (Isotopy Lemma)**.** *Let ϕ be a distance-like function and $r_1 < r_2$ two positive real numbers such that ϕ does not have critial points in the subset $\phi^{-1}([r_1, r_2])$. Then, all ϕ^r are isotopic for $r \in [r_1, r_2]$. In particular, inclusions $\phi^r \hookrightarrow \phi^{r'}$ for $r < r'$, $r, r' \in [r_1, r_2]$, are homotopy equivalences.*

Therefore, the topology of the r-sublevel sets of distance-like functions only changes when r crosses a critical value. Then, all the sublevel sets of ϕ between r and $r+\text{wfs}_\phi(r)$ have the same topology.

Once we have this, the next step is to obtain a connection between the sublevel sets of two different distance-like functions, in our case d_M and $d_{\mathbb{X}}$.

Theorem 1.3.8 (Reconstruction Theorem [17])**.** *Let ϕ, ψ be two distance-like functions such that $||\phi - \psi||_\infty \leq \varepsilon$ and $\text{reach}_\alpha(\phi) \geq R$ for some $\varepsilon > 0$ and $\alpha \in (0,1)$. Then, for every $r \in [4\varepsilon/\alpha^2, R - 3\varepsilon]$ and every $\eta \in (0, R)$, if $\varepsilon < \frac{R}{5+4/\alpha^2}$, the sublevel sets ψ^r and ϕ^η are homotopy equivalent.*

We need these previous results to prove the Reconstruction theorem.

Proposition 1.3.9 ([17])**.** *Let ϕ, ψ be two distance-like functions such that $||\phi - \psi||_\infty \leq \varepsilon$. Assume that the weak feature sizes satisfy $\text{wfs}_\phi(r) > 2\varepsilon$ and $\text{wfs}_\psi(r) > 2\varepsilon$ for some $r > 0$. Then, ϕ^η and ψ^η have the same homotopy type for all $\eta \in (r, r+2\varepsilon]$.*

Proof. Consider $\delta > 0$ such that $\text{wfs}_\phi(r) > 2\varepsilon + \delta$ and $\text{wfs}_\psi(r) > 2\varepsilon + \delta$. Then, $||\phi - \psi||_\infty \leq \varepsilon$ provides the following diagram, where each map is an inclusion.

$$\begin{array}{ccccc} \phi^{r+\delta} & \xrightarrow{f_0} & \phi^{r+\delta+\varepsilon} & \xrightarrow{f_1} & \phi^{r+\delta+2\varepsilon} \\ & \searrow^{c} \quad \nearrow & & \searrow^{d} \quad \nearrow & \\ \psi^{r+\delta} & \xrightarrow{g_0} & \psi^{r+\delta+\varepsilon} & \xrightarrow{g_1} & \psi^{r+\delta+2\varepsilon} \end{array}$$

Moreover, it is commutative, being the horizontal arrows f_0, f_1, g_0, g_1 homotopy equivalences thanks to the isotopy lemma 1.3.7. Denote their homotopic inverses by f_1', f_1', g_0', g_1' respectively. Then, a straightforward computation shows that $c : \psi^{r+\delta+\varepsilon} \to \phi^{r+\delta+2\varepsilon}$ is a homotopy equivalence with inverse $g_1' \circ d \circ f_1'$. ∎

Proposition 1.3.10 ([17])**.** *Let ϕ, ψ be two distance-like functions such that $||\phi - \psi||_\infty \leq \varepsilon$. Given x any α-critical point of ϕ, there exist x' an α'-critical point of ϕ such that*

$$||x - x'|| \leq 2\sqrt{\varepsilon\phi(x)}, \qquad \alpha' - \alpha \leq 2\sqrt{\varepsilon/\phi(x)}.$$

In addition, if $\text{reach}_\alpha(\phi) \geq R$ for some α, then ψ has no critical values in the interval $[4\varepsilon/\alpha^2, R - 3\varepsilon]$.

Proof of Reconstruction theorem 1.3.8. First of all, notice that

$$\varepsilon < \frac{R}{4/\alpha^2 + 5} < \frac{R}{4/\alpha^2 + 3} \Leftrightarrow \frac{4\varepsilon}{\alpha^2} < R - 3\varepsilon,$$

so the interval for r is nonempty. By definition of α-reach $(0, R]$ does not contain critical values for ϕ. Applying the isotopy lemma this means that all ϕ^η have the same homotopy type for $\eta \in (0, R]$. Therefore, it suffices to prove the result for a fix η. Let us choose $\eta = 4\varepsilon/\alpha^2$. It is clear that $\text{wfs}_\phi(\eta) \geq R - 4\varepsilon/\alpha^2$. Using proposition 1.3.10, ψ does not have critical values in the interval $[4\varepsilon/\alpha^2, R - 3\varepsilon]$, which implies that $\text{wfs}_\psi(\eta) \geq R - 3\varepsilon - 4\varepsilon/\alpha^2$. Again, isotopy lemma states that all ψ^r have the same homotopy type for $r \in [4\varepsilon/\alpha^2, R - 3\varepsilon]$. This means that, in order to finish the

proof, it is enough to show that $\phi^{r'}$ and $\psi^{r'}$ are homotopy equivalent for some r' in the given range.

Indeed, because of how we have chosen $\varepsilon > 0$, it satisfies

$$5\varepsilon + 4\varepsilon/\alpha^2 < R \Leftrightarrow R - 3\varepsilon - 4\varepsilon/\alpha^2 > 2\varepsilon.$$

By proposition 1.3.9 we can conclude that $\phi^{r'}$ is homotopy equivalent to $\psi^{r'}$ for every $r' \in (\eta, \eta + 2\varepsilon] \subset (4\varepsilon/\alpha^2, R - 3\varepsilon)$. ■

In our setting $\phi = d_M$ and $\psi = d_{\mathbb{X}}$. This means that if M and $\mathbb{X}$ are close with respect to the hausdorff distance, for well chosen values of r and η, M^η is homotopy equivalent to $\mathbb{X}^r$, which is again homotopy equivalent to $\check{\text{C}}\text{ech}_r(\mathbb{X})$. All in all, with the nerve theorem and the reconstruction theorem we are able to infer the topology of a compact space M from a simplicial complex $\check{\text{C}}\text{ech}_r(\mathbb{X})$ build on top of an approximate finite sample $\mathbb{X}$ of it. This is what we were looking for. However, there are some practical problems:

- The condition $\text{reach}_\alpha(\phi) = \text{reach}_\alpha(d_M) \geq R$, which could be interpreted as a regularity condition, may not always be satisfied.
- The choice of r is left to the user.
- Although the Čech complex provides an accurate topological summary of the data, it is not well-suited for data processing, as we have already mentioned in remark 1.1.3.

To overcome this last problem we only need to find topological descriptors that are simpler to handle and which could be easily computed from a complex. Algebraic topology, specially homology and, in particular, persistent homology, due to its balance between ease of computation and topological resolution is the answer. We will only consider homology groups over fields. For more information about computational aspects we refer to [7] and [8]. Besides, as it will be seen in the next chapter, this also let us overcome the second problem: persistent homology is able to handle homology at different scales simultaneously.

Having our eye on homology we recover the practical setting where $\mathbb{X} \subset \mathbb{R}^m$ is any finite set of points. Whenever $\mathbb{X}^r$ and $\check{\text{C}}\text{ech}_r(\mathbb{X})$ are homotopy equivalent their homology groups are isomorphic, and hence they have same Betti numbers. Therefore, the reconstruction theorem can be reformulated.

Theorem 1.3.11 ([1])**.** *Consider $M \subset \mathbb{R}^m$ be a compact set such that $\text{reach}_\alpha(d_M) \geq R > 0$ for some $\alpha \in (0, 1)$. Let $\mathbb{X}$ be a finite set of points such that*

$$\varepsilon = d_H(M, \mathbb{X}) < \frac{R}{5 + 4/\alpha^2}.$$

Then, for every $r \in [4\varepsilon/\alpha^2, R - 3\varepsilon]$ and every $\eta \in (0, R)$, the Betti numbers of $\check{\text{C}}\text{ech}_r(\mathbb{X})$ are equal to the Betti numbers of M^η.

Proof. The hypothesis are just the ones from the reconstruction theorem with $\phi = d_M$, $\psi = d_{\mathbb{X}}$ and $d_H(M, \mathbb{X}) = ||\phi - \psi||_\infty = \varepsilon$. Then, we use the fact that being homotopy equivalent implies having isomorphic homology groups. ■

Although it may happen that the homology groups of the sublevel sets M^η are different from those of M, this is a rare behaviour which does not included most practical cases. Hence, the above theorem provides a method to estimate the Betti numbers of M.

We still have to deal with the problem of the condition on the α-reach of M and the computation of the Čech complex. The following theorem gives a response to this. Basically, the condition on the α-reach is exchanged for a similar condition on the weak feature size, less strict in terms of regularity, and Rips complexes substitute Čech complexes. In the following denote $\mathbb{X}^\alpha = \bigcup_{x\in\mathbb{X}} B(x, \alpha) = d_{\mathbb{X}}^{-1}((-\infty, \alpha))$. Recall that, by the nerve theorem, the homology group of $\mathbb{X}^\alpha$ is isomorphic to that of $\check{\text{C}}\text{ech}_\alpha(\mathbb{X})$. The latter is a finite simplicial complex, so it has a finite dimensional homology group.

Theorem 1.3.12 ([13])**.** *Let $M \subset \mathbb{R}^m$ be a compact set such that $\text{wfs}(M) = \text{wfs}_{d_M}(0) \geq R > 0$ and let $\mathbb{X}$ be a finite set of points such that $d_H(M, \mathbb{X}) < \varepsilon$ for some $\varepsilon < \frac{1}{9}\text{wfs}(M)$. Then, for any $\alpha \in \left[2\varepsilon, \frac{1}{4}(\text{wfs}(M) - \varepsilon)\right]$ and any $\eta \in (0, R)$,*

$$\beta_n(M^\eta) = \text{rk}(j_*) \qquad \forall n \in \mathbb{N}, \tag{1.2}$$

where

$$j_* : H_n(\text{Rips}_\alpha(\mathbb{X})) \to H_n(\text{Rips}_{4\alpha}(\mathbb{X}))$$

is the homomorphism induced by the canonical inclusion $j : \text{Rips}_\alpha(\mathbb{X}) \hookrightarrow \text{Rips}_{4\alpha}(\mathbb{X})$.

We need two previous lemmas, being the first one a linear algebra standard result.

Lemma 1.3.13. *Given a sequence $A \to B \to C \to D \to E \to F$ of homomorphisms between finite dimensional vector spaces, if $\text{rk}(A \to F) = \text{rk}(C \to D)$, then it is also equal to $\text{rk}(B \to E)$. Analogously, if $A \to B \to C \to D \to E$ is a sequence of homomorphisms such that $\text{rk}(A \to E) = \dim C$, then $\text{rk}(B \to D) = \dim C$.*

Lemma 1.3.14 ([13])**.** *Let $M \subset \mathbb{R}^m$ be a compact set and $\mathbb{X} \subset \mathbb{R}^m$ a finite set of points such that $d_H(M, \mathbb{X}) < \varepsilon$ for some $\varepsilon < \frac{1}{4}\text{wfs}(M)$. Then, for all $\alpha, \alpha' \in [\varepsilon, \text{wfs}(M) - \varepsilon]$ such that $\alpha' - \alpha \geq 2\varepsilon$, and for all $\eta \in (0, \text{wfs}(M))$ we have the isomorphism*

$$H_n(M^\eta) \simeq \text{im}(i_*) \qquad \forall n \in \mathbb{N},$$

where $i_* : H_n(\mathbb{X}^\alpha) \to H_n(\mathbb{X}^{\alpha'})$ is the homomorphism induced by the canonical inclusion $i : \mathbb{X}^\alpha \hookrightarrow \mathbb{X}^{\alpha'}$.

Proof. Observe that by hypothesis $4\varepsilon < \text{wfs}(M)$, so the interval $[\varepsilon, \text{wfs}(M) - \varepsilon]$ does contain values α, α' that are at distance greater than 2ε. Without loss of generality assume that $\varepsilon < \alpha < \alpha' - 2\varepsilon < \text{wfs}(M) - 3\varepsilon$. Notice that this is always possible, since we can change ε for any other $\varepsilon' \in (d_H(M, \mathbb{X}), \varepsilon)$ and, once proven for this case,

the statement for ε follows. The inequality $d_H(M, \mathbb{X}) = ||d_M - d_{\mathbb{X}}||_\infty < \varepsilon$ provides a sequence of inclusions

$$M^{\alpha-\varepsilon} \hookrightarrow \mathbb{X}^{\alpha} \hookrightarrow M^{\alpha+\varepsilon} \hookrightarrow \mathbb{X}^{\alpha+2\varepsilon} \hookrightarrow \mathbb{X}^{\alpha'} \hookrightarrow M^{\alpha'+\varepsilon}. \tag{1.3}$$

This sequence transforms into a sequence of homomorphisms between homology groups, which are vector spaces. Thanks to the isotopy lemma and the fact that $\alpha' + \varepsilon < \text{wfs}(M)$, all the inclusions between $M^{\alpha-\varepsilon}, M^{\alpha+\varepsilon}$ and $M^{\alpha'+\varepsilon}$ induce isomorphisms. Hence, for any $n \in \mathbb{N}$,

$$\text{rk}(H_n(M^{\alpha-\varepsilon}) \to H_n(M^{\alpha'+\varepsilon})) = \dim H_n(M^{\alpha'+\varepsilon}) = \dim H_n(M^{\alpha+\varepsilon}). \tag{1.4}$$

In addition, the homology groups of $\mathbb{X}^{\alpha}, \mathbb{X}^{\alpha'}$ are finite dimensional. We can use then lemma 1.3.13 and deduce from equality (1.4) that

$$\text{rk}(i_*) = \text{rk}(H_n(\mathbb{X}^{\alpha}) \to H_n(\mathbb{X}^{\alpha'})) = \dim H_n(M^{\alpha+\varepsilon}).$$

However, for any $\eta \in (0, \text{wfs}(M))$ isotopy lemma gives an homotopy equivalence between M^{η} and $M^{\alpha+\varepsilon}$. That is, $\dim H_n(M^{\eta}) = \dim H_n(M^{\alpha+\varepsilon})$ and then equal to $\text{rk}(i_*)$. By finiteness of these vector spaces we can conclude that $\text{im}(i_*) \simeq H_n(M^{\eta})$. ■

Proof of theorem 1.3.12. First of all notice that $\varepsilon < \frac{1}{9}\text{wfs}(M)$ guarantees that the interval $[2\varepsilon, \frac{1}{4}(\text{wfs}(M) - \varepsilon)]$ is nonempty. Fix α in this interval. By equation (1.1) we have the next sequence of inclusions

$$\check{\text{C}}\text{ech}_{\alpha/2}(\mathbb{X}) \hookrightarrow \text{Rips}_{\alpha}(\mathbb{X}) \hookrightarrow \check{\text{C}}\text{ech}_{\alpha}(\mathbb{X}) \hookrightarrow \check{\text{C}}\text{ech}_{2\alpha}(\mathbb{X}) \hookrightarrow \text{Rips}_{4\alpha}(\mathbb{X}) \hookrightarrow \check{\text{C}}\text{ech}_{4\alpha}(\mathbb{X}), \tag{1.5}$$

which turns into a sequence of homomorphisms between finite vector spaces at the homology level. Consider the pairs (α_1, α_2) given by $(\alpha/2, 4\alpha)$ and $(\alpha, 2\alpha)$. By the persistent nerve theorem 1.2.2 we have the commutative diagrams

$$\begin{array}{ccc} H_n(\check{\text{C}}\text{ech}_{\alpha_1}(\mathbb{X})) & \xrightarrow{k_*} & H_n(\check{\text{C}}\text{ech}_{\alpha_2}(\mathbb{X})) \\ \uparrow & & \uparrow \\ H_n(\mathbb{X}^{\alpha_1}) & \xrightarrow{i_*} & H_n(\mathbb{X}^{\alpha_2}) \end{array}$$

where the vertical arrows are isomorphisms and the horizontal ones are induced by inclusions. Since $\varepsilon \leq \alpha_1 < \alpha_2 \leq \text{wfs}(M) - \varepsilon$ and $\alpha_2 - \alpha_1 \geq 2\varepsilon$, lemma 1.3.14 can be applied for the pairs (α_1, α_2). As a result, for any $n \in \mathbb{N}$ and $\eta \in (0, R)$, $\dim H_n(M^{\eta}) = \text{rk}(i_*) = \text{rk}(k_*)$, where the last equality follows from the commutative diagram. In particular, for the pairs of parameters chosen we obtain

$$\begin{aligned} \text{rk}(H_n(\check{\text{C}}\text{ech}_{\alpha/2}(\mathbb{X})) \to H_n(\check{\text{C}}\text{ech}_{4\alpha}(\mathbb{X}))) &= \dim H_n(M^{\eta}) \\ &= \text{rk}(H_n(\check{\text{C}}\text{ech}_{\alpha}(\mathbb{X})) \to H_n(\check{\text{C}}\text{ech}_{2\alpha}(\mathbb{X}))) \end{aligned}$$

This, together with the sequence of homomorphisms that comes from (1.5) and lemma 1.3.13, leads to $\dim H_n(M^{\eta}) = \text{rk}(H_n(\text{Rips}_{\alpha}(\mathbb{X})) \to H_n(\text{Rips}_{4\alpha}(\mathbb{X})))$ as wanted. ■

Notice that j_* relates the homology of two Vietoris-Rips complexes of the same data set but on different scales. In fact, $\text{Rips}_\alpha(\mathbb{X}) \subseteq \text{Rips}_{4\alpha}(\mathbb{X})$. Therefore, our next step should be to study the homology groups of nested simplicial complexes, for several values of r, and how they are related. This connects also with the problem of the choice of the scale r: it is a better strategy to consider some range of values. If we could encode this information we would be able to explain how homology evolves throughout the sequence of complexes: homology classes are born, can merge, split, die or persist, some are essential (those that persist over a significant parameter range), others might be omitted (those that die just after being born), ...

Chapter 2

Persistent homology

The main objective of this chapter is to explain how we can encode multiscale topological properties of a nested family of simplicial complexes by using the correspondent sequence of homology groups. The study of homology at different scales simultaneously is called persistent homology and it is done using persistence modules. This information can be summarized and visualized easily by introducing persistence diagrams or barcodes. Although the main works on persistent theory ([7], [18], [8]) assume finiteness on the nested family of simplicial complexes, which is understandable from a computational point of view, that restriction can be lifted. One of the reasons why a contiuous approach is desirable is that, ideally, the persistent homology of a sample will be an approximation to the persistent homology of its continuous underlying structure. We will focus on this last approach by following [4].

2.1 Filtrations

Definition 2.1.1. A **filtration** of a topological space X is a nested family of subspaces $(X_r)_{r\in T}$, where $T \subseteq \mathbb{R}$, such that for any $r, r' \in T$ with $r \leq r'$, then $X_r \subseteq X_{r'}$. Moreover, $X = \cup_{r\in T} X_r$.

In particular, when X is a simplicial complex, the subspaces X_r are considered to be subcomplexes. The usual filtrations are the following families.

1. When working with a data set $\mathbb{X}$, families of Vietoris-Rips complexes $(\mathrm{Rips}_r(\mathbb{X}))_{r\in\mathbb{R}}$ and Čech complexes $(\check{\mathrm{C}}\mathrm{ech}_r(\mathbb{X}))_{r\in\mathbb{R}}$. We assume $\mathrm{Rips}_r(\mathbb{X}) = \check{\mathrm{C}}\mathrm{ech}_r(\mathbb{X}) = \emptyset$ for $r < 0$. The parameter r could be understood as the scale at which we consider the data.

2. In general, given a topological space X and a continuous map $f : X \to \mathbb{R}$, the family of sublevel sets $X_r = f^{-1}((-\infty, r])$.

3. In the case of a simplicial complex K, with set of vertices V, the sublevel set filtration can be defined as follows. Let $f : V \to \mathbb{R}$ be a map defined on the vertices and extend it to all simplexes of K by $f([p_0, \ldots, p_n]) = \max\{f(p_i) \ : \ i = 0, \ldots, n\}$. Then, the family of subcomplexes given by the r-sublevel sets of f defines a filtration $(K_r)_{r\in\mathbb{R}}$, $K_r = \{\sigma \in K \ : \ f(\sigma) \leq r\}$.

Given a filtration, when the parameter r changes the homology of the subspaces X_r changes too. For example, concerning the zero homology groups (which are directly related to the number of path connected components), while r changes new path connected components could appear or existing ones could disappear (e.g. when two of them merge into one). We could say that the life of a path connected component C begins when it appears in the filtration and ends when it merges with a previous one. If this happens at X_{r_1} and X_{r_2} respectively, we say that the life time of C is the interval $[r_1, r_2)$. As a result, the information about the zero homology groups of the filtration is encoded as a set of intervals. This is called a persistent barcode. Moreover, these intervals can be interpreted as points in $\mathbb{R}^2$ obtaining a multiset[1] called persistence diagram. Obviously, this could be made for higher dimensional homology groups, tracking the changes of cycles and boundaries.

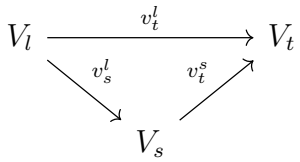
This idea of persistence barcodes and diagrams can be formally and rigorously defined in an algebraic framework using persistence modules.

2.2 Persistence modules

Let $(T, \leq)$ be any poset and k a field.

Definition 2.2.1. A **persistence module** $\mathbb{V}$ over T is functor from T, considered a category in the usual way, to the category of vector spaces.

Thus, it consists of an indexed family of k-vector spaces $(V_t)_{t\in T}$ and a doubly-indexed family of linear maps $(v_t^s : V_s \to V_t)_{s\leq t}$ that satisfy $v_t^s \circ v_s^l = v_t^l$, whenever $l \leq s \leq t$, and $v_t^t = id_{V_t}$.



In general, $T \subseteq \mathbb{R}$. When T is finite $\{t_0 < t_1 < \cdots < t_n\}$ the persistence module can be expressed as the sequence $V_{t_0} \to V_{t_1} \to \cdots \to V_{t_n}$.

Note that, given any filtration of a topological space X, a persistence module can be obtained by applying to it any functor from the category of topological spaces to the category of vector spaces. In particular, $H_n(\cdot, k)$ for k a field and any $n \in \mathbb{N}$, obtaining a persistence module of homology groups.

Definition 2.2.2. Consider a filtration $(X_t)_{t\in T}$ of a topological space X. Fix $n \in \mathbb{N}$. For any $s < t$, the **(s,t)-persistent homology group** of X is

$$H_n^{s\to t}(X) = \mathrm{im}(H_n(X_s, k) \to H_n(X_t, k)) = \frac{Z_n(X_s, k)}{Z_n(X_s, k) \cap B_n(X_t, k)}.$$

Its dimension is called **persistence Betti number** $\beta_t^s := \dim H_n^{s\to t}(X)$.

[1] A multiset is a set where points can be repeated.

A homology class in $H_n(X_t)$ is said to persist if its image in $H_n(X_{t+1})$ is also nonzero; otherwise it is said to die. A homology class in $H_n(X_t)$ is said to be born when it is not in the image of $H_n(X_{t-1})$. Hence, β_t^s can be interpreted as the number of independent homology classes of $H_n(X_t)$ born before $H_n(X_s)$.

Remark 2.2.3. It can happen that, given a persistence module $(V_t)_{t\in T}$, the vector spaces change only for some finite values of t, and they stay the same between these values. That is, if these values are $t_1 < t_2 < \cdots < t_n$, which are called **critical values**, then

$$\begin{aligned} &V_t = V_{t_1}, \quad t \leq t_1, \\ &V_t = V_{t_i}, \quad t_{i-1} < t \leq t_i \ \text{ for } i \in \{2, \dots, n\}, \\ &V_t = V_{t_n}, \quad t \geq t_n. \end{aligned}$$

Then, all the information is contained in the finite sequence $V_{t_1} \to V_{t_2} \to \cdots \to V_{t_n}$. ♢

Definition 2.2.4. A **homomorphism** ϕ between two persistence modules $\mathbb{V}$ and $\mathbb{U}$ over T is a natural transformation between functors. That is, a collection of linear maps $(\phi_t)_{t\in T}$ where $\phi_t : U_t \to V_t$ and such that $\phi_t \circ u_t^s = v_t^s \circ \phi_s$ for all $s \leq t$.

$$\begin{array}{ccc} U_s & \xrightarrow{u_t^s} & U_t \\ \downarrow{\scriptstyle \phi_s} & & \downarrow{\scriptstyle \phi_t} \\ V_s & \xrightarrow{v_t^s} & V_t \end{array} \tag{2.1}$$

Remark 2.2.5. Thanks to the above definition we can consider the category of persistence modules, which is actually an abelian category. ♢

Example 2.2.6. Let $\mathbb{X} \subset (\mathbb{R}^m, ||\cdot||_q)$ be a finite set of points and consider the filtration given by the Čech complexes $(\check{\text{C}}\text{ech}_\alpha(\mathbb{X}))_{\alpha\in\mathbb{R}_+}$. For each $n \in \mathbb{N}$ their n-th homology groups define a persistence module $(H_n(\check{\text{C}}\text{ech}_\alpha(\mathbb{X})))_{\alpha\in\mathbb{R}_+}$, where the linear maps are induced by the inclusions. Consider also the filtration $(X^\alpha)_{\alpha\in\mathbb{R}_+}$ for $\mathbb{X}^\alpha = \bigcup_{x\in\mathbb{X}} B(x, \alpha)$. Likewise, the n-th homology groups and the induced homomorphisms from the inclusions define a persistence module $(H_n(\mathbb{X}^\alpha))_{\alpha\in\mathbb{R}_+}$. Then, the persistent nerve theorem 1.2.2 gives an isomorphism between them. ♢

In order to define persistence diagrams, and to measure the life time of the appearing features, we need to introduce a special kind of persistence modules called interval modules.

Definition 2.2.7. An **interval** in a totally ordered set T is a subset $J \subseteq T$ such that, if $r \in J$ and $t \in J$ with $r < s < t$, then $s \in J$ too.

Definition 2.2.8. Given an interval in a totally ordered set $J \subseteq T$, the **interval module** $\mathbb{I}$ over T, denoted by k_T^J or simply k^J, is the persistence module defined by

$$I_t = \begin{cases} k & t \in J, \\ 0 & \text{otherwise}, \end{cases} \qquad i_t^s = \begin{cases} id_k & s, t \in J, \\ 0 & \text{otherwise}. \end{cases}$$

Remark 2.2.9. When T is a finite set $\{t_0 < t_1 < \cdots < t_n\}$ intervals over T are normally written as $J = [t_i, t_j]$. Hence, k^J is usually denoted by $k[t_i, t_j]$. It can be visualized as the sequence

$$0_{t_0} \to \cdots \to 0_{t_{i-1}} \to k_{t_i} \to k_{t_{i+1}} \to \cdots \to k_{t_j} \to 0_{t_{j+1}} \to \cdots \to 0_{t_n}.$$

♢

This type of persistence modules may be interpreted as the life time of a feature, which persists over the interval J and nowhere else. Consequently, if a persistence module could be decompose into interval modules, this decomposition would give us an idea of the life time of the different features of the module. These and the following results related to interval modules entitle them to be considered the building blocks of persistence.

Definition 2.2.10. Given two persistence modules $\mathbb{U}$ and $\mathbb{V}$ over T, we can define the **direct sum** $\mathbb{W} = \mathbb{U} \oplus \mathbb{V}$ (over T) as $W_t = U_t \oplus V_t$ and $w_t^s = u_t^s \oplus v_t^s$. We say that $\mathbb{W}$ is idecomposable if the only decompositions are the trivial ones, i.e. $\mathbb{W} = \mathbb{W} \oplus 0 = 0 \oplus \mathbb{W}$.

This generalizes to arbitrary direct sums.

Proposition 2.2.11 ([4])**.** *Let $J \subseteq T$ be an interval and $\mathbb{I} = k_T^J$ an interval module over T. Then,*

1. $\mathrm{End}(\mathbb{I}) = k$.
2. *Interval modules are idecomposable.*

Proof. 1. Let ϕ be an endomorphism of $\mathbb{I}$. For each nonzero $I_t = k$, i.e. $t \in J$, $\phi_t \in \mathrm{End}(k)$ so it acts on I_t by scalar multiplication. Besides, it also satisfies $\phi_t \circ i_t^s = i_t^s \circ \phi_s$ for all $s \leq t$. If both $s, t \in J$ then $i_t^s = id_k$, which implies that it is the same scalar for each $t \in J$. All in all, $\mathrm{End}(\mathbb{I}) = k$.

2. Suppose that $\mathbb{I}$ decomposes as $\mathbb{U} \oplus \mathbb{V}$. Consider the projections $\mathbb{U} \oplus \mathbb{V} \to \mathbb{U} \oplus \mathbb{V}$ onto $\mathbb{U}$ and $\mathbb{V}$. It is clear that they are idempotent[2] endomorphisms of $\mathbb{U} \oplus \mathbb{V} = \mathbb{I}$. However, we have just proved that $\mathrm{End}(\mathbb{I}) = k$, where the only idempotent elements are zero and the identity. We conclude that $\mathbb{I}$ can only decompose as $\mathbb{I} \oplus 0$ or $0 \oplus \mathbb{I}$, so it is idecomposable.

■

Given an indexed family of intervals $(J_l)_{l \in L}$ we can define the persistence module

$$\mathbb{V} = \bigoplus_{l \in L} k^{J_l}, \tag{2.2}$$

which can be thought as having an independent feature for each $l \in L$ that persist over the interval J_l. We could say then that each feature appears in the filtration at time $\inf(J_l)$ and disappears at time $\sup(J_l)$. That is why the left endpoint is also called birth time, while the right endpoint is called death time.Therefore, the next step is to study which persistence modules can be decompose as (2.2). Hereinafter all persistence modules will be considered over subsets of the real numbers $\mathbb{R}$.

[2] An endomorphism f is idempotent if $f \circ f = f$.

Theorem 2.2.12 ([4]). *Let $\mathbb{V}$ be a persistence module over $T \subseteq \mathbb{R}$. It can be decomposed as a direct sum of interval modules if one of the following is satisfied:*

1. *T is a finite set,*
2. *each V_t is finite-dimensional.*

In that case, we say that $\mathbb{V}$ is decomposable.

Proof. We are going to sketch the proof for the case of persistence modules of finite type over $\mathbb{N}$. That is, $\mathbb{V}$ where each V_i is finite dimensional and there exists $N \in \mathbb{N}$ such that, for all $i \geq N$, v_{i+1}^i are isomorphisms. It can be found in [8]. The extension to infinite-dimensional modules follows from [19]. On the other hand, the general proof of the second statement is fully developed in [20].

Let $\mathbb{V}$ be a persitence module of finite type over $\mathbb{N}$. Equip $k[t]$ with the standard grading. A graded module M over a graded ring R is a module such that $M \simeq \bigoplus_i M_i$, $i \in \mathbb{Z}$, and the action of R on M is defined by $R_n \otimes M_m \to M_{n+m}$. We say that M is non-negatively graded if $M_i = 0$ for all $i < 0$. Then, we can associate to $\mathbb{V}$ the following graded module over $k[t]$,

$$\alpha(\mathbb{V}) = \bigoplus_{i=0}^{\infty} V_i,$$

where the k-module structure is simply the sum of the structures on the individual components and the action of $k[t]$ is given by

$$t \cdot (u_0, u_1, \ldots) = (0, v_1^0(u_0), v_2^1(u_1), \ldots).$$

That is, it shifts elements up in the gradation thanks to the linear maps v_{i+1}^i. By Artin-Rees theory in commutative algebra, the correspondence α defines an equivalence of categories between the category of persistence modules of finite type over k and the category of finitely generated non-negatively graded modules over $k[t]$. Since k is a field, $k[t]$ is a PID and we can apply the structure theorem of graded modules over graded PID. This theorem states that, given M a graded module over D a PID, it decomposes uniquely as

$$\left(\bigoplus_{i=1}^{n} \Sigma^{\beta_i} D\right) \oplus \left(\bigoplus_{j=1}^{m} \Sigma^{\gamma_j} D/d_j D\right),$$

where $d_j \in D$ are homogeneous elements so that $d_j | d_{j+1}$; $\beta_i, \gamma_j \in \mathbb{Z}$ and Σ^β denotes a β-shift upward in the grading. In our case, the only graded ideals of $k[t]$ are homogeneous of the form $(t^i) = t^i \cdot k[t]$, $i \geq 0$, so we can conclude that $\alpha(\mathbb{V})$ decomposes as

$$\left(\bigoplus_{i=1}^{n} t^{\beta_i} k[t]\right) \oplus \left(\bigoplus_{j=1}^{m} t^{\gamma_j} k[t]/(t^{n_j})\right).$$

Finally, notice that, for interval modules, $\alpha(k[i, \infty))$ decomposes in the form $t^i k[t]$, while $\alpha(k[i, j))$ does as $t^i k[t]/(t^{j-i})$. ∎

Remark 2.2.13. The first part of the proof, up to the application of the structure theorem, is still true if we consider persistence modules $\mathbb{V}$ where each V_i is an R-module (R a commutative ring). Then, $\alpha(\mathbb{V})$ is a graded module over $R[t]$. However, $R[t]$ is not a PID, so we cannot apply the structure theorem and the classification of $R[t]$ graded modules is not simple. ♢

Lastly, we state and important property of decomposable modules: the multiset of intervals $(J_l)_{l\in L}$ is an invariant of $\mathbb{V}$, it does not depend on the decomposition chosen. It is deduced from the Krull-Remak-Schmidt-Azumaya theorem, that we introduce first.

Theorem 2.2.14 (Krull-Remak-Schmidt-Azumaya Theorem [21])**.** *Let M be a module and suppose it is decomposable into a direct sum of idecomposable submoduli M_λ,*

$$M \simeq \bigoplus_{\lambda\in\Lambda} M_\lambda.$$

In addition, suppose that the sum of any two non-automorphisms of M_λ is also a non-automorphism. Then, given a second direct decomposition of M into idecomposable submodules

$$M \simeq \bigoplus_{\delta\in\Delta} N_\delta,$$

there exists a one-to-one mapping $\sigma : \Lambda \ni \lambda \mapsto \sigma(\lambda) \in \Delta$ such that M_λ is isomorphic to $N_{\sigma(\lambda)}$ for each $\lambda \in \Lambda$.

Theorem 2.2.15 ([4])**.** *Let $\mathbb{V}$ be a persistence module over $T \subseteq \mathbb{R}$ which can be expressed as a direct sum of interval modules in two different ways*

$$\bigoplus_{m\in N} k^{K_n} \simeq \mathbb{V} \simeq \bigoplus_{l\in L} k^{J_l}.$$

Then, there is a bijection $\sigma : L \to N$ such that $J_l = K_{\sigma(l)}$ for all $l \in L$.

Proof. Notice that in our case $\mathbb{V}$ is decomposed as a direct sum of interval modules. We have seen that $\mathrm{End}(\mathbb{I}) = k$. Since the only non-isomorphism is the zero map, the condition on the non-automorphisms of Krull-Remak-Schmidt-Azumaya theorem is satisfied for both decompositions. We conclude that there exists a bijection $\sigma : L \to N$, $l \mapsto \sigma(l)$, such that k^{J_l} is isomorphic to $k^{K_{\sigma(l)}}$. Nevertheless, it is clear that if two interval modules are isomorphic $k^J \simeq k^K$, then $J = K$. All in all, $\sigma : L \to N$ is a bijection such that $J_l = K_{\sigma(l)}$ for each $l \in L$. ■

2.2.1 Persistence barcodes and diagrams

Let $\mathbb{V}$ be a decomposable persistence module over $T \subseteq \mathbb{R}$. As we have already pointed in remark 2.2.9, when the module is defined over a finite set T, the intervals can be written as closed intervals. However, in the general case, we have to distinguish between closed, open or half-open intervals. In order to keep a clear exposition

we introduce the following notation

$$\begin{aligned}(b^-, d^-) &:= [b, d),\\ (b^-, d^+) &:= [b, d],\\ (b^+, d^-) &:= (b, d),\\ (b^+, d^+) &:= (b, d],\end{aligned} \tag{2.3}$$

where $b, d \in \mathbb{R}$ and b^+, d^- could be infinite. These real numbers with a superscript are called **decorated reals** and they form a totally order set by establishing

$$b^- < b < b^+ < d^- < d < d^+, \tag{2.4}$$

whenever $b < d$. Usually, b^* means b^+ or b^-. Using this, a general decomposable persistence module is represented as

$$\bigoplus_{l \in L} k(b_l^*, d_l^*).$$

Thanks to the invariance of the multiset of intervals, given by theorem 2.2.14, the following notions are well-defined.

Definition 2.2.16. The **persistence barcode** of a decomposable module $\mathbb{V}$ is the collection of intervals of the decomposition.

Since the order of the intervals does not matter, one typically orders them in terms of the left endpoint (birth time) in increasing order.

Definition 2.2.17. The **decorated persistence diagram** is the multiset

$$\mathrm{Dgm}(\mathbb{V}) = \mathrm{Int}(\mathbb{V}) = \{(b_l^*, d_l^*) \ : \ l \in L\},$$

while the **undecorated persistence diagram** is defined to be the multiset

$$\mathrm{dgm}(\mathbb{V}) = \mathrm{int}(\mathbb{V}) = \{(b_l, d_l) \ : \ l \in L\} - \Delta,$$

where $\Delta = \{(r, r) \ : \ r \in \mathbb{R}\}$.

Remark 2.2.18. Isomorphic decomposable persistence modules have the same persistence diagrams. ♢

Observe that the persistence homology module obtained from filtrations of finite simplicial complexes satisfies the hypothesis of theorem 2.2.12. Hence, persistence barcodes and diagrams of such filtrations are always well-defined. The later are denoted by Dgm_n and dgm_n when they correspond to the n-th homology groups. Notice that distinct choices of the field k can lead to different persistence diagrams and barcodes.

Example 2.2.19. Let $\mathbb{X} \subset (\mathbb{R}^m, ||\cdot||_q)$ be a finite set of points and consider the persistence modules $(H_n(\check{\mathrm{C}}\mathrm{ech}_\alpha(\mathbb{X})))_{\alpha \in \mathbb{R}_+}$ and $(H_n(\mathbb{X}^\alpha))_{\alpha \in \mathbb{R}_+}$, where $\mathbb{X}^\alpha$ is the union of balls of radius α centered on the points of $\mathbb{X}$. Then, they have the same persistence barcode and diagram. ♢

When finite, persistence barcodes and diagrams can be represented in the plane. For persistence barcodes, intervals J are ploted as $\{(r_1, r_2) \in \mathbb{R}^2 : r_1 \in J\}$ for some fix $r_2 \in \mathbb{R}$. If we are studying the persistence modules obtained from homology, one represents the zero homology barcode first, followed by the first homology barcode and so on. An example is shown in figure 2.2. In the case of presistence diagrams (figure 2.3), the intervals are represented as a multiset of points in the extended upper half-plane

$$\bar{\mathbb{R}}^2_{\geq} = \{(r_1, r_2) \in \mathbb{R}^2 : r_2 \geq r_1\} \cup \{-\infty\} \times \mathbb{R} \cup \mathbb{R} \times \{+\infty\} \cup \{(-\infty, +\infty)\}.$$

For decorated persistence diagrams ticks pointing into the quadrant suggested by the superscripts are added, as it is shown in figure 2.1 and 2.4.

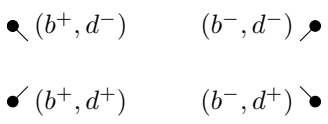


Figure 2.1: Decorated points of the plane and the associated ticks.

Example 2.2.20 ([4])**.** Consider a loop X in the plane, represented in figure 2.4, and the continuous map $g : X \to \mathbb{R}$ given by the projection onto the y-coordinate. Consider the sublevel set filtration $X_r = g^{-1}((-\infty, r])$. For each $n \geq 0$ this defines a persistence module $\mathbb{V}^n$ given by $V^n_r = H_n(X_r, k)$ and $v^s_r = H_n(i^s_r)$, where $i^s_r : X_s \hookrightarrow X_r$ is the inclusion. Let us see how they decompose into interval modules.

- For $n = 0$, the zero homology group is related to the path connected components of the topological space. In this case,

$$H_0(X_r, k) \simeq \begin{cases} 0 & r \in (-\infty, a), \\ k & r \in [a, b), \\ k^2 & r \in [b, c), \\ k & r \in [c, d), \\ k^2 & r \in [d, e), \\ k & r \in [e, \infty). \end{cases}$$

  Notice that there are five critical values of r at which the vector spaces change. Therefore, it is enough to consider the 5-term persistence modules obtained by restricting $\mathbb{V}^0$ to those five values

$$0 \to k \xrightarrow{v^a_b} k^2 \xrightarrow{v^b_c} k \xrightarrow{v^c_d} k^2 \xrightarrow{v^d_e} k.$$

  In order to decompose the persistence module, we need to specify the homomorphisms of the above sequence. Let p_a be a point of the path connected component born at $r = a$, p_b a point of the one born at $r = b$ and analogously for p_d and the one born at $r = d$. Then, when $r = c$ the fusion of the two components into one implies the equality of homology classes $[p_a] = [p_b]$ in

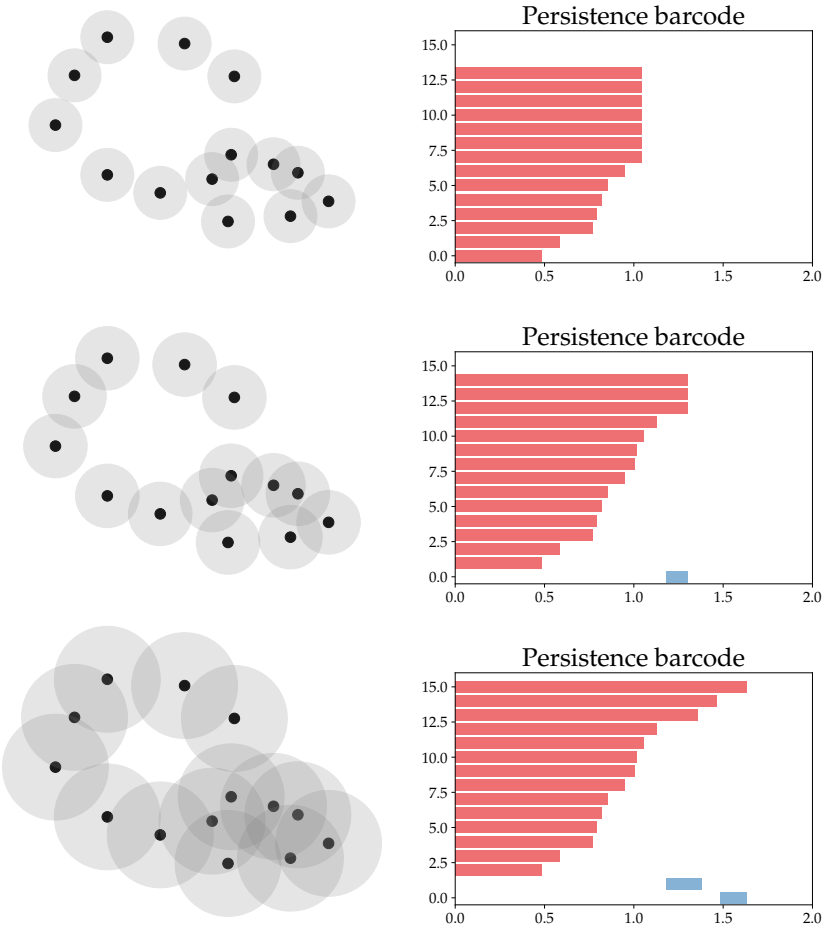


Figure 2.2: Example of persistence barcodes associated to the Rips filtration $(\text{Rips}_\alpha(\mathbb{X}))_{\alpha\in[0,\alpha']}$ of a data set of points $\mathbb{X}$ (left) at different scales: $\alpha' = 1, 1.2, 2$. The red intervals correspond to the zero homology, while the blue ones correspond to the first homology. The x-coordinate indicates the length of the interval.

$H_0(X_c, k)$. The same happens when $r = e$, and in this case $[p_a] = [p_b] = [p_d]$. Hence, the homomorphisms between vector spaces can be expressed as follows

$$\begin{array}{ccccccccccc} 0 & \to & k[p_a] & \to & k[p_a]\oplus k[p_b] & \to & k[p_a] & \to & k[p_a]\oplus k[p_d] & \to & k[p_a] \\ & & [p_a] & \mapsto & ([p_a],0) & \mapsto & [p_a] & \mapsto & ([p_a],0) & \mapsto & [p_a] \\ & & & & (0,[p_b]) & \mapsto & [p_a] & & & & \\ & & & & & & & & (0,[p_d]) & \mapsto & [p_a]. \end{array}$$

We can rewrite this so that the persistence module is the direct sum of interval modules, where remember that the maps between vector spaces could only be direct sums of the identity and zero map,

$$\begin{array}{ccccccccccc} 0 & \to & k[p_a] & \to & k[p_a]\oplus k([p_b]-[p_a]) & \to & k[p_a] & \to & k[p_a]\oplus k([p_d]-[p_a]) & \to & k[p_a] \\ & & [p_a] & \mapsto & ([p_a],0) & \mapsto & [p_a] & \mapsto & ([p_a],0) & \mapsto & [p_a] \\ & & & & (0,[p_b]-[p_a]) & \mapsto & 0 & & & & \\ & & & & & & & & (0,[p_d]-[p_a]) & \mapsto & 0 \end{array}$$

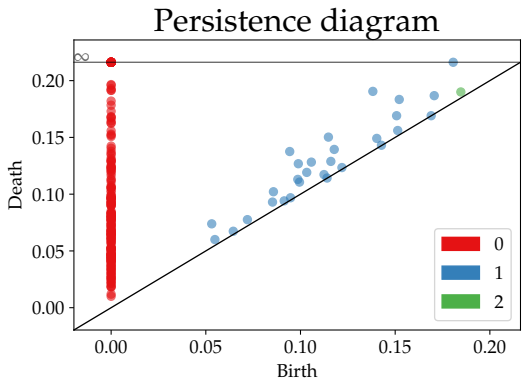


Figure 2.3: Example of persistence diagram. Red points correspond to the zero homology groups, blue to the first homology groups and the green one to the second homology group.

All in all, we have three persistence intervals

$$\begin{aligned} k[a,+\infty) &:0 \to k[p_a] \to k[p_a] \to k[p_a] \to k[p_a] \to k[p_a], \\ k[b,c) &:0 \to 0 \to k([p_b]-[p_a]) \to 0 \to 0 \to 0, \\ k[d,e) &:0 \to 0 \to 0 \to 0 \to k([p_d]-[p_a]) \to 0, \end{aligned}$$

and

$$\mathbb{V}^0 \simeq k[a,+\infty) \oplus k[b,c) \oplus k[d,e).$$

- For $n=1$, the homology groups are

$$H_1(X_r,k) \simeq \begin{cases} 0 & r \in (-\infty, f), \\ k & r \in [f,\infty). \end{cases}$$

  Therefore, in this case we already have a persistence interval, since all vector spaces are zero for $r<f$ and equal to k when $r \geq f$, with the identity as connecting homomorphism. Consequently,

$$\mathbb{V}^1 \simeq k[f,\infty).$$

- For $n \geq 2$ all homology groups are zero.

The decorated persistence diagram can be found in figure 2.4. Let us see what information can be obtained about X from it. For the zero homology group, figure 2.4 shows that it decomposes into three interval modules. One of them persists and the other two are born and die later, the last one at $r=e$. This implies that $H_0(X_r,k) \simeq k$ for $r>e$. Meanwhile, the first homology group only has one interval, which begins at f and persists. Therefore, $H_1(X_r,k) \simeq k$ for $r>f$. Any higher homology group is zero. Hence, X is path connected and there is a loop. ♢

Recall that each interval of a decomposition may be interpreted as the time that certain feature persists in the filtration. In this situation, long bars in a barcode,

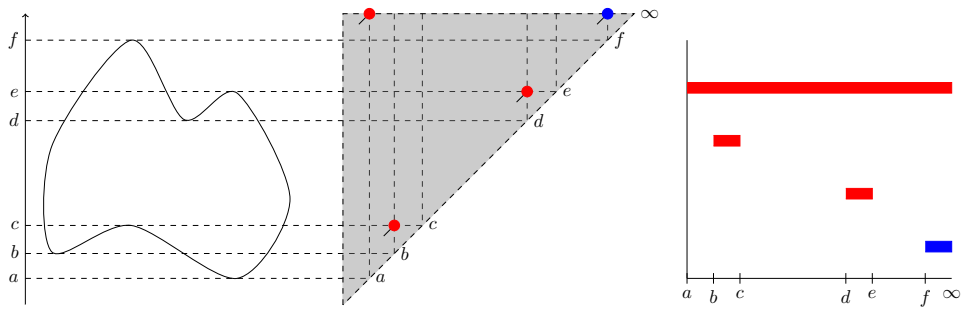


Figure 2.4: On the left, a loop X in the plane for which we consider the projection onto the y-coordinate. On the right, the decorated persistence diagram and barcode for the sublevel set filtration of X. The points and intervals corresponding to the zero homology groups are in red. The ones of the first homology group, in blue.

or points away from the diagonal in a the persistence diagram, represent topological features that persist through a large range of indexes, generally considered as being more significant. Meanwhile, short bars or points close to the diagonal are topological properties that appear and disappear almost instantly.

This gives us an idea of the importance of a barcode or a persistence diagram representation, due to its capability to capture significant features. In fact, we can establish a connection between the ranks of the homomorphisms of a decomposable persistence module and the barcode/diagram representation.

Assume that $\mathbb{V}$ is a persistence module, with finite dimensional vector spaces V_t, which is decomposable

$$\mathbb{V} \simeq \bigoplus_{l \in L} k(b_l^*, d_l^*) = \mathbb{W}.$$

Since $\mathbb{V}$ and $\mathbb{W}$ are isomorphic, the rank of $v_t^s : V_s \to V_t$ for $s \leq t$ is equal to the rank of $w_t^s : W_s \to W_t$. Recall that the latter are defined as direct sums of identity and zero maps. Therefore, $\mathrm{rk}(w_t^s)$ is the number of summands k of W_s that persist until W_t, i.e. that are not sent to zero. Notice that a summand satisfies this if and only if it belongs to an interval module k^J such that $[s,t] \subset J$. Therefore, $\mathrm{rk}(w_t^s)$ is equal to the number of interval modules k^J such that J contains $[s,t]$. In a persistence barcode representation this corresponds to the number of intervals that contain $[s,t]$. For a decorated persistence diagram this is equal to the number of points, counted with multiplicity, that belongs, together with their ticks, to $R = [-\infty, s] \times [t, \infty]$. Indeed, an interval $J \subset \mathbb{R}$ is represented by a point and a tick in R if and only if $J = [s,t]$; $J = (b,t]$ or $[b,t]$, with $b < s$; $J = [s,d)$ or $[s,d]$, with $d > t$; or any possible interval with endpoints $b < s \leq t < d$. It is clear that an interval J contains $[s,t]$ if and only if it is one of the above.

When working with homology persistence modules, $\beta_t^s = \dim(H_n^{s \to t}(X))$ is the rank of v_t^s. As a consequence, the persistence Betti number β_t^s is equal to the number of intervals in the barcode containing $[s,t]$ or, equivalently, to the cardinality of $\mathrm{Dgm}_n \cap ([-\infty, s] \times [t, \infty])$ counted with multiplicity.

Remark 2.2.21. This relation between ranks and cardinality of the persistence diagram restricted to a rectangle can be generalized, as we will see when studying

q-tame persistence modules in chapter 3 (see (3.9)). ♢

Going back to theorem 1.3.12, the homology of M^η can be recovered from the interval decomposition, the persistence barcode or the persistence diagram of the Rips complex filtration, since

$$\beta_n(M^\eta) = \mathrm{rk}(j_*) = \dim(H_n^{r\to 4r}(\mathrm{Rips}(\mathbb{X}))) = \beta_{4r}^r.$$

To sum up, we have managed to transform a set of complex observations into a collection of persistence diagrams (that are basically multisets in the extended upper half-plane), which are easier to handle and might be used to produce statistical analysis. In the next chapter we will extend the definition of persistence diagrams to persistence modules that are not necessarily decomposable into intervals.

Chapter 3

Q-tame persistence modules

It is possible to define persistence diagrams from the point of view of measure theory. In this framework, a persistence module over $T \subset \mathbb{R}$ defines an integer-valued measure on rectangles of $\mathbb{R}^2$, called persistence measure. When the measure is finite it is concentrated at a discrete set of points. These points, taken with their multiplicities, constitute the persistence diagram. It will be seen that this let us define persistence diagrams for certain persistence modules without assuming an interval decomposition, as it is the case of q-tame modules. For decomposable modules both definitions agree (theorem 3.1.4). The theory developed in this chapter can be found in chapter 3 of [4].

3.1 Persitence measures

This section is dedicated to introduce the principal aspects of this measure theory framework, before moving to q-tame modules. But first of all, some notation.

Persistence modules indexed over finite sets $\{t_0 < t_1 < \cdots < t_n\}$ are going to be understood as representations of quivers, due to their simplicity and clarity. Set theoretically, a quiver is a multigraph. A representation of a quiver consists in assigning a vector space to each vertex and a linear map to each edge. Therefore, the persistence module

$$V_{t_0} \to V_{t_1} \to \cdots \to V_{t_n}$$

is a representation of the quiver

$$\bullet \to \bullet \to \cdots \to \bullet.$$

By theorem 2.2.12 (1.), these persistence modules decompose as a finite sum of interval modules. So as to visualize the interval decomposition in the quiver diagram, we add the indexes $t \in T$ to the circles and we agree to use empty circles when the vector space associated is the zero one. Therefore, a map between two filled circles is always the identity; whereas all other maps are necessarily zero. We then write just $-$ instead of arrows. For instance,

$$k[t_0, t_2] = \bullet_{t_0} - \bullet_{t_1} - \bullet_{t_2}, \qquad k[t_0, t_1] = \bullet_{t_0} - \bullet_{t_1} - \circ_{t_2}.$$

Definition 3.1.1. The measure $\mu_{\mathbb{V}}$ associated to a persistence module $\mathbb{V}$ is a function defined on rectangles $R = [a, b] \times [c, d] \subset \mathbb{R}^2$, with $a < b \leq c < d$, and given by

$$\mu_{\mathbb{V}}(R) = \langle \circ_a - \bullet_b - \bullet_c - \circ_d | \mathbb{V} \rangle,$$

where $\langle \circ_a - \bullet_b - \bullet_c - \circ_d | \mathbb{V} \rangle$ is defined as the number of copies of $k[b, c]$ to occur in the decomposition of the finite indexed persistence module

$$V_a \to V_b \to V_c \to V_d.$$

Notice that, since we are considering a finite subset of indices, the interval decomposition exists and the above is well-defined. In general, for a finite index set $T = \{t_0 < \cdots < t_n\}$, the restricted persistence module $V_{t_0} \to \cdots \to V_{t_n}$ is denoted by $\mathbb{V}_T$. The number of copies of an arbitrary interval summand $\mathbb{I}$ in the decomposition of $\mathbb{V}_T$ is written as $\langle \mathbb{I} | \mathbb{V}_T \rangle$.

Obviously, two isomorphic persistence modules have the same associated measure.

Remark 3.1.2.

1. Consider the linear map $v_b^a : V_a \to V_b$. Then, the following is satisfied

$$\begin{aligned} \operatorname{rank}(v_b^a) &= \langle \bullet_a - \bullet_b | \mathbb{V} \rangle, \\ \operatorname{nullity}(v_b^a) &= \langle \bullet_a - \circ_b | \mathbb{V} \rangle, \\ \operatorname{conullity}(v_b^a) &= \langle \circ_a - \bullet_b | \mathbb{V} \rangle. \end{aligned}$$

   In fact, using Zorn's lemma when needed, choose a basis $\mathcal{B}^1$ of $\ker(v_b^a)$ and $\mathcal{P}^2$ a basis of $\operatorname{im}(v_b^a)$. Let $\mathcal{B}^2$ consist of preimages of the elements of $\mathcal{P}^2$ (choose one preimage per element). Then, $\mathcal{B}^1 \cup \mathcal{B}^2$ is a basis of V_a and we can complete $\mathcal{P}^2$ to a basis of V_b, $\mathcal{P}^2 \cup \mathcal{P}^3$. For each $e^1 \in \mathcal{B}^1$, $e^2 \in \mathcal{B}^2$ (which is sent to $l^2 \in \mathcal{P}^2$ by v_b^a) and $l^3 \in \mathcal{P}^3$ we have that

$$(\operatorname{span}(e^1) \to 0), \quad (\operatorname{span}(e^2) \to \operatorname{span}(l^2)), \quad (0 \to \operatorname{span}(l^3)),$$

   are respectively isomorphic to $\bullet_a - \circ_b$, $\bullet_a - \bullet_b$ and $\circ_a - \bullet_b$, which are the only three tipes of interval modules over $\{a, b\}$. This proves the assertion since, for instance, for the first one it implies that the cardinality of $\mathcal{B}^1$ is equal to the number of copies of $\bullet_a - \circ_b$ that occur in the decomposition of $V_a \to V_b$.

2. Let $a < b \leq c < d$ such that $r_c^b := \operatorname{rk}(v_c^b) < \infty$. Then,

$$\mu_{\mathbb{V}}(R) = \langle \circ_a - \bullet_b - \bullet_c - \circ_d | \mathbb{V} \rangle = r_c^b - r_c^a - r_d^b + r_d^a.$$

   In order to prove this let us see first that, for any finite index sets $S \subset T$ it is true that

$$\langle \mathbb{I} | \mathbb{V}_S \rangle = \sum_{\mathbb{J}} \langle \mathbb{J} | \mathbb{V}_T \rangle, \tag{3.1}$$

   where the sum is over those intervals $\mathbb{J} \subseteq T$ such that they restrict over S to $\mathbb{I}$. In fact, any interval decomposition of $\mathbb{V}_T$ induces and interval decomposition of $\mathbb{V}_S$. Then, in the latter, an interval summand is of certain type $\mathbb{I}$ if and only

if it comes from a summand $\mathbb{J}$ of $\mathbb{V}_T$ which restricts over S to $\mathbb{I}$. Therefore, the number of copies of $\mathbb{I}$ in $\mathbb{V}_S$ is equal to the sum, over those $\mathbb{J}$, of the number of copies of $\mathbb{J}$ in $\mathbb{V}_T$, proving (3.1). In the current case, $T = \{a, b, c, d\}$ and it follows that

$$\begin{aligned}
r_c^b &= \langle [b,c] | \mathbb{V}_{\{b,c\}} \rangle = \langle [a,c] | \mathbb{V}_T \rangle + \langle [a,d] | \mathbb{V}_T \rangle + \langle [b,c] | \mathrm{V}_T \rangle + \langle [b,d] | \mathbb{V}_T \rangle, \\
r_c^a &= \langle [a,c] | \mathbb{V}_{\{a,c\}} \rangle = \langle [a,c] | \mathbb{V}_T \rangle + \langle [a,d] | \mathbb{V}_T \rangle, \\
r_d^b &= \langle [b,d] | \mathbb{V}_{\{b,d\}} \rangle = \langle [b,d] | \mathbb{V}_T \rangle + \langle [a,d] | \mathbb{V}_T \rangle, \\
r_d^a &= \langle [a,d] | \mathbb{V}_{\{a,d\}} \rangle = \langle [a,d] | \mathbb{V}_T \rangle.
\end{aligned}$$

Being $r_c^b < \infty$ implies that the other three ranks are finite too. We can then compute $r_c^b - r_c^a - r_d^b + r_d^a$. On the right hand side all terms except $\langle [b,c] | \mathbb{V}_T \rangle = \langle \circ_a - \bullet_b - \bullet_c - \circ_d | \mathbb{V} \rangle$ cancel.

♢

The function $\mu_{\mathbb{V}}$ is called a measure because it is additive with respect to splitting a rectangle into two rectangles, i.e.

$$\begin{aligned}
\mu_{\mathbb{V}}([a,c] \times [d,f]) &= \mu_{\mathbb{V}}([a,b] \times [d,f]) + \mu_{\mathbb{V}}([b,c] \times [d,f]), \\
\mu_{\mathbb{V}}([a,c] \times [d,f]) &= \mu_{\mathbb{V}}([a,c] \times [d,e]) + \mu_{\mathbb{V}}([a,c] \times [e,f]),
\end{aligned} \tag{3.2}$$

whenever $a < b < c \leq d < e < f$.

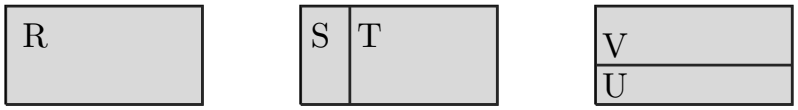


Figure 3.1: Vertical and horizontal splitting of a rectangle.

This is easy to prove using (3.1). We do it in quivers notation.

$$\begin{aligned}
\mu_{\mathbb{V}}([a,c] \times [d,f]) &= \langle \circ_a - - \bullet_c - \bullet_d - \circ_f | \mathbb{V} \rangle \\
&= \langle \circ_a - \bullet_b - \bullet_c - \bullet_d - \circ_f | \mathbb{V} \rangle + \langle \circ_a - \circ_b - \bullet_c - \bullet_d - \circ_f | \mathbb{V} \rangle \\
&= \langle \circ_a - \bullet_b - - \bullet_d - \circ_f - | \mathbb{V} \rangle + \langle - \circ_b - \bullet_c - \bullet_d - \circ_f | \mathbb{V} \rangle \\
&= \mu_{\mathbb{V}}([a,b] \times [d,f]) + \mu_{\mathbb{V}}([b,c] \times [d,f]), \\
\mu_{\mathbb{V}}([a,c] \times [d,f]) &= \langle \circ_a - \bullet_c - \bullet_d - - \circ_f | \mathbb{V} \rangle \\
&= \langle \circ_a - \bullet_c - \bullet_d - \circ_e - \circ_f | \mathbb{V} \rangle + \langle \circ_a - \bullet_c - \bullet_d - \bullet_e - \circ_f | \mathbb{V} \rangle \\
&= \langle \circ_a - \bullet_c - \bullet_d - \circ_e - | \mathbb{V} \rangle + \langle \circ_a - \bullet_c - - \bullet_e - \circ_f | \mathbb{V} \rangle \\
&= \mu_{\mathbb{V}}([a,c] \times [d,e]) + \mu_{\mathbb{V}}([a,c] \times [e,f]).
\end{aligned}$$

As a consequence, $\mu_{\mathbb{V}}$ satisfies the following properties.

Lemma 3.1.3 ([4]). *Consider rectangles $R_0, \ldots, R_k$ all of them of the form $[a_i, b_i] \times [c_i, d_i]$ with $a_i < b_i \leq c_i < d_i$:*

1. *Finitely additive: if R_0 can be written as a union $R_0 = R_1 \cup \cdots \cup R_k$ of rectangles with disjoint interiors, then $\mu_{\mathbb{V}}(R_0) = \mu_{\mathbb{V}}(R_1) + \cdots + \mu_{\mathbb{V}}(R_k)$.*

2. *Monotone: if $R_0 \subseteq R_1$ then $\mu_{\mathbb{V}}(R_0) \leq \mu_{\mathbb{V}}(R_1)$.*

3. *Subadditivity: if $R_0 \subseteq R_1 \cup \cdots \cup R_k$, then $\mu_{\mathbb{V}}(R_0) \leq \mu_{\mathbb{V}}(R_1) + \cdots + \mu_{\mathbb{V}}(R_k)$.*

Proof. 1. By induction and (3.2) the proof is immediate for product decompositions. Then, it suffices to consider a product decomposition of R by which each R_i is itself product-decomposed.

2. It is possible to decompose R_1 into finite rectangles with disjoint interiors such that R_0 is one of them. Then the result follows by finite aditivity and the fact that $\mu_{\mathbb{V}} \geq 0$.

3. Let us denote by $\alpha_1 < \alpha_2 < \cdots < \alpha_m$ all the x-coordinates of the corners of the rectangles. Do the same with $\beta_1 < \beta_2 < \cdots < \beta_n$ and the y-coordinates. It is clear that all pieces $[\alpha_l, \alpha_{l+1}] \times [\beta_t, \beta_{t+1}]$ have disjoint interiors and that each R_i is a union of some of those pieces. By finite aditivity, the measure of R_i is the sum of the measure of the pieces. As R_0 is contained in $R_1 \cup \cdots \cup R_k$, each of its pieces must belong, at least, to one of the R_j, for $j = 1, \ldots, k$. Hence, $\mu_{\mathbb{V}}(R_0) \leq \mu_{\mathbb{V}}(R_1) + \cdots + \mu_{\mathbb{V}}(R_k)$.

■

The next theorem links the measure $\mu_{\mathbb{V}}$ to the decorated persistence diagram defined for decomposable persistence modules.

Theorem 3.1.4 ([4])**.** *Suppose $\mathbb{V}$ is a decomposable persistence module over $\mathbb{R}$*

$$\mathbb{V} = \bigoplus_{l \in L} k(b_l^*, d_l^*).$$

Then

$$\mu_{\mathbb{V}}(R) = \operatorname{card}(\operatorname{Dgm}(\mathbb{V})|_R) \tag{3.3}$$

for every rectangle $R = [a, b] \times [c, d] \subset \mathbb{R}^2$ with $a < b \leq c < d$.

Remark 3.1.5. Seeing that points of $\operatorname{Dgm}(\mathbb{V})$ are decorated points, the restriction of this set to R consists of $(b_l^*, d_l^*) \in \operatorname{Dgm}(\mathbb{V}) \cap R$. In general, given a rectangle $R = [a, b] \times [c, d]$ with $a < b$ and $c < d$, we say that $(b_l^*, d_l^*) \in R$ if and only if

$$a < b_l^* < b \quad \& \quad c < d_l^* < d,$$

in the total order of (2.4). By studying the different cases of (2.3) it is easy to see that this is equivalent to the point with tick (b_l^*, d_l^*) lying in the closed rectangle R, as figure 3.2 shows. ♢

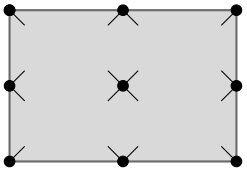


Figure 3.2: Different decorated points in a rectangle.

Proof. By decomposing each $k(b_l^*, d_l^*)_{\{a,b,c,d\}}$ into interval modules and putting them together we obtain an interval decomposition of $\mathbb{V}_{\{a,b,c,d\}}$. In addition, with this construction the number of copies of $k[b,c]$ in the latter is equal to the sum of the number of copies of $k[b,c]$ in each $k(b_l^*, d_l^*)$ decomposition. That is,

$$\mu_{\mathbb{V}}(R) = \langle [b,c] | \mathbb{V}_{\{a,b,c,d\}} \rangle = \sum_{l \in L} \langle [b,c] | k(b_l^*, d_l^*)_{\{a,b,c,d\}} \rangle.$$

On the other hand, since $k(b_l^*, d_l^*)_{\{a,b,c,d\}}$ is an interval module or zero, the number $\langle [b,c] | k(b_l^*, d_l^*)_{\{a,b,c,d\}} \rangle$ is 0 or 1. Moreover, it is 1 if and only if

$$k(b_l^*, d_l^*)_{\{a,b,c,d\}} = \circ_a - \bullet_b - \bullet_c - \circ_d,$$

that is, $a < b_l^* < b$ and $c < d_l^* < d$, or equivalently $(b_l^*, d_l^*) \in R$ by remark 3.1.5. All in all, $\mu_{\mathbb{V}}(R)$ is equal to the number of decorated points $(b_l^*, d_l^*) \in \mathrm{Dgm}(\mathbb{V})$ that also belong to R. ■

Given $\mathbb{V}$ decomposable such that r_c^b is finite, equality (3.3) and remark 3.1.2 lead to

$$\mathrm{card}(\mathrm{Dgm}(\mathbb{V})|_R) = r_c^b - r_d^b - r_c^a + r_d^a.$$

In the case of persistent homology these ranks are the persistence Betti numbers. Thus, if the corresponding persistence module is decomposable and β_c^b is finite, we have that

$$\mathrm{card}(\mathrm{Dgm}(\mathbb{V})|_R) = (\beta_c^b - \beta_d^b) - (\beta_c^a - \beta_d^a). \tag{3.4}$$

This can be interpreted as the number of independent homology classes born between $H_n(X_a)$ and $H_n(X_b)$ that die between $H_n(X_c)$ and $H_n(X_d)$.

Theorem 3.1.4 gives us an idea of how to define the persistence diagram when we do not have a decomposition into interval modules. Once the measure $\mu_{\mathbb{V}}$ associated to the persistence module is known, we look for a multiset $A \subset \mathbb{R}^2$ which satisfies (3.3) for all rectangles. In order to have a well-defined notion, we need to prove that such a multiset exists and is unique. Once this is done, the above corollary shows that both, this new definition and the old one, agree in the case of decomposable persistence modules.

Until now, we have restricted ourselves to finite rectangles. Nevertheless, we have seen that, even in the case of decomposable persistence modules, persistence diagrams can have points at infinity. Therefore, we should extend the measure $\mu_{\mathbb{V}}$ and give a definition for infinite rectangles.

Definition 3.1.6. Let $\mathbb{V}$ be a persistence module and $-\infty \leq a < b \leq c < d \leq +\infty$. Then,

$$\mu_{\mathbb{V}}([a,b] \times [c,d]) = \langle \circ_a - \bullet_b - \bullet_c - \circ_d | \mathbb{V} \rangle,$$

where $V_{-\infty} = 0 = V_{+\infty}$.

This extended version of $\mu_{\mathbb{V}}$ satisfies the same properties as before: additivity, monotonicity and subadditivity. There is also an analogue of remark 3.1.2,

$$\begin{aligned} \mu_{\mathbb{V}}([-\infty,b] \times [c,+\infty]) &= r_c^b, \\ \mu_{\mathbb{V}}([a,b] \times [c,+\infty]) &= r_c^b - r_c^a, \quad (\text{if } r_c^a < \infty) \\ \mu_{\mathbb{V}}([-\infty,b] \times [c,d]) &= r_c^b - r_d^b \quad (\text{if } r_c^b < \infty). \end{aligned} \tag{3.5}$$

Theorem 3.1.4 extends too.

Corollary 3.1.7 ([4])**.** *Given a decomposable persistence module $\mathbb{V}$, the measure $\mu_{\mathbb{V}}(R)$ of any rectangle R in the extended half-plane $\bar{\mathbb{R}}^2_{\geq}$ is equal to the number of interval summands corresponding to decorated points (b_l^*, d_l^*) which lie in R.*

Proof. Straightforward extension of the proof of theorem 3.1.4. ■

3.2 Rectangle measures

In this section we will introduce the notion of diagram of a given measure, in particular a rectangle measure. The measures $\mu_{\mathbb{V}}$ are just a particular case on the extended half-plane. For simplicity, we start working in $\mathbb{R}^2$. Later, we will expand the theory to the extended plane $\bar{\mathbb{R}}^2$.

Definition 3.2.1. Given $\mathcal{S}$ a subset of $\mathbb{R}^2$, define

$$\mathrm{Rect}(\mathcal{S}) = \{[a,b] \times [c,d] \subset \mathcal{S} \ : \ a < b \ \& \ c < d\}.$$

Definition 3.2.2. A **rectangle measure** or **r-measure** on $\mathcal{S}$ is a function $\mu : \mathrm{Rect}(\mathcal{S}) \to \{0,1,2,\ldots\} \cup \{\infty\}$ which is additive under vertical and horizontal splitting, as in (3.2).

Seeing that the proof of lemma 3.1.3 only uses vertical and horizontal additivity, it is also true for rectangle measures. Hence, they are finite additive, monotonous and subadditive.

Definition 3.2.3. The **r-interior** of a region $\mathcal{S} \subseteq \mathbb{R}^2$ is

$$\mathcal{S}^{\times} = \{(b^*,d^*) \ : \ \exists R \in \mathrm{Rect}(\mathcal{S}) \text{ such that } \ (b^*,d^*) \in R\}.$$

In particular, $R^{\times} = \{(b^*,d^*) \ : \ (b^*,d^*) \in R\}$ for any rectangle $R \subset \mathbb{R}^2$. The interior in the classical sense may be defined as

$$\mathcal{S}^{\circ} = \{(b,d) \ : \ \exists R \in \mathrm{Rect}(\mathcal{S}) \text{ such that } \ (b,d) \in R^{\circ}\},$$

where R° is the usual interior.

The relation between rectangle measures and multisets of $\mathbb{R}^2$ is stated by the following equivalence theorem.

Theorem 3.2.4 (Equivalence Theorem [4]). *Let $\mathcal{S} \subseteq \mathbb{R}^2$. There is a bijective correspondence between:*

1. *Finite r-measures μ on $\mathcal{S}$, i.e. $\mu(R) < \infty$ for every $R \in \mathrm{Rect}(\mathcal{S})$.*

2. *Locally finite multisets A in $\mathcal{S}^\times$, i.e. $\mathrm{card}(A|_R) < \infty$ for every $R \in \mathrm{Rect}(\mathcal{S})$.*

Any pair (μ, A) given by this correspondence satifies for all $R \in \mathrm{Rect}(\mathcal{S})$ the equality

$$\mu(R) = \mathrm{card}(A|_R), \tag{3.6}$$

Equivalently,

$$\mu(R) = \sum_{(b^*,d^*)\in R} n_{(b^*,d^*)},$$

where $n : \mathcal{S}^\times \to \{0, 1, 2, \ldots\}$ is the multiplicity function for A. Therefore, the measure can be expressed as

$$\mu = \sum_{x\in A} n_x \delta_x, \tag{3.7}$$

where δ_x denotes the Dirac mass located at $\{x\}$.

Proof. Let A be a locally finite multiset in $\mathcal{S}^\times$ and define the function $\mu : \mathrm{Rect}(\mathcal{S}) \to \{0, 1, \ldots\} \cup \{\infty\}$ as $\mu(R) = \mathrm{card}(A|_R)$. Since A is locally finite, μ is finite. In order to conclude that it is a rectangle measure, it suffices to prove additivity under splitting. Hence, consider $R \in \mathrm{Rect}(\mathcal{S})$ and split it vertically (horizontally) into two rectangles R_1, R_2. Note that every decorated point $(b^*, d^*) \in R$ belongs to exactly one of the R_i. Consequently,

$$\mu(R) = \mathrm{card}(A|_R) = \mathrm{card}(A|_{R_1}) + \mathrm{card}(A|_{R_2}) = \mu(R_1) + \mu(R_2).$$

For the reverse implication we refer to theorem 3.12 of [4]. ■

Thanks to this theorem we can finally define diagrams in the case of finite r-measures in $\mathbb{R}^2$.

Definition 3.2.5. Let μ be a finite rectangle measure on a region $\mathcal{S} \subseteq \mathbb{R}^2$. The **decorated diagram** of μ is the unique locally finite multiset $\mathrm{Dgm}(\mu)$ in $\mathcal{S}^\times$ such that

$$\mu(R) = \mathrm{card}(\mathrm{Dgm}(\mu)|_R),$$

for every $R \in \mathrm{Rect}(\mathcal{S})$. The **undecorated diagram** of μ is the corresponding locally finite multiset in $\mathcal{S}^\circ$

$$\mathrm{dgm}(\mu) = \{(b, d) \ : \ (b^*, d^*) \in \mathrm{Dgm}(\mu)\} \cap \mathcal{S}^\circ.$$

These notions can be generalized for non-finite rectangle measures and regions $\mathcal{S} \subseteq \bar{\mathbb{R}}^2$ in the extended plane. In this case, rectangles are of the form $R = [a, b] \times [c, d] \subset \bar{\mathbb{R}}^2$, with $-\infty \leq a < b \leq \infty$ and $-\infty \leq c < d \leq \infty$. The r-interior of R is defined analogously, whereas the interior R° is the relative interior of the rectangle as a subspace of the extended plane. This is done so as to not lose the points at infinity when we pass from a decorated diagram to its undecorated counterpart. A rectangle measure on a subset $\mathcal{S}$ of $\bar{\mathbb{R}}^2$ is defined likewise. The same happens with the r-interior and interior of $\mathcal{S}$. It can be proven as before that these measures are finite additive, monotonous and subadditive.

Example 3.2.6. For any persistence module $\mathbb{V}$, the measure $\mu_{\mathbb{V}}$ is an r-measure on $\bar{\mathbb{R}}^2_{\geq}$. ♢

To obtain an analogue of theorem 3.2.4 these two notions are required.

Definition 3.2.7. The **finite r-interior** of a rectangle measure μ on $\mathcal{S} \subset \bar{\mathbb{R}}^2$ is

$$\mathcal{F}^\times(\mu) = \{(b^*, d^*) \ : \ \exists R \in \mathrm{Rect}(\mathcal{S}) \ \text{ such that } \ (b^*, d^*) \in R, \ \mu(R) < \infty\}.$$

The **finite interior** is defined as

$$\mathcal{F}^\circ(\mu) = \{(b, d) \ : \ \exists R \in \mathrm{Rect}(\mathcal{S}) \ \text{ such that } \ (b, d) \in R^\circ, \ \mu(R) < \infty\}.$$

Notice that, whether μ is finite, then $\mathcal{F}^\times(\mu) = \mathcal{S}^\times$ and $\mathcal{F}^\circ(\mu) = \mathcal{S}^\circ$.
The following lemma will be needed in order to prove the next theorem.

Lemma 3.2.8 ([4])**.** *Let μ be a rectangle measure on $\mathcal{S} \subset \bar{\mathbb{R}}^2$ and consider any $R \in \mathrm{Rect}(\mathcal{S})$. Then, $R^\times \subseteq \mathcal{F}^\times(\mu)$ if and only if $\mu(R) < \infty$.*

Proof. Sufficiency is trivial. In order to prove necessity, let $R \in \mathrm{Rect}(\mathcal{S})$ such that $R^\times \subseteq \mathcal{F}^\times(\mu)$. Consider $(b, d) \in R$. If it is an interior point, then the four decorations (b^*, d^*) belong to $R^\times$, and hence to $\mathcal{F}^\times(\mu)$ by hypothesis. The definition of the finite r-interior implies that, for each of those decorations, there is a rectangle in $\mathrm{Rect}(\mathcal{S})$ of finite measure that contains the decorated point. Then, the union of these four rectangles contains a neighborhood of (b, d). It is clear that we can take a rectangle $S \subset R$ contained in these union with (b, d) in its relative interior, with respect to R. By subadditivity, S has finite measure. The same can be done for non-interior points. The only difference is that there will be only two or one decorations contained in $R^\times$, and then we will only need two or one rectangles of finite measure.

As a consequence, for each $(b, d) \in R$ we are able to find a rectangle $S \subset R$ of finite measure containing the point (b, d) in its relative interior. The rectangle R is contained in the union of all these rectangles. However, it is compact, so we can reduce it to a finite union. Again, subadditivity implies R has finite measure. ■

Theorem 3.2.9 ([4])**.** *Let μ be an r-measure on $\mathcal{S} \subset \bar{\mathbb{R}}^2$. Then there is a uniquely defined locally finite multiset $\mathrm{Dgm}(\mu)$ in $\mathcal{F}^\times(\mu)$ such that*

$$\mu(R) = \mathrm{card}(\mathrm{Dgm}(\mu)|_R), \tag{3.8}$$

for every $R \in \mathrm{Rect}(\mathcal{S})$ with $R^\times \subseteq \mathcal{F}^\times(\mu)$.

Proof. Firstly, consider the case $\mathcal{S} \subset \mathbb{R}^2$. For each rectangle R contained in $\mathcal{S}$ of finite measure, theorem 3.2.4 can be applied. This provides a multiset in $R^\times$ such that equation (3.8) holds. By the uniqueness from 3.2.4, multisets of rectangles that overlap should agree on the intersection. Note that the union of all the $R^\times$ of finite measure is exactly $\mathcal{F}^\times(\mu)$. As a result, we obtain a uniquely defined multiset in $\mathcal{F}^\times(\mu)$ satisfying (3.8) for any $R \in \mathrm{Rect}(\mathcal{S})$ whose μ-measure is finite. Applying lemma 3.2.8 we conclude the proof.

Secondly, let $\mathcal{S} \subset \bar{\mathbb{R}}^2$. Consider the embedding

$$\varphi : \bar{\mathbb{R}}^2 \to \mathbb{R}^2; \qquad (x, y) = (\arctan x, \arctan y)$$

It transforms homeomorphically the extended half-plane into the rectangle $[-\pi/2, \pi/2]^2$ contained in $\mathbb{R}^2$. Notice that the statements of theorem 3.2.4 and of the first proven part are invariant under φ. Indeed, consider $\varphi\#\mu$ defined as $\varphi\#\mu(R) = \mu(\varphi^{-1}(R))$ for any $R \in \mathrm{Rect}(\varphi(\mathcal{S}))$, where $\varphi^{-1} : [-\pi/2, \pi/2]^2 \to \bar{\mathbb{R}}^2$ is the inverse of φ. This is well-defined since $\varphi^{-1}(\mathrm{Rect}(\varphi(\mathcal{S}))) = \mathrm{Rect}(\mathcal{S})$. Moreover, this last equality also implies that, for every $R \in \mathrm{Rect}(\varphi(\mathcal{S}))$, there exists $R' \in \mathrm{Rect}(\mathcal{S})$ such that $\varphi^{-1}(R) = R'$. Therefore, $\varphi\#\mu(R) = \mu(R')$ and $\varphi\#\mu$ is a rectangle measure on $\varphi(\mathcal{S}) \subseteq [-\pi/2, \pi/2]^2 \subset \mathbb{R}^2$. Applying the first part of this proof to $\varphi\#\mu$, there is a unique locally finite multiset $\mathrm{Dgm}(\varphi\#\mu)$ in $\mathcal{F}^\times(\varphi\#\mu)$ such that $\varphi\#\mu(R) = \mathrm{card}(\mathrm{Dgm}(\varphi\#\mu)|_R)$, for all $R \in \mathrm{Rect}(\varphi(\mathcal{S}))$ with $R^\times \subseteq \mathcal{F}^\times(\varphi\#\mu)$.

However, it is easy to see that $\varphi^{-1}(\mathcal{F}^\times(\varphi\#\mu)) = \mathcal{F}^\times(\mu)$ and $\mathrm{card}(A|_R) = \mathrm{card}(\varphi^{-1}(A)|_{\varphi^{-1}(R)})$. All this, together with the fact that φ is a homeomorphism onto $[-\pi/2, \pi/2]^2$, implies that $\varphi^{-1}(\mathrm{Dgm}(\varphi\#\mu))$ is a locally finite multiset in $\mathcal{F}^\times(\mu)$ which satisfies the statement of the theorem. Moreover, it is unique. ■

We can finally present the wanted definitions.

Definition 3.2.10. The **decorated** and **undecorated diagrams** for an r-measure on $\mathcal{S} \subset \bar{\mathbb{R}}^2$ are the ordered pairs

$$(\mathrm{Dgm}(\mu), \mathcal{F}^\times(\mu)), \qquad (\mathrm{dgm}(\mu), \mathcal{F}^\circ(\mu)),$$

where $\mathrm{Dgm}(\mu)$ is given by the previous theorem and $\mathrm{dgm}(\mu)$ is

$$\mathrm{dgm}(\mu) = \{(b, d) \ : \ (b^*, d^*) \in \mathrm{Dgm}(\mu)\} \cap \mathcal{F}^\circ(\mu).$$

It should be pointed out that all the important information of μ is encoded in the persistence diagram. Indeed, the measure of any rectangle R with $R^\times \subseteq \mathcal{F}^\times(\mu)$ is recovered by counting decorated points of $\mathrm{Dgm}(\mu)$, and any other rectangle has infinite measure, as lemma 3.2.8 shows.

Remark 3.2.11. All in all, undecorated persistence diagrams $\mathrm{dgm}(\mu)$ of rectangle measures are locally finite multisets in $\mathcal{F}^\circ(\mu)$. If these multisets do not have points at infinity, they can be described by measures of the form $\nu = \sum_{x \in \mathrm{dgm}(\mu)} n_x \delta_x$ supported on $\mathcal{F}^\circ(\mu) \cap \mathbb{R}^2$, where $n_x \in \mathbb{N}$ is the multiplicity of the point x and δ_x is the Dirac mass located at $x \in \mathbb{R}^2$ (recall equality (3.7)). Note that ν is a Borel measure finite on compact sets of $\mathcal{F}^\circ(\mu) \cap \mathbb{R}^2$, i.e. a Radon measure. In fact, any compact set K of $\mathcal{F}^\circ(\mu) \cap \mathbb{R}^2$ can be covered by finitely many open rectangles R°

such that $\mu(R) < \infty$, by definition of the finite interior and by compactness. Then, by lemma 3.2.8 and theorem 3.2.9, $\text{card}(\text{Dgm}(\mu)|_R)$ coincide with $\mu(R)$, which is finite. Therefore, so is the cardinality of $\text{dgm}(\mu)$ on each of those rectangles. As a result, $\nu(K)$ is finite. This measure theory viewpoint has huge benefits when studying the properties of the space of persistence diagrams and it will be used in section 4.2. ♢

3.3 Measure persistence diagrams for persistence modules

As already mentioned in example 3.2.6, the persistence measure $\mu_{\mathbb{V}}$ of any persistence module $\mathbb{V}$ is an r-measure on the extended half-plane $\bar{\mathbb{R}}^2_{\geq}$. Consequently, its decorated diagram $\text{Dgm}(\mu_{\mathbb{V}})$ is defined in the subset of $\bar{\mathbb{R}}^2_{\geq}$ over which $\mu_{\mathbb{V}}$ is finite, the finite r-interior. In general, it does not have to be the whole extended half-plane. Notice that $\text{dgm}(\mu_{\mathbb{V}})$ is contained in $\bar{\mathbb{R}}^2_{>} = \{(r_1, r_2) \in \bar{\mathbb{R}}^2 \ : \ r_1 < r_2\}$.

Definition 3.3.1. Let $\mathbb{V}$ be a persistence module. Its **measure persistence diagrams** are the decorated and undecorated diagrams

$$\text{Dgm}(\mathbb{V}) = (\text{Dgm}(\mu_{\mathbb{V}}), \mathcal{F}^{\times}(\mu_{\mathbb{V}})), \qquad \text{dgm}(\mathbb{V}) = (\text{dgm}(\mu_{\mathbb{V}}), \mathcal{F}^{\circ}(\mu_{\mathbb{V}})),$$

Remark 3.3.2. Since two isomorphic persistence modules define the same measure, their persistence diagrams coincide. ♢

The following relationship between the two definitions of persistence diagrams is then clear.

Proposition 3.3.3 ([4])**.** *If $\mathbb{V}$ is a decomposable persistence module, then* $\text{Int}(\mathbb{V})$ *(the persistence diagram defined from the decomposition in 2.2.17) agrees with the measure persistence diagram* $\text{Dgm}(\mu_{\mathbb{V}})$ *where the latter is defined, that is, on* $\mathcal{F}^{\times}(\mu_{\mathbb{V}})$.

Proof. By corollary 3.1.7 we have that $\mu_{\mathbb{V}}(R) = \text{card}(\text{Int}(\mathbb{V})|_R)$ for all rectangles contained in the extended half-plane. On the other hand, by theorem 3.2.9 we also have $\mu_{\mathbb{V}}(R) = \text{card}(\text{Dgm}(\mu_{\mathbb{V}})|_R)$ for all rectangles with finite measure. Uniqueness implies that $\text{Int}(\mathbb{V})$ and $\text{Dgm}(\mu_{\mathbb{V}})$ must be the same multiset when restricted to $\mathcal{F}^{\times}(\mu_{\mathbb{V}})$. ■

Remark 3.3.4. Neither definition strictly outperforms the other. There are decomposable persistence modules for which Dgm is not defined but Int is. For instance, if $\mathbb{V} = \bigoplus_{l \in L} k(b_l^*, d_l^*)$ with (b_l, d_l) forming a dense subset of $\bar{\mathbb{R}}^2_{\geq}$, $\text{Int}(\mathbb{V})$ is defined but $\text{Dgm}(\mu_{\mathbb{V}})$ is not, since $\mu_{\mathbb{V}}(R)$ is infinite for every rectangle. Moreover, there are persistence modules not decomposable for which Dgm is defined almost everywhere (see example 3.31 of [4]). ♢

3.3.1 Persistence diagrams for q-tame modules

An interesting type of persistence modules for which persistence diagrams are well-defined are the q-tame modules.

Definition 3.3.5. A persistence module $\mathbb{V}$ over $T \subset \mathbb{R}$ is **q-tame** if $r_t^s := \mathrm{rk}(v_t^s)$ is finite, whenever $s < t$.

By remark 3.1.2 this implies that $\mu_{\mathbb{V}}(R)$ is finite for every $R = [a, b] \times [c, d] \subset \bar{\mathbb{R}}^2_{\geq}$ not touching the diagonal, i.e. $a < b < c < d$. That is, $\mu_{\mathbb{V}}$ is a finite r-measure on $\bar{\mathbb{R}}^2_{>}$ and $\mathcal{F}^{\circ}(\mu_{\mathbb{V}}) = \bar{\mathbb{R}}^2_{>}$. Therefore, the measure persistence diagram of a q-tame module is well-defined as a locally finite multiset on $\bar{\mathbb{R}}^2_{>}$, given by $\mathrm{dgm}(\mathbb{V})$. Nevertheless, $\mathrm{Dgm}(\mathbb{V})$ might have points in the diagonal (some rectangle touching the diagonal may have finite measure). For several reasons, which will become clear in the next chapter, we will not consider these points and therefore we will work with undecorated diagrams. Moreover, we will add to $\mathrm{dgm}(\mathbb{V})$ the diagonal with infinite multiplicity. This will allow us to define distances between persistence diagrams. Summarizing:

Theorem 3.3.6 ([1])**.** *If $\mathbb{V}$ is a q-tame persistence module, then it has a well-defined persistence diagram. Such a persistence diagram is the union of points in $\Delta = \{(r, r) : r \in \mathbb{R}\}$, counted with infinite multiplicity, and a locally finite multiset* $\mathrm{dgm}(\mathbb{V})$ *in $\bar{\mathbb{R}}^2_{>}$. That is, for any rectangle $R = [a, b] \times [c, d]$ with $a < b < c < d$, the number of points of* $\mathrm{dgm}(\mathbb{V})$ *contained in R, counted with multiplicity, is finite.*

Last but not least, the case of simplicial complexes is recovered and some results concerning their persistence modules at the homology level are stated.

Definition 3.3.7. A filtration $(X_t)_{t\in T}$ of topological spaces or simplicial complexes is **tame** if, for any integer n, the persistence module $(H_n(X_t, k))_{t\in T}$ is q-tame. If this filtration consists of sublevel sets of some function $f : X \to \mathbb{R}$, we say that f is tame.

In these cases, due to (3.5) and as long as $b < c$, the following equivalence between persistence Betti numbers and measures of rectangles $R = [-\infty, b] \times [c, \infty]$ is satisfied

$$\beta_c^b = \mu_{\mathbb{V}}(R) = \mathrm{card}(\mathrm{Dgm}(\mathbb{V})|_R). \tag{3.9}$$

This is the same relation we obtained before remark 2.2.21 for decomposable persistence modules.

Lastly, some examples of tame filtrations are given.

Example 3.3.8. The filtrations of finite simplicial complexes are always tame. ♢

Proposition 3.3.9 ([4])**.** *Let X be a compact polyhedron. That is, X is the realisation of a finite simplicial complex K as a topological space. Consider $f : X \to \mathbb{R}$ a continuous function. Then, f is tame.*

Proof. Consider $b < c$. By successive barycenter subdivision, X can be represented as the realisation of a finite simplicial complex where no simplex meets both $f^{-1}(b)$ and $f^{-1}(c)$. Let Y be the union of the closed simpleces which meet $X_b = f^{-1}(-\infty, b]$. Then, by construction, we have the inclusions $X_b \subseteq Y \subseteq X_c$. For any $n \in \mathbb{N}$, this translates into the commutative diagram

$$\begin{array}{ccc} H_n(X_b, k) & \longrightarrow & H_n(X_c, k) \\ & \searrow \quad \nearrow & \\ & H_n(Y, k) & \end{array}$$

Since Y is the realisation of a finite simplicial complex, $H_n(Y,k)$ has finite dimension. Therefore, the rank of $H_n(X_b,k) \to H_n(X_c,k)$ is finite. This proves that the homology persistence module is q-tame. ■

Corollary 3.3.10 ([4])**.** *Let X be a locally compact polyhedron. That is, X is the realisation of a locally finite*[1] *simplicial complex K. Consider $f : X \to \mathbb{R}$ a proper continuous funtion which is bounded below. Then, f is tame.*

Proof. Analogously to the previous proposition, we have to prove that, for any $b < c$, the rank of the homomorphism $H_n(X_b,k) \to H_n(X_c,k)$ is finite. Notice that, if we find a compact subpolyhedron Y of X that contains X_c ($X_b \subseteq X_c$), then we finish by applying the preceding proposition in Y. By hypothesis, f is bounded below, so $X_c = f^{-1}[\min(f), c]$; and proper, so X_c is compact. Therefore, any open cover of X_c by open neighborhoods of its points, has a finite subcover. Since K is locally finite, those neighborhoods can be chosen such that they only intersect finitely many closed simplices. As a result, there are only finitely many closed simplices meeting X_c. The union of them gives Y. ■

This is the case of distance functions, whence the homology persistence modules obtained from offsets of compact sets are q-tame too.

[1] A collection of subsets of a topological space X is said to be locally finite if each point in the space has a neighborhood that intersects only finitely many of the sets in the collection.

Chapter 4

Metrics on the space of persistence diagrams

In previous chapters it has been shown that persistence diagrams are hugely helpful tools to encode and interpret topological information of filtrations. It is clear that, in order to study this information, it would be extremely useful to compare them. We will focus on q-tame modules, whose persistence diagrams are well-defined and which include relevant practical cases. All in all, the objective of this chapter is to endow the set of persistence diagrams with a metric structure, i.e. to give a distance between them. The main reference is [6].

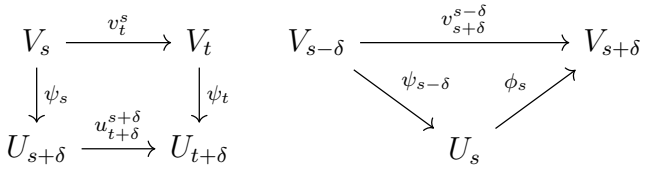
4.1 Bottleneck and p-th distances

Throughout this section we will consider persistence diagrams associated to q-tame modules. Owing to theorem 3.3.6, they consist of a locally finite multiset, in the open extended half-plane $\bar{\mathbb{R}}^2_>$, and the diagonal $\Delta = \{(r, r) : r \in \mathbb{R}\}$, with infinite multiplicity. Therefore, in order to define a distance between them, we need to specify the distance between any pair of points in $\bar{\mathbb{R}}^2_>$ and between any point and the diagonal. The standard way of doing this is the bottleneck distance d_∞, that we introduce below. This distance is related to the interleaving distance between two persistence modules $\mathbb{U}$, $\mathbb{V}$. The latter is defined as

$$d_i(\mathbb{U}, \mathbb{V}) = \inf\{\delta \geq 0 \ : \ \mathbb{U}, \mathbb{V} \ \text{ are } \delta\text{-interleaved}\}, \tag{4.1}$$

where two modules are δ-interleaved if, for all t, there are linear maps $\phi_t : U_t \to V_{t+\delta}$ and $\psi_t : V_t \to U_{t+\delta}$ such that the following diagrams commute for all $s \leq t$

$$\begin{array}{ccc} U_s & \xrightarrow{u_t^s} & U_t \\ \downarrow{\scriptstyle \phi_s} & & \downarrow{\scriptstyle \phi_t} \\ V_{s+\delta} & \xrightarrow{v_{t+\delta}^{s+\delta}} & V_{t+\delta} \end{array} \qquad \begin{array}{ccc} U_{s-\delta} & \xrightarrow{u_{s+\delta}^{s-\delta}} & U_{s+\delta} \\ & {\scriptstyle \phi_{s-\delta}} \searrow \quad \nearrow {\scriptstyle \psi_s} & \\ & V_s & \end{array}$$



In fact, the isometry theorem (theorem 5.14 of [4]) asserts that these two distances are equal for q-tame persistence modules and their undecorated persistence diagrams

$$d_\infty(\mathrm{dgm}(\mathbb{U}), \mathrm{dgm}(\mathbb{V})) = d_i(\mathbb{U}, \mathbb{V}). \tag{4.2}$$

It should be pointed out that the interleaving distance is not a true metric, $d_i(\mathbb{V}, \mathbb{U})$ does not imply $\mathbb{V} \simeq \mathbb{U}$ (see example 5.2 of [4])

Let then dgm^1 and dgm^2 be two persistence diagrams with locally finite multisets X and Y in $\bar{\mathbb{R}}^2_>$, respectively. Recall that, even though a multiset may have the same point repeated several times, all these repetitions are considered as different points of the set.

Definition 4.1.1. A **partial matching** between dgm^1 and dgm^2 is a subset $\gamma \subseteq (X \cup \Delta) \times (Y \cup \Delta)$ such that any $x \in X$ ($y \in Y$) appears exactly once as a first (second) coordinate in γ.

Denote by $\Gamma(\mathrm{dgm}^1, \mathrm{dgm}^2)$ the set of all partial matchings between dgm^1 and dgm^2. Notice it is nonempty, since we can always match all points to Δ.

Remark 4.1.2. To define a distance between persistence diagrams based in pairing points, it is essential to add the diagonal with infinite multiplicity, as we did in theorem 3.3.6. In this way, if they have different number of off-diagonal points, we can find a match for the remaining ones, which could be countable. ♢

Definition 4.1.3. The **bottleneck distance** d_∞ between two persistence diagrams dgm^1 and dgm^2 is

$$d_\infty(\mathrm{dgm}^1, \mathrm{dgm}^2) = \inf_{\gamma \in \Gamma(\mathrm{dgm}^1, \mathrm{dgm}^2)} \left(\max_{(x,y)\in\gamma} ||x - y||_\infty \right).$$

We can interpret $\max_{(x,y)\in\gamma} ||x - y||_\infty$ (the length of the longest edge in the matching) as the cost of γ, so that the bottleneck distance is defined as the minimal cost that can be achieved by matchings between both diagrams. A partial matching that realizes this infimum is said to be **optimal**. Indeed, there always exists an optimal matching.

Theorem 4.1.4. *Let X, Y be two locally finite multisets in the extended half-plane. Assume that, for every $\eta > \delta$, there exists a partial matching between X and Y with cost less or equal than η. Then, there exists a partial matching with cost less or equal than δ.*

Proof. For every integer $n \geq 1$ let γ_n be a matching with cost less or equal than $\delta + 1/n$. Each of these matchings defines an indicator function $\chi_n : X \times Y \to \{0, 1\}$. The proof is based on constructing an indicator function $\chi : X \times Y \to \{0, 1\}$ which

is the limit of the χ_n. Then, check that this χ actually defines a matching with cost less or equal than δ. This is the matching we were looking for. The proof can be found in theorem 5.12 of [4]. ■

Remark 4.1.5. Although in the standard definition of the bottleneck distance $||\cdot||_\infty$ is used, any norm $||\cdot||_q$ may be used instead, where $1 \leq q \leq \infty$. Due to the equivalence of norms in $\mathbb{R}^2$, this change has almost no consequences on the results. ♢

Given a partial matching γ notice that, if $(x, y) \in \gamma$ where $x \in X$ has finite coordinates and $y \in \Delta$, matching x with its orthogonal projection onto the diagonal instead would only reduce the cost. Hence, we can always assume y to be the orthogonal projection of x. Similarly, the points of the diagonal can always be matched to themselves with a null cost.

Remark 4.1.6. The definition of bottleneck distance can also be applied to the decorated persistence diagram of a q-tame module, adding to it the diagonal with infinite multiplicity. However, in this case, the persistence diagram $\mathrm{Dgm}(\mathbb{V})$ may include diagonal points. Therefore, Dgm^1 and Dgm^2 with the same off-diagonal points, but with different diagonal ones, would satisfy $d_\infty(\mathrm{Dgm}^1, \mathrm{Dgm}^2) = 0$, whereas $\mathrm{Dgm}^1 \neq \mathrm{Dgm}^2$. By considering undecorated diagrams this problem is avoided. From a practical point of view it is also understandable to forget about boundary points, since they represent features that do not persist in the filtration. ♢

Furthermore, persistence diagrams may contain points of the form $(b, +\infty)$ or $(-\infty, d)$. The set of these points is called **essential part** of the diagram and they are compared in the expected way

$$\begin{aligned} ||(-\infty, d) - (-\infty, q)||_\infty &= |d - q|, \\ ||(b, +\infty) - (p, +\infty)||_\infty &= |b - p|, \\ ||(-\infty, +\infty) - (-\infty, +\infty)||_\infty &= 0, \\ ||(b, d) - (-\infty, q)||_\infty &= \infty, \\ ||(b, d) - (p, +\infty)||_\infty &= \infty. \end{aligned}$$

Hence, if the cardinalities of the essential parts of two persistent diagrams differ, the bottleneck distance would be infinite. That is, it would not define a true metric. To be finite we need

$$\begin{aligned} \mathrm{card}(\mathrm{dgm}^1|_{\{-\infty\}\times\mathbb{R}}) &= \mathrm{card}(\mathrm{dgm}^2|_{\{-\infty\}\times\mathbb{R}}), \\ \mathrm{card}(\mathrm{dgm}^1|_{\mathbb{R}\times\{+\infty\}}) &= \mathrm{card}(\mathrm{dgm}^2|_{\mathbb{R}\times\{+\infty\}}), \\ \mathrm{card}(\mathrm{dgm}^1|_{\{-\infty\}\times\{+\infty\}}) &= \mathrm{card}(\mathrm{dgm}^2|_{\{-\infty\}\times\{+\infty\}}), \end{aligned}$$

and that the points of the essential parts are matched with each other. It is easy to see that an optimal matching consists in sorting the points with respect to their finite coordinate and then using increasing matching. We can conclude that the essential parts behave independently of the rest of the diagram. Whence, in the following persistence diagrams with empty essential parts will be considered. This means that dgm will be contained in $\{(r_1, r_2) \in \mathbb{R}^2 \ : \ r_1 < r_2\}$.

Remark 4.1.7. When computing persistence diagrams with numerical tools, filtrations need to be finite. As a consequence, it may happen that essential points are not such, and they disappear when larger filtrations are consider, or that they have not been studied yet, and will appear if we study a larger filtration. ♢

A distance between persistence diagrams is also useful to study how perturbations of the data set affect them. Obviously, these perturbations can change the filtration and, consequently, the persistence module and persistet homology, and they will be visible in the diagram. We would like that persistence diagrams are stable under some perturbations. That is, that two close data sets induce close diagrams in the bottleneck distance. Regarding this we have the two results below.

Theorem 4.1.8 ([1])**.** *Let M be a topological space and $f, g : M \to \mathbb{R}$ two real-valued functions. Suppose they are tame. Then, for any $n \in \mathbb{N}$,*

$$d_\infty(\mathrm{dgm}_n(f), \mathrm{dgm}_n(g)) \leq ||f - g||_\infty = \sup_{x \in M} |f(x) - g(x)|,$$

where $\mathrm{dgm}_n(f)$ is the persistence diagram associated to the persistence module $(H_n(f^{-1}((-\infty, r])))_{r \in \mathbb{R}}$. In particular, the operator $f \mapsto \mathrm{dgm}_n(f)$ is 1-Lipschitz.

Proof. Denote $\delta = ||f - g||_\infty$. Then, for any $r \in \mathbb{R}$, we have

$$F_r = f^{-1}(-\infty, r] \subseteq g^{-1}(-\infty, r + \delta] = G_{r+\delta},$$
$$G_r = g^{-1}(-\infty, r] \subseteq f^{-1}(-\infty, r + \delta] = F_{r+\delta}.$$

For every $s \leq t$ this yields to the following commutative diagrams, where all maps involved are inclusions

$$\begin{array}{ccc} F_s & \hookrightarrow & F_t \\ \downarrow{\scriptstyle \phi_s} & & \downarrow \\ G_{s+\delta} & \hookrightarrow & G_{t+\delta} \end{array} \qquad \begin{array}{ccc} F_{s-\delta} & \hookrightarrow & F_{s+\delta} \\ & \searrow \quad \nearrow & \\ & G_s & \end{array}$$

$$\begin{array}{ccc} G_s & \hookrightarrow & G_t \\ \downarrow & & \downarrow \\ F_{s+\delta} & \hookrightarrow & F_{t+\delta} \end{array} \qquad \begin{array}{ccc} G_{s-\delta} & \hookrightarrow & G_{s+\delta} \\ & \searrow \quad \nearrow & \\ & F_s & \end{array}$$

It is clear then that the induced commutative diagrams between n-th homology groups imply that the corresponding persistence modules $(H_n(F_r, k))_{r \in \mathbb{R}}$ and $(H_n(G_r, k))_{r \in \mathbb{R}}$ are δ-interleaved. Moreover, by hypothesis, these modules are q-tame. Therefore, the isometry theorem of (4.2) provides the inequality

$$d_\infty(\mathrm{dgm}_n(f), \mathrm{dgm}_n(g)) = d_i(H_n(F_r, k), H_n(G_r, k)) \leq \delta = ||f - g||_\infty.$$

■

If we consider simplicial complexes built on top of point clouds, the above statement transforms as follows.

Theorem 4.1.9 ([22]). *Let $(\mathbb{X}, d^{\mathbb{X}})$ and $(\mathbb{Y}, d^{\mathbb{Y}})$ be two finite metric spaces. Then, for any $n \in \mathbb{N}$,*

$$d_\infty(\mathrm{dgm}_n(\mathbb{X}), \mathrm{dgm}_n(\mathbb{Y})) \leq 2d_{GH}(\mathbb{X}, \mathbb{Y}),$$

where d_{GH} is the Gromov-Hausdorff distance and $\mathrm{dgm}_n(\mathbb{X}), \mathrm{dgm}_n(\mathbb{Y})$ are the persistence diagrams of the Vietoris-Rips or Čech filtrations.

Definition 4.1.10. The **Gromov-Hausdorff distance** between two compact metric spaces (M_1, d_1), (M_2, d_2) is

$$d_{GH}((M_1, d_1), (M_2, d_2)) = \inf_{\varphi_{M_1}, \varphi_{M_2}, M} d_H(\varphi_{M_1}(M_1), \varphi_{M_2}(M_2)),$$

where the infimum is taken over all isometric embeddings $\varphi_{M_1}, \varphi_{M_2}$ of M_1, M_2 into some same metric space (M, d). Indeed, this infimum is a minimum. When both compact spaces belong to the same metric space, $d_{GH} = d_H$.

Proof. Let $\delta = d_{GH}((\mathbb{X}, d^{\mathbb{X}}), (\mathbb{Y}, d^{\mathbb{Y}}))$. By definition, there is a metric space (M, d^M) and two isometric embeddings $\varphi_{\mathbb{X}} : \mathbb{X} \to M$ and $\varphi_{\mathbb{Y}} : \mathbb{Y} \to M$ such that

$$d_H(\varphi_{\mathbb{X}}(\mathbb{X}), \varphi_{\mathbb{Y}}(\mathbb{Y})) = d_{GH}((\mathbb{X}, d^{\mathbb{X}}), (\mathbb{Y}, d^{\mathbb{Y}})) = \delta.$$

It is a standard result that any finite metric space of cardinality m can be isometrically embedded into $(\mathbb{R}^m, l^\infty)$, where l^∞ is the distance defined by the infinity norm. Since both $\mathbb{X}$ and $\mathbb{Y}$ are finite, so is the isometric embedding $\varphi_{\mathbb{X}}(\mathbb{X}) \cup \varphi_{\mathbb{Y}}(\mathbb{Y})$ in M. Denote by m its cardinality. Then, there exists an isometric embedding $\varphi : (\varphi_{\mathbb{X}}(\mathbb{X}) \cup \varphi_{\mathbb{Y}}(\mathbb{Y}), d^M) \to (\mathbb{R}^m, l^\infty)$. It is clear that it satisfies

$$d_H(\varphi \circ \varphi_{\mathbb{X}}(\mathbb{X}), \varphi \circ \varphi_{\mathbb{Y}}(\mathbb{Y})) = d_H(\varphi_{\mathbb{X}}(\mathbb{X}), \varphi_{\mathbb{Y}}(\mathbb{Y})) = \delta.$$

Recall that, by definition of Hausdorff distance 1.3.1, the left hand side is equal to $||l^\infty_{\varphi\circ\varphi_{\mathbb{X}}(\mathbb{X})} - l^\infty_{\varphi\circ\varphi_{\mathbb{Y}}(\mathbb{Y})}||_\infty$, so

$$||l^\infty_{\varphi\circ\varphi_{\mathbb{X}}(\mathbb{X})} - l^\infty_{\varphi\circ\varphi_{\mathbb{Y}}(\mathbb{Y})}||_\infty = \delta.$$

On the other hand, consider the filtration given by $(l^\infty_{\varphi\circ\varphi_{\mathbb{X}}(\mathbb{X})})^{-1}((-\infty, r))$, which is just the union of open l^∞-balls of radius r centered on points of $\varphi\circ\varphi_{\mathbb{X}}(\mathbb{X})$. Note that these balls are convex sets. Therefore, we can apply the persistent nerve theorem as in examples 2.2.6 and 2.2.19. We conclude that this filtration defines a homology persistence module isomorphic to the one of the Čech filtration, $\check{\mathrm{C}}\mathrm{ech}_r(\varphi \circ \varphi_{\mathbb{X}}(\mathbb{X}))$. In particular, both filtrations are tame and have identical persistence diagrams. The same can be deduced for $\varphi\circ\varphi_{\mathbb{Y}}(\mathbb{Y})$. Being tame filtrations implies that, using similar argument to those of the previous theorem, we can prove the following inequality

$$d_\infty(\mathrm{dgm}(l^\infty_{\varphi\circ\varphi_{\mathbb{X}}(\mathbb{X})}), \mathrm{dgm}(l^\infty_{\varphi\circ\varphi_{\mathbb{Y}}(\mathbb{Y})})) \leq ||l^\infty_{\varphi\circ\varphi_{\mathbb{X}}(\mathbb{X})} - l^\infty_{\varphi\circ\varphi_{\mathbb{Y}}(\mathbb{Y})}||_\infty = \delta.$$

Having the same persistence diagrams means that the diagrams of the Čech filtrations of $\varphi \circ \varphi_{\mathbb{X}}(\mathbb{X})$ and $\varphi \circ \varphi_{\mathbb{Y}}(\mathbb{Y})$ are also at bottleneck distance δ. As φ, $\varphi_{\mathbb{X}}$ and $\varphi_{\mathbb{Y}}$ are isometric embeddings, they preserve distances. Consequently, $\check{\mathrm{C}}\mathrm{ech}_r(\varphi \circ \varphi_{\mathbb{X}}(\mathbb{X}), l^\infty) = \check{\mathrm{C}}\mathrm{ech}_r(\mathbb{X}, d^{\mathbb{X}})$. The same equality is deduced for $\mathbb{Y}$. All in all, the Čech filtrations of $\mathbb{X}$ and $\mathbb{Y}$ are at bottleneck distance δ, as wanted.

Lastly, by lemma 2.9 of [22], Čech complexes of parameter r coincide with Rips complexes of parameter $2r$ in $(\mathbb{R}^m, l^\infty)$. Therefore, the filtration $\check{\text{C}}\text{ech}_r(\varphi \circ \varphi_{\mathbb{X}}(\mathbb{X}))$ is exactly the same as $\text{Rips}_{2r}(\varphi \circ \varphi_{\mathbb{X}}(\mathbb{X}))$. Arguing as before, the latter coincide with $\text{Rips}_{2r}(\mathbb{X}, d^{\mathbb{X}})$. Consequently, the persistence diagrams of all these filtrations are the same. Proceeding analogously with $\mathbb{Y}$ we conclude that persistence diagrams of the Rips filtrations of $\mathbb{X}$ and $\mathbb{Y}$ are at bottleneck distance 2δ, as wanted. ■

As a consequence, similar point clouds have similar persistence diagrams. If we interpret persistence diagrams as multiscale topological features, it means that these features are robust with respect to perturbations of the data in the Gromov-Hausdorff metric. Indeed, the lower bound provided by theorem 4.1.9 can be used for classification tasks, as done in [22].

Although the bottleneck distance is extremely useful when studying the stability properties of persistence diagrams, it only depends on the largest distance among pairs of the matching. Hence, the closeness of the remaining points is not taken into account. As a consequence, if γ is an optimal matching, any other matching that does not change the distance between the furthest pair is also optimal. In particular, optimal matchings are not unique in general. Likewise, a perturbation that leaves the length of the longest edge of the matching intact will not be detected by the bottleneck distance. This is not desirable, since discriminating information used in applications may lie in these undetected properties. All in all, a new distance must be introduced.

Definition 4.1.11. The **p-th distance**, or **Wasserstein distance**, between two persistence diagrams dgm^1 and dgm^2 is defined as

$$d_p(\text{dgm}^1, \text{dgm}^2) = \left(\inf_{\gamma \in \Gamma(\text{dgm}^1, \text{dgm}^2)} \sum_{(x,y)\in\gamma} ||x - y||_q^p \right)^{\frac{1}{p}},$$

where $1 \leq p < \infty$.

The infimum is also attained in this case (propositions 4.3.9 and 4.3.10). Now, the cost of a partial matching involves all the edges of the matching, solving the previous problems. For simplicity we will denote both distances as d_p for $1 \leq p \leq \infty$.

There is also a weaker stability result for the p-th distance. Before the statement we need some definitions.

Definition 4.1.12. Given a persistence diagram dgm, the **total persistence** of parameter p is

$$\begin{aligned} \text{Pers}_p(\text{dgm}) &= d_p(\text{dgm}, 0)^p, \qquad \text{if} \quad 1 \leq p \leq \infty, \\ \text{Pers}_\infty(\text{dgm}) &= d_\infty(\text{dgm}, 0) \end{aligned}$$

where 0 denotes the empty diagram (which consists only of the diagonal points with infinite multiplicity).

Definition 4.1.13. Let M be a triangulable, compact metric space. It is said to imply a **bounded degree-q total persistence**, with $1 \leq q < \infty$, if there exists a constant C_M such that

$$\mathrm{Pers}_q(\mathrm{dgm}_n(f)) \leq C_M \mathrm{Lip}(f)^q,$$

for every tame Lipschitz function $f : M \to \mathbb{R}$.

Example 4.1.14 ([23])**.** $\mathbb{S}^n$ implies bounded degree-q total persistence for $q = n + \delta$, for any $\delta > 0$. In particular, whether $n = 1, 2$, it implies bounded degree-n total persistence. ♢

Theorem 4.1.15 ([23])**.** *Let M be a triangulable, compact metric space that implies bounded degree-q total persistence for some real number $1 \leq q < \infty$. Consider $f, g : M \to \mathbb{R}$ two tame Lipschitz functions.*

Then, there exists a constant D_M, that depends on M, such that for all $p \geq q$ and for any $n \in \mathbb{N}$

$$d_p(\mathrm{dgm}_n(f), \mathrm{dgm}_n(g)) \leq C^{1/p} ||f - g||_\infty^{1-\frac{q}{p}},$$

where $C = D_M \max\{\mathrm{Lip}(f)^q, \mathrm{Lip}(g)^q\}$.

Proof. Consider without loss of generality the p-th distance with the infinity norm. Let γ be a matching that realizes the bottleneck distance

$$d_\infty(\mathrm{dgm}_n(f), \mathrm{dgm}_n(g)) = \max_{x \in \mathrm{dgm}_n(f)} ||x - \gamma(x)||_\infty,$$

where $\gamma(x)$ is the point of $\mathrm{dgm}_n(g)$ paired with x. Denote $\varepsilon = ||f - g||_\infty$. Applying theorem 4.1.8 we obtain $\max_{x \in \mathrm{dgm}_n(f)} ||x - \gamma(x)||_\infty \leq \varepsilon$. Therefore, for every $x \in \mathrm{dgm}_n(f)$ we have that

$$||x - \gamma(x)||_\infty \leq \varepsilon. \tag{4.3}$$

Moreover, it is possible to assume that all these points also satisfy

$$||x - \gamma(x)||_\infty \leq ||x - \Delta_x||_\infty + ||\gamma(x) - \Delta_{\gamma(x)}||_\infty, \tag{4.4}$$

where by $||x - \Delta_x||_\infty$ we mean the distance between a point and its orthogonal projection onto the diagonal. Indeed, if a point $x \in \mathrm{dgm}_n(f)$ does not satisfy inequality (4.4), then

$$||x - \Delta_x||_\infty + ||\gamma(x) - \Delta_{\gamma(x)}||_\infty < ||x - \gamma(x)||_\infty,$$

which implies that $||x - \Delta_x||_\infty < ||x - \gamma(x)||_\infty$ and $||x - \Delta_{\gamma(x)}||_\infty < ||x - \gamma(x)||_\infty$. But then, matching x and $\gamma(x)$ with their projections costs less than pairing x with $\gamma(x)$. Thus, without risking optimality, we can assume (4.4) is satisfied for every $x \in \mathrm{dgm}_n(f)$. Then, given any $p \geq q$ and using inequalities (4.3) and (4.4), we can write

$$\begin{aligned} d_p(\mathrm{dgm}_n(f),\mathrm{dgm}_n(g))^p &\leq \sum_{x \in \mathrm{dgm}_n(f)} ||x - \gamma(x)||_\infty^p \leq \sum_{x \in \mathrm{dgm}_n(f)} \varepsilon^{p-q} ||x - \gamma(x)||_\infty^q \\ &\leq \varepsilon^{p-q} \sum_{x \in \mathrm{dgm}_n(f)} \left(||x - \Delta_x||_\infty + ||\gamma(x) - \Delta_{\gamma(x)}||_\infty\right)^q \\ &= 2^q \varepsilon^{p-q} \sum_{x \in \mathrm{dgm}_n(f)} \left(\frac{||x - \Delta_x||_\infty + ||\gamma(x) - \Delta_{\gamma(x)}||_\infty}{2}\right)^q \end{aligned}$$

For $q \geq 1$ the map $\varphi(u) = u^q$ is convex in $[0, \infty)$, thus

$$\varphi\left(\frac{u+v}{2}\right) \leq \frac{\varphi(u)+\varphi(v)}{2}.$$

Applying this we obtain

$$\begin{aligned} d_p(\mathrm{dgm}_n(f),\mathrm{dgm}_n(g))^p &\leq 2^q \varepsilon^{p-q} \sum_{x \in \mathrm{dgm}_n(f)} \frac{1}{2}(||x - \Delta_x||_\infty^q + ||\gamma(x) - \Delta_{\gamma(x)}||_\infty^q) \\ &= 2^{q-1}\varepsilon^{p-q} \left(\sum_{x \in \mathrm{dgm}_n(f)} ||x - \Delta_x||_\infty^q + \sum_{x \in \mathrm{dgm}_n(f)} ||\gamma(x) - \Delta_{\gamma(x)}||_\infty^q \right) \\ &= 2^{q-1}\varepsilon^{p-q} \left(\mathrm{Pers}_q(\mathrm{dgm}_n(f)) + \mathrm{Pers}_q(\mathrm{dgm}_n(g))\right). \end{aligned}$$

The last equality follows from the fact that an optimal matching between a persistence diagram and the empty one for the infinite norm consists in pairing each point with its orthogonal projection onto the diagonal. By hypothesis on the space M and definition 4.1.13, there exists a constant C_M such that

$$\begin{aligned} d_p(\mathrm{dgm}_n(f), \mathrm{dgm}_n(g))^p &\leq 2^{q-1}\varepsilon^{p-q} C_M [\mathrm{Lip}(f)^q + \mathrm{Lip}(g)^q] \\ &\leq 2^q \varepsilon^{p-q} C_M \max\{\mathrm{Lip}(f)^q, \mathrm{Lip}(g)^q\} \end{aligned}$$

Finally, recalling that $\varepsilon = ||f - g||_\infty$, we get

$$d_p(\mathrm{dgm}_n(f), \mathrm{dgm}_n(g)) \leq C^{1/p} ||f - g||_\infty^{1-\frac{q}{p}},$$

where $C = 2^q C_M \max\{\mathrm{Lip}(f)^q, \mathrm{Lip}(g)^q\}$. ■

Note that this theorem just establishes Hölder continuity, not Lipschitz. Moreover, a constant upper bound is obtained only when $p = 1$ (and hence $q = 1$).

4.2 The space of persistence diagrams

So far, we have studied persistence diagrams of q-tame modules, which consists of a locally finite multiset of $\{(r_1, r_2) \in \mathbb{R}^2 \ : \ r_1 < r_2\}$ together with the diagonal with infinite multiplicity. Recall that the essential part has been removed, since it can be treated separately when comparing diagrams. Let us introduce the notation

$$\Omega := \{(r_1, r_2) \in \mathbb{R}^2 \ : \ r_1 < r_2\}, \qquad \partial\Omega := \{(r, r) \ : \ r \in \mathbb{R}\}.$$

Recovering the measure-theoretic point of view of remark 3.2.11, these persistence diagrams can be expressed as point measures supported on Ω

$$\nu = \sum_{x \in X} n_x \delta_x,$$

where $n_x \in \mathbb{N}$ is the multiplicity of the point, δ_x denotes the Dirac mass located on $\{x\}$ and X are the points of the locally finite multiset $\mathrm{dgm}(\mathbb{V}) \cap \Omega$, counted

without multiplicity. Furthermore, most of the definitions and results of the previous subsection are general enough to be applied to any point measure on Ω. Indeed, partial matchings between point measures are well-defined, considering the locally finite sets X counted with multiplicity. Hence, so are distances d_p for $1 \leq p \leq \infty$, where the infimum is also attained. As a consequence, given two persistence diagrams of q-tame modules dgm^1 and dgm^2 and their associated point measures μ_1 and μ_2, we have that $d_p(\mathrm{dgm}^1, \mathrm{dgm}^2) = d_p(\mu_1, \mu_2)$, for any $1 \leq p \leq \infty$.

In conclusion, it is possible to work in the more general framework of point measures, with d_p distances between them. Until the end of this chapter we will adopt this measure-theoretic viewpoint. As happened with persistence modules, this generalization will lead to interesting results useful in the practical case of studying a data set.

Firstly, some notions must be introduced.

Definition 4.2.1. The set $\mathcal{D}$ is the set of all point measures supported on Ω. Its elements will be called persistence diagrams too.

As we have explained before, this definition is consistent with persistence diagrams of q-tame modules without their essential parts.

Definition 4.2.2. The **space of persistence diagrams** of parameter p is defined by

$$\mathcal{D}^p = \{\mu \in \mathcal{D} \ : \ \mathrm{Pers}_p(\mu) < \infty\},$$

where $1 \leq p \leq \infty$. The total persistence of μ of parameter p, $\mathrm{Pers}_p(\mu)$, is defined as in definition 4.1.12.

Observe that, given $\mu = \sum_{x \in X} n_x \delta_x$, for $1 \leq p < \infty$ and $q \geq 1$, we have

$$\mathrm{Pers}_p(\mu) = \sum_{x \in X} n_x ||x - \partial\Omega||_q^p,$$

where $||x - \partial\Omega||_q$ denotes the distance between $x \in \Omega$ and its orthogonal projection onto the diagonal. If $p = \infty$,

$$\mathrm{Pers}_\infty(\mu) = \sup_{x \in X} ||x - \partial\Omega||_q.$$

Therefore, $\mathcal{D}^p$ is the set of persitence diagrams that are at finite d_p distance from the empty diagram. As we will see, this ensures that the space of persistence diagrams is a metric space.

Remark 4.2.3. In practice, persistence diagrams coming from data sets are finite, so they belong to $\mathcal{D}^p$. However, not all persistence diagrams of q-tame modules do. For instance, $\mathbb{V} = \bigoplus_{n \in \mathbb{N}} k[n, 2n]$ is q-tame, but its persistence diagram consists of all the points $(n, 2n) \in \mathbb{R}^2$. Hence, it is not at finite distance from the empty diagram if $p < \infty$. $\diamondsuit$

Lemma 4.2.4. *$(\mathcal{D}^p, d_p)$ is a metric space, where $1 \leq p \leq \infty$.*

Proof. Given two persistence diagrams $\mu, \nu \in \mathcal{D}^p$, it is clear that a subset γ of $(X \cup \partial\Omega) \times (Y \cup \partial\Omega)$ is a partial matching between μ and ν if and only if, as a subset of $(Y \cup \partial\Omega) \times (X \cup \partial\Omega)$, it is a partial matching between ν and μ. Hence, $\mathrm{d}_p(\mu, \nu) = \mathrm{d}_p(\nu, \mu)$.

Let us show that $\mathrm{d}_p(\mu, \nu) = 0$ if and only if $\mu = \nu$. On the one hand, if $\mu = \nu$ we can always take the partial matching $\gamma = \{(x, x) \ : \ x \in X \cup \partial\Omega\}$. Since $||x - x||_q = 0$ and $\mathrm{d}_p(\mu, \nu) \geq 0$, we conclude that $\mathrm{d}_p(\mu, \nu) = 0$. On the other hand, suppose $d_p(\mu, \nu) = 0$. That means there exists a partial matching $\gamma \in \Gamma(\mu, \nu)$ such that $||x - y||_q = 0$ for every $(x, y) \in \gamma$. That is, $x = y$ and hence $\mu = \nu$.

For the triangle inequality, consider three persistence diagrams $\mu, \nu, \tau \in \mathcal{D}^p$ with locally finite multisets X, Y and Z respectively. Given any partial matchings $\gamma_{\mu,\nu}, \gamma_{\nu,\tau}$, between μ and ν and between ν and τ respectively, we can build a partial matching $\gamma_{\mu,\tau}$ between μ and τ. Indeed, for any $x \in X$ (points are counted with multiplicity) there exists a unique pair $(x, y_x) \in \gamma_{\mu,\nu}$. If $y_x \in \partial\Omega$, take it as unique pair of $\gamma_{\mu,\tau}$ with x in the first coordinate. If $y_x \in Y$, there exists a unique $(y_x, z_{y_x}) \in \gamma_{\nu,\tau}$. Choose then (x, z_{y_x}) as the x-pair of $\gamma_{\mu,\tau}$. For the remaining $z \in Z$ that do not appear in any chosen pair for $\gamma_{\mu,\tau}$, repeat the process in the other direction. There is a unique $(y_z, z) \in \gamma_{\nu,\tau}$. If $y_z \in \partial\Omega$, take it as z-pair. If not, it exists a unique pair $(x_{y_z}, y_z) \in \gamma_{\mu,\nu}$. Then, take $(x_{y_z}, z) \in \gamma_{\mu,\tau}$. By this construction any $x \in X$ and $z \in Z$ appears exactly once in the chosen pairs. Therefore, $\gamma_{\mu,\tau}$ is a partial matching.

Let $\gamma_{\mu,\nu}$ and $\gamma_{\nu,\tau}$ be optimal partial matchings. In the case of the bottleneck distance, for any $(x, z) \in \gamma_{\mu,\tau}$ we can write

$$\begin{aligned} ||x - z||_\infty &\leq ||x - y||_\infty + ||y - z||_\infty \\ &\leq \max_{(x,y)\in\gamma_{\mu,\nu}} ||x - y||_\infty + \max_{(y,z)\in\gamma_{\nu,\tau}} ||y - z||_\infty = d_\infty(\mu, \nu) + d_\infty(\nu, \tau). \end{aligned}$$

Therefore,

$$d_\infty(\mu, \tau) \leq \max_{(x,z)\in\gamma_{\mu,\tau}} ||x - z||_\infty \leq d_\infty(\mu, \nu) + d_\infty(\nu, \tau),$$

as wanted. For $1 \leq p < \infty$, a similar result is obtained using Minkowski inequality

$$\begin{aligned} d_p(\mu, \tau) &\leq \left(\sum_{(x,z)\in\gamma_{\mu,\tau}} ||x - z||_q^p \right)^{1/p} \\ &\leq \left(\sum_{(x,y)\in\gamma_{\mu,\nu}} ||x - y||_q^p \right)^{1/p} + \left(\sum_{(y,z)\in\gamma_{\nu,\tau}} ||y - z||_q^p \right)^{1/p} = d_p(\mu, \nu) + d_p(\nu, \tau). \end{aligned}$$

The assumption on the total persistence ensures that $\mathrm{d}_p(\mu, \nu)$ is finite for any $\mu, \nu \in \mathcal{D}^p$. Indeed,

$$d_p(\mu, \nu) \leq d_p(\mu, 0) + d_p(0, \nu) = \mathrm{Pers}_p(\mu)^{1/p} + \mathrm{Pers}_p(\nu)^{1/p} < \infty.$$

All in all, $d_p : \mathcal{D}^p \times \mathcal{D}^p \to [0, \infty)$ is a metric and $(\mathcal{D}^p, \mathrm{d}_p)$ is a metric space. ■

Whence, we equip the space $\mathcal{D}^p$ with the distance d_p.

Remark 4.2.5. Consider a data set and the induced q-tame module at the homology level. Its persistence diagram is an element of $\mathcal{D}^p$. Therefore, we have managed to map the entiere data set to a single point in a metric space. As a result, qualitative information (topology) has transformed into quantitaive (distances). ♢

Proposition 4.2.6 ([24])**.** *For $1 \leq p < \infty$ the space $(\mathcal{D}^p, d_p)$ is a Polish metric space, i.e. is complete and separable in the metric d_p.*

Proof. Completness can be found in theorem 6 of [24]. Let us prove separability.

In order to show this, we have to find a countable dense subset in $\mathcal{D}^p$. Denote by $|\mu|$ the cardinality of the off-diagonal part of a persistence diagram $\mu \in \mathcal{D}^p$. For $\alpha > 0$, consider the map $u_\alpha : \mathcal{D}^p \to \mathcal{D}^p$ such that $x \in u_\alpha(\mu)$ if and only if $x \in \mu$ and $||x - \partial\Omega||_q \geq \alpha$. Observe that, since $\text{Pers}_p(\mu) < \infty$, the cardinality of $u_\alpha(\mu)$ is finite for every $\alpha > 0$. Similarly, define $l_\alpha : \mathcal{D}^p \to \mathcal{D}^p$, where $x \in l_\alpha(\mu)$ if and only if $x \in \mu$ and $||x - \partial\Omega||_q < \alpha$. Now, let

$$S = \{\mu \in \mathcal{D}^p \ : \ |\mu| < \infty, \quad x \in \mathbb{Q}^2 \ \ \forall x \in \mu\}.$$

Notice that $S = \bigcup_{m=0}^{\infty} S_m$, where $S_m = \{\mu \in S \ : \ |\mu| = m\}$. Each S_m is isomorphic to a subset of $\mathbb{Q}^{2m}$, so countable. Hence, S is countable. We just have to prove that it is dense in $\mathcal{D}^p$ for the p-th distance. That is, given $\nu \in \mathcal{D}^p$, for all $\varepsilon > 0$ we have to find $\mu \in S$ such that $d_p(\nu, \mu) < \varepsilon$.

Fix $\nu = \sum_{x \in X} n_x \delta_x$ and $\varepsilon > 0$. Notice that

$$d_p(l_\alpha(\nu), 0)^p = \sum_{\{x \in X : ||x - \partial\Omega||_q < \alpha\}} ||x - \partial\Omega||_q^p.$$

Therefore, there exists $\alpha > 0$ such that $d_p(l_\alpha(\nu), 0) < \varepsilon/2$. Since the points of $u_\alpha(\nu)$ belong also to ν, a possible matching between $u_\alpha(\nu)$ and ν is to pair together the commom points and the remaining ones (i.e. $l_\alpha(\nu)$) with the projection onto the diagonal. Note that the cost of this matching is equal to $d_p(l_\alpha(\nu), 0)$. When computing the p-th distance we search for the minimum, so

$$d_p(\nu, u_\alpha(\nu)) \leq d_p(l_\alpha(\nu), 0) < \varepsilon/2.$$

On the other hand, the persistence diagram $u_\alpha(\nu)$ can be seen as a point in $\mathbb{R}^{2|u_\alpha(\nu)|}$ (recall $|u_\alpha(\nu)|$ is finite). As $\mathbb{Q}^{2|u_\alpha(\nu)|}$ is dense in $\mathbb{R}^{2|u_\alpha(\nu)|}$, we can find a point in the former, which corresponds to a persistence diagram $\mu \in S$ and, such that $d_p(\mu, u_\alpha(\nu)) < \varepsilon/2$. All in all,

$$d_p(\nu, \mu) \leq d_p(\nu, u_\alpha(\nu)) + d_p(u_\alpha(\nu), \mu) < \varepsilon/2 + \varepsilon/2 = \varepsilon,$$

as wanted. ■

In spite of the fact that persistence diagrams must remain locally finite and have finite total persistence, they may have infinitely many points. Indeed, this assumption is necessary for $\mathcal{D}^p$ to be complete, as the next example shows.

Example 4.2.7. The limit of a sequence of finite persistence diagrams can have infinitely many points. Consider $\mu_n = \sum_{k=1}^{n} \delta_{(0,2^{-k})}$. This is a Cauchy sequence because $d_p(\mu_n, \mu_m) \leq \sum_{k=n+1}^{m} 2^{-k} = 2^{-n} - 2^{-m}$ and the sequence $\{2^{-k}\}_{k\in\mathbb{N}}$ is Cauchy. However, the number of off-diagonal points in μ_n grows to infinity as $n \to \infty$. That is, in the limit we have $\sum_{k\geq 1} \delta_{(0,2^{-k})}$ which has infinitely many points, though it belongs to $\mathcal{D}^p$. ♢

Remark 4.2.8 ([6])**.** Although the space $(\mathcal{D}^\infty, d_\infty)$ is complete (corollary 4.3.16), it is not separable for the bottleneck distance. In fact, let $I \subset \mathbb{N}$ and define the persistence diagram $a_I = \sum_{i\in I} \delta_{(i,i+1)}$. For all $I \subset \mathbb{N}$, $a_I \in \mathcal{D}^\infty$. Consider the uncountable family $\mathcal{A} = \{a_I \ : \ I \subset \mathbb{N}\} \subset \mathcal{D}^\infty$. Notice that $d_\infty(a_I, a_{I'}) = \sqrt{2}/2$ for any two distinct $I, I' \subset \mathbb{N}$, as an optimal matching consists in pairing all points with the diagonal. Then, let S be an arbitrary dense subset in $\mathcal{D}^\infty$. By density, for each $a_I \in \mathcal{A}$ there exists $\mu_I \in S$ such that $d_\infty(\mu_I, a_I) < \sqrt{2}/4$. However, for any $I' \neq I$, $d_\infty(\mu_I, a_{I'}) \geq d_\infty(a_I, a_{I'}) - d_\infty(a_I, \mu_I) > \sqrt{2}/4$. Therefore, inside S we must find a different persistence diagram μ_I for each a_I. This implies that S cannot be countable, so $\mathcal{D}^\infty$ is not separable. ♢

4.3 Optimal transport

The natural next stage is to introduce distances between more general measures (not only pointed ones) and use them to compare persistence diagrams. To this aim, we will present a distance based on optimal transport. Firstly, the idea behind optimal transport and the formulation that concerns us are explained. After this brief introduction, we will deal directly with the case which allows us to compare persistence diagrams. For further reading we refer to [25], [5] and [6].

4.3.1 Unbalanced optimal transport

Intuitively, two measures μ and ν represent two ways of distributing mass. Consider the case where transporting mass from the first configuration to the second has some cost. The goal is to find the optimal way of transporting the mass from the first to the second configuration, minimizing this cost. The classical formulation of the transport problem is the Kantorovich formulation: consider $\mathcal{X}$ and $\mathcal{Y}$ two Polish spaces and μ, ν two Borel non-negative finite measures supported on $\mathcal{X}, \mathcal{Y}$, respectively. Let $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_+$ be a lower-semicontinuous non-negatively valued cost function. The optimal transport problem reads

$$\inf_{\pi\in\Pi(\mu,\nu)} \left(\int_{\mathcal{X}\times\mathcal{Y}} c(x,y) d\pi(x,y) \right), \tag{4.5}$$

where

$$\Pi(\mu,\nu) = \{\pi \ \text{ measure on } \mathcal{X}\times\mathcal{Y} \ : \ p_1\#\pi = \mu, \ p_2\#\pi = \nu\},$$

is the set of **transport plans** between μ and ν, for $p_1 : \mathcal{X} \times \mathcal{Y} \to \mathcal{X}$ and $p_2 : \mathcal{X} \times \mathcal{Y} \to \mathcal{Y}$ the natural projections. The pushforward conditions are equivalent to the fact that, for any Borel subset $A \subset \mathcal{X}, B \subset \mathcal{Y}$,

$$\pi(A \times \mathcal{Y}) = \mu(A), \qquad \pi(\mathcal{X} \times B) = \nu(B).$$

Observe that both measures are forced to have the same mass

$$\mu(\mathcal{X}) = p_1\#\pi(\mathcal{X}) = \pi(\mathcal{X}\times\mathcal{Y}) = p_2\#\pi(\mathcal{Y}) = \nu(\mathcal{Y}).$$

As a matter of fact, transport plans always exist, it is enough to consider the product measure $\frac{1}{\mu(\mathcal{X})}\mu\otimes\nu$. The advantage of this formulation is that optimal transport plans do always exist, i.e. the minimum is attained (see theorem 4.1 of [25]). Besides, in some important situations, optimal transport plans and maps coincide, as stated in theorem 9.4 of [25].

Theorem 4.3.1. *Consider two probability measures ρ, μ in $\mathbb{R}^2$ with finite second moment and let $c(x,y) = ||x-y||^2$ be the cost function. Assume ρ is absolutely continuous with respect to the Lebesgue measure. Then, the optimal transport plan $\pi \in \Pi(\rho,\mu)$ is unique and it is induced by a transport map. That is, there exists a measurable map $T_\mu : \mathbb{R}^2 \to \mathbb{R}^2$, which is the gradient of a convex function $T_\mu = \nabla\phi$, such that $\pi = (id, T_\mu)\#\rho$.*

With this formulation, distances between measures can be defined, e.g. the p-Wasserstein distance, where the cost function is $d(x,y)^p$,

$$W_p(\mu,\nu) = \left(\inf_{\pi\in \mathrm{Adm}(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} d(x,y)^p d\pi(x,y)\right)^{1/p}. \tag{4.6}$$

Nevertheless, persistence diagrams may have different number of off-diagonal points, and even infinite. Therefore, Kantorovich formulation is not enough for us, since it is only capable of comparing measures of same finite mass. Although some tries have been made in order to extend this theory to measures with different mass, the more interesting is the one developed by Figalli, A. in [5], since it also deals with Radon measures of infinite mass and so overcomes the two problems. This approach is developed for Radon measures supported on a bounded open proper subset $\mathcal{X}$ of $\mathbb{R}^m$. The idea is to consider this set together with its boundary $\bar{\mathcal{X}} = \mathcal{X} \sqcup \partial\mathcal{X}$. Intuitively, when transporting the mass from the first configuration to the second, we can use $\partial\mathcal{X}$ as an infinite reservoir. In this way, we can take as much mass as we wish from the boundary or give it back, as long as we pay the cost. The formulation of this optimal transport problem is as follows: given μ, ν two Radon measures supported on $\mathcal{X}$, the problem reads

$$\inf_{\pi\in \mathrm{Adm}(\mu,\nu)} \left(\int_{\bar{\mathcal{X}}\times\bar{\mathcal{X}}} c(x,y) d\pi(x,y)\right), \tag{4.7}$$

where $\mathrm{Adm}(\mu,\nu)$ is the set of **admissible transport plans** defined as

$$\mathrm{Adm}(\mu,\nu) = \left\{\pi \ \text{ measure on } \bar{\mathcal{X}}\times\bar{\mathcal{X}} \ : \ p_1\#\pi|_{\mathcal{X}} = \mu,\ p_2\#\pi|_{\mathcal{X}} = \nu\right\},$$

being $p_i : \bar{\mathcal{X}}\times\bar{\mathcal{X}} \to \bar{\mathcal{X}}$ the natural projections. Again, the pushforward restrictions are equivalent to the fact that, for any Borel sets $A, B \subset \mathcal{X}$,

$$\pi(A\times\bar{\mathcal{X}}) = \mu(A), \qquad \pi(\bar{\mathcal{X}}\times B) = \nu(B). \tag{4.8}$$

It is always possible to define an admissible transport plan, the trivial one which sends all mass of μ and ν to the boundary. In addition, optimal plans always exists too, as shown in [5]. Similarly to Kantorovich formulation and the p-Wasserstein distance, it is also possible to define a distance between measures using (4.7), which is called optimal transport distance.

Notice that the approach taken in [5] is developed for $\mathcal{X}$ bounded, while persistence diagrams are supported on Ω, obviously not bounded. Luckily for us, this optimal transport problem can be extended to Ω with additional care. This is done in the next subsection for a larger set of measures, not only $\mathcal{D}^p$. Of course, implications of the results obtained will be highlighted for the particular case of persistence diagrams. Consequently, it will be possible to compare the optimal transport distance with the d_p distances defined in the previous sections.

4.3.2 Optimal transport distance

Given Ω the open half-plane and $\partial\Omega$ the diagonal $\{(r,r) \ : \ r \in \mathbb{R}\}$, consider $\bar{\Omega} = \Omega \sqcup \partial\Omega$. Denote by $d : \bar{\Omega} \times \bar{\Omega} \to [0,\infty)$ the metric $||\cdot||_q$ on $\bar{\Omega}$ with $1 \leq q \leq \infty$, though other distances may be used. Consider $p_{\partial\Omega} : \Omega \to \partial\Omega$ the orthogonal projection onto the diagonal. Letting $\mathcal{M}(\Omega)$ be the set of non-negative Radon measures supported on Ω, we introduce the following notion.

Definition 4.3.2. Given $\mu \in \mathcal{M}(\Omega)$ the **total persistence** of parameter p is defined as

$$\mathrm{Pers}_p(\mu) = \int_\Omega d(x,\partial\Omega)^p d\mu(x), \qquad \text{if } \ 1 \leq p < \infty,$$

and[1]

$$\mathrm{Pers}_\infty(\mu) = \sup_{x \in \mathrm{spt}(\mu)} d(x,\partial\Omega), \qquad \text{if } \ p = \infty.$$

Remark 4.3.3. Let μ be a Borel measure on Ω satisfying $\mathrm{Pers}_p(\mu) < \infty$, where $1 \leq p < \infty$. Then, for any Borel set $A \subset \Omega$ such that $d(A,\partial\Omega) := \inf_{x\in A} d(x,\partial\Omega) > 0$, we have

$$\mu(A) d(A,\partial\Omega)^p \leq \int_A d(x,\partial\Omega)^p d\mu(x) \leq \int_\Omega d(x,\partial\Omega)^p d\mu(x) = \mathrm{Pers}_p(\mu) < \infty. \quad (4.9)$$

This implies that $\mu(A) < \infty$. In particular, μ is automatically a Radon measure, because for any compact set K of Ω, $d(K,\partial\Omega) > 0$. ♢

Definition 4.3.4. The space of **persistence measures** of parameter p, with $1 \leq p \leq \infty$, is the space of non-negative Radon measures supported on Ω that have finite total persistence, i.e.

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

[1]The support of a measure μ are the points x such that if U_x is any open neighborhood of x, then $\mu(U_x) > 0$.

Notice that $\mathcal{D}^p$, the space of persistence diagrams of parameter p given by definition 4.2.2, is the subspace of $\mathcal{M}^p$ consisting only of point measures. As it happened with $\mathcal{D}^p$, the finiteness assumption on the total persistence of $\mu \in \mathcal{M}^p$ will ensure that the optimal transport distance, defined below, is finite.

Definition 4.3.5. The **optimal transport distance** between two measures $\mu, \nu \in \mathcal{M}^p$, with $1 \leq p < \infty$, is

$$\mathrm{OT}_p(\mu,\nu) = \left(\inf_{\pi \in \mathrm{Adm}(\mu,\nu)} \int_{\bar{\Omega}\times\bar{\Omega}} d(x,y)^p d\pi(x,y) \right)^{1/p} = \left(\inf_{\pi \in \mathrm{Adm}(\mu,\nu)} C_p(\pi) \right)^{1/p}.$$

For $p = \infty$, given two measures $\mu, \nu \in \mathcal{M}^\infty$ we define

$$\mathrm{OT}_\infty(\mu,\nu) = \inf_{\pi \in \mathrm{Adm}(\mu,\nu)} \left(\sup_{(x,y)\in \mathrm{spt}(\pi)} d(x,y) \right) = \inf_{\pi \in \mathrm{Adm}(\mu,\nu)} C_\infty(\pi).$$

Admisible transport plans for which the infimum is attained are called **optimal**. The set of these optimal transport plans is denoted by $\mathrm{Opt}_p(\mu,\nu)$.

Observe that OT_p is defined for Radon measures of infinite mass, as long as their total persistence is finite.

Remark 4.3.6. Consider a finite collection of finite point measures $\{\nu_i\}_{i=1}^n$ on Ω (e.g. comes from the computation of the persistence diagrams of a collection of samples). Thanks to finiteness, an open bounded subset $\mathcal{Y}$ of Ω containing the support of all the measures can be obtained. It is possible to choose $\mathcal{Y}$ as a triangle with one of the boundaries on $\partial\Omega$ and such that $d(x,\partial\mathcal{Y}) = d(x,\partial\Omega)$ for all x that belongs to the supports. This choice implies that, computing the optimal transport distance between ν_i and ν_j as in [5] (using all the boundary $\partial\mathcal{Y}$ as a reservoir) gives the same result and same optimal admissible transport plan as computing the optimal transport in Ω following definition 4.3.5, since sending mass to $\partial\mathcal{Y} \cap \partial\Omega$ is the cheapest. Hence, when analyzing data sets, persistence diagrams may be considered as finite point measures on a bounded subset $\mathcal{Y}$ of Ω and results from [5] may be directly applied. ♢

Optimal admissible transport plans $\pi \in \mathrm{Opt}_p(\mu,\nu)$ can be assumed to be supported on $E_\Omega := \bar{\Omega}\times\bar{\Omega}\backslash\partial\Omega\times\partial\Omega$. That is, $\pi|_{\partial\Omega\times\partial\Omega} = 0$. Indeed, it is easy to check that the measure $\pi - \pi|_{\partial\Omega\times\partial\Omega}$ is also an admissible transport plan for μ and ν. For example, for any $A \subseteq \Omega$ Borel set

$$(\pi - \pi|_{\partial\Omega\times\partial\Omega})(A\times\bar{\Omega}) = \pi(A\times\bar{\Omega}) - \pi((A\times\bar{\Omega})\cap(\partial\Omega\times\partial\Omega)) = \pi(A\times\bar{\Omega}) - \pi(\emptyset) = \mu(A).$$

Moreover, it is clear that $C_p(\pi - \pi|_{\partial\Omega\times\partial\Omega}) \leq C_p(\pi)$ for any $1 \leq p \leq \infty$.

A relevant characteristic of this formulation is that optimal admissible transport plans always exist.

Proposition 4.3.7 ([6])**.** *Let $\mu, \nu \in \mathcal{M}^p$ with $1 \leq p \leq \infty$. Then, $\mathrm{Opt}_p(\mu,\nu)$ is nonempty.*

Proof. Assume for the moment that $\mathrm{Adm}(\mu,\nu)$ is sequentially compact (1) for the vague topology and $C_p : \mathrm{Adm}(\mu,\nu) \to \mathbb{R}$ is lower semicontinuous w.r.t vague topology (2). Let $\{\pi_n\}_n$ be a minimizing sequence of admissible plans, i.e.

$$\liminf_{n\to\infty} C_p(\pi_n) = \mathrm{OT}_p^p(\mu,\nu).$$

Since $\mathrm{Adm}(\mu,\nu)$ is sequentially compact, there exists a subsequence $\{\pi_{n_k}\}_k$ that converges vaguely to an admissible plan π. By lower semicontinuity of C_p, we have that

$$C_p(\pi) \le \liminf_{k\to\infty} C_p(\pi_{n_k}) = \mathrm{OT}_p^p(\mu,\nu).$$

But the optimal transport distance is defined as the infimum of the $C_p(\pi)$. Therefore, $\mathrm{OT}_p^p(\mu,\nu) = C_p(\pi)$, and then $\pi \in \mathrm{Opt}_p(\mu,\nu)$. To finish the proof we need to show the assumptions we made at the beginning:

1. Admissible transport plans may be assumed to be supported on E_Ω, so it is enough to prove sequentially compactness there. We first show that $\mathrm{Adm}(\mu,\nu)$ is relatively compact for the vague topology, using proposition B.0.10. Secondly, we prove that it is closed. This concludes the proof.

   Let $K \subset E_\Omega$ be an arbitrary compact set. It is easy to see that there always exist two compact sets $K_1, K_2 \subset \Omega$ such that $K \subset (K_1 \times \bar{\Omega}) \cup (\bar{\Omega} \times K_2)$. Let $\pi \in \mathrm{Adm}(\mu,\nu)$. Then,

   $$\begin{aligned} \pi(K) &\le \pi((K_1 \times \bar{\Omega}) \cup (\bar{\Omega} \times K_2)) \\ &\le \pi(K_1 \times \bar{\Omega}) + \pi(\bar{\Omega} \times K_2) = \mu(K_1) + \nu(K_2) < \infty. \end{aligned} \tag{4.10}$$

   Therefore,

   $$\sup\{\pi(K) \ : \ \pi \in \mathrm{Adm}(\mu,\nu)\} < \infty$$

   is satisfied for any compact set K of E_Ω. By proposition B.0.10 we conclude that $\mathrm{Adm}(\mu,\nu)$ is relatively compact.

   On the other hand, let $\{\pi_n\}_n \subset \mathrm{Adm}(\mu,\nu)$ such that $\pi_n \xrightarrow{v} \pi \in \mathcal{M}(E_\Omega)$. In order to prove that π is admissible, we need to verify that both marginals of π on Ω are still μ and ν. Notice that $p_1\#\pi_n|_\Omega = \mu$ for all $n \in \mathbb{N}$. Thus, if $p_1\#\pi_n|_\Omega$ converges vaguely to $p_1\#\pi|_\Omega$ in $\mathcal{M}(E_\Omega)$, then $p_1\#\pi|_\Omega = \mu$ as wanted. By (B.2) we just have to show that, given any $f \in C_c(\Omega)$,

   $$\lim_{n\to\infty} \varphi_f(p_1\#\pi_n|_\Omega) = \varphi(p_1\#\pi|_\Omega).$$

   In fact, since π_n converges vaguely to π, using (B.2) again we get

   $$\begin{aligned} \lim_{n\to\infty} \varphi_f(p_1\#\pi_n|_\Omega) &= \lim_{n\to\infty} \int_\Omega f(x) d(p_1\#\pi_n|_\Omega)(x) = \lim_{n\to\infty} \int_{E_\Omega} f(x) d\pi_n(x,y) \\ &= \int_{E_\Omega} f(x) d\pi(x,y) = \int_\Omega f(x) d(p_1\#\pi|_\Omega)(x) = \varphi_f(p_1\#\pi|_\Omega). \end{aligned}$$

   Analogously we obtain that the second marginal of π on Ω is ν. All in all, $\pi \in \mathrm{Adm}(\mu,\nu)$ and the set of admissible transport plans is closed in $\mathcal{M}(E_\Omega)$.

2. We start proving lower semicontinuity for $p \in [1, \infty)$. Let $\{\pi_n\}_n \subset \mathrm{Adm}(\mu, \nu)$ such that $\pi_n \xrightarrow{v} \pi$. We have seen that $\pi \in \mathrm{Adm}(\mu, \nu)$. Consider the measure on $\bar{\Omega} \times \bar{\Omega}$ given by

$$\pi_n'(A) = \int_A d(x, y)^p d\pi_n(x, y).$$

Indeed, we can assume $\pi_n \in \mathcal{M}(E_\Omega)$, so π_n' is also supported on E_Ω. Moreover, it is a Radon measure. This is deduced from the fact that, for any compact set $K \subset E_\Omega$, since K is bounded and inequality (4.10) is satisfied,

$$\pi_n'(K) = \int_K d(x, y)^p d\pi_n(x, y) \leq D\pi_n(K) < \infty.$$

All in all, $\pi_n' \in \mathcal{M}(E_\Omega)$ for every $n \in \mathbb{N}$. Notice that

$$\pi_n'(E_\Omega) = C_p(\pi_n), \qquad \pi'(E_\Omega) = C_p(\pi).$$

Using the characterization of vague convergence (B.2) we have that, for any bounded Borel set $A \subset E_\Omega$ such that $\pi(\partial A) = 0$,

$$\begin{aligned}\lim_{n\to\infty} \pi_n'(A) &= \lim_{n\to\infty} \int_A d(x, y)^p d\pi_n(x, y) = \lim_{n\to\infty} \int_{E_\Omega} 1_A d(x, y)^p d\pi_n(x, y) \\ &= \int_{E_\Omega} 1_A d(x, y)^p d\pi(x, y) = \int_A d(x, y)^p d\pi(x, y) = \pi'(A),\end{aligned}$$

being 1_A the characteristic function for $A \subset E_\Omega$. Therefore, Portmanteau theorem B.0.11 implies $\pi_n' \xrightarrow{v} \pi'$. Note that E_Ω is open so, applying again theorem B.0.11, the following holds

$$\liminf_{n\to\infty} C_p(\pi_n) = \liminf_{n\to\infty} \pi_n'(E_\Omega) \geq \pi'(E_\Omega) = C_p(\pi).$$

That is, C_p is lower semicontinuous.

For $p = \infty$, consider as before $\{\pi_n\}_n \subset \mathrm{Adm}(\mu, \nu)$ such that $\pi_n \xrightarrow{v} \pi \in \mathrm{Adm}(\mu, \nu)$. Let $r \in \mathbb{R}$ which satisfies

$$r > \liminf_{n\to\infty} C_\infty(\pi_n). \tag{4.11}$$

and consider $U_r = \{(x, y) \in E_\Omega \,:\, d(x, y) > r\}$. Observe that, by definition of $C_\infty(\pi_n)$, $\liminf_{n\to\infty} \pi_n(U_r) = 0$. Since U_r is open, we can apply Portmanteau theorem B.0.11 and obtain

$$\pi(U_r) \leq \liminf_{n\to\infty} \pi_n(U_r) = 0.$$

As a result, the support of π is contained in the complementary of U_r. Therefore, $C_\infty(\pi) \leq r$, and this holds for any $r \in \mathbb{R}$ satisfying (4.11). If C_∞ is not lower semicontinuous, then $C_\infty(\pi) > \liminf_{n\to\infty} C_\infty(\pi_n)$. However, in that case we can find $r \in \mathbb{R}$ such that $C_\infty(\pi) > r > \liminf_{n\to\infty} C_\infty(\pi_n)$, in contradiction with $C_\infty(\pi) \leq r$.

■

After that, the next lemma presents the expected relation between the optimal transport distance of a measure $\mu \in \mathcal{M}^p$ to the zero measure and the total persistence of μ.

Lemma 4.3.8. *For any $\mu \in \mathcal{M}^p$, $1 \leq p \leq \infty$,*

$$\mathrm{OT}_p^p(\mu, 0) = \mathrm{Pers}_p(\mu), \qquad \mathrm{OT}_\infty(\mu, 0) = \mathrm{Pers}_\infty(\mu).$$

Proof. We begin characterizing the admissible plans between μ and 0. Consider $\pi \in \mathrm{Adm}(\mu, 0)$. In particular it satisfies

$$\pi(\Omega \times \bar{\Omega}) = \mu(\Omega), \qquad \pi(\bar{\Omega} \times \Omega) = 0.$$

Notice that $\bar{\Omega} \times \Omega = \Omega \times \Omega \sqcup \partial\Omega \times \Omega$. Therefore, the last equality implies that

$$\pi(\Omega \times \Omega) = 0, \qquad \pi(\partial\Omega \times \Omega) = 0.$$

As mentioned before, we assume that $\pi(\partial\Omega \times \partial\Omega) = 0$. Hence, if we write the total space as the disjoint union

$$\bar{\Omega} \times \bar{\Omega} = \Omega \times \Omega \sqcup \partial\Omega \times \partial\Omega \sqcup \Omega \times \partial\Omega \sqcup \partial\Omega \times \Omega,$$

we can conclude that $\pi(\bar{\Omega} \times \bar{\Omega}) = \pi(\Omega \times \partial\Omega)$, and this is true for any admissible plan. Consequently, fixing $p \in [1, +\infty)$,

$$\mathrm{OT}_p^p(\mu, 0) = \inf_{\pi \in \mathrm{Adm}(\mu,0)} \int_{\bar{\Omega}\times\bar{\Omega}} d(x,y)^p d\pi(x,y) = \inf_{\pi \in \mathrm{Adm}(\mu,0)} \int_{\Omega\times\partial\Omega} d(x,y)^p d\pi(x,y).$$

On the other hand, being admissible implies

$$p_1 \# \pi|_\Omega = \mu, \qquad p_2 \# \pi|_\Omega = 0,$$

so that if $p_{\partial\Omega} : \bar{\Omega} \to \partial\Omega$ is the orthogonal projection onto the diagonal, we have for any $\pi \in \mathrm{Adm}(\mu, 0)$,

$$\begin{aligned}\int_{\Omega\times\partial\Omega} d(x, p_{\partial\Omega}(x))^p d\pi(x,y) &= \int_{\bar{\Omega}\times\bar{\Omega}} d(x, p_{\partial\Omega}(x))^p d\pi(x,y) \\ &= \int_{\bar{\Omega}\times\bar{\Omega}} (d(\cdot, p_{\partial\Omega}(\cdot))^p \circ p_1)(x,y) d\pi(x,y) \\ &= \int_{\bar{\Omega}} d(x, p_{\partial\Omega}(x))^p d(p_1 \# \pi)(x).\end{aligned}$$

Observe that the map $d(\cdot, p_{\partial\Omega}(\cdot))^p$ is zero on the diagonal. Therefore,

$$\int_{\Omega\times\partial\Omega} d(x, p_{\partial\Omega}(x))^p d\pi(x,y) = \int_\Omega d(x, p_{\partial\Omega}(x))^p d(p_1\#\pi)(x) = \int_\Omega d(x, p_{\partial\Omega}(x))^p d\mu.$$

In particular, it does not depend on π. For any $(x,y) \in \Omega \times \partial\Omega$, $d(x, p_{\partial\Omega}(x))^p \leq d(x,y)^p$. This implies that

$$\int_\Omega d(x, p_{\partial\Omega}(x))^p d\mu = \int_{\Omega\times\partial\Omega} d(x, p_{\partial\Omega}(x))^p d\pi(x,y) \leq \int_{\Omega\times\partial\Omega} d(x,y)^p d\pi(x,y)$$

for any admissible transport plan between μ and 0. That is, it is an inferior bound, and as a result

$$\int_\Omega d(x, p_{\partial\Omega}(x))^p d\mu \leq \inf_{\pi\in \mathrm{Adm}(\mu,0)} \int_{\Omega\times\partial\Omega} d(x,y)^p d\pi(x,y) = \mathrm{OT}_p^p(\mu, 0).$$

In order to finish the proof we just need to find and admissible transport plan $\tilde{\pi} \in \mathrm{Adm}(\mu, 0)$ such that

$$\int_{\Omega\times\partial\Omega} d(x,y)^p d\tilde{\pi}(x,y) = \int_\Omega d(x, p_{\partial\Omega}(x))^p d\mu(x) = \int_\Omega d(x,\partial\Omega)^p d\mu(x)$$

To define $\tilde{\pi}$ consider the following. Equip $\bar{\Omega}$ with the measure $\tilde{\mu}$ defined by $\tilde{\mu}(A) = \mu(A\cap\Omega)$ for any Borel set $A \subseteq \bar{\Omega}$. Notice that $\tilde{\mu}(\partial\Omega) = 0$. On the other hand, equip $\bar{\Omega}$ with the measure $\tilde{\nu}$ defined on the Borel sets of $\bar{\Omega}$ as $\tilde{\nu}(A) = \nu(A \cap \partial\Omega)$, where ν is any probability measure on $\partial\Omega$. Notice that in this case $\tilde{\nu}(\Omega) = 0$. Consider the map

$$\begin{aligned} F : \bar{\Omega}\times\bar{\Omega} &\to \bar{\Omega}\times\bar{\Omega} \\ (x,y) &\mapsto (p_1(x,y), p_{\partial\Omega}\circ p_1(x,y)), \end{aligned}$$

and the measure $\tilde{\pi} = F\#(\tilde{\mu}\otimes\tilde{\nu})$ on $\bar{\Omega}\times\bar{\Omega}$, which is given by

$$\tilde{\pi}(A\times B) = \tilde{\mu}\otimes\tilde{\nu}(F^{-1}(A\times B)) = \tilde{\mu}\otimes\tilde{\nu}(p_1^{-1}(A)\cap(p_{\partial\Omega}\circ p_1)^{-1}(B)).$$

It is an admissible transport plan for μ and 0. Indeed, $\tilde{\pi} \in \mathcal{M}(\bar{\Omega}\times\bar{\Omega})$, since $\tilde{\mu}, \tilde{\nu} \in \mathcal{M}(\bar{\Omega})$. Moreover, for any Borel sets $A, B \subseteq \Omega$,

$$\begin{aligned} \tilde{\pi}(A\times\bar{\Omega}) &= \tilde{\mu}\otimes\tilde{\nu}(p_1^{-1}(A)\cap(p_{\partial\Omega}\circ p_1)^{-1}(\bar{\Omega})) = \tilde{\mu}\otimes\tilde{\nu}(A\times\bar{\Omega}) \\ &= \tilde{\mu}(A)\tilde{\nu}(\bar{\Omega}) = \mu(A)\nu(\partial\Omega) = \mu(A), \\ \tilde{\pi}(\bar{\Omega}\times B) &= \tilde{\mu}\otimes\tilde{\nu}(p_1^{-1}(\bar{\Omega})\cap(p_{\partial\Omega}\circ p_1)^{-1}(B)) = \tilde{\mu}\otimes\tilde{\nu}(\emptyset) = 0. \end{aligned}$$

All in all, $\tilde{\pi} \in \mathrm{Adm}(\mu, 0)$. Let us see that it is the transport plan we were looking for. Using the properties of the pushforward and the fact that $\tilde{\mu}|_{\partial\Omega} = 0$ and $\tilde{\nu}|_\Omega = 0$ we obtain

$$\begin{aligned} \int_{\Omega\times\partial\Omega} d(x,y)^p d\tilde{\pi}(x,y) &= \int_{\bar{\Omega}\times\bar{\Omega}} d(x,y)^p d\tilde{\pi}(x,y) = \int_{\bar{\Omega}\times\bar{\Omega}} d(x,y)^p d\left(F\#\tilde{\mu}\otimes\tilde{\nu}\right) \\ &= \int_{\bar{\Omega}\times\bar{\Omega}} \left(d(\cdot,\cdot)^p\circ F\right)(x,y) d(\tilde{\mu}\otimes\tilde{\nu}) = \int_{\bar{\Omega}\times\bar{\Omega}} d(x, p_{\partial\Omega}(x))^p d(\tilde{\mu}\otimes\tilde{\nu}) \\ &= \int_{\bar{\Omega}}\int_{\bar{\Omega}} d(x, p_{\partial\Omega}(x))^p d\tilde{\nu}(y) d\tilde{\mu}(x) = \int_\Omega\int_{\partial\Omega} d(x, p_{\partial\Omega}(x))^p d\tilde{\nu}(y) d\tilde{\mu}(x) \\ &= \int_\Omega d(x, p_{\partial\Omega}(x))^p \left(\int_{\partial\Omega} d\tilde{\nu}(y)\right) d\tilde{\mu}(x) = \int_\Omega d(x, p_{\partial\Omega}(x))^p d\tilde{\mu}(x) \\ &= \int_\Omega d(x, p_{\partial\Omega}(x))^p d\mu(x) = \int_\Omega d(x,\partial\Omega)^p d\mu(x). \end{aligned}$$

The case $p = \infty$ is very similar. Given $\pi \in \mathrm{Adm}(\mu, 0)$ we have seen that $\mathrm{spt}(\pi) \subset \Omega\times\partial\Omega$. Notice that, for any $x \in \mathrm{spt}(\mu)$, exists $y \in \partial\Omega$ such that $(x,y) \in \mathrm{spt}(\pi)$. Therefore,

$$\mathrm{Pers}_\infty(\mu) = \sup_{x\in\mathrm{spt}(\mu)} d(x,\partial\Omega) \leq \sup_{(x,y)\in\mathrm{spt}(\pi)} d(x,\partial\Omega) \leq \sup_{(x,y)\in\mathrm{spt}(\pi)} d(x,y),$$

since $d(x, \partial\Omega) \leq d(x, y)$ for all $(x, y) \in \Omega \times \partial\Omega$. This is true for any admissible transport plan, so we conclude that

$$\mathrm{Pers}_\infty(\mu) \leq \mathrm{OT}_\infty(\mu, 0).$$

Analogously to the previous part, to finish we just need to prove that there exists and admissible transport plan $\tilde{\pi}$ such that $C_\infty(\tilde{\pi}) = \mathrm{Pers}_\infty(\mu)$. Consider $\tilde{\pi} = F\#(\tilde{\mu} \otimes \tilde{\nu})$ as before. Let $(x, y) \in \mathrm{spt}(\tilde{\pi})$ and suppose $y \neq p_{\partial\Omega}(x)$. Then, we can find neighborhoods U and V in $\bar{\Omega}$, of x and y respectively, such that $p_{\partial\Omega}(U) \cap V = \emptyset$. Therefore, $p_1^{-1}(U) \cap (p_{\partial\Omega} \circ p_1)^{-1}(V) = \emptyset$. But this provides the contradiction

$$0 = \tilde{\mu} \otimes \tilde{\nu}(p_1^{-1}(U) \cap (p_{\partial\Omega} \circ p_1)^{-1}(V)) = \tilde{\pi}(U \times V) > 0,$$

where the right hand side is positive because $(x, y) \in \mathrm{spt}(\tilde{\pi})$. All in all, $(x, y) \in \mathrm{spt}(\tilde{\pi})$ implies $y = p_{\partial\Omega}(x)$. Since the projection of a boundary point is itself, we can write

$$\sup_{(x,y)\in\mathrm{spt}(\tilde{\pi})} d(x, y) = \sup_{(x,y)\in\mathrm{spt}(\tilde{\pi}) \,:\, x\in\Omega} d(x, \partial\Omega) = \sup_{x\in\mathrm{spt}(\mu)} d(x, \partial\Omega),$$

where the last equality follows from $p_1\#\tilde{\pi}|_\Omega = \mu$. ■

As we expected, OT_p defines a true metric on $\mathcal{M}^p$.

Proposition 4.3.9 ([6]). *Let $\mu, \nu \in \mathcal{M}^p$ with $1 \leq p \leq \infty$. Then, OT_p is a distance on $\mathcal{M}^p$.*

Proof. Let $\mu, \nu, \tau \in \mathcal{M}^p$, we divide the proof into two parts according to the value of p. Firstly, consider $1 \leq p < \infty$. For the symmetry property, note that given $\pi \in \mathrm{Adm}(\mu, \nu)$ the homeomorphism $\phi : (x, y) \mapsto (y, x)$ provides an admisible plan in $\mathrm{Adm}(\nu, \mu)$, and viceversa. Therefore, it is clear that $\mathrm{OT}_p(\mu, \nu) = \mathrm{OT}_p(\nu, \mu)$.

Assume $\mathrm{OT}_p(\mu, \nu) = 0$. Then, there exists $\pi \in \mathrm{Adm}(\mu, \nu)$ such that

$$\int_{E_\Omega} d(x, y)^p d\pi(x, y) = 0.$$

Hence, $d(x, y)^p = 0$ π-a.e. This, together with the fact that $d(x, y)^p = 0$ if and only if $x = y$, implies that π is supported on $\{(x, x) \,:\, x \in \Omega\}$. Therefore, for any Borel set $A \subset \Omega$,

$$\mu(A) = \pi(A \times \bar{\Omega}) = \pi(A \times A) = \pi(\bar{\Omega} \times A) = \nu(A).$$

That is, $\mu = \nu$. On the other hand, suppose $\mu = \nu$. Consider the measure $\tilde{\mu}$ on $\bar{\Omega}$ defined by $\tilde{\mu}(A) = \mu(A \cap \Omega)$. Let $F : \bar{\Omega} \to \bar{\Omega} \times \bar{\Omega}$; $F(x) = (x, x)$. Then, it is easy to see that $\pi = F\#\tilde{\mu}$ defines an admissible transport plan between μ and itself such that $\int_{\bar{\Omega}\times\bar{\Omega}} d(x, y)^p d\pi(x, y) = 0$. Hence, $\mathrm{OT}_p(\mu, \mu) = 0$.

To prove the triangle inequality we need an adaptation of the glueing lemma 2.1 of [5]. In our case it states that, for any $\pi_{12} \in \mathrm{Opt}_p(\mu, \nu)$, $\pi_{23} \in \mathrm{Opt}_p(\nu, \tau)$ there exists $\gamma \in \mathcal{M}(\bar{\Omega} \times \bar{\Omega} \times \bar{\Omega})$ such that

$$p_{12}\#\gamma|_{E_\Omega} = \pi_{12}, \quad p_{23}\#\gamma|_{E_\Omega} = \pi_{23}, \tag{4.12}$$

where p_{ij} are the canonical projections, and the restrictions are concentrated on $\{(x,x) \,:\, x \in \partial\Omega\}$ on $\partial\Omega \times \partial\Omega$. Define on $\bar{\Omega} \times \bar{\Omega}$ the measure

$$\eta(E) = \int_{(E,x,z)} \int_{(\bar{\Omega},y)} d\gamma(x,y,z).$$

for any Borel $E \subset \bar{\Omega} \times \bar{\Omega}$. Then, for any Borel subsets $A, B \subset \Omega$,

$$\eta(A \times \bar{\Omega}) = \int_A \int_{\bar{\Omega}} \int_{\bar{\Omega}} d\gamma(x,y,z) = \gamma(A \times \bar{\Omega} \times \bar{\Omega}) = \pi_{12}(A \times \bar{\Omega}) = \mu(A),$$
$$\eta(\bar{\Omega} \times B) = \int_{\bar{\Omega}} \int_{\bar{\Omega}} \int_B d\gamma(x,y,z) = \gamma(\bar{\Omega} \times \bar{\Omega} \times B) = \pi_{23}(\bar{\Omega} \times B) = \tau(B).$$

Therefore, $\eta \in \mathrm{Adm}(\mu,\tau)$. By the triangle and Minkowski inequality we obtain

$$\begin{aligned}
\mathrm{OT}_p(\mu,\tau) &\leq \left(\int_{\bar{\Omega}\times\bar{\Omega}} d(x,z)^p d\eta(x,z) \right)^{1/p} = \left(\int_{\bar{\Omega}\times\bar{\Omega}\times\bar{\Omega}} d(x,z)^p d\gamma(x,y,z) \right)^{1/p} \\
&\leq \left(\int_{\bar{\Omega}^3} d(x,y)^p d\gamma(x,y,z) \right)^{1/p} + \left(\int_{\bar{\Omega}^3} d(y,z)^p d\gamma(x,y,z) \right)^{1/p} \\
&= \left(\int_{\bar{\Omega}^2} d(x,y)^p d(p_{12}\#\gamma)(x,y) \right)^{1/p} + \left(\int_{\bar{\Omega}^2} d(y,z)^p d(p_{23}\#\gamma)(y,z) \right)^{1/p}.
\end{aligned}$$

Since on $\partial\Omega \times \partial\Omega$ the measures $p_{12}\#\gamma$ and $p_{23}\#\gamma$ are concentrated on $\{(x,x) \,:\, x \in \partial\Omega\}$ (where the distance function is zero) and they satify (4.12), the last two integrals can be written as

$$\begin{aligned}
\mathrm{OT}_p(\mu,\tau) &\leq \left(\int_{E_\Omega} d(x,y)^p d(p_{12}\#\gamma)(x,y) \right)^{1/p} + \left(\int_{E_\Omega} d(y,z)^p d(p_{23}\#\gamma)(y,z) \right)^{1/p} \\
&= \left(\int_{E_\Omega} d(x,y)^p d\pi_{12}(x,y) \right)^{1/p} + \left(\int_{E_\Omega} d(y,z)^p d\pi_{23}(y,z) \right)^{1/p} \\
&= \mathrm{OT}_p(\mu,\nu) + \mathrm{OT}_p(\nu,\tau).
\end{aligned}$$

Secondly, let $p = \infty$. Given $\pi \in \mathrm{Adm}(\mu,\nu)$, the same homeomorphism as before provides and element of $\mathrm{Adm}(\nu,\mu)$ with the same support as π, and viceversa. It is clear then that $\mathrm{OT}_\infty(\mu,\nu) = \mathrm{OT}_\infty(\nu,\mu)$.

If $\mathrm{OT}_\infty(\mu,\nu) = 0$, there exists $\pi \in \mathrm{Adm}(\mu,\nu)$ such that $\sup_{(x,y)\in\mathrm{spt}(\pi)} d(x,y) = 0$. This implies that $\mathrm{spt}(\pi) \subseteq \{(x,x) \,:\, x \in \Omega\}$. We conclude as before $\mu = \nu$. Suppose $\mu = \nu$ and define $\pi = F\#\mu$ as previously. This measure is concentrated on $E_\Omega \cap \{(x,x) \,:\, x \in \Omega\}$. Hence, $C_\infty(\pi) = 0$, which implies $\mathrm{OT}_\infty(\mu,\mu) = 0$.

For the triangle inequality we proceed analogously. Given $\pi_{12} \in \mathrm{Opt}_p(\mu,\nu)$ and $\pi_{23} \in \mathrm{Opt}_p(\nu,\tau)$ we use again the glueing lemma to obtain γ and we define the same measure η. Notice that, for any $(x,z) \in \mathrm{spt}(\eta)$, there exists at least one $y \in \bar{\Omega}$ such that $(x,y,z) \in \mathrm{spt}(\gamma)$. Therefore,

$$\begin{aligned}
\mathrm{OT}_\infty(\mu,\tau) \leq C_\infty(\eta) &= \sup_{(x,z)\in\mathrm{spt}(\eta)} d(x,z) \leq \sup_{(x,z)\in\mathrm{spt}(\eta):(x,y,z)\in\mathrm{spt}(\gamma)} d(x,y) + d(y,z) \\
&\leq \sup_{(x,y)\in\mathrm{spt}(p_{12}\#\gamma)} d(x,y) + \sup_{(y,z)\in\mathrm{spt}(p_{23}\#\gamma)} d(y,z).
\end{aligned}$$

Again, the fact that the marginals of γ are concentrated on the diagonal on $\partial\Omega\times\partial\Omega$ (where d is zero) and equalities (4.12) lead to

$$\begin{aligned}\mathrm{OT}_\infty(\mu,\tau) &\leq \sup_{(x,y)\in\mathrm{spt}(\pi_{12})} d(x,y) + \sup_{(y,z)\in\mathrm{spt}(\pi_{23})} d(y,z)\\ &= C_\infty(\pi_{12}) + C_\infty(\pi_{23}) = \mathrm{OT}_\infty(\mu,\nu) + \mathrm{OT}_\infty(\nu,\tau).\end{aligned}$$

To finish the proof notice that finite total persistence and triangle ineguality implies $\mathrm{OT}_p : \mathcal{M}^p \times \mathcal{M}^p \to [0,\infty)$. ■

Hence, we equip $\mathcal{M}^p$ with the optimal transport distance OT_p. In the particular case of persistence diagrams, the following is satisfied.

Proposition 4.3.10 ([6]). *For $\mu,\nu\in\mathcal{D}^p\subset\mathcal{M}^p$, with $1\leq p\leq\infty$,*

$$\mathrm{OT}_p(\mu,\nu) = d_p(\mu,\nu).$$

That is, the optimal transport distance is an extension of the distance d_p.

Proof. Let us sketch the proof when $p=\infty$. For full arguments we refer to proposition 3.5 and 3.23 of [6].

Let $a,b\in\mathcal{D}^\infty$ such that $a=\sum_{i\in I}\delta_{x_i}$ and $b=\sum_{j\in J}\delta_{y_j}$, where I,J are the sets of indices, possibly infinite. Consider $\pi\in\mathrm{Adm}(a,b)$. It is easy to see that the marginal constraints imply $\mathrm{spt}(\pi)\subset\{x_i\}_{i\in I}\cup\partial\Omega\times\{y_j\}_{j\in J}\cup\partial\Omega$. Moreover, we can always assume that the mass $\pi(\{x_i\}\times\partial\Omega)$ (resp. $\pi(\partial\Omega\times\{y_j\})$ is sent on the projection of x_i (resp. y_j). Therefore, an element of $\mathrm{Opt}(a,b)$ can be written as a bystochastic matrix indexed on $(-J\cup I)\times(-I\cup J)$. Denoting by S the set of all these matrices and defining

$$\begin{aligned}&C_{ij}=d(x_i,y_j) \qquad i,j>0, \qquad C_{ij}=d(x_i,p_{\partial\Omega}(x_i)) \qquad i>0,j<0,\\ &C_{ij}=d(p_{\partial\Omega}(y_j),y_j) \qquad i<0,j>0, \qquad C_{ij}=0 \qquad i,j<0,\end{aligned}$$

it is immediate that $\mathrm{OT}_\infty(a,b)=\inf_{Q\in S}\sup\{C_{ij}\ :\ (i,j)\in\mathrm{spt}(Q)\}$. Notice that, for any $Q\in S$, $k\in\mathbb{N}$ and different indices $\{i_1,\dots,i_k\}\subset -J\cup I$, we have

$$k=\sum_{k'=1}^{k}\sum_{j\in-I\cup J}Q_{i_{k'}j}=\sum_{j\in-I\cup J}\sum_{k'=1}^{k}Q_{i_{k'}j}\Rightarrow\mathrm{Card}\{j\ :\ \exists k'\ (i_{k'},j)\in\mathrm{spt}(Q)\}\geq k.$$

By Hall's marriage theorem exists a permutation matrix P with $\mathrm{spt}(P)\subset\mathrm{spt}(Q)$. Consequently, if S' is the set of all permutation matrices indexed on $(-J\cup I)\times(-I\cup J)$,

$$\begin{aligned}\sup\{C_{ij}\ :\ (i,j)\in\mathrm{spt}(Q)\}&\geq\inf_{P\in S'}\sup\{C_{ij}\ :\ (i,j)\in\mathrm{spt}(P)\}=d_\infty(a,b)\\ &\Rightarrow\mathrm{OT}_\infty(a,b)\geq d_\infty(a,b).\end{aligned}$$

Finally, since $S'\subset S$ we conclude that $\mathrm{OT}_\infty(a,b)=d_\infty(a,b)$. ■

Thanks to the equality $\mathrm{OT}_p = d_p$ on $\mathcal{D}^p$, all stability results for persistence diagrams developed in section 4.1 are true when using optimal transport distance instead. Those statements can be seen as an approach to characterize convergence between persistence diagrams. In the case of the space of persistence measures that characterization is clearer, and relates vague and weak convergence to optimal transport convergence.

Theorem 4.3.11. *Let $\{\mu_n\}_{n\in\mathbb{N}}$ be a sequence of measures in $\mathcal{M}^p$ and $\mu \in \mathcal{M}^p$, where $1 \leq p < \infty$. Then,*

$$\lim_{n\to\infty} \mathrm{OT}_p(\mu_n, \mu) = 0$$

if and only if

$$\mu_n \xrightarrow{v} \mu \quad \textit{and} \quad \lim_{n\to\infty} \mathrm{Pers}_p(\mu_n) = \mathrm{Pers}_p(\mu).$$

Proof. Let us prove the direct implication. The proof of the converse implication can be found in theorem 3.7 of [6].

Assume that $\lim_{n\to\infty} \mathrm{OT}_p(\mu_n, \mu) = 0$. Since OT_p is a distance in $\mathcal{M}^p$ we can apply the triangle inequality and obtain

$$\mathrm{OT}_p(\mu_n, 0) \leq \mathrm{OT}_p(\mu_n, \mu) + \mathrm{OT}_p(\mu, 0) \Rightarrow \mathrm{OT}_p(\mu_n, 0) - \mathrm{OT}_p(\mu, 0) \leq \mathrm{OT}_p(\mu_n, \mu).$$

Likewise,

$$-\mathrm{OT}_p(\mu_n, \mu) \leq \mathrm{OT}_p(\mu_n, 0) - \mathrm{OT}_p(\mu, 0).$$

Therefore,

$$0 \leq |\mathrm{OT}_p(\mu_n, 0) - \mathrm{OT}_p(\mu, 0)| \leq \mathrm{OT}_p(\mu_n, \mu) \to 0,$$

which implies

$$\lim_{n\to\infty} \mathrm{OT}_p(\mu_n, 0) = \mathrm{OT}_p(\mu, 0).$$

By lemma 4.3.8 we conclude that

$$\lim_{n\to\infty} \mathrm{Pers}_p(\mu_n) = \mathrm{Pers}_p(\mu).$$

To prove the vague convergence of μ_n to μ it is enough to show that, for any $f \in C_c(\Omega)$, $\lim_{n\to\infty} \varphi_f(\mu_n) = \varphi_f(\mu)$. Let then $f \in C_c(\Omega)$, with support contained in some compact set K of Ω. For this compact set K we have that $0 < d(K, \partial\Omega) < \infty$. Hence, by remark 4.3.3, $\mu_n(K) < \infty$ for every $n \in \mathbb{N}$.

Thanks to Stone–Weierstrass theorem, for any $\varepsilon > 0$ there exists a Lipschitz function f_ε, with constant L_ε, whose support is also contained in K and such that $||f - f_\varepsilon||_\infty \leq \varepsilon$. Fixing $\varepsilon > 0$,

$$\begin{aligned}
|\varphi_f(\mu_n) - \varphi_f(\mu)| &= \left| \int_\Omega f d\mu_n - \int_\Omega f d\mu \right| \\
&= \left| \int_\Omega f d\mu_n - \int_\Omega f_\varepsilon d\mu_n + \int_\Omega f_\varepsilon d\mu_n - \int_\Omega f d\mu + \int_\Omega f_\varepsilon d\mu - \int_\Omega f_\varepsilon d\mu \right| \\
&\leq \left| \int_\Omega (f - f_\varepsilon) d\mu_n \right| + \left| \int_\Omega (f_\varepsilon - f) d\mu \right| + \left| \int_\Omega f_\varepsilon d\mu_n - \int_\Omega f_\varepsilon d\mu \right|.
\end{aligned}$$

Because both f and f_ε have support contained in K,

$$\begin{aligned}\left|\int_\Omega (f-f_\varepsilon)d\mu_n\right| &= \left|\int_K (f-f_\varepsilon)d\mu_n\right| \le \int_K |f-f_\varepsilon|\,d\mu_n \\ &\le \mu_n(K)\sup_{x\in K}|f(x)-f_\varepsilon(x)| \le \mu_n(K)\varepsilon.\end{aligned}$$

Proceeding analogously for the term $(f-f_\varepsilon)d\mu$ we get

$$|\varphi_f(\mu_n)-\varphi_f(\mu)| \le \mu_n(K)\varepsilon + \mu(K)\varepsilon + \left|\int_\Omega f_\varepsilon d\mu_n - \int_\Omega f_\varepsilon d\mu\right|$$

Suppose $\sup_k \mu_k(K)$ is not finite. Denote $1/\delta = d(K,\partial\Omega)$ so that $\mu_n(K) \le \delta \mathrm{Pers}_p(\mu_n)$ by (4.9). Then, for every $M>0$ exists $n\in\mathbb{N}$ such that $\mu_n(K)>M$. That is, we can find a subsequence $\{\mu_{n_k}(K)\}_k$ which tends to infinity. But then the subsequence $\{\mathrm{Pers}_p(\mu_{n_k})\}_k$ tends to infinity too. However, we have already proven that the sequence $\mathrm{Pers}_p(\mu_n)$ has finite limit, and the same should satisfy any of its subsequences. Therefore, we get a contradiction and conclude that $\sup_k \mu_k(K)$ is finite. Hence,

$$\begin{aligned}|\varphi_f(\mu_n)-\varphi_f(\mu)| &\le \sup_k \mu_k(K)\varepsilon + \mu(K)\varepsilon + \left|\int_\Omega f_\varepsilon d\mu_n - \int_\Omega f_\varepsilon d\mu\right| \\ &= (\sup_k \mu_k(K) + \mu(K))\varepsilon + A_n.\end{aligned} \tag{4.13}$$

Let us study the last term A_n more carefully. For each $n\in\mathbb{N}$, consider $\pi_n \in \mathrm{Opt}_p(\mu_n,\mu)$. In particular, π_n are admissible plans, hence

$$p_1\#\pi_n|_\Omega = \mu_n, \qquad p_2\#\pi_n|_\Omega = \mu \qquad \forall n\in\mathbb{N},$$

where $p_i : \bar\Omega\times\bar\Omega \to \bar\Omega$ are the canonical projections. By properties of pushforwards and knowing that $f_\varepsilon|_{\partial\Omega}=0$, we obtain

$$\begin{aligned}\int_\Omega f_\varepsilon(x)d\mu_n(x) &= \int_\Omega f_\varepsilon(x)d(p_1\#\pi_n)(x) = \int_{\bar\Omega} f_\varepsilon(x)d(p_1\#\pi_n)(x) \\ &= \int_{\bar\Omega\times\bar\Omega}(f_\varepsilon\circ p_i)(x,y)d\pi_n(x,y) = \int_{\bar\Omega\times\bar\Omega} f_\varepsilon(x)d\pi_n(x,y).\end{aligned}$$

Similarly,

$$\int_\Omega f_\varepsilon(y)d\mu(y) = \int_{\bar\Omega\times\bar\Omega} f_\varepsilon(y)d\pi_n(x,y).$$

As a result, since f_ε has support in K and Lipschitz constant L_ε, the last term is bounded by

$$\begin{aligned}A_n &= \left|\int_\Omega f_\varepsilon d\mu_n - \int_\Omega f_\varepsilon d\mu\right| = \left|\int_{\bar\Omega\times\bar\Omega} f_\varepsilon(x)d\pi_n(x,y) - \int_{\bar\Omega\times\bar\Omega} f_\varepsilon(y)d\pi_n(x,y)\right| \\ &= \left|\int_{\bar\Omega\times\bar\Omega}(f_\varepsilon(x)-f_\varepsilon(y))d\pi_n(x,y)\right| \le \int_{\bar\Omega\times\bar\Omega}|f_\varepsilon(x)-f_\varepsilon(y)|\,d\pi_n(x,y) \\ &= \int_{(K\times\bar\Omega)\cup(\bar\Omega\times K)}|f_\varepsilon(x)-f_\varepsilon(y)|\,d\pi_n(x,y) \le L_\varepsilon\int_{(K\times\bar\Omega)\cup(\bar\Omega\times K)} d(x,y)d\pi_n(x,y).\end{aligned}$$

Let us now apply Hölder's inequality, which is given by

$$||fg||_1 \leq ||f||_p ||g||_q, \tag{4.14}$$

where $\frac{1}{p} + \frac{1}{q} = 1$ and

$$||f||_p = \left(\int |f|^p \right)^{1/p},$$

for $f, g \in \mathcal{L}^p$[2]. In our case $f = id$ and $g = d(x, y)$. The inequality for A_n becomes

$$\begin{aligned} A_n &\leq L_\varepsilon \left(\int_{(K\times\bar{\Omega})\cup(\bar{\Omega}\times K)} |id(x,y)|^{p/(p-1)} d\pi_n \right)^{1-\frac{1}{p}} \left(\int_{(K\times\bar{\Omega})\cup(\bar{\Omega}\times K)} |d(x,y)|^p d\pi_n \right)^{\frac{1}{p}} \\ &= L_\varepsilon \left[\pi_n((K\times\bar{\Omega})\cup(\bar{\Omega}\times K)) \right]^{1-\frac{1}{p}} \left(\int_{(K\times\bar{\Omega})\cup(\bar{\Omega}\times K)} d(x,y)^p d\pi_n \right)^{\frac{1}{p}}. \end{aligned}$$

On the one hand,

$$\pi_n((K\times\bar{\Omega})\cup(\bar{\Omega}\times K)) \leq \pi_n(K\times\bar{\Omega}) + \pi_n(\bar{\Omega}\times K) = \mu_n(K) + \mu(K) \leq \sup_k \mu_k(K) + \mu(K).$$

On the other hand, due to the optimality of the transport plans π_n,

$$\left(\int_{(K\times\bar{\Omega})\cup(\bar{\Omega}\times K)} d(x,y)^p d\pi_n \right)^{\frac{1}{p}} \leq \left(\int_{\bar{\Omega}\times\bar{\Omega}} d(x,y)^p d\pi_n \right)^{\frac{1}{p}} = \mathrm{OT}_p(\mu_n, \mu).$$

All in all,

$$A_n \leq L_\varepsilon [\sup_k \mu_k(K) + \mu(K)]^{1-\frac{1}{p}} \mathrm{OT}_p(\mu_n, \mu).$$

Notice that $L_\varepsilon[\sup_k \mu_k(K) + \mu(K)]^{1-\frac{1}{p}}$ is finite while $\mathrm{OT}_p(\mu_n, \mu)$ tends to zero as n tends to infinity. As a result, A_n tends to zero as $n \to \infty$, for fixed ε. Recovering inequality (4.13), for fixed $\varepsilon > 0$ we have

$$\begin{aligned} \limsup_{n\to\infty} |\varphi_f(\mu_n) - \varphi_f(\mu)| &\leq \limsup_{n\to\infty} \left[(\sup_k \mu_k(K) + \mu(K))\varepsilon + A_n \right] \\ &= (\sup_k \mu_k(K) + \mu(K))\varepsilon + \lim_{n\to\infty} A_n = (\sup_k \mu_k(K) + \mu(K))\varepsilon. \end{aligned}$$

Taking the limit $\varepsilon \to 0$, and recalling that $\sup_k \mu_k(K) < \infty$, we conclude that

$$\limsup_{n\to\infty} |\varphi_f(\mu_n) - \varphi_f(\mu)| = \lim_{\varepsilon\to 0} \limsup_{n\to\infty} |\varphi_f(\mu_n) - \varphi_f(\mu)| \leq \lim_{\varepsilon\to 0} (\sup_k \mu_k(K) + \mu(K))\varepsilon = 0.$$

Consequently,

$$\limsup_{n\to\infty} |\varphi_f(\mu_n) - \varphi_f(\mu)| = 0 \Rightarrow \lim_{n\to\infty} |\varphi_f(\mu_n) - \varphi_f(\mu)| = 0.$$

This implies $\varphi_f(\mu_n) \to \varphi_f(\mu)$ as wanted. ■

[2]Recall that $L^p(\mathcal{S}) = \mathcal{L}^p(\mathcal{S})/\sim$, where $f \sim g$ if and only if $||f - g||_p = 0$ and $\mathcal{L}^p = \{f : \mathcal{S} \to \mathbb{R} : ||f||_p < \infty\}$ with $1 \leq p \leq \infty$

Due to the previous theorem we can deduce that the topology of the optimal transport metric is stronger than the vague topology. In the case $p = \infty$ only one of the implications holds, whose proof is similar to the one above and can be found in proposition 3.26 of [6].

Theorem 4.3.12. *Let $\mu, \mu_1, \mu_2, \ldots \in \mathcal{M}^\infty$. If*

$$\lim_{n\to\infty} \mathrm{OT}_\infty(\mu_n, \mu) = 0,$$

then $\mu_n \xrightarrow{v} \mu$.

Regarding weak convergence, we have the following.

Corollary 4.3.13 ([6])**.** *Consider the measures $\mu, \mu_1, \mu_2, \ldots \in \mathcal{M}^p$, with $p < \infty$. Then,*

$$\lim_{n\to\infty} \mathrm{OT}_p(\mu_n, \mu) = 0,$$

if and only if $\mu_n^{[p]} \xrightarrow{w} \mu^{[p]}$, where these measures are defined for every Borel subset $A \subset \Omega$ as

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

Proof. Firstly, assume that $\mathrm{OT}_p(\mu_n, \mu)$ converges to zero. By theorem 4.3.11 this is equivalent to

$$\mu_n \xrightarrow{v} \mu \quad \text{and} \quad \mu_n^{[p]}(\Omega) = \mathrm{Pers}_p(\mu_n) \to \mathrm{Pers}_p(\mu) = \mu^{[p]}(\Omega).$$

By proposition B.0.6, it is enough to prove $\mu_n^{[p]} \xrightarrow{v} \mu^{[p]}$. Observe that $\mu_n^{[p]} \ll \mu_n$ and $\mu^{[p]} \ll \mu$. Besides, given any compactly supported continuous function $f \in C_c(\Omega)$, the map $x \mapsto d(x, \partial\Omega)^p f(x)$ is also continuous and compactly supported. Therefore, using vague convergence of μ_n to μ and Radon-Nikodym theorem B.0.13,

$$\int_\Omega f(x) d\mu_n^{[p]} = \int_\Omega f(x) d(x, \partial\Omega)^p d\mu_n \longrightarrow \int_\Omega f(x) d(x, \partial\Omega)^p d\mu = \int_\Omega f(x) d\mu^{[p]},$$

which, by (B.2), is equivalent to vague convergence.

Conversely, assume $\mu_n^{[p]} \xrightarrow{w} \mu^{[p]}$, which by proposition B.0.6 is equivalent to

$$\mu_n^{[p]} \xrightarrow{v} \mu^{[p]} \quad \text{and} \quad \mu_n^{[p]}(\Omega) = \mathrm{Pers}_p(\mu_n) \to \mathrm{Pers}_p(\mu) = \mu^{[p]}(\Omega).$$

Thanks to theorem 4.3.11, it suffices to prove that $\mu_n \xrightarrow{v} \mu$. Likewise, given $f \in C_c(\Omega)$, the map $x \mapsto \frac{f(x)}{d(x,\partial\Omega)^p}$ is continuous and compactly supported. Hence, $\mu_n^{[p]} \xrightarrow{v} \mu^{[p]}$ implies

$$\int_\Omega f(x) d\mu_n = \int_\Omega \frac{f(x)}{d(x, \partial\Omega)^p} d\mu_n^{[p]} \longrightarrow \int_\Omega \frac{f(x)}{d(x, \partial\Omega)^p} d\mu^{[p]} = \int_\Omega f(x) d\mu.$$

By (B.2) this is equivalent to vague convergence. ∎

These results on convergence are pretty useful in order to derive characterizations on the continuity of some vectorizations. As we will see in the next chapter, vectorizations are used to embed non linear spaces (as the space of persistence diagrams) into linear ones. This is necessary if we want to use statistical descriptors and machine learning techniques, since many of them requiere the input to live in a finite-dimensional vector space.

Lastly, we present some properties of the metric space of persistence measures $(\mathcal{M}^p, \mathrm{OT}_p)$. Those are coherent with the ones obtained for the space of persistence diagrams $(\mathcal{D}^p, d_p)$.

Proposition 4.3.14 ([6])**.** *The space $(\mathcal{M}^p, \mathrm{OT}_p)$ is a Polish space for $1 \leq p < \infty$.*

Remark 4.3.15. As we saw in remark 4.2.8, $\mathcal{D}^\infty$ is not separable. Analogously, the space $(\mathcal{M}^\infty, \mathrm{OT}_\infty)$ is not separable, although it can be proven to be complete (proposition 3.24 of [6]). ♢

We recover the following result, which can be seen as a consequence of the optimal transport topology being stronger than the vague topology.

Corollary 4.3.16 ([6])**.** *For $1 \leq p \leq \infty$ the subspace $\mathcal{D}^p$ is closed in $\mathcal{M}^p$ for the metric OT_p. In particular, this implies that:*

- *If $p < \infty$, $(\mathcal{D}^p, \mathrm{OT}_p) = (\mathcal{D}^p, d_p)$ is a Polish space.*
- *If $p = \infty$, $(\mathcal{D}^\infty, \mathrm{OT}_\infty) = (\mathcal{D}^\infty, d_\infty)$ is complete.*

Proof. Let $1 \leq p \leq \infty$ and consider any sequence $\{\mu_n\}_n \subset \mathcal{D}^p$ such that it converges in $\mathcal{M}^p$ to a measure μ for the OT_p distance. That is, $\lim_{n\to\infty} \mathrm{OT}_p(\mu_n, \mu) = 0$. To prove that $\mathcal{D}^p$ is closed in $\mathcal{M}^p$ we need to show that $\mu \in \mathcal{D}^p$. By theorems 4.3.11 and 4.3.12 we have convergence in the vague topology $\mu_n \xrightarrow{v} \mu$. Notice that $\mu_n \in \mathcal{D}$ for every $n \in \mathbb{N}$. Therefore, proposition B.0.7 implies that $\mu \in \mathcal{D}$. As it also belongs to $\mathcal{M}^p$, we conclude that $\mu \in \mathcal{D} \cap \mathcal{M}^p = \mathcal{D}^p$ as wanted. Hence, $\mathcal{D}^p$ is closed in $\mathcal{M}^p$. Since $\mathrm{OT}_p = d_p$ and taking into consideration proposition 4.3.14 and remark 4.3.15, this proves that $(\mathcal{D}^p, d_p)$ is complete for every value of p (a closed subspace of a complete space is complete). Moreover, we have seen in proposition 4.2.6 that for $p < \infty$ the space $\mathcal{D}^p$ is separable. Therefore, $(\mathcal{D}^p, d_p)$ is Polish for $p < \infty$. ■

Chapter 5

Vectorizations of persistence diagrams

For some scientific data sets, particular geometric structures or patterns can be related to topological features. For example, in the research of structures in chemistry and biology or in the study of convection in fluid dynamics. In those cases, their persistence diagrams are on their own enough to interpret and understand important properties of the phenomena from which the data has been generated.

On the other hand, in general it is necessary further processing in order to interpret the topological information encoded in persistence diagrams. Nevertheless, the space of persistence diagrams is not a linear space. That is, there is no simple way to define the sum of two diagrams, or the multiplication of a diagram by a real number, such that it is compatible with the bottleneck distance. This represents a major drawback and could endanger the use of persistence diagrams in practical applications: many statistical descriptors and machine learning techniques need the input to be elements of a finite-dimensional vector space.

An almost immediate way to solve this problem is to transform persistence diagrams into vectors, with which we can perform linear tasks. This can be done by embeding the space of persistence diagrams into a linear space using vectorizations.

Definition 5.0.1. A **vectorization** is a map $\Phi : \mathcal{D} \to \mathcal{B}$ from the space of persistence diagrams to some Banach space $\mathcal{B}$.

This chapter is dedicated to introduce the principal vectorizations used in topological data analysis. The main references are [26] and [6].

In general, what interest us is that the properties encoded in persistence diagrams are transferred in some way to the vectorizations, and can be used to study the data set. Moreover, stability with respect to input noise and efficiency in computation are also important. However, this is not always possible. To begin with, the choice of the vectorization is arbitrary. Secondly, for the moment it is not known whether, in general, close diagrams have close representations and vice-versa. Lastly, the interpretability we had with persistence diagrams is lost when using vectorizations. For example, once the vectorization is applied we can define a mean m. Nervertheless, a priori m has nothing to do with a notion of mean in $\mathcal{D}$, it may not even be the image of a persistence diagram.

5.1 Persistence landscapes

Since the aim of vectorizations is to be able to use numerical techniques, we restrict ourselves to the practical case of finite persistence diagrams. Hence, let $\mu = \sum_{i=1}^{n} \delta_{x_i} \in \mathcal{D}$. The **persistence landscape** is defined as follows (see figure 5.1). Firstly, consider the change of coordinates

$$h : \mathbb{R}^2 \to \mathbb{R}^2; \qquad h(r_1, r_2) = \left(\frac{r_1 + r_2}{2}, \frac{r_2 - r_1}{2} \right). \tag{5.1}$$

The outcome of applying h is the rotated and rescaled persistence diagram $h(X)$ located in $\{(r_1, r_2) \in \mathbb{R}^2 \ : \ r_2 \geq 0\}$. The diagonal $\partial\Omega$ is sent by h to the r_1-axis. After that, for each x_i we tent the point $h(x_i)$ by applying the triangle function, which is given by

$$\Lambda_r : \mathbb{R} \to \mathbb{R}; \qquad \Lambda_r(t) = \max\{0, r_2 - |t - r_1|\}, \tag{5.2}$$

obtaining a family of continuous functions $\{\Lambda_{h(x_i)}\}_{i=1}^{n}$. Finally, the persistence landscape is defined as

$$\lambda_\mu : \mathbb{N} \times \mathbb{R} \to \mathbb{R}; \qquad \lambda_\mu(k, t) = k\text{max}_{i=1,\dots,n} \Lambda_{h(x_i)}(t),$$

where $k$max is the k-th largest value in the set. That is, λ_μ is a summary of the arrangement of the piecewise linear curves that are obtained when overlaying the graphs of the functions $\{\Lambda_{h(x_i)}\}_{i=1}^{n}$. For each $k \in \mathbb{N}$ the function $\lambda_\mu(k, \cdot) : \mathbb{R} \to \mathbb{R}$ is called **k-th landscape** of μ.

Remark 5.1.1. For any $t \in \mathbb{R}$ and $k \in \mathbb{N}$, $\lambda_\mu(k, t) \geq 0$ and $\lambda_\mu(k, t) \geq \lambda_\mu(k+1, t)$. Moreover, from the fact that Λ_P is 1-Lipschitz and the definition of landscape, it is easy to deduce that k-th landscapes are 1-Lipschitz too. ♢

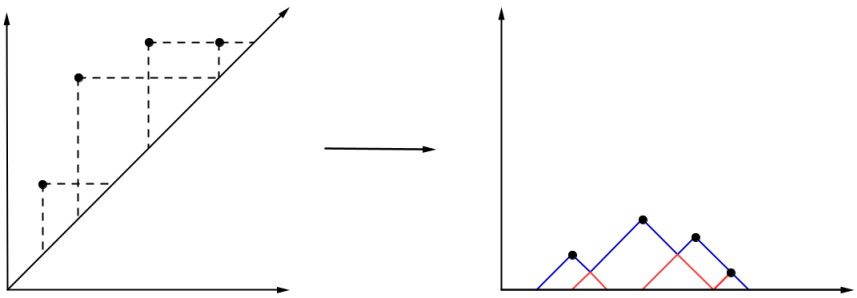


Figure 5.1: Construction of the persistence landscape (right) from a finite persistence diagram (left). The first landscape is in blue and the second one is in red. All the other landscapes are zero.

Notice that there exists $k_0 \in \mathbb{N}$ such that, for any $k \leq k_0$, the k-th landscape is zero. Hence, persistence landscapes are elements of the Banach space $L^p(\mathbb{N} \times \mathbb{R})$, with the counting measure on $\mathbb{N}$ and the Lebesgue measure on $\mathbb{R}$. We can then define the vectorization $\mu \mapsto \lambda_\mu$, which in fact is injective. As a consequence, the

representation of persistence diagrams through landscapes does not imply loss of information. Another advantage of persistence landscapes is that they are easily invertible. Moreover, they share similar stability properties with persistence diagrams.

Proposition 5.1.2. *Let $\mathbb{X}$ and $\mathbb{Y}$ be two finite compact sets. Let $\lambda_{\mathbb{X}}$ and $\lambda_{\mathbb{Y}}$ be the persitence landscapes associated to the Vietoris-Rips or Čech filtrations. For any $t \in \mathbb{R}$ and any $k \in \mathbb{N}$, we have*

$$|\lambda_{\mathbb{X}}(k,t) - \lambda_{\mathbb{Y}}(k,t)| \leq d_{\infty}(\mathrm{dgm}(\mathbb{X}), \mathrm{dgm}(\mathbb{Y})).$$

For the proof of this proposition we refer to theorem 5.2 of [26]

On the other hand, landscapes highlight points far from the diagonal, which represent persistent features of the data. This implies that secondary and tertiary features are not given as much importance. However, they are usually useful when discriminating between data sets.

5.2 Linear representations

In this subsection we will focus on a particular case of vectorizations that are defined as integrals of certain functions. Indeed, they do not rely on the fact that the measure must be a persistence diagram, so we will work on the metric space $(\mathcal{M}^p, \mathrm{OT}_p)$, with $1 \leq p < \infty$.

Definition 5.2.1. A **linear representation** is a vectorization $\Phi : \mathcal{M}^p \to \mathcal{B}$ of the form

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

for some fixed function $f : \Omega \to \mathcal{B}$.

Typically $\mathcal{B} = \mathbb{R}^m$. One of the reasons why this type of vectorizations are interesting is because they are simple to define and compute. For instance, in the practical case of finite persistence diagrams, they transform into a sum $\Phi\left(\sum_{i=1}^n \delta_{x_i}\right) = \sum_{i=1}^n f(x_i)$, where f sends each point of the persistence diagram to an element of the Banach space. Moreover, they include many vectorization methods of topological data analysis, as we will see.

Thanks to how Φ is defined, the study of its continuity is directly related with the vague topology on $\mathcal{M}^p$, and consequently with optimal transport distance. Therefore, chapter 4 may be used to characterize the continuity of these linear representations in terms of the function f, as proposition below shows. The map $d : \bar{\Omega} \times \bar{\Omega} \to [0, \infty)$ denotes the metric $||\cdot||_q$, with $1 \leq q \leq \infty$.

Proposition 5.2.2 ([6])**.** *Consider a Banach space $\mathcal{B}$ and a function $f : \Omega \to \mathcal{B}$. Let $\Phi : \mathcal{M}^p \to \mathcal{B}$ be the linear representation defined by $\Phi(\mu) = \int_{\Omega} f(x) d\mu(x)$. Then, it is continuous with respect to the optimal transport distance OT_p and the norm $||\cdot||_{\mathcal{B}}$ if and only if f belongs to the class of functions*

$$C^0_{b,p} = \left\{ f : \Omega \to \mathcal{B} \ : \ f \ \ \textit{continuous and} \ \ x \mapsto \frac{||f(x)||_{\mathcal{B}}}{d(x, \partial\Omega)^p} \ \ \textit{bounded} \right\}.$$

Proof. Let $f \in C^0_{b,p}$ and consider the measures $\mu, \mu_1, \mu_2, \ldots \in \mathcal{M}^p$ such that $\mathrm{OT}_p(\mu_n, \mu)$ tends to zero. To prove continuity of Φ we have to check that $\lim_{n\to\infty} \Phi(\mu_n) = \Phi(\mu)$. Thanks to theorem 2 of [27], it is enough to study the case where $\mathcal{B} = \mathbb{R}$. Corollary 4.3.13 implies that convergence in optimal transport distance is equivalent to weak convergence of the measures $\mu_n^{[p]} \xrightarrow{w} \mu^{[p]}$. By assumption, $\frac{f(x)}{d(x,\partial\Omega)^p}$ is bounded in Ω, so thanks to the characterization of weak convergence (B.1),

$$\int_\Omega \frac{f(x)}{d(x,\partial\Omega)^p} d\mu_n^{[p]}(x) \to \int_\Omega \frac{f(x)}{d(x,\partial\Omega)^p} d\mu^{[p]}(x).$$

Notice that $\mu_n^{[p]} \ll \mu_n$ and $\mu^{[p]} \ll \mu$. Applying Radon-Nikodym theorem B.0.13 the limit above transforms to,

$$\begin{aligned}\Phi(\mu_n) = \int_\Omega f(x) d\mu_n(x) &= \int_\Omega f(x) \frac{1}{d(x,\partial\Omega)^p} d\mu_n^{[p]} \\ &\longrightarrow \int_\Omega f(x) \frac{1}{d(x,\partial\Omega)^p} d\mu^{[p]}(x) = \int_\Omega f(x) d\mu(x) = \Phi(\mu),\end{aligned}$$

as wanted.

For the other implication we proceed by contradiction. Suppose first that $f : \Omega \to \mathcal{B}$ is not continuous at some $x \in \Omega$. Then, there exists a sequence of points $\{x_n\}_n \subset \Omega$ such that $x_n \to x$ but $f(x_n)$ does not converge to $f(x)$. Consider now the measures $\mu_n = \delta_{x_n}$ and $\mu = \delta_x$, both in $\mathcal{M}^p$. It is clear, by theorem 4.3.11, that $\lim_{n\to\infty} \mathrm{OT}_p(\mu_n, \mu) = 0$. However, $\Phi(\mu_n) = f(x_n)$ does not converge to $\Phi(\mu) = f(x)$. That is, Φ is not continuous for the optimal transport distance. On the other hand, assume f is continuous but $x \to \frac{f(x)}{d(x,\partial\Omega)^p}$ is not bounded. This implies that we can find a sequence $\{x_n\}_n \subset \Omega$ such that

$$\lim_{n\to\infty} \left|\left| \frac{f(x_n)}{d(x_n,\partial\Omega)^p} \right|\right| = \infty.$$

Consider the persistence measures $\mu_n = \frac{1}{||f(x_n)||}\delta_{x_n}$. Notice that the above limit implies

$$\lim_{n\to\infty} \mathrm{OT}_p^p(\mu_n, 0) = \lim_{n\to\infty} \mathrm{Pers}_p(\mu_n) = \lim_{n\to\infty} \frac{1}{||f(x_n)||} d(x_n,\partial\Omega)^p = 0.$$

Therefore, in order to be Φ continuous it should satisty $\Phi(\mu_n) \to \Phi(0) = 0$. But $\Phi(\mu_n) = \frac{f(x_n)}{||f(x_n)||}$, which has norm equal to one for every $n \in \mathbb{N}$. ■

Remark 5.2.3. Note that proposition 5.2.2 not only provides a method to built continuous linear representations, but asserts that it is the only way to do so. It consists in constructing the functions $f : \Omega \to \mathcal{B}$ as follows. Firstly, a function ϕ on Ω with enough good characteristics is considered. Then, this function is weighted with respect to the distance to the diagonal $d(x,\partial\Omega)^p$, obtaining $f(x) = \phi(x)d(x,\partial\Omega)^p$. Examples below exhibit that this produces relevant vectorization methods. ♢

Apart from continuity, we would like, as we have said, that this linear representations satisfies some kind of stability,

$$||\Phi(\mu) - \Phi(\nu)|| \leq C \cdot \mathrm{OT}_p(\mu, \nu),$$

for some constant C. Fortunately, that is the case for $p = 1$.

Proposition 5.2.4 ([6]). *Let $T \subset \mathbb{R}$ and consider a family $(f_t)_{t\in T}$ such that each $f_t : \bar{\Omega} \to \mathbb{R}$ is a Lipschitz continuous function with Lipschitz constant less than one and satisfying $f_t(\partial\Omega) = 0$. Then, the linear representation*

$$\Phi : \mathcal{M}^1 \to \mathbb{R}^T; \qquad \Phi(\mu) = (\varphi_{f_t}(\mu))_{t\in T} = \left(\int_\Omega f_t(x)d\mu(x)\right)_{t\in T}$$

satisfies for all $\mu, \nu \in \mathcal{M}^1$

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

Proof. Let $\mu, \nu \in \mathcal{M}^1$ and consider and optimal transport plan $\pi \in \mathrm{Opt}(\mu, \nu)$. For any $t \in T$, since f_t is zero on the boundary and using the marginal constraints of π,

$$\begin{aligned}
|\varphi_{f_t}(\mu) - \varphi_{f_t}(\nu)| &= \int_\Omega f_t(x)d\mu(x) - \int_\Omega f_t(y)d\nu(y) = \int_{\bar{\Omega}} f_t(x)d\mu(x) - \int_{\bar{\Omega}} f_t(y)d\nu(y) \\
&= \int_{\bar{\Omega}\times\bar{\Omega}} f_t(x)d\pi(x, y) - \int_{\bar{\Omega}\times\bar{\Omega}} f_t(y)d\pi(x, y) \\
&= \int_{\bar{\Omega}\times\bar{\Omega}} (f_t(x) - f_t(y))d\pi(x, y).
\end{aligned}$$

Since f_t has Lipschitz constant less than one

$$\int_{\bar{\Omega}\times\bar{\Omega}} (f_t(x) - f_t(y))d\pi(x, y) \leq \int_{\bar{\Omega}\times\bar{\Omega}} d(x, y)d\pi(x, y) = \mathrm{OT}_1(\mu, \nu).$$

All in all, $|\varphi_{f_t}(\mu) - \varphi_{f_t}(\nu)| \leq \mathrm{OT}_1(\mu, \nu)$ for every $t \in T$, which implies

$$||\Phi(\mu) - \Phi(\nu)||_\infty \leq \mathrm{OT}_1(\mu, \nu).$$

■

Finally, we present some examples of continuous linear representations that are used in topological data analysis (figure 5.2). As before, these definitions do not rely on the fact that we are working with persistence diagrams, and actually are well-defined for any persistence measure in $\mathcal{M}^p$.

Persistence surfaces For this linear representation a non-negative Lipschitz continuous bounded function $K : \mathbb{R}^4 \to \mathbb{R}$ is chosen. For example,

$$K(x, y) = \exp\left(-\frac{||x - y||^2}{2}\right).$$

The map $f : \Omega \to C_b(\mathbb{R}^2)$ is defined as $f(x) = d(x, \partial\Omega)^p K(x, \cdot)$, so $f(x) : \mathbb{R}^2 \to \mathbb{R}$. Hence, the persistence surface is

$$\Phi(\mu) = \int_\Omega f(x)d\mu(x) = \int_\Omega d(x, \partial\Omega)^p K(x, \cdot)d\mu(x),$$

which sends each persistence diagram to an element of the space of continuous bounded functions $(C_b(\mathbb{R}^2), ||\cdot||_\infty)$. Obviously, K being bounded and $f(x)$ being proportional to $d(x, \partial\Omega)^p$ implies that $x \mapsto \frac{||f(x)||_\infty}{d(x,\partial\Omega)^p}$ is bounded. Thus, Φ is continuous. In particular, for a finite persistence diagram we obtain

$$\Phi\left(\sum_{i=1}^n \delta_{x_i}\right) = \sum_{i=1}^n f(x_i) = \sum_{i=1}^n d(x_i, \partial\Omega)^p K(x_i, \cdot).$$

Persistence silhouettes In this case consider the function

$$\Lambda : \Omega \times \mathbb{R} \to \mathbb{R}; \qquad \Lambda(x,t) = \Lambda_{h(x)}(t),$$

where h and Λ_r are given by (5.1) and (5.2) respectively. Notice that, for a fixed $x = (x_1, x_2) \in \Omega$, the map $\Lambda(x, \cdot) : \mathbb{R} \to \mathbb{R}$ is bounded. Take $f : \Omega \to C_b(\mathbb{R})$ as $f(x) = d(x, \partial\Omega)^{p-1}\Lambda(x, \cdot)$. Then, the persistence silhouette

$$\Phi(\mu) = \int_\Omega f(x)d\mu(x) = \int_\Omega d(x, \partial\Omega)^{p-1}\Lambda(x, \cdot)d\mu(x),$$

takes values in the space $(C_b(\mathbb{R}), ||\cdot||_\infty)$. Notice that the maximum value of $\Lambda(x, \cdot)$ is $\frac{x_2 - x_1}{2}$, which is proportional to $d(x, \partial\Omega)$. Therefore, $||f(x)||_\infty$ is proportional to $d(x, \partial\Omega)^p$. By proposition 5.2.2, Φ is continuous. In the case of finite persistence diagrams, this representation consists of a weighted sum of the different functions of the persistence landscape,

$$\Phi\left(\sum_{i=1}^n \delta_{x_i}\right) = \sum_{i=1}^n f(x_i) = \sum_{i=1}^n d(x_i, \partial\Omega)^{p-1}\Lambda(x_i, \cdot).$$

Weighted Betti curves The construction of the weighted Betti curves is similar to the previous ones. Let $p, p' \geq 1$ and for each $t \in \mathbb{R}$ consider the rectangle $R_t = (-\infty, t] \times [t, +\infty)$. Define the function $f : \Omega \to L_{p'}(\mathbb{R})$ as

$$x \mapsto \left[f(x) : t \mapsto d(x, \partial\Omega)^{p-\frac{1}{p'}} 1_{R_t}(x)\right],$$

where $1_{R_t}(x) = 1$ if $x \in R_t$ and zero otherwise. Observe that, for a given $x = (x_1, x_2) \in \Omega$, the function $f(x)(t)$ is nonzero if and only if $t \in [x_1, x_2]$. The linear representation is then given by

$$\Phi(\mu)(t) = \int_\Omega f(x)(t)d\mu(x) = \int_\Omega d(x, \partial\Omega)^{p-\frac{1}{p'}} 1_{R_t}(x)d\mu(x),$$

with $\Phi(\mu) \in (L_{p'}(\mathbb{R}), ||\cdot||_{p'})$. It is easy to see that $||f(x)||_{p'}$ is proportional to $d(x, \partial\Omega)^p$. Indeed,

$$\begin{aligned} ||f(x)||_{p'}^{p'} &= \int_{\mathbb{R}} |f(x)(t)|^{p'} dt = \int_{[x_1,x_2]} (f(x)(t))^{p'} dt = \int_{[x_1,x_2]} \left(d(x, \partial\Omega)^{p-\frac{1}{p'}}\right)^{p'} dt \\ &= \int_{[x_1,x_2]} d(x, \partial\Omega)^{pp'-1} dt = d(x, \partial\Omega)^{pp'-1}(x_2 - x_1). \end{aligned}$$

As noticed before, $d(x, \partial\Omega)$ is proportional to $x_2 - x_1$. Consequently, we conclude that $||f(x)||_{p'}^{p'}$ is proportional to $d(x, \partial\Omega)^{pp'}$. By proposition 5.2.2, Φ is continuous.

When $p = 1 = p'$ is chosen, this representation is called Betti curve and, in the finite case, is given by a sum of step functions

$$\Phi\left(\sum_{i=1}^n \delta_{x_i}\right) = \sum_{i=1}^n f(x_i)(t) = \sum_{i=1}^n 1_{R_t}(x_i).$$

For each $t \in \mathbb{R}$ the sum is equal to the number of points x_i of the persistence diagram that lie in $R_t = (-\infty, t] \times [t, +\infty)$. If the persistence diagram comes from the n-th homology groups of a filtration $(X_t)_{t\in\mathbb{R}}$, that is equal to the Betti number of X_t.

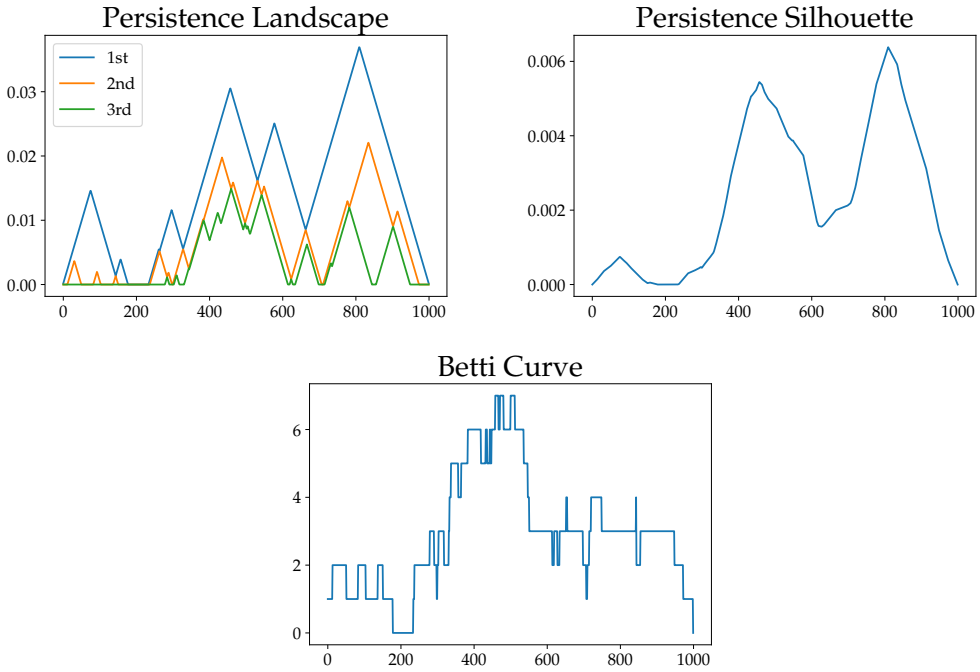


Figure 5.2: Examples of vectorizations of the blue persistence diagram in 2.3. From right to left: the first 3 persistence landscapes, the persistence silhouette and the Betti curve.

5.3 Point transformations

Let us recover a more applied point of view by letting $\mathcal{B} = \mathbb{R}^m$ and by considering only finite persistence diagrams, since those are the ones obtained from any practical setting. The building method presented in remark 5.2.3 can be generalized. As before, we begin with a function $\phi : \mathbb{R}^2 \to \mathbb{R}^m$, which is called **point transformation** because it transforms each point of the persistence diagram to a vector in $\mathbb{R}^m$. Then, we weight this function, but instead of using the distance to the diagonal, we let the weight be any map $w : \mathbb{R}^2 \to \mathbb{R}$, called **weight function**. If we considered the linear representation associated to this construction $f(x) = w(x)\phi(x)$, for any finite persistence diagram $\mu = \sum_{i=1}^n \delta_{x_i}$ we would get a sum $\sum_{i=1}^n w(x_i)\phi(x_i)$. However, more general operators may be used in order to construct the vectorization. Hence, let op be a permutation invariant operator, i.e. $\mathsf{op}(x_1, \cdots, x_n) = \mathsf{op}(x_{\sigma(1)}, \cdots, x_{\sigma(n)})$ for any permutation $\sigma \in \mathfrak{S}_n$. Some examples are minimum, maximum, sum, $k$max,.... Finally, we define

$$\Phi(\mu) = \mathsf{op}\left(\{w(x_i)\phi(x_i)\}_{1 \leq i \leq n}\right) \in \mathbb{R}^m, \tag{5.3}$$

for any $\mu = \sum_{i=1}^n \delta_{x_i}$ finite persistence diagram.

Remark 5.3.1. Although the continuity provided by proposition 5.2.2 is lost in general, this setup has relevant properties regarding machine learning techniques, which will be underlined. ♢

Example 5.3.2. An example of point transformation is the **triangle point trans-**

formation, defined as

$$\phi_\Lambda(x) = [\Lambda_x(t_1), \Lambda_x(t_2), \cdots, \Lambda_x(t_m)]^T,$$

where $t_1, \cdots, t_m \in \mathbb{R}$ and $\Lambda_x(\cdot)$ is defined by equation (5.2). The **Gaussian point transformation** is given by

$$\phi_\Gamma(x) = [\Gamma(x, y_1), \Gamma(x, y_2), \cdots, \Gamma(x, y_m)]^T,$$

where $y_i \in \mathbb{R}^2$ for $1 \leq i \leq m$ and, for a given $\sigma > 0$, the map $\Gamma(x, \cdot) : \mathbb{R} \to \mathbb{R}$ is

$$\Gamma(x, y) = \exp\left(-\frac{||x - y||_2^2}{2\sigma^2}\right), \tag{5.4}$$

♢

The generality of the formalism given by (5.3) lets us recover standard vectorization methods.

- Take as point transformation the triangle one with $m = 1$ composed with the change of coordinates $h : \mathbb{R}^2 \to \mathbb{R}^2$ given in (5.1), i.e. $\phi_\Lambda \circ h$. As weight function let $w = 1$ and $\mathsf{op} = k\text{max}$. Then, $\Phi(\mu)$ gives the k-th landscape evaluated at $t \in \mathbb{R}$.
- Consider again the composition $\phi_\Lambda \circ h$, together with $\mathsf{op} = \text{sum}$ and an arbitrary weight function w. Then, $\Phi(\mu)$ is equivalent to evaluate the persistence silhouette, weighted by w, on $t_1, \cdots, t_m \in \mathbb{R}$.
- In this case, consider the Gaussian point transformation ϕ_Γ, $\mathsf{op} = \text{sum}$ and an arbitrary w. Then, $\Phi(\mu)$ provides the evaluation of the persistence surface, weighted by w, on $y_1, \cdots, y_m \in \mathbb{R}^2$ and being $K(x, \cdot) = \Gamma(x, \cdot)$.

Remark 5.3.3. When transforming a persistence diagram into a vector, the weight function $w : \mathbb{R}^2 \to \mathbb{R}$ let us hightlight certain points or areas of the diagram that are useful to achieve a given task. In the case of linear representations, $w(x)$ was a power of $d(x, \partial\Omega)$, which is greater when the point $x \in \Omega$ is far from the diagonal, i.e. it represents a feature that persists more. Hence, those features are weighted more than the ones who last little in time, and which are represented by points close to the diagonal in the persistence diagram. Although this provides continuity of the representation, depending on the setup it may be better to underline other properties by choosing different w. The formulation presented in this section let the user choose the weight function that suits him best. ♢

A general drawback of this freedom of choice is that, since the function f has to be fixed in advance, it might not be optimal to perform a given task. Regarding this problem, the success of deep neural networks has shown that learning representations is a strong recommended approach. Actually, the formulation given by (5.3) can be used to optimize $f(x) = w(x)\phi(x)$, constructing a learnable vectorization. In order to do this we parametrized the weight $w = w_{\theta_1}$ and point $\phi = \phi_{\theta_2}$ functions, where $(\theta_1, \theta_2) \in \mathbb{R}^D$. The idea is that the expression of $\Phi(\mu)$ could be used as a generic

neural network layer for persistence diagrams and then the parameters could be learned during a training phase. In practice, this is done by back-propagation, assuming that we know and can implement the gradients of $\theta_1 \mapsto w_{\theta_1}$ and $\theta_2 \mapsto \phi_{\theta_2}$. A network architecture is illustrated in [6]. Besides, approaches towards learning task-optimal stable representations of persistence diagrams have been studied and developed, e.g. [28].

5.4 Linearized optimal transport

In the classical transport problem setting, linearized optimal transport provides a method to embed measures into a vector space. The latter consists in the following. Consider the Kantorovich formulation (4.5) for two probability measures ρ, μ in $\mathbb{R}^2$ with finite second moment and cost function $c(x, y) = ||x - y||^2$. By theorem 4.3.1, whether ρ is absolutely continuous with respect to the Lebesgue measure, the optimal thansport plan is unique and induced by a transport map $\pi = (id, T_\mu)\#\rho$, where $T_\mu = \nabla\phi$ for a convex function ϕ. If we fix ρ, it is possible to define a vectorization as $\mu \mapsto T_\mu \in L^2(\rho, \mathbb{R}^2)$. Injectivity is clear because in this case $T_\mu\#\rho = \mu$, being T_μ the unique optimal transport map between ρ and μ. Besides, stability results of this vectorization with respect to the Wasserstein distance (4.6) and the $L^2(\rho)$ norm have been proven in [29] and [30].

In this section we mimic this idea using a similar result stated in [5], which connects optimal admissible transport plans and transport maps. The advantage of this formulation, as we have seen, is that it can be used with measures that have different, or even infinite, mass as long as they have finite total persistence. As it has been done before, we focus on the practical case of finite persistence diagrams, which are represented as finite point measures on Ω. Following remark 4.3.6, we can assume them to be supported on a bounded subset $\mathcal{Y}$ of Ω and then apply directly what has been proved in [5].

In particular, the result that concerns us in this section is corollary 2.5 of [5]. We state the general result.

Proposition 5.4.1 (Uniqueness of optimal plans)**.** *Consider $\mathcal{X}$ a bounded open set of $\mathbb{R}^2$ with the euclidean distance d. Let $\rho, \mu \in \mathcal{M}^2(\mathcal{X})$ and fix $\gamma \in \mathrm{Opt}(\rho, \mu)$.*

1. *If ρ is absolutely continuous with respect to the Lebesgue measure, then $\gamma|_{\mathcal{X}\times\bar{\mathcal{X}}}$ is unique, and it is given by $(id, T)\#\rho$, where $T : \mathcal{X} \to \bar{\mathcal{X}}$ is the gradient of a convex function. However, γ as a whole may be not uniquely defined since there might be multiple ways of bringing the mass from the boundary to μ if no hypothesis on μ are made.*

2. *If moreover μ is absolutely continuous with respect to the Lebesgue measure, then γ is unique.*

The proof follows exactly as in the classical transport problem due to the equivalence $(i) \Leftrightarrow (iii)$ of proposition 2.3 in [5]. The uniqueness of the first case only over $\mathcal{X} \times \bar{\mathcal{X}}$ comes from the fact that $p_1\#\gamma = \rho$ only on $\mathcal{X}$.

Since we want to apply this to persistence diagrams, assume from now on that ν is a finite point measure. By fixing $\rho \in \mathcal{M}^2(\mathcal{X})$ satisfying *1*, it is possible to associate to each ν the map T_ν coming from $\gamma_\nu|_{\mathcal{X}\times\bar{\mathcal{X}}} = (id, T_\nu)\#\rho$.

Remark 5.4.2. Observe that, in spite of the fact that for two different measures $\nu \neq \nu'$ optimal transport plans $\gamma_\nu \in \mathrm{Opt}(\rho, \nu)$ and $\gamma_{\nu'} \in \mathrm{Opt}(\rho, \nu')$ are also distinct, a priori, the proposition does not assure us that T_ν and $T_{\nu'}$ are not equal. That is, in general it may happen that γ_ν and $\gamma_{\nu'}$ bring in the same way the mass from $\mathcal{X}$ to $\bar{\mathcal{X}}$, but differ on how they bring the mass from the boundary. In order to obtain injectivity for the map $\nu \mapsto T_\nu$, let us study if this holds for some choice of ρ. Suppose $\nu \neq \nu'$ but $T_\nu = T_{\nu'}$. Then, for any $x \in \mathcal{X}$, $\gamma_\nu(\mathcal{X} \times \{x\}) = \gamma_{\nu'}(\mathcal{X} \times \{x\})$. Recall that $\gamma_\nu(\bar{\mathcal{X}} \times \{x\}) = \nu(x)$ and the same equality holds for $\gamma_{\nu'}$ and ν'. Since ν and ν' does not have the same support, this implies the existence of a point $x \in \mathrm{spt}(\nu)$ (or ν') but $x \notin \mathrm{spt}(\nu')$ (or ν) such that the optimal transport plan γ_ν does not bring mass from $\mathcal{X}$ to x, i.e. $\gamma_\nu(\mathcal{X} \times \{x\}) = 0$ and $\gamma_\nu(\bar{\mathcal{X}} \times \{x\}) = \gamma(\partial\mathcal{X} \times \{x\}) = 1$. As a result, to have injectivity it is sufficient to choose ρ such that for any $\nu = \sum_{i=1}^n \delta_{x_i}$ the optimal transport plan satisfies $\gamma_\nu(\mathcal{X} \times \{x_i\}) \neq 0$ for all $i = 1, \ldots, n$. If this happens, $T_\nu = T_{\nu'}$ implies $\nu = \nu'$. ♢

Henceforth ρ is fixed as the uniform measure. Similar to the classical optimal transport problem, it can be proven that $T_\nu \in L^2(\rho, \mathcal{X})$. Indeed, for any $y \in \bar{\mathcal{X}}$, using Minkowski inequality

$$\left(\int_{\mathcal{X}} d^2(T_\nu(x), y)d\rho(x)\right)^{1/2} \leq \left(\int_{\mathcal{X}} d^2(T_\nu(x), x)d\rho(x)\right)^{1/2} + \left(\int_{\mathcal{X}} d^2(x, y)d\rho(x)\right)^{1/2}.$$

Because $\mathcal{X}$ is bounded and ρ is finite, the second integral is finite too. On the other hand,

$$\int_{\mathcal{X}} d^2(T_\nu(x), x)d\rho(x) = \int_{\mathcal{X}} d^2(x, T_\nu(x))d\rho(x) = \int_{\mathcal{X}} d^2(\cdot,\cdot) \circ (id, T_\nu)(x)d\rho(x) = \int_{\mathcal{X}\times\bar{\mathcal{X}}} d^2(x, y)d((id, T_\nu)\#\rho)(x, y) = \int_{\mathcal{X}\times\bar{\mathcal{X}}} d^2(x, y)d\gamma_\nu(x, y).$$

The space $\mathcal{X}\times\bar{\mathcal{X}} \subset \mathbb{R}^4$ is bounded and $\gamma_\nu(\mathcal{X}\times\bar{\mathcal{X}}) = \rho(\mathcal{X}) = 1$. As a consequence, the initial integral is finite and $T_\nu \in L^2(\rho, \mathcal{X})$. In conclusion, we propose the following definition.

Definition 5.4.3. Consider finite point measures ν in $\mathcal{M}^2(\mathcal{X})$, where $\mathcal{X} \subset \Omega$ is an open bounded subset of $\mathbb{R}^2$, and fix the uniform measure ρ on $\mathcal{X}$. Then, we define a vectorization

$$\nu \in (\mathcal{M}^2(\mathcal{X}), \mathrm{OT}_2) \mapsto T_\nu \in (L^2(\rho, \mathcal{X}), ||\cdot||_{L^2}),$$

where T_ν is given by proposition 5.4.1.

Proposition 5.4.4. *Given two finite point measures $\nu_1, \nu_2 \in \mathcal{M}^2(\mathcal{X})$ consider the maps T_1, T_2 from the above definition. Then,*

$$\mathrm{OT}_2^2(\nu_1, \nu_2) \leq ||T_1 - T_2||^2_{L^2(\rho)} + \mathrm{Pers}_2(\nu_1) + \mathrm{Pers}_2(\nu_2).$$

Proof. Let $\gamma_1 \in \mathrm{Opt}(\nu_1, \rho)$ and $\gamma_2 \in \mathrm{Opt}(\rho, \nu_2)$. Consider the measure $\gamma \in \mathcal{M}(\bar{\mathcal{X}} \times \bar{\mathcal{X}})$ defined as

$$\gamma = (T_1, T_2)\#\rho + \gamma_2|_{\partial\mathcal{X}\times\bar{\mathcal{X}}} + \gamma_1|_{\bar{\mathcal{X}}\times\partial\mathcal{X}},$$

It is an admissible plan between ν_1 and ν_2. In fact, for every $A \subset \mathcal{X}$,

$$\begin{aligned} p_1\#\gamma(A) &= \rho(T_1^{-1}(A)) + \gamma_2(\partial\mathcal{X}\times\bar{\mathcal{X}} \cap A\times\bar{\mathcal{X}}) + \gamma_1(\bar{\mathcal{X}}\times\partial\mathcal{X}\cap A\times\bar{\mathcal{X}}) \\ &= \rho(T_1^{-1}(A)) + \gamma_1(A\times\partial\mathcal{X}), \\ p_2\#\gamma(A) &= \rho(T_2^{-1}(A)) + \gamma_2(\partial\mathcal{X}\times\bar{\mathcal{X}} \cap \bar{\mathcal{X}}\times A) + \gamma_1(\bar{\mathcal{X}}\times\partial\mathcal{X}\cap\bar{\mathcal{X}}\times A) \\ &= \rho(T_2^{-1}(A)) + \gamma_2(\partial\mathcal{X}\times A). \end{aligned}$$

On the other hand,

$$\begin{aligned} \nu_1(A) &= \gamma_1(A\times\bar{\mathcal{X}}) = \gamma_1(A\times\mathcal{X}) + \gamma_1(A\times\partial\mathcal{X}) = \rho(T_1^{-1}(A)) + \gamma_1(A\times\partial\mathcal{X}), \\ \nu_2(A) &= \gamma_2(\bar{\mathcal{X}}\times A) = \gamma_2(\mathcal{X}\times A) + \gamma_2(\partial\mathcal{X}\times A) = \rho(T_2^{-1}(A)) + \gamma_2(\partial\mathcal{X}\times A). \end{aligned}$$

Thus, they coincide. Therefore,

$$\begin{aligned} \mathrm{OT}_2^2(\nu_1,\nu_2) &\le \int_{\bar{\mathcal{X}}\times\bar{\mathcal{X}}} d^2(x,y)d\gamma(x,y) = \int_{\bar{\mathcal{X}}\times\bar{\mathcal{X}}} d^2(x,y)d((T_1,T_2)\#\rho) \\ &+ \int_{\bar{\mathcal{X}}\times\bar{\mathcal{X}}} d^2(x,y)d(\gamma_2|_{\partial\mathcal{X}\times\bar{\mathcal{X}}}) + \int_{\bar{\mathcal{X}}\times\bar{\mathcal{X}}} d^2(x,y)d(\gamma_1|_{\bar{\mathcal{X}}\times\partial\mathcal{X}}) \\ &= \int_{\mathcal{X}} d^2(T_1(x),T_2(x))d\rho(x) + \int_{\partial\mathcal{X}\times\bar{\mathcal{X}}} d^2(x,y)d\gamma_2(x,y) + \int_{\bar{\mathcal{X}}\times\partial\mathcal{X}} d^2(x,y)d\gamma_1(x,y) \\ &= ||T_1 - T_2||^2_{L^2(\rho)} + \int_{\partial\mathcal{X}\times\mathcal{X}} d^2(x,y)d\gamma_2(x,y) + \int_{\mathcal{X}\times\partial\mathcal{X}} d^2(x,y)d\gamma_1(x,y), \end{aligned}$$

where the last equality follows from the fact that any optimal admissible transport plan can be assumed to satisfy $\gamma|_{\partial\mathcal{X}\times\partial\mathcal{X}} = 0$. Let us now study the last two integrals. Proposition 2.3 (ii) of [5] states that optimality of γ_i (for $i = 1, 2$) implies that they are concentrated on

$$\{(x,y) \in \bar{\mathcal{X}}\times\bar{\mathcal{X}} \ : \ d^2(x,y) \le d^2(x,\partial\mathcal{X}) + d^2(y,\partial\mathcal{X})\}.$$

Consequently,

$$\begin{aligned} \int_{\partial\mathcal{X}\times\mathcal{X}} d^2(x,y)d\gamma_2(x,y) &\le \int_{\partial\mathcal{X}\times\mathcal{X}} d^2(y,\partial\mathcal{X})d\gamma_2(x,y) \\ &= \int_{\partial\mathcal{X}\times\mathcal{X}} d^2(\cdot,\partial\mathcal{X})\circ p_2(x,y)d\gamma_2(x,y) \\ &= \int_{\mathcal{X}} d^2(y,\partial\mathcal{X})d(p_2\#(\gamma_2|_{\partial\mathcal{X}\times\mathcal{X}}))(y) \\ &\le \int_{\mathcal{X}} d^2(y,\partial\mathcal{X})d(p_2\#\gamma_2)(y) = \int_{\mathcal{X}} d^2(y,\partial\mathcal{X})d\nu_2(y) = \mathrm{Pers}_2(\nu_2). \end{aligned}$$

Likewise,

$$\int_{\mathcal{X}\times\partial\mathcal{X}} d^2(x,y)d\gamma_1(x,y) \le \mathrm{Pers}_2(\nu_1).$$

■

Chapter 6

Numerical results

Finally, we include results on real data in order to show computations of the different notions of TDA we have introduced (Rips complexes, persistence barcodes, diagrams, etc.) and to exemplify the use of vectorizations when performing linear tasks. In particular, we implement PCA. In the following, we present a summary of what can be found in this Python Jupyter Notebook.

The data set used for these computations can be found in GUDHI TDA tutorial. It has been obtained from the walk of 3 persons A, B and C. The walk has been recorded using the accelerometer sensor of a smartphone in their pocket, giving rise to 3 time series, each of them representing the 3 coordinates of the acceleration of the corresponding walker in a coordinate system attached to the sensor. Each series has been splitted in a list of 100 times series made of 200 consecutive coordinate points.

We compute a Rips complex for each time series, obtaining 100 complexes per walker. As maximum radius we choose $r = 0.2$, not considering simplexes of dimension greater than 3. In particular, this means that all n-th homology groups with $n > 3$ are zero. Then, from those Rips complexes, we compute the persistence barcode and persistence diagram. In total, we obtain 300 persistence diagrams, each of them corresponding to a splitted series of one of the walkers. Figure 2.3 shows the persistence diagram associated to the first time series of walker A, containing the first 200 acceleration coordinates.

As we have seen, it is possible to transform each of these persistence diagrams into a vector. In this case, we obtain the first five landscapes and the Betti curve of each diagram. The former corresponds to a vector in $\mathbb{R}^{5000}$, whereas the latter is a vector in $\mathbb{R}^{1000}$. In figure 5.2 we represent the first three landscapes, the Betti curve and the persistence silhouette associated to the first time series of walker A. That is, they are the vectorizations of the persistence diagram of figure 2.3.

Due to the high dimensionality of this vectorizations, we execute principal component analysis to reduce their dimension and visualize them. After doing this, we observe whether the three walkers can be differentiated from the information encoded in the landscapes or the Betti curves. It seems that Betti curves perform better for this task and that, in both cases, walker C stands out above the other two, which are more intermingled.

Appendix A

Simplicial homology

In this chapter we review the bases of homology theory, which is needed to encode and study the topological features on topological data analysis. We focus on simplicial homology, since it is the one developed specially for simplicial complexes and used in applications. We refer to [12] and [31] for more details.

A.1 Simplicial complexes

A simplicial complex is basically a generalization of a graph where, instead of connecting only pairs of points with edges, freedom is left to define higher n-dimensional structures connecting n-tuples "properly".

Definition A.1.1. Given a set $\{x_0, \ldots, x_n\} \subset \mathbb{R}^d$ of affine independent points, its convex hull is called **n-simplex** and is defined as

$$\sigma = [x_0, \ldots, x_n] = \left\{ \sum_{i=0}^{n} a_i x_i \ : \ \sum_{i=0}^{n} a_i = 1, \ a_i \leq 0 \right\}.$$

A face of σ is any simplex which can be built by a subset of $\{x_0, \ldots, x_n\}$. The 0-simplexes $[x_i]$ are called vertices.

Definition A.1.2. A finite **geometric simplicial complex** K in $\mathbb{R}^d$ is a finite collection of simplexes in $\mathbb{R}^d$ such that

1. If a simplex $\sigma \in K$, then all the faces of σ belong to K;
2. if some simplexes $\sigma, s \in K$, then $\sigma \cap s$ is empty or a common face of both.

In particular, all the vertices belong to K.

Definition A.1.3. Given $T \subset K$, it is a **subsimplicial complex** or **subcomplex** if T is a geometric simplicial complex and for every $t \in T$, $t \in K$.

Remark A.1.4. Notice that simplexes are fully characterized by its vertices. Each n simplex has $n+1$ distinct vertices, and no other n simplex has this same set of vertices. Thus, K can be described combinatorially as a set V of vertices together

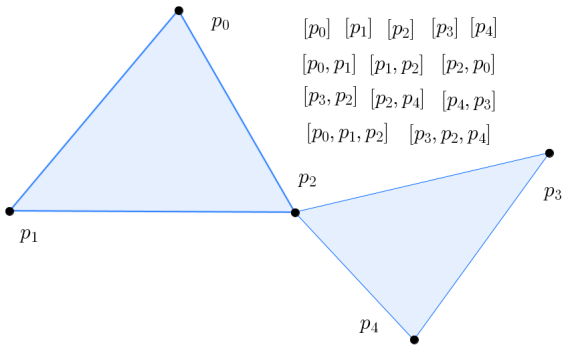


Figure A.1: Visualization of a simplicial complex and its simplices.

with sets X_n of n-simplixes σ_n, which are $(n+1)$-element subsets of V. The only requirement is that, each $(k+1)$-element subset of the vertices of an n-simplex σ_n in X_n, is a k-simplex in X_k . From this combinatorial data, and once we have chosen a partial ordering of the vertices V, K can be constructed. This ordering restricts to a linear ordering on the vertices of each simplex in X_n. ♢

The dimension of a simplicial complex K is given by

$$\dim K = \sup\{n \ : \ \exists \sigma \in K, \ \sigma \text{ is a } n\text{-simplex}\}.$$

Definition A.1.5. Given a finite geometric simplicial complex K, we can associate to it a topological space

$$|K| = \bigcup_{\sigma \in K} \sigma \subset \mathbb{R}^d,$$

which inherits the topology of $\mathbb{R}^d$.

Note that this topological space is compact. The notion of geometric simplicial complex can be generalized. By doing this we obtain objects which are purely combinatorial structures and do not need to be embedded in any Euclidean space, although it is possible if needed.

Definition A.1.6. Given a set V, an **abstract simplicial complex** with vertex set V is a family $\tilde{K}$ of finite nonempty subsets σ of V satisfying:

1. the elements of V belongs to $\tilde{K}$;
2. if $\sigma \in \tilde{K}$, then any subset of σ also belongs to $\tilde{K}$.

These finite subsets σ of V are called **abstract simplexes**.

In practice, since we are working with data sets which are finite, we will only have to deal with finite abstract simplicial complexes.

The dimension of an abstract simplex is its cardinality minus one. The dimension of an abstract simplicial complex is

$$\dim \tilde{K} = \sup\{\dim \sigma \ : \ \sigma \in \tilde{K}\}.$$

Remark A.1.7. Observe that any geometric simplicial complex is an abstract simplicial complex by its combinatorial description. ♢

As we anticipated, to any finite abstract simplicial complex $\tilde{K}$ it is possible to associate a topological space $|\tilde{K}|$ as a subset of $\mathbb{R}^d$, where $d = \dim \tilde{K}$, as follows. We identify each $v_i \in V = \{v_0, \ldots, v_n\}$ with some $x_i \in \mathbb{R}^d$ such that $\{x_0, \ldots, x_n\} \subset \mathbb{R}^d$ are affine independent. Then, abstract simplexes $\sigma = \{v_{i_0}, \ldots, v_{i_k}\}$ are identified with the corresponding simplexes of $\mathbb{R}^d$, $[x_{i_0}, \ldots, x_{i_k}]$. The topological space is the union of all them.

Notice that, if K is a geometric complex whose combinatorial description is the same as $\tilde{K}$, then $|K|$ is homeomorphic to $|\tilde{K}|$. In this case K is called **geometric realization** of $\tilde{K}$. Isomorphic abstract simplicial complexes have homeomorphic geometric realizations. Thus, we will drop the adjective abstract/geometric from now on and only write simplicial complex.

Example A.1.8 (Nerve complex)**.** Given a topological space X and $\mathcal{U} = \{U_i\}_{i \in I}$ an open cover, an important example of abstract simplicial complex is the **nerve** of $\mathcal{U}$, denoted by $N(\mathcal{U})$. The set of vertices is $V = \mathcal{U}$ and $\sigma = \{U_{i_1}, \ldots, U_{i_l}\} \in N(\mathcal{U})$ if and only if $U_{i_1} \cap \cdots \cap U_{i_l} \neq \emptyset$. ♢

A.2 Simplicial homology

Homology identifies topological features on a given simplicial complex K: connected components, loops, cavities, etc. This is done by developing an algebraic structure on K that encodes this topological information. In order to do it we need to define an orientation on the simplicial complex.

Definition A.2.1. A simplicial complex K is **oriented** if there is a partial order on the set of vertices of K, whose restriction to the vertices of any simplex is a linear order.

On an oriented simplicial complex K we define the following equivalence relation between simplexes $\{p_0, \ldots, p_n\} \in K$:

1. $\{p_0, \ldots, p_n\} = 0$ if some vertex is repeated,
2. $\{p_0, \ldots, p_n\} = \operatorname{sgn}(\pi)\{p_{\pi 0}, \ldots, p_{\pi n}\}$ where π is a permutation of $\{0, 1, \ldots, n\}$.

We will denote each class by $[p_0, \ldots, p_n]$, where $\{p_0, \ldots, p_n\}$ is the element of the class such that $p_0 < p_1 < \cdots < p_n$. Let k be a field.

Definition A.2.2. If K is an oriented simplicial complex, we define the **space of n-chains** $C_n(K)$ with coefficients in k as the set whose elements are the formal (finite) sums of n-simpleces $\sigma = [p_0, \ldots, p_n]$ of K. That is, any n-chain can be written as $c = \sum_{i=1}^{k} n_i \sigma_i$, where σ_i are n-simplexes and $n_i \in k$.

Notice that $C_n(K) = 0$ for $n > \dim K$. A sum and a product can be defined on $C_n(K)$, making it a k-vector space.

Definition A.2.3. The **boundary** operator $\partial_n : C_n(K) \to C_{n-1}(K)$ is defined by setting

$$\partial_n([p_0, \dots, p_n]) = \sum_{i=0}^{n} (-1)^i [p_0, \dots, \hat{p}_i, \dots, p_n],$$

and extending by linearity, where $\hat{p}_i$ means to remove the vertex p_i.

It can be proven that, whether K is an oriented simplicial complex of dimension m, the following is a chain complex, i.e. $\partial \circ \partial = 0$,

$$0 \to C_m(K) \xrightarrow{\partial} \cdots \xrightarrow{\partial} C_1(K) \xrightarrow{\partial} C_0(K) \to 0.$$

Consider $Z_n(K) := \ker \partial_n$, which is the group of **simplicial n-cycles**, and $B_n(K) := \text{im}\partial_{n+1}$, the group of **simplicial n-boundaries**. Then, the following is well-defined

$$H_n(K) = \frac{Z_n(K)}{B_n(K)} = \frac{\ker \partial_n}{\text{im}\partial_{n+1}},$$

and it is called **n-th simplicial homology group**. Notice that only finitely many homology groups are nonzero and that any of them is a finite dimensional k-vector space (since $C_n(K)$ are finitely generated). The dimension of this vector space is called **n-th Betti number** of K,

$$\beta_n(K) = \dim H_n(K).$$

Homology theory can be extended to more general spaces. Namely, given X topological space, we can consider the continuous map $\sigma : [p_0, \dots, p_n] \to X$, where $[p_0, \dots, p_n]$ is a geometric realization of a oriented n-simplex. It is called singular simplex of dimension n. As before, it is possible to define the space of n-chains as the free abelian group generated by the singular simpleces of X. Therefore, the notions of boundary operator, cycles, boundaries and homology groups extend straightforward. In this case, $H_n(X)$ is called singular homology group.

Given any topological spaces X and Y, a continuous map $f : X \to Y$ induces a homomorphism between homology groups

$$H_n(f) : H_n(X) \to H_n(Y);\ H_n(f)([\sigma]) = [f \circ \sigma].$$

This map $f : X \to Y$ is a homotopy equivalence if there exists another continuous map $g : Y \to X$ such that $f \circ g$ is homotopic to id_Y and $g \circ f$ is homotopic to id_X. In that case we say that X and Y are homotopy equivalent. It can be proven that any homotopy equivalence induces an isomorphism $H_n(f)$ between singular homology groups.

Theorem A.2.4 ([31])**.** *Let K be an (oriented finite) simplicial complex. Then, the simplicial homology group $H_n(K)$ is isomorphic to the singular homology group $H_n(|K|)$.*

This implies that $H_n(K)$ is independent of the orientation defined on K. In addition, it let us talk indifferently about simplicial or singular homology.

Corollary A.2.5. *Let K, K' be two finite simplicial complexes such that $|K|$ and $|K'|$ are homotopy equivalent. Then, $H_n(K)$ is isomorphic to $H_n(K')$.*

When studying simplicial complexes it could be useful to subdivide each simplex into small pieces, so that each of them is totally contained in some required space. A way of doing this, while preserving the topological properties, is the barycenter subdivision. After the subdivision we still have a simplicial complex, with more simpleces than the original but defining the same topological space, and hence having the same homology.

Definition A.2.6. Let $\{p_0, \ldots, p_n\}$ be affine independent. Then, the **barycenter** of $[p_0, \ldots, p_n]$ is $\frac{1}{n+1}(p_0 + \cdots + p_n)$

If $\{p_0, \ldots, p_n\}$ is affine independent with barycenter b then $\{b, p_0, \ldots \hat{p}_i, \ldots, p_n\}$ is affine independent for each i. In the affine case $\mathbb{R}^d$, the computation of the barycenter subdivision can be done as we show below.

Definition A.2.7. Let σ_n be an affine n-simplex. Its barycenter subdivision, denoted by $\mathrm{Sd}(\sigma^n)$, is a family of affine n-simplexes defined inductively on $n \geq 0$ as follows:

1. $\mathrm{Sd}(\sigma^0) = \sigma^0$, i.e. is the identity on vertices;
2. if $\lambda_0, \lambda_1, \ldots, \lambda_{n+1}$ are the faces of dimension n of σ^{n+1} and b is the barycenter of σ^{n+1}, then $\mathrm{Sd}(\sigma^{n+1})$ cosists of all the $(n+1)$-simplexes spanned by b and the n-simplexes in $\mathrm{Sd}(\lambda_i)$, for each $i = 0, \ldots, n+1$.

Notice that each n-simplex of $\mathrm{Sd}(\sigma^n)$ comes equipped with an orientation induced by the original ones. This construction can be applied successively, $\mathrm{Sd}^m(\sigma^n)$.

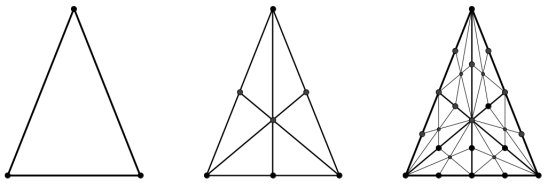


Figure A.2: Successive barycenter subdivision of a simplicial complex. From left to right, σ, $\mathrm{Sd}(\sigma)$ and $\mathrm{Sd}^2(\sigma)$.

In the general case of a simplicial complex we can proceed as follows. For any simplex σ let b^σ denote its barycenter

Definition A.2.8. Given K a simplicial complex, the barycenter subdivision $\mathrm{Sd}(K)$ is defined to be the simplicial complex with $\mathrm{Vert}(\mathrm{Sd}(K)) = \{b^\sigma \ : \ \sigma \in K\}$ and with simplexes $[b^{\sigma_0}, \ldots, b^{\sigma_n}]$, where the σ_i are simplexes in K with $\sigma_0 < \sigma_1 \ldots < \sigma_n$.

It can be proven that, for any simplicial complex, $|K| = |\mathrm{Sd}(K)|$. Therefore, K and $\mathrm{Sd}(K)$ have isomorphic homology groups.

Appendix B

Measure theory

Consider $\mathcal{X}$ a locally compact Polish metric space, i.e. a separable, complete, metric space. Let us introduce some basic notions in measure theory, that will be used throughout the text.

Definition B.0.1. A Radon measure μ supported on $\mathcal{X}$ is a Borel measure such that $\mu(K)$ is finite for every compact set K of $\mathcal{X}$. Denote by $\mathcal{M}(\mathcal{X})$ the space of all these measures.

The discrete case will be of special interest when studying persistence diagrams.

Definition B.0.2. A **point measure** on $\mathcal{X}$ is a Radon measure with discrete support and integer mass on each point, i.e. of the form $\nu = \sum_{x \in X} n_x \delta_x$, where $n_x \in \mathbb{N}$ and $X \subset \mathcal{X}$ is a locally finite[1] set. The space of all point measures on $\mathcal{X}$ is $\mathcal{D}(\mathcal{X})$.

An important notion in measure theory, that is going to be exploit in the last chapters, is that of pushforward measure.

Definition B.0.3. Let $\mathcal{X}$ and $\mathcal{Y}$ be two measurable spaces and μ a measure on $\mathcal{X}$. Consider $f : \mathcal{X} \to \mathcal{Y}$ a measurable map. Then, $f\#\mu(A) = \mu(f^{-1}(A))$, for any measurable set $A \subset \mathcal{Y}$, defines a measure $f\#\mu$ on $\mathcal{Y}$, called **pushforward measure**.

An important property of pushforwards is that a measurable map g on $\mathcal{Y}$ is integrable with respect to $f\#\mu$ if and only if $g \circ f$ is integrable with respect to μ. Moreover, it satisfies

$$\int_{\mathcal{Y}} g(y) d(f\#\mu)(y) = \int_{\mathcal{X}} g \circ f(x) d\mu(x).$$

The space of Radon measures may be equipped with the weak and the vague topology.

Definition B.0.4. The **weak topology** on $\mathcal{M}(\mathcal{X})$ is the coarsest topology such that, for any $f \in C_b(\mathcal{X})$ in the space of continuous and bounded functions, the following map is continuous

$$\varphi_f : \mathcal{M}(\mathcal{X}) \to \mathbb{R}; \qquad \mu \mapsto \int_{\mathcal{X}} f(x) d\mu(x).$$

[1] A collection of subsets of a topological space X is said to be locally finite if each point in the space has a neighborhood that intersects only finitely many of the sets in the collection.

Convergence in the weak topology will be denoted by $\mu_n \xrightarrow{w} \mu$. Observe that it is equivalent to the convergence of $\varphi_f(\mu_n)$ to $\varphi_f(\mu)$ for all $f \in C_b(\mathcal{X})$,

$$\mu_n \xrightarrow{w} \mu \Leftrightarrow \lim_{n\to\infty} \varphi_f(\mu_n) = \varphi_f(\mu) \quad \forall f \in C_b(\mathcal{X}). \tag{B.1}$$

Definition B.0.5. The **vague topology** on $\mathcal{M}(\mathcal{X})$ is the coarsest topology such that the maps

$$\varphi_f : \mathcal{M}(\mathcal{X}) \to \mathbb{R}; \qquad \mu \mapsto \int_{\mathcal{X}} f(x) d\mu(x)$$

are continuous for every $f \in C_c(\mathcal{X})$ continuous function with compact support in $\mathcal{X}$.

We say that a sequence of measures $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathcal{M}(\mathcal{X})$ converges vaguely to μ if they converge with respect to the vague topology. It will be denoted by $\mu_n \xrightarrow{v} \mu$. Likewise,

$$\mu_n \xrightarrow{v} \mu \Leftrightarrow \lim_{n\to\infty} \varphi_f(\mu_n) = \varphi_f(\mu) \quad \forall f \in C_c(\mathcal{X}). \tag{B.2}$$

In the following, we present some important results regarding the weak and vague topology.

Proposition B.0.6. *Consider measures $\mu, \mu_1, \mu_2, \ldots \in \mathcal{M}(\mathcal{X})$. Then, $\mu_n \xrightarrow{w} \mu$ if and only if $\mu_n \xrightarrow{v} \mu$ and $\mu_n(\mathcal{X}) \to \mu(\mathcal{X})$.*

Proposition B.0.7. *The set $\mathcal{D}(\mathcal{X})$ is closed in $\mathcal{M}(\mathcal{X})$ for the vague topology*

Before the next statements we need some definitions.

Definition B.0.8. A topological space X is said to be **sequentially compact** if every sequence has a converging subsequence to a point of X.

Definition B.0.9. A subspace Y of a topological space X is **relatively compact** if its closure is compact.

Proposition B.0.10. *A set $S \subset \mathcal{M}(\mathcal{X})$ is relatively compact for the vague topology if and only if $\sup\{\mu(K) \,:\, \mu \in S\}$ is finite for all compact sets $K \subset \mathcal{X}$.*

Theorem B.0.11 (Portmanteau Theorem)**.** *Consider a sequence of measures $\mu, \{\mu_n\} \in \mathcal{M}(\mathcal{X})$. Then, μ_n converges vaguely to μ if and only if one of the following is satisfied:*

- *for all open sets $U \subset \mathcal{X}$ and all bounded closed sets $F \subset \mathcal{X}$,*

$$\limsup_{n\to\infty} \mu_n(F) \le \mu(F) \qquad \text{and} \qquad \liminf_{n\to\infty} \mu_n(U) \ge \mu(U). \tag{B.3}$$

- *for all bounded Borel sets A with $\mu(\partial A) = 0$,*

$$\lim_{n\to\infty} \mu_n(A) = \mu(A). \tag{B.4}$$

Last but not least, we recall the Radon-Nikodym theorem.

Definition B.0.12. Given two non-negative Radon measures on $\mathcal{X}$, μ is said to be **absolutely continuous** with respect to ν if $\nu(A) = 0$ implies $\mu(A) = 0$ for any $A \subset \mathcal{X}$ Borel set. This is denoted by $\mu \ll \nu$.

Theorem B.0.13 (Radon-Nikodym Theorem)**.** *Consider μ, ν two non-negative Radon measures on $\mathcal{X}$. Assume $\mu \ll \nu$. Then, there exists a measurable non-negatice function $g : \mathcal{X} \to [0, \infty)$, unique ν-a.e, such that*

$$\mu(A) = \int_A g d\nu. \tag{B.5}$$

*This function is denoted by $\frac{d\mu}{d\nu}$ and is called the **Radon-Nikodym derivative** of μ with respect to ν.*

Bibliography

[1] Chazal, F. and Michel, B. (2017) An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists. *arXiv:1710.04019.*

[2] Ghrist, R. (2008) Barcodes: the persistent topology of data. *Bull. Amer. Math. Soc.* 45, 61-75.

[3] Ghrist, R. (2018). Homological algebra and data.

[4] Chazal, F., de Silva, V., Glisse, M., Oudot, S. (2016). *The Structure and Stability of Persistence Modules.* SpringerBriefs in Mathematics. Springer.

[5] Figalli, A. and Gigli, N. (2010). 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.

[6] Lacombe, T. (2020). Statistics for Topological Descriptors using optimal transport. *Metric Geometry [math.MG]. Institut Polytechnique de Paris.* English. NNT : 2020IPPAX036.

[7] Edelsbrunner, H., Letscher, D. and Zomorodian, A. (2002). Topological persistence and simplification. *Discrete Comput. Geom.*, 28:4, 511-533.

[8] Zomorodian, A. and Carlsson, G. (2005). Computing Persistent Homology. *Discrete Comput. Geom.*, 33, 249–274.

[9] Boissonnat, J-D. and Maria, C. (2014). The Simplex Tree: an Efficient Data Structure for General Simplicial Complexes. *Algorithmica, Springer Verlag*, 70 (3), pp.20. 10.1007/s00453-014-9887- 3ff. ffhal-00707901v3f

[10] Zomorodian, A. (2010). The tidy set: a minimal simplicial set for computing homology of clique complexes. In *SoCG*, pages 257–266, 1.

[11] Singh, G., Mémoli, F., and Carlsson, G. E. (2007). Topological methods for the analysis of high dimensional data sets and 3d object recognition. *In SPBG*, pages 91–100.

[12] Hatcher, A. 1(2001). *Algebraic Topology.* Cambridge Univ. Press.

[13] Chazal, F. and Oudot, S. Y. (2008). Towards persistence-based reconstruction in euclidean spaces. In *Proceedings of the twenty-fourth annual symposium on Computational geometry*, SCG '08, pages 232–241, New York, NY, USA. ACM.

[14] Cannarsa, P. and Sinestrari, C. (2004). *Semiconcave functions, Hamilton–Jacobi equations, and optimal control*. Progress in Nonlinear Differential Equations and THeir Applications. Birkhäuser.

[15] Chazal, F., Cohen-Steiner, D., Lieutier, A. (2009). A sampling theory for compact sets in Euclidean space. *Discrete Comput. Geom.* 41(3), 461-479.

[16] Grove, K. (1993). Critical point theory for distance functions. In *Proc. of Symposia in Pure Mathematics*, volume 54.

[17] Chazal, F., Cohen-Steiner, D., and Mérigot, Q. (2011). Geometric inference for probability measures. *Foundations of Computational Mathematics*, 11(6):733-751.

[18] Cohen-Steiner, D., Edelsbrunner, H., and Harer., J. (2005). Stability of persistence diagrams. In *Proc. 21st ACM Sympos. Comput. Geom.*, pages 263-271.

[19] Auslander, M. (1974). Representation theory of Artin algebras. II. *Commun. Algebr.* 1(4), 269-310.

[20] Crawley-Boevey, W. (2015). Decomposition of pointwise finite-dimensional persistence modules. *J. Algeb. Appl.* 14(5).

[21] Azumaya, G. (1950). Corrections and supplementaries to my paper concerning Krull-Remak-Schmidt's theorem. *Nagoya Math. J.* 1, 117-124.

[22] Chazal, F., Cohen-Steiner, D., Guibas, L. J., Mémoli, F., and Oudot, S. Y. (2009). Gromov-hausdorff stable signatures for shapes using persistence. *Computer Graphics Forum (proc.SGP 2009)*, pages 1393-1403.

[23] Cohen-Steiner, D., Edelsbrunner, H., Harer, J., Mileyko, Y., (2010). Lipschitz Functions Have L_p-Stable Persistence. *Foundations of Computational Mathematics*, 10: 127-139.

[24] Mileyko, Y., Mukherjee, S. and Harer, J. (2011). Probability measures on the space of persistence diagrams. *Inverse Problems,* 27(12):124007.

[25] Villani, C. (2008). *Optimal transport: old and new*, volume 338. Springer Science & Business Media.

[26] Bubenik, P. (2015). Statistical Topological Data Analysis using Persistence Landscapes. *Journal of Machine Learning Research*, 16:77-102.

[27] Nielsen, L. (2011). Weak convergence and banach sapce-valued functions:improving the stability theory of feynman's operational calculi. *Mathematical Physics, Analysis and Geometry*, 14(4):279.

[28] Hofer, C., Kwitt, R., Niethammer, M., and Uhl, A. (2017). Deep learning with topological signatures. *arXiv preprint arXiv:1707.04041*.

[29] Mérigot, Q., Delalande, A., Frédéric, C. (2019). Quantitative Stability of Optimal Transport Maps and Linearization of the 2-Wasserstein Space. *[stat.ML] arXiv:1910.05954*

[30] Delalande, A., Mérigot, Q. (2021). Quantitative stability of optimal transport maps under variations of the target measure. *[math.FA] arXiv:2103.05934*

[31] Rotman, J.J. (1998). *An Introduction to Algebraic Topology.* Springer.

Index