



HAL
open science

The Stability of Persistence Diagrams Revisited

Frédéric Chazal, David Cohen-Steiner, Leonidas J. Guibas, Steve Oudot

► **To cite this version:**

Frédéric Chazal, David Cohen-Steiner, Leonidas J. Guibas, Steve Oudot. The Stability of Persistence Diagrams Revisited. [Research Report] RR-6568, 2008, pp.24. inria-00292566v1

HAL Id: inria-00292566

<https://inria.hal.science/inria-00292566v1>

Submitted on 5 Jul 2008 (v1), last revised 28 Nov 2008 (v4)

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

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



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

The Stability of Persistence Diagrams Revisited

Frédéric Chazal — David Cohen-Steiner — Leonidas J. Guibas — Steve Y. Oudot

N° 6568

June 2008

Thème COM



*R*apport
de recherche

The Stability of Persistence Diagrams Revisited

Frédéric Chazal ^{*}, David Cohen-Steiner [†], Leonidas J. Guibas [‡],
Steve Y. Oudot [§]

Thème COM — Systèmes communicants
Équipe-Projet Geometrica

Rapport de recherche n° 6568 — June 2008 — 24 pages

Abstract: The concept of topological persistence introduced independently by several groups [9, 12, 10] is a rather general tool providing an efficient way to encode the qualitative and quantitative behavior of real-valued functions defined over topological spaces. Since its introduction this encoding, known as the persistence diagram or barcode, has been extensively studied, specifically in topological data analysis where its stability properties allow to infer robust topological information on the studied data sets. Motivated by problems coming from topological data analysis, we extend the notions of persistence and persistence diagrams to a larger setting than the one classically considered. In this paper we prove new stability results for the persistence diagrams that lead to new applications in topological and geometric data analysis.

Key-words: Topological persistence, Stability, Persistence diagram, Topological and geometric data analysis

* frederic.chazal@inria.fr

† David.Cohen-Steiner@sophia.inria.fr

‡ guibas@cs.stanford.edu

§ steve.oudot@inria.fr

La stabilité des diagrammes de persistences revisitée

Résumé : Le concept de persistance topologique, introduit indépendamment par différents groupes [9, 12, 10], est un outil général permettant de coder de façon qualitative et quantitative le comportement de fonctions à valeurs réelles définies sur des espaces topologiques. Depuis son introduction ce codage, connu sous le nom de diagramme de persistance ou de code-barre, a été largement étudié, en particulier pour l'analyse topologique de données où ses propriétés de stabilité permettent d'inférer de manière fiable les propriétés topologiques des données étudiées. Dans cet article, motivés par des problèmes issus de l'analyse topologique des données, nous étendons les notions de persistance et de diagrammes de persistance à un cadre plus large que celui classiquement considéré. On prouve de nouveaux résultats de stabilité pour les diagrammes de persistance qui conduisent à de nouvelles applications pour l'étude des propriétés topologiques et géométriques des données.

Mots-clés : Persistance topologique, Stabilité, Diagramme de persistance, Analyse topologique et géométrique de données

1 Introduction

The concept of topological persistence introduced independently by several groups [9, 12, 10] is a rather general tool to study the qualitative and quantitative behavior of real-valued functions defined over topological spaces. More specifically, it provides a way to encode the topological changes in the sequence of the sublevel sets of the considered functions. This encoding is classically represented by a diagram in the plane \mathbb{R}^2 (the so-called persistence diagram) or a set of intervals (the so-called barcode). Since the introduction of the persistence algorithm to compute such diagrams by Edelsbrunner *et al.* [9], topological persistence and its applications have been an extensively studied topic. First designed for simplicial complexes in \mathbb{R}^3 , the persistence algorithm was later extended to the persistent homology of discrete functions over arbitrary finite simplicial complexes [13]. A number of variants were also proposed, for instance to cope with changes in the function over time [6] or to handle pairs of functions defined over nested pairs of spaces [5]. All these methods deal with continuous functions or functions defined over simplicial complexes for which it has been proven that the persistence diagrams are stable under perturbations of the functions [4]. This stability property appears as a fundamental result to ensure the relevance of the use of topological persistence in topological and geometric data analysis (see [1] for example). Nevertheless some recent works in topological data analysis [5, 2] have shown that the functional setting is not large enough to address the problems encountered in practical applications. In this paper we introduce a new larger algebraic setting, encompassing the classical functional one, in which the persistence can still be defined.

In the classical setting the persistence diagram of a real-valued function $f : \mathbb{X} \rightarrow \mathbb{R}$ is built from the family of homology groups of its sublevel sets $\{H_k(f^{-1}((-\infty, \alpha]))\}_{\alpha \in \mathbb{R}}$ related by the homomorphisms induced by the canonical inclusion maps $f^{-1}((-\infty, \alpha]) \hookrightarrow f^{-1}((-\infty, \beta])$, for $\alpha \leq \beta$. To be well-defined and stable with respect to perturbations of f the notion of persistence diagram requires some topological assumptions on f and X : all the considered homology groups have to be finite dimensional and to satisfy some additional rather restrictive hypothesis called tameness in [4] and X needs to be triangulable. Here we introduce a purely algebraic notion of *persistence module* \mathcal{F} which is a family of vector spaces (or R -modules for some commutative ring R) $\mathcal{F} = \{F^\alpha\}_{\alpha \in \mathbb{R}}$ related by a family of homomorphisms $\{\phi_{\alpha, \beta} : F^\alpha \rightarrow F^\beta\}_{\alpha \leq \beta}$ such that $\forall \alpha \leq \beta \leq \gamma$, $\phi_{\alpha, \gamma} = \phi_{\beta, \gamma} \circ \phi_{\alpha, \beta}$ and $\phi_{\alpha, \alpha} = id_{F^\alpha}$. Under a rather weak hypothesis, called δ -tameness¹, we provide a notion of persistence diagrams for persistence modules. This notion appears to be equal to the one introduced in [4] in the classical setting. Our approach is based on an intuitive approximation strategy: we obtain the persistence diagram of \mathcal{F} as a limit of easy-to-define persistence diagrams of its discretizations. Here, by a discretization of \mathcal{F} we mean a subsequence of the family $\{F^\alpha\}_{\alpha \in \mathbb{R}}$ of the form $\{F^{x+n\varepsilon}\}_{n \in \mathbb{Z}}$. Our main result shows that these newly introduced persistence diagrams are stable under perturbations of the persistence modules (in a sense defined in section 4).

The paper is organized as follows. In section 2 we briefly recall classical results of topological persistence and we introduce the notion of 0-tame persistence module which extend the notion of persistence module of the sublevel sets

¹ \mathcal{F} is said to be δ -tame for some $\delta \geq 0$ if for any $\alpha < \alpha + \delta < \beta$, the rank of the homomorphism $\phi_{\alpha, \beta}$ is finite.

filtration of a continuous function. Section 3 is dedicated to the definition of the persistence diagrams using a discretization strategy. The stability of persistence diagrams is proven in section 4. Some applications of the stability are briefly mentioned in section 5. More detailed applications are presented in another paper [2].

2 Background and definitions

2.1 Extended plane, multisets and Bottleneck distance

Throughout the paper we denote $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ the extended real line and we use the following rules: $\forall x \in \mathbb{R}$, $x + \infty = +\infty$, $x - \infty = -\infty$, and $\forall x \in \mathbb{R} \setminus \{0\}$, $x \times (+\infty) = +\infty$ if $x > 0$ and $-\infty$ otherwise, $x \times (-\infty) = -\infty$ if $x > 0$ and $+\infty$ otherwise. The extended plane $\bar{\mathbb{R}}^2$ is endowed with the d^∞ distance defined by

$$d^\infty((x, y), (x', y')) = \max(|x - x'|, |y - y'|)$$

We call Δ the diagonal $\delta = \{(x, x), x \in \bar{\mathbb{R}}\}$ and $\Delta_+ = \{(x, y) \in \bar{\mathbb{R}}^2 : y > x\}$ the open half-plane above the diagonal Δ . For any $\varepsilon > 0$, we denote $\Delta_+^\varepsilon = \{(x, y) \in \bar{\mathbb{R}}^2 : y > x + \varepsilon\}$ the open half-plane at distance $\frac{\varepsilon}{2}$ above Δ (see figure 1).

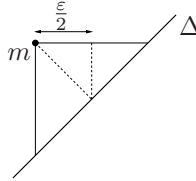


Figure 1: d_∞ distance to the diagonal.

The main mathematical objects considered in topological persistence are the so called persistence diagrams (or barcodes) that are multisets in $\bar{\mathbb{R}}^2$. A *multiset* D in $\bar{\mathbb{R}}^2 = \bar{\mathbb{R}} \times \bar{\mathbb{R}}$ is a subset of $\bar{\mathbb{R}}^2$ where to each point p of D is associated a multiplicity $\text{mult}(p) \in \mathbb{N} \cup \{+\infty\}$. The *support* of D , i.e. the subset considered without the multiplicities is denoted by $|D|$. Equivalently, D can be represented as a disjoint union

$$D = \coprod_{p \in |D|} \coprod_{i=1}^{\text{mult}(p)} p$$

A *multibijection* Φ between two multisets D, D' is a bijection

$$\Phi : \coprod_{p \in |D|} \coprod_{i=1}^{\text{mult}(p)} p \rightarrow \coprod_{p' \in |D'|} \coprod_{i=1}^{\text{mult}(p')} p'$$

Given two multisets D and D' the Hausdorff distance between their support is defined by

$$d_H(|D|, |D'|) = \max\left(\sup_{p \in |D|} \inf_{p' \in |D'|} d^\infty(p, p'), \sup_{p' \in |D'|} \inf_{p \in |D|} d^\infty(p', p)\right)$$

We also denote by $d_H(D, D')$ the Hausdorff distance $d_H(|D|, |D'|)$ (but take care that this is no longer a distance on the set of multisets). A more interesting distance between multisets is the so-called *Bottleneck distance* d_B^∞ introduced in [4]:

$$d_B^\infty(D, D') = \inf_{\Phi: D \rightarrow D'} \inf_{\text{multibijection}} \left(\sup_{p \in D} d^\infty(p, \Phi(p)) \right)$$

In the following of the paper, to avoid heavy notations, we use the same notation for a multiset and its support.

2.2 Filtrations and persistence modules

The homology theory used throughout the paper is the singular homology with coefficients in a commutative ring R with unity (see [11] for an introduction to the subject) which is assumed to be a field in the following.

Topological persistence has been introduced as a useful and robust tool to encode the topological behavior of real valued functions defined on topological spaces. Its construction strongly relies on the filtrations defined by the sublevel sets of the considered functions.

Definition 2.1 *Given a topological space \mathbb{X} and a function $f : \mathbb{X} \rightarrow \mathbb{R}$, the sublevel-sets filtration of f is defined by:*

- i) the family of subsets $\mathcal{F} = \{F^\alpha\}_{\alpha \in \mathbb{R}}$ where $\forall \alpha \in \mathbb{R}, F^\alpha = f^{-1}((-\infty, \alpha])$,*
- ii) the family of canonical inclusion maps $i_{\alpha, \beta} : F^\alpha \hookrightarrow F^\beta, \forall \alpha \leq \beta$.*

The previous family of inclusion maps induces a family of homomorphisms between the homology groups $H_k(F^\alpha)$ of the sublevel sets of f , known as the *persistence module* of the filtration \mathcal{F} [13], which plays an essential role in the definition of topological persistence. Persistence modules can be defined independently of the sublevel sets filtration of a function in a formal way that allows to generalize the notion of persistence to a larger setting than the one considered in [8, 13].

Definition 2.2 *Let R be a commutative ring with unity. A persistence module $\mathcal{F} = \{F^\alpha\}_{\alpha \in \mathbb{R}}$ is a family of R -modules F^α together with a family $\{\phi_{\alpha, \beta} : F^\alpha \rightarrow F^\beta\}_{\alpha \leq \beta}$ of homomorphisms such that*

$$\forall \alpha \leq \beta \leq \gamma, \quad \phi_{\alpha, \gamma} = \phi_{\beta, \gamma} \circ \phi_{\alpha, \beta} \quad \text{and} \quad \phi_{\alpha, \alpha} = id_{F^\alpha}$$

If $\mathcal{F} = \{F^\alpha\}_{\alpha \in \mathbb{Z}}$ is indexed by the set \mathbb{Z} of the integers, it is called a discrete persistence module.

The persistence module associated to the sublevel sets filtration of a function f obviously satisfies the above conditions. Since the ring R is assumed to be a fixed field, the modules F^α are vector spaces and the homomorphisms $\phi_{\alpha, \beta}$ are linear maps between vector spaces. In particular, the rank of $\phi_{\alpha, \beta}$ is a well-defined integer or $+\infty$. To avoid heavy notations we use the same notation F^α to refer to a R -module or the sublevel sets of a given function f . This abuse of notation will not be confusing in the following.

2.3 Tameness and topological persistence

The original goal of topological persistence was to study and quantify the topological changes in the sublevel sets filtrations of functions. These changes are encoded in the so-called persistence diagram (or barcode). One important property of the persistence diagram is that it is stable under some perturbations of the considered function making it a robust tool for topological analysis (see [4]). Nevertheless, this stability property has been proven for a rather restricting class of function (the so-called tame functions) that appears to be not sufficient in some applications in topological data analysis (see for example [2, 5]). In this section we briefly recall the classical definitions and properties of topological persistence and we introduce the larger setting in which we define the persistence in the next sections.

Definition 2.3 ([4]) *Let $f : \mathbb{X} \rightarrow \mathbb{R}$ be a topological space and let $f : \mathbb{X} \rightarrow \mathbb{R}$ be a function. A homological critical value of f is a real number a for which there exists an integer k such that for all sufficiently small $\varepsilon > 0$ the maps $H_k(F^{a-\varepsilon}) \rightarrow H_k(F^{a+\varepsilon})$ induced by the inclusion is not an isomorphism. The function f is said to be tame if it has only a finite number of homological critical values and if all the homology groups $H_k(F^\alpha)$ are finite dimensional for all $k \in \mathbb{N}$ and all $\alpha \in \mathbb{R}$.*

The persistence diagram of a tame function f is defined as a multiset in the following way. Let $a_1 < a_2 < \dots < a_n$ be the homological critical values of f and let b_0, \dots, b_n be a sequence of real numbers such that $b_{i-1} < a_i < b_i$ for all $i = 1, \dots, n$. We set $b_{-1} = a_0 = -\infty$ and $b_{n+1} = a_{n+1} = +\infty$. For any $0 \leq i < j \leq n+1$, the multiplicity of the point (a_i, a_j) is defined by

$$\begin{aligned} \text{mult}(a_i, a_j) &= \text{rk}(H_*(F^{b_{i-1}}) \rightarrow H_*(F^{b_j})) - \text{rk}(H_*(F^{b_i}) \rightarrow H_*(F^{b_j})) \quad (1) \\ &+ \text{rk}(H_*(F^{b_i}) \rightarrow H_*(F^{b_{j-1}})) - \text{rk}(H_*(F^{b_{i-1}}) \rightarrow H_*(F^{b_{j-1}})) \quad (2) \end{aligned}$$

where $H_*(\cdot) = \bigoplus_{k \in \mathbb{Z}} H_k(\cdot)$ is the direct sum of all the k -dimensional homology groups and the arrows are the homomorphisms induced by the inclusion between the sublevel sets of f ². The *persistence diagram* $Df \subset \overline{\mathbb{R}^2}$ of f is the multiset defined as the union of the pairs (a_i, a_j) , $0 \leq i < j \leq n+1$ with multiplicity $\text{mult}(a_i, a_j)$ and the diagonal Δ where all its points have infinite multiplicity.

The main result on persistence diagrams is the following stability result.

Theorem 2.4 ([4]) *Let \mathbb{X} be a triangulable space and let $f, g : \mathbb{X} \rightarrow \mathbb{R}$ be two continuous tame functions. Then*

$$d_B^\infty(Df, Dg) \leq \|f - g\|_\infty = \sup_{x \in \mathbb{X}} |f(x) - g(x)|$$

The main assumptions needed to apply the previous theorem are that \mathbb{X} has to be sufficiently nice (triangulable) and f, g have to be continuous, with a finite number of homological critical values and with sublevel sets having finite dimensional homology groups. These hypotheses appear to be too restrictive in some applications [2, 5]. It is thus appealing to generalize such a stability

²By convention, if $i = -\infty$ then $\text{rk}(H_*(F^{b_i}) \rightarrow H_*(F^{b_j})) = \lim_{\alpha \rightarrow -\infty} \text{rk}(H_*(F^\alpha) \rightarrow H_*(F^{b_j}))$ and if $j = +\infty$ then $\text{rk}(H_*(F^{b_i}) \rightarrow H_*(F^{b_j})) = \lim_{\alpha \rightarrow +\infty} \text{rk}(H_*(F^{b_i}) \rightarrow H_*(F^\alpha))$.

result to a larger setting. In the following we show that the notion of persistence diagrams can be extended to a larger class of persistence modules than the ones induced by sublevel sets filtrations. We also prove a stability result for this larger class.

Note that the above definitions of multiplicity and persistence diagrams of continuous tame functions do not require the homology groups to be finite dimensional. Indeed, it is sufficient to assume that the ranks of all the homomorphisms between the homology groups are finite. This immediately leads to the following generalisation of tameness.

Definition 2.5 *A persistence module \mathcal{F} is said to be 0-tame if*

$$\forall \alpha < \beta, \quad \text{rank } \phi_{\alpha,\beta} < +\infty$$

The persistence modules associated to continuous tame functions or to filtrations of simplicial complexes [13] are obviously 0-tame. For the filtrations defined by the sublevel sets of functions, the main advantage of 0-tameness with respect to the previously mentioned notion of tameness is that it neither requires the vector spaces to be finite dimensional nor the number of critical values to be finite.

In the following section, we show that 0-tameness is a sufficient assumption to define the persistence diagram of a persistence module. Indeed we introduce an even weaker notion in section 3.3, called δ -tameness, which is also sufficient to define the persistence diagrams. But for the sake of clarity we first start by considering 0-tame persistence modules. δ -tameness plays an important role in the study of the stability of persistence diagrams in section 4.3.

3 Discretizing persistence modules

To define the persistence diagram of a persistence module, we proceed in two steps: first, we consider the discretized persistence modules obtained by restricting the family of vector spaces F^α to discrete sequences of indices. The persistence diagram of such discretizations is easily defined in the same way as in the classical setting [8, 13]. Second, we show that the persistence diagram of the whole persistence module \mathcal{F} is obtained as a well-defined limit of the persistence diagrams of its discretizations.

Let $\mathcal{F} = \{F^\alpha\}_{\alpha \in \mathbb{R}}$ be a persistence module. Given two real numbers $\varepsilon > 0$ and $x \in [0, \varepsilon)$, the (ε, x) -discretization of \mathcal{F} is the persistence module $\mathcal{F}_{\varepsilon,x} = \{F^{x+n\varepsilon}\}_{n \in \mathbb{Z}}$. The regular grid $G_{\varepsilon,x} \subset \overline{\mathbb{R}}^2$ defined by

$$G_{\varepsilon,x} = \{(x + i\varepsilon, x + j\varepsilon) \in \mathbb{R}^2 : i, j \in \mathbb{Z}\} \cup \{(x + i\varepsilon, +\infty) \in \mathbb{R} \times \overline{\mathbb{R}}\}$$

is called the *pixelization grid* associated to ε, x . Similarly, given any sequence of real numbers $(\alpha) = (\alpha_n)_{n \in \mathbb{Z}}$ with no accumulation point we can define the (α) -discretization $\mathcal{F}_{(\alpha)}$ of \mathcal{F} and the associated (irregular) pixelization grid $G_{(\alpha)}$ consisting of the points (α_i, α_j) and $(\alpha_i, +\infty)$, $i, j \in \mathbb{Z}$.

Associated to a pixelization grid $G_{\varepsilon,x}$, we associate the (ε,x) -pixelization map $\text{pix}_{\varepsilon,x} : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+^2$ defined by:

$$\forall \alpha \leq \beta, \text{pix}_{\varepsilon,x}(\alpha, \beta) = \begin{cases} \left(x + \lceil \frac{\alpha-x}{\varepsilon} \rceil \varepsilon, x + \lceil \frac{\beta-x}{\varepsilon} \rceil \varepsilon \right) & \text{if } \lceil \frac{\beta-x}{\varepsilon} \rceil > \lceil \frac{\alpha-x}{\varepsilon} \rceil \\ \left(\frac{\alpha+\beta}{2}, \frac{\alpha+\beta}{2} \right) & \text{if } \lceil \frac{\beta-x}{\varepsilon} \rceil = \lceil \frac{\alpha-x}{\varepsilon} \rceil \end{cases}$$

Given any sequence of real numbers $(\alpha) = (\alpha_n)_{n \in \mathbb{Z}}$ with no accumulation point the pixelization map $\text{pix}_{(\alpha)}$ is defined similarly.

The pixelization map $\text{pix}_{\varepsilon,x}$ performs the following snapping operations: each point $(\alpha, \beta) \in \mathbb{R}_+^2$ lying in a grid cell C that does not intersect the diagonal Δ is snapped onto the upper-right corner of C , whereas every point lying in a grid cell that intersects Δ is snapped onto the nearest point of Δ . In particular, diagonal points are left unchanged. Here, each finite grid cell is of the form $(x + m\varepsilon, x + (m+1)\varepsilon] \times (x + n\varepsilon, x + (n+1)\varepsilon]$, where $m, n \in \mathbb{Z}$, *i.e.* it does not include its left and bottom edges. Similarly, each infinite grid cell is either of the form $\{(\pm\infty, \pm\infty)\}$, or of the form $\{\pm\infty\} \times (x + n\varepsilon, x + (n+1)\varepsilon]$, or of the form $(x + n\varepsilon, x + (n+1)\varepsilon] \times \{\pm\infty\}$, where $n \in \mathbb{Z}$.

3.1 The persistence diagram of the discretization of a persistence module

Let now consider a 0-tame persistence module \mathcal{F} and let $\varepsilon > 0$ and $x \in [0, \varepsilon)$ be two real numbers. To simplify the notations in the following, when there is no ambiguity we use the notations F^i instead of $F^{x+i\varepsilon}$ and $F^i \rightarrow F^j$ instead of $F^{x+i\varepsilon} \xrightarrow{\phi_{x+i\varepsilon, x+j\varepsilon}} F^{x+j\varepsilon}$.

Definition 3.1 *The persistence diagram of $\mathcal{F}_{\varepsilon,x}$ is the multi-subset $D\mathcal{F}_{\varepsilon,x}$ of \mathbb{R}^2 defined by:*

- i) $D\mathcal{F}_{\varepsilon,x}$ is contained in $(G_{\varepsilon,x} \cap \Delta_+^\varepsilon) \cup \Delta$ and*
- ii) the multiplicity of each node $m_{i,j} = (x + i\varepsilon, x + j\varepsilon)$, $i < j \leq +\infty$ is given by*

$$\text{mult}(m_{i,j}) = \text{rk}(F^i \rightarrow F^{j-1}) - \text{rk}(F^i \rightarrow F^j) + \text{rk}(F^{i-1} \rightarrow F^j) - \text{rk}(F^{i-1} \rightarrow F^{j-1})$$

By convention, if $j = +\infty$, then $\text{rk}(F^i \rightarrow F^j) = \lim_{k \rightarrow +\infty} \text{rk}(F^i \rightarrow F^k)$ ³.

- iii) the multiplicity of the points on the diagonal Δ is equal to $+\infty$.*

The multiplicity of each point is a non negative integer and (by an elementary computation) the persistence diagram of $\mathcal{F}_{\varepsilon,x}$ satisfies the following property which is illustrated on figure 3:

Lemma 3.2 *For any $i_1, i_2, j_1, j_2 \in \mathbb{Z}$ such that $i_1 < i_2 < j_1 < j_2 \leq +\infty$, we have*

$$\begin{aligned} \sum_{i_1 < i \leq i_2, j_1 < j \leq j_2} \text{mult}(x + i\varepsilon, x + j\varepsilon) &= \text{rank}(F^{i_2} \rightarrow F^{j_1}) - \text{rank}(F^{i_2} \rightarrow F^{j_2}) \\ &\quad + \text{rank}(F^{i_1} \rightarrow F^{j_2}) - \text{rank}(F^{i_1} \rightarrow F^{j_1}) \end{aligned}$$

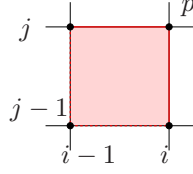


Figure 2: The multiplicity of a node p is determined by the ranks of the four homeomorphisms corresponding to the corners of the bottom left cell adjacent to p .

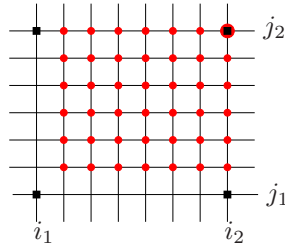


Figure 3: The sum of the multiplicities of the nodes (in red) contained in the box $(x + i_1\varepsilon, x + i_2\varepsilon] \times (x + j_1\varepsilon, x + j_2\varepsilon]$ is equal to the alternate sum of the ranks of the linear maps corresponding to the vertices of the box (black squares).

A consequence of this lemma is that in any half-open quadrant $(-\infty, \alpha] \times (\beta, +\infty]$, the number of points (counted with multiplicity) with non zero multiplicity is finite. It follows that the set of the points of $D\mathcal{F}_{\varepsilon, x}$ with non zero multiplicity does not contain any accumulation point. As a consequence we have the following corollary.

Corollary 3.3 *For any $a < b$, the number of non zero multiplicity points contained in the vertical band $B_{a,b} = \{(u, v) \in \mathbb{R}^2 : a \leq u \leq b\} \cap \Delta_+^\varepsilon$ is finite.*

Proof. Just remark that $B_{a,b}$ is contained in a finite union of boxes of the form $(x + i_1\varepsilon, x + i_2\varepsilon] \times (x + j_1\varepsilon, x + j_2\varepsilon]$. \square

To be able to define the persistence diagram of the persistence module \mathcal{F} it is necessary to compare the persistence diagrams of its different discretizations.

Theorem 3.4 *Let \mathcal{F} be a 0-tame filtration. For any $\varepsilon > 0$ and any $x, y \in \mathbb{R}$ we have*

$$d_B^\infty(D\mathcal{F}_{\varepsilon, x}, D\mathcal{F}_{\varepsilon, y}) < \varepsilon$$

Note that ε can be considered as a scale parameter at which the persistence module is considered. Intuitively the theorem 3.4 states that the knowledge of \mathcal{F} at a scale ε leads to a knowledge of its persistence diagram (that we have not already formally defined) with an uncertainty of ε .

Proof. If $x = y$, the result is obvious. We can now assume without loss of generality that $0 \leq x < y < \varepsilon$. Let $G_{\varepsilon, x, y} \subset \overline{\mathbb{R}^2}$ be the (irregular) grid obtained

³Note that such a limit always exists since the general inequality $\text{rk}(g \circ f) \leq \text{rk} f$ implies that for any fixed i , $\text{rk}(F^i \rightarrow F^k)$ is non increasing. In particular, i being fixed, since the ranks are non negative integers, the sequence $\text{rk}(F^i \rightarrow F^k)$ is constant for k sufficiently large.

as the union of the two grids $G_{\varepsilon,x}$ and $G_{\varepsilon,y}$ and let $\mathcal{F}_{\varepsilon,x,y}$ be the corresponding (irregular) discretization of \mathcal{F} . The proof of the theorem consists in considering $\mathcal{F}_{\varepsilon,x}$ and $\mathcal{F}_{\varepsilon,y}$ as two discretizations of $\mathcal{F}_{\varepsilon,x,y}$ and showing that the persistence diagram of $\mathcal{F}_{\varepsilon,x}$ (respectively $\mathcal{F}_{\varepsilon,y}$) is the image of the persistence diagram of $\mathcal{F}_{\varepsilon,x,y}$ by the pixelization map $\text{pix}_{\varepsilon,x}$ (respectively $\text{pix}_{\varepsilon,y}$). The following lemma follows from an immediate computation using the definition of multiplicity.

Lemma 3.5 *Let $a < c$ and $d < f$ be two pairs of two consecutive values of the sequence $(x + i\varepsilon)_{i \in \mathbb{Z}}$ such that $c < d$ and let (b, e) be the unique node of the grid $G_{\varepsilon,y}$ in the half-open square $(a, c] \times (d, f]$ (see figure 4). We have*

$$\text{mult}_{\mathcal{F}_{\varepsilon,x}}(c, f) = \text{mult}_{\mathcal{F}_{\varepsilon,x,y}}(c, f) + \text{mult}_{\mathcal{F}_{\varepsilon,x,y}}(c, e) + \text{mult}_{\mathcal{F}_{\varepsilon,x,y}}(b, e) + \text{mult}_{\mathcal{F}_{\varepsilon,x,y}}(b, f)$$

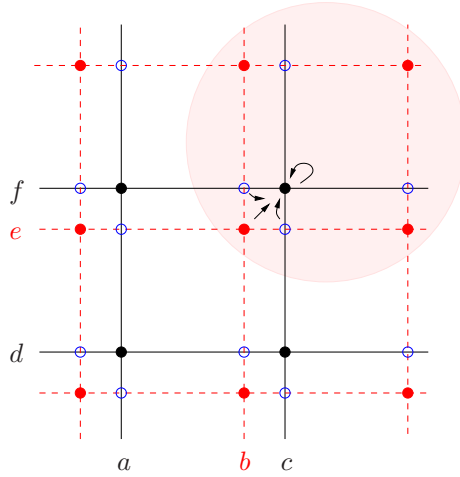


Figure 4: The snapping argument. The closed curve encloses the possible pre-images of the point (c, f) by the snapping bijection between $D\mathcal{F}_{\varepsilon,x,y}$ and $D\mathcal{F}_{\varepsilon,x}$. The possible images of these enclosed points by the pixelization map $\text{pix}_{\varepsilon,y}$ are the corners of the cell of $G_{\varepsilon,y}$ that contains (c, f) .

As a consequence of the lemma 3.5 the restriction to $D\mathcal{F}_{\varepsilon,x,y}$ of the pixelization map $\text{pix}_{\varepsilon,x}$ that “snaps” (c, f) , (c, e) , (b, e) and (b, f) in $G_{\varepsilon,x,y}$ to (c, f) in $G_{\varepsilon,x}$ defines a (multi-)bijection between the persistence diagrams of $\mathcal{F}_{\varepsilon,x,y}$ and $\mathcal{F}_{\varepsilon,x}$ that moves the points by at most ε ⁴. Using the same argument with $\mathcal{F}_{\varepsilon,y}$ and the pixelization map $\text{pix}_{\varepsilon,y}$ we immediately deduce that $d_B^\infty(D\mathcal{F}_{\varepsilon,x}, D\mathcal{F}_{\varepsilon,y}) < 2\varepsilon$. Looking at the possible images of a point of $D\mathcal{F}_{\varepsilon,x}$ by the composition of the two previously defined snapping bijections it appears that the image of a point of $D\mathcal{F}_{\varepsilon,x}$ is contained in the union of the corners of the cell of $G_{\varepsilon,y}$ that contains it (see figure 4). It follows that $d_B^\infty(D\mathcal{F}_{\varepsilon,x}, D\mathcal{F}_{\varepsilon,y}) < \varepsilon$. \square

3.2 The persistence diagram of 0-tame persistence modules

Intuitively the theorem 3.4 states that the persistence diagram of a tame filtration \mathcal{F} at a given scale ε is well-defined up to a multibijection that moves the

⁴in fact, the points move by no more than $\max(|x - y|, \varepsilon - |x - y|)$

points by no more than ε . This makes possible to define the persistence diagram of \mathcal{F} as a limit of the persistence diagrams of its pixelizations.

Let $\mathcal{F}_{\varepsilon,x}$ be a pixelization of a 0-tame filtration \mathcal{F} . The pixelization $\mathcal{F}_{\varepsilon,x}$ can be considered as a $\frac{\varepsilon}{2}$ -pixelization of $\mathcal{F}_{\frac{\varepsilon}{2},x}$. The multiplicity in $D\mathcal{F}_{\varepsilon,x}$ of $(u, v) \in G_{\varepsilon,x} \cap \Delta_+^\varepsilon$ is related to the multiplicities of its neighbors in $D\mathcal{F}_{\frac{\varepsilon}{2},x}$ by the formula

$$\text{mult}_{\mathcal{F}_{\varepsilon,x}}(u, v) = \text{mult}_{\mathcal{F}_{\frac{\varepsilon}{2},x}}(u, v) + \text{mult}_{\mathcal{F}_{\frac{\varepsilon}{2},x}}(u, v - \frac{\varepsilon}{2}) + \text{mult}_{\mathcal{F}_{\frac{\varepsilon}{2},x}}(u - \frac{\varepsilon}{2}, v - \frac{\varepsilon}{2}) + \text{mult}_{\mathcal{F}_{\frac{\varepsilon}{2},x}}(\frac{\varepsilon}{2}, v)$$

which is obtained by applying the lemma 3.5 to (u, v) where y is chosen to be equal to $x - \frac{\varepsilon}{2}$. Let $M_\varepsilon = D\mathcal{F}_{\varepsilon,x} \cap \Delta_+^\varepsilon = \{m_1, m_2, \dots\}$ be the nodes of $G_{\varepsilon,x}$ with non-zero multiplicity that are contained in Δ_+^ε . It follows from the previous equality that the nodes of $G_{\frac{\varepsilon}{2},x}$ with non zero multiplicities have to be contained in an $\frac{\varepsilon}{2}$ -neighborhood of the points of M_ε represented on the figure 5. More precisely, they have to be contained in the corners of the bottom left half-open cells C_i of the points m_i . Moreover, the sum of the multiplicities of the points in the cell C_i has to be equal to the multiplicity of m_i by lemma 3.2. Since the points of $D\mathcal{F}_{\varepsilon,x}$ and $D\mathcal{F}_{\frac{\varepsilon}{2},x}$ that are contained in the complement of Δ_+^ε are at distance at most $\frac{\varepsilon}{2}$ from the diagonal Δ we obtain

$$d_B^\infty(D\mathcal{F}_{\varepsilon,x}, D\mathcal{F}_{\frac{\varepsilon}{2},x}) < \frac{\varepsilon}{2} \tag{3}$$

Moreover the restriction of the pixelization map $\text{pix}_{\varepsilon,x}$ to $D\mathcal{F}_{\frac{\varepsilon}{2},x}$ is a multi-bijection between $D\mathcal{F}_{\frac{\varepsilon}{2},x}$ and $D\mathcal{F}_{\varepsilon,x}$ that moves the points by at most $\frac{\varepsilon}{2}$.

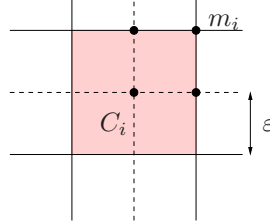


Figure 5: The black discs represents the points of $D\mathcal{F}_{\frac{\varepsilon}{2},x}$ with possibly non-zero multiplicity that are located around a point m_i of M .

By subdividing $\mathcal{F}_{\frac{\varepsilon}{2},x}$ we can now iterate the previous construction. Since the sum of the multiplicities in each C_i is preserved at each step of the iteration and since the multiplicities are finite non-negative integers the multisets $M_{\frac{\varepsilon}{2^n}}$ have to converge to a multiset in Δ_+ . So the persistence diagrams $D\mathcal{F}_{\frac{\varepsilon}{2^n}}$ converges to a limit diagram $D\mathcal{F}$:

$$\lim_{n \rightarrow +\infty} d_B^\infty(D\mathcal{F}, D\mathcal{F}_{\frac{\varepsilon}{2^n},x}) = 0$$

Definition 3.6 *The previously defined diagram $D\mathcal{F}$ is called the persistence diagram of the 0-tame filtration \mathcal{F} . If \mathcal{F} is the persistence module associated to the sublevel sets filtration of a function f , we denote its persistence diagram by Df .*

Proposition 3.7 *The persistence diagram $D\mathcal{F}$ does not depend on the choice of x and ε .*

Proof. First, if $x \neq y$, then it follows from theorem 3.4 that for all $n \in \mathbb{N}$

$$d_B^\infty(D\mathcal{F}_{\frac{\varepsilon}{2^n}, x}, D\mathcal{F}_{\frac{\varepsilon}{2^n}, y}) < \frac{\varepsilon}{2^n} \xrightarrow{n \rightarrow \infty} 0$$

So the two limit diagrams obtained from $\mathcal{F}_{\varepsilon, x}$ and $\mathcal{F}_{\varepsilon, y}$ have to be the same. Now, if $\varepsilon \neq \varepsilon'$ are two positive real numbers, consider the irregular grids $G_{\frac{\varepsilon}{2^n}, \frac{\varepsilon'}{2^n}, x}$ obtained as the union of the grids $G_{\frac{\varepsilon}{2^n}, x}$ and $G_{\frac{\varepsilon'}{2^n}, x}$ for any $n \in \mathbb{N}$. Applying the same argument as the previous ones shows that these discretizations converge to a limit diagram. Denoting by $\mathcal{F}_{\frac{\varepsilon}{2^n}, \frac{\varepsilon'}{2^n}, x}$ the corresponding discretization of \mathcal{F} , $\mathcal{F}_{\frac{\varepsilon}{2^n}, x}$ and $\mathcal{F}_{\frac{\varepsilon'}{2^n}, x}$ are discretizations of $\mathcal{F}_{\frac{\varepsilon}{2^n}, \frac{\varepsilon'}{2^n}, x}$. As a consequence the two limit diagrams obtained from $\mathcal{F}_{\varepsilon, x}$ and $\mathcal{F}_{\varepsilon', x}$ have to be the same. \square

Another immediate but important property is that the bottleneck distance between \mathcal{F} and any of its ε -pixelization is bounded by ε .

Theorem 3.8 *Let \mathcal{F} be a 0-tame persistence module. Then for any $\varepsilon > 0$ and $x \in \mathbb{R}$ we have*

$$d_B^\infty(D\mathcal{F}, D\mathcal{F}_{\varepsilon, x}) < \varepsilon$$

Moreover, the restriction of $\text{pix}_{\varepsilon, x}$ to $D\mathcal{F}$ is a multi-bijection between $D\mathcal{F}$ and $D\mathcal{F}_{\varepsilon, x}$ that moves the points by no more than ε .

In other words this result means that if only an ε -discretization of a persistence module is known, then its persistence diagram is known up to an ε -approximation (for the bottleneck distance).

Proof. The inequality is an immediate consequence of the inequality 3 since $\sum_{i=1}^{\infty} \frac{\varepsilon}{2^i} = 1$. For the second part of the theorem, note that we have

$$\text{pix}_{\varepsilon, x} = \text{pix}_{\varepsilon, x} \circ \text{pix}_{\frac{\varepsilon}{2}, x} \circ \cdots \circ \text{pix}_{\frac{\varepsilon}{2^n}, x}$$

So the restriction of $\text{pix}_{\varepsilon, x}$ to $D\mathcal{F}_{\frac{\varepsilon}{2^n}, x}$ is a multi-bijection between $D\mathcal{F}_{\frac{\varepsilon}{2^n}, x}$ and $D\mathcal{F}_{\varepsilon, x}$. As a consequence, the restriction of $\text{pix}_{\varepsilon, x}$ to $D\mathcal{F}$ is a multi-bijection between $D\mathcal{F}$ and $D\mathcal{F}_{\varepsilon, x}$ that moves the points by no more than ε . \square

Remark 1 *If $f : \mathbb{X} \rightarrow \mathbb{R}$ is a tame function in the sense of definition 2.3, then its persistence diagram $D'f$ as defined in [4] is equal to the above defined persistence diagram Df . To prove it, it is sufficient to choose $x \in \mathbb{R}$ and $\varepsilon > 0$ such that the homological critical values a_1, \dots, a_m of f are never equal to one of the coordinates of the nodes of the grids $G_{\frac{\varepsilon}{2^n}, x}$, $n \in \mathbb{N}$ ⁵. So, at any step n of the subdivision process each point (a_i, a_j) of $D'f \setminus \Delta$ is contained in the interior of a cell $C_{i, j, n}$ of $G_{\frac{\varepsilon}{2^n}, x}$ and the points of $Df_{\frac{\varepsilon}{2^n}, x}$ with non zero multiplicity are contained in the upper right corners of such cells $C_{i, j, n}$. Moreover, the multiplicity of (a_i, a_j) is equal to the multiplicity of the upper right corner of $C_{i, j, n}$. As a consequence, $Df = D'f$.*

⁵such a choice is possible since the set of homological critical values of f is finite and the set of coordinates of the nodes is numerable.

3.3 δ -tameness and persistence diagrams

The previous construction of persistence diagram immediately extends to a slightly weaker notion of tameness that will reveal useful in section 4.3. First note that the construction of the persistence diagrams of 0-tame persistence modules only needs the multiplicities of the points to be finite, that is the rank of the maps $\phi_{\alpha,\beta}$ to be finite. Now, assume that we are just interested in the part of the persistence diagram located in Δ_+^δ for some fixed $\delta > 0$. Then the multiplicity of any node in $G_{\varepsilon,x} \cap \Delta_+^{\delta+\varepsilon}$ is well-defined and finite as soon as the ranks of the maps $\phi_{\alpha,\beta}$ are finite for all α,β such that $\beta - \alpha > \delta + \varepsilon$.

Definition 3.9 *Given $\delta > 0$, a persistence module \mathcal{F} is said to be δ -tame if*

$$\forall \alpha < \alpha + \delta < \beta, \quad \text{rank } \phi_{\alpha,\beta} < +\infty$$

The constructions done in the previous section immediately apply to δ -tame persistence modules: the persistence diagram of any ε -discretization of a δ -tame persistence module is well defined in $\Delta_+^{\delta+\varepsilon}$ and it converges to a well defined multiset on Δ_+^δ when $\varepsilon \rightarrow 0$. We thus obtain the following notion of persistence diagram.

Definition 3.10 *Let \mathcal{F} be a δ -tame persistence module for some $\delta \geq 0$. The δ -persistence diagram of \mathcal{F} is the multiset $D_\delta \mathcal{F}$ obtained as the union of the persistence diagram of \mathcal{F} on Δ_+^δ and of the line $\Delta_\delta = \{(u,v) \in \mathbb{R}^2 : v - u = \delta\}$ where the points of Δ_δ have infinite multiplicity. For any $\alpha \geq \delta$, the α -persistence diagram of \mathcal{F} is the multiset defined by*

$$D_\alpha \mathcal{F} = (D_\delta \mathcal{F} \cap \Delta_+^\alpha) \cup \Delta_\alpha$$

where the points of the line $\Delta_\alpha = \{(u,v) \in \mathbb{R}^2 : v - u = \alpha\}$ have infinite multiplicity.

Intuitively, considering the δ -persistence diagram of a persistence module consists in discarding the “events” which life no longer than δ . So δ can be seen as a scale parameter quantifying a minimum relevant “life time” for the changes in the persistence module. It is important to note that this scale parameter is different from the parameter ε used for the discretization in the previous sections: ε quantify an uncertainty on the life time of the “events” in the persistence module.

The persistence diagrams of δ -tame filtrations share the same properties as the persistence diagrams of 0-tame filtrations. First, we have the following obvious property.

Lemma 3.11 *If \mathcal{F} is a δ -tame filtration then for any $\alpha, \alpha' \geq \delta$ we have*

$$d_H(D_\alpha \mathcal{F}, D_{\alpha'} \mathcal{F}) = d_B^\infty(D_\alpha \mathcal{F}, D_{\alpha'} \mathcal{F}) = |\alpha' - \alpha|$$

Proof. Assuming that $\alpha' > \alpha$ the multibijection is realized by just translating $\Delta_{\alpha'}$ onto Δ_α and snapping the points of $D_{\alpha'} \mathcal{F}$ that are below $\Delta_{\alpha'}$ onto their nearest neighbor on Δ_α . \square

The lemma 3.2 still holds for the persistence diagrams of δ -tame filtrations.

Lemma 3.12 *Let $\delta \geq 0$ and let $\mathcal{F} = \{F^\alpha\}_{\alpha \in \mathbb{R}}$ be a δ -tame filtration. For any $a < b < b + \delta < c < d$ the sum of the multiplicities of the points of $D_\delta \mathcal{F}$ in the box $[a, b] \times [c, d]$ is equal to*

$$\text{rank}(F^b \rightarrow F^c) - \text{rank}(F^b \rightarrow F^d) + \text{rank}(F^a \rightarrow F^d) - \text{rank}(F^a \rightarrow F^c)$$

Proof. Since the definition of the persistence diagram of \mathcal{F} does not depend on the choice of the initial discretization, we can choose x and ε such that the point (b, d) is a node of the pixelization grid $G_{\varepsilon, x}$ and $0 < \varepsilon < (c - b) - \delta$. As a consequence, the lemma 3.2 holds for all the pixelizations $D\mathcal{F}_{\frac{\varepsilon}{2^n}, x}$. Since the pixelization maps snap the points to the upper-right corner of their cell, the lemma 3.2 also holds for the limit diagram $D\mathcal{F}$. \square

As in the case of pixelized persistence diagrams, the previous lemma admits the following corollary.

Corollary 3.13 *For any $a < b$ and any $\varepsilon > 0$, the number of non zero multiplicity points contained in the vertical band $B_{a, b} = \{(u, v) \in \mathbb{R}^2 : a \leq u \leq b\} \cap \Delta_+^{\delta + \varepsilon}$ is finite.*

Generalizing the theorem 3.8, the distance between the persistence diagram of a δ -tame persistence module and any of its ε -discretization is still bounded by ε .

Theorem 3.14 *Let \mathcal{F} be a δ -tame persistence module. Then for any $\varepsilon > 0$ and $x \in \mathbb{R}$ we have*

$$d_B^\infty(D_\delta \mathcal{F}, D_\delta \mathcal{F}_{\varepsilon, x}) < \varepsilon$$

Moreover, the restriction of $\text{pix}_{\varepsilon, x}$ to $D_\delta \mathcal{F}$ is a multi-bijection between $D_\delta \mathcal{F}$ and $D_\delta \mathcal{F}_{\varepsilon, x}$ that moves the points by no more than ε .

This last result plays an important role in the study of the stability of persistence diagrams in section 4.

4 Stability of persistence diagrams

In this section, we provide an equivalent of the stability result of [4] that holds in the more general setting of δ -tame filtrations. For this purpose, we need to introduce a quantitative notion of closeness between persistence modules. We first consider the case of the persistence modules associated to sublevel sets filtrations of functions. This motivates a definition of closeness between general persistence modules.

4.1 Interleaving persistence modules

Let first consider how the sublevel sets of two functions defined on the same topological space include in each other.

Definition 4.1 *Given $\varepsilon > 0$ and two real-valued functions f, g defined over a same topological space \mathbb{X} ,*

- *f, g are weakly ε -interleaved if there exist some $a \in \mathbb{R}$ such that $\forall n \in \mathbb{Z}$, $F^{a+2n\varepsilon} \subseteq G^{a+(2n+1)\varepsilon} \subseteq F^{a+2(n+1)\varepsilon}$,*

- f, g are strongly ε -interleaved if for all $a \in \mathbb{R}$, $\forall n \in \mathbb{Z}$, $F^{a+2n\varepsilon} \subseteq G^{a+(2n+1)\varepsilon} \subseteq F^{a+2(n+1)\varepsilon}$.

where $F^\alpha = f^{-1}((-\infty, \alpha])$, $G^\alpha = g^{-1}((-\infty, \alpha])$.

The following lemma relate the two notions of interleaved filtrations to the distance (for the $\|\cdot\|_\infty$ -norm) between the functions.

Lemma 4.2 *Let f, g be two real-valued functions defined over a same topological space \mathbb{X} . Then,*

- (i) *if f, g are strongly ε -interleaved, then they are weakly ε -interleaved,*
- (ii) *if f, g are weakly ε -interleaved, then they are strongly 3ε -interleaved,*
- (iii) *f, g are strongly ε -interleaved if and only if $\|f - g\|_\infty \leq \varepsilon$.*

Proof. Assertion (i) is a direct consequence of Definition 4.1.

Assume now that f, g are weakly ε -interleaved. Let $a \in \mathbb{R}$ be such that f, g are (ε, a) -interleaved. Given $\alpha \in \mathbb{R}$, let $n \in \mathbb{Z}$ be such that $a + 2n\varepsilon < \alpha \leq a + 2(n+1)\varepsilon$. Then, we have $F^\alpha \subseteq F^{a+2(n+1)\varepsilon} \subseteq G^{a+(2n+3)\varepsilon}$, which is included in $G^{\alpha+3\varepsilon}$ since $\alpha > a + 2n\varepsilon$. Symmetrically, we also have $G^\alpha \subseteq F^{\alpha+3\varepsilon}$. Since this is true for all $\alpha \in \mathbb{R}$, f and g are strongly 3ε -interleaved. This proves (ii).

Assume now that f, g are strongly ε -interleaved. For all $x \in \mathbb{X}$, we have $x \in f^{-1}((-\infty, f(x)]) = F^{f(x)} \subseteq G^{f(x)+\varepsilon} = g^{-1}((-\infty, f(x) + \varepsilon])$, which implies that $g(x) \leq f(x) + \varepsilon$. By symmetry, we also have $f(x) \leq g(x) + \varepsilon$. Hence, $|f(x) - g(x)| \leq \varepsilon$. Since this is true for all $x \in \mathbb{X}$, we deduce that $\|f - g\|_\infty \leq \varepsilon$. Conversely, assume that $\|f - g\|_\infty \leq \varepsilon$. Then, for all $\alpha \in \mathbb{R}$ and all $x \in F^\alpha$, we have $f(x) \leq \alpha$, and therefore $g(x) \leq f(x) + \varepsilon \leq \alpha + \varepsilon$, which implies that $x \in G^{\alpha+\varepsilon}$. Thus, $F^\alpha \subseteq G^{\alpha+\varepsilon}$. Symmetrically, we also have $G^\alpha \subseteq F^{\alpha+\varepsilon}$. Since this is true for all $\alpha \in \mathbb{R}$, we conclude that f, g are strongly ε -interleaved. This proves (iii). \square

Now, the canonical inclusion maps between the sublevel sets of two ε -interleaved functions f and g induce the following commutative diagram between the ε -pixelizations of their persistence modules

$$\begin{array}{ccccccc}
 \dots & \longrightarrow & H_k(F^{a+2n\varepsilon}) & \longrightarrow & H_k(F^{a+(2n+1)\varepsilon}) & \longrightarrow & H_k(F^{a+(2n+2)\varepsilon}) & \longrightarrow & \dots \\
 & & \nearrow & & \searrow & & \nearrow & & \searrow \\
 \dots & \longrightarrow & H_k(G^{a+2n\varepsilon}) & \longrightarrow & H_k(G^{a+(2n+1)\varepsilon}) & \longrightarrow & H_k(G^{a+(2n+2)\varepsilon}) & \longrightarrow & \dots
 \end{array}$$

where the arrows represents the homomorphisms induced by the inclusion maps. The persistence modules of f and g are said to be *weakly ε -interleaved* (resp. *strongly ε -interleaved*) if the previous diagram is commutative for some $a \in \mathbb{R}$ (resp. for all $a \in \mathbb{R}$). This definition immediately extends to general persistence modules

Definition 4.3 *Two persistence modules $\mathcal{F} = \{F^\alpha\}_{\alpha \in \mathbb{R}}$ and $\mathcal{G} = \{G^\alpha\}_{\alpha \in \mathbb{R}}$ are said to be weakly (resp. strongly) ε -interleaved if for some $a \in \mathbb{R}$ (resp. if for all $a \in \mathbb{R}$) and for all $n \in \mathbb{Z}$ there exist homomorphisms $\Phi_{a+2n\varepsilon} : F^{a+2n\varepsilon} \rightarrow G^{a+(2n+1)\varepsilon}$ and $\Psi_{a+(2n+1)\varepsilon} : G^{a+(2n+1)\varepsilon} \rightarrow F^{a+(2n+2)\varepsilon}$ that make the following*

diagram commutative

$$\begin{array}{ccccccc}
 \cdots & \longrightarrow & F^{a+2n\varepsilon} & \longrightarrow & F^{a+(2n+1)\varepsilon} & \longrightarrow & F^{a+(2n+2)\varepsilon} & \longrightarrow & \cdots \\
 & & \nearrow & & \searrow & & \nearrow & & \searrow \\
 \cdots & \longrightarrow & G^{a+2n\varepsilon} & \longrightarrow & G^{a+(2n+1)\varepsilon} & \longrightarrow & G^{a+(2n+2)\varepsilon} & \longrightarrow & \cdots
 \end{array} \quad (4)$$

The two pixelizations $\mathcal{F}_{\varepsilon,a}$ and $\mathcal{G}_{\varepsilon,a}$ are also said to be ε -interleaved.

4.2 Stability of persistence diagrams of δ -tame persistence modules

We are now able to prove that closely interleaved tame persistence modules have close persistence diagrams.

Theorem 4.4 *Let \mathcal{F} and \mathcal{G} be two weakly (or strongly) ε -interleaved persistence modules that are δ -tame for some $\delta \geq 0$. Then*

$$d_B^\infty(D_\delta \mathcal{F}, D_\delta \mathcal{G}) \leq 3\varepsilon$$

Proof. Let $a \in \mathbb{R}$ be such that the pixelizations $\mathcal{F}_{\varepsilon,a}$ and $\mathcal{G}_{\varepsilon,a}$ are ε interleaved. Let $\mathcal{H} = \{H^n\}_{n \in \mathbb{Z}}$ be the discrete persistence module defined by:

- $H^{2n} = F^{a+2n\varepsilon}$ and $H^{2n+1} = G^{a+(2n+1)\varepsilon}$ for all $n \in \mathbb{Z}$,
- $\phi_{2n,2n+1} = \Phi_{a+2n\varepsilon}$ and $\phi_{2n+1,2n+2} = \Psi_{a+(2n+1)\varepsilon}$ for all $n \in \mathbb{Z}$

so that $\mathcal{F}_{2\varepsilon,a}$ and $\mathcal{G}_{2\varepsilon,a+\varepsilon}$ are 2ε -discretization of \mathcal{H} as well as 2ε -discretization of \mathcal{F} and \mathcal{G} respectively. Note that \mathcal{H} is not necessarily δ -tame but $(\delta + 2\varepsilon)$ -tame. Using theorem 3.14 we have

$$d_B^\infty(D_\delta \mathcal{F}, D_\delta \mathcal{F}_{2\varepsilon,a}) < 2\varepsilon, \quad d_B^\infty(D_\delta \mathcal{G}, D_\delta \mathcal{G}_{2\varepsilon,a+\varepsilon}) < 2\varepsilon$$

and, since $\mathcal{F}_{2\varepsilon,a}$ and $\mathcal{G}_{2\varepsilon,a+\varepsilon}$ are 2ε -pixelizations of \mathcal{H} ,

$$d_B^\infty(D_{\delta+2\varepsilon} \mathcal{F}_{2\varepsilon,a}, D_{\delta+2\varepsilon} \mathcal{G}_{2\varepsilon,a+\varepsilon}) < 2\varepsilon$$

Since $\mathcal{F}_{2\varepsilon,a}$ and $\mathcal{G}_{2\varepsilon,a+\varepsilon}$ are δ -tame and the points of their δ -persistence diagrams that are contained in the complement of $\Delta_+^{\delta+2\varepsilon}$ are at distance less than 2ε of the line $\Delta_\delta = \{(u, v) : v - u = \delta\}$, we indeed have

$$d_B^\infty(D_\delta \mathcal{F}_{2\varepsilon,a}, D_\delta \mathcal{G}_{2\varepsilon,a+\varepsilon}) < 2\varepsilon$$

As an immediate consequence using the triangular inequality we obtain that

$$d_B^\infty(D_\delta \mathcal{F}, D_\delta \mathcal{G}) < 6\varepsilon$$

To improve the bound from 6ε to 3ε we need to consider how the points of the above diagrams are moved by the multibijections induced by the pixelization maps. Let denote by γ_1 (resp. γ_2) the multibijection between $D_{\delta+2\varepsilon} \mathcal{F}$ and $D_{\delta+2\varepsilon} \mathcal{F}_{2\varepsilon,a}$ (resp. between $D_{\delta+2\varepsilon} \mathcal{G}$ and $D_{\delta+2\varepsilon} \mathcal{G}_{2\varepsilon,a+\varepsilon}$) induced by the pixelization map $\text{pix}_{2\varepsilon,a}$ (resp. $\text{pix}_{2\varepsilon,a+\varepsilon}$). Let denote by γ_3 (resp. γ_4) the multibijection between $D_{\delta+2\varepsilon} \mathcal{H}$ and $D_{\delta+2\varepsilon} \mathcal{F}_{2\varepsilon,a}$ (resp. between $D_{\delta+2\varepsilon} \mathcal{H}$ and

$D_{\delta+2\varepsilon}\mathcal{G}_{2\varepsilon,a+\varepsilon}$) induced by the pixelization map $\text{pix}_{2\varepsilon,a}$ (resp. $\text{pix}_{2\varepsilon,a+\varepsilon}$). So the map $\gamma = \gamma_2^{-1} \circ \gamma_4 \circ \gamma_3^{-1} \circ \gamma_1$ is a multibijection between $D_{\delta+2\varepsilon}\mathcal{F}$ and $D_{\delta+2\varepsilon}\mathcal{G}$. Now consider the possible images of a point $p \in D_{\delta+2\varepsilon}\mathcal{F}$ by this multibijection. We have that $\gamma_1(p) = (u, v) \in G_{2\varepsilon,a}$ and $\gamma_3^{-1} \circ \gamma_1(p)$ is contained in the four corners of the bottom left cell of p in $G_{\varepsilon,a}$: (u, v) , $(u, v - \varepsilon)$, $(u - \varepsilon, v)$ and $(u - \varepsilon, v - \varepsilon)$ (see figure 6). Now, the images of these four corners by γ_4 are contained in the four points $(u - \varepsilon, v - \varepsilon)$, $(u - \varepsilon, v + \varepsilon)$, $(u + \varepsilon, v - \varepsilon)$ and $(u + \varepsilon, v + \varepsilon)$. Since γ_2 is the restriction of the pixelization map onto the grid $G_{2\varepsilon,a+\varepsilon}$ the possible images of $\gamma_4 \circ \gamma_3^{-1} \circ \gamma_1(p)$ by γ_2^{-1} are contained in the union of the bottom left cells of the four previously defined points in $G_{2\varepsilon,a+\varepsilon}$, i.e. γ is contained in the box $[u - 3\varepsilon, u + \varepsilon] \times [v - 3\varepsilon, v + \varepsilon]$ (see figure 6). As a consequence p is moved by no more than 3ε . To conclude the proof, we can extend γ to a multibijection between $D_\delta\mathcal{F}$ and $D_\delta\mathcal{G}$ by snapping the points that are between Δ_δ and $\Delta_{\delta+2\varepsilon}$ onto Δ_δ . This last operation does not move the points by more than ε . Moreover, it follows from the above analysis that the points of $D_{\delta+2\varepsilon}\mathcal{F}$ that can be mapped onto $\Delta_{\delta+2\varepsilon}$ by the multibijection γ are at distance less than 1.5ε from $\Delta_{\delta+2\varepsilon}$. As a consequence, composing γ with the snapping map between Δ_δ and $\Delta_{\delta+2\varepsilon}$ onto Δ_δ cannot move the points of $D_\delta\mathcal{F}$ by more than 3ε , so we obtain $d_B^\infty(D_\delta\mathcal{F}, D_\delta\mathcal{G}) \leq 3\varepsilon$. \square

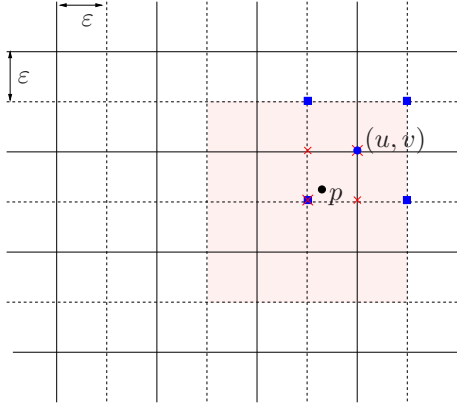


Figure 6: Possible images of p by the multibijection $\gamma_2^{-1} \circ \gamma_4 \circ \gamma_3^{-1} \circ \gamma_1$: $\gamma_1(p)$ is the blue disc with coordinates (u, v) ; $\gamma_3^{-1} \circ \gamma_1(p)$ is contained in the four red crosses; $\gamma_4 \circ \gamma_3^{-1} \circ \gamma_1(p)$ is contained in the four blue squares and thus the image of p is contained in the pink box.

Note that, if the persistence modules are only weakly interleaved, the upper bound in Theorem 4.4 is tight. Indeed, given $\varepsilon > 0$, choose any arbitrary $\eta \in (0, \varepsilon)$ and consider the segment $\mathbb{X} = [v_0, v_1]$ where $v_0 < v_1$ are two real numbers and the two functions $f, g : \mathbb{X} \rightarrow \mathbb{R}$ defined by $f(v_0) = -\infty$, $f(v_1) = \eta$, $g(v_0) = -\infty$, $g(v_1) = 3\varepsilon - \eta$ and $f(v) = g(v) = 10\varepsilon$ for all $v \in (v_0, v_1)$. Denote by \mathcal{F}_η and \mathcal{G}_η the persistence modules induced by these functions at the 0-dimensional homology level. Clearly, \mathcal{F}_η and \mathcal{G}_η are 0-tames and $(\varepsilon, 0)$ -interleaved. Thus, they satisfy the hypotheses of Theorem 4.4. Now, we have $D\mathcal{F}_\eta = \{(-\infty, +\infty), (\eta, 10\varepsilon)\} \cup \Delta$ and $D\mathcal{G}_\eta = \{(-\infty, +\infty), (3\varepsilon - \eta, 10\varepsilon)\} \cup \Delta$, which implies that $d_B^\infty(D\mathcal{F}_\eta, D\mathcal{G}_\eta) = 3\varepsilon - 2\eta$. Since such pairs of filtrations $\mathcal{F}_\eta, \mathcal{G}_\eta$ exist for all $\eta \in (0, \varepsilon)$, the upper bound of 3ε in Theorem 4.4 is tight.

Open problem: Although the bound of 3ε is tight in Theorem 4.4 for weakly interleaved persistence modules, we do not know whether it is tight for strongly interleaved persistence modules. In the next section we prove that it can indeed be improved to ε for persistence modules associated to sublevel sets filtrations of functions. Unfortunately, the proof given in such a setting does not seem to generalize to the case of general persistence modules.

Considering, the Hausdorff distance between the persistence diagrams of strongly ε -interleaved filtrations instead of the bottleneck distance allows to obtain a bound of ε .

Lemma 4.5 (Box lemma [4]) *Let \mathcal{F} and \mathcal{G} be two strongly ε -interleaved persistence modules that are δ -tame for some $\delta \geq 0$. For $a < b < b + \delta < c < d$ let $R = [a, b] \times [c, d] \subset \overline{\mathbb{R}^2}$ and let $R_\varepsilon = [a + \varepsilon, b - \varepsilon] \times [c + \varepsilon, d - \varepsilon]$ be the box obtained by shrinking R by ε . Then the sum of the multiplicities of the points of $D_\delta \mathcal{F}$ contained in R_ε is not greater than the sum of the multiplicities of the points of $D_\delta \mathcal{G}$ contained in R .*

Proof. Using the lemma 3.12, the proof of this box lemma is exactly the same as the one given in [4]⁶. \square

As a consequence, we obtain an upper bound on the Hausdorff distance between strongly interleaved persistence modules.

Theorem 4.6 *Let \mathcal{F} and \mathcal{G} be two strongly ε -interleaved persistence modules that are δ -tame for some $\delta \geq 0$. Then*

$$d_H(D_\delta \mathcal{F}, D_\delta \mathcal{G}) < \varepsilon$$

Proof. This is an immediate consequence of the Box Lemma 4.5: since \mathcal{F} and \mathcal{G} are strongly ε -interleaved, for any point (u, v) of $D_\delta \mathcal{F}$ the sum of the multiplicities of the points of $D_\delta \mathcal{G}$ in the box $[u - \varepsilon, u + \varepsilon] \times [v - \varepsilon, v + \varepsilon]$ is positive. As a consequence there exists a point p' of $D_\delta \mathcal{G}$ at distance less than ε from p . Symmetrically, any point of $D_\delta \mathcal{G}$ is at distance at most ε from a point of $D_\delta \mathcal{F}$. This concludes the proof. \square

4.3 The case of persistence modules associated to sublevel sets filtrations

In this section we improve the theorem 4.4 for the persistence modules associated to sublevel sets filtrations of functions.

Definition 4.7 *Given $\delta \geq 0$, a function $f : \mathbb{X} \rightarrow \mathbb{R}$ defined on a topological space \mathbb{X} is said to be δ -tame, if the persistence module induced at the homology level by its sublevel sets filtration is δ -tame. The δ -persistence diagram of the persistence module induced by f at the k -dimensional homology level is denoted by $D_\delta^k f$.*

⁶Note that as in [4], we can assume first that the boundaries of R and R_ε do not contain any point of the persistence diagrams of \mathcal{F} and \mathcal{G} . Second, if some points of the persistence diagrams are contained in the boundaries of R and R_ε , the same perturbation argument as the one given in [4] leads to the result.

The main result of this section is the following upper bound on the bottleneck distance between persistence diagrams of δ -tame functions.

Theorem 4.8 *Let $\delta \geq 0$ and let $f, g : \mathbb{X} \rightarrow \mathbb{R}$ be two δ -tame functions such that $\|g - f\|_\infty < \varepsilon$ for some $\varepsilon > 0$. Then for any $k \in \mathbb{N}$ we have*

$$d_B^\infty(D_\delta^k f, D_\delta^k g) < \varepsilon$$

Note that this theorem extends the stability result of [4]. Its main advantage is that any assumption on the continuity of f and g or the triangulability of \mathbb{X} is required.

Up to technical difficulties, the thread of the proof is the same as the one of the stability theorem given in [4]. For the sake of simplicity and to avoid too heavy notations, we assume for the proof of the theorem 4.8 that f and g are 0-tame (the proof in the general case being exactly the same). We also assume that $k \in \mathbb{N}$ is fixed and we omit it in the notations. At last recall that we use the following (non confuding) abuse of notation: F^α denotes indifferently the sublevel set $f^{-1}((-\infty, \alpha])$ and its k -th homology group.

Lemma 4.9 *Let $f, g : \mathbb{X} \rightarrow \mathbb{R}$ be two 0-tame functions such that $\|g - f\|_\infty < \varepsilon$ for some $\varepsilon > 0$ and let $h_s : \mathbb{X} \rightarrow \mathbb{R}$ be defined by $h_s = (1-s)f + sg = f + s(g-f)$ for $s \in [0, 1]$. We denote by \mathcal{F} , \mathcal{G} and \mathcal{H}_s the corresponding persistence modules induced by the sublevel sets filtrations. We have the following properties:*

- (i) $\mathcal{H}_0 = \mathcal{F}$ and $\mathcal{H}_1 = \mathcal{G}$;
- (ii) the filtration \mathcal{H}_s is $2s\varepsilon$ -tame (and thus 2ε -tame);
- (iii) $\forall s, s' \in [0, 1]$, \mathcal{H}_s and $\mathcal{H}_{s'}$ are strongly $|s - s'|\varepsilon$ -interleaved.

Proof. The only non obvious point is (ii). It follows from the following remark. Since $\|f - h_s\|_\infty = \|s(g - f)\|_\infty < s\varepsilon$, for any $\alpha \in \mathbb{R}$ there exists $\eta > 0$ such that

$$F^{\alpha - s\varepsilon} \rightarrow H_s^\alpha \rightarrow F^{\alpha + s\varepsilon} \rightarrow F^{\alpha + s\varepsilon + \eta} \rightarrow H_s^{\alpha + 2s\varepsilon}$$

Using that f is 0-tame, $\text{rk}(F^{\alpha + s\varepsilon} \rightarrow F^{\alpha + s\varepsilon + \delta}) < +\infty$ and thus $\text{rk}(H_s^\alpha \rightarrow H_s^{\alpha + 2s\varepsilon}) < +\infty$. This proves that \mathcal{H}_s is $2s$ -tame. \square

The following lemma gives an upper bound on the Hausdorff distance between the persistence diagrams of the functions h_s .

Lemma 4.10 *Let $f, g : \mathbb{X} \rightarrow \mathbb{R}$ and let h_s be defined as in lemma 4.9. For any $s \leq s' \in [0, 1]$ and for any $\alpha \geq 2s'\varepsilon$, we have*

$$d_H(D_\alpha h_s, D_\alpha h_{s'}) \leq (s' - s)\varepsilon$$

Proof. Since \mathcal{H}_s and $\mathcal{H}_{s'}$ are strongly $(s' - s)\varepsilon$ -interleaved the lemma follows from the theorem 4.6. \square

Let $a < b$ be two fixed real numbers and let $B_{a,b} = \{(u, v) \in \overline{\mathbb{R}}^2 : a \leq u \leq b\}$ be the vertical band between the vertical lines $\{u = a\}$ and $\{u = b\}$. Note that since $\|f - g\|_\infty < \varepsilon$, there exists $0 < \zeta < \varepsilon$ such that $\|f - g\|_\infty < \zeta$, so the functions h_s are indeed $2s\zeta$ -tame. As a consequence, it follows from the corollary 3.13 that, for any $s \in [0, 1]$ the number of points (counted with multiplicity) of $D_{2s\varepsilon} \mathcal{H}_s \cap B_{a,b} \cap \Delta_+^{2s\varepsilon}$ is finite. So,

$$\eta_s := \frac{1}{6} \min\{\|p - q\|_\infty : p, q \in D_{2s\varepsilon} \mathcal{H}_s \cap \Delta_+^{2s\varepsilon} \cap B_{a,b} \text{ and } p \neq q\} > 0 \quad (5)$$

A function $h_{s'}$ is said to be *very close* to h_s if $\|h_{s'} - h_s\|_\infty < \eta_s$ (or equivalently $|s' - s| < \eta_s/\varepsilon$).

The following lemma, which is an adaptation of the easy bijection lemma of [4] shows that for very close functions h_s and $h_{s'}$ there exist multibijections between their persistence diagrams that moves the points in $B_{a,b}$ by no more than $|s' - s|\varepsilon$.

Lemma 4.11 (Easy bijection Lemma) *Let $s, s' \in [0, 1]$ be such that $h_{s'}$ is very close to h_s . Then for any $\alpha \geq 2 \max(s, s')\varepsilon$ there exists a multibijection $\phi_{s,s'}$ between $D_\alpha \mathcal{H}_s$ and $D_\alpha \mathcal{H}_{s'}$ such that:*

- (i) for any $p \in D_\alpha \mathcal{H}_s \cap B_{a,b}$, $d^\infty(p, \phi_{s,s'}(p)) < |s' - s|\varepsilon$;
- (ii) for any $p \in D_\alpha \mathcal{H}_s$, $d^\infty(p, \phi_{s,s'}(p)) < 3|s' - s|\varepsilon$;

Proof. The proof which follows from the Box Lemma 4.5 is exactly the same as the proof of the Easy Bijection Lemma in [4]. \square

Now, to conclude the proof of the theorem 4.8, we proceed in two steps: first we compose the multibijections (given by the easy bijection lemma) between the persistence diagrams of the persistence modules \mathcal{H}_s that are very close and we use a compactness argument to conclude that we can join f and g by composition of a finite number of such multibijections. This will provide us with a multibijection between the persistence diagrams of f and g that moves the points in $B_{a,b}$ by no more than ε and the point in the complement by no more than 3ε . In a second step, we use that a and b have been chosen arbitrarily to build a multibijection that move all the points by no more than ε . In the first step, we have to overcome an additional issue because we want to obtain a bijection between the 0-persistence diagrams of f and g while the functions h_s do not (necessarily) admit 0-persistence diagrams for $s \in (0, 1)$.

Proposition 4.12 *There exists a multibijection $\Phi : D_0 \mathcal{H}_1 = D_0 g \rightarrow D_0 \mathcal{H}_0 = D_0 f$ such that:*

- (i) for any $p \in D_0 \mathcal{H}_1 \cap B_{a+\varepsilon, b-\varepsilon}$, $d^\infty(p, \Phi(p)) < \varepsilon$;
- (ii) for any $p \in D_0 \mathcal{H}_1$, $d^\infty(p, \Phi(p)) < 3\varepsilon$;

Proof. Let $\eta > 0$ be a fixed positive real number such that $\|f - g\|_\infty + \eta < \varepsilon$ and for any $s \in [0, 1]$, let $\tilde{\eta}_s = \min(\eta_s, \eta) > 0$ where η_s is defined by Eq. (5). The union for all $s \in [0, 1]$ of the intervals $I_s = (s - \tilde{\eta}_s/\varepsilon, s + \tilde{\eta}_s/\varepsilon)$ is an open covering of the segment $[0, 1]$ and for any $s', s'' \in I_s$, $h_{s'}$ and $h_{s''}$ are very close. By compactness of $[0, 1]$, there exists a finite sequence $s_0 = 1 > s_2 > \dots > s_n = 0$ such that for all $0 \leq i < n$, $h_{s_{i+1}}$ is very close to h_{s_i} (or h_{s_i} is very close to $h_{s_{i+1}}$). For each $0 \leq i < n$, we denote by ϕ_i a multibijection between $D_{2s_i\varepsilon} \mathcal{H}_{s_i}$ and $D_{2s_{i+1}\varepsilon} \mathcal{H}_{s_{i+1}}$ given by the lemma 4.11. We also denote by ψ_i a multibijection realising the bottleneck distance between $D_{2s_i\varepsilon} \mathcal{H}_{s_{i+1}}$ and $D_{2s_{i+1}\varepsilon} \mathcal{H}_{s_{i+1}}$.

Composing alternatively the multibijections ϕ_i and ψ_i we obtain a multibijection between the 2ε -persistence diagram of \mathcal{H}_1 and the 0-persistence diagram of \mathcal{H}_0 . More precisely, this bijection is defined as $\Phi = \psi_{n-1} \circ \phi_{n-1} \circ \dots \circ \psi_0 \circ \phi_0$. Applying directly lemma 3.11 and lemma 4.11 is not sufficient to conclude the proof of the proposition since this would produce a bound of 2ε . Nevertheless we claim that the bijection Φ does not move the points of $D_{2\varepsilon} \mathcal{H}_1 \cap B_{a+\varepsilon, b-\varepsilon}$ by more than ε . To see this we have to consider the ‘‘trajectories’’ of the points of $D_{2\varepsilon} \mathcal{H}_1 \cap B_{a+\varepsilon, b-\varepsilon}$ under the successive bijections ϕ_i and ψ_i . Let p be a point

(“counted with multiplicity”) of $D_{2\varepsilon}\mathcal{H}_1 \cap B_{a+\varepsilon, b-\varepsilon}$. If p is on the line $\Delta_{2\varepsilon}$ then $\phi_0(p) = p$ and ψ_0 just projects p to the line Δ_{2s_1} . Repeating this argument for all ϕ_i, ψ_i , it follows that p is just projected by Φ to the diagonal. Now assume that p is at distance greater than ε from the diagonal and consider the images of p under the successive multibijections $\phi_0, \psi_0, \phi_1 \cdots$ and let i_p the smallest index (if it exists) such that $\phi_{i_p} \circ \psi_{i_p-1} \circ \cdots \circ \phi_0(p)$ is contained in $\Delta_{2s_{i_p}}$. For all $i < i_p$ the point $\phi_i \circ \psi_{i-1} \circ \cdots \circ \phi_0(p)$ remains unchanged by the multibijection ψ_i . For all $i > i_p$, the point $\psi_{i-1} \circ \cdots \circ \phi_0(p)$ is contained in the line $\Delta_{2s_{i-1}}$ and thus remains unchanged by the multibijection ϕ_i . As consequence, the multibijection $\psi_i \circ \phi_i$ moves the point $\psi_{i-1} \circ \cdots \circ \phi_0(p)$ by at most $(s_{i-1} - s_i)\|f - g\|_\infty$ for all i except for $i = i_p$ where the point can be moved by $2(s_{i_p-1} - s_{i_p})\varepsilon < (s_{i_p-1} - s_{i_p})\|f - g\|_\infty + \eta$. It follows that the multibijection Φ does not move the points of $D_{2\varepsilon}\mathcal{H}_1$ by more than $\|f - g\|_\infty + \eta < \varepsilon$. By the same arguments, we obtain that if $p \in D_{2\varepsilon}\mathcal{H}_1 \setminus B_{a+\varepsilon, b-\varepsilon}$, then $\Phi(p)$ is at distance less than 3ε from p . Now we can extend the multibijection Φ to $D_0\mathcal{H}_1$ by projecting the points that are at distance less than ε from the diagonal (i.e. contained between $\Delta_{2\varepsilon}$ and Δ) onto the diagonal. As a consequence for any $p \in D_0\mathcal{H}_1 \cap B_{a+\varepsilon, b-\varepsilon}$, $d^\infty(p, \Phi(p)) < \varepsilon$ and for any $p \in D_0\mathcal{H}_1$, $d^\infty(p, \Phi(p)) < 3\varepsilon$. \square

Proof. of theorem 4.8

Let $(a_n)_{n \in \mathbb{N}}$ and $(b_n)_{n \in \mathbb{N}}$ be two sequences of real numbers such that:

- $a_0 < b_0 + 2\varepsilon$;
- $(a_n)_{n \in \mathbb{N}}$ is strictly decreasing and $\lim a_n = -\infty$;
- $(b_n)_{n \in \mathbb{N}}$ is strictly increasing and $\lim b_n = +\infty$.

We denote by $\Phi_n : D_0f \rightarrow D_0g$ the multibijection given by the proposition 4.12 for the band B_{a_n, b_n} . Let $\eta > 0$ be a fixed real number. Since $D_0g \cap B_{a_0, b_0} \cap \Delta_+^\eta$ contains a finite number (counted with multiplicities) of points, the set of the restrictions of the multibijections Φ_n to $D_0f \cap B_{a_0+\varepsilon, b_0-\varepsilon} \cap \Delta_+^{\eta+2\varepsilon}$ is finite. So taking a subsequence of (a_n, b_n) if necessary, we can assume without loss of generality that all the restrictions of Φ_n to $D_0f \cap B_{a_0+\varepsilon, b_0-\varepsilon} \cap \Delta_+^{\eta+2\varepsilon}$ are equal. By the same argument, taking a subsequence if necessary, we can assume that all the restrictions of Φ_n to $D_0f \cap B_{a_1+\varepsilon, b_1-\varepsilon} \cap \Delta_+^{\eta+2\varepsilon}$ are equal. We can iterate this process for all n and by a diagonal argument, we obtain a subsequence $(a_{k(n)}, b_{k(n)})$ of (a_n, b_n) such that the restrictions of the multibijections $\Phi_{k(n)}$ to $D_0f \cap \Delta_+^{\eta+2\varepsilon}$ converge to a multibijection Φ that moves the points by no more than ε . This multibijection can be extended to a multibijection between D_0f and D_0g by snapping onto the diagonal Δ the points of D_0f that are between Δ and $\Delta_{\eta+2\varepsilon}$ and the point of D_0g that are not in the image of Φ (note that these last points cannot be at distance more than ε from Δ). Since η can be chosen arbitrarily small, the bottleneck distance between D_0f and D_0g is upper bounded by ε . \square

5 Applications

In this section we briefly present two immediate applications of the stability theorems obtained in the previous section. More developed applications are presented in [2].

5.1 Relating Čech, Rips-Vietoris and witness complex filtrations

Given a finite point cloud L lying in an arbitrary metric space \mathbb{X} , Čech, Rips-Vietoris and witness complexes are filtered simplicial complexes built from the set L whose some parts of the persistence diagrams have been proven to be very useful to infer the topology of the space \mathbb{X} under some mild conditions (see for example [3] and the references there). Our stability result allows to prove that the whole persistence diagrams of these filtrations are closely related in any metric space.

Let L be a finite set of points lying in an arbitrary metric space \mathbb{X} and let α be a positive number. We call L^α the union of the open balls of radius α centered at the points of L : $L^\alpha = \bigcup_{x \in L} B(x, \alpha)$. We also denote by $\{L^\alpha\}$ the open cover of L^α formed by the open balls of radius α centered at the points of L . The *Čech complex* of L of parameter α , or $C_\alpha(L)$ for short, is the *nerve* of this cover, *i.e.* it is the abstract simplicial complex whose vertex set is L , and such that, for all $k \in \mathbb{N}$ and all $x_0, \dots, x_k \in L$, $[x_0, \dots, x_k]$ is a k -simplex of $C_\alpha(L)$ if and only if $B(x_0, \alpha) \cap \dots \cap B(x_k, \alpha) \neq \emptyset$. The *Rips complex* of L of parameter α , or $R_\alpha(L)$ for short, is the abstract simplicial complex whose k -simplices correspond to unordered $(k + 1)$ -tuples of points of L which are pairwise within distance α of one another.

Now assume that W is another (possibly infinite) subset of \mathbb{X} , identified as the witness set. Given a point $w \in W$ and a k -simplex σ with vertices in L , w is an α -*witness* of σ (or, equivalently, w α -*witnesses* σ) if the vertices of σ lie within distance $(d_k(w) + \alpha)$ of w , where $d_k(w)$ denotes the distance between w and its $(k + 1)$ th nearest landmark. The α -*witness complex* of L relative to W , or $W_\alpha(L)$ for short, is the maximum abstract simplicial complex, with vertices in L , whose faces are α -witnessed by points of W . When $\alpha = 0$, the α -witness complex coincides with the standard witness complex $W(L)$, introduced in [7].

It is known from [3] that Čech and Rips filtrations are strongly 1-interleaved on a logarithmic scale, namely:

$$\forall \alpha \in [0, 2), \forall n \in \mathbb{N}, C_{\alpha+2n}^{\log_2}(L) \subseteq R_{\alpha+2n+1}^{\log_2}(L) \subseteq C_{\alpha+2(n+1)}^{\log_2}(L), \quad (6)$$

where \mathcal{F}^{\log_2} stands for filtration \mathcal{F} on a logarithmic scale: $\forall \alpha \in \mathbb{R}, F_\alpha^{\log_2} = F_{2^\alpha}$. If the points of L are densely sampled from some compact set \mathbb{X} in \mathbb{R}^d , then, according to [3], the Čech and witness complex filtrations are also strongly 3-interleaved on a logarithmic scale:

$$\forall \alpha \in [0, 6), \forall n \in \mathbb{N}, C_{\alpha+6n}^{\log_2}(L) \subseteq W_{\alpha+6n+3}^{\log_2}(L) \subseteq C_{\alpha+6(n+1)}^{\log_2}(L). \quad (7)$$

Combining Eqs. (6) and (7) with Theorem 4.8⁷, we get the following:

Theorem 5.1 *Let L be a finite point cloud in an arbitrary metric space. Then, the bottleneck distance between $DC^{\log_2}(L)$ and $DR^{\log_2}(L)$ is at most 1. If in addition the points of L are densely sampled from a compact set \mathbb{X} in \mathbb{R}^d (see [3] for explicit conditions), then the bottleneck distance between $DC^{\log_2}(L)$ and $DW^{\log_2}(L)$ is at most 3.*

⁷remark that the above filtrations are indeed the sublevel sets filtrations of the function taking a constant value on each simplex equal to its time of appearance in the filtration

This theorem refines the results of [3] by stating that the whole persistence diagram of the Čech filtration is approximated by the Rips and witness complex filtrations, not only the initial part that is used for estimating the homology of the space \mathbb{X} underlying the data set L .

5.2 Building simplicial filtrations from noisy distance matrices

Existing work on Rips or witness complex filtrations generally assumes that distances between sample points can be computed exactly, which is rarely the case in practice. A typical example, inspired from applications in sensor networks, is when the actual locations of the sample points remain unknown and the geodesic distances between the samples in the underlying domain are approximated through their distances in a certain neighborhood graph. In such situations, noise appears in the entries of the distance matrix, and therefore a noisy variant of the actual simplicial filtration (whether Rips- or witness-based) is built. It is then important to ensure that the added noise does not significantly perturb the induced persistence diagram, which our main result (Theorem 4.8) does.

Specifically, given a symmetric $n \times n$ matrix M with positive coefficients, as well as a slightly perturbed variant \tilde{M} of this matrix, call ε the max norm $\|M - \tilde{M}\|_{\max}$. Then, for every set of indices $1 \leq i_0 < i_1 < \dots < i_m \leq n$, the respective times $t_M, t_{\tilde{M}}$ at which the simplex $\{i_0, \dots, i_m\}$ appears in the Rips filtration built from M and in the Rips filtration built from \tilde{M} satisfy $|t_M - t_{\tilde{M}}| \leq \varepsilon$. Indeed, by definition of the Rips complex we have $t_M = \max\{M_{i_k, i_l}, 0 \leq k < l \leq m\}$ and $t_{\tilde{M}} = \max\{\tilde{M}_{i_k, i_l}, 0 \leq k < l \leq m\}$, where M_{i_k, i_l} and \tilde{M}_{i_k, i_l} satisfy $|M_{i_k, i_l} - \tilde{M}_{i_k, i_l}| \leq \varepsilon$. Since this is true for every set of indices in the range $\{1, \dots, n\}$, the time-of-appearance functions $t_M, t_{\tilde{M}} : 2^{\{1, \dots, n\}} \rightarrow \mathbb{R}$ of the Rips filtrations satisfy $\|t_M - t_{\tilde{M}}\|_{\infty} \leq \varepsilon$. Hence, by Theorem 4.8, the bottleneck distance between the persistence diagrams of the Rips filtrations built from M and from \tilde{M} is at most 3ε . A similar argument applies to the case of the witness complex filtration.

Theorem 5.2 *Given two symmetric $n \times n$ matrices M and \tilde{M} with positive coefficients, the bottleneck distance between the persistence diagrams of the Rips (resp. witness complex) filtrations built respectively from M and from \tilde{M} is at most $3\|M - \tilde{M}\|_{\max}$.*

References

- [1] F. Chazal and D. Cohen-Steiner. *Geometric Inference*. submitted as a book chapter, 2007.
- [2] F. Chazal, L. J. Guibas, S. Y. Oudot, and P. Skraba. Analysis of scalar fields over point cloud data. INRIA research report - submitted, 2008.
- [3] F. Chazal and S. Y. Oudot. Towards persistence-based reconstruction in Euclidean spaces. In *Proc. 24th ACM Sympos. Comput. Geom.*, 2008.

- [4] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. In *Proc. 21st ACM Sympos. Comput. Geom.*, pages 263–271, 2005.
- [5] D. Cohen-Steiner, H. Edelsbrunner, J. Harer, and D. Morozov. Persistent homology for kernels and images. Preprint, 2008.
- [6] D. Cohen-Steiner, H. Edelsbrunner, and D. Morozov. Vines and vineyards by updating persistence in linear time. In *Proc. 22nd Sympos. on Comput. Geom.*, pages 119–126, 2006.
- [7] V. de Silva. A weak characterisation of the Delaunay triangulation. *Geometriae Dedicata*, 2008. to appear.
- [8] H. Edelsbrunner, J. Harer, and A. Zomorodian. Hierarchical Morse complexes for piecewise linear 2-manifolds. In *Proc. 17th Annu. ACM Sympos. Comput. Geom.*, pages 70–79, 2001.
- [9] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discrete Comput. Geom.*, 28:511–533, 2002.
- [10] P. Frosini. Discrete computation of size functions. *Journal of Combinatorics, Information, and System Science*, 17:232–250, 1992.
- [11] A. Hatcher. *Algebraic Topology*. Cambridge University Press, 2001.
- [12] V. Robins. Towards computing homology from approximations. *Topology*, 24:503–532, 1999.
- [13] A. Zomorodian and G. Carlsson. Computing persistent homology. *Discrete Comput. Geom.*, 33(2):249–274, 2005.



Centre de recherche INRIA Saclay – Île-de-France
Parc Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 Orsay Cedex (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
<http://www.inria.fr>
ISSN 0249-6399