On C^0-persistent homology and trees

Daniel Perez*[1,2,3]

[1]Département de mathématiques et applications, École normale supérieure, CNRS, PSL University, 75005 Paris, France
[2]Laboratoire de mathématiques d'Orsay, Université Paris-Saclay, CNRS, 91405 Orsay, France
[3]DataShape, Centre Inria Saclay, 91120 Palaiseau, France

November 23, 2022

Abstract

In this paper we revisit and extend some classical results on persistent homology. We start by extending the notion of merge trees to all continuous functions on some general topological spaces. We revisit the concept of homological dimension, previously introduced by other authors and show that the suprema in the definitions of these concepts is attained generically in the sense of Baire. We then generalize the Wasserstein stability theorem to irregular settings, giving explicit bounds on the constants in the theorem and sharp bounds on its regime of validity. Finally, we use this generalized Wasserstein stability theorem to show a stochastic stability theorem for persistence diagrams.

Keywords— Persistent homology, continuous functions, barcodes, persistence diagrams, trees, Wasserstein stability, persistence index, homological dimension, stochastic processes
MSC2020 Classification— 55N31, 55M10, 62R40, 54C50

Contents

*Email: daniel.perez@ens.fr

1 Introduction

1.1 State of the art

The topology of superlevel sets of a function has been a subject of study in different mathematical communities. In the probability theory, the introduction of trees lead to the understanding of connected components of superlevel sets (called *excursions* in the probabilistic setting) of (irregular) random functions on $[0, 1]$ [22, 28, 29, 51].

More recently, so-called *merge trees* have also appeared in topological data analysis (TDA) (*cf.* the books by Chazal *et al.* [14] and Oudot's book [48] for an introduction to TDA). As in the probabilistic case, these trees carry important information about the connected components of superlevel sets and moreover about the 0th degree homology persistence diagram of a Morse function f defined on a compact manifold X [23, 24, 43, 47, 58] (an explicit construction and correspondence between trees and barcodes can be found in [23]).

The construction of these trees is different in both cases: the approach of the probabilists is analytic [22, 28], whereas merge trees are more algebraic in nature [23, 24, 47]. Since these trees capture essentially the same information about the connected components of superlevel sets, one can ask whether both constructions coincide where their regimes of validity intersect.

Parallel to this development, Wasserstein p distances on the space of diagrams (denoted d_p) [30, Chapter VIII.2] have been widely used and studied by the TDA community in different contexts [12, 21, 25, 46, 57]. Recently, Wasserstein distances have been formalized through the use of optimal *partial* transport by Divol and Lacombe [25]. In this formalism, persistence diagrams are regarded as measures and use optimal transport theory to extend the notion of Wasserstein p distances, previously defined on persistence diagrams, to Radon measures on the upper-half plane $\mathcal{X} \subset \mathbb{R}^2$. It is to be noted that regarding persistence diagrams as measures is a point of view which had already been introduced [14, 48] and which has proved fruitful independently from the considerations regarding Wasserstein distances.

The extension to all Radon measures comes with certain advantages, such as having an easily definable and computable notion of "average diagram", defined by duality. This notion was originally introduced by Chazal and Divol in [15] as follows. If f is a random function, seeing $\mathrm{Dgm}(f)$ as a measure, it is possible to define the average diagram of the process by duality in the following way. For every measurable set $B \subset \mathcal{X}$,

$$\mathbb{E}[\mathrm{Dgm}(f)]\,(B) := \mathbb{E}[\mathrm{Dgm}(f)(B)]\ . \tag{1.1}$$

From the definition, $\mathbb{E}[\mathrm{Dgm}(f)]$ encodes every linear functional of the diagram and is easily computed, motivating its introduction. This definition contrasts the Fréchet means approach of other authors (*e.g.* Turner *et al.* [57]), which is non-linear, depends on p and requires a proof of existence and unicity, but does not require the extension of the space of persistence diagrams to the space of arbitrary measures on $\mathcal{X}$.

This dual approach of Chazal and Divol inscribes itself in the more general context of the study of the persistence diagrams of stochastic processes, which have been studied by a wide variety of authors, for instance in [1, 3, 4, 7, 15–17, 49, 57]. Some of the previously cited results discuss different aspects of random field persistence theory, which include, but are not limited to, computations for canonical processes [7, 49], stability of certain linear functionals with respect to the bottleneck distance [16], the Euler characteristic [3], random complexes [3] and notions of central tendency [15, 57].

Given the widespread use of Wasserstein p distances, it is important to understand whether this notion is stable with respect to perturbations to filtration functions on X. This so-called "Wasserstein stability" of persistence diagrams of functions $f : X \to \mathbb{R}$ has been widely discussed by the TDA community in the context where the space X is triangulable. There are many results in this direction [18, 21, 55], valid with different degrees of generality, covering both X compact [21, 55] and X non-compact [18], but mainly focusing on Lipschitz functions (however, the work of Chen and Edelsbrunner [18] does not require the Lipschitz condition). The first result in this direction was obtained by Cohen-Steiner *et al.* [21] and depends on the following *ad-hoc* condition on X.

Definition 1.1. [21] A (triangulable) metric space X **implies bounded q-total persistence** if, for all $k \in \mathbb{N}$, there exists a constant C_X that depends only on X such that

$$\mathrm{Pers}_q^q(\mathrm{Dgm}_k(f)) < C_X \tag{1.2}$$

for every tame function f with Lipschitz constant $\mathrm{Lip}(f) \leq 1$.

The Pers_p-functional of the definition above is the usual p-persistence used in TDA (a non-exhaustive list of uses of this functional includes [2, 12, 25, 46, 57]) and is defined as the ℓ^p-norm of the length of the bars of the barcode of f. The results obtained thereafter rely heavily on this condition, which is not rendered quantitative (in particular, given X, no upper bound for C_X or lower bound for q had been established in general). Nonetheless, this condition allowed the authors to show a Wasserstein stability result.

Theorem 1.2 (Cohen-Steiner, Edelsbrunner, Harer, [21])**.** *Let X be a triangulable space implying bounded q-total persistence and let f and g be two $\mathbb{R}$-valued Lipschitz functions on X. Then, for all $p > q$, we have*

$$d_p(\mathrm{Dgm}(f), \mathrm{Dgm}(g)) \leq C_X (\mathrm{Lip}(f) \vee \mathrm{Lip}(g))^{\frac{q}{p}} \|f - g\|_\infty^{1-\frac{q}{p}} , \tag{1.3}$$

where $\mathrm{Lip}(f)$ *denotes the Lipschitz constant of f and* $a \vee b := \max\{a, b\}$.

Further results in this direction, such as the ones in [55], also rely on the bounded q-total persistence condition, but give lower bounds on admissible q, finding that $q \geq d$, where d is the maximal dimension of simplices in the triangulation of X. It is also known that, for distance functions to point clouds in $\mathbb{R}^d$, $q = d$.

We will later see that the lower bound for the validity of Wasserstein stability is closely related to a different question regarding the link between the so-called *homological dimensions* of X and the upper-box dimension of X, which we will denote $\overline{\dim}(X)$ (analogously, we will denote $\underline{\dim}(X)$ the lower-box dimension). To the best knowledge of the author, although Yuliy Baryshnikov and Shmuel Weinberger had previously obtained results in this direction (but never published them), this question was first opened and studied by Schweinhart and MacPherson [41] and later studied in more detail by Schweinhart in [54], but has also been addressed by other authors (*cf.* [1] and the references therein).

Definition 1.3 (Schweinhart's definition of $\dim_{\mathrm{PH}}^k$, [54])**.** Let X be a bounded subset of a metric space. The PH_k-dimension of X is

$$\dim_{\mathrm{PH}}^k(X) := \inf_p \{\sup_{\mathbf{x}} \mathrm{Pers}_p(\mathrm{Dgm}_k(d(-, \mathbf{x}))) < \infty\} , \tag{1.4}$$

where the supremum is taken over all finite sets of points $\mathbf{x}$ of X.

There are open problems stated in Schweinhart's paper regarding the relation between these notions of dimension and $\overline{\dim}(X)$, some of which we will give a partial answer to in this paper.

1.2 Our contribution

This paper mainly extends previously known results about the persistence theory of filtrating functions f to more irregular settings. More precisely, we focus on relaxing the assumptions on not only the regularity of the underlying functions f, but also on the nature of the compact metric space X over which they are defined. The main contributions of this paper are mostly contained under the scope of the four following theorems, which are not stated in their full generality here for the sake of brevity and clarity, but which contain the general ideas hereby explored. We kindly refer the reader to the corresponding theorems for the full generality of the statements.

Theorem 1 (Merge tree extension to C^0-functions (section 2.1 and theorem 2.20))**.** Let X be a compact, connected, locally path connected topological space and k be a field. There exists a map $T : C^0(X,\mathbb{R}) \to \mathbf{Tree}$ associating $f \mapsto T_f$ and a functor $\mathrm{Alg} : \mathbf{Tree} \to \mathbf{PersMod}_k$ such that

$$\mathrm{Alg}(T_f) = H_0(X,f)\,, \tag{1.5}$$

where **Tree** is the category of rooted $\mathbb{R}$-trees seen as metric spaces, whose morphisms are isometric embeddings preserving the roots and where $\mathbf{PersMod_k}$ is the observable category of q-tame persistence modules over a field k. ■

This result extends the results of Curry linking merge trees and persistence barcodes [23], by extending the regime of validity of the theory of merge trees from a Morse setting to a C^0 one. The theorem draws its inspiration from the constructions previously made by Le Gall and Curien [22, 28]. We subsequently show that the map assigning a continuous function $f : [0,1] \to \mathbb{R}$ to its constructed tree T_f is a surjection onto the space of trees of finite upper-box dimension and provide an explicit construction of an inverse image (section 2.3). The latter is quite technical, and may be skipped in a first reading.

Following Picard [51] and Schweinhart [54], we introduce the so-called *persistence index of degree* k, $\mathcal{L}_k(f)$, of a function $f : X \to \mathbb{R}$ (definition 3.3) as

$$\mathcal{L}_k(f) := \inf\{p \,|\, \mathrm{Pers}_p(H_k(X,f)) < \infty\}\,, \tag{1.6}$$

and show the following.

Theorem 2 (Generic saturation of persistence indices (theorems 3.19 and 3.23))**.** Let X be a compact Riemannian manifold of dimension d and let $f \in C^\alpha(X,\mathbb{R})$. Then, for any $0 \le k < d$,

$$\mathcal{L}_k(f) \le \frac{d}{\alpha}\,. \tag{1.7}$$

Moreover, this bound is saturated generically in the sense of Baire within $C^\alpha(X,\mathbb{R})$. ■

Modifying Schweinhart's definition for the kth degree homological dimension of X (definition 3.32), we define

$$\dim^k_{\mathrm{PH}}(X) := \sup_{f \in \mathrm{Lip}_1(X)} \mathcal{L}_k(f)\,. \tag{1.8}$$

The generic saturation theorem shows that the supremum over all 1-Lipschitz functions in the definition of $\dim^k_{\mathrm{PH}}$ can be replaced by a supremum over the class of α-Hölder class with bounded Hölder constant for any α, up to a factor of α. In so doing, we answer a question by Schweinhart [54] regarding bounds on homological dimensions and regularity conditions on X for this bound to be sharp (section 3.3).

We give a quantitative version of the Wasserstein stability theorem valid for all degrees of Čech homology on regular enough metric spaces (which in particular include compact smooth manifolds) and for a wider class of regularity than what was previously considered.

Theorem 3 (General Wasserstein Stability (theorem 4.13))**.** Let X be a compact Riemannian manifold of dimension d, then for every $0 \leq k < d$, and all $p > q > \frac{d}{\alpha}$,

$$d_p^p(H_k(X,f), H_k(X,g)) \leq C_{X,\alpha,k}(\|f\|_{C^\alpha}^q + \|g\|_{C^\alpha}^q)\,\|f-g\|_\infty^{p-q}\;, \tag{1.9}$$

with

$$C_{X,\alpha,k} \leq 4^q(M^{k+1} - M^k)\alpha q \int_0^{\mathrm{diam}(X)} \varepsilon^{\alpha q - 1}(\mathcal{N}_X(\varepsilon) \vee \mathcal{N}_X(r_C))\, d\varepsilon\,, \tag{1.10}$$

where $\mathcal{N}_X(\varepsilon)$ is the minimal covering number of X by balls of radius ε, r_C denotes the radius of convexity of X and M is the doubling constant of X. ■

Notice that we retrieve the usual stability theorem with respect to the bottleneck distance d_∞ by taking $p \to \infty$ in the above expression. We also show an annex result of stability for the trees constructed from the functions f in terms of the Gromov-Hausdorff distance between the trees (theorem 4.21).

Finally, we discuss some consequences of these results to the stochastic setting (section 5) and prove Chazal *et al.*-like results [17] for the d_p-stability of average diagrams of stochastic processes with an *a priori* hypothesis of regularity, which can in particular be used to infer distances between distributions of diagrams of close stochastic processes.

Theorem 4 (Stochastic Wasserstein Stability (theorem 5.9))**.** Let f and g be two $\mathbb{R}$-valued a.s. C^α stochastic processes on a d-dimensional compact Riemannian manifold X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, for any $0 \leq k < d$, every $\frac{d}{\alpha} < q < p < \infty$ and any $r, s \in \,]1, \infty[$ satisfying $\frac{1}{r} + \frac{1}{s} = 1$ and $(p-q)s \geq 1$, there exists a constant C_X depending only on X such that

$$d_p(\mathbb{E}[\mathrm{Dgm}_k(f)]\,, \mathbb{E}[\mathrm{Dgm}_k(g)]) \leq W_{p,d_p}((\mathrm{Dgm}_k \circ f)_\sharp\mathbb{P}, (\mathrm{Dgm}_k \circ g)_\sharp\mathbb{P}) \tag{1.11}$$

$$\leq C_X \left[\mathbb{E}\left[\|f\|_{C^\alpha}^{qr}\right]^{\frac{1}{r}} + \mathbb{E}\left[\|g\|_{C^\alpha}^{qr}\right]^{\frac{1}{r}}\right]^{\frac{1}{p}} W_{(p-q)s,\infty}^{1-\frac{q}{p}}(f_\sharp\mathbb{P}, g_\sharp\mathbb{P}) \tag{1.12}$$

$$\leq C_X \left[\mathbb{E}\left[\|f\|_{C^\alpha}^{qr}\right]^{\frac{1}{r}} + \mathbb{E}\left[\|g\|_{C^\alpha}^{qr}\right]^{\frac{1}{r}}\right]^{\frac{1}{p}} \|f-g\|_{L^{(p-q)s}(\Omega, L^\infty(X,\mathbb{R}))}^{1-\frac{q}{p}}\,. \tag{1.13}$$

■

2 Barcodes, diagrams and trees

2.1 Trees stemming from a continuous function

Unless otherwise specified, throughout this section, let X denote a connected, locally path-connected, compact topological space and let $f : X \to \mathbb{R}$ be a continuous function. Let us denote $(X_r)_{r\in\mathbb{R}}$ the filtration of X by the **superlevels** of f, that is

$$X_r := \{x \in X \mid f(x) \geq r\}\,. \tag{2.14}$$

Notation 2.1. We will denote the open superlevel sets by $X_{>r}$ whenever necessary and denote X_r^z the connected component of X_r containing z.

There exists a pseudo-distance on X, denoted d_f, given by:

Definition 2.2. Let X and f be defined as above. The H_0**-distance**, d_f, is the pseudo-distance

$$d_f(x,y) := f(x) + f(y) - 2 \sup_{\gamma : x \mapsto y} \inf_{t\in[0,1]} f(\gamma(t))\,, \tag{2.15}$$

where the supremum runs over every path γ linking x to y.

Remark 2.3. Notice there are different ways of writing this distance. In particular, the sup above is also characterized by

$$\sup_{\gamma:x\mapsto y}\inf_{t\in[0,1]} f(\gamma(t)) = \sup\{r \mid [x]_{H_0(X_r)} = [y]_{H_0(X_r)}\} \tag{2.16}$$

$$= \sup\{r \mid \exists\gamma \in C_1(X_r) \text{ such that } \partial\gamma = x - y\}\,. \tag{2.17}$$

These equalities hold, since we take the coefficients of homology with respect to $\mathbb{Z}/2\mathbb{Z}$, so we can interpret 1-chains as sums of paths on X.

This pseudo-distance is a generalization of the distance introduced by Curien, Le Gall and Miermont in [22]. d_f has the following properties:

- **(P1) Identification of the connected components of superlevel sets**: $d_f(x, y) = 0$ if and only if there exists $t \in \mathbb{R}$ such that $x, y \in \{f = t\}$ and for every $\varepsilon > 0$, x and y lie in the same connected component of $X_{>t-\varepsilon}$;
- **(P2) Compatibility with the filtration induced by** f: Let $x, y \in X$ and suppose that $f(x) < f(y)$, then if $[x]_{H_0(X_{f(x)})} = [y]_{H_0(X_{f(x)})}$,

$$d_f(x, y) := |f(x) - f(y)|\,. \tag{2.18}$$

(P2) is immediate from the definition of d_f. It remains to show the two following propositions.

Proposition 2.4. The function $d_f : X^2 \to \mathbb{R}^+$ of definition 2.2 is a pseudo-distance.

Proof. Checking symmetry and positivity is easy. The only non-obvious point is that the triangle inequality is satisfied by this expression. Let $x, y, z \in X$ and denote

$$[x \mapsto y] := \sup_{\gamma:x\mapsto y}\inf_{t\in[0,1]} f\circ\gamma(t)\,. \tag{2.19}$$

It suffices to show the following inequality

$$[x \mapsto z] + [z \mapsto y] \le [x \mapsto y] + f(z)\,. \tag{2.20}$$

Let γ be a path from x to z and η be a path from z to y and let $\gamma * \eta$ be the concatenation of these two paths. By definition,

$$\inf_{t\in[0,1]} f\circ(\gamma*\eta)(t) \le [x \mapsto y]\,, \tag{2.21}$$

from which it follows that

$$[x \mapsto z] \wedge [z \mapsto y] \le [x \mapsto y]\,. \tag{2.22}$$

Without loss of generality, suppose that $[x \mapsto z]$ achieves the above minimum and note that

$$[z \mapsto y] \le f(z) \tag{2.23}$$

by definition of $[z \mapsto y]$. Adding the two last inequalities together,

$$[x \mapsto z] + [z \mapsto y] \le [x \mapsto y] + f(z)\,, \tag{2.24}$$

as desired. ∎

Proposition 2.5. Let f be a continuous function as above, then **(P1)** holds.

Proof. The ($\Leftarrow$) direction is immediate, so let us show ($\Rightarrow$).

Suppose that $d_f(x,y) = 0$ and that $f(x) \neq f(y)$, then,

$$\sup_{\gamma:x\mapsto y} \inf_{t\in[0,1]} f(\gamma(t)) = \frac{f(x)+f(y)}{2} > f(x) \wedge f(y)\,. \tag{2.25}$$

However,

$$\sup_{\gamma:x\mapsto y} \inf_{t\in[0,1]} f(\gamma(t)) \leq f(x) \wedge f(y)\,, \tag{2.26}$$

which leads to a contradiction, so $f(x) = f(y)$. The condition $d_f(x,y) = 0$ becomes:

$$f(x) = \sup_{\gamma:x\mapsto y} \inf_{t\in[0,1]} f(\gamma(t))\,. \tag{2.27}$$

This is only possible if for every $\varepsilon > 0$ there is a path γ lying entirely in $X^x_{>f(x)-\varepsilon}$, so

$$x, y \in \bigcap_{\varepsilon>0} X^x_{>f(x)-\varepsilon} \tag{2.28}$$

finishing the proof. ■

With these technicalities out of the way, let us consider the metric space

$$(T_f, d_f) := (X/\{d_f = 0\}, d_f)\,, \tag{2.29}$$

where $X/\{d_f = 0\}$ denotes the quotient of X where we identify all points x and y on X satisfying $d_f(x,y) = 0$. Slightly abusing the notation, let d_f denote the distance induced on T_f by the pseudo-distance d_f on X.

The metric structure of T_f turns out to be simple, as T_f is an $\mathbb{R}$-tree. Let us briefly recall the definition of an $\mathbb{R}$-tree.

Definition 2.6 (Chiswell, [19])**.** An **$\mathbb{R}$-tree** (T,d) is a connected metric space such that any of the following equivalent conditions hold:

- T is a geodesic connected metric space and there is no subset of T which is homeomorphic to the circle, $\mathbb{S}_1$;
- T is a geodesic connected metric space and the Gromov 4-point condition holds, *i.e.* :

$$\forall x,y,z,t \in T \quad d(x,y) + d(z,t) \leq \max\left[d(x,z) + d(y,t), d(x,t) + d(y,z)\right]\,;$$

- T is a geodesic connected 0-hyperbolic space.

A **rooted $\mathbb{R}$-tree** (T,O,d) is an $\mathbb{R}$-tree along with a marked point $O \in T$.

A first important remark is that since X is connected, so is T_f. To show T_f is an $\mathbb{R}$-tree, we will use the first characterization of the definition above and show both conditions, *i.e.* that there are no subspaces of T_f which are homeomorphic to $\mathbb{S}_1$ and that T_f is in fact a geodesic metric space, to be satisfied separately.

Before showing this, it is helpful to introduce some notation.

Notation 2.7. Let $\pi_f : X \to T_f$ denote the canonical projection onto T_f and let O denote the root of T_f (*i.e.* $f(O) = \min f$), let us define the following quantity

$$\ell(\tau) := \inf_X f + d_f(O,\tau)\,, \tag{2.30}$$

where $X^\tau_{f(\tau)}$ denotes the connected component of the superlevel set $X_{f(\tau)}$ containing a preimage of τ.

Remark 2.8. These objects are well-defined by definition of d_f and render π_f continuous.

Definition 2.9. The **pseudo-distance topology on** X or the **topology of** d_f is the topology on X generated by the open balls:

$$B(x,r) := \{z \in X \mid d_f(x,z) < r\} \tag{2.31}$$

Despite the fact that the pseudo-distance topology is not in general Hausdorff, it is nonetheless fine enough to be useful, as shown by the two following technical lemmas.

Lemma 2.10. $X_{>r}$ has the same connected components for the topology of d_f on X and the usual topology of X.

Proof. By continuity of f, $X_{>r}$ is open in X for both topologies. For the usual topology, it is trivial. For the topology of d_f it is the complement in X of the closed ball $\overline{B(p, r - \inf f)}$, where p is a point on X achieving the infimum of f, which exists by compactness of X (in the usual topology).

Let Y now denote a connected component of $X_{>r}$ for the usual topology. The set Y is connected for the topology of d_f. Otherwise, we could write $Y = U \sqcup V$ for some open sets U and V, but since open sets of the topology of d_f are also open for the usual topology, this leads to a contradiction, as we assumed Y was connected for the usual topology. We will now show that Y is both open and closed in $X_{>r}$ for the topology of d_f. Y is open, since it can be written as the union of open balls

$$Y = \bigcup_{y \in Y} B(y, f(y) - r)\,. \tag{2.32}$$

Additionally, Y is closed since its complement is open, as it can be similarly written as the union of open balls. It follows that Y is also a connected component of $X_{>r}$ for the topology of d_f.

Now, suppose that Y is a connected component of $X_{>r}$ for the topology of d_f. Any ball of the covering above is path connected, but since Y is connected, this implies that Y is path connected (and the paths are completely included within Y), it is thus a path connected component of $X_{>r}$. Since X is connected and locally path connected for the usual topology, Y is a path connected component of the usual topology, rendering it a connected component for the usual topology. ■

Lemma 2.11. Denote $T_{>r}$ the open superlevel set on T_f. For the topology of d_f, π_f induces a bijective correspondence between the connected components of $X_{>r}$ and those of $T_{>r}$.

Proof. Since π_f is surjective and is both open and closed for the topology of d_f (on both X and T_f), lemma 2.10 implies that the map π_f surjectively sends the connected components of $X_{>r}$ onto connected components of $T_{>r}$, since the connected components of $X_{>r}$ for the topology of d_f and the usual topology of X are the same.

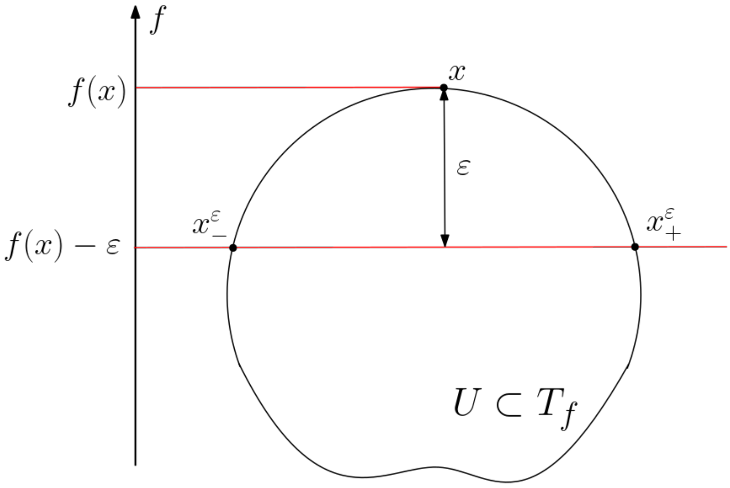
It remains to show the injectivity. Note that π_f is open and closed for the topology of d_f on X. The connected components of $X_{>r}$ are either disjoint or equal and, in fact, so are the images by π_f of these connected components. Otherwise, there exists some $\tau \in T_{>r}$ such that there is a preimage of τ lying in two different connected components of $X_{>r}$, which is impossible, as every preimage of τ must lie in the same connected component of $X_{>r}$ in accordance to proposition 2.5. This is equivalent to stating that if Y and Z are two connected components of $X_{>r}$ and $Y \neq Z$, then $\pi_f(Y) \cap \pi_f(Z) = \emptyset$, in particular $\pi_f(Y) \neq \pi_f(Z)$. Symbolically,

$$Y \neq Z \Rightarrow \pi_f(Y) \neq \pi_f(Z)\,, \tag{2.33}$$

which is the contrapositive of the statement of injectivity. ■

From the above lemmata, we get the following proposition.

Proposition 2.12. The metric space $T_f := X/\{d_f = 0\}$ equipped with distance d_f possesses no subspace homeomorphic to $\mathbb{S}_1$.



Proof. We will reason by contradiction. Suppose that T_f contains $U \subset T_f$ such that U is homeomorphic to the circle, $\mathbb{S}_1$. The function f descends to a function on T_f which is not locally constant anywhere by definition of d_f and in particular not locally constant anywhere on U, as the level-sets of f in T are totally discontinuous.

It follows that there exists an element $x \in U$ such that the maximum of f on U is attained at x. For $\varepsilon > 0$ small enough, there are two distinct points x_-^ε and x_+^ε such that $f(x_+^\varepsilon) = f(x) - \varepsilon = f(x_-^\varepsilon)$. Without loss of generality, we pick these points to be the closest ones to x along an arbitrary parametrization of U where this equality occurs. Since U is homeomorphic to $\mathbb{S}_1$, there is a path γ linking x_+ and x_- lying entirely above $f(x)-\varepsilon$ (equal at the endpoints) and passing through x. The image of γ in $T_{>f(x)-\varepsilon}$ is contained within one and only one connected component of $T_{>f(x)-\varepsilon}$, which we will denote S. By lemma 2.11, S corresponds to a unique connected component of $X_{>f(x)-\varepsilon}$ with respect to the topology of d_f, which we will denote X^S. By lemma 2.10, X^S is a connected component of $X_{>f(x)-\varepsilon}$ for the usual topology.

For every $0 < \varepsilon' < \varepsilon$, we can pick points $x_\pm^{\varepsilon'}$ on U. The connected component S contains $x_\pm^{\varepsilon'}$ for every such ε' and since inverse images of these two points are connected in $X_{>f(x)-\varepsilon}$ by a path $\gamma : x_+^{\varepsilon'} \mapsto x_-^{\varepsilon'}$,

$$d_f(x_+^{\varepsilon'}, x_-^{\varepsilon'}) < 2(f(x) - \varepsilon') - 2 \inf_{t\in[0,1]} f \circ \gamma < 2(\varepsilon - \varepsilon') \tag{2.34}$$

Letting $x_\pm^{\varepsilon'} \to x_\pm^\varepsilon$ in U as $\varepsilon' \to \varepsilon$, we have that $d_f(x_-^\varepsilon, x_+^\varepsilon) = 0$, leading to a contradiction, since we supposed that x_+^ε and x_-^ε were disjoint in T_f (and therefore not a distance zero away from one-another). ∎

Proposition 2.13. The metric space (T_f, d_f) is a rooted $\mathbb{R}$-tree whose root is the unique point point O in the image in T_f of a point $x \in X$ for which the function f is minimal.

Proof. The only thing left to show is that T_f is a geodesic space. As before, f descends to the quotient and induces a non-locally constant function on T_f. Let $x, y \in X$, if $f(x) = f(y)$ and x and y are in the same path connected component of $X_{>f(x)-\varepsilon}$ for all $\varepsilon > 0$ and there is nothing to show.

Suppose that $f(x) < f(y)$ and that x and y are in the same path connected component of $X_{f(x)}$ and consider a path in $X_{f(x)}$ going from y to x, $\gamma : [0,1] \to X$. The path γ can be modified into a path

$$\tilde{\gamma}(t) := \pi_f\left(\gamma\left(\arg\min_{s\in[0,t]} f \circ \gamma(s)\right)\right) . \tag{2.35}$$

On this modified path f is decreasing implying that $\tilde{\gamma}$ does not self-intersect, although it may be locally constant. The length of $\tilde{\gamma}$ is defined as

$$L(\tilde{\gamma}) = \sup_{(t_i)} \sum_{(t_i)} d_f(\tilde{\gamma}(t_{i+1}), \tilde{\gamma}(t_i)) , \tag{2.36}$$

where the supremum is taken over all finite partitions of $[0, 1]$. For any finite partition, this sum is always bounded by $f(y) - f(x)$, since along $\tilde{\gamma}$

$$f(\tilde{\gamma}(t_i)) \geq f(\tilde{\gamma}(t_{i+1})) \implies d_f(\tilde{\gamma}(t_{i+1}), \tilde{\gamma}(t_i)) = f(\tilde{\gamma}(t_i)) - f(\tilde{\gamma}(t_{i+1})) \tag{2.37}$$

by monotonicity of f along $\tilde{\gamma}$. This leads to pairwise cancellation of terms in the sum of equation 2.36. And so,

$$L(\tilde{\gamma}) = d_f(x, y) . \tag{2.38}$$

Now, suppose that x and y are two points on X, such that $f(x) \leq f(y)$ but such that x and y no longer lie in the same path connected component of $X_{f(x)}$ and pick a maximizer γ of the supremum (*cf.* remark 2.14)

$$\sup_{\gamma: x \mapsto y} \inf_{t \in [0,1]} f \circ \gamma(t) . \tag{2.39}$$

Since y is not connected to x in $X_{f(x)}$, by continuity of f, the path γ must eventually go under the level $f(x)$. Let us set

$$t^* := \sup \left\{ \arg\min_{s \in [0,1]} f \circ \gamma(s) \right\} \tag{2.40}$$

and note that $f(\gamma(t^*)) < f(x)$.

On $[0, t^*]$, the path γ lies entirely in $X_{f(\gamma(t^*))}$ and similarly, entirely in $X_{f(\gamma(t^*))}$ on $[t^*, 1]$. On $[0, t^*]$, we can define a modification of γ, $\tilde{\gamma} : [0, t^*] \to T_f$ by

$$\tilde{\gamma}(t) := \pi_f \left(\gamma \left(\arg\min_{s \in [0,t]} f \circ \gamma(s) \right) \right) . \tag{2.41}$$

Analogously, if we define $\eta(t) := \gamma(1 - t)$ – the reversed version of γ – it is possible to define a modification of η, $\tilde{\eta} : [0, 1 - t^*] \to T_f$, by

$$\tilde{\eta}(s) := \pi_f \left(\eta \left(\arg\min_{r \in [0,s]} f \circ \eta(r) \right) \right) . \tag{2.42}$$

In particular, $\tilde{\eta}(1 - t^*) = \tilde{\gamma}(t^*)$. If $\tilde{\eta}_-$ denotes the reversed path along $\tilde{\eta}$, the concatenation (without reparametrization),

$$\zeta := \tilde{\gamma} * \tilde{\eta}_- \tag{2.43}$$

is a path going from $\pi_f(x)$ to $\pi_f(y)$ monotone decreasing on $[0, t^*]$ and monotone increasing on $]t^*, 1]$.

For all $\varepsilon > 0$, $\zeta(t^* + \varepsilon)$ does not lie in the same connected component of $X_{f(\zeta(t^*+\varepsilon))}$ as $\pi_f(x)$, but lies in the same connected component of $X_{f(\zeta(t^*+\varepsilon))}$ as $\pi_f(y)$. We are thus reduced to examine the length of the path along two different sections of ζ, each lying in the same connected component as either $\pi_f(x)$ and $\pi_f(y)$. By the previous argument for points of T_f lying in the same connected component of a superlevel set, the length of ζ is

$$L(\zeta) = f(x) - f(\zeta(t^*)) + f(y) - f(\zeta(t^*)) = d_f(x, y) \tag{2.44}$$

by definition of $d_f(x, y)$. Thus, T_f is indeed geodesic and it is an $\mathbb{R}$-tree, by virtue of proposition 2.12.

Finally, the tree is rooted since for any $r < \inf f$, every single point of $X_r = X$ is identified in the quotient (since X was supposed to be connected), so we can identify the root with the point of T_f achieving this infimum. ■

Remark 2.14. If the supremum in the proof of proposition 2.13 is not achieved, it is still possible to construct a geodesic path in T_f. Let us denote

$$r := \sup_{\gamma: x \mapsto y} \inf_{t \in [0,1]} f \circ \gamma \tag{2.45}$$

and X_r^x and X_r^y the path connected components of of $X_r = \{f \geq r\}$ containing x and y respectively. The sets $X_r^x \cap \{f = r\}$ and $X_r^y \cap \{f = r\}$ are not empty, since we can always find an element in these sets by taking the first and last instances where any path $\gamma : x \mapsto y$ hits the level set $\{f = r\}$. On the T_f we have

$$\pi_f(X_r^x \cap \{f = r\}) = \pi_f(X_r^y \cap \{f = r\}) . \tag{2.46}$$

To see this, consider $z \in X_r^x \cap \{f = r\}$ and $z' \in X_r^y \cap \{f = r\}$, then there exist two paths $\eta : x \mapsto z$ and $\eta' : y \mapsto z'$ which lie entirely above r. Furthermore, for any $\varepsilon > 0$, there exists a path γ_ε linking x and y whose image lies entirely in $X_{>r-\varepsilon}$. The concatenation $\eta * \gamma_\varepsilon * \overline{\eta}'$ yields a path whose image lies entirely in $X_{>r-\varepsilon}$ which links z to z'. We conclude that $\pi_f(z) = \pi_f(z')$, since these points are at zero d_f-distance away from one another. The geodesic path in T_f can be found by taking a path γ linking x and y in X, stopping γ as soon as it hits an element of $X_r^x \cap \{f = r\}$ and resume following it at the last instance where γ intersects $X_r^y \cap \{f = r\}$. Taking the images of this stopped version of γ as per the construction of the proof above yields a geodesic path in T_f linking $\pi_f(x)$ and $\pi_f(y)$.

Remark 2.15. If $X = [0, 1]$, there is only one possible path between any two points $x < y$, so the definition above boils down to

$$d_f(x, y) := f(x) + f(y) - 2 \inf_{t \in [x,y]} f , \tag{2.47}$$

which is exactly the distance originally introduced by Le Gall *et al.* [28].

2.2 From trees to barcodes

Given a tree stemming from a continuous function $f : X \to \mathbb{R}$, it is possible to reconstruct the H_0-barcode of f from T_f. If T_f has a finite number of leaves, the relation between the barcode of $H_0(X, f)$ with respect to the superlevel filtration and the tree T_f is given by algorithm 1.

Algorithm 1: A functorial relation between persistence modules and $\mathbb{R}$-trees

Result: $\mathbb{V}$
$\mathcal{F} \leftarrow T$;
$\mathbb{V} \leftarrow 0$;
$i \leftarrow 0$;
while $\mathcal{F} \neq \varnothing$ **do**
  Find γ the longest path in $\mathcal{F}$ starting from a root α and ending in a leaf β ;
  if $i = 0$ **then**
    $\mathbb{V} \leftarrow \mathbb{V} \oplus k[\ell(\alpha), \infty[$;
  else
    $\mathbb{V} \leftarrow \mathbb{V} \oplus k[\ell(\alpha), \ell(\beta)[$;
  end
  $\mathcal{F} \leftarrow \overline{\mathcal{F} \setminus \mathrm{Im}(\gamma)}$;
  $i \leftarrow i + 1$;
end
return $\mathbb{V}$

Definition 2.16. We say T_f is finite if it has a finite number of leaves and say it is infinite otherwise.

If T_f is infinite, we can still give a correspondence between the barcode and the tree proceeding by approximation. This approximation procedure requires the introduction of so-called ε-trimmings of T_f, of which we briefly recall the definition. Since the results of this section can be easily extended to any compact tree, we formulate the rest of this section in full generality.

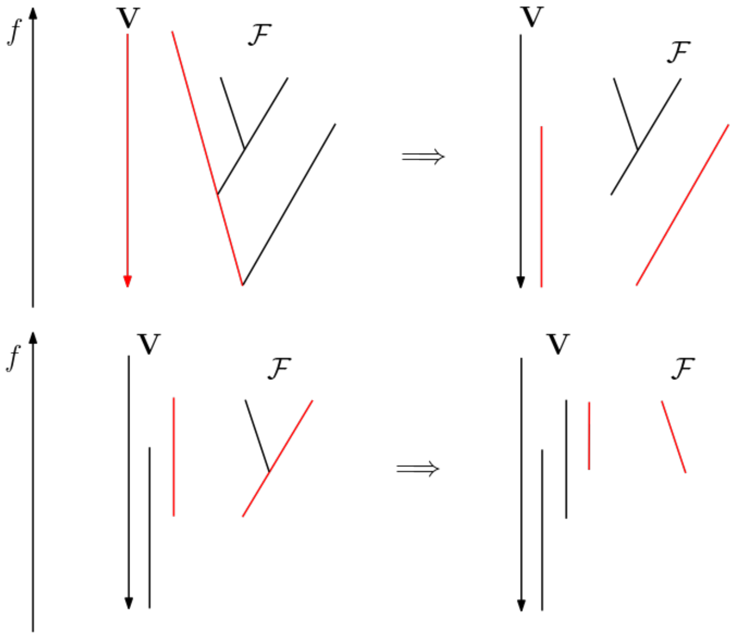


Figure 1: The first four iterations of algorithm 1. For every step, in red is the longest branch of the tree, which we use to progressively construct the persistent module $\mathbb{V}$ by associating an interval module whose ends correspond exactly to the values of the endpoints of the branches.

For any rooted $\mathbb{R}$-tree (T, d, O), we can define a filtering function $\ell : T \to \mathbb{R}$ by setting

$$\ell(\tau) := d(O, \tau)\,. \tag{2.48}$$

This allows us to define the height above a point τ as follows.

Definition 2.17. The function of **the height above** τ on a rooted $\mathbb{R}$-tree T is a function $h : T \mapsto \mathbb{R}$, defined as

$$h(\tau) := \sup_{\eta \in T^{\tau}_{\ell(\tau)}} d(O, \eta) - \ell(\tau)\,, \tag{2.49}$$

The height above τ allows us to define so-called ε*-trimmings* or ε*-simplifications* of T.

Definition 2.18. The ε**-simplified tree of** T**,** T^{ε} or the ε**-trimmed tree of** T, is the subtree of T defined as

$$T^{\varepsilon} := \{\tau \in T \mid h(\tau) \geq \varepsilon\} \tag{2.50}$$

Provided T is compact, its ε-trimmings are always finite. For a monotone decreasing sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ such that $\varepsilon_n \to 0$, we have the following chain of inclusions

$$T^{\varepsilon_1} \hookrightarrow T^{\varepsilon_2} \hookrightarrow T^{\varepsilon_3} \hookrightarrow \cdots\,. \tag{2.51}$$

Applying algorithm 1, we get a set of maps on the persistence modules induced by these inclusions. More precisely, denoting $\mathrm{Alg}(T^{\varepsilon_n})$ the output of the algorithm

$$\mathrm{Alg}(T^{\varepsilon_1}) \to \mathrm{Alg}(T^{\varepsilon_2}) \to \mathrm{Alg}(T^{\varepsilon_3}) \to \cdots\,. \tag{2.52}$$

where the morphisms are the maps induced at the level of the interval modules generating $\mathrm{Alg}(T^{\varepsilon_n})$. The interval modules $k[\alpha, \beta_n[$ of $\mathrm{Alg}(T^{\varepsilon_n})$ satisfy that there is exactly one interval module of $\mathrm{Alg}(T^{\varepsilon_m})$ $(m > n)$ such that $[\alpha, \beta_n[\subset [\alpha, \beta_m[$. A natural definition for infinite T is thus

$$\mathrm{Alg}(T) := \varinjlim \mathrm{Alg}(T^{\varepsilon_n})\,. \tag{2.53}$$

In categoric terms, the algorithm above in fact is a functor

$$\mathrm{Alg} : \mathbf{Tree} \to \mathbf{PersMod_k}\,, \tag{2.54}$$

where **Tree** is the category of rooted $\mathbb{R}$-trees seen as metric spaces, whose morphisms are isometric embeddings (which are not required to be surjective) preserving the roots, and where $\mathbf{PersMod_k}$ is the category of q-tame persistence modules over a field k (*cf.* Oudot's book for details on the category of persistence modules [48]). The action of Alg on morphisms between two trees $\zeta : T \to T'$ is defined as follows. Let $\bigoplus_i k[\alpha_i, \beta_i[$ be the interval module decomposition stemming from $\text{Alg}(T)$. By construction, the intervals $[\alpha_i, \beta_i[$ are in bijective correspondence with branches of T. Since both T and T' are finite and since ζ is an isometric embedding and it preserves the root, we can look at the intesection $\text{Im}(\zeta) \cap [\alpha_i, \beta_i[$, where we somewhat abuse the notation, by regarding the intervals as embedded in T'. Somewhat abusing the notation once again to look at these intersections as simple intervals, we define

$$\text{Alg}(\zeta) := \bigoplus_i \text{id}_{k[\alpha_i,\beta_i[\cap \text{Im}(\zeta)} \ . \tag{2.55}$$

If T is infinite, we extend the above definition by taking successive ε_n-simplifications of T and taking the direct limit of the construction above. This procedure is well-defined since ε_n-simplifications only depend on the function h, which in turn can be taken to only depend on the distance to the root.

2.2.1 Trees stemming from a function

Let us now consider a tree T_f stemming from a function f and show that $\text{Alg}(T_f) = H_0(X, f)$.

Proposition 2.19. Let τ and η be elements of T_f such that $f(\tau) < f(\eta)$ and let $x \in \pi_f^{-1}(\tau)$ and $y \in \pi_f^{-1}(\eta)$, then

$$\forall \varepsilon > 0, \exists \text{ path } \gamma : x \mapsto y \text{ s.t. } \forall t,\ f(\gamma(t)) > f(\tau) - \varepsilon \iff h(\tau) \geq f(\eta) - f(\tau) \text{ and } \forall \varepsilon > 0\, x, y \in X^{\tau}_{>f(\tau)-\varepsilon} \ . \tag{2.56}$$

Proof. Since there exists γ connecting x and y and since γ always stays above $f(\tau) - \varepsilon$ for all $\varepsilon > 0$, we conclude naturally that $\text{Im}(\gamma) \subset X^{\tau}_{>f(\tau)-\varepsilon}$, which implies that for all $\varepsilon > 0$, $h(\tau) \geq f(\eta) - f(\tau) + \varepsilon$ by definition of $h(\tau)$.

The implication ($\Leftarrow$) is clear since if for all $\varepsilon > 0$, $x, y \in X^{\tau}_{>f(\tau)-\varepsilon}$ and since $X^{\tau}_{>f(\tau)-\varepsilon}$ is connected, by path connectedness of X there exists a path between x and y which stays above $f(\tau) - \varepsilon$ for all $\varepsilon > 0$. ■

This proposition suffices to prove the following theorem on the validity of algorithm 1.

Theorem 2.20. *Let X be a compact, connected, locally path connected topological space and let $f : X \to \mathbb{R}$ be continuous. Then $\text{Alg}(T_f) = H_0(X, f)$ in the observable category of persistence modules (cf. [13]).*

Remark 2.21. This theorem is a slight improvement on the result of Curry in [23, Theorem §2.13]. In the language of [23], this constitutes a proof of the "Elder rule" with less assumptions of regularity. Indeed, in [23], the assumption of a Morse set (or that f is a Morse function) is necessary for the proof, whereas the functions hereby considered are merely required to be continuous.

Remark 2.22. By setting $X = T$ and $\ell = f$, theorem 2.20 states that $\text{Alg}(T) = H_0(T, \ell)$.

Proof. Suppose that T_f is finite, then $\text{Alg}(T_f)$ is a decomposable persistence module $\text{Alg}(T_f) := \mathbb{V}$. The fact that $\mathbb{V}$ is pointwise isomorphic to $H_0(X, f)$ (after Serre localization, *i.e.* up to evanescent modules of the form $k[\alpha, \alpha]$) holds since d_f correctly identifies the connected components of the (open) superlevel sets. This guarantees the existence of a pointwise isomorphism since both spaces have the same (finite) dimension.

Let us now check that $\operatorname{rank}(\mathbb{V}(r \to s)) = \operatorname{rank}(H_0(X_r \to X_s))$. The inclusion $X_r \hookrightarrow X_s$ induces the following long exact sequence in homology

$$\begin{array}{c} \cdots \longrightarrow H_1(X_s) \longrightarrow H_1(X_s, X_r) \longrightarrow H_0(X_r) \\ \downarrow \\ H_0(X_s) \longrightarrow H_0(X_s, X_r) \longrightarrow 0 \end{array}$$

Since this sequence is exact

$$\operatorname{rank}(H_0(X_r \to X_s)) = \dim\ker(H_0(X_s) \to H_0(X_s, X_r)) \,. \tag{2.57}$$

For notational simplicity, let us denote $\phi : H_0(X_s) \to H_0(X_s, X_r)$. Note that $\phi[c] = [0]$ if and only if there is a path γ between the representative $c \in X_s$ and an element $b \in X_r$ such that γ stays within X_s. Without loss of generality, let us take c such that $c \in \{f = s\}$. Finding such a path γ is only possible if c and b lie in the same connected component of X_r. By proposition 2.19, this can happen if and only if $h([c]_{T_f}) \geq r - s$. It follows that

$$\dim\ker\phi = \#\{\tau \in T_f \,|\, h(\tau) \geq r - s \ \text{ and } \ f(\tau) = s\} \,, \tag{2.58}$$

which concludes the proof for the finite case.

If T_f is infinite, we consider a sequence of ε_n-trimmings of T_f such that $\varepsilon_n \xrightarrow[n\to\infty]{} 0$. For any $r > s$, there exists n such that $r - s > \varepsilon_n$. But $T_f^{\varepsilon_n}$ is finite, so we are reduced to the previous case. ■

2.3 The inverse problem

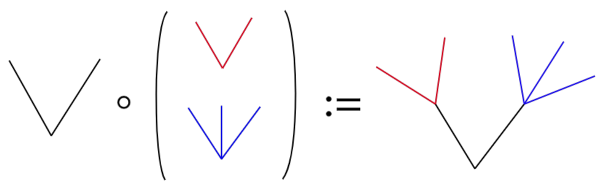
An interesting question is whether every (compact) tree stems from a function $f : X \to \mathbb{R}$. If the tree is a so-called *merge tree* (in particular, we require that it be locally finite and 1-dimensional), a solution has been provided by Curry in [23, §6]. We will now positively answer this question under the assumptions that $\overline{\dim}\, T < \infty$ and that $X = [0, 1]$ by constructing a function $f : [0, 1] \to \mathbb{R}$, which constitute a wider class of trees than merge trees. The rest of this section will focus on proving the following theorem:

Theorem 2.23. *Let T be a compact $\mathbb{R}$-tree such that $\overline{\dim}\, T < \infty$. Then, for any $\delta > 0$ it is possible to construct a continuous function $f : [0, 1] \to \mathbb{R}$ of finite $(\overline{\dim}\, T + \delta)$-variation such that $T = T_f$. In particular, up to a reparametrization, f can be taken to be $\frac{1}{\overline{\dim}\, T+\delta}$-Hölder continuous.*

The idea is to once again use ε-simplifications T^ε for which we can construct a function by taking the contour of the tree. Such a construction is referred to as the Dyck path in the terminology of [56].

2.3.1 Finite trees

We can regard a rooted discrete tree as being an operator with N inputs, where N is the number of leaves of the tree. There is a natural operation on the space of discrete trees which composes these operations by:



These objects are called **operads** and originated in the study of iterated loop spaces [8,9,42]. Since then, these objects have been studied in different fields for a variety of purposes [35,40]. We

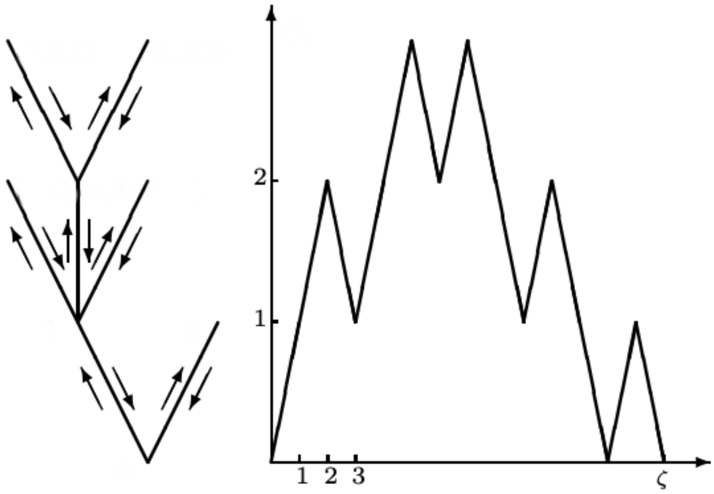


Figure 2: The Dyck path is the function f which assigns the height (the distance from the root) of each vertex of the tree as we wrap around the tree following a clockwise contour around it. There is a map $\phi : T \mapsto [0, \zeta]$ where $[0, \zeta]$ is now marked at the points at which f achieves its local maxima. The figure is taken from [28].

will not give the explicit definition of an operad here, as a rigorous introduction is unnecessary for our purposes. However, we introduce this notion of composition of trees for notational simplicity.

Given a discrete $\mathbb{R}$-tree T, if we have an embedding of T in $\mathbb{R}^2$, or equivalently, a partial order on its vertices, we can assign to T an interval I of a certain length with N marked points as well as a function $f_T : I \to \mathbb{R}$, where N is the number of leaves of T. Using the terminology of [56], a way to do this is by considering the so-called **Dyck path** or **contour path** where the path around T parametrized by arclength in T. The construction of the Dyck path has been carefully detailed in [28, 56], but it is better understood by looking at figure 2. By construction the equality: $T_{f_T} = T$ holds for any discrete $\mathbb{R}$-tree T. Here, equality is taken up to isometry.

As per the description of figure 2, the construction of the Dyck path yields a map ϕ which to T assigns an interval $\phi(T)$ with N marked points. An example of the action of ϕ is illustrated in figure 3.

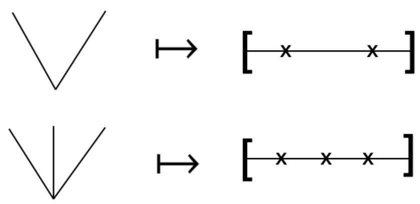


Figure 3: The action of ϕ on trees with two and three leaves respectively. The length of the intervals assigned is exactly the length of the contour around the trees and the marked points are the points at which f_T achieves its maxima.

This operation ϕ is in fact a "morphism" with respect to a composition operation on the intervals, defined as follows. If we have an interval I with N marked points and N intervals J_k each with M_j marked points, the result of the operation $I \circ (J_1, \cdots, J_N)$ is the insertion of the marked interval J_k at the kth marked point of I. The length of $I \circ (J_1, \cdots, J_N)$ is

$$|I \circ (J_1, \cdots, J_N)| = |I| + \sum_{k=1}^{n} |J_k| \, , \tag{2.59}$$

where $|\cdot|$ denotes the lengths of the intervals. The fact that ϕ is a "morphism" results from the definitions of compositions for trees and intervals. We can also define a variant of this morphism

ϕ, which we will call ϕ_λ, which for any tree T simply scales the (marked) interval $\phi(T)$ by a factor λ.

Given a tree T the Dyck path $f_T : \phi(T) \to \mathbb{R}$ can be transformed into a function $f_T^\lambda : \phi_\lambda(T) \to \mathbb{R}$ by setting

$$f_T^\lambda(x) := f_T(x/\lambda)\,. \tag{2.60}$$

This is a rescaling of the x-axis which means that $T_{f_T^\lambda} = T_{f_T} = T$ still holds. Once again, these equalities are taken up to isometry.

Remark 2.24. The definition of f_T^λ is readily generalizable to forests. If $\mathcal{F}$ denotes a forest, then we define $f_\mathcal{F}^\lambda = \bigsqcup_{T\in\mathcal{F}} f_T^\lambda$.

For discrete trees, there is an upper bound of the number of vertices of the tree given its number of leaves.

Lemma 2.25. Let T be a rooted discrete tree, $N \geq 2$ be its number of leaves and V be its number of vertices, then

$$V \leq 2N - 1\,. \tag{2.61}$$

In particular, if the edges of T all have length 1, the contour of the tree can be done over an interval of length at most $4N - 2$

Proof. For binary trees, it is known that [28, 56]

$$V = 2N - 1\,. \tag{2.62}$$

Given a tree with N leaves, we can obtain a binary tree with N leaves by blowing up the vertices which are non-binary. The inequality of the lemma follows. On a binary tree, the Dyck path passes through almost every point in T twice, so the length of the interval is exactly $4N - 2$. Since binary trees are the extremal case, a bound for all trees with N leaves follows. ■

The results above show the result of theorem 2.23 for finite trees, since their upper-box dimension is equal to 1.

2.3.2 Infinite trees

The concatenation of trees can be defined for $\mathbb{R}$-trees too in the obvious way. Given an infinite number of compositions, we can define a limit tree by defining it to be the limit of the partial compositions in the Gromov-Hausdorff sense. Ideally, we would like to have an equality of the following type

$$T = T^a \circ \overline{(T \setminus T^a)}\,, \tag{2.63}$$

where $T \setminus T^a$ now denotes the rooted forest corresponding to the set $T \setminus T^a$. This equality is desirable because by taking infinitely many compositions, we can eventually recover the original tree T, by composing successive ε_n-simplifications with each other. However, this equality does not hold since T^a might not have the right amount of leaves for this operation to be well-defined. Nonetheless, we can decide to count the vertices $T^a \cap \overline{(T \setminus T^a)}$ as leaves with multiplicity, so that the equality above holds.

For an infinite compact tree with $\overline{\dim}\, T < \infty$, the idea is to take some appropriate rapidly decreasing (monotonous) sequence $(\varepsilon_n)_{n\in\mathbb{N}^*}$ such that the interval

$$I = \phi_{\varepsilon_1}(T^{\varepsilon_1}) \circ \phi_{\varepsilon_2}(T^{\varepsilon_2} \setminus T^{\varepsilon_1}) \circ \phi_{\varepsilon_3}(T^{\varepsilon_3} \setminus T^{\varepsilon_2}) \circ \cdots \tag{2.64}$$

has finite length. On each $\phi_{\varepsilon_k}(T^{\varepsilon_k} \backslash T^{\varepsilon_{k-1}})$ we can consider the Dyck path on the forest $T^{\varepsilon_k} \backslash T^{\varepsilon_{k-1}}$. Defining a correct superposition of these Dyck paths, we would be done (*cf.* figure 4).

For an infinite tree, it suffices to show that the sequence generated by the procedure of figure 4 converges in the Gromov-Hausdorff sense to an interval of finite length I and that $(f_i)_i$ converge in $L^\infty(I)$ to some function f.

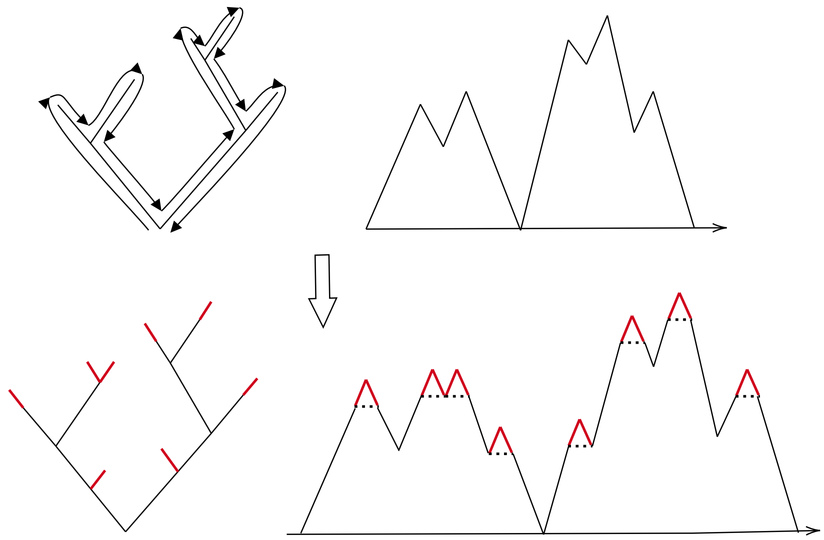


Figure 4: Starting from a tree $T^{a/2^k}$ (black) we construct the Dyck path around it in the first step. Then, we look at $T^{a/2^{k+1}}$ which leads to the addition of intervals (dotted), and a correction of the function at the kth step f_k (which is the function depicted in black, extended linearly over the new intervals). We can further define a function by pasting the Dyck paths of the forest over the corresponding leaves, which leads to the function depicted in the second step (red and black).

Detailed construction of the approximants

Definition 2.26. Let $I \subset \mathbb{R}_+$ be a marked interval with n marked points, which we will denote $(i_k)_{\{1\leq k\leq n\}}$. Furthermore, let $(J_k)_{\{1\leq k\leq n\}}$ be a set of n marked intervals of $\mathbb{R}_+$, each with j_k marked points. Define $\sigma_I : I \to I \circ (J_1, \cdots J_n)$ by

$$\sigma_I(x; J_1, \cdots, J_n) := \left[x + \sum_{i=1}^{\arg\max_k \{i_k < x\}} |J_i| \right] \in I \circ (J_1, \cdots, J_n)\,. \tag{2.65}$$

This definition naturally extends to the whole interval I and we can define the image $\sigma_I(I; J_1 \cdots J_n)$, which is diagramatically represented in figure 5.

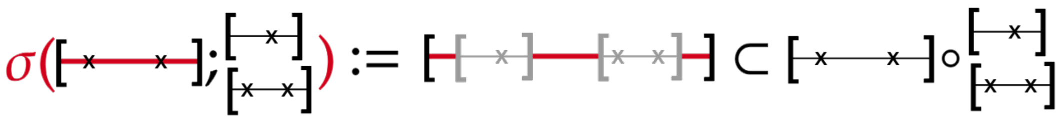


Figure 5: $\sigma_I(I; J_1 \cdots J_n)$

Remark 2.27. Fixing $J_1, \cdots, J_n$, σ_I is a bijective map onto its image, meaning every point $y \in \sigma_I(I; J_1, \cdots, J_n)$ admits a preimage in I, which we will denote by $\sigma_I^{-1}(y; J_1, \cdots, J_n)$.

Definition 2.28. Let $f : I \to \mathbb{R}$ be a continuous function from an interval I with n marked points and let $(J_1, \cdots, J_n)$ be intervals with each with j_i marked points as before. Abusing the notation, we define another function $\sigma(-; J_1, \cdots, J_n)$ which assigns a function on I to a function on $\sigma_I(I; J_1, \cdots, J_n)$ via the following formula

$$\sigma_I(f; J_1, \cdots, J_n)(x) := \begin{cases} f(\sigma_I^{-1}(x; J_1, \cdots, J_n)) & x \in \sigma_I(I; J_1, \cdots, J_n) \\ \text{Linearly extend elsewhere} & \end{cases} \tag{2.66}$$

Remark 2.29. By continuity of $f : I \to \mathbb{R}$, this linear extension on $I \circ (J_1, \cdots, J_n)$ is in fact constant everywhere outside $\sigma_I(I; J_1, \cdots, J_n)$ (this is the dotted region in figure 4). Note also that $\sigma_I(f; J_1, \cdots, J_n)$ is continuous.

Definition 2.30. Given a tree T_f associated to a continuous function $f : I \to \mathbb{R}$, defined on an interval $I \subset \mathbb{R}$. Let $\tau \in T_f$, define the **left preimage of** τ**,** $\overleftarrow{\tau}$ and the **right preimage of** τ **by** π**,** $\overrightarrow{\tau}$ as

$$\overleftarrow{\tau} := \inf \pi^{-1}(\tau) \tag{2.67}$$

$$\overrightarrow{\tau} := \sup \pi^{-1}(\tau)\,. \tag{2.68}$$

Definition 2.31. Let T be a discrete rooted tree and $T' \subset T$ be a subtree sharing roots with T and suppose that we have chosen some embedding of T. Suppose there is a function $f : I \to \mathbb{R}$ on a certain interval I such that $T_f = T'$. Then, the marking of I induced by T is the marking induced by marking the preimage $\pi_f^{-1}(T' \cap \overline{(T \setminus T')})$ chosen in the following way:

- If $\tau \in T' \cap \overline{(T \setminus T')}$ admits a single preimage, choose this preimage;
- Else, if the connected component of τ in $\overline{T \setminus T'}$ is smaller (with respect to the partial order on the tree induced by the embedding of T) than every vertex strictly greater than $\tau \in T'$, choose $\overleftarrow{\tau}$. Otherwise, choose $\overrightarrow{\tau}$. In simpler terms, we choose $\overrightarrow{\tau}$ or $\overleftarrow{\tau}$ depending on whether the subtree of $\overline{T \setminus T'}$ containing τ branches to the right or to the left respectively of T', with the convention that we say that it branches to the left if lies at the top of a leaf of T' (*cf.* figure 6).

We will denote this marking operation by $\mu(I; T', T, f)$.

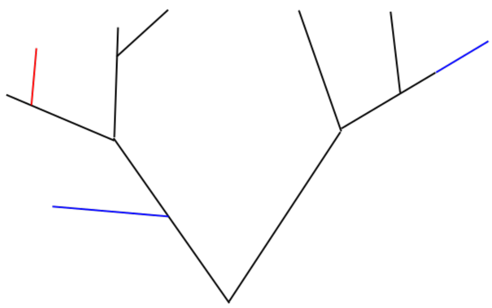


Figure 6: A tree T embedded in $\mathbb{R}^2$ with a subtree T' in black, the subtrees highlighted in red branch to the right and those in blue to the left.

We can also define analogous maps to σ_I, but this time on the intervals J_k as follows.

Definition 2.32. Let $I \subset \mathbb{R}_+$ be a marked interval with n marked points, which we will denote $(i_k)_{\{1 \leq k \leq n\}}$. Furthermore, let $(J_k)_{\{1 \leq k \leq n\}}$ be a set of n marked intervals of $\mathbb{R}_+$, each with j_k marked points. Define $\eta_I^{J_k} : J_k \to I \circ (J_1, \cdots, J_n)$ by

$$\eta_I^{J_k}(x; J_1, \cdots, J_n) := x + i_k + \sum_{j=1}^{k-1} |J_j|\,. \tag{2.69}$$

These maps define a map $\eta_I = \bigsqcup_k \eta_I^{J_k}$ on $\bigsqcup_k J_k$ and η_I also induces a map on the functions $f : \bigsqcup_k J_k \to \mathbb{R}$, defined analogously to σ_I, which we shall also denote η_I.

With this notation, the construction is made in accordance to algorithm 2. A depiction of the mechanism of algorithm 2 can be found in figure 4. For an infinite tree, it suffices to show that the sequence generated by this algorithm converges in the Gromov-Hausdorff sense to an interval of finite length I and that $(f_i)_i$ converge in $L^\infty(I)$ to some function f.

Algorithm 2: Construction of approximants

Output: A set of unions of intervals $(I_i)_{i\in\{1,\cdots,n\}}$ and a set of functions on I_n, $(f_i : I_n \to \mathbb{R})_{i\in\{1,\cdots,n\}}$

Input: An infinite tree T and $a > 0$.

$I_1 \leftarrow \phi(T^a)$;
$f_1 \leftarrow f_{T^a}$;
$I \leftarrow I_1$;
$i \leftarrow 1$;
while $i \leq n$ **do**
  $I_{i+1} := I_i \circ \phi_{\lambda^i}(\overline{T^{a/2^{i+1}} \setminus T^{a/2^i}})$;
  $f \leftarrow \eta_{I_{i+1}}(f^{\lambda^{i+1}}_{T^{a/2^{i+1}}\setminus T^{a/2^i}}; I_1, \cdots I_i)$;
  $I_i \leftarrow \mu(I_i; T^{a/2^{i-1}}, T^{a/2^i}, f_i)$;
  for j=1; $j \leq i$ **do**
    $I_j \leftarrow \sigma(I_j; \phi_{\lambda^i}(\overline{T^{a/2^{i+1}} \setminus T^{a/2^i}}))$;
    $f_j \leftarrow \sigma(f_j; \phi_{\lambda^i}(\overline{T^{a/2^{i+1}} \setminus T^{a/2^i}}))$;
    $j \leftarrow j + 1$;
  end
  $f_{i+1} := f_i + f$;
  $i \leftarrow i + 1$;
end
return $(I_i)_{i\in\{1,\cdots,n\}}$, $(f_i)_{i\in\{1,\cdots,n\}}$.

End of the proof

To get the desired convergence we must show the two following lemmata.

Lemma 2.33. If T is a compact $\mathbb{R}$-tree of finite upper-box dimension, there exist a and λ such that I defined by the construction above has finite length.

We need to show the convergence of the corresponding functions $(f_n)_n$. This can be done by proving that the sequence is Cauchy.

Lemma 2.34. Given the definition of functions f_n above, then the sequence $(f_n)_{n\in\mathbb{N}^*}$ is Cauchy in $C^0(I)$, we have

$$\|f_n - f_m\|_{C^0} \leq a2^{-(n\wedge m)} \tag{2.70}$$

for any n and $m \in \mathbb{N}^*$.

By completeness of C^0, the sequence $(f_n)_{n\in\mathbb{N}^*}$ uniformly converges to a continuous function f. By virtue of stability theorem for trees (theorem 4.21) it follows that T is isometric to T_f. Using Picard's theorem (theorem 3.6)

$$\mathcal{V}(f) = \overline{\dim}\, T_f = \overline{\dim}\, T \tag{2.71}$$

which concludes the proof of theorem 2.23.

Proof of lemma 2.33. Recall that, according to the proof of theorem 3.9, the following equality holds for any tree T

$$\limsup_{\varepsilon\to 0} \frac{\log N^\varepsilon}{\log(1/\varepsilon)} \vee 1 \leq \overline{\dim}\, T := \alpha\,. \tag{2.72}$$

Unpacking the definition of the limit, for any $\delta > 0$ there is a $a > 0$ such that for all $\varepsilon < a$, we have that

$$N^\varepsilon < C\varepsilon^{-\alpha-\delta}\,. \tag{2.73}$$

Let us fix such a δ and pick a small enough so that the condition above holds. For any $n \in \mathbb{N}^*$, the partial composition of intervals has length

$$|I_n| = |\phi(T^a)| + \sum_{k=1}^{n} \left| \phi_{\lambda^k}(T^{a/2^k} \setminus T^{a/2^{k-1}}) \right| . \tag{2.74}$$

However, we can bound $\left| \phi_{\lambda^k}(T^{a/2^k} \setminus T^{a/2^{k-1}}) \right|$ by

$$\begin{aligned} \left| \phi_{\lambda^k}(T^{a/2^k} \setminus T^{a/2^{k-1}}) \right| &= \lambda^k \left| \phi(T^{a/2^k} \setminus T^{a/2^{k-1}}) \right| \\ &\leq \lambda^k \left(\frac{a}{2^k} \right) (4N^{a/2^k}) , \end{aligned} \tag{2.75}$$

since on $T^{a/2^k} \setminus T^{a/2^{k-1}}$ the distances between the vertices of each tree are at most $a/2^k$ and there are at most $4N^{a/2^k}$ such edges by virtue of lemma 2.25. Thus,

$$\left| \phi_{\lambda^k}(T^{a/2^k} \setminus T^{a/2^{k-1}}) \right| < 4\lambda^k \left(\frac{a}{2^k} \right)^{1-\alpha-\delta} = 4a^{1-\alpha-\delta} \left(2^{\alpha+\delta-1} \lambda \right)^k . \tag{2.76}$$

Setting $\lambda < 2^{1-\alpha-\delta}$ I_n converges to some interval of finite length I, since the partial sums $|I_n|$ converge. ■

Proof of lemma 2.34. Suppose that $n < m$. It is sufficient to show that on I_m the equality holds, since in all further iterations of the algorithm, the functions f_n and f_m are locally constant over the intervals introduced. By definition of f_n, f_n and f_m agree on I_n. Outside of this set, f_n is constant and the difference in the L^∞-norm depends only on what happens above $T^{a/2^n}$, thus we can write

$$\|f_n - f_m\|_{L^\infty} \leq \left\| f_{T^{a/2^m} \setminus T^{a/2^n}} \right\|_{L^\infty} \tag{2.77}$$

by definition of f_n. However, the Dyck path on $T^{a/2^m} \setminus T^{a/2^n}$ can at most reach a height of $a(2^{-n} - 2^{-m}) < a2^{-n}$, which finishes the proof. ■

3 Regularity, persistence index and metric properties of trees

Throughout this section X will be a compact, connected and locally path-connected metric space. On general topological spaces, it is important to specify which homological theory we are using to compute the homology of X. For nice enough spaces, this choice has little to no importance, as most homological theories coincide. However, for abstract metric spaces this is no longer necessarily the case. For our purposes, we will always consider the homology of the space X to be its Čech homology. *A priori*, this might pose some problems, as Čech homology does not always satisfy the axioms of a proper homological theory in the sense of Eilenberg-Steenrod. For this to be the case, a sufficient condition is to consider X to be compact and the homology to be taken over a field. These are not the only conditions for which Čech homology gives rise to a proper homological theory, as in general the exactness axiom might fail, but suffices for our purposes. For more on these technical details, we encourage the reader to consult Eilenberg's book [31, Chapter 7].

Remark 3.1. If we wish to consider more general topological spaces where the exactness axiom does indeed fail for the Čech homology, there are multiple options. We could either consider more elaborate homology theories such as singular homology or strong homology (which fixes the issue with the exactness axiom of Čech homology), or we could rewrite this paper in cohomological terms and use Čech cohomology, for which this problem doesn't present itself.

With this technicality out of the way, let us now define the main objects which will concern us for the rest of this paper.

Definition 3.2. Let X be a compact, connected, locally path connected topological space and consider $f : X \to \mathbb{R}$ be a continuous function. The kth Pers_p**-functional of** f is

$$\mathrm{Pers}_p(H_k(X,f)) := \left(\sum_{b \in H_k(X,f)} \ell(b \cap [\inf(f), \sup(f)])^p \right)^{1/p} , \tag{3.78}$$

where $\ell(b)$ denotes the length of the bar b and $H_k(X,f)$ denotes the H_k-barcode (or diagram) stemming from the superlevel filtration. Abusing the notation, we will denote $\mathrm{Pers}_p(f) := \mathrm{Pers}_p(H_0(X,f))$. If we further assume that there exists n such that for all $m > n$, $H_m(X) = 0$, we define the **total** Pers_p **functional of** f as

$$\mathrm{TPers}_p(f) := \sum_{k=0}^{n} \mathrm{Pers}_p(H_k(X,f)) . \tag{3.79}$$

Definition 3.3. Let $f : X \to \mathbb{R}$ be a continuous function. The k**th-persistence index of** f is defined as

$$\mathcal{L}_k(f) := \inf\{p \geq 1 \mid \mathrm{Pers}_p(H_k(X,f)) < \infty\} . \tag{3.80}$$

We will sometimes write $\mathcal{L}(f) := \mathcal{L}_0(f)$. Provided that higher degrees of homology identically vanish, we may also talk about the **total persistence index of** f, defined as

$$\mathcal{L}_{Tot}(f) := \inf\{p \geq 1 \mid \sum_k \mathrm{Pers}_p(H_k(X,f)) < \infty\} . \tag{3.81}$$

3.1 1D case: a connection with the p-variation

Definition 3.4. Let $f : [0,1] \to \mathbb{R}$ be a continuous function. The **true** p**-variation of** f is defined as

$$\|f\|_{p-\mathrm{var}} := \left[\sup_D \sum_{t_k \in D} |f(t_k) - f(t_{k-1})|^p \right]^{1/p} , \tag{3.82}$$

where the supremum is taken over all finite partitions D of the interval $[0,1]$.

Remark 3.5. We talk about *true* p-variation to make the distinction with the notion of variation typically considered in probabilistic contexts (more precisely, stochastic calculus), where instead of the supremum over all partitions, we have a probable limit as the mesh of the partition considered tends to zero.

Proposition 3.6 (Picard, §3 [51])**.** Let $f : [0,1] \to \mathbb{R}$ be a continuous function, then $\|f\|_{p-\mathrm{var}}$ is finite as soon as $\mathrm{Pers}_p(f)$ is finite. In fact, for any p

$$\|f\|_{p-\mathrm{var}}^p \leq 2\,\mathrm{Pers}_p^p(f)^p . \tag{3.83}$$

Furthermore, if $\|f\|_{(p-\delta)\text{-var}}$ is finite for some $\delta > 0$, $\mathrm{Pers}_p(f)$ is also finite.

In fact, Picard showed that on the interval $[0,1]$, the persistence index of f is linked to the regularity of f.

Theorem 3.7 (Picard, §3 [51])**.** *Let $f : [0,1] \to \mathbb{R}$ be a continuous function and denote*

$$\mathcal{V}(f) := \inf\{p \mid \|f\|_{p\text{-}var} < \infty\} . \tag{3.84}$$

Then,

$$\mathcal{V}(f) = \mathcal{L}(f) = \limsup_{\varepsilon \to 0} \frac{\log(\lambda(T_f^\varepsilon)/\varepsilon)}{\log(1/\varepsilon)} + 1 = \limsup_{\varepsilon \to 0} \frac{\log N^\varepsilon}{\log(1/\varepsilon)} \vee 1 = \overline{\dim}\, T_f \tag{3.85}$$

where $a \vee b := \max\{a, b\}$, N^ε is the number of leaves of the ε-trimmed tree T_f^ε, $\lambda(T_f^\varepsilon)$ denotes the length of T_f^ε and $\overline{\dim}$ denotes the upper-box dimension.

Remark 3.8. More generally, we can define λ as the unique atomless Borel measure on T_f characterized by the fact that the measure of a geodesic is given by the length of the geodesic [51].

3.2 More general spaces

3.2.1 Connected, locally path-connected, compact topological spaces

Theorem 3.9. *Let X be a connected, locally path-connected, compact topological space and let $f : X \to \mathbb{R}$ be a continuous function. With the same notation as above and supposing that $\overline{\dim}\, T_f$ is finite, the following chain of equalities holds*

$$\mathcal{L}(f) = \limsup_{\varepsilon\to 0} \frac{\log N^\varepsilon}{\log(1/\varepsilon)} \vee 1 = \limsup_{\varepsilon\to 0} \frac{\log(\lambda(T_f^\varepsilon)/\varepsilon)}{\log(1/\varepsilon)} + 1 = \overline{\dim}\, T_f\,. \tag{3.86}$$

Furthermore,

$$\liminf_{\varepsilon\to 0} \frac{\log N^\varepsilon}{\log(1/\varepsilon)} \vee 1 \le \underline{\dim}\, T_f \le \liminf_{\varepsilon\to 0} \frac{\log(\lambda(T_f^\varepsilon)/\varepsilon)}{\log(1/\varepsilon)} + 1\,, \tag{3.87}$$

where $\underline{\dim}$ is the lower-box dimension. For $\underline{\dim}\, T_f > 1$, these inequalities turn into equalities if either:

$$\limsup_{\varepsilon\to 0} \frac{N^{2\varepsilon}}{N^\varepsilon} < 1 \quad \text{or} \quad \limsup_{\varepsilon\to 0} \frac{\lambda(T_f^{2\varepsilon})}{\lambda(T_f^\varepsilon)} < 1\,. \tag{3.88}$$

Remark 3.10. The study of N^ε is in fact completely equivalent to the study of $\mathrm{Pers}_p^p(f)$. Indeed,

$$\mathrm{Pers}_p^p(f) = p \int_0^\infty \varepsilon^{p-1} N^\varepsilon \, d\varepsilon\,, \tag{3.89}$$

which is finite as soon as $p > \mathcal{L}(f)$. This is nothing other than the Mellin transform of N^ε. By the Mellin inversion theorem, for any $c > \mathcal{L}(f)$, we have

$$N^\varepsilon = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \mathrm{Pers}_p^p(f)\, \varepsilon^{-p}\, \frac{dp}{p}\,. \tag{3.90}$$

Proof of theorem 3.9. By the procedure detailed in section 2.3, since $\overline{\dim}\, T_f$ is finite we can construct a function $\hat{f} : [0,1] \to \mathbb{R}$ such that T_f and $T_{\hat{f}}$ are isometric. Applying Picard's theorem to $T_{\hat{f}}$ and noting that $\mathcal{L}(f)$ depends only on the T_f, we have that

$$\mathcal{L}(f) = \limsup_{\varepsilon\to 0} \frac{\log(\lambda(T_f^\varepsilon)/\varepsilon)}{\log(1/\varepsilon)} + 1 = \limsup_{\varepsilon\to 0} \frac{\log N^\varepsilon}{\log(1/\varepsilon)} \vee 1 = \overline{\dim}\, T_f\,. \tag{3.91}$$

Let us now show the inequalities for the $\liminf$. Since

$$\lambda(T_f^\varepsilon) = \int_\varepsilon^\infty N^a \, da\,, \tag{3.92}$$

the following inequality holds

$$\liminf_{\varepsilon\to 0} \frac{\log N^\varepsilon}{\log(1/\varepsilon)} \vee 1 \le \liminf_{\varepsilon\to 0} \frac{\log(\lambda(T_f^\varepsilon)/\varepsilon)}{\log(1/\varepsilon)} + 1\,. \tag{3.93}$$

Additionally,

$$N^\varepsilon \le \mathcal{N}(\varepsilon/2) \tag{3.94}$$

where $\mathcal{N}(\varepsilon)$ denotes the minimal number of balls of radius ε necessary to cover T_f. This inequality holds as above each leaf of T_f^ε, at least one ball of radius $\frac{\varepsilon}{2}$ is necessary to cover this section of the tree. It follows that

$$\liminf_{\varepsilon\to 0} \frac{\log N^\varepsilon}{\log(1/\varepsilon)} \vee 1 \le \underline{\dim}\, T_f\,. \tag{3.95}$$

We can bound this minimal number of balls $\mathcal{N}(\varepsilon)$ by the following

$$\mathcal{N}(\varepsilon) \le N^{\varepsilon/2} + \frac{\lambda(T_f^{\varepsilon/2})}{\varepsilon/2} \le 2\, N^{\varepsilon/2} \vee \left\lceil \frac{\lambda(T_f^{\varepsilon/2})}{\varepsilon/2} \right\rceil\,, \tag{3.96}$$

which holds since, at most N^ε balls are needed to cover $T_f \setminus T_f^\varepsilon$. To cover T_f^ε, at most: $\left\lceil \lambda(T_f^{\varepsilon/2})/(\varepsilon/2) \right\rceil$ balls are needed, so the inequality above follows by further majorizing the terms. This implies that

$$\underline{\dim}\, T_f \leq \left[\liminf_{\varepsilon \to 0} \frac{\log N^\varepsilon}{\log(1/\varepsilon)} \vee 1 \right] \vee \left[\liminf_{\varepsilon \to 0} \frac{\log(\lambda(T_f^\varepsilon)/\varepsilon)}{\log(1/\varepsilon)} + 1 \right], \tag{3.97}$$

but by inequality 3.95 this means that

$$\underline{\dim}\, T_f \leq \liminf_{\varepsilon \to 0} \frac{\log(\lambda(T_f^\varepsilon)/\varepsilon)}{\log(1/\varepsilon)} + 1\,. \tag{3.98}$$

Finally,

$$\begin{aligned} \frac{\lambda(T_f^\varepsilon) - \lambda(T_f^{2\varepsilon})}{\varepsilon} &= \frac{1}{\varepsilon} \left[\int_\varepsilon^\infty N^a \, da - \int_{2\varepsilon}^\infty N^a \, da \right] \\ &= \frac{1}{\varepsilon} \int_\varepsilon^{2\varepsilon} N^a \, da \leq N^\varepsilon \,, \end{aligned} \tag{3.99}$$

since N^ε is monotone decreasing. This reasoning also gives a lower bound

$$N^{2\varepsilon} \leq \frac{\lambda(T_f^\varepsilon) - \lambda(T_f^{2\varepsilon})}{\varepsilon} \leq N^\varepsilon \,, \tag{3.100}$$

which entails that

$$\liminf_{\varepsilon \to 0} \frac{\log N^\varepsilon}{\log(1/\varepsilon)} = \liminf_{\varepsilon \to 0} \frac{\log \left[\frac{\lambda(T_f^\varepsilon) - \lambda(T_f^{2\varepsilon})}{\varepsilon} \right]}{\log(1/\varepsilon)} \,. \tag{3.101}$$

Suppose that this limit is larger than 1. Rearranging, we get

$$\frac{\varepsilon N^{2\varepsilon}}{\lambda(T_f^\varepsilon)} \leq 1 - \frac{\lambda(T_f^{2\varepsilon})}{\lambda(T_f^\varepsilon)} \leq \frac{\varepsilon N^\varepsilon}{\lambda(T_f^\varepsilon)} \,, \tag{3.102}$$

from which it follows that if any of these quantities admits a $\liminf$ which is stricly greater than zero, we have

$$\liminf_{\varepsilon \to 0} \frac{\log N^\varepsilon}{\log(1/\varepsilon)} = \liminf_{\varepsilon \to 0} \frac{\log \lambda(T_f^\varepsilon)}{\log(1/\varepsilon)} + 1\,. \tag{3.103}$$

Noticing another equivalent condition for the validity of this equality is whether

$$\limsup_{\varepsilon \to 0} \frac{N^{2\varepsilon}}{N^\varepsilon} < 1\,, \tag{3.104}$$

finishes the proof. ■

Remark 3.11. If $\overline{\dim} = \underline{\dim}$, all the limits of the above theorem are well-defined, yielding exact asymptotics for $\lambda(T_f^\varepsilon)$ and N^ε. This is in particular the case if $\overline{\dim} = \dim_H$, where $\dim_H$ denotes the Hausdorff dimension.

The functional $\lambda(T_f^\varepsilon)$ is what some authors [52, 53] refer to as the Banach indicatrix and its asymptotics have a topological interpretation as described in the statement of the theorem. It is interesting to note that the study of the upper-box dimension is natural in the tree approach. Additionally, $\overline{\dim}$ has also been used in the context of persistent homology by Schweinhart [54], Schweinhart and MacPherson [41] and by Adams *et al.* [1] in a probabilistic setting.

3.2.2 LLC metric spaces

It is possible to further extend Picard's theorem by some rudimentary considerations and by imposing the so-called locally linearly connected condition on X.

Definition 3.12. A **locally linearly connected (LLC) metric space** (X, d), is a connected metric space such that for all $r > 0$ and for all $z \in X$, for all $x, y \in B(z, r)$, there exists an arc connecting x and y such that the diameter of this arc is linear in $d(x, y)$.

With this extra assumption, we can prove the following lemma.

Lemma 3.13 (Regularity-dimension)**.** Let X be a compact LLC metric space. Keeping the same notations as in theorem 3.9, the following inequality holds

$$\mathcal{L}(f) = \overline{\dim}\, T_f \leq \mathcal{H}(f)\, \overline{\dim}\, X\,, \tag{3.105}$$

where:

$$\mathcal{H}(f) := \inf \left\{ \frac{1}{\alpha} \;\middle|\; \exists \lambda \in \mathrm{Homeo}(X)\,,\ \|f \circ \lambda\|_{C^\alpha} < \infty \right\} \tag{3.106}$$

The proof of this lemma relies itself on two lemmata, which are interesting in and of themselves.

Lemma 3.14. Let X and Y be two metric spaces such that there is a surjective map $\pi : X \to Y$ such that $\pi \in C^\alpha(X, Y)$, then

$$\overline{\dim}\, Y \leq \frac{1}{\alpha}\, \overline{\dim}\, X \quad \text{and} \quad \underline{\dim}\, Y \leq \frac{1}{\alpha}\, \underline{\dim}\, X\,. \tag{3.107}$$

and if we denote K the Hölder constant of π, the following inequality holds

$$\mathcal{N}_Y(\varepsilon) \leq \mathcal{N}_X\left(\left(\frac{\varepsilon}{K} \right)^{1/\alpha} \right)\,. \tag{3.108}$$

Lemma 3.15. Let X be a compact locally linearly connected (LLC) metric space (*cf.* definition 3.12) and let $f : X \to \mathbb{R}$ be a continuous function, then

$$f \in C^\alpha(X, \mathbb{R}) \Longrightarrow \pi_f \in C^\alpha(X, T_f)\,. \tag{3.109}$$

Let us show that lemmata 3.14 and 3.15 imply lemma 3.13.

Proof of lemma 3.13. If, up to precomposition, $f \notin C^\alpha(X, \mathbb{R})$ for any α, there is nothing to show, since the statement is vacuous. Otherwise, since T_f is preserved by precomposition by a homeomorphism, we may suppose without loss of generality that $f \in C^\alpha(X, \mathbb{R})$. The projection onto the tree of f, $\pi_f : X \to T_f$ is in $C^\alpha(X, T_f)$ according to lemma 3.15. It follows from lemma 3.14 that

$$\overline{\dim}\, T_f \leq \frac{1}{\alpha}\, \overline{\dim}\, X\,. \tag{3.110}$$

The statement of the theorem follows by taking the infimum over $\frac{1}{\alpha}$. ■

All that remains to show is the two remaining lemmata.

Proof of lemma 3.14. Since $\pi : X \to Y$ is surjective and $C^\alpha(X, Y)$, for any $x \in X$

$$\pi\left(B_X\left(x, \left(\frac{\varepsilon}{K} \right)^{1/\alpha} \right) \right) \subset B_Y(\pi(x), \varepsilon) \tag{3.111}$$

for some constant K. It follows that the minimal number of balls needed to cover X, $\mathcal{N}_X$ dominates the minimal number of balls needed to cover Y, $\mathcal{N}_Y$. More precisely

$$\mathcal{N}_Y(\varepsilon) \leq \mathcal{N}_X\left(\left(\frac{\varepsilon}{K} \right)^{1/\alpha} \right) \iff \alpha\, \frac{\mathcal{N}_Y(\varepsilon)}{\log(1/\varepsilon) + \log(K)} \leq \frac{\mathcal{N}_X\left(\left(\frac{\varepsilon}{K} \right)^{1/\alpha} \right)}{\log\left(\left(\frac{K}{\varepsilon} \right)^{1/\alpha} \right)}\,.$$

The statement of the lemma follows. ■

Proof of lemma 3.15. Suppose that $f : X \to \mathbb{R}$ is in $C^\alpha(X, \mathbb{R})$ with Hölder constant Λ and let $x, y \in X$. Without loss of generality, suppose that $f(x) < f(y)$. Since T_f is a geodesic space, the distance $d_f(\pi_f(x), \pi_f(y))$ is the length of the geodesic arc in T_f linking $\pi_f(x)$ and $\pi_f(y)$. By compactness of this geodesic path, there is a point $\tau \in T_f$ where f achieves its minimum, thus

$$d_f(\pi_f(x), \pi_f(y)) = f(x) - f(\tau) + f(y) - f(\tau) \,. \tag{3.112}$$

This minimum $f(\tau)$ has the particularity that

$$f(\tau) = \sup_{\gamma : x \mapsto y} \inf_{t \in [0,1]} f \circ \gamma \,, \tag{3.113}$$

where the supremum is taken over all paths on X linking x and y. From the LLC condition, we know that there is a path $\eta : x \mapsto y$ whose diameter is controlled by $d_X(x, y)$ and $z \in X$ achieving the minimum of f over η. In particular,

$$f(\tau) \geq \inf_{t \in [0,1]} f \circ \eta =: f(z) \,. \tag{3.114}$$

Since f is α-Hölder on X,

$$f(x) - f(\tau) \leq f(x) - f(z) \leq \Lambda \, d(x, z)^\alpha \leq \Lambda \, \mathrm{diam}(\eta)^\alpha \leq C\Lambda \, d(x, y)^\alpha \tag{3.115}$$

for some constant C determined by the LLC condition and we have an analogous inequality for $f(y) - f(\tau)$. Putting everything together we have that:

$$d_f(\pi_f(x), \pi_f(y)) \leq 2C\Lambda \, d_X(x, y)^\alpha \,, \tag{3.116}$$

which finishes the proof. ■

Lemma 3.13 is sharp, since Brownian sample paths almost surely saturate this inequality. However, there is no hope to prove equality for every f. Indeed, for any $f \in C^1(\mathbb{T}^2, \mathbb{R})$ having a finite amount of bars, T_f is a finite tree and has upper-box dimension 1, but

$$\overline{\dim}\, T_f = 1 < 2 = \mathcal{H}(f)\, \overline{\dim}\, \mathbb{T}^2 \,. \tag{3.117}$$

Nonetheless, it is possible to show that lemma 3.13 holds generically. This is a consequence of a generalization of work never published by Weinberger and Baryshnikov. We extend their result to homogenous enough spaces in the following sense.

Definition 3.16. A metric space (X, d) is said to **admit a homogeneous set** (for a certain property) if there exists an open set $U \subset X$ where for every ball $B(x, r) \subset U$, the property of the ball is the same as the property of the space X.

Remark 3.17. In the previous definition, one can for instance take any notion of dimension, entropy, *etc.*

The following proposition will be useful in simplifying the assumptions of the theorem.

Proposition 3.18. Let (X, d) be a compact metric space and $N_P(\varepsilon)$ denote the cardinality of the maximal packing of X by balls of radius ε. Then,

$$\mathcal{N}_X(2\varepsilon) \leq N_P(\varepsilon) \leq \mathcal{N}_X(\varepsilon) \tag{3.118}$$

and in particular,

$$\underline{\dim}(X) = \liminf_{\varepsilon \to 0} \frac{\log(N_P(\varepsilon))}{\log(1/\varepsilon)} \quad \text{and} \quad \overline{\dim}(X) = \limsup_{\varepsilon \to 0} \frac{\log(N_P(\varepsilon))}{\log(1/\varepsilon)} \,. \tag{3.119}$$

Proof. Let M_ε be a maximal packing of X by balls of radius ε. For every $x \in X \setminus (\cup_{V \in M_\varepsilon} V)$ there exists $U \in M_\varepsilon$ such that $d(x,U) \le \varepsilon$, otherwise, $B(x,\varepsilon) \cup M_\varepsilon$ would also be a packing of X with cardinality strictly greater than $|M_\varepsilon|$. It follows that the balls of radius 2ε of centers that of the maximal packing of radius ε is a covering of X, proving the first inequality.

For the second inequality, we reason by contradiction. Suppose there is a maximal packing P_ε and a minimal covering C_ε such that $|P_\varepsilon| \ge |C_\varepsilon| + 1$. Then, since C_ε covers X, by the pigeonhole principle there are at least two centers of balls of P_ε inside a ball of C_ε. But the triangle inequality implies that the balls around these two centers of radius ε have non-empty intersection (as the center of the ball of C_ε in which they are contained is in the intersection), thereby contradicting that P_ε is a packing, showing the result. ■

Theorem 3.19. *Let X be a compact LLC metric space admitting a set of homogeneous upper-box dimension, then for any $0 < \alpha \le 1$*

$$\sup_{f \in C^\alpha(X,\mathbb{R})} \alpha \mathcal{L}(f) = \overline{\dim}(X)\,. \tag{3.120}$$

Moreover, the supremum is attained generically in the sense of Baire, i.e. the set over which $\alpha\mathcal{L}(f) < \overline{\dim}(X)$ is meagre in $C^\alpha(X,\mathbb{R})$.

Once again, we split the proof along key lemmata.

Lemma 3.20. Let X be a compact LLC space, then the functional $\mathrm{Pers}^p_{p,\varepsilon} : C^\alpha_\Lambda(X,\mathbb{R}) \to \mathbb{R}_+$ defined by

$$f \mapsto \sum_{\substack{b \in \mathcal{B}(f) \\ \ell(b) \ge \varepsilon}} \ell(b)^p \tag{3.121}$$

is continuous.

Proof. We start by noting that the total number of bars of length $\ge \varepsilon$ that a function $f \in C^\alpha_\Lambda(X,\mathbb{R})$ can have is uniformly bounded above by virtue of the proof of lemma 3.14 by a constant $C_{X,\alpha,\varepsilon}$. By lemma 3.15, we know $\pi_f : X \to T_f$ is α-Hölder, with Hölder constant K depending only on Λ and X. This fact, combined with the inequality $N^\varepsilon_f \le \mathcal{N}_{T_f}(\varepsilon/2)$ entails that for any f,

$$N^\varepsilon_f \le \mathcal{N}_{T_f}(\varepsilon/2) \le \mathcal{N}_X\left(\left(\frac{\varepsilon}{2K}\right)^{1/\alpha}\right) =: C_{X,\alpha,\varepsilon}\,. \tag{3.122}$$

It follows that for any $f, g \in C^\alpha_\Lambda(X,\mathbb{R})$, by choosing to sum along the d_∞-matching, we have

$$\begin{aligned}
\left|\mathrm{Pers}^p_{p,\varepsilon}(f) - \mathrm{Pers}^p_{p,\varepsilon}(g)\right| &\le \sum_{\substack{b_f \in \mathcal{B}(f)\,,\, b_g \in \mathcal{B}(g) \\ \ell(b_f),\ell(b_g) \ge \varepsilon}} |\ell(b_f)^p - \ell(b_g)^p| \\
&\le \sum p \underbrace{|\ell(b_f) - \ell(b_g)|}_{\le \|f-g\|_\infty \text{ by stability}} \max\{\ell(b_f)^{p-1}, \ell(b_g)^{p-1}\} \\
&\le p\,\|f-g\|_\infty \underbrace{\sum \max\{\ell(b_f)^{p-1}, \ell(b_g)^{p-1}\}}_{\le C_{X,\alpha,\varepsilon}\Lambda^{p-1}\,\mathrm{diam}(X)^{\alpha(p-1)} \text{ by global } \alpha\text{-Hölderness}} \\
&\le C_{X,\alpha,\varepsilon}\,\Lambda^{p-1}\mathrm{diam}(X)^{\alpha(p-1)}\,p\,\|f-g\|_\infty\,.
\end{aligned}$$

■

Lemma 3.21. Let X be a compact, LLC, admitting a set of homogeneous lower-box dimension. Then, for all $p < \overline{\dim}(X)$ and $M \ge 0$, the set of functions

$$\{f \in C^\alpha(X,\mathbb{R}) \mid \mathrm{Pers}^p_p(f) > M\} \tag{3.123}$$

is dense in $C^\alpha(X,\mathbb{R})$.

Proof. Without loss of generality, suppose that the uniform set is a ball of radius 1 inside X, denoted $B \subset X$ and construct a function h of persistence $> M$ on this ball. Noting $d = \overline{\dim}(X)$, by proposition 3.18 and the definition of the upper-box dimension, for some subsequence of $(\varepsilon_n)_n$ decreasing to 0, we have

$$\tilde{C}\varepsilon_n^{-(d-\delta)} \leq N_P(\varepsilon) \leq C\varepsilon_n^{-(d+\delta)} \tag{3.124}$$

for some constants C and $\tilde{C}$. Note E_ε the centers of the balls of a maximal packing of radius ε and define $h_n : B \to \mathbb{R}$ as

$$h_n(x) := d^\alpha(x, E_{\varepsilon_n}) \tag{3.125}$$

The Pers_p-functional of these functions can be bounded below by

$$\mathrm{Pers}_p^p(h_n) \geq N_P(\varepsilon_n)\varepsilon_n^p \geq \tilde{C}\varepsilon_n^{p\alpha-d+\delta} \tag{3.126}$$

for all $\delta > 0$. Since $\alpha p < d$, this quantity can be made as large as we want and in particular $> M$ by picking a large enough n. By the assumptions of the theorem, it is possible to choose the original ball of the construction to have as small a radius as we wish. Note we may perturb any function $f \in C^\alpha(X, \mathbb{R})$ by a function close to it which is locally constant on a small enough ball and on this ball, add h_n for n large enough. Since the ball of the construction can be chosen as small as we want, any neighborhood of f contains a function satisfying the condition of the lemma. ∎

Proof of theorem 3.19. We are interested in showing that for $p < \overline{\dim}(X)$, the set

$$\mathcal{S}(p) := \{f \in C^\alpha(X, \mathbb{R}) \,|\, \mathrm{Pers}_p^p(f) < \infty\} \tag{3.127}$$

is meager in $C^\alpha(X, \mathbb{R})$. Let us start by noticing that

$$\mathcal{S}(p) = \bigcup_{\Lambda \geq 0} \bigcup_{M \geq 0} \mathcal{S}(p, \Lambda, M)\,, \tag{3.128}$$

where the union is taken over an increasing diverging sequences of Λ and M and

$$\mathcal{S}(p, \Lambda, M) := \{f \in C^\alpha_\Lambda(X, \mathbb{R}) \,|\, \mathrm{Pers}_p^p(f) \leq M\}\,. \tag{3.129}$$

Furthermore,

$$\mathcal{S}(p, \Lambda, M) = \bigcap_{k \geq 1} \{f \in C^\alpha_\Lambda(X, \mathbb{R}) \,|\, \mathrm{Pers}_{p,\frac{1}{k}}^p(f) \leq M\}\,. \tag{3.130}$$

By lemma 3.20, $\mathrm{Pers}_{p,\frac{1}{k}}$ is continuous, thereby guaranteeing that these sets are closed in $C^\alpha(X, \mathbb{R})$, and therefore so is their intersection. It remains to show that the $\mathcal{S}(p, \Lambda, M)$ are nowhere dense, but this amounts to finding a dense set of functions for which

$$\mathrm{Pers}_{p,\frac{1}{k}}^p(f) \leq M \tag{3.131}$$

is violated for infinitely many k. It suffices to find a dense set of functions for which the total $\mathrm{Pers}_p^p(f) > M$ (for $p < \overline{\dim}(X)$), but the existence of such a dense family is given by lemma 3.21, showing the result. ∎

Remark 3.22. The space defined by

$$E_p = \{f \in C^0(X, \mathbb{R}) \,|\, \mathcal{L}(f) \leq p\} \tag{3.132}$$

is **not** a linear space.

3.2.3 Doubling spaces with small convex balls

One could ask whether the results of genericity of theorem 3.19 hold in every degree of homology for f within some class of regularity. This question has been considered in [21] and more recently in [55] with different degrees of generality. The following theorem is a slight generalization of the two cited results.

Theorem 3.23. *Let X be a compact connected geodesic doubling space whose small enough balls are geodesically convex. Denote $d = \overline{\dim}(X)$, $k \in \mathbb{N}$ and let $f \in C^\alpha(X, \mathbb{R})$, then $\mathcal{L}_k(f) \leq \frac{d}{\alpha}$.*

Remark 3.24. The doubling assumption is satisfied for Riemannian manifolds whose Ricci curvature is bounded below, by the Bishop-Gromov inequality. By considering Gromov-Hausdorff limits of Riemannian manifolds with Ricci curvature bounded below, we obtain spaces satisfying the doubling property. Spaces included in this class include, but are not limited to, Riemannian manifolds with conic singularities. In general, it is also possible to obtain less well-behaved spaces. For more on poorly behaved examples, we refer the reader to the works of Xavier Menguy [44, 45] and to even more recent and poorly behaved examples, such as those described in [37].

The proof relies on the two following well-known lemmata.

Lemma 3.25 (Nerve lemma, Lemma 4.11 [48])**.** Let X be a paracompact space, and let $\mathcal{U}$ be an open cover of X such that the $(k+1)$-fold intersections of elements of $\mathcal{U}$ are either empty or contractible for all $k \in \mathbb{N}$. Then, there is a homotopy equivalence between the nerve of $\mathcal{U}$ and X.

Lemma 3.26. Let (X, d) be a geodesic metric space whose balls of radius $\leq \varepsilon$ are geodesically convex. Then, minimal coverings of X by balls of radius $\leq r_C$ are such that the $(k+1)$-fold intersections of elements of $\mathcal{U}$ are either empty or contractible for all $k \in \mathbb{N}$. We call the maximal radius for which balls are geodesically convex the **convexity radius** of X and we denote it r_C.

Proof of theorem 3.23. The proof is an immediate consequence of the proof of theorem 3.19, where we only need to modify the proof of lemma 3.20. For this, it is sufficient to bound the number of bars in the persistence diagram of the kth degree in homology of length $\geq \varepsilon$, N_k^ε. On a given a minimal covering $\mathcal{U}$ of X by balls of radius $\left(\frac{\varepsilon}{4\|f\|_{C^\alpha}}\right)^{1/\alpha}$, f varies by at most $\frac{\varepsilon}{2}$ inside each ball. Given any $r \in \mathbb{R}$, construct the set $\mathcal{U}_r$ consisting in the union of all balls of $\mathcal{U}$ which intersect X_r. From this, we get a chain of inclusions

$$X_r \hookrightarrow \mathcal{U}_r \hookrightarrow X_{r-\varepsilon}\,, \tag{3.133}$$

which induces a chain of maps at the homology level. More precisely, these inclusions entail that the map $H_*(X_r \hookrightarrow X_{r-\varepsilon})$ factorizes through $H_*(\mathcal{U}_r)$, *i.e.*

$$H_*(X_r \hookrightarrow X_{r-\varepsilon}) = H_*(X_r \hookrightarrow \mathcal{U}_r) \circ H_*(\mathcal{U}_r \hookrightarrow X_{r-\varepsilon})\,. \tag{3.134}$$

In particular, for all $0 \neq [\alpha] \in \mathrm{Im}(H_*(X_r \to X_{r-\varepsilon}))$, there exists a non-trivial cycle in $H_*(\mathcal{U}_r)$ homologous to $[\alpha]$, which is representable as a cycle of the nerve of the minimal covering by virtue of the isomorphism provided by the nerve lemma (lemma 3.25), here applicable by virtue of lemma 3.26 for $\varepsilon \leq 4\,\|f\|_{C^\alpha} r_C^\alpha$. In what follows, we will always identify

$$H_*(\mathcal{U}_r) \cong H_*(\mathcal{N}(\mathcal{U}_r))\,, \tag{3.135}$$

where $\mathcal{N}(\mathcal{U}_r)$ denotes the nerve of the covering $\mathcal{U}_r$ via this isomorphism.

Denote $([\alpha_i])_{1\leq i\leq N_k^\varepsilon}$ the homology classes represented by the N_k^ε bars (noting each of these is itself represented by a persistent cycle α_i). We may in particular order these classes or cycles by their births $b(\alpha_i)$ as follows

$$b(\alpha_{N_k^\varepsilon}) \leq \cdots \leq b(\alpha_1)\,. \tag{3.136}$$

We note that if $b(\alpha_i) = b(\alpha_{i+1}) = \cdots = b(\alpha_k)$, then the $([\alpha_j])_{i\leq j\leq k}$ are independent homology classes and thus the α_j are homologically independent cycles.

We now show that the representations of these cycles as cycles in the k-skeleton of the nerve of the covering are themselves independent, *i.e.* that the family $\{H_k(X_{b(\alpha_i)} \to \mathcal{U}_{b(\alpha_i)})(\alpha_i)\}_{1\leq i\leq m}$ is independent.

By induction, suppose that for a certain i, $b(\alpha_i) < b(\alpha_{i-1})$ and that the family

$$\{H_k(X_{b(\alpha_j)} \to \mathcal{U}_{b(\alpha_j)})(\alpha_j)\}_{1\leq j\leq i} \tag{3.137}$$

is dependent, then

$$H_k(X_{b(\alpha_i)} \to \mathcal{U}_{b(\alpha_i)})(\alpha_i) = \sum_{j=1}^{i-1} c_j\, H_k(X_{b(\alpha_i)} \to \mathcal{U}_{b(\alpha_i)})(\alpha_j)\,. \tag{3.138}$$

Composing both sides by $H_k(\mathcal{U}_{b(\alpha_i)} \to X_{b(\alpha_i)-\varepsilon})$, we have

$$H_k(X_{b(\alpha_i)} \to X_{b(\alpha_i)-\varepsilon})(\alpha_i) = \sum_{j=1}^{i-1} c_j\, H_k(X_{b(\alpha_i)} \to X_{b(\alpha_i)-\varepsilon})(\alpha_j), \tag{3.139}$$

which contradicts that the cycle α_i has persistence $\geq \varepsilon$. Suppose now that $b(\alpha_i) = b(\alpha_{i-1})$ and consider the minimal index $k < i$ for which $b(\alpha_k) = b(\alpha_i)$. We suppose dependence once again, so that after composing with $H_k(\mathcal{U}_{b(\alpha_i)} \to X_{b(\alpha_i)-\varepsilon})$ we get

$$\sum_{j=k}^{i} a_j\, H_k(X_{b(\alpha_i)} \to X_{b(\alpha_i)-\varepsilon})(\alpha_j) = \sum_{j=1}^{k-1} c_j\, H_k(X_{b(\alpha_i)} \to X_{b(\alpha_i)-\varepsilon})(\alpha_j)\,, \tag{3.140}$$

which once again contradicts that the cycles have persistence $\geq \varepsilon$. Since the cycles $(\alpha_j)_{k\leq j\leq i}$ are independent, we conclude that the family up to i is independent. This finishes showing the independence of the representations of the cycles in the k-skeleton of the nerve of the cover.

Since the family of N_k^ε k-cycles is independent in the nerve, we conclude that N_k^ε is bounded above by the cardinatily of the k-skeleton of the nerve of the minimal covering. The M-doubling property of the space yields the upper bound

$$N_k^\varepsilon \leq (M^{k+1} - M^k)\, \mathcal{N}_X\left(\left(\frac{\varepsilon}{4\,\|f\|_{C^\alpha}}\right)^{1/\alpha}\right)\,, \tag{3.141}$$

for $\varepsilon \leq 4\,\|f\|_{C^\alpha}\, r_C^\alpha$. The rest of the proof follows from previous arguments without extra difficulty. ■

From the proof of theorem 3.23, we extract the following useful lemma.

Lemma 3.27. Let X be a compact connected geodesic doubling space whose small enough balls are geodesically convex. Denote $d = \overline{\dim}(X)$, $k \in \mathbb{N}$ and let $f \in C^\alpha(X, \mathbb{R})$ and r_C denotes the convexity radius of X. Then,

$$N^\varepsilon_{H_k(X,f)} \leq (M^{k+1} - M^k)\left[\mathcal{N}_X\left(\left(\frac{\varepsilon}{4\,\|f\|_{C^\alpha}}\right)^{1/\alpha}\right) \vee \mathcal{N}_X\left(r_C\right)\right]$$

$$\mathrm{Pers}_p^p(H_k(X,f)) \leq 4^p\,\|f\|^p_{C^\alpha}\,\alpha p \int_0^{\mathrm{diam}(X)} z^{\alpha p-1}\left[\mathcal{N}_X(z) \vee \mathcal{N}_X\left(r_C\right)\right]\, dz\,.$$

Remark 3.28. If the space is not supposed to be doubling, the only bound we have on N_k^ε is given by $\mathcal{N}_X^{k+1}$, which yields an analogous statement for $\mathcal{L}_k(f) \leq \frac{d(k+1)}{\alpha}$.

Under a supplementary assumption, we can show that the inequality obtained in theorem 3.23 is in fact generically an equality. As before, the genericity result relies on the existence of functions whose Pers_p functional for $p < \frac{d}{\alpha}$ is arbitrarily large. For this we rely on the following theorem of Divol and Polonik.

Theorem 3.29 (Divol and Polonik, [27])**.** *Let μ be a bounded probability measure on $[0,1]^d$ and let $\mathbf{X}_n := (X_1, \cdots, X_n)$ be a vector of i.i.d. samples of μ, then for $0 < p < d$ and $0 \leq k < d$ then almost surely,*

$$\lim_{n\to\infty} n^{-1+\frac{p}{d}} \operatorname{Pers}_p^p(H_k([0,1]^d, d(-,\mathbf{X}_n)) \to \operatorname{Pers}_p^p(\nu_p^\mu)\,, \tag{3.142}$$

for some non-degenerate Radon measure depending on p and the probability measure μ, ν_p^μ on $\mathcal{X}$.

With this result we are now ready to prove the following theorem.

Theorem 3.30. *Let X be a compact Riemannian manifold of dimension d. Then, generically in the sense of Baire in $C^\alpha(X,\mathbb{R})$, for any $0 \leq k < d$, $\mathcal{L}_k(f) = \frac{d}{\alpha}$.*

Proof. The proof of genericity is essentially the same as that of theorem 3.19, with the exception that we now need to modify lemma 3.21. The existence of a function h with arbitrarily large Pers_p-functional for $p < \frac{d}{\alpha}$ on any small ball is given by Divol and Polonik's construction by tweaking the filtration in their proofs from being the distance d to d^α. As before, this entails the genericity result for the set of functions of C^α satisfying $\mathcal{L}_k(f) \geq \frac{d}{\alpha}$. Compact Riemannian manifolds have strictly positive convexity radii and Ricci curvature bounded below, and so satisfy the hypotheses of theorem 3.23, applying the theorem yields the desired equality. ∎

3.3 A partial answer to a question by Schweinhart

In [54], Schweinhart introduces a notion of persistent homology dimension of a metric space X, defined as follows.

Definition 3.31 (Schweinhart's definition of $\dim_{\mathrm{PH}}^k$, [54])**.** Let X be a bounded subset of a metric space. The kth homological dimension of X is

$$\dim_{\mathrm{PH}}^k(X) := \inf_p \{\sup_{\mathbf{x}} \operatorname{Pers}_p(H_k(X, d(-,\mathbf{x}))) < \infty\}\,, \tag{3.143}$$

where the supremum is taken over all finite sets of points $\mathbf{x}$ of X.

Given our previous results, we suggest the following modification to this definition, for reasons which will become apparent later.

Definition 3.32 (kth homological dimension of X)**.** Let X be a bounded subset of a metric space. The kth homological dimension of X is defined as

$$\dim_{\mathrm{PH}}^k(X) := \sup_{f\in \mathrm{Lip}_1(X)} \mathcal{L}_k(f)\,, \tag{3.144}$$

where $\mathrm{Lip}_1(X)$ denotes the set of Lipschitz functions with Lipschitz constant ≤ 1.

Theorem 3.23 already allows us to partially answer Schweinhart's Question 5 [54]. However, this is not a complete answer, because one should make sure that there are Lipschitz functions on X on the class of metric spaces as those of those of theorem 3.23 such that the inequality $\mathcal{L}_k(f) \leq d$ is saturated, or saturated to within δ for all $\delta > 0$. Without the assumption that X is doubling, an interesting question is whether the bound found is optimal: the proof of the theorem suggests that if such metric spaces exist, they cannot be of "bounded geometry" and are relatively pathological.

As we saw in theorem 3.30, this bound is saturated for any integer $0 \leq k < \overline{\dim}(X)$ under the assumption that X is a compact manifold. Thereby entailing

$$\dim_{\mathrm{PH}}^k(X) = \overline{\dim}(X) \tag{3.145}$$

for such X. Here, the notions of homological dimension of Schweinhart and our own coincide exactly, as the genericity result is proven via distance functions to point clouds. This thus establishes sufficient conditions for this equality to hold, albeit not necessary ones.

4 Distance notions and stability properties of trees and diagrams

4.1 Some elements of optimal transport

4.1.1 Defining optimal partial transport

Let us follow the exposition by Divol and Lacombe [25], and quickly introduce optimal *partial* transport, which extends optimal transport to measures of *a priori* different masses (which may be potentially infinite), for a detailed account of the theory, we refer the reader to the cited article, but also to the works of different authors [20, 33, 39]. Divol and Lacombe build on the work of Figalli [34] and extend Wasserstein distances to Radon measures supported on open proper subsets $\mathcal{X}$ of $\mathbb{R}^n$, whose boundary is denoted by $\partial\mathcal{X}$ (and $\overline{\mathcal{X}} := \mathcal{X} \sqcup \partial\mathcal{X}$). The general idea is that we should look at $\partial\mathcal{X}$ as a reservoir of infinite mass, capable of accomodating for any disparity in the mass of the measures considered. In this way, if two Radon measures μ and ν have different mass, we can form still define a transport map from one measure to the other by sending the mass surplus to the boundary $\partial\mathcal{X}$. Symbolically,

Definition 4.1. [34, Problem 1.1] Let $p \in [1, +\infty)$. Let μ, ν be two Radon measures supported on $\mathcal{X}$ satisfying

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

The set of **admissible transport plans** $\Gamma(\mu, \nu)$ is defined as the set of Radon measures π on $\overline{\mathcal{X}} \times \overline{\mathcal{X}}$ satisfying

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

for all Borel sets $A, B \subset \mathcal{X}$. Furthermore, the cost of $\pi \in \Gamma(\mu, \nu)$ is defined as

$$C_p(\pi) := \int_{\overline{\mathcal{X}} \times \overline{\mathcal{X}}} d(x, y)^p \, d\pi(x, y). \tag{4.146}$$

The **optimal transport distance** $d_p(\mu, \nu)$ is defined as

$$d_p(\mu, \nu) := \left(\inf_{\pi \in \Gamma(\mu,\nu)} C_p(\pi) \right)^{1/p}. \tag{4.147}$$

Plans $\pi \in \Gamma(\mu, \nu)$ realizing the infimum in equation 4.147 are called **optimal**.

Definition 4.2. The space of Radon measures on $\mathcal{X}$ will be denoted $\mathcal{D}(\mathcal{X})$ (or simply $\mathcal{D}$ if $\mathcal{X}$ is clear from context). We also introduce the following spaces

$$\mathcal{D}_p := \left\{ \mu \in \mathcal{D} \,\middle|\, \int_{\mathcal{X}} d^p(x, \partial\mathcal{X}) \, d\mu(x) < \infty \right\}. \tag{4.148}$$

We further define $\mathcal{D}_\infty$ as the space of Radon measures with compact support.

Remark 4.3. A proof by Théo Lacombe shows that for optimal partial transport distances d_p also satisfy $d_p \xrightarrow{p\to\infty} d_\infty$. Indeed, for any $\pi \in \Gamma(\mu, \nu)$

$$C_p(\pi) \xrightarrow{p\to\infty} C_\infty(\pi) \tag{4.149}$$

The space $\Gamma(\mu, \nu)$ is sequentially compact [25, Proposition 3.2], so up to extraction of a subsequence, $(\pi_p)_p$ admits a limit π_∞. Finally, if π^* is an optimal transport for the cost function C_∞, then

$$C_\infty(\pi^*) = \lim_{p\to\infty} C_p(\pi^*) \geq \lim_{p\to\infty} C_p(\pi_p) = C_\infty(\pi_\infty), \tag{4.150}$$

so π_∞ also achieves $\inf_\pi C_\infty(\pi)$, showing the desired result.

When considering optimal partial transport, there may be complications with respect to the conventional theory of optimal transport, because the measures may have infinite mass. This poses some problems, among others because of the unavalaibility of Jensen's inequality, which may render certain results of the classical theory false, or require alternative proofs. Luckily, most classical results we will need can be adapted to this more general setting.

4.1.2 Some results on optimal transport distances

To distinguish the theory of optimal transport from that of optimal *partial* transport, let us introduce the following notation.

Notation 4.4. Let (X, δ) be a Polish metric space. Denote $\mathcal{P}(X)$ (or simply $\mathcal{P}$ is X is clear from context) the set of **probability measures on** X and define

$$\mathcal{P}_p(X) := \left\{ \mu \in \mathcal{P} \,\middle|\, \int_X \delta^p(x, x_0)\, d\mu(x) < \infty \right\} \tag{4.151}$$

for some $x_0 \in X$ (note that this definition does not depend on x_0). Once again, we may omit X if it is clear from context. For any two measures $\mu, \nu \in \mathcal{P}(X)$, slightly abusing then notation, we may define the **space of transport maps** $\Gamma(\mu, \nu)$ to be the space of probability measures on X^2 having marginals μ and ν. We equip the space $\mathcal{P}_p(X)$ with a Wasserstein distance, defined as

$$W_{p,\delta}(\mu, \nu) := \inf_{\pi \in \Gamma(\mu,\nu)} \|\delta\|_{L^p(\pi)}\,. \tag{4.152}$$

For the rest of this paper, the distance indicated by W will always reserved to classical Wasserstein distances between *probability* measures, whereas the distance denoted d_p will always refer to the notion of Wasserstein distances between general Radon measures, previously described in the context of optimal *partial* transport.

Many statements are valid whether we are in the optimal transport or the optimal *partial* transport setting. For this reason, we introduce the following generic notation along with the following dictionary to transpose statements to one setting or another.

Generic notation	Optimal transport	Optimal partial transport
(Y, d)	(X, δ)	$(\mathcal{X}, d)$
∂Y	$x_0 \in X$	$\partial \mathcal{X}$
OT_p	$W_{p,\delta}$	d_p
$\mathcal{M}(Y)$	$\mathcal{P}(X)$	$\mathcal{D}(\mathcal{X})$
$\mathcal{M}_p(Y)$	$\mathcal{P}_p(X)$	$\mathcal{D}_p(\mathcal{X})$

Table 1: Dictionary between optimal and optimal partial transport.

Proposition 4.5. For any $1 \le p < \infty$, OT_p^p is convex, in the sense that for every $\mu_1, \mu_2, \nu \in \mathcal{M}_p$ and $t \in [0, 1]$,

$$\mathrm{OT}_p^p(t\mu_1 + (1-t)\mu_2, \nu) \le t\,\mathrm{OT}_p^p(\mu_1, \nu) + (1-t)\,\mathrm{OT}_p^p(\mu_2, \nu)\,. \tag{4.153}$$

Moreover, if $\nu_1, \nu_2 \in \mathcal{M}_p$,

$$\mathrm{OT}_p^p(t\mu_1 + (1-t)\mu_2, t\nu_1 + (1-t)\nu_2) \le t\,\mathrm{OT}_p^p(\mu_1, \nu_1) + (1-t)\,\mathrm{OT}_p^p(\mu_2, \nu_2)\,. \tag{4.154}$$

Proof. For every $\pi_i \in \Gamma(\mu_i, \nu)$, $t\pi_1 + (1-t)\pi_2 \in \Gamma(t\mu_1 + (1-t)\mu_2, \nu)$, so

$$\mathrm{OT}_p^p(t\mu_1 + (1-t)\mu_2, \nu) \le t \int_{Y^2} d(x, y)^p\, d\pi_1(x, y) + (1-t) \int_{Y^2} d(x, y)^p\, d\pi_2(x, y)\,, \tag{4.155}$$

which yields the result by taking the infimum over π_1 and π_2 on the right-hand side. The second convexity result is obtained by an analogous proof. ■

Remark 4.6. Convexity does not hold for $p = \infty$. By taking the $\frac{1}{p}$-th power of both sides and letting $p \to \infty$ in the inequality above, all that we may conclude is that

$$\mathrm{OT}_\infty(t\mu_1 + (1-t)\mu_2, \nu) \leq \max\{\mathrm{OT}_\infty(\mu_1, \nu), \mathrm{OT}_\infty(\mu_2, \nu)\}\,. \tag{4.156}$$

Theorem 4.7 (OT_p for $p = \infty$, [36])**.** *The distance obtained on $\mathcal{M}_\infty(Y)$ from OT_p by taking $p \to \infty$ is well-defined and coincides with the distance defined by*

$$\mathrm{OT}_\infty(\mu, \nu) = \inf_{\pi \in \Gamma(\mu,\nu)} \|d\|_{L^\infty(\pi)}\,. \tag{4.157}$$

Furthermore, we have the following characterization of OT_∞

$$\mathrm{OT}_\infty(\mu, \nu) = \inf\{r > 0 \mid \forall U \subset Y \text{ open, } \mu(U) \leq \nu(U^r) \text{ and } \nu(U) \leq \mu(U^r)\}\,, \tag{4.158}$$

where U^r denotes an open tubular neighborhood of radius r around U.

Remark 4.8. The topology of OT_∞ is finer than that of weak convergence.

Proposition 4.9. Let $f : (Y, \delta) \to (Y', \delta')$ be an α-Hölder map with Hölder constant Λ and let $\mu, \nu \in \mathcal{P}(Y)$ then

$$W_{p,\delta'}(f_\sharp\mu, f_\sharp\nu) \leq \Lambda\, W^\alpha_{p\alpha,\delta}(\mu, \nu)\,. \tag{4.159}$$

Proof. The inequality is an immediate consequence of the Hölder continuity of f. ∎

4.1.3 Persistence measures

Coming back to persistence theory, recall that it is possible to see persistence diagrams as measures on

$$\mathcal{X} := \{(x, y) \in \mathbb{R}^2 \mid y > x\}\,. \tag{4.160}$$

Henceforth, $\mathcal{X}$ will always refer to this half space. Seen as measures, persistence diagrams are nothing other than a sum of Dirac measures. Closing this space with respect to the topology of vague convergence, we retrieve the set of Radon measures on $\mathcal{X}$.

Definition 4.10. The set of **persistence measures** $\mathcal{D}$ is the set of Radon measures (of potentially infinite mass) on $\mathcal{X} := \{(x, y) \in \mathbb{R}^2 \mid y > x\}$.

Equipping $\mathcal{X}$ with the ℓ^∞-distance on $\mathbb{R}^2$ defined by

$$d((p, q), (r, s)) = \max\{|p - r|\,, |q - s|\}\,, \tag{4.161}$$

optimal partial transport distances d_p between persistence measures become definable. The repercussions of this have been explored by Divol and Lacombe in [25].

The extension from the space of persistence diagrams to the space of persistence measures has three main advantages. First, that, as shown in [25], it is possible to use the machinery of optimal transport to address problems in persistence theory. Second, that $\mathcal{D}$ is a linear space, which renders taking means and combinations of diagrams possible and easy. Finally, that it is well-adapted to the stochastic setting, because of the linearity property and Tonelli's theorem: two key properties which we will exploit repeatedly.

Remark 4.11. The notion of average as defined in the linear space of persistence measures in general exits the space of persistence diagrams. This can for instance be seen by considering a sequence of measures which vaguely tend to a measure which is absolutely continuous with respect to the Lebesgue measure on $\mathcal{X}$. In this case, it is impossible to reconstruct a function whose diagram agrees with the desired measure. This is obvious in the 1D case where it is impossible to construct any tree from such a persistence measure, and so by extension, to construct any function. Nonetheless, this notion of average has the advantage of encoding the averages of all linear functionals of the diagrams (one can in fact see this as a definition of this notion of average by adopting a dual point of view). Some authors have considered alternative notions of central tendencies adapted to metric spaces (and in particular the space of diagrams), such as Fréchet means defined on the spaces of diagrams (*cf.* for instance the work of Turner *et al.* [57]). While this notion stays in the space where persistence diagrams are defined, it depends on the distance chosen on $\mathcal{D}$ and moreover also on the exponent chosen for the cost function in the definition of Fréchet means.

4.2 Stability of Wasserstein p-distances on diagrams

With respect to optimal transport distances, we have some "stability theorems" the most classical of which is

Theorem 4.12 (Bottleneck stability with respect to L^∞, Corollary 3.6 [48])**.** *Let $f, g : X \to \mathbb{R}$ be two continuous functions, then*

$$d_\infty(\mathrm{Dgm}(f), \mathrm{Dgm}(g)) \leq \|f - g\|_\infty \tag{4.162}$$

where $\mathrm{Dgm}(f)$ and $\mathrm{Dgm}(g)$ denote the diagrams of f and g respectively.

Theorem 4.13 (Wasserstein p stability)**.** *Let X be a compact LLC metric space of $\overline{\dim}(X) = d$ and consider $f, g \in C^\alpha(X, \mathbb{R})$. Then, for all $p > q > \frac{d}{\alpha}$,*

$$d_p^p(H_0(X, f), H_0(X, g)) \leq C_{X,\alpha} \left(\|f\|_{C^\alpha} \vee \|g\|_{C^\alpha}\right)^q \|f - g\|_\infty^{p-q} \,. \tag{4.163}$$

If X is assumed to be geodesic and is such that small enough balls of X are geodesically convex, then for every $k \in \mathbb{N}^*$, and all $p > q > \frac{d(k+1)}{\alpha}$

$$d_p^p(H_k(X, f), H_k(X, g)) \leq C_{X,\alpha,k} \left(\|f\|_{C^\alpha} \vee \|g\|_{C^\alpha}\right)^q \|f - g\|_\infty^{p-q} \,. \tag{4.164}$$

Finally, if X is further supposed to be doubling, then the inequality above holds for all $p > q > \frac{d}{\alpha}$.

Proof. Let $\Lambda = (\|f\|_{C^\alpha} \vee \|g\|_{C^\alpha})$. The first part of the proof is essentially as in [21]. Start by picking the bottleneck matching between the diagrams of f and g and denote it by $\gamma : \mathrm{Dgm}(f) \to \mathrm{Dgm}(g)$. Then for any $p > q > \frac{d}{\alpha}$,

$$\begin{aligned} d_p^p(\mathrm{Dgm}(f), \mathrm{Dgm}(g)) &\leq \sum_{b \in \mathrm{Dgm}(f)} d_{\mathcal{X},\infty}(b, \gamma(b))^p \\ &\leq \|f - g\|_\infty^{p-q} \sum_{b \in \mathrm{Dgm}(f)} d_{\mathcal{X},\infty}(b, \gamma(b))^q \\ &\leq 2^q \|f - g\|_\infty^{p-q} \sum_{b \in \mathrm{Dgm}(f)} d_{\mathcal{X},\infty}(b, \Delta)^q + d_{\mathcal{X},\infty}(\gamma(b), \Delta)^q \\ &= 2^q \|f - g\|_\infty^{p-q} \left(\mathrm{Pers}_q^q(f) + \mathrm{Pers}_q^q(g)\right) \end{aligned}$$

But both $\mathrm{Pers}_q^q(f)$ and $\mathrm{Pers}_q^q(g)$ are bounded above by a global constant for the class $C_\Lambda^\alpha(X, \mathbb{R})$, since by the proof of lemma 3.20

$$N_f^\varepsilon \leq \mathcal{N}_X\left(\left(\frac{\varepsilon}{2C\Lambda}\right)^{1/\alpha}\right) , \tag{4.165}$$

where C is a constant stemming from the quantitative LLC condition on X. This inequality entails that

$$\begin{aligned} \mathrm{Pers}_q^q(f) = q \int_0^\infty \varepsilon^{q-1} N_f^\varepsilon \, d\varepsilon &\leq q \int_0^{\Lambda \,\mathrm{diam}(X)^\alpha} \varepsilon^{q-1} \mathcal{N}_X\left(\left(\frac{\varepsilon}{2C\Lambda}\right)^{1/\alpha}\right) \, d\varepsilon \\ &= (2C\Lambda)^q \alpha q \int_0^{\frac{\mathrm{diam}(X)}{(2C)^{1/\alpha}}} \varepsilon^{q\alpha - 1} \mathcal{N}_X(\varepsilon) \, d\varepsilon \,, \end{aligned}$$

which is finite as soon as $q > \frac{d}{\alpha}$ since $\mathcal{N}_X(\varepsilon) = O(\varepsilon^{-d-\delta})$ as $\varepsilon \to 0$ for all $\delta > 0$, by definition of the upper-box dimension. The constant in the statement of the theorem is bounded above by the above estimate. The statements for with the supplementary assumptions of the theorem, the proof follows from the same reasononing by using the proof of theorem 3.23 and remark 3.28 ∎

Remark 4.14. More generally, the proof of the theorem adapts with ease to accomodate any compact set of $C^0(X, \mathbb{R})$ admitting a global modulus of continuity dominated by a Hölder modulus of continuity. It is worth mentioning that such a theorem is impossible to prove for any regularity strictly worse than Hölder, as in such a class of regularity, there are functions f of infinite persistence index, so the theorem is vacuous.

Wasserstein stability results are common in the literature and are typically stated by making the following assumption on the underlying metric space X.

Definition 4.15. [21] A metric space X **implies bounded q-total persistence** if, for all $k \in \mathbb{N}$, there exists a constant C_X that depends only on X such that

$$\mathrm{Pers}_q^q(H_k(X, f)) < C_X \tag{4.166}$$

for every tame function f with Lipschitz constant $\mathrm{Lip}(f) \leq 1$.

The regime of validity of Wasserstein stability thus depends solely on this condition on X. We can thus see theorem 4.13 as a theorem giving explicit bounds on the q such that X implies bounded q-total persistence (in fact, it does so for every degree in homology independently). Following [21], it follows clearly from the proof of Wasserstein stability that this definition implies bounded persistence stability for Lipschitz functions.

Corollary 4.16. *Let X be a compact LLC metric space of $\overline{\dim}(X) = d$ of LLC constant C. Then, for all $f \in \mathrm{Lip}_1(X)$ and $p > q > d$,*

$$\mathrm{Pers}_q^q(H_0(X, f)) \leq (2C)^q q \int_0^{\frac{diam(X)}{2C}} \varepsilon^{q-1} \mathcal{N}_X(\varepsilon)\, d\varepsilon\,. \tag{4.167}$$

If X is assumed to be geodesic and is such that small enough balls of X are geodesically convex, then for every $k \in \mathbb{N}^*$, and all $p > q > d(k+1)$

$$\mathrm{Pers}_q^q(H_k(X, f)) \leq 4^q q \int_0^{diam(X)} \varepsilon^{q-1} (\mathcal{N}_X(\varepsilon) \vee \mathcal{N}_X(r_C))^k\, d\varepsilon\,, \tag{4.168}$$

where r_C denotes the convexity radius of X. Finally, if X is further supposed to be M-doubling, then for all $p > q > d$,

$$\mathrm{Pers}_q^q(H_k(X, f)) \leq 4^q q (M^{k+1} - M^k) \int_0^{diam(X)} \varepsilon^{q-1} (\mathcal{N}_X(\varepsilon) \vee \mathcal{N}_X(r_C))\, d\varepsilon\,. \tag{4.169}$$

Some other Wasserstein p stability results have been reported in the literature: Chen and Edelsbrunner [18] studied functions on non-compact domains of $\mathbb{R}^d$, obtaining a stability result which holds for $p > d$. The condition $p > d$ also appears in stability results for Čech filtrations for point clouds in $\mathbb{R}^d$ and the case of Vietoris-Rips filtrations was recently addressed in [55] by Skraba and Turner.

4.3 Distance notion and stability for trees

Definition 4.17. Let X and Y be two compact metric spaces, the **Gromov-Hausdorff distance,** $d_{GH}(X, Y)$ between X and Y, is defined as

$$d_{GH}(X, Y) := \inf_{\substack{f: X \to Z \\ g: Y \to Z}} \max \left\{ \sup_{x \in X} \inf_{y \in Y} d_Z(f(x), g(y)), \sup_{y \in Y} \inf_{x \in X} d_Z(f(x), g(y)) \right\}. \tag{4.170}$$

where the infimum is taken over all metric spaces Z and all isometric embeddings $f : X \to Z$ and $g : Y \to Z$.

The Gromov-Hausdorff distance quantifies how far away two metric spaces X and Y are from being isometric to each other. However, it is practically impossible to compute this distance with the above definition. To somewhat alleviate this, we will use the following characterization of the Gromov-Hausdorff distance:

Proposition 4.18 (Burago *et al.*, §7 [11])**.** The Gromov-Hausdorff distance is characterized by

$$d_{GH}(X,Y) = \frac{1}{2}\inf_{\mathfrak{R}} \sup_{\substack{(x,y)\in\mathfrak{R}\\(x',y')\in\mathfrak{R}}} \left|d_X(x,x') - d_Y(y,y')\right| , \tag{4.171}$$

where the infimum is taken over all *correspondences*, *i.e.* subsets $\mathfrak{R} \subset X \times Y$ such that for every $x \in X$ there is at least one $y \in Y$ such that $(x,y) \in \mathfrak{R}$ and a symmetric condition for every $y \in Y$.

Remark 4.19. Given two surjective maps $\pi_X : Z \to X$ and $\pi_Y : Z \to Y$, it is possible to build a correspondence between X and Y by considering the set $\{(\pi_X(z), \pi_Y(z)) \in X \times Y \,|\, z \in Z\}$.

A natural question is to ask whether we have an equivalent statement about the stability of d_{GH} with respect to $\|\cdot\|_{L^\infty}$ and whether the two notions of distances are in some sense "compatible". We will positively answer this first question. In general d_∞ and d_{GH} are not compatible, in the sense that no inequality between the two holds in all generality (*cf.* remark 4.22). Le Gall and Duquesne [28] gave a first stability result of d_{GH} with respect to the L^∞-norm on continuous functions on $[0,1]$:

Theorem 4.20 (L^∞-stability of trees, [28])**.** *Let $f, g : [0,1] \to \mathbb{R}$ be two continuous functions. Then*

$$d_{GH}(T_f, T_g) \le 2\,\|f-g\|_{L^\infty} . \tag{4.172}$$

This result for functions on $[0,1]$ generalizes to more general topological spaces.

Theorem 4.21 (Stability theorem for trees)**.** *Let X be a compact, connected and locally path connected topological space and let f and $g : X \to \mathbb{R}$ be two continuous functions, then*

$$d_{GH}(T_f, T_g) \le 2\,\|f-g\|_{L^\infty} . \tag{4.173}$$

Proof. We will use the distortion characterization of the Gromov-Hausdorff distance, which yields the following inequality

$$d_{GH}(T_f, T_g) \le \frac{1}{2}\sup_{x,y\in X} |d_f(x,y) - d_g(x,y)| . \tag{4.174}$$

Following the logic of the proof of lemma 3.15, the distance between $\pi_f(x)$ and $\pi_f(y)$ is of the form

$$d_f(\pi_f(x),\tau) + d_f(\tau,\pi_f(y)) = f(x) - f(\tau) + f(y) - f(\tau) \tag{4.175}$$

where τ is the lowest point of the geodesic path in T_f between $\pi_f(x)$ and $\pi_f(y)$. This geodesic path on T_f admits preimages by π_f which are paths connecting x to y. These paths achieve the following supremum

$$\sup_{\gamma : x\mapsto y}\inf_{t\in[0,1]} f\circ\gamma = f(\tau) \le f(x)\wedge f(y) \tag{4.176}$$

where $a \wedge b := \min\{a,b\}$ since by construction γ must always stay above $f(\tau)$ and since for $r > f(\tau)$, x and y lie in different connected components of X_r. If ν is the analogous vertex to τ on T_g,

$$\begin{aligned}
d_{GH}(T_f,T_g) &\le \frac{1}{2}\sup_{x,y\in X}|d_f(x,y) - d_g(x,y)| \\
&= \frac{1}{2}\sup_{x,y\in X}|f(x) - g(x) + f(y) - g(y) - 2f(\tau) + 2g(\nu)| \\
&\le \|f-g\|_{L^\infty} + \sup_{x,y\in X}\left|\sup_{\gamma:x\mapsto y}\inf_{t\in[0,1]} f\circ\gamma - \sup_{\eta:x\mapsto y}\inf_{t\in[0,1]} g\circ\eta\right| \\
&\le 2\,\|f-g\|_{L^\infty} ,
\end{aligned} \tag{4.177}$$

as desired. ∎

Remark 4.22. One can be tempted to establish a general inequality between d_{GH} and d_∞ since both of these distances are bounded by the L^∞-norm. However, this is not possible.

Indeed, there is a simple counter-example to $d_{GH} \geq d_\infty$. To illustrate this consider two barcodes over a field k, $k[s, -\infty[$ and $k[s+\varepsilon, -\infty[$. The bottleneck distance between these two is clearly $\geq \varepsilon$. But supposing that the functions f and g generating these barcodes are such that $f = g + \varepsilon$ the trees T_g and T_f are isometric, so $d_{GH}(T_f, T_g) = 0 < \varepsilon \leq d_\infty(\mathcal{B}(f), \mathcal{B}(g))$.

Conversely, there are also counter-examples to $d_\infty \geq d_{GH}$, as this inequality would imply that two trees which have the same barcode are isometric. This is clearly false, as one can "glue" the bars of a given barcode is many different ways to give a tree, which generically will not be isometric.

5 Stochastic processes

As we have previously seen, the study of diagrams of continuous functions involves understanding their regularity. Many stochastic processes are *almost* Hölder continuous in the following sense.

Definition 5.1. The class of **almost α-Hölder continuous functions from X to $\mathbb{R}$**, denoted $E^\alpha(X, \mathbb{R})$ is the class of functions defined by

$$E^\alpha(X, \mathbb{R}) := \bigcap_{0 \leq \beta < \alpha} C^\beta(X, \mathbb{R}) \tag{5.178}$$

For example, Brownian motion and fractional Brownian motion are in a certain E^α for some value of α and moreover, as shown by Kahane [38, Chapter 7], random subgaussian Fourier series on torii of any dimension also tend to have E^α regularities. The ubiquity of E^α-regularities in the context of stochastic processes partially motivate this definition.

Notation 5.2. In what will follow, we will denote $f_\sharp \mathbb{P}$ the pushforward measure of $\mathbb{P}$ by f.

5.1 A change in perspective

Remark 5.3. Slightly abusing the notation, throughout this section, when we talk about a (continuous) stochastic process, we will talk about a measurable function $f : \Omega \to C^0(X, \mathbb{R})$ (where $(\Omega, \mathcal{F}, \mathbb{P})$ is some probability space).

Random diagrams, or more precisely, probability measures on the space of diagrams (or on the space of persistence measures) have been studied under many different contexts in the persistence theory literature [15, 17, 26, 32, 57]. Since ultimately we are interested in studying random processes on some base space X, the space of probability measures on the space of diagrams is far too large, as not all diagrams stem from (continuous) functions. In all practical applications, we are never given an abstract persistence diagram. Rather, we compute the persistence diagram from a certain continuous function (on which we may postulate further regularity assumptions, typically that the function is inside some $E^\alpha(X, \mathbb{R})$). This motivates studying subspaces of the full space of persistence diagrams of the form $\cup_k \operatorname{Dgm}_k(E^\alpha(X, \mathbb{R})) \subset \mathcal{D}$. This perspective turns out to have notable advantages. For instance, it is known that $(\mathcal{D}, d_\infty)$ is not a separable space [10, Theorem 5], but adopting this point of view we can show the opposite.

Proposition 5.4. Let $K \subset (C^0(X, \mathbb{R}), \|\cdot\|_{L^\infty})$, be a closed subset, then $(\overline{\operatorname{Dgm}(K)}, d_\infty)$ is a Polish metric space.

Proof. We start by noticing that the map Dgm is continuous and that the continuous image of a separable metric space is separable [59, Theorem 16.4a]. Moreover, $\overline{\operatorname{Dgm}(K)}$ remains separable, since the countable dense subset of $\operatorname{Dgm}(K)$ remains dense in the completion. ■

Remark 5.5. If the subset K is compact, then $\operatorname{Dgm}(K) = \overline{\operatorname{Dgm}(K)}$. Notice also that the compact subsets of $C^0(X, \mathbb{R})$ are sets having a uniform modulus of continuity, by virtue of Ascoli's theorem. In particular, spaces such as $C^\alpha_\Lambda(X, \mathbb{R})$ are compact.

Consider now continuous $\mathbb{R}$-valued stochastic processes on X, f, defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, the space of probability measures on diagrams is also too large, as the probability measures we are concerned with must be of the form $(\mathrm{Dgm}_k \circ f)_\sharp \mathbb{P}$. For convenience, we could take the closure of this space induced by measures of this form with respect to the topology of vague convergence, or with respect to some Wasserstein distance $W_{p,\delta}$ (on the space of probability measures on diagrams). This is a technical point, but allows us to avoid making hypotheses on the probability measures on the space of diagrams, which are in practice almost never verifiable, and instead give hypotheses on the stochastic processes from which the diagrams stem from.

This point of view is particularly well-suited to look at stochastic processes supported on compact subsets of $C^0(X, \mathbb{R})$ (in fact, $E^\alpha(X, \mathbb{R})$, for reasons which will become apparent later). An easy first result in this direction is that

Proposition 5.6. Let K be a compact subset of $C^0(X, \mathbb{R})$, then $\mathrm{Dgm}_k(K) \subset \mathcal{D}_\infty$.

This restriction to compact sets can be seen as a considerable limitation. For example, Brownian motion on the interval $[0, 1]$ does not satisfy this hypothesis of compactness. However, by virtue of the tightness of probability measures on $C^0(X, \mathbb{R})$, we may restrict ourselves to a compact K_ε of $C^0(X, \mathbb{R})$ in which the process lies with probability $1 - \varepsilon$ and make probable statements there, or, alternatively, make conditional statements.

Furthermore,

Proposition 5.7. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and f be a $\mathbb{R}$-valued, a.s. E^α stochastic process on a d-dimensional compact manifold X. Then, for all $\varepsilon > 0$, $(\mathrm{Dgm}_k \circ f)_\sharp \mathbb{P} \in \mathcal{P}(\mathcal{D}_{\frac{d}{\alpha}+\varepsilon} \cap \mathcal{D}_\infty)$ and *a fortiori* in $\mathcal{P}(\mathcal{D}_r)$ for every $\frac{d}{\alpha} < r < \infty$. Furthermore, if $\frac{d}{\alpha} < q < \infty$ and for all $\beta < \alpha$, $\mathbb{E}\left[\|f\|^q_{C^\beta(X,\mathbb{R})}\right] < \infty$, then, $\mathbb{E}[\mathrm{Dgm}_k(f)] \in \bigcap_{\frac{d}{\alpha} < p \leq q} \mathcal{D}_p$.

Proof. Since $f \in E^\alpha(X, \mathbb{R})$ a.s., it is a.s. $C^\beta(X, \mathbb{R})$ for every $\beta < \alpha$, and so a.s. bounded by compactness of X. By theorem 3.23 and the previous remark, it follows that for every $k \in \mathbb{N}$, $\mathrm{Dgm}_k(f) \in \mathcal{D}_{\frac{d}{\beta}} \cap \mathcal{D}_\infty$, proving the first result.

Next, we remark that if $\mathbb{E}\left[\|f\|^q_{C^\beta(X,\mathbb{R})}\right]$ is finite so is the pth moment of the norm for every $1 \leq p \leq q$ by a simple application of Jensen's inequality. To show the result, it suffices to show that for such p,

$$\mathbb{E}\left[\mathrm{Pers}_p^p(f)\right] < \infty \,. \tag{5.179}$$

But applying Tonelli's theorem and using lemma 3.27,

$$\begin{aligned} \mathrm{Pers}_p^p(f) &= p \int_0^\infty \varepsilon^{p-1} \mathbb{E}\left[N_f^\varepsilon\right] \, d\varepsilon \\ &\leq 4^p \beta p \, \|f\|^p_{C^\beta} \int_0^{\mathrm{diam}(X)} \varepsilon^{p\beta - 1} [\mathcal{N}_X(\varepsilon) \vee \mathcal{N}_X(r_C)] \, d\varepsilon \,. \end{aligned}$$

The integral on $[0, 1]$ is finite as soon as $p > \frac{d}{\alpha}$ since the dimension of X is d. Taking the expectation of both sides,

$$\mathbb{E}\left[\mathrm{Pers}_p^p(f)\right] \leq \tilde{C}_{X,p,\beta} \mathbb{E}\left[\|f\|^p_{C^\beta}\right] \,, \tag{5.180}$$

which is finite as soon as the moments of the C^β-norm of f are finite, exactly as supposed in the proposition. Finally, the *a fortiori* inclusion in $\mathcal{D}_r$ is a consequence of the Wasserstein interpolation theorem [50]. ■

Corollary 5.8. *With the same hypotheses for f and p as in the previous proposition, for every $r \geq 1$ and every $\beta < \alpha$,*

$$\mathbb{E}\left[\mathrm{Pers}_p^r(f)\right] \leq C_{X,p,\beta} \mathbb{E}[\|f\|^r_{C^\beta}] \tag{5.181}$$

5.2 Consequences of stability

Equipped with some of the elementary facts from optimal transport theory, we may come back to persistence measures and diagrams. The main goal of this section will be to prove the following theorem.

Theorem 5.9 (Stability of random fields under Wasserstein perturbations). *Let f and g be two $\mathbb{R}$-valued a.s. E^α stochastic processes on a d dimensional compact Riemannian manifold X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, for any $k \in \mathbb{N}$ and any $1 \leq p \leq \infty$,*

$$W_{p,d_\infty}((\mathrm{Dgm}_k \circ f)_\sharp \mathbb{P}, (\mathrm{Dgm}_k \circ g)_\sharp \mathbb{P}) \leq W_{p,L^\infty}(f_\sharp \mathbb{P}, g_\sharp \mathbb{P}) \,. \tag{5.182}$$

Moreover, for every $\frac{d}{\alpha} < q < p < \infty$ and any $r, s \in]1, \infty[$ satisfying $\frac{1}{r} + \frac{1}{s} = 1$ and $(p-q)s \geq 1$, there exists a constant C_X depending only on X such that

$$\begin{aligned} d_p(\mathbb{E}[\mathrm{Dgm}_k(f)] \,, \mathbb{E}[\mathrm{Dgm}_k(g)]) &\leq W_{p,d_p}((\mathrm{Dgm}_k \circ f)_\sharp \mathbb{P}, (\mathrm{Dgm}_k \circ g)_\sharp \mathbb{P}) && (5.183)\\ &\leq C_X \left[\mathbb{E}\left[\|f\|_{C^\beta}^{qr}\right]^{\frac{1}{r}} + \mathbb{E}\left[\|g\|_{C^\beta}^{qr}\right]^{\frac{1}{r}} \right]^{\frac{1}{p}} W_{(p-q)s,\infty}^{1-\frac{q}{p}}(f_\sharp \mathbb{P}, g_\sharp \mathbb{P}) && (5.184)\\ &\leq C_X \left[\mathbb{E}\left[\|f\|_{C^\beta}^{qr}\right]^{\frac{1}{r}} + \mathbb{E}\left[\|g\|_{C^\beta}^{qr}\right]^{\frac{1}{r}} \right]^{\frac{1}{p}} \|f-g\|_{L^{(p-q)s}(\Omega, L^\infty(X,\mathbb{R}))}^{1-\frac{q}{p}} \,. && (5.185) \end{aligned}$$

Finally, if the supports of $f_\sharp \mathbb{P}$ and $g_\sharp \mathbb{P}$ are compact in $E^\alpha(X, \mathbb{R})$, then

$$d_\infty(\mathbb{E}[\mathrm{Dgm}_k(f)] \,, \mathbb{E}[\mathrm{Dgm}_k(g)]) \leq W_{\infty,d_\infty}((\mathrm{Dgm}_k \circ f)_\sharp \mathbb{P}, (\mathrm{Dgm}_k \circ g)_\sharp \mathbb{P}) \,. \tag{5.186}$$

Remark 5.10. The proof of this theorem uses some of the techniques from [17, Lemma 15]. It differs from this result, as it concerns the d_p-stability as opposed to simply d_∞-stability, but also because the statement of theorem 5.9 gives a bound on the distance between expected diagrams, as opposed to a linear functional of the latter. However, necessary and sufficient conditions for the continuity of linear functionals of $\mathbb{E}[\mathrm{Dgm}(f)] \in (\mathcal{D}_p, d_p)$ has been studied by Divol and Lacombe in [25].

Proof of theorem 5.9. The first inequality is a simple consequence of a change of variables and an application of the bottleneck stability theorem. Next, notice that if $f_\sharp \mathbb{P}$ and $g_\sharp \mathbb{P}$ have compact support in E^α, then f and g are almost surely uniformly bounded functions, so $\mathbb{E}[\mathrm{Dgm}(f)]$ and $\mathbb{E}[\mathrm{Dgm}(g)]$ are both in $\mathcal{D}_\infty$.

Notice that,

$$\mathbb{E}[\mathrm{Dgm}(f)] = \int_{E^\alpha} \mathrm{Dgm}(h) \; df_\sharp \mathbb{P}(h) = \int_{(E^\alpha)^2} \mathrm{Dgm}(h) \; d\pi(h, \tilde{h}) \,, \tag{5.187}$$

for any $\pi \in \Gamma(f_\sharp \mathbb{P}, g_\sharp \mathbb{P})$ and an analogous equality holds for $\mathbb{E}[\mathrm{Dgm}(g)]$. Since d_p^p is convex, applying Jensen's inequality

$$\begin{aligned} d_p^p(\mathbb{E}[\mathrm{Dgm}(f)] \,, \mathbb{E}[\mathrm{Dgm}(g)]) &= d_p^p \left(\int_{(E^\alpha)^2} \mathrm{Dgm}(h) \; d\pi(h, \tilde{h}), \int_{(E^\alpha)^2} \mathrm{Dgm}(\tilde{h}) \; d\pi(h, \tilde{h}) \right) \\ &\leq \int_{(E^\alpha)^2} d_p^p(\mathrm{Dgm}(h), \mathrm{Dgm}(\tilde{h})) \; d\pi(h, \tilde{h}) \\ &= \int_{(\mathrm{Dgm}(E^\alpha))^2} d_p^p(x, y) \; d\mathrm{Dgm}_\sharp^{\otimes 2} \pi(x, y) \,. \end{aligned}$$

Taking the infimum over every π of this inequality and taking the pth root,

$$d_p(\mathbb{E}[\mathrm{Dgm}(f)] \,, \mathbb{E}[\mathrm{Dgm}(g)]) \leq W_{p,d_p}((\mathrm{Dgm} \circ f)_\sharp \mathbb{P}, (\mathrm{Dgm} \circ g)_\sharp \mathbb{P}) \,.$$

Under the hypothesis of compactness, the result for $p = \infty$ is obtained by taking the limit $p \to \infty$, justified by remark 4.3 and the fact that the stochastic processes and their distributions in E^α are uniformly bounded.

Going back to the non-compact setting, keeping the same notation, let $\pi \in \Gamma((\mathrm{Dgm} \circ f)_\sharp \mathbb{P}, (\mathrm{Dgm} \circ g)_\sharp \mathbb{P})$ be an optimal transport. For any $\beta < \alpha$, applying the Wasserstein p stability theorem, for all $p > q > \frac{d}{\beta}$,

$$\int_{(E^\alpha)^2} d_p^p(\mathrm{Dgm}(h), \mathrm{Dgm}(k))\, d\pi(h,k) \leq C_X \int_{(E^\alpha)^2} (\|h\|_{C^\beta}^q + \|k\|_{C^\beta}^q)\, \|h-k\|_\infty^{p-q}\, d\pi(h,k)\,. \quad (5.188)$$

By virtue of Hölder's inequality, for any $r, s \in]1, \infty[$ satisfying $\frac{1}{r} + \frac{1}{s} = 1$ and $(p-q)s \geq 1$,

$$\begin{aligned}
\int_{(E^\alpha)^2} \|h\|_{C^\beta}^q\, \|h-k\|_\infty^{p-q}\, d\pi(h,k) &\leq \left[\int_{(E^\alpha)^2} \|h\|_{C^\beta}^{qr}\, d\pi(h,k)\right]^{\frac{1}{r}} \left[\int_{(E^\alpha)^2} \|h-k\|_\infty^{(p-q)s}\, d\pi(h,k)\right]^{\frac{1}{s}} \\
&= \left[\int_{E^\alpha} \|f\|_{C^\beta}^{qr}\, d\mathbb{P}(\omega)\right]^{\frac{1}{r}} \left[\int_{(E^\alpha)^2} \|h-k\|_\infty^{(p-q)s}\, d\pi(h,k)\right]^{\frac{1}{s}} \\
&= \mathbb{E}\big[\|f\|_{C^\beta}^{qr}\big]^{\frac{1}{r}}\, W_{(p-q)s,\infty}^{p-q}(f_\sharp\mathbb{P}, g_\sharp\mathbb{P})\,,
\end{aligned}$$

where the equality on the second line is valid since we know the marginals of π. Putting everything together we retrieve the statement of the theorem, namely that for a universal constant C_X depending only on X,

$$W_{p,d_p}((\mathrm{Dgm} \circ f)_\sharp \mathbb{P}, (\mathrm{Dgm} \circ g)_\sharp \mathbb{P}) \leq C_X \left[\mathbb{E}\big[\|f\|_{C^\beta}^{qr}\big]^{\frac{1}{r}} + \mathbb{E}\big[\|g\|_{C^\beta}^{qr}\big]^{\frac{1}{r}}\right]^{\frac{1}{p}} W_{(p-q)s,\infty}^{1-\frac{q}{p}}(f_\sharp\mathbb{P}, g_\sharp\mathbb{P})\,.$$

The last inequality in the theorem is obtained by virtue of proposition 5.12. ■

This shows the following proposition.

Proposition 5.11. Let $\mathcal{B}$ be a Banach space and $\Psi : \mathcal{D}_p \to \mathcal{B}$ be an α-Hölder continuous functional. Let $\mathbb{P}, \mathbb{Q} \in \mathcal{P}_{\alpha q}(\mathcal{D}_p)$, then

$$\|\mathbb{E}_{\mathbb{P}}[\Psi] - \mathbb{E}_{\mathbb{Q}}[\Psi]\|_{\mathcal{B}} \leq W_{q,\|\cdot\|_{\mathcal{B}}}(\Psi_\sharp\mathbb{P}, \Psi_\sharp\mathbb{Q}) \leq \|\Psi\|_{C^\alpha(\mathcal{D}_p,\mathcal{B})}\, W_{q\alpha,d_p}^\alpha(\mathbb{P}, \mathbb{Q})\,. \quad (5.189)$$

Proposition 5.12 (Control of W_{p,L^∞})**.** Let f and g be two $\mathbb{R}$-valued a.s. E^α stochastic processes on a d dimensional compact Riemannian manifold X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, the following inequality holds

$$W_{p,L^\infty}(f_\sharp\mathbb{P}, g_\sharp\mathbb{P}) \leq \|f-g\|_{L^p(\Omega, L^\infty(X,\mathbb{R}))} \quad (5.190)$$

Proof. The map $F : \Omega \to E^\alpha(X,\mathbb{R})^2$ which sends $\omega \mapsto (f(\omega), g(\omega))$ induces a transport map $F_\sharp\mathbb{P} \in \Gamma(f_\sharp\mathbb{P}, g_\sharp\mathbb{P})$ and

$$\begin{aligned}
W_{p,L^\infty}^p(f_\sharp\mathbb{P}, g_\sharp\mathbb{P}) &\leq \int_{E^\alpha(X,\mathbb{R})^2} \|h-k\|^p\, dF_\sharp\mathbb{P}(h,k) = \int_\Omega \|f(\omega) - g(\omega)\|_\infty^p\, d\mathbb{P}(\omega) \\
&= \|f-g\|_{L^p(\Omega, L^\infty(X,\mathbb{R}))}^p\,,
\end{aligned}$$

which finishes the proof. ■

Remark 5.13. Proposition 5.12 yields an easy way to estimate the value of Wasserstein distances between stochastic processes. Using the results of [49] and other results on rates of convergence of random processes (which could be obtained by using results such as those of Kahane [38]), this instantly gives estimates for Wasserstein distances between distributions for a panoply of processes.

Corollary 5.14 (A remark on discretization)**.** *Keeping the same notation, fix a triangulation P of X whose 0-skeleton has n points and such that the 0-skeleton of P is an ε-net of X (this constrains $n \geq \mathcal{N}_X(\varepsilon)$) and define a new process $\hat{f}$ which is equal to f on the 0-skeleton of P and linearly interpolate in between. Then,*

$$W_{p,L^\infty}(f_\sharp\mathbb{P}, \hat{f}_\sharp\mathbb{P}) \leq \mathbb{E}\big[\|f\|_{C^\beta}^p\big]\,\varepsilon^{\beta p}\,. \tag{5.191}$$

If $p = \infty$ and that $\|f\|_{C^\beta}$ is uniformly bounded by L, then

$$W_{\infty,L^\infty}(f_\sharp\mathbb{P}, \hat{f}_\sharp\mathbb{P}) \leq L\varepsilon^\alpha \tag{5.192}$$

and theorem 5.9 applies.

Proof. Clearly, $\hat{f} : \Omega \to \mathrm{Lip}_{\Lambda_\varepsilon}(X, \mathbb{R})$ of law $\hat{f}_\sharp\mathbb{P}$. By proposition 5.12, for any $\beta < \alpha$,

$$W_{p,L^\infty}^p(f_\sharp\mathbb{P}, \hat{f}_\sharp\mathbb{P}) \leq \mathbb{E}\left[\left\|f - \hat{f}\right\|_\infty^p\right] \leq \mathbb{E}\big[\|f\|_{C^\beta}^p\big]\,\varepsilon^{\beta p}\,.$$

Taking $p \to \infty$, provided that the distribution of $\|f\|_{C^\beta}$ has bounded support, we can bound the support of this distribution by L, we get $W_{\infty,L^\infty}(f_\sharp\mathbb{P}, \hat{f}_\sharp\mathbb{P}) \leq L\varepsilon^\alpha$. In particular, the expected diagrams differ from less than $L\varepsilon^\alpha$ in d_∞. ■

Remark 5.15. The topology on the measures on $C^0(X, \mathbb{R})$ defined by Wasserstein distances may be too weak. Indeed, note that W_{p,L^∞}-balls around any measure μ supported on some $E^\alpha(X, \mathbb{R})$ include probability measures whose support intersects sets of $C^0(X, \mathbb{R})$ whose number of small bars grows faster than any polynomial (or indeed any computable function!). To see why, it suffices to exhibit an example of such a function (let us denote it h), and notice that if a stochastic process f has law μ, if ξ denotes a standard gaussian random variable, then $f + \varepsilon\xi h$ is (up to rendering f locally constant on some small ball) an arbitrarily small L^∞-perturbation of f whose number of small bars grows arbitrarily fast. In particular, this perturbation is not in any $\mathcal{D}_p$ for any p, but the law of this perturbed process is included within a W_{p,L^∞}-ball of arbitrarily small radius.

However, by changing topology to that of a Sobolev space which injects itself onto some $C^\alpha(X, \mathbb{R})$, we can avoid this problem. With this change in topology, it might be superfluous to require that the processes lie in $E^\alpha(X, \mathbb{R})$, as it might follow from an argument ressembling that of the proof of the Kolmogorov-Chentsov theorem (theorem 5.16).

5.3 Establishing classes of regularity

A sufficient and easily verifiable condition for a stochastic process to be almost surely E^α is given by the Kolmogorov-Chentsov theorem.

Theorem 5.16 (Kolmogorov-Chentsov Theorem for compact manifolds, [5,6])**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{B}$ be a Banach space, X be a d-dimensional compact Riemannian manifold (without boundary) with distance d_X and $f : \Omega \times X \to \mathcal{B}$ be a $\mathcal{B}$-valued separable stochastic process. Suppose there exists constants $C > 0$, $\varepsilon > 0$ and $\delta > 1$ such that for all $x, y \in X$,*

$$\mathbb{E}\left[\|f(x) - f(y)\|_{\mathcal{B}}^\delta\right] \leq C d_X(x, y)^{d+\varepsilon}\,, \tag{5.193}$$

then there exists a modification of f such that for all $\alpha \in [0, \frac{\varepsilon}{\delta}[$, f is almost surely α-Hölder continuous.

The proof uses the same idea of [5] to use the Sobolev embedding theorem. For compact Riemannian manifolds, the required Sobolev embedding theorem is given by [6, Theorem 2.20] (in fact, within [6], one can actually find Sobolev embedding theorems valid for wider classes of manifolds). Let us give a sketch of the proof.

Sketch of proof of theorem 5.16. First, by virtue of Markov's inequality, the estimation on the moments above entails that the process is continuous in probability. We may therefore assume that, up to taking a modification of f, the process f is measurable on $\Omega \times X$. Fix γ a real number, then Tonelli's theorem and the estimation of the moments above implies that

$$\mathbb{E}\left[\int_X \int_X \frac{\|f(x) - f(y)\|_{\mathcal{B}}^\delta}{d_X(x,y)^{d+\gamma\delta}}\, dx\, dy\right] = \int_X \int_X \frac{\mathbb{E}\left[\|f(x) - f(y)\|_{\mathcal{B}}^\delta\right]}{d_X(x,y)^{d+\gamma\delta}}\, dx\, dy$$
$$\leq C \int_X \int_X d_X(x,y)^{\varepsilon - \gamma\delta}\, dx\, dy$$

which is finite as soon as $\gamma < \frac{d+\varepsilon}{\delta}$. Notice that the bounded quantity is nothing other than the norm of f in $L^\delta(\Omega, W^{\gamma,\delta}(X, \mathcal{B}))$, so that almost surely, $f_\omega \in W^{\gamma,\delta}(X, \mathcal{B})$. There is a Sobolev injection of $W^{\gamma,\delta}(X, \mathcal{B}) \hookrightarrow C^\alpha(X, \mathcal{B})$ for all $\alpha < \gamma - \frac{d}{\delta}$, so for every $\alpha < \frac{\varepsilon}{\delta}$, there is a measurable set $\Omega_0 \subset \Omega$ of probability measure 1 on which for every $\omega \in \Omega_0$, f_ω is α-Hölder almost everywhere on X. The corresponding modification can be obtained by making the trajectories continuous everywhere. Since the process f is measurable on $\Omega \times X$, we can set

$$g_\omega(h, x) := \frac{1}{\mathrm{Vol}(B(x,h))} \int_{B(x,h)} f_\omega(y)\, dy\,, \tag{5.194}$$

and consider the set

$$B = \{(\omega, x) \in \Omega \times X \mid (g_\omega(h,x))_h \text{ converges as } h \to 0\} \tag{5.195}$$

and set the continuous modification of f to be

$$g_\omega(x) := \begin{cases} \lim_{h\to 0} g_\omega(h,x) & (\omega, x) \in B \\ 0 & \text{else} \end{cases}. \tag{5.196}$$

Finally, it is easy to check this function is indeed α-Hölder everywhere on Ω_0 and to check that $\mathbb{P}(g(x) = f(x)) = 1$ almost everywhere on X. ■

Remark 5.17. If $\mathcal{B} = \mathbb{R}$, the same idea works (as shown in [5]) to prove results on the existence of modifications of processes such that the modification is almost surely of class C^k.

Provided that we have control over all moments of $\|f(x) - f(y)\|$, the Kolmogorov-Chentsov theorem constrains the regularity of the process to live within some family

$$\bigcap_{0 \leq \alpha < \alpha^*} C^\alpha(X, \mathbb{R}) \tag{5.197}$$

for some α^*. As an immediate corollary,

Corollary 5.18. *With the same hypotheses and notation of theorem 5.16 where now $\mathcal{B} = \mathbb{R}$, denoting $\alpha^* := \sup_{\varepsilon,\delta} \frac{\varepsilon}{\delta}$, almost surely,*

$$\mathcal{L}_{Tot}(f) \leq \frac{d}{\alpha^*}\,. \tag{5.198}$$

6 Acknowledgements

The author would like to thank Pierre Pansu and Claude Viterbo for helping with the redaction of the manuscript as well as their guidance. Many thanks are also owed to Shmuel Weinberger, Yuliy Baryshnikov, David Cohen-Steiner, Jean-François Le Gall and Nicolas Curien for the fruitful discussions without which some of this work would not have been possible.

References

[1] H. Adams, M. Aminian, E. Farnell, M. Kirby, J. Mirth, R. Neville, C. Peterson, and C. Shonkwiler. A fractal dimension for measures via persistent homology. *Abel Symposia*, pages 1–31, 2020.

[2] H. Adams, T. Emerson, M. Kirby, R. Neville, C. Peterson, P. Shipman, S. Chepushtanova, E. Hanson, F. Motta, and L. Ziegelmeier. Persistence images: A stable vector representation of persistent homology. *Journal of Machine Learning Research*, 18(8):1–35, 2017.

[3] R. J. Adler, O. Bobrowski, M. S. Borman, E. Subag, and S. Weinberger. Persistent homology for random fields and complexes. In *Institute of Mathematical Statistics Collections*, pages 124–143. Institute of Mathematical Statistics, 2010.

[4] R. J. Adler and J. E. Taylor. *Random Fields and Geometry.* Springer New York, 2007.

[5] R. Andreev and A. Lang. Kolmogorov-Chentsov theorem and differentiability of random fields on manifolds. *Potential Analysis*, 41(3):761–769, feb 2014.

[6] T. Aubin. *Some Nonlinear Problems in Riemannian Geometry.* Springer Berlin Heidelberg, 1998.

[7] Y. Baryshnikov. Time series, persistent homology and chirality. *arXiv:1909.09846*, 2019.

[8] J. M. Boardman and R. M. Vogt. Homotopy-everything h-spaces. *Bulletin of the American Mathematical Society*, 74(6):1117–1123, nov 1968.

[9] J. M. Boardman and R. M. Vogt. *Homotopy Invariant Algebraic Structures on Topological Spaces.* Springer Berlin Heidelberg, 1973.

[10] P. Bubenik and T. Vergili. Topological spaces of persistence modules and their properties. *Journal of Applied and Computational Topology*, 2(3-4):233–269, dec 2018.

[11] D. Burago, Y. Burago, and S. Ivanov. *A course in metric geometry*, volume 33 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, 2001.

[12] M. Carriere, S. Oudot, and M. Ovsjanikov. Sliced Wasserstein Kernel for Persistence Diagrams. In *ICML 2017 - Thirty-fourth International Conference on Machine Learning*, pages 1–10, Sydney, Australia, Aug. 2017.

[13] F. Chazal, W. Crawley-Boevey, and V. de Silva. The observable structure of persistence modules. *Homology, Homotopy and Applications*, 18(2):247–265, 2016.

[14] F. Chazal, V. de Silva, M. Glisse, and S. Oudot. *The Structure and Stability of Persistence Modules.* Springer International Publishing, 2016.

[15] F. Chazal and V. Divol. The density of expected persistence diagrams and its kernel based estimation. In B. Speckmann and C. D. Tóth, editors, *34th International Symposium on Computational Geometry (SoCG 2018)*, volume 99 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 26:1–26:15, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[16] F. Chazal, B. Fasy, F. Lecci, B. Michel, A. Rinaldo, and L. Wasserman. Subsampling methods for persistent homology. In F. Bach and D. Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 2143–2151, Lille, France, 07–09 Jul 2015. PMLR.

[17] F. Chazal, M. Glisse, C. Labruère, and B. Michel. Convergence rates for persistence diagram estimation in topological data analysis. In E. P. Xing and T. Jebara, editors, *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 163–171, Bejing, China, 22–24 Jun 2014. PMLR.

[18] C. Chen and H. Edelsbrunner. Diffusion runs low on persistence fast. In *2011 International Conference on Computer Vision*, pages 423–430, 2011.

[19] I. Chiswell. *Introduction to Λ-Trees.* World Scientific, feb 2001.

[20] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard. Unbalanced optimal transport: Dynamic and kantorovich formulation, 2015.

[21] D. Cohen-Steiner, H. Edelsbrunner, J. Harer, and Y. Mileyko. Lipschitz functions have L^p-stable persistence. *Foundations of Computational Mathematics*, 10(2):127–139, Jan 2010.

[22] N. Curien, J.-F. Le Gall, and G. Miermont. The Brownian cactus I. Scaling limits of discrete cactuses. *Ann. Inst. H. Poincaré Probab. Statist.*, 49(2):340–373, 05 2013.

[23] J. Curry. The fiber of the persistence map for functions on the interval. *Journal of Applied and Computational Topology*, 2(3-4):301–321, dec 2018.

[24] J. Curry, H. Hang, W. Mio, T. Needham, and O. B. Okutan. Decorated merge trees for persistent topology, 2021.

[25] V. Divol and T. Lacombe. Understanding the topology and the geometry of the persistence diagram space via optimal partial transport. *CoRR*, abs/1901.03048, 2019.

[26] V. Divol and T. Lacombe. Estimation and quantization of expected persistence diagrams. 2021.

[27] V. Divol and W. Polonik. On the choice of weight functions for linear representations of persistence diagrams. *Journal of Applied and Computational Topology*, 3(3):249–283, aug 2019.

[28] T. Duquesne and J.-F. Le Gall. *Random trees, Lévy processes and spatial branching processes.* Number 281 in Astérisque. Société mathématique de France, 2002.

[29] T. Duquesne and J.-F. Le Gall. Probabilistic and fractal aspects of Lévy trees. *Probability Theory and Related Fields*, 131(4):553–603, Nov 2004.

[30] H. Edelsbrunner and J. Harer. *Computational Topology: An Introduction.* 01 2010.

[31] S. Eilenberg and N. Steenrod. *Foundations of Algebraic Topology.* Princeton University Press, dec 1952.

[32] B. T. Fasy, F. Lecci, A. Rinaldo, L. Wasserman, S. Balakrishnan, and A. Singh. Confidence sets for persistence diagrams. *The Annals of Statistics*, 42(6), dec 2014.

[33] A. Figalli. The optimal partial transport problem. *Archive for Rational Mechanics and Analysis*, 195(2):533–560, jan 2009.

[34] A. Figalli and N. Gigli. A new transportation distance between non-negative measures, with applications to gradients flows with dirichlet boundary conditions. *Journal de Mathématiques Pures et Appliquées*, 94(2):107–130, 2010.

[35] V. Ginzburg and M. Kapranov. Koszul duality for operads. *Duke Mathematical Journal*, 76(1):203–272, oct 1994.

[36] C. R. Givens and R. M. Shortt. A class of Wasserstein metrics for probability distributions. *Michigan Mathematical Journal*, 31(2), jan 1984.

[37] H. Jiang and Y.-H. Yang. Manifolds of positive ricci curvature with quadratically asymptotically nonnegative curvature and infinite topological type. *Communications in Analysis and Geometry*, 29(5):1233–1253, 2021.

[38] J.-P. Kahane. *Some random series of functions.* Cambridge University Press, Feb 1986.

[39] S. Kondratyev, L. Monsaingeon, and D. Vorotnikov. A new optimal transport distance on the space of finite Radon measures. *Adv. Differ. Equ.*, 21(11-12):1117–1164, 2016.

[40] J.-L. Loday. La renaissance des opérades. In *Séminaire Bourbaki. Volume 1994/95. Exposés 790-804*, pages 47–74, ex. Paris: Société Mathématique de France, 1996.

[41] R. MacPherson and B. Schweinhart. Measuring shape with topology. *Journal of Mathematical Physics*, 53(7):073516, Jul 2012.

[42] J. P. May. *The Geometry of Iterated Loop Spaces.* Springer Berlin Heidelberg, 1972.

[43] F. Mémoli and O. B. Okutan. Reeb posets and tree approximations. *Discrete Mathematics*, 343(2):111658, 2020.

[44] X. Menguy. Examples of strictly weakly regular points. *Geometric and Functional Analysis*, 11(1):124–131, apr 2001.

[45] X. C. Menguy. *Examples of manifolds and spaces with positive Ricci curvature.* PhD thesis, New York University, 2000.

[46] Y. Mileyko, S. Mukherjee, and J. Harer. Probability measures on the space of persistence diagrams. *Inverse Problems*, 27(12):124007, Nov 2011.

[47] E. Munch and A. Stefanou. The ℓ^∞-cophenetic metric for phylogenetic trees as an interleaving distance. In *Association for Women in Mathematics Series*, pages 109–127. Springer International Publishing, 2019.

[48] S. Y. Oudot. *Persistence Theory - From Quiver Representations to Data Analysis*, volume 209 of *Mathematical surveys and monographs.* American Mathematical Society, 2015.

[49] D. Perez. On the persistent homology of almost surely C^0 stochastic processes. `https://arxiv.org/abs/2012.09459`, Dec. 2020.

[50] D. Perez. ζ-functions and the topology of superlevel sets of stochastic processes. *arXiv e-prints*, page arXiv:2110.10982, Oct. 2021.

[51] J. Picard. A tree approach to p-variation and to integration. *The Annals of Probability*, 36(6):2235–2279, Nov 2008.

[52] I. Polterovich, L. Polterovich, and V. Stojisavljević. Persistence barcodes and Laplace eigenfunctions on surfaces. *Geometriae Dedicata*, 201(1):111–138, Aug 2018.

[53] L. Polterovich, D. Rosen, K. Samvelyan, and J. Zhang. Topological Persistence in Geometry and Analysis. *arXiv e-prints*, page arXiv:1904.04044, Apr 2019.

[54] B. Schweinhart. Persistent homology and the upper box dimension. *Discrete & Computational Geometry*, Nov 2019.

[55] P. Skraba and K. Turner. Wasserstein stability for persistence diagrams, 2020.

[56] R. P. Stanley. *Enumerative Combinatorics.* Cambridge University Press, 2009.

[57] K. Turner, Y. Mileyko, S. Mukherjee, and J. Harer. Fréchet means for distributions of persistence diagrams. *Discrete & Computational Geometry*, 52(1):44–70, Jul 2014.

[58] S. Wang, Y. Wang, and R. Wenger. The JS-graphs of join and split trees. In *Proceedings of the thirtieth annual symposium on Computational geometry.* ACM, jun 2014.

[59] S. Willard. *General topology.* Addison-Wesley Publishing Co., 1970.